Mariano Giaquinta
Giuseppe Modica

Mathematical Analysis

Foundations and Advanced Techniques for Functions of Several Variables

Birkhäuser

Mariano Giaquinta • Giuseppe Modica

Mathematical Analysis

Foundations and Advanced Techniques for Functions of Several Variables

Birkhäuser

Mariano Giaquinta
Scuola Normale Superiore
Piazza dei Cavalieri, 7
I-56100 Pisa
Italy
giaquinta@sns.it

Giuseppe Modica
Dipartimento di Sistemi e Informatica
Università di Firenze
Via S. Marta, 3
I-50139 Firenze
Italy
giuseppe.modica@unifi.it

ISBN 978-0-8176-8309-2 e-ISBN 978-0-8176-8310-8
DOI 10.1007/978-0-8176-8310-8
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011940806

Mathematics Subject Classification: 28-01, 35-01, 49-01, 52-01, 58A10

Birkhäuser Boston is part of Springer Science+Business Media (www.birkhauser.com)

Preface

This volume[1], that adds to the four volumes[2] that already appeared, complements the study of ideas and techniques of the differential and integral calculus for functions of several variables with the presentation of several specific topics of particular relevance from which the calculus of functions of several variables has originated and in which it has its most natural context. Some chapters have to be seen as introductory to further developments that proceed autonomously and that cannot be treated here because of space and complexity. However, we believe that a discussion at an elementary level of some aspects is surely part of a basic mathematical education and helps to understand the context in which the study of abstract functions of many variables finds its true motivation.

Chapter 1 aims at illustrating in concrete situations the abstract treatment of the geometry of Hilbert spaces that we presented in [GM3]. After a short illustration of Lebesgue's spaces, in particular of L^2, and a brief introduction to Sobolev spaces, we present some complements to the theory of Fourier series, the method of separation of variables for the Laplace, heat and wave equations, and the Dirichlet principle and we conclude with some results concerning the Sturm–Liouville theory. Chapter 2 is dedicated to the theory of convex functions and to illustrating several instances in which it naturally shows. Among these, the study of inequalities, the Farkas lemma, and the linear and the convex programming with the theorem of Kuhn–Tucker and of von Neumann and Nash in the theory of games. Chapter 3 is an introduction to calculus of variations. Our aim is of just presenting the Lagrangian and Hamiltonian formalism, hinting at some of its connections with geometrical optics, mechanics, and some geometrical examples. Chapter 4 deals with the general theory of differential

[1] This book is a translation and a revised edition of M. Giaquinta, G. Modica *Analisi Matematica, V. Funzioni di più variabili: ulteriori sviluppi*, Pitagora Ed., Bologna, 2005.

[2] M. Giaquinta, G. Modica, *Mathematical Analysis, Functions of One Variable*, Birkhauser, Boston, 2003,
M. Giaquinta, G. Modica, *Mathematical Analysis, Approximation and Discrete Processes*, Birkhauser, Boston, 2004,
M. Giaquinta, G. Modica, *Mathematical Analysis, Linear and Metric Structures and Continuity*, Birkhauser, Boston, 2007,
M. Giaquinta, G. Modica, *Mathematical Analysis, An Introduction to Functions of several variables*, Birkhauser, Boston, 2009.
We shall refer to these books as to [GM1], [GM2], [GM3] and [GM4], respectively.

forms with the Stokes theorem, the Poincaré lemma, and some applications of geometrical character. The final two chapters, 5 and 6, are dedicated to the general theory of measure and integration, only outlined in [GM4], and includes the study of Borel, Radon and Hausdorff measures and of the theory of derivation of measures.

The study of this volume requires a strong effort compared to the one requested for the first four volumes, both for the intrinsinc difficulties and for the width and varieties of the topics that appear. On the other hand, we believe that it is very useful for the reader to have a wide spectrum of contexts in which the ideas have developed and play an important role and some reasons for an analysis of the formal and structural foundations that at first sight might appear excessive. However, we have tried to keep a simple style of presentation, always providing examples, enlightening remarks and exercises at the end of each chapter. The illustrations and the bibliographical note provide suggestions for further readings.

We are greatly indebted to Cecilia Conti for her help in polishing our first draft and we warmly thank her. We would also like to thank Paolo Acquistapace, Timoteo Carletti, Giulio Ciraolo, Roberto Conti, Giovanni Cupini, Matteo Focardi, Pietro Majer and Stefano Marmi for their comments and their invaluable help in catching errors and misprints, and Stefan Hildebrandt for his comments and suggestions concerning especially the choice of illustrations. Our special thanks also go to all members of the editorial technical staff of Birkhäuser for the excellent quality of their work and especially to Katherine Ghezzi and the executive editor Ann Kostant.

Note: We have tried to avoid misprints and errors. However, as most authors, we are imperfect. We will be very grateful to anybody who is willing to point out errors or misprints or wants to express criticism or comments. Our e-mail addresses are

giaquinta@sns.it modica@dma.unifi.it

We will try to keep up an errata corrige at the following webpages:

http://www.sns.it/~giaquinta

http://www.dma.unifi.it/~modica

http://www.dsi.unifi.it/~modica

Mariano Giaquinta
Giuseppe Modica
Pisa and Firenze
March 2011

Contents

1. Spaces of Summable Functions and Partial Differential Equations

This chapter aims at substantiating the abstract theory of Hilbert spaces developed in [GM3]. After introducing the Laplace, heat and wave equations we present the classical method of *separation of variables* in the study of partial differential equations. Then we introduce *Lebesgue's spaces* of p-summable functions and we continue with some elements of the theory of *Sobolev spaces*. Finally, we present some basic facts concerning the notion of *weak solution*, the *Dirichlet principle* and the *alternative theorem*.

1.1 Fourier Series and Partial Differential Equations

1.1.1 The Laplace, Heat and Wave Equations

In our previous volumes [GM2, GM3, GM4] we discussed time by time *partial differential equations*, i.e., equations involving functions of several variables and some of their partial derivatives.

Among linear equations, i.e., equations for which the superposition principle holds, the following equations are particularly relevant, for instance, in classical physics: the *Laplace equation*, the *heat equation* and the *wave equation*. They are respectively the prototypes of the so-called *elliptic*, *parabolic* and *hyperbolic* partial differential equations.

a. Laplace's and Poisson's equation

Laplace's equation for a function $u : \Omega \to \mathbb{R}$ defined on an open set $\Omega \subset \mathbb{R}^n$, $n \geq 2$, is

$$\Delta u := \operatorname{div} \nabla u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial {x^i}^2} = \sum_{i=1}^{n} u_{x^i x^i} = 0.$$

The operator Δ is called *Laplace's operator* and the solutions of $\Delta u = 0$ are called *harmonic functions*.

M. Giaquinta and G. Modica, *Mathematical Analysis, Foundations and Advanced Techniques for Functions of Several Variables*, DOI 10.1007/978-0-8176-8310-8_1,

Several "equilibrium" situations reduce or can be reduced to Laplace's equation. For instance, a system is often subject to "internal forces" represented by a field $E : \Omega \to \mathbb{R}^n$, and, at the equilibrium, the outgoing flux from each domain is zero, i.e.,

$$\int_{\partial A} E \bullet \nu_A \, d\mathcal{H}^{n-1} = 0 \qquad \forall A \subset\subset \Omega.$$

If E is smooth, $E \in C^1(\Omega)$, we may use the Gauss–Green formulas, see e.g., [GM4], to deduce

$$0 = \int_{\partial B(x,r)} E \bullet \nu_{B(x,r)} \, dx = \int_{B(x,r)} \operatorname{div} E(y) \, dy$$

for every ball $B(x,r) \subset\subset \Omega$, and, letting $r \to 0$, conclude that

$$\operatorname{div} E(x) = 0 \qquad \forall x \in \Omega, \tag{1.1}$$

on account of the integral mean theorem. Often the field E has a potential $u : \Omega \to \mathbb{R}$, $E = -\nabla u$. In this case the potential u solves *Laplace's equation*

$$\Delta u(x) = \operatorname{div} \nabla u(x) = 0 \qquad \forall x \in \Omega. \tag{1.2}$$

In mathematical physics, quantities are often functions of densities $f : \Omega \to \mathbb{R}$ (so that $\int_A f(x) \, dx$ is the quantity related to $A \subset \Omega$) that are related with a force field $E : \Omega \to \mathbb{R}^n$. For instance, in electrostatics $f(x)$ is the density of charge and $E(x)$ is the induced electric field at $x \in \Omega$. The interaction is then expressed as proportionality of the quantities

$$\int_A f(x) \, dx \qquad \text{and} \qquad \int_{\partial A} E \bullet \nu_A \, d\mathcal{H}^{n-1}$$

for every subset $A \subset\subset \Omega$. Assuming that $f \in C^0(\Omega)$, $E \in C^1(\Omega)$ and the constant of proportionality equals 1, as previously, Gauss–Green formulas yield

$$\int_{B(x,r)} f(y) \, dy = \int_{\partial B(x,r)} E \bullet \nu_{B(x,r)} \, dx = \int_{B(x,r)} \operatorname{div} E(y) \, dy$$

for every ball $B(x,r) \subset\subset \Omega$, hence, letting $r \to 0$,

$$\operatorname{div} E(x) = f(x) \qquad \forall x \in \Omega. \tag{1.3}$$

If E has a potential, $E = -\nabla u$, then (1.3) reads as *Poisson's equation*

$$-\Delta u(x) = f(x) \qquad \forall x \in \Omega. \tag{1.4}$$

We have seen in [GM4] that for harmonic functions $u : \Omega \to \mathbb{R}$ of class $C^2(\Omega) \cap C^0(\overline{\Omega})$ the following *maximum principle* holds:

$$\sup_{\Omega} |u| \le \sup_{\partial\Omega} |u|.$$

A consequence is a *uniqueness result* for the the so-called *Dirichlet problem*.

1.1 Proposition (Uniqueness). *Dirichlet's problem for Poisson's equation, i.e., the problem of finding $u : \Omega \to \mathbb{R}$ satisfying*

$$\begin{cases} \Delta u = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega, \end{cases} \tag{1.5}$$

has at most a solution of class $C^2(\Omega) \cap C^0(\overline{\Omega})$.

Proof. In fact, the difference u of two solutions of (1.5) satisfies

$$\begin{cases} \Delta u = 0 & \text{in } \Omega, \\ u = 0 & \text{su } \partial\Omega, \end{cases} \tag{1.6}$$

hence $\sup_\Omega |u| \le \sup_{\partial\Omega} |u| = 0$ by the maximum principle.

Alternatively, one can use the so-called *energy method*, for instance if the difference u of two solutions of (1.5) is of class $C^2(\overline{\Omega})$. In fact, if $u \in C^2(\overline{\Omega})$ is a solution of (1.6), we have $u\Delta u = 0$ in Ω, and, integrating by parts, we get

$$\begin{aligned} 0 = \int_\Omega u\Delta u\, dx &= \int_\Omega \sum_{i=1}^n D_i(uD_iu)\, dx - \int_\Omega |Du|^2\, dx \\ &= \int_{\partial\Omega} u\frac{\partial u}{\partial \nu}\, d\sigma - \int_\Omega |Du|^2\, dx = -\int_\Omega |Du|^2\, dx, \end{aligned}$$

hence $Du = 0$ in Ω and, consequently, $u = 0$ in $\overline{\Omega}$ since $u = 0$ on $\partial\Omega$. □

b. The heat equation

The *heat equation* for $u = u(x,t)$, $x \in \Omega \subset \mathbb{R}^n$, $t \in \mathbb{R}$, is

$$u_t - \Delta u = 0.$$

It is also known as the *diffusion equation*, and it is supposed to describe the time evolution of a quantity such as the temperature or the density of a population under suitable *viscosity conditions*.

Let $u(x,t) : \Omega \times \mathbb{R} \to \mathbb{R}$ be a function and let $F(x,t) : \Omega \times \mathbb{R} \to \mathbb{R}^n$ be a field. It often happens that the time variation of u in $A \subset\subset \Omega$ is balanced by the outgoing flux of F through ∂A,

$$\frac{\partial}{\partial t}\int_A u(x,t)\, dx = -\int_{\partial A} F \bullet \nu_A\, ds \qquad \forall A \subset\subset \Omega,\ \forall t.$$

Assuming u and F sufficiently smooth (for instance, u continuous in x for all t and C^1 in t for all x, and $F(x,t)$ of class C^1 in x for all t), Gauss–Green formulas and the theorem of integration under the integral sign allow us to conclude

$$\begin{aligned} \int_{B(x,r)} \frac{\partial u}{\partial t}(x,t)\, dx &= \frac{\partial}{\partial t}\int_{B(x,r)} u(x,t)\, dx \\ &= -\int_{\partial B(x,r)} F \bullet \nu_{B(x,r)}\, ds = -\int_{B(x,r)} \operatorname{div} F(x,t)\, dx \end{aligned}$$

for all $B(x,r) \subset \Omega$ and $\forall t$. Letting $r \to 0$, we deduce the so-called *continuity equation* or *balance equation*

$$\frac{\partial u}{\partial t}(x,t) + \operatorname{div} F(x,t) = 0 \qquad \text{in } \Omega \times \mathbb{R}. \tag{1.7}$$

The physical characteristics of the system are now expressed by adding to (1.7) a *constitutive equation* that relates the field F to u,

$$F = F[u]. \tag{1.8}$$

In the simplest case, one assumes that $F(x,t)$ is proportional to the spatial gradient of u at the same instant, $F(x,t) = -k\nabla u(x,t)$. The internal forces tend to *diffuse* u if $k > 0$ and to *concentrate* u if $k < 0$ (if $u(x,t)$ represents the temperature at point x in the body Ω at instant t, we have diffusion). For simplicity, if $k = 1$, the constitutive equation is $F(x,t) = -\nabla u(x,t)$, and from (1.7) and (1.8) we infer the *heat equation* for u:

$$u_t = \operatorname{div} \nabla u = \Delta u \qquad \text{in } \Omega \times \mathbb{R}.$$

1.2 Parabolic equations. The model, continuity equation plus constitutive law (1.7) and (1.8), is sufficiently flexible to be adapted to several situations. For instance, the variation in time of u may be caused by the field F but also by a volume effect determined by a density $f(x,t)$. The equation becomes then

$$\int_A \frac{\partial u}{\partial t}(x,t)\,dx = -\int_A \operatorname{div} F\,dx + \int_A f(x,t)\,dx \qquad \forall A \subset\subset \Omega,$$

that, assuming sufficient regularity for u, F and f, can be written as

$$u_t = -\operatorname{div} F + f \qquad \text{in } \Omega \times \mathbb{R}.$$

Additionally, the field F may take into account external effects. For instance, we may add a privileged direction

$$F(x,t) = -\nabla u(x,t) + g(x,t), \qquad g : \Omega \to \mathbb{R}^n,$$

or some intrinsic nonhomogeneity of the system (even in time)

$$F(x,t) = -k(x,t)\nabla u(x,t),$$

or some anisotropy

$$F = (F_i), \qquad F_i(x,t) = -\sum_{j=1}^{n} a_{ij}(x,t) D_j u(x,t),$$

or a dependence on u,

$$F(x,t) = -k\nabla u(x,t) + c(x,t)\,u(x,t),$$

or imagine that all these effects act at the same time.

$$F = (F_i), \qquad F_i = \sum_{j=1}^{n} a_{ij} D_j u + b_i u + g_i.$$

If all quantities are sufficiently regular, we end up with the *parabolic equation*

$$u_t = \sum_{ij=1}^{n} D_i(a_{ij} D_j u) - \sum_{i=1}^{n} D_i(b_i u) - \operatorname{div} g + f \qquad \text{in } \Omega \times \mathbb{R}.$$

A maximum principle holds also for parabolic equations.

1.3 ¶ Maximum principle for the heat equation. Prove the following parabolic maximum principle: Let $u = u(x,t)$ be a solution of $u_t - \Delta u = 0$ in $\Omega\times]0,T[$ of class $C^2(\Omega\times]0,T[) \cap C^0(\overline{\Omega} \times [0,T[)$. Then

$$\sup_{\Omega\times[0,T[} |u| \le \sup_{\Gamma} |u|$$

where $\Gamma := (\Omega \times \{0\}) \cup (\partial\Omega \times [0,T[)$. More precisely, show that maximum and minimum points of u lie on the base or on the lateral walls of the cylinder $\Omega \times [0,T[$: For instance, if u denotes the temperature of a body Ω, the maximum principle tells us that $u(x,t)$ cannot be higher than the initial temperature of the body or of the temperature that we apply to the walls.

Also on the basis of Exercise 1.3, it is natural to consider the following problem in which initial and boundary values are prescribed: Given f, g and h, find a function $u(x,t)$ such that

$$\begin{cases} u_t - \Delta u = f & \text{in } \Omega\times]0,T[, \\ u(x,0) = g(x) & \forall x \in \Omega, \\ u(x,t) = h(x,t) & \forall x \in \partial\Omega,\ \forall t \in]0,T[. \end{cases} \tag{1.9}$$

We then have the following uniqueness for the parabolic problem.

1.4 Proposition (Uniqueness). *Problem* (1.9) *has at most a solution of class* $C^2(\Omega\times]0,T[) \cap C^0(\overline{\Omega} \times [0,T[)$.

Proof. In fact, since the *difference* u between two solutions of (1.9) satisfies

$$\begin{cases} u_t - \Delta u = 0 & \text{in } \Omega\times]0,T[, \\ u(x,0) = 0 & \forall x \in \Omega, \\ u(x,t) = 0 & \forall x \in \partial\Omega,\ \forall t \in]0,T[, \end{cases} \tag{1.10}$$

the maximum principle for the heat equation implies $u = 0$ on $\Omega \times [0,T[$.

Alternatively, we may get the result using the *energy method*, at least for sufficiently regular solutions in $\Omega \times [0,T]$. In fact, if u denotes the difference between two solutions, and $u \in C^2(\overline{\Omega}\times]0,T[) \cap C^0(\overline{\Omega} \times [0,T[)$, then u satisfies (1.10). Thus, multiplying (1.10) by u and integrating, we obtain

$$0 = \int_0^T \int_\Omega u(u_t - \Delta u)\, dx\, dt$$
$$= \int_0^T dt \int_\Omega \frac{d}{dt}\Big(\frac{|u|^2}{2}\Big) - \int_0^T dt \int_{\partial\Omega} u \frac{du}{d\nu}\, d\sigma + \int_0^T dt \int_\Omega |Du|^2\, dx,$$

and, using the initial and boundary conditions, this reduces to

$$\frac{1}{2}\int_\Omega |u(x,T)|^2\, dx + \int_0^T \int_\Omega |Du|^2\, dx = 0,$$

i.e., $u = 0$ in $\Omega \times [0, T[$. □

c. The wave equation

The *wave equation* is

$$\Box u := u_{tt} - \Delta u = 0. \tag{1.11}$$

The operator $\Box$ is called the *operator of D'Alembert*. If $u(x,t)$ represents the deviation on a direction of a vibrating string or a membrane at point x and time t and if the "force" acting on a piece A of the membrane is given by

$$-\int_{\partial A} F \bullet \nu_A \, d\mathcal{H}^{n-1},$$

according to Newton's law, we deduce

$$\frac{d^2}{dt^2}\int_A u(x)\, dx = -\int_{\partial A} F \bullet \nu_A \, d\mathcal{H}^{n-1}$$

for all $A \subset\subset \Omega$. Assuming that the constitutive law is

$$F = -\nabla u$$

and that u is sufficiently smooth, as previously, using differentiation under the integral sign, Gauss–Green formulas and the integral mean theorem, we deduce the *wave equation* for u:

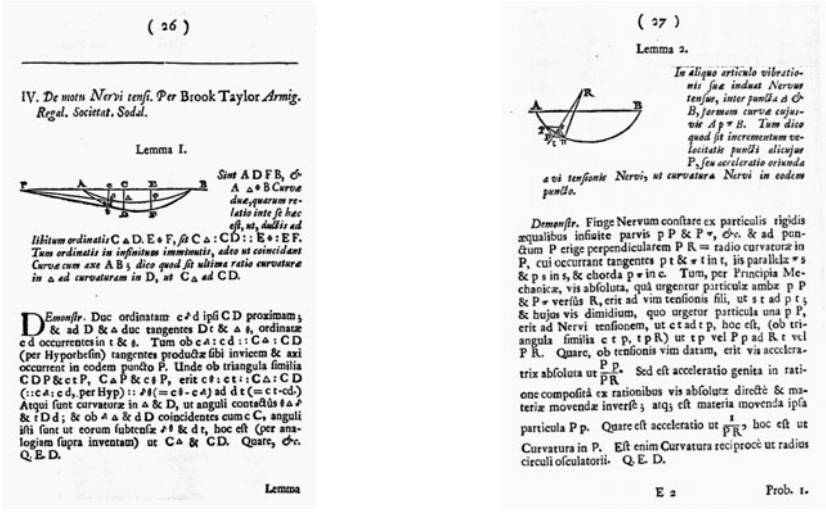
$$u_{tt} = \operatorname{div} \nabla u = \Delta u \qquad \text{in } \Omega.$$

Given f, g_0, g_1 and h, we consider the initial value problem for the wave equation which consists in finding u sufficiently regular so that

$$\begin{cases} u_{tt} - \Delta u = f & \text{in } \Omega \times [0, T[, \\ u(x,0) = g_0(x),\ u_t(x,0) = g_1(x), & \forall x \in \Omega, \\ u(x,t) = h(x,t), & \forall x \in \partial\Omega,\ \forall t \in [0, T[\end{cases} \tag{1.12}$$

and we prove the following uniqueness result.

1.5 Proposition (Uniqueness). *The initial value problem* (1.12) *has at most one solution.*



(26)

IV. *De motu Nervi tensi.* Per Brook Taylor *Armig. Regal. Societat. Sodal.*

Lemma I.

Sint A D F B, & A △ ✱ B *Curvæ duæ, quarum relatio inter ſe hac eſt, ut, ductis ad libitum ordinatis* C △ D. E ✱ F, *ſit* C △ : C D :: E ✱ : E F. *Tum ordinatis in infinitum imminutis, adeo ut coincidant Curvæ cum axe* A B ; *dico quod ſit ultima ratio curvaturæ in* △ *ad curvaturam in* D, *ut* C △ *ad* C D.

Demonſtr. Duc ordinatam c ✱ d ipſi C D proximam ; & ad D & △ duc tangentes D t & △ ϑ, ordinatæ c d occurrentes in t & ϑ. Tum ob c △ : c d :: C △ : C D (per Hypotheſin) tangentes producta ſibi invicem & axi occurrent in eodem puncto P. Unde ob triangula ſimilia C D P & c t P, C △ P & c ϑ P, erit c ϑ : c t :: C △ : C D (:: c △ : c d, per Hyp.) :: ✱ (= c ϑ - c ✱) ad d t (= c t - c d.) Atqui ſunt curvaturæ in ✱ & D, ut anguli contactûs ϑ ✱ & t D d ; & ob ✱ & △ & d D coincidentes cum c C, anguli iſti ſunt ut eorum ſubtenſæ ϑ ✱ & d t, hoc eſt (per analogiam ſupra inventam) ut C △ & C D. Quare, &c. Q. E. D.

Lemma

(27)

Lemma 2.

In aliquo articulo vibrationis ſuæ induat Nervus tenſuræ, inter puncta A *&* B, *fermem curvæ conjungens* A ϑ ✱ B. *Tum dico quod ſit incrementum velocitatis puncti alicujus* P, *ſeu acceleratio oriunda a vi tenſionis Nervi, ut curvatura Nervi in eodem puncto.*

Demonſtr. Finge Nervum conſtare ex particulis rigidis æqualibus infinite parvis p P & P ✱, &c. & ad punctum P erige perpendicularem P R, = radio curvaturæ in P, cui occurrant tangentes p t & ✱ t in t, iis parallelæ ✱ s & p s in s, & chorda p ✱ in c. Tum, per Principia Mechanicæ, vis abſoluta, quâ urgentur particulæ ambæ p P & P ✱ verſûs R, erit ad vim tenſionis fili, ut s t ad p t ; & hujus vis dimidium, quo urgetur particula una p P, erit ad Nervi tenſionem, ut c t ad t p, hoc eſt, (ob triangula ſimilia c t p, t p R) ut t p vel P p ad R t vel P R. Quare, ob tenſionis vim datam, erit vis acceleratrix abſoluta ut $\frac{P\,p}{P\,R}$. Sed eſt acceleratio genita in ratione compoſitâ ex rationibus vis abſolutæ directè & materiæ movendæ inverſè ; atq; eſt materia movenda ipſa particula P p. Quare eſt acceleratio ut $\frac{1}{P\,R}$, hoc eſt ut Curvatura in P. Eſt enim Curvatura reciprocè ut radius circuli oſculatorii. Q. E. D.

E 2 Prob. 1.

Figure 1.1. Two pages of *De Motu Nervi Tensi* by Brook Taylor (1685–1731) from the Philosophical Transactions, 1713.

Proof. We proved the claim in [GM3] if $\Omega = [a, b]$. In the general case, we use the so-called *energy method*. The difference $u(x,t)$ of two solutions of (1.12) satisfies

$$\begin{cases} u_{tt} - \Delta u = 0 & \text{in } \Omega \times [0, T[, \\ u(x,0) = 0,\ u_t(x,0) = 0 & \forall x \in \Omega, \\ u(x,t) = 0 & \forall x \in \partial\Omega,\ \forall t \in [0, T[. \end{cases} \tag{1.13}$$

Multiplying by u_t and integrating in t and x, we find for all $\tau \in [0, T]$

$$\begin{aligned} 0 &= \int_0^\tau dt \int_\Omega u_{tt} u_t \, dx - \int_0^\tau dt \int_{\partial\Omega} u_t \frac{du}{d\nu} \, d\sigma + \int_0^\tau dt \int_\Omega \frac{d}{dt} \frac{|u_x|^2}{2} \, dx \\ &= \frac{1}{2} \int_\Omega \Big(|u_t(x,\tau)|^2 + |u_x(x,\tau)|^2 \Big) \, dt, \end{aligned}$$

which yields $u = 0$ in $\Omega \times [0, T[$. □

However, how can we find solutions (or even prove that there exist solutions) of the previous boundary and initial problems for the Poisson, heat and wave equations? This is part of the *theory of partial differential equations* which, of course, we are not going to get into. However, in the next subsection we shall describe a method that, in some cases and in the presence of a simple geometry of the domain Ω, allows us to find solutions.

1.1.2 The method of separation of variables

In this subsection we shall illustrate how to get solutions of the previous partial differential equation (PDE) in some simple cases, without aiming at generality and systematization.

a. Laplace's equation in a rectangle

We consider Laplace's equation in a rectangle of $\mathbb{R}^2$ with boundary value g. First we notice that it suffices to solve the Dirichlet problem when g is nonzero only on one of the sides of the rectangle. In fact, by superposition we are then able to find a solution $u_0(x, y)$ of the Dirichlet problem for the Laplace equation on a rectangle when the boundary datum vanishes at the vertices of the rectangle. For an arbitrary datum g, it suffices then to choose α, β, γ and δ in such a way that $g_0 := g - \alpha - \beta x - \gamma y - \delta xy$ vanishes at the four vertices of the rectangle and, if u_0 is a solution with boundary value g_0, then

$$u(x, y) := u_0(x, y) + (\alpha + \beta x + \gamma y + \delta xy)$$

solves our original problem with boundary value g.

Therefore, let us consider the problem of finding a solution $u(x, y)$ of

$$\begin{cases} u_{xx} + u_{yy} = 0 & \text{in }]0, \pi[\times]0, a[, \\ u(0, y) = u(\pi, y) = 0 & \forall y \in [0, a], \\ u(x, a) = 0 & \forall x \in [0, \pi], \\ u(x, 0) = g(x) & \forall x \in [0, \pi]. \end{cases} \tag{1.14}$$

We shall use the so-called *method of separation of variables*.

Our first step is to look for nonzero solutions $u(x, y)$ of the problem

$$\begin{cases} u_{xx} + u_{yy} = 0 & \text{in }]0, \pi[\times]0, a[, \\ u(0, y) = u(\pi, y) = 0 & \forall y \in [0, a], \\ u(x, a) = 0 & \forall x \in [0, \pi] \end{cases} \tag{1.15}$$

of the type

$$u(x, y) = X(x)Y(y). \tag{1.16}$$

It is easily seen that such solutions exist if there is a constant $\lambda \in \mathbb{R}$ for which there exist nonzero solutions X and Y of

$$\begin{cases} X'' + \lambda X = 0, \\ X(0) = X(\pi) = 0, \end{cases} \quad \text{and} \quad \begin{cases} Y'' - \lambda Y = 0, \\ Y(a) = 0. \end{cases} \tag{1.17}$$

Let us look for solutions $X(x)$ of the boundary value problem

$$\begin{cases} X'' + \lambda X = 0, \\ X(0) = X(\pi) = 0. \end{cases} \tag{1.18}$$

If $\lambda < 0$, there are no solutions. In fact, the equation and the condition $X(0) = 0$ imply that $X(x)$ is a multiple of $\sinh(\sqrt{-\lambda}x)$, and among these functions, only $X = 0$ vanishes at $x = \pi$ because $\sinh(\sqrt{-\lambda}\pi) \neq 0$. If $\lambda = 0$, the unique solution of the problem is clearly $X = 0$; hence there are

no nonzero solutions. If $\lambda > 0$, the equation and the condition $X(0) = 0$ imply that $X(x)$ is a multiple of $\sin(\sqrt{\lambda}x)$. Therefore, there exist solutions of (1.18) if and only if $\sin(\sqrt{\lambda}\pi) = 0$. In conclusion, (1.18) has nonzero solutions if and only if

$$\lambda = n^2, \qquad n = \pm 1, \pm 2, \dots$$

and, for every n, the solutions of

$$\begin{cases} X'' + n^2 X = 0, \\ X(0) = X(\pi) = 0 \end{cases}$$

are exactly the multiples of

$$X_n(x) := \sin(nx).$$

Having found the sequence of λ's that produce nonzero solutions of the first problem in (1.17), let us look for solutions of

$$\begin{cases} Y'' - n^2 Y = 0, \\ Y(a) = 0. \end{cases}$$

For each n, these are multiples of $\sinh(n(a-y))$.

Returning to problem (1.15), for all $n \geq 1$ the functions

$$X_n(x)Y_n(y) = \sin(nx)\sinh(n(a-y)), \qquad x \in [0,\pi],\ y \in [0,a],$$

solve (1.15) and, because of the superposition principle, for every $N \geq 1$ and for any choice of constants $c_1, c_2, \dots, c_N$,

$$u_N(x,y) := \sum_{n=1}^{N} c_n \sin(nx)\sinh(n(a-y))$$

is again a solution of (1.15). Therefore, if $\{c_n\}$ is a sequence of real numbers for which the series

$$u(x,y) := \sum_{n=1}^{\infty} c_n \sin(nx)\sinh(n(a-y)) \tag{1.19}$$

converges uniformly together with its first and second derivatives on the compact sets of $]0,\pi[\times]0,a[$, then

$$D^2\Big(\sum_{n=1}^{\infty} \dots\Big) = \sum_{n=1}^{\infty} D^2\Big(\dots\Big)$$

and the function $u(x,y)$ in (1.19) solves (1.15). This concludes the first step in which we have found a family of solutions, the functions in (1.19), of (1.15).

The second step consists now in selecting from this family the solution of (1.14). In order to do this, we need some regularity on the boundary datum g.

Let $g(x) : [0,1] \to \mathbb{R}$ be of class $C^{0,\alpha}([0,\pi])$, i.e., let us assume that there exists a constant $C > 0$ such that

$$|g(x+t) - g(x)| \le C\, t^\alpha \qquad \forall x, x+t \in [0,\pi], \tag{1.20}$$

and let $g(0) = g(\pi) = 0$. Denote still by g its odd extension to $[-\pi,\pi]$. It follows from Dini's criterium for Fourier series, see e.g., [GM3], that g has an expansion in Fourier series of sines that converges pointwise to $g(x)$ for every $x \in [0,\pi]$,

$$g(x) = \sum_{n=1}^{\infty} b_n \sin(nx), \qquad b_n := \frac{2}{\pi}\int_0^\pi g(x)\sin(nx)\,dx.$$

Trivially,

$$|b_n| \le \frac{2}{\pi}\int_0^\pi |g(x)|\,dx \qquad \forall n$$

and, from (1.20), we infer that the convergence of the Fourier series of g is uniform in $[0,\pi]$, see e.g., [GM3].

1.6 Theorem. *The function*

$$u(x,y) = \sum_{n=1}^{\infty} b_n \frac{\sinh n(a-y)}{\sinh na}\sin(nx) \tag{1.21}$$

is of class $C^\infty(]0,\pi[\times]0,a[)$, *continuous in* $C^0([0,\pi]\times[0,a])$, *harmonic and solves* (1.14).

Proof. Since $\{b_n\}$ is bounded and

$$\frac{\sinh n(a-y)}{\sinh na} = \frac{e^{n(a-y)} - e^{-n(a-y)}}{e^{na} - e^{-na}} = e^{-ny}\frac{1 - e^{-2n(a-y)}}{1 - e^{-2na}} \le \frac{e^{-ny}}{1 - e^{2a}},$$

we infer that the series (1.21) is totally (hence uniformly) convergent together with the series of its derivatives of any order in $[0,\pi]\times[\overline{y},a]$ for all $\overline{y} > 0$. It follows that u is of class $C^\infty(]0,\pi[\times]0,a[)$ and harmonic in $(]0,\pi[\times]0,a[)$.

Writing

$$s_N(x,y) := \sum_{n=1}^{N} b_n \frac{\sinh n(a-y)}{\sinh na}\sin nx,$$

we have $s_N(x,0) = \sum_{n=1}^N b_n \sin nx = S_N(g)(x)$. Since the Fourier series of g converges uniformly to g, we infer for all $\epsilon > 0$

$$|s_M(x,0) - s_N(x,0)| < \epsilon \qquad \text{for some } N, M \ge N_\epsilon.$$

Trivially,

$$s_M(x,y) - s_N(x,y) = 0 \qquad \text{if } (x,y) = (0,y) \text{ or } (0,\pi) \text{ or } (x,a)$$

and $s_N(x,y) - s_M(x,y)$ is harmonic in $]0,\pi[\times]0,a[$ and continuous in $[0,\pi]\times[0,a]$. From the maximum principle it follows that

$$|s_M(x,y) - s_N(x,y)| < \epsilon \qquad \text{in } [0,\pi]\times[0,a] \qquad \text{for } N, M \ge N_\epsilon.$$

In conclusion, the series (1.21) converges uniformly in $[0,\pi]\times[0,a]$. It follows that $u(x,y) \in C^0([0,\pi]\times[0,a])$ and $u(x,0) = g(x)\ \forall x \in [0,\pi]$. □

b. Laplace's equation on a disk

The Dirichlet problem for Laplace's equation on the unit disk writes, see e.g., [GM4], as

$$\begin{cases} u_{rr} + \frac{1}{r}u_r + \frac{1}{r^2}u_{\theta\theta} = 0 & \text{in }]0,1[\times[0,2\pi], \\ u(1,\theta) = f(\theta), & \forall\theta \in [0,2\pi[. \end{cases} \tag{1.22}$$

By applying the method of separation of variables, we begin by seeking nonzero solutions of Laplace's equations in the disk of the form $u(r,\theta) = R(r)\Theta(\theta)$, finding for R and Θ

$$\frac{r^2R''}{R} + \frac{rR'}{R} = -\frac{\Theta''}{\Theta}.$$

Therefore, there exist nonzero solutions of Laplace's equation in the disk of the form $u(r,\theta) = R(r)\Theta(\theta)$ if and only if there is $\lambda \in \mathbb{R}$ for which the two problems

$$\begin{cases} \Theta'' + \lambda\Theta = 0, \\ \Theta \ 2\pi\text{-periodic}, \end{cases} \qquad \text{and} \qquad \begin{cases} r^2R'' + rR' = \lambda R, \ 0 \le r \le 1, \\ R(0) \in \mathbb{R} \end{cases}$$

have solutions. The first equation, $\Theta'' + \lambda\Theta = 0$, has nontrivial 2π-periodic solutions if and only if $\lambda = n^2$, $n = 0, \pm1, \pm2, \ldots$. Moreover, the solutions are the constants for $\lambda = 0$ and the vector space generated by $\sin n\theta$ and $\cos n\theta$ for $n \neq 0$. Solving the second equation for $\lambda = n^2$, we find that $R(r)$ has to be a multiple of r^n or of r^{-n}. Since $R(0) \in \mathbb{R}$, we find $R(r) = r^n$ when $\lambda = n^2$. In conclusion, for all $n \ge 1$, the functions

$$r^n \cos n\theta, \qquad r^n \sin n\theta$$

solve Laplace's equation in $B(0,1)$ and, because of the superposition principle, for all choices of $\{a_n\}$ and $\{b_n\}$ the function

$$u_N(r,\theta) := \frac{a_0}{2} + \sum_{n=1}^{N} r^n(a_n \cos n\theta + b_n \sin n\theta)$$

is harmonic. Moreover, if $\{a_n\}$ and $\{b_n\}$ are equibounded, then the series

$$u(r,\theta) := \frac{a_0}{2} + \sum_{n=1}^{\infty} r^n(a_n \cos n\theta + b_n \sin n\theta) \tag{1.23}$$

converges totally, hence uniformly, in $B(0,r_0)$ for every $r_0 < 1$ together with the series of its derivatives of any order. It follows that the function u in (1.23) is of class $C^\infty(B(0,1))$ and harmonic. It remains to select the solution of (1.22) from the family (1.23).

Following the same path as for Theorem 1.6, we conclude the following.

1.7 Theorem. *Let $f \in C^{0,\alpha}(\partial B(0,1))$ and $\{a_n\}$, $\{b_n\}$ be the Fourier coefficients of f so that*

$$f(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos n\theta + b_n \sin n\theta)$$

uniformly in $[0, 2\pi]$. Then the function

$$u(r,\theta) := \frac{a_0}{2} + \sum_{n=1}^{\infty} r^n(a_n \cos n\theta + b_n \sin n\theta) \tag{1.24}$$

is of class $C^0(\overline{B(0,1)})$, agrees with f on $\partial B(0,1)$ and solves (1.22).

1.8 Poisson's formula. We now give an integral representation of the solution u in (1.24). Since the series (1.24) converges uniformly, we have

$$\begin{aligned} u(r,\theta) :&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\varphi)\,d\varphi \\ &\quad + \frac{1}{\pi}\sum_{n=1}^{\infty} r^n \int_{-\pi}^{\pi} f(\varphi)[\cos n\theta \cos n\varphi + \sin n\theta \sin n\varphi]\,d\varphi \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(\varphi)\Big(\frac{1}{2} + \sum_{n=1}^{\infty} r^n \cos n(\theta-\varphi)\Big)\,d\varphi, \end{aligned}$$

and, since

$$(r^2 + 1 - 2r\cos\theta)\sum_{n=1}^{\infty} r^n \cos n\theta = r\cos\theta - r^2,$$

i.e.,

$$(r^2 + 1 - 2r\cos\theta)\Big(\frac{1}{2} + \sum_{n=1}^{\infty} r^n \cos n\theta\Big) = \frac{1}{2}(1 - r^2),$$

we conclude that u is given by *Poisson's formula*

$$u(r,\theta) = \frac{1-r^2}{2\pi}\int_{-\pi}^{\pi} \frac{f(\varphi)}{1 + r^2 - 2r\cos(\theta-\varphi)}\,d\varphi \qquad \forall (r,\theta) \in \overline{B(0,1)}. \tag{1.25}$$

In particular, we infer the so-called *formula of the mean*:

$$u(0) := u(0,\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\varphi)\,d\varphi.$$

1.9 Continuous boundary data. If the boundary data f is only continuous, we cannot use the method of separation of variables to solve (1.22) due to the difficulties with the expansion in Fourier series of merely continuous functions, see [GM3]. It turns out that Poisson's formula is very useful. Let $f \in C^0(\partial B(0,1))$ and let

$$u(r,\theta) := \frac{1-r^2}{2\pi} \int_{-\pi}^{\pi} \frac{f(\varphi)}{1+r^2-2r\cos(\theta-\varphi)}\, d\varphi \qquad \forall (r,\theta) \in B(0,1). \tag{1.26}$$

If we reverse the computation to get (1.25) from (1.24) in $B(0,r)$, $r<1$, we see that (1.26) defines a harmonic function in $B(0,1)$. Moreover, the following proposition holds.

1.10 Proposition. *The function $u(r,\theta)$ defined by (1.25) for $r<1$ and by $u(1,\theta) := f(\theta)$ is the unique solution in $C^2(B(0,1)) \cap C^0(\overline{B(0,1)})$ of*

$$\begin{cases} \Delta u = 0 & \text{in } B(0,1), \\ u = f & \text{on } \partial B(0,1). \end{cases}$$

Proof. It suffices to show that $u(r,\theta) \to f(\theta_0)$ as $(r,\theta) \to (1,\theta_0)$. Since the unique harmonic function with boundary value 1 is the function 1, we have

$$\frac{1-r^2}{2\pi} \int_{-\pi}^{\pi} \frac{1}{1+r^2-2r\cos\varphi}\, d\varphi = 1,$$

hence

$$\begin{aligned} u(r,\theta) - f(\theta_0) &= \frac{1-r^2}{2\pi} \int_{-\pi}^{\pi} \frac{f(\varphi)-f(\theta_0)}{1+r^2-2r\cos(\theta-\varphi)}\, d\varphi \\ &= \frac{1-r^2}{2\pi} \int_{-\pi}^{\pi} \frac{f(\theta_0+\psi)-f(\theta_0)}{1+r^2-2r\cos(\theta-\theta_0-\psi)}\, d\psi. \end{aligned} \tag{1.27}$$

Let $\epsilon > 0$. By assumption there is $\delta > 0$ such that $|f(\theta_0+\psi)-f(\theta_0)| < \epsilon/2$ if $|\psi| < \delta$. We rewrite the last integral in (1.27) as the sum of the three integrals $\int_{-\pi}^{\delta} + \int_{-\delta}^{\delta} + \int_{\delta}^{\pi}$. We have

$$\begin{aligned} &\left| \frac{1-r^2}{2\pi} \int_{-\delta}^{\delta} \frac{f(\theta_0+\psi)-f(\theta_0)}{1+r^2-2r\cos(\theta-\theta_0-\psi)}\, d\psi \right| \\ &\qquad \le \frac{\epsilon}{2} \frac{1-r^2}{2\pi} \int_{-\pi}^{\pi} \frac{1}{1+r^2-2r\cos(\theta-\theta_0-\psi)}\, d\psi = \frac{\epsilon}{2}. \end{aligned}$$

On the other hand, if $|\theta-\theta_0| < \delta/2$ and $|\psi| > \delta$, we have $1+r^2-2r\cos(\theta-\theta_0-\psi) > r^2+1-2r\cos\delta/2$. Therefore, we may estimate the other two integrals with

$$4\, \frac{1-r^2}{1+r^2-2r\cos(\delta/2)} \sup_{z\in\partial B(0,1)} |f(z)|,$$

that tends to zero when $r \to 1$. □

1.11 Hadamard's example. The series

$$u(r,\theta) := \frac{a_0}{2} + \sum_{n=1}^{\infty} r^n (a_n \cos n\theta + b_n \sin n\theta)$$

defines a function u of class $C^\infty(B(0,1))\cap C^0(\overline{B(0,1)})$ harmonic in $B(0,1)$ if

$$\sum_{n=1}^{\infty}(|a_n|+|b_n|)<+\infty.$$

On the other hand

$$\begin{aligned}\frac{1}{2}\int_{B(0,\rho)}|Du|^2\,dx&=\frac{1}{2}\int_0^{2\pi}d\theta\int_0^{\rho}(|u_r|^2+\frac{1}{r^2}|u_\theta|^2)r\,dr\\&=\pi\sum_{n=1}^{\infty}n\rho^{2n}(a_n^2+b_n^2).\end{aligned}$$

Therefore, we conclude that there exist harmonic functions in $C^2(B(0,1))\cap C^0(\overline{B(0,1)})$ with divergent Dirichlet's integral, if, for instance, we consider

$$u(r,\theta):=\sum_{n=1}^{\infty}r^{(2n)!}n^{-2}\sin n\theta,\qquad 0\le r<1,\ 0\le\theta\le 2\pi.$$

c. The heat equation

By applying the method of separation of variables to the equation $u_t-ku_{xx}=0$, it is not difficult to find that

$$u(x,t)=\sum_{n=1}^{\infty}c_ne^{-n^2kt}\sin nx$$

is smooth in $]0,\pi[\times]0,T[$ and solves

$$\begin{cases}u_t-ku_{xx}=0, & \text{in }]0,\pi[\times]0,+\infty[,\\ u(0,t)=0,\ u(\pi,t)=0, & \forall t>0\end{cases}$$

provided the coefficients $\{c_n\}$ do not increase too fast. Let f be Hölder-continuous with $f(0)=f(\pi)=0$. We may develop it into a series of sines

$$f(x)=\sum_{n=1}^{\infty}b_n\sin nx,\qquad b_n:=\frac{2}{\pi}\int_0^{\pi}f(t)\sin nt\,dt$$

that converges uniformly in $[0,\pi]$ and conclude that the function

$$u(x,t)=\sum_{n=1}^{\infty}b_ne^{-n^2kt}\sin nx$$

is smooth in $]0,\pi[\times]0,+\infty[$, continuous on $[0,\pi]\times[0,+\infty[$ and solves the initial boundary-value problem

$$\begin{cases} u_t - ku_{xx} = 0, & \text{in }]0,\pi[\times]0,\infty[, \\ u(0,t) = 0,\ u(\pi,t) = 0 & \forall t > 0, \\ u(x,0) = f(x), & x \in [0,\pi]. \end{cases}$$

We leave to the reader the task of justifying the claims along the same lines of what we have done for the Laplace equation.

d. The wave equation

Similarly to the above, given $a \geq 0$ and $f \in C^{0,\alpha}([0,\pi])$ with $f(0) = f(\pi) = 0$, one can find that (at least formally) the solution of the problem

$$\begin{cases} u_{tt} + 2au_t - c^2 u_{xx} = 0 & \text{in }]0,\pi[\times]0,+\infty[, \\ u(x,0) = f(x) & \forall x \in]0,\pi[, \\ u_t(x,0) = 0 & \forall x \in]0,\pi[, \\ u(0,t) = u(\pi,t) = 0 & \forall t > 0 \end{cases}$$

for the *wave equation with viscosity* is given by

$$u(x,t) := \sum_{n=1}^{\infty} b_n T_n(t) \sin nx,$$

where

$$b_n := \frac{2}{\pi} \int_0^{\pi} f(x) \sin nx \, dx,$$

and

$$T_n(t) = \begin{cases} e^{-at}\left[\cosh\sqrt{a^2 - n^2c^2}\,t + \frac{a}{\sqrt{a-n^2c^2}} \sinh\sqrt{a^2 - n^2c^2}\,t\right) & \text{if } n < \frac{a}{c}, \\ e^{-at}(1 + at) & \text{if } n = \frac{a}{c}, \\ e^{-at}\left[\cos\sqrt{a^2 - n^2c^2}\,t + \frac{a}{\sqrt{a-n^2c^2}} \sin\sqrt{a^2 - n^2c^2}\,t\right) & \text{if } n > \frac{a}{c}. \end{cases}$$

We leave to the reader the task of discussing the convergence and of proving in particular that

(i) $u(x,t)$ converges uniformly in $0 \leq t \leq t_0$ for all t_0, since f is Hölder-continuous,
(ii) u is of class C^2 if the second derivatives of f are Hölder-continuous,
(iii) $u(x,t) = \frac{1}{2}(f(x+ct) + f(x-ct))$ if $a = 0$.

1.2 Lebesgue's Spaces

We say that two measurable functions f and g on E are equivalent, and we write $f \sim g$, if the set $\{x \in E \,|\, f(x) \neq g(x)\}$ has zero Lebesgue measure, that is, if they agree *almost everywhere*, *a.e.* in short. This is, actually, an *equivalence relation*, i.e., it is reflexive, symmetric and transitive. Thus, functions that agree a.e. may be identified. However, in the presence of extra structures, for instance, when taking the sum of functions or limits, we need to check that these structures are compatible with the meaning of equality. Fortunately, it is easy to show that operations on measurable functions are compatible with the a.e. equality; for example

(i) if $f_1 \sim f_2$ and $g_1 \sim g_2$, then $f_1 + g_1 \sim f_2 + g_2$,
(ii) if $f_k \sim g_k$, $f \sim g$ and $f_k \to f$ a.e., then $g_k \to g$ a.e.,

and so on.

From now on we shall understand *equality in the sense of a.e. equality* and we shall make use of the equivalence class $[f]$ of f only if it is necessary.

1.2.1 The space L^∞

If $f : E \to \overline{\mathbb{R}}$ is measurable on $E \subset \mathbb{R}^n$, that from now on we assume to be measurable, we define the *essential supremum* of f on E to be

$$\begin{aligned} ||f||_{\infty,E} :&= \operatorname*{esssup}_E |f| := \inf\Big\{t \in \mathbb{R} \,\Big|\, |\{x \in E \,|\, |f(x)| > t\}| = 0\Big\} \\ &= \inf\Big\{t \in \mathbb{R} \,\Big|\, |f(x)| < t \text{ for a.e. } x \in E\Big\} \end{aligned}$$

and, of course, $||f||_{\infty,E} = +\infty$ if $|\{x \in E \,|\, f(x) > t\}| > 0 \ \forall t$. When the set E is clear from the context, we write $||f||_\infty$ instead of $||f||_{\infty,E}$. Notice that

$$|f(x)| \leq ||f||_{\infty,E} \qquad \text{for a.e. } x \in E. \tag{1.28}$$

In fact, if

$$\begin{aligned} A_k &:= \Big\{x \in E \,\Big|\, |f(x)| \geq ||f||_{\infty,E} + \frac{1}{k}\Big\}, \\ A &:= \Big\{x \in E \,\Big|\, |f(x)| > ||f||_{\infty,E}\Big\}, \end{aligned}$$

we have $|A_k| = 0$ for all k, hence we have $|A| = 0$ since $A = \cup_k A_k$. A trivial consequence of (1.28) is that for measurable functions f and g we have

$$\int_E |f(x)|\,|g(x)|\,dx \leq ||f||_{\infty,E} \int_E |g(x)|\,dx; \tag{1.29}$$

in particular,

$$\int_E |f(x)|\,dx \leq ||f||_{\infty,E}\,|E|. \tag{1.30}$$

1.12 Proposition. *Let f and g be measurable on $E \subset \mathbb{R}^n$. Then we have*

(i) $||f||_\infty = 0$ *if and only if* $f = 0$ *a.e.,*
(ii) $||f||_\infty = ||g||_\infty$ *if* $f = g$ *a.e.,*
(iii) $||\lambda f||_\infty = |\lambda|\,||f||_\infty \ \forall \lambda \in \mathbb{R}$,
(iv) $||f + g||_\infty \le ||f||_\infty + ||g||_\infty$.

Proof. (i) and (ii) follow from the definition. (iii) is trivial. For (iv) it suffices to observe that since $|f(x)| \le ||f||_\infty$ and $|g(x)| \le ||g||_\infty$ a.e., then $|f + g|(x) \le ||f||_\infty + ||g||_\infty$ a.e. □

1.13 Definition. *We denote by $L^\infty(E)$ the space of (the classes of a.e. equivalence of) measurable functions on E with $||f||_{\infty,E} < +\infty$,*

$$L^\infty(E) = \Big\{ [f] \,\Big|\, f \text{ measurable}, \ ||f||_{\infty,E} < +\infty \Big\}.$$

Proposition 1.12 yields that $L^\infty(E)$ is a vector space with $||f||_{\infty,E}$ as norm. Actually, we have the following.

1.14 Theorem. *$L^\infty(E)$ is a Banach space.*

Proof. Consider a sequence $\{f_k\}$ of measurable functions with $||f_h - f_k||_\infty \to 0$ as $h, k \to \infty$. We have $|f_h(x) - f_k(x)| \le ||f_h - f_k||_\infty$ except on a set $Z_{h,k}$ of zero measure. If $Z := \cup_{h,k} Z_{h,k}$, then, again, $|Z| = 0$ and $|f_h(x) - f_k(x)| \le ||f_h - f_k||_\infty$ at every point of $E \setminus Z$. Therefore, $\{f_k\}$ is a Cauchy sequence for the uniform convergence on $E \setminus Z$; thus, it converges to a function $f : E \setminus Z \to \mathbb{R}$ that is measurable on E. Moreover, for every $\epsilon > 0$ there exists $\overline{k}$ such that $|f_k(x) - f(x)| \le \epsilon \ \forall x \in E \setminus Z$ and $k \ge \overline{k}$; therefore $f \in L^\infty(E)$. □

1.15 Remark. In general, $L^\infty(E)$ is not separable. For instance, the family $\{f_t\}$ of functions $f_t(x) := \chi_{[0,t]}(x)$ in $L^\infty([0,1])$ is not denumerable and is not dense in $L^\infty([0,1])$ since $||f_t - f_s||_\infty = 1$ when $t \ne s$.

1.16 ¶. The convergence in $L^\infty(E)$ is the *a.e. uniform convergence*. Show that $||f_k - f||_\infty \to 0$ if and only if there exists a set $N \subset E$ with $|N| = 0$ such that $\{f_k\}$ converges to f uniformly on $E \setminus N$.

1.17 Theorem (Egorov). *Let $\{f_n\}$ and f be measurable on A. Suppose that $|A| < \infty$ and that $f_n \to f$ a.e. on A. Then, for every positive $\epsilon > 0$ there is a measurable subset A_ϵ of A with $|A_\epsilon| < \epsilon$ such that $f_n \to f$ uniformly on $A \setminus A_\epsilon$.*

Proof. Since $f_n \to f$ for a.e. $x \in A$, the set

$$C_j := \Big\{ x \in A \,\Big|\, \exists \{k_n\} \text{ such that } |f_{k_n}(x) - f(x)| > 2^{-j} \ \forall n \Big\}$$

has zero measure for all j. Set

$$C_{ij} := \bigcup_{n=i}^{\infty} \Big\{ x \in A \,\Big|\, |f_n(x) - f(x)| > 2^{-j} \Big\};$$

we have $\cap_i C_{ij} = C_j$, hence $|C_{ij}| \to 0$ as $i \to \infty$, since $|A| < \infty$. For every integer j, choose now $i = i(j)$ in such a way that $|C_{i(j)j}| < \epsilon 2^{-j}$ and set $A_\epsilon := \cup_j C_{i(j)j}$. Clearly $|A_\epsilon| \le \epsilon$ and $f_n \to f$ uniformly on $A \setminus A_\epsilon$. □

1.2.2 L^p spaces, $1 \le p < +\infty$

a. The L^p norm

For $p \in \mathbb{R}$, $1 \le p < +\infty$, and f measurable on $E \subset \mathbb{R}^n$, we set

$$||f||_{p,E} := \left(\int_E |f|^p \, dx \right)^{1/p}$$

and shorten it to $||f||_p$ if E is clear from the context. Notice that

(i) $||f||_p = 0$ if and only if $f = 0$ a.e.,
(ii) $||\lambda f||_p = |\lambda| \, ||f||_p \ \forall \lambda \in \mathbb{R}$.

Let $1 \le p \le +\infty$. The number $p' \in [1, +\infty]$ such that $1/p + 1/p' = 1$ is called the *conjugate exponent* of p, and

$$p' := \begin{cases} +\infty & \text{if } p = 1, \\ \frac{p}{p-1} & \text{if } 1 < p < \infty, \\ 1 & \text{if } p = \infty. \end{cases}$$

1.18 Proposition (Hölder's inequality). *Let $1 \le p \le +\infty$ and let f and g be measurable functions on E. Then*

$$\int_E |f(x)g(x)| \, dx \le ||f||_{p,E} \, ||g||_{p',E} \tag{1.31}$$

where p' is the conjugate exponent of p.

Proof. If $p = 1$, then $|f(x)g(x)| \le |f(x)| \, ||g||_{\infty,E} \ \forall x \in E$, and the claim $||fg||_{1,E} \le ||f||_{1,E} \, ||g||_{\infty,E}$ follows by integration. If $1 < p < +\infty$, the claim follows from Young's inequality $ab \le a^p/p + b^{p'}/p' \ \forall a, b > 0$, see [GM1]. In fact, if $||f||_{p,E}$ or $||g||_{p,E}$ is infinite, or $f = 0$ or $g = 0$, the claim is trivial; otherwise, it suffices to apply Young's inequality with

$$a = \frac{f(x)}{||f||_{p,E}}, \qquad b = \frac{g(x)}{||g||_{p,E}}$$

and integrate. □

From Hölder's inequality, we infer Minkowski's inequality.

1.19 Proposition (Minkowski's inequality). *Let $1 \le p \le +\infty$ and let f and g be measurable on E. Then*

$$||f + g||_{p,E} \le ||f||_{p,E} + ||g||_{p,E}.$$

Proof. The claim is trivial when $p = 1$. Assume now $p > 1$.

(i) If $||f + g||_{p,E} = 0$ the claim is again trivial.

(ii) If $||f + g||_{p,E} = \infty$, by applying the inequality

$$|a|^p \le (|a - b| + |b|)^p \le 2^{p-1}(|a - b|^p| + |b|^p)$$

with $a = f(x)+g(x)$ and $b = -g(x)$, we infer that either $||f||_{\infty,E} = \infty$ or $||g||_{\infty,E} = \infty$, or both, hence the claim holds.

(iii) When $0 < ||f+g||_{p,E} < +\infty$, from Hölder's inequality we get

$$\begin{aligned} ||f+g||_{p,E}^p &= \int_E |f+g|^{p-1}|f+g|\,dx \\ &\le \int_E |f+g|^{p-1}|f|\,dx + \int_E |f+g|^{p-1}|g|\,dx \\ &\le ||f+g||_{p,E}^{p-1}(||f||_{p,E} + ||g||_{p,E}). \end{aligned}$$

Then the claim follows dividing by $||f+g||_{p,E}^{p-1}$. □

1.20 Definition. *Let $E \subset \mathbb{R}^n$ be a measurable set. We denote by $L^p(E)$ the space of (classes of a.e. equivalence of) measurable functions on E with $||f||_p < +\infty$,*

$$L^p(E) = \Big\{ [f] \,\Big|\, f \text{ measurable},\ ||f||_p < +\infty \Big\};$$

*we say that f is p-*summable *on E if $f \in L^p(E)$.*

From Proposition 1.19, clearly $L^p(E)$ is a vector space and $||f||_p$ is a norm on it. Moreover, we have the following theorem.

1.21 Theorem. *$L^p(E)$ endowed with the norm $||\ ||_{p.E}$ is a Banach space.*

Proof. We show that if $f_k \in L^p(E)$ and $\sum_{k=1}^\infty ||f_k||_p < +\infty$, then there exists $f \in L^p(E)$ such that $||f - \sum_{j=1}^k f_j||_p \to 0$ as $k \to \infty$. As we know, see Proposition 9.15 of [GM3], this property is equivalent to the completeness of $L^p(E)$.

Thus, let $\sum_{k=1}^\infty f_k(x)$ be a series in $L^p(E)$ that totally converges in $L^p(E)$, that is,

$$\sum_{k=1}^\infty ||f_k||_{p,E} < +\infty.$$

Set $g(x) := \sum_{k=1}^\infty |f_k(x)|$. The triangle inequality and Beppo Levi's theorem yield

$$||g||_{p,E} \le \sum_{k=1}^\infty ||f_k||_{p,E} < +\infty;$$

in particular, $g \in L^p(E)$ and $g(x) < +\infty$ per a.e. $x \in E$. Therefore, the series $\sum_{k=1}^\infty f_k(x)$ converges absolutely for a.e. $x \in E$ to a measurable function f on E

$$\sum_{k=1}^n f_k(x) - f(x) = -\sum_{k=n+1}^\infty f_k(x) \to 0 \qquad \text{for a.e. } x \in E.$$

Since

$$\sup_n \Big| \sum_{k=n+1}^\infty f_k(x) \Big| \le \sup_n \sum_{k=n+1}^\infty |f_k(x)| \le g(x) \in L^p(E),$$

the claim follows from Lebesgue's dominated convergence in Exercise 1.22 below. □

1.22 ¶. Prove the following variant of Lebesgue's dominated convergence theorem, see [GM4].

Theorem (Lebesgue's dominated convergence theorem). *Let $1 \le p < \infty$, let $E \subset \mathbb{R}^n$ be measurable, and let $\{f_k\}$ and f be functions in $L^p(E)$. If*

(i) *$f_k \to f$ a.e. on E,*
(ii) *there is $g \in L^p(E)$ such that $|f_k(x)| \le g(x)$ for all k and a.e. $x \in E$,*

then $f_k \to f$ in $L^p(E)$.

1.23 ¶. Notice that the proof of the completeness of L^p is nothing but a theorem of integration term by term for a series, see [GM4].

1.24 Proposition. *Let $\{f_n\}$ be a Cauchy sequence in $L^p(E)$, $1 < p < \infty$. Then $\{f_k\}$ has a* subsequence *which converges a.e. on E.*

Proof. We can extract a subsequence $\{g_k\}$ of $\{f_k\}$, $g_k := f_{n_k}$, such that

$$||g_{k+1} - g_k||_{p,E} \le 2^{-k} \qquad \forall k.$$

Set

$$F(x) := |g_1(x)| + \sum_{k=1}^{\infty} |g_{k+1}(x) - g_k(x)|.$$

Beppo Levi's theorem yields

$$||F||_{p,E} \le ||g_1||_{p,E} + \sum_{k=1}^{\infty} ||g_{k+1} - g_k||_{p,E} < +\infty,$$

hence $F(x) < +\infty$ a.e. We conclude that the series $g_1(x) + \sum_{k=1}^{\infty}(g_{k+1}(x) - g_k(x))$ converges absolutely for a.e. $x \in E$ to a function $f(x)$, and

$$|f(x) - g_k(x)| = \sum_{h=k+1}^{\infty} |g_{h+1}(x) - g_h(x)| \to 0 \qquad \text{for a.e. } x \in E.$$

□

b. Approximation

As a consequence of Lusin's theorem, see [GM4], we now prove the density of smooth functions in L^p.

1.25 Theorem. *Let $1 \le p < +\infty$. The space $C_c^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$.*

Proof. Since for every bounded open set $\Omega \subset \mathbb{R}^n$, $C_c^\infty(\Omega)$ is dense in $C_c^0(\Omega)$ with respect to the uniform convergence, see [GM3], (and, a fortiori, with respect to the L^p-convergence), it suffices to show that if $f \in L^p(\mathbb{R}^n)$ and $\epsilon > 0$, then there exists a function $g \in C_c^0(\mathbb{R}^n)$ such that $\int_{\mathbb{R}^n} |f - g| < 2\epsilon$. Fix $\epsilon > 0$ and choose N large enough so that for

$$f_N(x) := \begin{cases} N & \text{if } f(x) > N \text{ and } |x| \le N, \\ f(x) & \text{if } |f(x)| \le N \text{ and } |x| \le N, \\ -N & \text{if } f(x) < -N \text{ and } |x| \le N, \\ 0 & \text{if } |x| > N \end{cases}$$

we have $\int_\Omega |f - f_N|^p \, dx < \epsilon^p$. We can do this since $\int_\Omega |f - f_N|^p \, dx \to 0$ as $N \to \infty$ because of Lebesgue's dominated convergence.

Lusin's theorem, see [GM4], yields the existence of a function $g \in C_c^0(\Omega)$ such that

$$||g||_\infty \le ||f_N||_\infty \le N \qquad \text{and} \qquad \Big|\Big\{x \in \Omega \,\Big|\, g(x) \ne f_N(x)\Big\}\Big| < \Big(\frac{\epsilon}{2N}\Big)^p.$$

We therefore find

$$||f-g||_{p,\mathbb{R}^n} \le ||f-f_N||_{p,\mathbb{R}^n} + ||f_N-g||_{p,\mathbb{R}^n} \le \epsilon + 2N\frac{\epsilon}{2N} = 2\epsilon.$$

□

As in [GM4] we can also prove the following.

1.26 Proposition (Continuity in the mean). *Let $1 \le p < +\infty$ and $f \in L^p(\mathbb{R}^n)$. Then*

$$\int_{\mathbb{R}^n} |f(x+h) - f(x)|^p \, dx \to 0 \qquad \text{as } h \to 0.$$

Let k be a (symmetric) smoothing kernel. For $f \in L^p(\mathbb{R}^n)$ and $\epsilon > 0$ denote by

$$f_\epsilon(x) := \int_{\mathbb{R}^n} f(y) k_\epsilon(x-y)\, dy$$

the *ϵ-regularized* of f. We have the following theorem.

1.27 Theorem. *Let $f \in L^p(\mathbb{R}^n)$, $1 \le p < +\infty$. Then f_ϵ is well-defined and of class $C^\infty(\mathbb{R}^n)$. Moreover,*

$$\int_{\mathbb{R}^n} |f_\epsilon(x)|^p \, dx \le \int_{\mathbb{R}^n} |f(x)|^p \, dx$$

and

$$\int_{\mathbb{R}^n} |f_\epsilon - f|^p \, dx \to 0 \qquad \text{as } \epsilon \to 0.$$

Proof. For $p = 1$ see [GM4]. We proceed similarly for $p > 1$ by using the Hölder inequality. We have

$$\begin{aligned} |f_\epsilon(x)|^p &= \left| \int_{\mathbb{R}^n} f(y) k_\epsilon(x-y)\, dy \right|^p = \left| \int_{\mathbb{R}^n} k_\epsilon^{1/p}(x-y) f(y) k_\epsilon^{1-1/p}(x-y)\, dy \right|^p \\ &\le \left(\int_{\mathbb{R}^n} k_\epsilon(x-y)\, dy \right)^{p-1} \int_{\mathbb{R}^n} |f(y)|^p k_\epsilon(x-y)\, dy = \int_{\mathbb{R}^n} |f(y)|^p k_\epsilon(x-y)\, dy. \end{aligned}$$

This proves that f_ϵ is well-defined and that $f_\epsilon \in C^\infty(\mathbb{R}^n)$ as for $p = 1$, see [GM4]. Integrating the previous estimate, changing variables, and interchanging the order of integration with Fubini's theorem, we find

$$\begin{aligned} \int_{\mathbb{R}^n} |f_\epsilon(x)|^p \, dx &\le \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} k_\epsilon(z) |f(x-z)|^p \, dz \right) dx \\ &= \int_{\mathbb{R}^n} k_\epsilon(z)\, dz \left(\int_{\mathbb{R}^n} |f(x-z)|^p \, dx \right) = \int_{\mathbb{R}^n} |f(x)|^p \, dx. \end{aligned}$$

In order to prove the convergence of f_ϵ to f, we notice that

$$|f_\epsilon(x) - f(x)| \le \int_{\mathbb{R}^n} |f(y) - f(x)| k_\epsilon(x-y)\, dy;$$

taking the power p, using Hölder's inequality and integrating in x, we get

$$\int_{\mathbb{R}^n} |f_\epsilon(x) - f(x)|^p \, dx \le \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |f(x-z) - f(x)|^p k_\epsilon(z)\, dx\, dz$$

and the conclusion follows from Proposition 1.26 as in the case $p = 1$. □

c. Separability

1.28 Proposition. *Let $1 \le p < +\infty$. The class S_0 of measurable simple functions with supports of finite measure is dense in $L^p(\mathbb{R}^n)$.*

Proof. We may and do restrict ourselves to considering nonnegative functions $f \in L^p(\mathbb{R}^n)$. Consider an increasing sequence $\{\varphi_k\}$ of measurable simple functions converging pointwise to f. Of course, $\varphi_k \in L^p(\mathbb{R}^n)$ for all k since $f \in L^p(\mathbb{R}^n)$ and the support of each φ_k's has finite measure since φ_k take a finite number of values. Finally, Beppo Levi's theorem yields $||f - \varphi_k||_p \to 0$. □

1.29 Theorem (Separability of L^p). *Let $1 \le p < +\infty$ and let $E \subset \mathbb{R}^n$ be a measurable set. Then $L^p(E)$ is separable.*

Proof. First consider a measurable set $A \subset \mathbb{R}^n$ of finite measure. As we know, for every $\epsilon > 0$ we can find a *finite* union P of intervals such that $|A \Delta P| < \epsilon$, see Proposition 5.12. Moreover, we may assume that the coordinates of the vertices of the intervals of P are rational and still $|A \Delta P| < \epsilon$, or, in terms of characteristic functions, $||\chi_A - \chi_P||_p < \epsilon^{1/p}$. Therefore, we conclude that the denumerable class $\mathcal{R}$ of characteristic functions of finite unions of intervals with rational vertices is dense, with respect to the the L^p distance, in the class S_0 of simple functions with support of finite measure.

Since S_0 is dense in $L^p(\mathbb{R}^n)$, so is $\mathcal{R}$, thus the claim is proved for $E = \mathbb{R}^n$. To conclude, in the general case it suffices to notice that the family $\mathcal{R}'$ of restrictions of the functions of $\mathcal{R}$ to E is dense to $L^p(E)$. □

d. Duality

1.30 Theorem. *Let $f \in L^p(E)$, $E \subset \mathbb{R}^n$. Suppose that*

- *either $1 \le p < +\infty$,*
- *or $p = +\infty$ and $\{x \,|\, |f(x)| > t\}$ has finite measure $\forall t > 0$.*

Then

$$||f||_{p,E} = \sup\Big\{ \int_E fg\,dx \,\Big|\, g \in L^{p'}(E),\ ||g||_{p',E} \le 1 \Big\}.$$

Proof. If $||f||_p = 0$, the claim is trivial. Set

$$L_p := \sup\Big\{ \int_E fg\,dx \,\Big|\, g \in L^{p'}(E), ||g||_{p',E} \le 1 \Big\}.$$

From Hölder's inequality we infer $L_p \le ||f||_{p,E}$ $\forall p$. Moreover:
(i) If $p = 1$, by choosing $g(x) := \operatorname{sgn} f(x)$ we get $||g||_\infty \le 1$ and $||f||_1 := \int f(x)g(x)\,dx$.
(ii) If $1 < p < \infty$ and $||f||_{p,E} < +\infty$, by choosing

$$g(x) := \operatorname{sgn}(f(x)) \Big(\frac{|f(x)|}{||f||_p} \Big)^{p-1}$$

we get $||g||_{p',E} = 1$ and $||f||_p = \int fg\,dx$.
(iii) If $p = \infty$ and $0 < t < ||f||_\infty$, since $E_t := \{|f(x)| > t\}$ has nonzero and finite measure, then $g(x) := |E_t|^{-1} \operatorname{sgn}(f(x)) \chi_{E_t}(x) \in L^1(E)$ is well-defined, $||g||_{1,E} = 1$ and

$$L_\infty \ge \int_E f(x)g(x)\,dx = \frac{1}{|E_t|} \int_{E_t} |f|\,dx \ge t \frac{|E_t|}{|E_t|} = t.$$

Since t is arbitrary, then $L_\infty = ||f||_\infty$. □

Of course, we may extend the previous notions to vector-valued measurable functions. For instance, we say that $f : E \subset \mathbb{R}^n \to \mathbb{R}^k$ is in $L^p(E, \mathbb{R}^k)$ if its components are in $L^p(E)$. It is readily seen that $L^p(E, \mathbb{R}^k)$ is a Banach space with respect to the norm

$$||f||_{p,E} := \left(\int_E ||f(x)||^p \, dx \right)^{1/p}$$

where $||f(x)||$ denotes the norm in $\mathbb{R}^k$ of the vector $f(x)$.

e. L^2 is a separable Hilbert space

Of special interest is the separable Banach space $L^2(E)$. In fact, it is a separable Hilbert space because its norm is induced by the *inner product*

$$(f|g)_2 := \int_E f(x)g(x) \, dx, \qquad f, g \in L^2(E).$$

Consequently, we may use all means specific to separable Hilbert spaces such as the Dirichlet principle, the projection theorem, the existence of complete orthonormal basis and the isomorphism with the space of sequences

$$\ell^2(\mathbb{R}) := \Big\{ \{a_n\} \Big| \sum_{n=0}^{\infty} |a_n|^2 < +\infty \Big\},$$

see [GM3] and Section 1.4 of this chapter.

Similarly, the space $L^2(E, \mathbb{C})$ of complex-valued functions with square integrable modulus is a separable Hilbert space over $\mathbb{C}$ with *hermitian product* given by

$$(f|g)_2 := \int_E f(x)\overline{g}(x) \, dx, \qquad f, g \in L^2(E, \mathbb{C}).$$

f. Means

The *integral mean* of a nonnegative measurable function on a measurable set E of finite measure is defined by

$$f_E = \fint_E f(x) \, dx := \frac{1}{|E|} \int_E f(x) \, dx,$$

and for $p \geq 1$, we set

$$\phi_p(f) := \left(\frac{1}{|E|} \int_E f(x)^p \, dx \right)^{1/p}. \tag{1.32}$$

1.31 Proposition. *Let $f : E \subset \mathbb{R}^n \to \overline{\mathbb{R}}$ be nonnegative and measurable on a measurable set E with $|E| < +\infty$. Then $\phi_p(f) \to ||f||_{\infty,E}$ as $p \to +\infty$.*

Proof. For $M < ||f||_{\infty,E}$, the set $A := \{x \,|\, |f(x)| > M\}$ has positive measure, and

$$\phi_p(f) \geq \left(\frac{|A|}{|E|}\int_A f^p\,dx\right)^{1/p} \geq M\left(\frac{|A|}{|E|}\right)^{1/p};$$

hence $\liminf_{p\to\infty} ||f||_{p,E} \geq M$, consequently $\liminf_{p\to\infty}\phi_p(f) \geq ||f||_{\infty,E}$. On the other hand, by (1.28)

$$\phi_p(E) \leq \left(\frac{1}{|E|}\int_E ||f||^p_{\infty,E}\,dx\right)^{1/p} = ||f||_{\infty,E}$$

and the claim follows. □

By Hölder's inequality

$$\left(\fint_E |f|^q\,dx\right)^{1/q} \leq \left(\fint_E |f|^p\right)^{1/p} \qquad \forall p,q,\ 1 \leq q \leq p \leq \infty, \tag{1.33}$$

or, equivalently, for a fixed f, the map $p \to \phi_p(f)$, $p \geq 1$, is nondecreasing. Notice that (1.33) with $q = 1$ and $p = 2$ is the well-known inequality between the mean value and the root-mean-square value of f:

$$\frac{1}{|E|}\int_E |f|\,dx \leq \left(\frac{1}{|E|}\int_E |f|^2\,dx\right)^{1/2}.$$

From Hölder's inequality, we can also deduce the following *interpolation inequality*: For $q \leq r \leq p \leq \infty$ we have

$$||f||_r \leq ||f||_q^{\lambda}||f||_p^{1-\lambda}$$

where λ is defined by the equality $\frac{1}{r} = \lambda\,\frac{1}{q} + (1-\lambda)\,\frac{1}{p}$. The last inequality is equivalent to saying that *the function $p \mapsto \log\phi_{1/p}(f)$ is convex.*

Inequality (1.33) is a special case of *Jensen's inequality.*

1.32 Proposition (Jensen's inequality). *Let $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous convex function and let f be an integrable*[1] *function on a measurable set E of finite measure. Then $\phi(f(x))$ is integrable on E and*

$$\phi\left(\frac{1}{|E|}\int_E f(x)\,dx\right) \leq \frac{1}{|E|}\int_E \phi(f(x))\,dx. \tag{1.34}$$

Moreover, if f is summable, ϕ is strictly convex and both terms in (1.34) *are finite, then equality holds if and only if f is constant.*

Proof. First we observe that $\phi \circ f$ is measurable since ϕ is lower semicontinuous. Next, see Theorem 2.109 and Exercise 2.140, $\phi(y) = \sup_{\varphi\in\mathcal{S}}\varphi(y)\ \forall y \in \mathbb{R}$, where $\mathcal{S}$ is the class of linear affine minorants of ϕ. For every affine map φ we clearly have

$$\varphi\left(\frac{1}{|E|}\int_E f(x)\,dx\right) = \frac{1}{|E|}\int_E \varphi(f(x))\,dx,$$

[1] Recall that "integrable" means either summable or nonnegative and measurable.

hence $\forall \varphi \in \mathcal{S}$

$$\varphi\Big(\frac{1}{|E|}\int_E f(x)\,dx\Big) \le \frac{1}{|E|}\int_E \phi(f(x))\,dx.$$

It follows that $\int_E \phi(f(x))\,dx > -\infty$, hence $\phi(f(x))$ is integrable, and taking the supremum we deduce (1.34).

Suppose now that f is summable, ϕ is strictly convex and both terms in (1.34) are finite and equality holds. Let $L := \frac{1}{|E|}\int_E f(x)\,dx \in \mathbb{R}$ and let $z = m(y-L)+\phi(L)$ be a line of support for ϕ at L. The function

$$\psi(x) := \phi(f(x)) - \phi(L) - m(f(x)-L)$$

is nonnegative and its integral is zero. Hence $\psi = 0$ a.e. in E. Since ψ is strictly convex, for $x \in E$ such that $\psi(x) = 0$ we have $f(x) = L$. □

1.33 ¶ Jensen's inequality for vector-valued maps. Jensen's inequality extends to vector-valued functions. Show that, if $\phi : \mathbb{R}^k \to \mathbb{R}$ is a lower semicontinuous convex function, then $\frac{1}{|E|}\int_E f(x)\,dx$ is in the convex envelope of $f(E)$ and the conclusion of Proposition 1.32 holds.

1.34 ¶ Some important properties of means. Let f be a nonnegative measurable function on a measurable set of finite measure. We already have proved that $\phi_p(f) \to ||f||_{\infty,E}$ as $p \to +\infty$. Extend now $\phi_p(f)$ to a function defined on $\mathbb{R}$ by

$$\phi_p(f) := \begin{cases} \Big(\frac{1}{|E|}\int_E f^p(x)\,dx\Big)^{1/p} & \text{if } p \neq 0, \\ \exp\Big(\frac{1}{|E|}\int_E \log|f|\,dx\Big) & \text{if } p = 0. \end{cases}$$

Show that

(i) $\phi_p(f)$ is well-defined for every $p \in \mathbb{R}$,
(ii) $\phi_p(f)$ is increasing on $\{p > 0\}$ and $\{p < 0\}$,
(iii) $\phi_p(f)$ is continuous on $\mathbb{R}$, hence increasing on $\mathbb{R}$,
(iv) $\phi_p(f) \to \operatorname{essinf}_{x\in E}|f|$ as $p \to -\infty$, where

$$\operatorname*{essinf}_{x\in E}|f| := \sup\Big\{t \;\Big|\; |\{x \in E \,|\, |f(x)| < t\}| = 0\Big\},$$

(v) if $\phi_p(f) = \phi_q(f)$ for some $p \neq q$, then $|f|$ is a.e. constant,
(vi) $p \to \log \phi_{1/p}(f)$ is convex.

1.2.3 Trigonometric series in L^2

Consider the complex Hilbert space $L^2(]-\pi,\pi[,\mathbb{C})$ endowed with the inner product

$$(f|g) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\overline{g}(t)\,dt$$

and the trigonometric system $\{e^{ikt}\}_{k\in\mathbb{Z}}$. It is trivial to show that the trigonometric system is orthonormal in L^2:

$$\frac{1}{2\pi}\int_{\pi}^{\pi} e^{ikt}e^{-iht}\,dt = \delta_{h,k}.$$

1.35 Theorem. *The trigonometric system* $\{e^{ikt}\}$ *is a complete orthonormal system in* $L^2(]-\pi,\pi[)$*, that is, the finite linear combinations of the trigonometric system, i.e., the trigonometric polynomials, are dense in* $L^2(]-\pi,\pi[)$*.*

Proof. In [GM3] we proved that for every 2π-periodic function f of class $C^1(\mathbb{R})$ the Fourier series of f

$$\sum_{k=-\infty}^{\infty} c_k e^{ikt}, \qquad c_k := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}\,dt,$$

converges uniformly to f on $\mathbb{R}$. In particular, the class T of (the restrictions to $[-\pi,\pi]$ of) trigonometric polynomials is dense in the class P of (the restrictions to $[-\pi,\pi]$ of) 2π-periodic functions of class C^1 with respect to the uniform convergence on $[-\pi,\pi]$. In particular, T is dense in P with respect to the L^2 convergence, $\overline{T} = \overline{P}$. On the other hand, it is easy to show that $C_c^1(]-\pi,\pi[)$ is dense in the class of 2π-periodic functions of class C^1 with respect to the L^2 convergence, $\overline{C_c^1} = \overline{P}$. Finally, by Theorem 1.25, $C_c^1(]-\pi,\pi[)$ is dense in $L^2(]-\pi,\pi[)$, $\overline{C_c^1} = L^2$. In conclusion $\overline{T} = \overline{P} = \overline{C_c^1} = L^2$, i.e., the claim. □

Moreover, by rewriting the abstract Riesz–Fisher theorem, see [GM3], for the Hilbert space $L^2(]-\pi,\pi[)$ and the trigonometric system, the following holds.

1.36 Theorem. *The following claims are equivalent:*

(i) $\{e^{ikt}\}$ *is a complete orthonormal system in* L^2*.*

(ii) *Every* $f \in L^2(]-\pi,\pi[)$ *writes as*

$$f(t) = \sum_{k=-\infty}^{+\infty} c_k e^{ikt} \qquad \text{in the } L^2(]-\pi,\pi[) \text{ sense}$$

where for $k \in \mathbb{Z}$

$$c_k := (f|e^{ikt})_{L^2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}\,dt.$$

(iii) *If* $\{c_k\}_{k\in\mathbb{Z}}$ *is such that* $\sum_{k=-\infty}^{+\infty} |c_k|^2 < \infty$*, then the* trigonometric series

$$\sum_{k=-\infty}^{+\infty} c_k e^{ikt}$$

converges in $L^2(]-\pi,\pi[)$ *to a function* $f \in L^2(]-\pi,\pi[)$*.*

(iv) *If* $f \in L^2$*,* $f(t) = \sum_{k=-\infty}^{+\infty} c_k e^{ikt}$*, then the energy equality*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(t)|^2\,dt = ||f||_2^2 = \sum_{k=-\infty}^{+\infty} |c_k|^2$$

holds.

(v) *Let* $f(t) = \sum_{k=-\infty}^{+\infty} c_k e^{ikt}$ *and* $g(t) = \sum_{k=-\infty}^{+\infty} d_k e^{ikt}$ *be in* $L^2(]-\pi,\pi[)$. *Then*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\overline{g}(t)\,dt = \sum_{k=-\infty}^{\infty} c_k\overline{d_k}.$$

(vi) $\int_{-\pi}^{\pi} f(t)e^{ikt}\,dt = 0\ \forall k \in \mathbb{Z}$ *if and only if* $f = 0$ *a.e.*

1.37 Remark. It is possible to show that the trigonometric system $\{e^{ikt}\}_{k\in\mathbb{Z}}$ is complete in $L^2(]-\pi,\pi[,\mathbb{C})$ (or, equivalently, that $\{1, \cos t, \sin t, \cos 2t, \sin 2t, \dots\}$ is complete in $L^2(]-\pi,\pi[,\mathbb{R})$) by using (vi) of Theorem 1.36. In fact, let $f \in L^2(]-\pi,\pi[,\mathbb{C})$ and suppose that for every element $\varphi(t)$ of the trigonometric system we have

$$\int_{-\pi}^{\pi} f(t)\overline{\varphi(t)}\,dt = 0.$$

Then for every trigonometric polynomial $P(t) \in \mathcal{P}_{n,2\pi}$ we also have

$$\int_{-\pi}^{\pi} f(t)\overline{P(t)}\,dt = 0.$$

Since trigonometric polynomials are dense among continuous 2π-periodic functions with respect to the uniform convergence, see the Weierstrass theorem in [GM3], and continuous periodic functions are dense in L^2, we conclude that

$$\int_{-\pi}^{\pi} f(t)\overline{g(t)}\,dt = 0 \qquad \forall g \in L^2(]-\pi,\pi[,\mathbb{C});$$

in particular,

$$\int_{-\pi}^{\pi} |f(t)|^2\,dt = 0,$$

i.e., $f = 0$ in $L^2(]-\pi,\pi[)$.

One can also prove, but we refer to the specialized literature for this, that the Fourier series of $f \in L^p$ converges to f in L^p if $1 < p < \infty$. Much more delicate and complex is the pointwise and the a.e. convergence of the partial sums $S_n f(t)$ of the Fourier series of f to $f(t)$ if $f \in L^p$, similarly to the case of continuous functions, see [GM3]. Although the L^p convergence implies the a.e. convergence for a subsequence, the following holds.

1.38 Theorem (Kolmogorov). *There exist periodic functions in the space* $L^1(]-\pi,\pi[)$ *such that*

$$\limsup_{n\to\infty} |S_n f(t)| = +\infty \qquad \forall t \in]-\pi,\pi[.$$

1.39 Theorem (Carleson). *If* $f \in L^p(]-\pi,\pi[)$, $p \geq 2$, *then* $S_n f(t) \to f(t)$ *for a.e.* t.

1.40 Theorem (Kahane–Katznelson). *For every $E \subset [-\pi, \pi[$ with $|E| = 0$, there exists a continuous 2π-periodic function such that*

$$\limsup_{n\to\infty} |S_n f(t)| = +\infty \qquad \forall t \in E.$$

1.2.4 The Fourier transform

Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function, and let f_T be the restriction of f to $]-T, T]$. We now think of f_T as extended periodically in $\mathbb{R}$. We may write f_T as a sum of waves with frequencies that are integer multiples of $2\pi/T$ and amplitudes given by the Fourier coefficients of f_T, i.e.,

$$f_T(x) := \sum_{k=-\infty}^{+\infty} \left(\frac{1}{T} \int_{-T/2}^{T/2} f_T(y) e^{-i\frac{2\pi}{T}ky}\, dy \right) e^{i\frac{2\pi}{T}kx}.$$

If we let T tend to infinity, we find, at least formally,

$$f(x) := \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(y) e^{-i\xi y}\, dy \right) \frac{e^{i\xi x}}{2\pi}\, d\xi.$$

In other words, nonperiodic functions can be represented as superposition of a continuous family of waves $e^{i\xi x}$ of frequencies ξ and corresponding amplitude

$$\widehat{f}(\xi) := \int_{-\infty}^{+\infty} f(x) e^{-i\xi x}\, dx.$$

When it makes sense, we define the *Fourier transform* of $f : \mathbb{R}^n \to \mathbb{C}$ as

$$\widehat{f} : \mathbb{R}^n \to \mathbb{R}, \qquad \widehat{f}(\xi) := \int_{\mathbb{R}^n} f(x) e^{-i\,\xi \bullet x}\, dx.$$

It is easily seen that, if $f \in L^1(\mathbb{R})$, then

(i) $|\widehat{f}(\xi)| \le ||f||_{L^1}\ \forall \xi \in \mathbb{R}^n$,

(ii) as a consequence of the Riemann–Lebesgue lemma, see [GM3], $\widehat{f}$ is uniformly continuous and

$$\widehat{f}(\xi) \to 0 \qquad \text{per} \qquad \xi \to \pm\infty,$$

(iii) if f is the impulse

$$f(x) = \begin{cases} 1 & \text{if } x \in [-1, 1], \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\widehat{f}(\xi) = \frac{2\sin\xi}{\xi}.$$

Notice that the Fourier transform of the impulse is not summable.

a. The Fourier transform in $\mathcal{S}(\mathbb{R}^n)$

The space $\mathcal{S}(\mathbb{R}^n)$ of *rapidly decreasing functions* is defined as the space of functions $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$\sup_{x\in\mathbb{R}^n} |x^\alpha D^\beta f(x)| < +\infty$$

for all multiindices α and β. Clearly $\mathcal{S}(\mathbb{R}^n)$ is a vector space, and $x^\alpha f(x) \in \mathcal{S}(\mathbb{R}^n)$ and $D^\alpha f \in \mathcal{S}(\mathbb{R}^n)$ for all α; moreover, $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$ and for every $f \in \mathcal{S}(\mathbb{R}^n)$, there exists a constant $C > 0$ such that

$$|f(x)| \le C\,(1 + |x|)^{-(n+1)} \qquad \forall x \in \mathbb{R}^n. \tag{1.35}$$

1.41 Definition. *For $f \in \mathcal{S}(\mathbb{R}^n)$ we call* Fourier transform *$\widehat{f}$ of f the well-defined function*

$$\widehat{f}(\xi) := \int_{\mathbb{R}^n} f(x) e^{-i\,x\,\bullet\,\xi}\,dx, \qquad \xi \in \mathbb{R}^n.$$

We leave to the reader the task of proving the following two propositions.

1.42 Proposition. *For all $f \in \mathcal{S}(\mathbb{R}^n)$ and all multiindices α*

(i) *the Fourier transform of $D^\alpha f(x)$ is $(i\xi)^\alpha \widehat{f}(\xi)$,*
(ii) *the Fourier transform of $x^\alpha f(x)$ is $(iD)^\alpha \widehat{f}(\xi)$.*

In particular, $\widehat{f} \in \mathcal{S}(\mathbb{R}^n)$.

1.43 Proposition. *If f and $g \in \mathcal{S}(\mathbb{R}^n)$, then the convolution*

$$f * g(x) := \int_{\mathbb{R}^n} f(x-y)g(y)\,dy$$

of f and g is a rapidly decreasing function and $\widehat{f * g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi)$ $\forall \xi \in \mathbb{R}^n$.

1.44 ¶. Prove that for $a > 0$ we have

$$\widehat{f}(\xi) = \frac{1}{a + i\xi} \qquad \text{if } f(x) = \begin{cases} 0 & \text{if } x < 0, \\ e^{-ax} & \text{if } x > 0, \end{cases}$$

$$\widehat{f}(\xi) = \frac{2a}{a^2 + \xi^2} \qquad \text{if } f(x) = e^{-a|x|},$$

$$\widehat{f}(\xi) = \frac{-2i\xi}{a^2 + \xi^2} \qquad \text{if } f(x) = \begin{cases} -e^{ax} & \text{if } x < 0, \\ e^{-ax} & \text{if } x > 0, \end{cases}$$

$$\widehat{f}(\xi) = (2\pi)^{n/2} e^{-|\xi|^2/2} \qquad \text{if } f(x) = e^{-|x|^2/2}.$$

[*Hint.* Concerning the last claim: From $D_j f(x) = -x_j f(x)$ infer that $D_j \widehat{f}(\xi) = -\xi_j \widehat{f}(\xi)$, $j = 1, \dots, n$, and integrate.]

We have the following *inversion formula*.

1.45 Theorem (Fourier's inversion formula). *The Fourier transform is a linear automorphism of* $\mathcal{S}(\mathbb{R}^n)$*. Its inverse, called the* inverse Fourier transform*, is given by*

$$f(x) := (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i\,x\,\bullet\,\xi}\, \widehat{f}(\xi)\, d\xi. \tag{1.36}$$

Proof. We need to compute

$$\int_{\mathbb{R}^n} e^{i\,x\,\bullet\,\xi} \left(\int_{\mathbb{R}^n} f(y) e^{-i\,y\,\bullet\,\xi}\, dy \right) d\xi.$$

Since the double integral is not absolutely convergent, we are not allowed to change the order of integration. For this reason we proceed as follows: We choose $\psi \in \mathcal{S}(\mathbb{R}^n)$ with $\psi(0) = 1$ and we compute, using the Lebesgue dominated convergence theorem and (1.35),

$$\begin{aligned}
\int_{\mathbb{R}^n} \widehat{f}(\xi) e^{i\,x\,\bullet\,\xi}\, d\xi &= \lim_{\epsilon \to 0} \int_{\mathbb{R}^n} \psi(\epsilon\xi) \widehat{f}(\xi) e^{i\,x\,\bullet\,\xi}\, d\xi \\
&= \lim_{\epsilon \to 0} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \psi(\epsilon\xi) f(y) e^{i\,x-y\,\bullet\,\xi}\, d\xi \\
&= \lim_{\epsilon \to 0} \int_{\mathbb{R}^n} f(y) \widehat{\psi}\Big(\frac{y-x}{\epsilon}\Big) \frac{1}{\epsilon^n}\, dy \\
&= \lim_{\epsilon \to 0} \int_{\mathbb{R}^n} f(x+\epsilon z) \widehat{\psi}(z)\, dz = f(x) \int_{\mathbb{R}^n} \widehat{\psi}(z)\, dz.
\end{aligned}$$

□

1.46 Remark. The inversion formula (1.36) now states (not heuristically) that every $f \in \mathcal{S}(\mathbb{R}^n)$ is the superposition of a continuum of *plane waves* $\widehat{f}(\xi) e^{i\,x\,\bullet\,\xi}$, $\xi \in \mathbb{R}^n$, each with velocity of propagation ξ and amplitude $\widehat{f}(\xi)$.

Notice that the wave $e^{i\,x\,\bullet\,\xi}$ is up to a constant the eigenfunction of the differentiation operator $\mathbf{D}$ associated to the purely imaginary eigenvalue $i\xi$. In fact, if $f \in C^1(\mathbb{R}^n, \mathbb{C})$ is such that $\mathbf{D}f(x) = i\xi f(x)$, then $f(x) = Ce^{i\,x\,\bullet\,\xi}$.

1.47 Example (Heat equation). Consider once more in $\mathbb{R}^n \times \mathbb{R}$ Cauchy's problem for the heat equation

$$\begin{cases} u_t(x,t) = k\Delta u(x,t) & \text{on } \mathbb{R}^n \times]0, +\infty[, \\ u(x,0) = f(x) & \forall x \in \mathbb{R}^n, \end{cases} \tag{1.37}$$

where we assume $f \in \mathcal{S}(\mathbb{R}^n)$ and $u(\cdot, t) \in \mathcal{S}(\mathbb{R}^n)$ for all t. By taking the Fourier transformation of u with respect to x, (1.37) becomes

$$\begin{cases} \dfrac{\partial \widehat{u}}{\partial t}(\xi,t) = -k|\xi|^2 \widehat{u}(\xi,t) & (\xi,t) \in \mathbb{R}^n \times]0, \infty[, \\ \widehat{u}(\xi,0) = \widehat{f}(\xi) & \forall \xi \in \mathbb{R}^n, \end{cases}$$

hence

$$\widehat{u}(\xi,t) = \widehat{f}(\xi)e^{-tk|\xi|^2}.$$

By setting

$$g_{(t)}(x) = g(x,t) := (4\pi kt)^{-n/2}\exp\Big(-\frac{|x|^2}{4kt}\Big),$$

we see that $\widehat{g_{(t)}}(\xi) = e^{-tk|\xi|^2}$, hence

$$\widehat{u}(\xi,t) = \widehat{f}(\xi)\widehat{g_{(t)}}(\xi) = \widehat{f * g_{(t)}}(\xi),$$

therefore

$$\begin{aligned} u(x,t) = (f * g_{(t)})(x) &= (4\pi kt)^{-n/2}\int_{\mathbb{R}^n} f(x-y)e^{-\frac{|y|^2}{4kt}}\,dy \\ &= \pi^{-n/2}\int_{\mathbb{R}^n} f(x-2\sqrt{kt}y)e^{-|y|^2}\,dy. \end{aligned} \tag{1.38}$$

If $f \geq 0$ has compact support and $x \in \mathbb{R}^n$ and $t > 0$, clearly we can find $y \in \mathbb{R}^n$ such that $x - 2\sqrt{kt}y$ is in the support of f, i.e., $f(x-2\sqrt{kt}y) > 0$: The last integral in (1.38) tells us that $u(x,t) > 0\ \forall t > 0$, although $u(x,0)$ may vanish. One says that the velocity of propagation of the data is infinite.

b. The Fourier transform in L^2

It is also easy to check the following equalities, the second of which is known as *Parseval's formula*.

1.48 Proposition. *Let ϕ, ψ be in $\mathcal{S}(\mathbb{R}^n)$. Then*

(i) $\int_{\mathbb{R}^n} \widehat{\phi}\,\psi\,dx = \int_{\mathbb{R}^n} \phi\,\widehat{\psi}\,dx$,

(ii) $\int_{\mathbb{R}^n} \phi\,\overline{\psi}\,dx = (2\pi)^{-n}\int_{\mathbb{R}^n} \widehat{\phi}\,\overline{\widehat{\psi}}\,dx$,

(iii) $\widehat{\phi * \psi} = \widehat{\phi}\,\widehat{\psi}$,

(iv) $\widehat{\phi\psi} = (2\pi)^{-n}\widehat{\phi} * \widehat{\psi}$.

Let $f \in L^2(\mathbb{R}^n)$ and let $\{f_k\}$ be a sequence in $C_c^0(\mathbb{R}^n)$ with $f_k \to f$ in $L^2(\mathbb{R}^n)$. Parseval's formula yields

$$||\widehat{f_j} - \widehat{f_k}||_{L^2} = (2\pi)^{-n}||f_j - f_k||_{L^2}.$$

Therefore, the sequence $\{\widehat{f_k}\}$ is a Cauchy sequence and converges in L^2 to some function $\widehat{f}$ that is easily seen to be independent on the sequence that approximates f. We again call $\widehat{f}$ the Fourier transform of $f \in L^2(\mathbb{R}^n)$. In other words, Parseval's formula allows us to extend by continuity the operator

$$\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n), \qquad \mathcal{F}(f)(\xi) := \widehat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-i\,x\,\bullet\,\xi}\,dx,$$

to a continuous operator $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. If we denote $\widehat{f} := \mathcal{F}(f)$, clearly the claims of Proposition 1.48 still hold; in particular, the following identity, called *formula of Plancherel*, holds: For all $f \in L^2(\mathbb{R}^n)$ we have

$$\int_{\mathbb{R}^n} |f(x)|^2\,dx = (2\pi)^{-n} \int_{\mathbb{R}^n} |\widehat{f}(\xi)|^2\,d\xi,$$

i.e., $\mathcal{F}$ *is an isometry of* $L^2(\mathbb{R}^n)$ *with respect to the norm*

$$\frac{1}{(2\pi)^{n/4}} ||f||_{L^2(\mathbb{R}^n)}.$$

1.49 The principle of indeterminacy. For the sake of simplicity, suppose that f has compact support and belongs to $L^2(\mathbb{R})$. The two quantities

$$E_x^2 := \frac{\int_{\mathbb{R}} |x|^2 |f(x)|^2\,dx}{\int_{\mathbb{R}} |f(x)|^2\,dx}, \qquad E_\xi^2 := \frac{\int_{\mathbb{R}} |\xi|^2 |\widehat{f}(\xi)|^2\,d\xi}{\int_{\mathbb{R}} |\widehat{f}(\xi)|^2\,d\xi}$$

are the expected values of x and ξ with respect to the densities $f/||f||_{L^2}$ and $\widehat{f}/||\widehat{f}||_{L^2}$.

Proposition. *We have*

$$E_x\, E_\xi \geq \frac{1}{2}.$$

Proof. From Plancherel's formula

$$2\pi \int_{\mathbb{R}} |f(x)|^2\,dx = \int_{\mathbb{R}} |\widehat{f}(\xi)|^2\,d\xi,$$

and the identities

$$2\pi \int_{\mathbb{R}} |f'(x)|^2\,dx = \int_{\mathbb{R}} |\xi|^2\, |\widehat{f}(\xi)|^2\,d\xi$$

and

$$\int_{\mathbb{R}} x\, f(x) f'(x)\,dx = x \frac{f^2(x)}{2}\bigg|_{-\infty}^{+\infty} - \int_{\mathbb{R}} \frac{f^2(x)}{2}\,dx = -\frac{1}{2} \int_{\mathbb{R}} |f(x)|^2\,dx$$

we see that

$$\begin{aligned}
\left(\frac{1}{2} \int_{\mathbb{R}} |f(x)|^2\,dx\right)\left(\frac{1}{2} \int_{\mathbb{R}} |\widehat{f}(\xi)|^2\,d\xi\right) &= 2\pi \left(\frac{1}{4} \int_{\mathbb{R}} |f(x)|^2\,dx\right)^2 \\
&= 2\pi \left(\int_{\mathbb{R}} x f(x) f'(x)\,dx\right)^2 \\
&\leq 2\pi \left(\int_{\mathbb{R}} |x|^2\, |f(x)|^2\,dx\right)\left(\int_{\mathbb{R}} |f'(x)|^2\,dx\right) \\
&= \left(\int_{\mathbb{R}} |x|^2\, |f(x)|^2\,dx\right)\left(\int_{\mathbb{R}} |\xi|^2\, |\widehat{f}(\xi)|^2\,d\xi\right),
\end{aligned}$$

i.e.,

$$E_x\, E_\xi \geq \frac{1}{2}.$$

□

1.3 Sobolev Spaces

The theory of Sobolev spaces plays a fundamental role in the study of partial differential equations. Here we confine ourselves to illustrating a few definitions and some basic facts that will allow us to substantiate the Dirichlet principle.

a. Strong derivatives

Let $\Omega \subset \mathbb{R}^n$ and $1 \le p < +\infty$. We say that $u \in L^p(\Omega)$ has *strong derivatives* $v_1, v_2, \dots, v_n \in L^p(\Omega)$ if there exists a sequence $\{u_n\}$ of functions in $C^1(\Omega) \cap L^p(\Omega)$ such that $u_n \to u$ in $L^p(\Omega)$ and $D_i u_n \to v_i$ in $L^p(\Omega)$ for all $i = 1, \dots, n$. Proposition 1.50 shows that the strong derivatives of u, if they exist, are unique and depend only on u and not on the approximating smooth sequence used to define them. For this reason we denote them by $Du = (D_1 u, D_2 u, \dots, D_n u)$.

1.50 Proposition. *Let $\{u_n\}$ and $\{v_n\}$ be two sequences of functions in $C^1(\Omega) \cap L^p(\Omega)$ converging to $u \in L^p(\Omega)$. If $Du_n \to g$ and $Dv_n \to h$ in $L^1_{loc}(\Omega)$, then $g = h$ a.e.*

Proof. Let $\varphi \in C_c^\infty(\Omega)$. By Gauss–Green formulas

$$\int_\Omega (D_i u_n - D_i v_n)\varphi \, dx = -\int_\Omega (u_n - v_n) D_i \varphi \, dx,$$

hence

$$\left| \int_\Omega (D_i u_n - D_i v_n)\varphi \, dx \right| \le ||u_n - v_n||_{L^p} |||D\varphi|||_{L^q},$$

where $\frac{1}{q} = 1 - \frac{1}{p}$ if $p > 1$ and $q = \infty$ if $p = 1$. Taking the limit we conclude

$$\int_\Omega (g - h)\, \varphi \, dx = 0 \qquad \forall \varphi \in C_c^\infty(\Omega).$$

The claim then follows from the following lemma. □

The following lemma is often referred to as to the *fundamental lemma of the calculus of variations.*

1.51 Lemma. *Let $u \in L^1_{loc}(\Omega)$. If $\int_\Omega u\varphi \, dx = 0$ for all $\varphi \in C_c^\infty(\Omega)$, then $u = 0$ a.e. in Ω.*

Proof. First suppose that u is continuous and that $u(x_0) > 0$ for some $x_0 \in \Omega$. Then there is $\delta > 0$ such that $u > u(x_0)/2$ in $B(x_0, \delta)$. If $\varphi \in C_c^\infty(B(x_0, \delta))$ is nonnegative and has nonzero integral, then

$$0 = \int_\Omega u\varphi \, dx > \frac{u(x_0)}{2} \int_{B(x_0,\delta)} \varphi(x) \, dx \neq 0,$$

a contradiction. If u is just in $L^1_{loc}(\Omega)$, we extend it to be zero outside Ω and choose a symmetric regularization kernel ρ. The function $u * \rho_\epsilon$ is in $C^\infty(\mathbb{R}^n)$ and for all $\varphi \in C_c^\infty(\Omega)$ and $\epsilon << 1$, $\varphi * \rho_\epsilon$ is still in $C_c^\infty(\mathbb{R}^n)$ and

$$\int_{\mathbb{R}^n} (u * \rho_\epsilon)\, \varphi \, dx = \int_{\mathbb{R}^n} u\, (\varphi * \rho_\epsilon) \, dx = 0.$$

Thus $u * \rho_\epsilon = 0$ in Ω; consequently, $u = 0$ a.e. since $u * \rho_\epsilon \to u$ in L^1_{loc}. □

It is convenient to state a variant of Lemma 1.51 which will be very useful in the sequel.

1.52 Lemma (du Bois–Reymond). *Let $u \in L^1_{loc}(\Omega)$ where $\Omega \subset \mathbb{R}^n$ is a connected open set. If*

$$\int_\Omega u D_i\varphi \, dx = 0 \qquad \forall i = 1, \dots, n, \ \forall \varphi \in C_c^\infty(\Omega),$$

then u is constant a.e. in Ω.

Proof. Let $u \in C^1(\Omega)$. We have, integrating by parts,

$$\int_\Omega D_i u \, \varphi \, dx = -\int_\Omega u \, D_i\varphi \, dx = 0 \qquad \forall \varphi \in C_c^\infty(\Omega).$$

Lemma 1.51 then yields $Du = 0$, i.e., u constant in Ω.

Let $u \in L^1_{loc}(\Omega)$. Again, we extend u to be zero outside Ω and choose a symmetric regularizing kernel ρ. The function $u_\epsilon := u * \rho_\epsilon$ is then in $C^\infty(\mathbb{R}^n)$, and, for all $\varphi \in C_c^\infty(\mathbb{R}^n)$ and $\epsilon << 1$, $\varphi_\epsilon := \varphi * \rho_\epsilon$ is in $C_c^\infty(\mathbb{R}^n)$, too, and $(D_i\varphi) * \rho_\epsilon = D_i\varphi_\epsilon$. Now we compute

$$\int_\Omega D_i u_\epsilon \, \varphi \, dx = -\int_\Omega u_\epsilon D_i\varphi \, dx = -\int_\Omega u \, (D_i\varphi) * \rho_\epsilon \, dx = -\int_\Omega u \, D_i\varphi_\epsilon \, dx = 0.$$

Consequently, u_ϵ is constant in Ω, since it is of class C^1. It follows that u is constant a.e. in Ω since $u_\epsilon \to u$ in L^1_{loc}. □

Clearly, every function $u \in C^1(\Omega) \cap L^p(\Omega)$ with $Du \in L^p(\Omega)$ has strong derivatives in L^p that coincide with the classical derivatives. By approximation it is easily seen that the following holds:

(i) If $u, v \in L^p(\Omega)$ have strong derivatives in $L^p(\Omega)$, then also $u + v$ and λu for all $\lambda \in \mathbb{R}$ have strong derivatives in $L^p(\Omega)$.
(ii) If u has strong derivatives in $L^p(\Omega)$ and v has strong derivatives in $L^q(\Omega)$ where $p, q > 1$ and $1/p + 1/q = 1$, then uv has strong derivatives in $L^1(\Omega)$ and $D(uv) = vDu + uDv$.
(iii) If $f \in C^1(\mathbb{R})$ is bounded and u has strong derivatives in L^p, then $v := f(u)$ has strong derivatives in L^p and $D(f(u)) = f'(u)Du$.
(iv) If $u \in L^p(\Omega)$ has strong derivatives in L^p and $\varphi \in C_c^1(\Omega)$, then

$$\int_\Omega D_i u \, \varphi \, dx = -\int_\Omega u \, D_i\varphi \, dx. \tag{1.39}$$

1.53 Definition. *The Sobolev space $H^{1,p}(\Omega)$, $1 \le p < +\infty$, is the subspace of $L^p(\Omega)$ given by*

$$H^{1,p}(\Omega) := \left\{ u \in L^p(\Omega) \,\middle|\, u \text{ has strong derivatives in } L^p(\Omega) \right\}$$

and the map $u \to ||u||_{1,p}$ defined by

$$||u||_{1,p}^p := \int_\Omega (|u|^p + |Du|^p) \, dx \tag{1.40}$$

is a norm on $H^{1,p}(\Omega)$. *The closure of* $C_c^\infty(\Omega)$ *(with respect to the* $||\ ||_{1,p}$ *norm) is denoted by* $H_0^{1,p}(\Omega)$. *For* $p = 2$, $H^{1,2}$ *and* $H_0^{1,2}$ *are pre-Hilbert spaces with respect to the inner product*

$$(u|v)_{1,2} := \int_\Omega (uv + (Du|Dv))\, dx. \tag{1.41}$$

$H^{1,2}(\Omega)$ *and* $H_0^{1,2}(\Omega)$ *are often abbreviated as* $H^1(\Omega)$ *and* $H_0^1(\Omega)$.

1.54 Theorem. $H^{1,p}(\Omega)$ *and* $H_0^{1,p}(\Omega)$ *endowed with the norm defined by* (1.40) *are Banach spaces. In particular, for* $p = 2$, $H^{1,2}(\Omega)$ *and* $H_0^{1,2}(\Omega)$ *are Hilbert spaces with respect to the inner product in* (1.41).

Proof. Let $\{u_n\} \subset H^{1,p}(\Omega)$ be a Cauchy sequence with respect to the $||\ ||_{1,p}$ norm. Then $\{u_n\}$ and $\{Du_n\}$ are Cauchy's sequences in L^p, hence there exist u and $g \in L^p$ such that $u_n \to u$ and $Du_n \to g$. By a diagonal process, we find a sequence $\{v_n\}$ of functions of class $C^1(\Omega) \cap L^p(\Omega)$ such that $v_n \to u$ and $Dv_n \to g$ in $L^p(\Omega)$. Hence $u \in H^{1,p}(\Omega)$ and $Du = g$. Finally, $H_0^{1,p}(\Omega)$ is also a Banach space since it is a closed linear subspace of $H^{1,p}(\Omega)$. □

b. Weak derivatives

Let $\Omega \subset \mathbb{R}^n$ and $u \in L^1_{loc}(\Omega)$. We say that u has *weak derivatives* $v_1, v_2, \dots, v_n \in L^1_{loc}(\Omega)$ if for all $i = 1, \dots, n$ we have

$$\int_\Omega u D_i\varphi\, dx = -\int_\Omega v_i\varphi\, dx \qquad \forall \varphi \in C_c^\infty(\Omega). \tag{1.42}$$

If $u \in C^1(\Omega)$, then $u\varphi \in C_c^1(\Omega)$, and the Gauss–Green formulas allow us to conclude that the weak derivatives of u exist and are the classical derivatives of u. Formula (1.39) shows that the weak derivatives of a function $u \in H^{1,p}(\Omega)$ exist and coincide with the corresponding strong derivatives. Finally, it follows from Lemma 1.51 that the weak derivatives, if they exist, are uniquely defined by u via (1.42). For these reasons, also the weak derivatives of u, if they exist, are denoted by $D_1u, D_2u, \dots, D_nu$.

1.55 Definition. *We say that a function* $u \in L^p(\Omega)$ *is in the class* $W^{1,p}(\Omega)$ *if* u *has weak derivatives in* $L^p(\Omega)$. *The closure of* $C_c^1(\Omega)$ *in* $W^{1,p}(\Omega)$ *is denoted by* $W_0^{1,p}(\Omega)$.

Let ρ be a symmetric smoothing kernel, $\varphi \in C_c^1(\Omega)$, where Ω is an open set in $\mathbb{R}^n$, and, as usual, for $\epsilon > 0$ set $\varphi_\epsilon(x) := \epsilon^{-n}\varphi(x/\epsilon)$. Let us recall, see Proposition 2.47 of [GM4], that for $\Omega_\epsilon := \{x \in \Omega \mid \operatorname{dist}(x, \partial\Omega) > \epsilon\}$ we have

$$(D_i\varphi) * \rho_\epsilon(x) = D_i(\varphi * \rho_\epsilon)(x) \qquad \forall x \in \Omega_\epsilon,$$

and for $f \in L^1(\Omega)$

$$\int_\Omega f(x)\, (\varphi * \rho_\epsilon)(x)\, dx = \int_\Omega (f * \rho_\epsilon)(x)\, \varphi(x)\, dx$$

if φ has support in $\Omega_{2\epsilon}$.

Now suppose $u \in W^{1,p}(\Omega)$. Then for all $\varphi \in C_c^1(\Omega)$ and all ϵ sufficiently small we have

$$\begin{aligned}\int_\Omega ((D_i u) * \rho_\epsilon)\varphi \, dx &= \int_\Omega D_i u(\varphi * \rho_\epsilon) \, dx = -\int_\Omega u(D_i(\varphi * \rho_\epsilon)) \, dx \\ &= -\int_\Omega u((D_i\varphi) * \rho_\epsilon) \, dx = -\int_\Omega (u * \rho_\epsilon) D_i\varphi \, dx \\ &= \int_\Omega (D_i(u * \rho_\epsilon))\varphi \, dx,\end{aligned}$$

hence

$$D(u * \rho_\epsilon)(x) = (Du) * \rho_\epsilon(x) \qquad \text{if } x \in \Omega_{2\epsilon}.$$

From the convergence properties in L^p of the mollified sequence, we therefore infer that for any open set $\widetilde{\Omega} \subset\subset \Omega$, the mollified sequence $\{u_\epsilon\}$ converges to u in $W^{1,p}(\widetilde{\Omega})$.

1.56 ¶. Prove that $H_0^1(\Omega) = W_0^1(\Omega)$.

1.57 Theorem (Meyers–Serrin). *$H^{1,p}(\Omega) = W^{1,p}(\Omega)$. This means that for every $u \in W^{1,p}(\Omega)$ there exists a sequence $\{u_n\}$ in $W^{1,p}(\Omega) \cap C^1(\Omega)$ such that $u_n \to u$ in $W^{1,p}(\Omega)$.*

The reader can find the proof of this result in any of the many books on Sobolev spaces. Here we prove a stronger result for a restricted class of domains.

1.58 Theorem. *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set that is star-shaped with respect to one of its points. Then $C^\infty(\overline{\Omega})$ is dense in $W^{1,p}(\Omega)$.*

Proof. Suppose that Ω is star-shaped with respect to the origin and, for $0 < \tau < 1$, set

$$u_\tau(x) := u(\tau x) \qquad \text{and} \qquad \tau^{-1}(\Omega) = \Big\{ y = \tau^{-1}x \,\Big|\, x \in \Omega \Big\}.$$

According to the definition of the weak derivative, $u_\tau \in W^{1,p}(\tau^{-1}\Omega)$ and $Du_\tau(x) = \tau Du(\tau x)$. Moreover,

$$||D(u - u_\tau)||_{L^p(\Omega)} \le (1 - \tau)||Du||_{L^p} + ||Du - (Du)_\tau||_{L^p}.$$

Hence $u_\tau \to u$ in $W^{1,p}(\Omega)$ as $\tau \to 1$ because of the continuity in the mean, see Proposition 1.26. Mollifying u_τ with a mollifying parameter $\epsilon = \epsilon(\tau)$ sufficiently small, the claim follows at once. □

c. Absolutely continuous functions

Let $I \subset \mathbb{R}$ be an interval of $\mathbb{R}$, $I =]a, b[$.

1.59 Theorem. *Let $u \in H^{1,p}(]a,b[)$, $p \geq 1$. There is a continuous representative $\widetilde{u} : [a,b] \to \mathbb{R}$ of u, that is,*

$$\widetilde{u}(x) - \widetilde{u}(y) = \int_x^y u'(s)\, ds \qquad \forall x, y \in [a,b]$$

where u' denotes the weak derivative of u and $u = \widetilde{u}$ a.e. in $[a,b]$. Moreover, if $p > 1$, then $\widetilde{u}$ is Hölder-continuous with exponent $\alpha := 1 - 1/p$.

Proof. Consider a sequence $\{u_k\} \in C^1([a,b])$ that converges to u in $H^{1,p}(]a,b[)$, see Theorem 1.58. The fundamental theorem of calculus yields

$$u_k(y) - u_k(x) = \int_x^y u_k'(s)\, ds \qquad \forall x, y \in [a,b]. \tag{1.43}$$

It follows that the sequence $\{u_k\}$ is equibounded and equicontinuous in $C^0([a,b])$. On account of the Ascoli–Arzelà theorem, see [GM3], a subsequence $\{u_{k_n}\}$ of $\{u_k\}$ converges uniformly in $[a,b]$ to a continuous function $\widetilde{u} : [a,b] \to \mathbb{R}$. The first part of the claim follows by letting $n \to \infty$ in (1.43) with $k = k_n$.

If, moreover, $p > 1$, because of Hölder's inequality we have

$$|\widetilde{u}(x) - \widetilde{u}(y)| = \left| \int_x^y u'(s)\, ds \right| \leq \left(\int_a^b |u'(s)|^p\, ds \right)^{1/p} |x - y|^{1-1/p}$$

for all $x, y \in [a,b]$. □

More precisely, a celebrated theorem due to Giuseppe Vitali (1875–1932), see Theorem 6.52 and Proposition 6.55, states the following.

1.60 Theorem. *Let $u \in L^1(]a,b[)$. Then $u \in H^{1,1}(]a,b[)$ if and only if u has an absolutely continuous representative defined on $[a,b]$. Moreover, if $u' \in L^1$ is the weak derivative of u, then*

$$u'(s) = \lim_{h \to 0} \frac{\widetilde{u}(s+h) - \widetilde{u}(s)}{h} \qquad \text{for a.e. } s \in]a,b[.$$

d. H^1-periodic functions

As a consequence of the Riesz–Fischer theorem and the completeness of the trigonometric system in L^2 we may state the following proposition.

1.61 Proposition. *A function u belongs to $H^{1,2}(]-\pi,\pi[)$ if and only if there exist two sequences of real numbers $\{a_k\}$ and $\{b_k\}$ with $\sum_{k=1}^{\infty}(1 + k^2)(a_k^2 + b_k^2) < +\infty$ such that*

$$u(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \qquad \text{in } L^2(]-\pi,\pi[). \tag{1.44}$$

In this case, we have

$$u'(x) = \sum_{k=1}^{\infty} (kb_k \cos kx - ka_k \sin kx) \qquad \text{in } L^2(]-\pi,\pi[),$$
$$a_k := \frac{1}{\pi}\int_{-\pi}^{\pi} u(t)\cos kt\,dt, \qquad b_k := \frac{1}{\pi}\int_{-\pi}^{\pi} u(t)\sin kt\,dt$$

and,

$$||u||_2^2 = \pi\Big(\frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2)\Big) \qquad \text{and} \qquad ||u'||_2^2 = \pi\sum_{k=1}^{\infty} k^2(a_k^2 + b_k^2).$$

1.62 The isoperimetric inequality in the plane. We have seen in [GM1] that Steiner's argument allows us to prove that the circle is the unique curve of prescribed length enclosing maximal area. We present here an analytic proof due to Adolf Hurwitz (1859–1919).

Let C be a continuous closed curve that is piecewise of class C^1, of finite length L and parametrized by $\gamma(t) := (x(t), y(t))$, where $t = 2\pi s/L \in [0, 2\pi]$, and s be its arclength. Because of the choice of parametrization, we have

$$\sqrt{x'(t)^2 + y'^2(t)} = \frac{L}{2\pi} \qquad \forall t \text{ for which } \gamma'(t) \text{ is defined.}$$

Since $x(t)$, $y(t) \in H^{1,2}(I)$, if a_k, b_k and A_k, B_k denote respectively the Fourier coefficients of $x(t)$ and $y(t)$, we infer from Proposition 1.61 that

$$\frac{L^2}{2\pi} = \int_0^{2\pi} (x'^2(t) + y'^2(t))\,dt = \pi\sum_{k=1}^{\infty} k^2(a_k^2 + b_k^2 + A_k^2 + B_k^2).$$

On the other hand, we may compute the area enclosed by C by means of Stokes's formula in the plane, see [GM4], and, from Proposition 1.61, we find

$$A := \{\text{enclosed area}\} = -\int_C y\,dx = -\int_0^{2\pi} y(t)x'(t)\,dt$$
$$= -\pi\sum_{k=1}^{\infty} k(A_k b_k - a_k B_k).$$

In conclusion,

$$L^2 - 4\pi A = 2\pi^2\sum_{k=1}^{\infty}\Big((ka_k - B_k)^2 + (kb_k + A_k)^2 + (k^2-1)(A_k^2 + B_k^2)\Big);$$

in particular, we get the *isoperimetric inequality*

$$L^2 \geq 4\pi A,$$

where equality $L^2 = 4\pi A$ holds if and only if

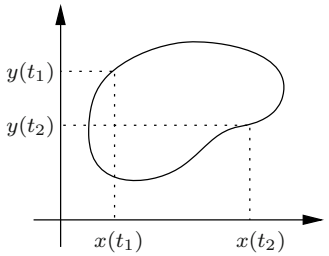


Figure 1.2. Arc-length parametrization is Lipschitz-continuous.

$$\begin{cases} ka_k - B_k = 0, \\ kb_k + A_k = 0, \\ a_k = b_k = A_k = B_k = 0 \quad \forall k \geq 2. \end{cases}$$

In other words, the equality $L^2 = 4\pi A$ holds if and only if

$$x(t) = \frac{a_0}{2} + a_1 \cos t + b_1 \sin t, \qquad y(t) = \frac{A_0}{2} - b_1 \cos t + a_1 \sin t,$$

that is,

$$\left(x(t) - \frac{a_0}{2}\right)^2 + \left(y(t) - \frac{A_0}{2}\right)^2 = \frac{L^2}{4\pi^2},$$

i.e., if and only if C is the circle of center $(a_0/2, A_0/2)$ and radius $\frac{L}{2\pi}$.

1.63 Remark. The previous proof shows that the circle encloses maximal area among all closed curves of the same length that are piecewise of class C^1. Actually, the proof generalizes to show that indeed the circle encloses maximal area among all continuous curves of finite length.

Let C be a continuous and closed curve of finite length L. As we know, see [GM3], we may reparametrize C by means of the arc-length parameter s and the resulting parametrization $\gamma(s) = (x(s), y(s))$, $s \in [0, L]$, is Lipschitz-continuous, since

$$|x(s_2) - x(s_1)| \leq \overline{P_1P_2} \leq |s_2 - s_1|,$$
$$|y(s_2) - y(s_1)| \leq \overline{P_1P_2} \leq |s_2 - s_1|,$$

see Figure 1.2.; consequently its components are absolutely continuous. By Vitali's theorem $x(s)$ and $y(s)$ are in $H^{1,1}([0, L])$, and the weak derivatives $x'(s)$ and $y'(s)$ are the classical derivatives of x and y in a.e. point. In particular, $|x'(s)|, |y'(s)| \leq 1$ for a.e. s and $x(s)$ and $y(s)$ belong to $H^{1,2}([0, L])$. Moreover, by Tonelli's theorem, Theorem 6.56, we have $x'^2 + y'^2 = 1$ for a.e. s. In terms of the original parametrization $(x(t), y(t))$ of C with $t = 2\pi s/L \in [0, 2\pi]$ of C, we have $x(t)$ and $y(t) \in H^{1,2}([0, 2\pi])$ and

$$x'(t)^2 + y'(t)^2 = \frac{L^2}{4\pi} \qquad \text{for a.e. } t \in [0, 2\pi].$$

From this point on, we can repeat word by word the argument in 1.62 for piecewise-C^1 curves to conclude.

e. Poincaré's inequality

1.64 Theorem (Poincaré's inequality). *Let Ω be a bounded open set of $\mathbb{R}^n$, $n \geq 1$. For all $u \in H_0^{1,p}(\Omega)$ we have*

$$\int_\Omega |u|^p \, dx \leq (\operatorname{diam}\Omega)^p \int_\Omega |Du|^p \, dx. \tag{1.45}$$

In particular,

$$|u|_{1,p} := \left(\int_\Omega |Du|^p \, dx \right)^{1/p}$$

is an equivalent norm to $||u||_{1,p}$ in $H_0^{1,p}(\Omega)$,

$$\frac{1}{1 + (\operatorname{diam}\Omega)^p} ||u||_{1,p}^p \leq |u|_{1,p}^p \leq ||u||_{1,p}^p.$$

Proof. Since $C_c^\infty(\Omega)$ is dense in $H_0^{1,p}$, it suffices to prove the inequality for $u \in C_c^\infty(\Omega)$. Let $a > 0$ be such that

$$\Omega \subset \left\{ x = (x^1, x^2, \dots, x^n) \,\middle|\, -a < x^1 < a \right\}.$$

We have

$$\varphi(x) = \int_{-\infty}^{x^1} \frac{\partial \varphi}{\partial x_1}(\xi, x^2, \dots, x^n) \, d\xi$$

and, using Hölder's inequality,

$$|\varphi(x)|^p \leq (2a)^{p-1} \int_{-a}^{a} \left| \frac{\partial \varphi}{\partial x^1}(\xi, x^2, \dots, x^n) \right|^p d\xi \qquad \forall x \in \mathbb{R}^n.$$

Integrating first with respect to x^1 and then with respect to the other variables, we get (1.45). □

1.65 ¶. Poincaré's inequality is false if Ω is unbounded in one direction, see Figure 1.3.

A second Poincaré-type inequality, again called *Poincaré's inequality* or *Poincaré–Wirtinger's inequality*, is the following.

1.66 Proposition (Poincaré–Wirtinger's inequality). *Let $\Omega \subset \mathbb{R}^n$ be a cube (or a ball) in $\mathbb{R}^n$ and $p \geq 1$. There is a constant $c = c(n, p)$ such that*

$$\int_\Omega |u - u_\Omega|^p \, dx \leq c(n, p) \, (\operatorname{diam}\Omega)^p \int_\Omega |Du|^p \, dx \tag{1.46}$$

for all $u \in H^{1,p}(\Omega)$, where

$$u_\Omega := \frac{1}{|\Omega|} \int_\Omega u(x) \, dx =: \fint_\Omega u \, dx$$

denotes the mean value of u in Ω.

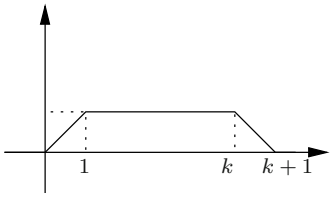


Figure 1.3. Poincaré's inequality does not hold in domains that are unbounded in a direction.

Proof. Let $\Omega := \left\{x = (x^1, x^2, \dots, x^n) \,|\, |x^i| \le a \ \forall i\right\}$ and $u \in C^1(\overline{\Omega})$. For $x, y \in \Omega$ we have

$$u(x) - u(y) = \int_{y^1}^{x^1} \frac{\partial u}{\partial x^1}(\xi, y^2, \dots, y^n) + \int_{y^2}^{x^2} \frac{\partial u}{\partial x^2}(x^1, \xi, y^3, \dots, y^n)\, d\xi + \dots$$
$$+ \int_{y^n}^{x^n} \frac{\partial u}{\partial x^n}(x^1, x^2, \dots, x^{n-1}, \xi)\, d\xi.$$

By taking the power p, using Hölder's inequality and integrating in $(x, y) \in \Omega \times \Omega \subset \mathbb{R}^n \times \mathbb{R}^n$, we get

$$\int_\Omega dx \int_\Omega |u(x) - u(y)|^p\, dy \le c(n,p)|\Omega| a^p \int_\Omega |Du|^p\, dx$$

from which the result follows, since

$$\int_\Omega |u - u_\Omega|^p\, dx = \frac{1}{|\Omega|^p} \int_\Omega dx \left| \int_\Omega (u(x) - u(y))\, dy \right|^p \le \frac{1}{|\Omega|} \int_\Omega dx \int_\Omega |u(x) - u(y)|^p\, dy.$$

□

f. Rellich's compactness theorem

The following theorem is a key result in the theory of Sobolev spaces.

1.67 Theorem (Rellich). *Let Ω be a bounded open set in $\mathbb{R}^n$ and $p \ge 1$. The embedding $j : H_0^{1,p}(\Omega) \to L^p(\Omega)$, $j(u) := u$, is compact. Consequently, for every $\widetilde{\Omega} \subset\subset \Omega$ the embedding $j : H^{1,p}(\Omega) \to L^p(\widetilde{\Omega})$, $j(u) = u_{|\widetilde{\Omega}}$ is compact, too.*

Proof. It suffices to prove that the embedding $j : H^{1,p}(Q) \to L^p(Q)$, $j(u) = u$, is compact if Q is a cube, since every function in $H_0^{1,p}(\Omega)$ extends with zero value to a function in $H_0^{1,p}(Q)$ preserving the norm, Q being a cube containing Ω. Finally, a covering argument in conjunction with a diagonal process then easily leads to the proof of the second part of the theorem.

Let us prove that the embedding $j : H^{1,p}(Q) \to L^p(Q)$, $j(u) = u$, is a compact operator. Let ℓ be the side of Q. Consider a subdivision in cubes $Q_1, Q_2, \dots, Q_s$ of Q with disjoint interiors and sides σ. Trivially,

$$|u_{Q_j}| = \left| \frac{1}{|Q_j|} \int_{Q_j} u(x)\, dx \right| \le \frac{c}{\sigma^n}.$$

Consider now the map $j_\sigma : H^{1,p}(Q) \to L^p(Q)$ defined by

$$j_\sigma(u)(x) := \sum_{j=1}^{s} u_{Q_j}\chi_{Q_j}(x)$$

where χ_{Q_j} denotes the characteristic function of Q_j. Clearly, j_σ is linear and compact, since its range is finite. Moreover, by Poincaré's inequality

$$||u - A_\sigma u||_{L^p}^p = \int_Q |u - A_\sigma(u)|^p\,dx = \sum_{j=1}^{s}\int_{Q_j} |u(x) - u_{Q_j}|^p\,dx$$

$$\leq \sigma^p \sum_{j=1}^{s}\int_{Q_j} |Du|^p\,dx = \sigma^p \int_Q |Du|^p dx \leq \sigma^p ||u||_{1,p,Q}^p,$$

hence

$$||j_\sigma - j||_{\mathcal{B}(H^{1,p},L^p)} := \sup_{||u||_{1,p,Q}\leq 1} |j_\sigma(u) - j(u)| \leq \sigma^p.$$

Therefore, the compact operators j_σ converge to j as bounded linear operators from $H^{1,p}(Q)$ into $L^p(Q)$ as $\sigma \to 0$. Theorem 9.140 of [GM3] then yields that j is compact. □

Another proof of Theorem 1.67. An alternative and more direct proof is the following. It suffices to prove that every sequence $\{u_k\} \subset H^{1,p}(Q)$ with $\sup_k ||u_k||_{1,p,Q} < \infty$ as a sequence $\{u_{k_n}\}$ which is convergent in L^p. Equivalently, it suffices to prove that the set $S := j(\{u_n\}) \subset L^p(Q)$ is relatively compact in L^p. Since $\overline{S}$ is a closed set of a Banach space, then $\overline{S}$ is complete. Let us prove that $j(S)$, and consequently $\overline{j(S)}$, are totally bounded. This implies that $\overline{j(S)}$ is compact, because of Theorem 6.15 of [GM3].

Let ℓ be the side of Q. Fix $\epsilon > 0$ and consider a subdivision in cubes $Q_1, Q_2, \ldots, Q_s$ of Q with disjoint interiors and sides σ with $\sigma < \epsilon$. Trivially,

$$|u_{k,Q_j}| = \left|\frac{1}{|Q_j|}\int_{Q_j} u_k(x)\,dx\right| \leq \frac{c}{\sigma^n}.$$

Next, consider the finite family $G \subset L^p(Q)$ of simple functions of the type

$$g(x) = n_1\epsilon\,\chi_{Q_1}(x) + \cdots + n_s\epsilon\,\chi_{Q_s}(x),$$

where $n_1, \ldots, n_s$ are integers in $]-N, N[$, $N > c/(\epsilon\sigma^n)$ and χ_{Q_j} denotes the characteristic function of Q_j. It suffices to show that each u_k has distance in $L^p(Q)$ less than ϵ from a suitable function $g \in G$. Define

$$u_k^*(x) := \sum_{j=1}^{s} u_{k,Q_j}\chi_{Q_j}(x).$$

By Poincaré's inequality we have

$$\int_Q |u_k - u_k^*|^p\,dx \leq \sum_{j=1}^{s}\int_{Q_j} |u_k - u_{k,Q_j}|^p\,dx \leq c(n)\sigma^p \sum_{j=1}^{s}\int_{Q_j} |Du_k|^p\,dx$$

$$\leq c(n)\sigma^p \int_Q |Du_k|^p \leq C\,\sigma^p.$$

On the other hand, according to the definition of G, there exists $g \in G$ such that

$$|g(x) - u_k^*(x)| < \epsilon \qquad \forall x \in Q,$$

hence

$$||u_k - g||_{L^p(Q)} \leq ||u_k - u_k^*||_{L^p(Q)} + ||u_k^* - g||_{L^p(Q)} \leq C\,\sigma^p + \ell^n\epsilon^p \leq C_1\,\epsilon^p.$$

□

1.68 Remark. We notice the following:

(i) The embedding $H^{1,p}(\Omega) \to L^p(\Omega)$ is not compact if Ω is unbounded: Think of $\Omega = \mathbb{R}$ and of a wave that moves toward infinity.

(ii) In general, the embedding $H^{1,p}(\Omega) \to L^p(\Omega)$ is not compact even if Ω is bounded (we shall not dwell on this); instead, it is compact if Ω is a bounded set with the the following *extension property*: Every function $u \in H^{1,p}(\Omega)$ extends to a function $\widetilde{u} \in H_0^{1,p}(\Lambda)$, $\Lambda \supset\supset \Omega$ with

$$||\widetilde{u}||_{1,p,\Lambda} \leq c\, ||u||_{1,p,\Omega}.$$

For instance, a star-shaped domain enjoys such a property, compare with Theorem 1.58.

g. Traces

Let Ω be a bounded open set satisfying the extension property, in such a way that for every $u \in H^{1,p}(\Omega)$ we can find a sequence of smooth functions $\{u_n\}$, defined in an open set that strictly contains $\overline{\Omega}$ and converging to u in $H^{1,p}(\Omega)$. In this case one can show that there exists a linear operator $T : H^{1,p}(\Omega) \to L^p(\partial\Omega)$ that is continuous,

$$||Tu||_{L^p(\partial\Omega)} \leq ||u||_{1,p}$$

and that agrees with the trace operator $Tu = u_{|\partial\Omega}$ on functions $C^0(\overline{\Omega}) \cap H^{1,p}(\Omega)$. The operator T, called the *trace operator*, extends the usual restriction operator to $\partial\Omega$. Moreover, by approximating $u \in H^{1,p}(\Omega)$ with a sequence of functions of class $C^1(\Lambda)$, $\Lambda \supset\supset \Omega$, it is not difficult to show that the Gauss–Green formulas for the approximating functions

$$\int_\Omega D_i u_k \, dx = \int_{\partial\Omega} u_k(x)\, \nu_\Omega^i(x)\, d\mathcal{H}^{n-1}(x), \qquad i = 1, \ldots, n$$

pass to the limit as $k \to \infty$ to get

$$\int_\Omega D_i u \, dx = \int_{\partial\Omega} Tu(x)\, \nu_\Omega^i(x)\, d\mathcal{H}^{n-1}(x), \qquad i = 1, \ldots, n.$$

1.4 Existence Theorems for PDE's

1.4.1 Dirichlet's principle

We shall now read again some general results concerning the geometry of abstract Hilbert spaces, see [GM3], in the case of the Hilbert space $H_0^1(\Omega)$. This allows us to state existence results for the Dirichlet boundary value problem for Poisson's equation.

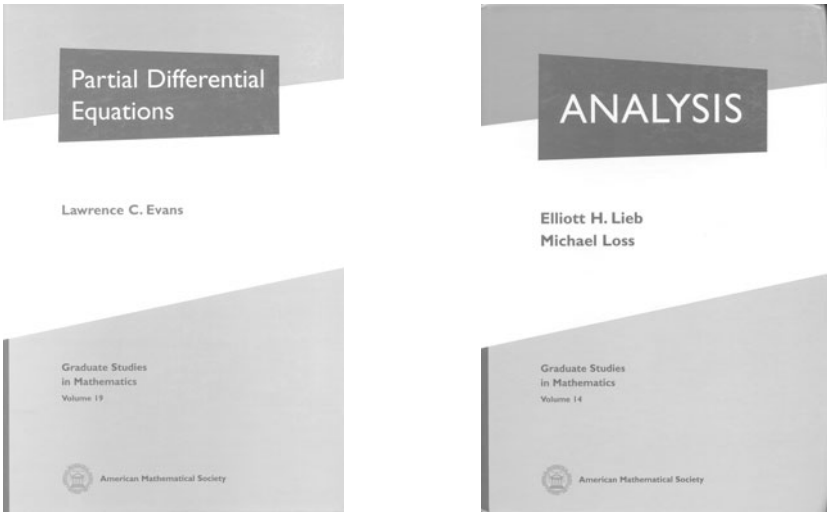


Figure 1.4. Frontispieces of two volumes on Sobolev spaces and PDE's.

a. The weak form of the equilibrium equation

Let Ω be a bounded open set of $\mathbb{R}^n$, let E be a vector field in Ω of class $C^1(\Omega)$ and let $f \in C^0(\Omega)$. On account of the Gauss–Green formulas, we have seen in Section 1.1 that the equations

$$\int_A f(x)\,dx = \int_{\partial A} E \bullet \nu_A \; d\mathcal{H}^{n-1} \qquad \forall A \subset\subset \Omega \text{ admissible} \tag{1.47}$$

and

$$f(x) = \operatorname{div} E(x) \qquad \text{in } \Omega \tag{1.48}$$

are equivalent. Since (1.47) is meaningful even when E is just continuous, we may regard (1.47) as a *weak version* of the equation $\operatorname{div} E = f$ in Ω.

There is a different way of writing (1.48). Suppose $E \in C^1(\Omega)$, $f \in C^0(\Omega)$, and $\operatorname{div} E = f$ in Ω. Multiplying (1.48) by $\varphi \in C_c^\infty(\Omega)$ and integrating in Ω we find

$$\begin{aligned} 0 &= \int_\Omega (\operatorname{div} E - f)\varphi \, dx \\ &= \int_\Omega \operatorname{div}(E\varphi)\, dx - \int_\Omega E \bullet D\varphi \; dx - \int_\Omega f\varphi \, dx \end{aligned} \tag{1.49}$$

and, since $E\varphi$ vanishes near $\partial\Omega$,

$$\int_\Omega E \bullet D\varphi \; dx + \int_\Omega f\,\varphi \, dx = 0 \qquad \forall \varphi \in C_c^\infty(\Omega). \tag{1.50}$$

Of course, we may proceed conversely: If $E \in C^1(\Omega)$ and $f \in C^0(\Omega)$ satisfy (1.50), from the Gauss–Green formulas we get

$$\int_\Omega \operatorname{div}(E\varphi)\,dx = 0,$$

hence

$$\int_\Omega (\operatorname{div} E - f)\varphi\,dx = 0 \qquad \forall \varphi \in C_c^\infty(\Omega)$$

and, consequently, $\operatorname{div} E = f$ in Ω.

In conclusion, (1.48) and (1.50) are equivalent when $E \in C^1(\Omega)$ and $f \in C^0(\Omega)$; on the other hand, (1.50) is meaningful even when $E, f \in L^1(\Omega)$. This motivates the following definition.

1.69 Definition. *Let $E \in L^1(\Omega, \mathbb{R}^n)$ and $f \in L^1(\Omega)$. We say that E and f satisfy the equation* $\operatorname{div} E = f$ *in the* weak sense *or in the* sense of distributions *if*

$$\int_\Omega (E \bullet D\varphi + f\varphi)\,dx = 0 \qquad \forall \varphi \in C_c^\infty(\Omega).$$

Summarizing, we have the following.

1.70 Theorem. *The equilibrium equation* (1.48) *and its integral forms* (1.47) *and* (1.50) *are equivalent if $E \in C^1(\Omega)$ and $f \in C^0(\Omega)$. Moreover,* (1.50) *makes sense if E and f are in $L^1(\Omega)$, and* (1.47) *makes sense if E and f are of class C^0. Finally, the two weak forms* (1.47) *and* (1.50) *are equivalent if E and f are of class C^0.*

Proof. We now have to prove the last claim. Suppose (1.50) holds with E and f of class C^0, and let $A \subset\subset \Omega$ be admissible. Let $\epsilon_0 := \frac{1}{2}\operatorname{dist}(A, \partial\Omega)$ and choose a symmetric regularization kernel $\rho(x) = r(|x|)$. If $\varphi \in C_c^\infty(A)$, then $\varphi * \rho_\epsilon$ belongs to $C_c^\infty(\Omega)$ for every $\epsilon < \epsilon_0$, hence

$$\begin{aligned}\int_\Omega (E * \rho_\epsilon) \bullet D\varphi\; dx &= \int_\Omega E \bullet ((D\varphi) * \rho_\epsilon)\; dx = \int_\Omega E \bullet D(\varphi * \rho_\epsilon)\; dx \\ &= -\int_\Omega f\varphi * \rho_\epsilon\,dx = -\int_\Omega (f * \rho_\epsilon)\varphi\,dx\end{aligned}$$

by the assumption and since the kernel is symmetric. Since $E_\epsilon := E * \rho_\epsilon$ and $f_\epsilon := f * \rho_\epsilon \in C^\infty(\Omega)$ and since $\varphi \in C_c^\infty(A)$ is arbitrary, we have

$$\operatorname{div} E_\epsilon(x) = f_\epsilon(x) \qquad \forall x \in A,\ \forall \epsilon < \epsilon_0.$$

Hence, integrating in A and using Gauss–Green formulas in A, we get

$$\int_{\partial A} E_\epsilon \bullet \nu_A\; d\mathcal{H}^{n-1} + \int_A f_\epsilon(x)\,dx = 0.$$

Since $E_\epsilon \to E$ and $f_\epsilon \to f$ uniformly on the compact subsets of Ω, passing to the limit as $\epsilon \to 0$ we conclude

$$\int_{\partial A} E \bullet \nu_A\; d\mathcal{H}^{n-1} + \int_A f(x)\,dx = 0,$$

and since A is arbitrary, E and f satisfy (1.47).

Conversely, suppose (1.47) holds and let $\varphi \in C_c^\infty(\Omega)$. For $t \in \mathbb{R}$, set $A_t := \{x \in \Omega \,|\, \varphi(x) \le t\}$ so that

$$\partial A_t := \{x \in \Omega \,|\, \varphi(x) = t\},$$

and set $R := \{x \in \Omega \,|\, |D\varphi(x)| > 0\}$. R is open and $R \cap \partial A_t$ is a smooth submanifold with exterior unit vector given by

$$\nu_{A_t}(x) = \frac{D\varphi(x)}{|D\varphi(x)|}, \qquad \forall x \in \partial A_t \cap R,$$

according to the implicit function theorem. Moreover, $\partial A_t \cap R^c$ is closed, and by the coarea formula, see Theorem 6.82, we have

$$0 = \int_{R^c} |D\varphi|\, dx = \int_{-\infty}^{+\infty} \mathcal{H}^{n-1}(\partial A_t \cap R^c)\, dt,$$

hence $\partial A_t \cap R^c$ has zero $\mathcal{H}^{n-1}$ measure for a.e. $t \in \mathbb{R}$.

Therefore A_t is an admissible domain for $\mathcal{H}^1$-a.e. t, and, for these t's, we may apply (1.47) to A_t and ∂A_t for $t < 0$: For a.e. $t < 0$ we have $A_t \subset\subset \Omega$, thus

$$\begin{aligned}\int_{\partial A_t \cap R} E \bullet \frac{D\varphi}{|D\varphi|}\, d\mathcal{H}^{n-1} &= \int_{\partial A_t \cap R} E \bullet \nu_{A_t}\, d\mathcal{H}^{n-1} \\ &= \int_{\partial A_t} E \bullet \nu_{A_t}\, d\mathcal{H}^{n-1} = \int_{A_t} f(x)\, dx,\end{aligned}$$

whereas for a.e. $t > 0$ we have $\Omega \setminus A_t \subset\subset \Omega$, thus

$$\begin{aligned}\int_{\partial A_t \cap R} E \bullet \frac{D\varphi}{|D\varphi|}\, d\mathcal{H}^{n-1} &= -\int_{\partial A_t \cap R} E \bullet \nu_{\Omega\setminus A_t}\, d\mathcal{H}^{n-1} \\ &= -\int_{\partial A_t} E \bullet \nu_{\Omega\setminus A_t}\, d\mathcal{H}^{n-1} = -\int_{\Omega\setminus A_t} f(x)\, dx.\end{aligned}$$

Again by the coarea formula, see Theorem 6.82,

$$\begin{aligned}\int_\Omega E \bullet D\varphi\, dx &= \int_R E \bullet \frac{D\varphi}{|D\varphi|}\, |D\varphi|\, dx = \int_{-\infty}^{+\infty} dt \int_{\partial A_t \cap R} E \bullet \frac{D\varphi}{|D\varphi|}\, d\mathcal{H}^{n-1} \\ &= -\int_0^{+\infty} dt \int_{\Omega\setminus A_t} f(x)\, dx + \int_{-\infty}^0 dt \int_{A_t} f(x)\, dx \\ &= \int_\Omega f(x)\Big(-\int_0^{+\infty} \chi_{\Omega\setminus A_t}(x)\, dt + \int_{-\infty}^0 \chi_{A_t}(x)\, dt\Big)\, dx \\ &= \int_\Omega f(x)(\max(\varphi(x), 0) + \min(\varphi(x), 0))\, dx \\ &= \int_\Omega f(x)\varphi(x)\, dx.\end{aligned}$$

□

b. The space H^{-1}

Denote by H^{-1} the *dual space* of $H_0^1(\Omega)$, i.e., the Hilbert space of linear bounded applications $L : H_0^1(\Omega) \to \mathbb{R}$ normed by

$$||L|| := \sup_{\varphi \neq 0} \frac{|L(\varphi)|}{||\varphi||_{1,2}}.$$

1.71 Proposition. *F belongs to $H^{-1}(\Omega)$ if and only if there are functions $f_0, f_1, \dots f_n \in L^2(\Omega)$ such that*

$$F(\varphi) = \int_\Omega \Big(f_0\varphi + \sum_{i=1}^n f_i D_i\varphi \Big)\, dx \qquad \forall \varphi \in H_0^1(\Omega), \tag{1.51}$$

moreover,

$$||F|| := \min\Big\{ \int_\Omega \Big(f_0^2 + \sum_{i=1}^n f_i^2 \Big) dx \;\Big|\; f_0, f_1,\, f_2, \dots,\, f_n \text{ satisfy (1.51)} \Big\}. \tag{1.52}$$

Proof. Let $f_0, f_1,\, f_2, \dots,\, f_n \in L^2(\Omega)$ and

$$F(\varphi) := \int_\Omega \Big(f_0\varphi + \sum_{i=1}^n f_i D_i\varphi \Big)\, dx.$$

Then

$$|F(\varphi)| \le ||f_0||_2 ||\varphi||_2 + \sum_{i=1}^n ||f_i||_2 ||D_i\varphi||_2 \le \Big(||f_0||_2^2 + \sum_{i=1}^n ||f_i||_2^2 \Big)^{1/2} ||\varphi||_{1,2},$$

i.e., $F \in H^{-1}(\Omega)$ and

$$||F||^2 \le \Big(\int_\Omega \Big(f_0^2 + \sum_{i=1}^n f_i^2 \Big) dx \Big)^{1/2}. \tag{1.53}$$

Conversely, if $F \in H^{-1}(\Omega)$, by Riesz's theorem there exists $u \in H_0^1$ such that

$$F(\varphi) = \int_\Omega (u\varphi + Du \bullet D\varphi)\, dx, \qquad ||u||_{1,2} = ||F||;$$

consequently, F can be written as in (1.51) with $f_0 := u$ and $f_i = D_i u$, and (1.52) holds. □

c. The abstract Dirichlet principle

We recall from [GM3] the following *Dirichlet's abstract principle* and *Riesz's theorem.*

1.72 Theorem. *Let H be a Hilbert space with inner product $(\ |\ )$ and let $L : H \to \mathbb{R}$ be a linear bounded functional on H. The functional*

$$\mathcal{F}(u) := \frac{1}{2}(u|u) - L(u)$$

has a unique minimum point $\overline{u} \in H$. Moreover, $||\overline{u}|| = ||L||$ and $\overline{u}$ is the unique solution of the equation

$$(\varphi|\overline{u}) = L(\varphi) \qquad \forall \varphi \in H. \tag{1.54}$$

In particular, every element $L \in H^*$ can be uniquely represented as inner product via (1.54).


PARTIELLE

DIFFERENTIALGLEICHUNGEN

UND

DEREN ANWENDUNG AUF PHYSIKALISCHE FRAGEN.

VORLESUNGEN

VON

BERNHARD RIEMANN

FÜR DEN DRUCK BEARBEITET UND HERAUSGEGEBEN

VON

K. HATTENDORFF

MIT IN DEN TEXT EINGEDRUCKTEN HOLZSTICHEN

BRAUNSCHWEIG,
DRUCK UND VERLAG VON FRIEDRICH VIEWEG UND SOHN.
1869.

Figure 1.5. Jacques Hadamard (1865–1963) and the frontispiece of the lectures on differential equations by Georg F. Bernhard Riemann (1826–1866).

d. The Dirichlet problem

Let Ω be an open and bounded set of $\mathbb{R}^n$. Because of Poincaré's inequality (1.45) the bilinear form

$$a(u,\varphi) := \int_\Omega Du \bullet D\varphi \; dx \tag{1.55}$$

is an inner product on $H_0^1(\Omega)$ that is equivalent to the standard one:

$$(u|\varphi) = \int_\Omega (u\varphi + \; Du \bullet D\varphi \,)\, dx.$$

Moreover, given $f_0 \in L^2(\Omega,\mathbb{R})$ and $f \in L^2(\Omega,\mathbb{R}^n)$, the linear functional $L : H_0^1(\Omega) \to \mathbb{R}$ given by

$$L(\varphi) := \int_\Omega f_0\, u\, dx + \int_\Omega f \bullet Du \; dx \tag{1.56}$$

is continuous, $L \in \mathcal{L}(H_0^1,\mathbb{R})$. Theorem 1.72 with $H = H_0^1(\Omega)$, $(u|\varphi) := a(u,\varphi)$, and L given by (1.56) then reads as follows, taking also into account Proposition 1.71.

1.73 Theorem. *For every $f_0 \in L^2(\Omega,\mathbb{R})$ and $f \in L^2(\Omega,\mathbb{R}^n)$ there exists a unique minimum point $\overline{u}$ of the integral functional*

$$\mathcal{F}(u) := \frac{1}{2}\int_\Omega |Du|^2\, dx - \int_\Omega f_0\, u\, dx - \int_\Omega f \bullet Du \; dx$$

in $H_0^1(\Omega)$. *The minimum point* $\overline{u}$ *is also the unique solution of*

$$\int_\Omega D\overline{u} \bullet D\varphi \, dx = \int_\Omega \Big(f_0\varphi + f \bullet D\varphi \Big) \, dx \qquad \forall \varphi \in H_0^1(\Omega). \tag{1.57}$$

Furthermore,

$$\int_\Omega |D\overline{u}|^2 \, dx \le \int_\Omega \Big(f_0^2 + |f|^2 \Big) \, dx.$$

As we have seen, (1.57) is a weak form of

$$-\Delta \overline{u} = f_0 - \operatorname{div} f \qquad \text{in } \Omega,$$

and the trace of $\overline{u}$ on $\partial\Omega$ is zero since $u \in H_0^1(\Omega)$. Therefore we call a function $u \in H_0^1(\Omega)$ that satisfies (1.57) a *weak solution of the Dirichlet problem*

$$\begin{cases} -\Delta u = f_0 - \operatorname{div} f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{1.58}$$

With this terminology, Theorem 1.73 can be rephrased as follows.

1.74 Corollary. *For every* $f_0 \in L^2(\Omega, \mathbb{R})$ *and* $f \in L^2(\Omega, \mathbb{R}^n)$ *there exists a unique minimum point* $\overline{u}$ *of the integral functional*

$$\mathcal{F}(u) := \frac{1}{2} \int_\Omega |Du|^2 \, dx - \int_\Omega f_0 \, u \, dx - \int_\Omega f \bullet Du \, dx$$

in $H_0^1(\Omega)$. *Moreover,* $\overline{u}$ *is the unique* weak solution *of the Dirichlet problem* (1.58). *Furthermore,*

$$\int_\Omega |D\overline{u}|^2 \, dx \le \int_\Omega \Big(f_0^2 + |f|^2 \Big) \, dx.$$

1.75 Nonzero boundary data. To the previous case we may subsume the (weak) solvability of the problem

$$\begin{cases} \Delta u = 0, & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega \end{cases}$$

when $g \in H^1(\Omega)$. In fact, the *Dirichlet integral*

$$\mathcal{F}(u) := \frac{1}{2} \int_\Omega |Du|^2 \, dx$$

has a minimum point in the class $g + H_0^1(\Omega)$. Setting $v := u - g$, this amounts to minimizing among the $v \in H_0^1(\Omega)$ the functional

$$\mathcal{F}(u) = \mathcal{F}(v+g) = \frac{1}{2}\int_\Omega |D(g+v)|^2\,dx$$
$$= \frac{1}{2}\int_\Omega |Dv|^2 + \int_\Omega Dv \bullet Dg\; dx + \frac{1}{2}\int_\Omega |Dg|^2\,dx$$

thus, g being given, to minimize

$$\mathcal{F}_1(v) := \frac{1}{2}\int_\Omega |Dv|^2\,dx + \int_\Omega Dv \bullet Dg\; dx$$

in $H_0^1(\Omega)$. By the Dirichlet principle we conclude that $\mathcal{F}_1$ has a minimizer $\overline{v}$ in $H_0^1(\Omega)$, and that $\overline{v}$ is the unique solution of

$$\int_\Omega Dv \bullet D\varphi\; dx + \int_\Omega Dg \bullet D\varphi\; dx = 0 \qquad \forall \varphi \in H_0^1(\Omega). \tag{1.59}$$

Since (1.59) is the weak form of

$$\begin{cases} \Delta v = -\sum_{i=1}^n D_i D_i g = -\Delta g & \text{in } \Omega, \\ v = 0 & \text{on } \partial\Omega, \end{cases}$$

we say that the unique function $\overline{u} = g + \overline{v} \in g + H_0^1(\Omega)$ satisfying (1.59) is the unique *weak solution* of

$$\begin{cases} \Delta u = 0 & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega. \end{cases}$$

1.76 Regularity. Deciding whether $\overline{u}$ is smooth or not according to the regularity of the data is now a question answered by the so-called *theory of* (*elliptic*) *regularity*, that we cannot discuss here. We only state a theorem without further comments.

Theorem. *Let Ω be a bounded open set with smooth boundary and let $g \in H^1(\Omega) \cap C^0(\overline{\Omega})$. The unique weak solution $u \in H^1(\Omega)$ of the minimum problem*

$$\begin{cases} \frac{1}{2}\int_\Omega |Du|^2\,dx \to \min, \\ u - g \in H_0^1(\Omega) \end{cases}$$

or, equivalently, of (1.59)*, is of class $C^\infty(\Omega) \cap C^0(\overline{\Omega})$. Consequently, $\overline{u}$ is the classical solution of the boundary value problem*

$$\begin{cases} \Delta u = 0 & \textit{in } \Omega, \\ u = g & \textit{on } \partial\Omega. \end{cases}$$

1.77 Approximated solutions. Of course, the abstract methods of Ritz and Faedo–Galerkin are both useful for approximating the solution of (1.57).

Assume that $\{u_n\} \subset H_0^1(\Omega)$ is an orthonormal basis of $H_0^1(\Omega)$ with respect to the inner product (1.55). Then, for every $\varphi \in H_0^1(\Omega)$, $\varphi = \sum_{n=0}^{\infty} a(\varphi, u_n)u_n$ in $H_0^1(\Omega)$, we write the operator L in (1.56) as

$$L(\varphi) = \sum_{n=0}^{\infty} L(u_n)\, a(\varphi, u_n) \qquad \forall \varphi \in H_0^1,$$

and the unique solution of (1.45) is

$$\overline{u} = \sum_{n=0}^{\infty} L(u_n)\, u_n \qquad \text{in } H_0^1(\Omega)$$

by the Ritz method, see [GM3].

Alternatively, select a sequence of finite vector spaces $V_n \subset H_0^1$ such that $\cup_n V_n$ is dense in $H_0^1(\Omega)$, and for each V_n, let $\{e_1^n, e_2^n, \ldots, e_{p(n)}^n\}$ be a basis of V_n. Then solve for $u_n \in V_n$ the finite-dimensional linear system

$$a(u_n, e_i^n) = L(e_i^n), \qquad i = 1, \ldots, p(n).$$

By the Faedo–Galerkin result, see [GM3], the sequence $\{u_n\}$ converges in H_0^1 to the unique solution $\overline{u}$ of (1.57).

e. Neumann problem

By applying the abstract Dirichlet principle to the Hilbert space $H^1(\Omega)$ we conclude the following.

1.78 Proposition. *Let $f_0 \in L^2(\Omega, \mathbb{R})$ and $f \in L^2(\Omega, \mathbb{R}^n)$. Then there exists a unique minimum point $\overline{u} \in H^1(\Omega)$ of the functional*

$$\mathcal{F}(u) := \frac{1}{2}\int_\Omega |Du|^2\, dx + \int_\Omega |u|^2\, dx + \int_\Omega \Big(f_0 u - f \bullet Du \Big)\, dx$$

in $H^1(\Omega)$ which is also the unique solution of the equation

$$\int_\Omega Du \bullet D\varphi \; dx + \int_\Omega u\varphi \, dx = -\int_\Omega \Big(f_0\varphi - f \bullet D\varphi \Big)\, dx \qquad \forall \varphi \in H^1(\Omega). \tag{1.60}$$

Moreover,

$$||u||_{1,2}^2 \le C \int_\Omega (|f_0|^2 + |f|^2)\, dx.$$

Let us try to interpret (1.60) assuming f_0, f and u are sufficiently regular, for instance $f \in C^1(\overline{\Omega}, \mathbb{R}^n)$, $f_0 \in C^0(\Omega)$ and $u \in C^2(\overline{\Omega})$.

Since (1.60) holds in particular for all $\varphi \in C_c^\infty(\Omega)$, u is a weak solution of

$$-\Delta u + u = -f_0 - \operatorname{div} f, \tag{1.61}$$

and, according to the regularity assumptions, it is also a classical solution of (1.61). From (1.60) and (1.61) we then get

$$\begin{aligned} 0 &= \int_\Omega Du \bullet D\varphi \, dx + \int_\Omega u\varphi \, dx + \int_\Omega f_0\varphi - \int_\Omega f \bullet D\varphi \, dx \\ &= \int_\Omega \Big(Du \bullet D\varphi + u\varphi + f_0\varphi - f \bullet D\varphi \Big) \, dx \\ &\quad - \int_\Omega (-\operatorname{div} Du + u + f_0 + \operatorname{div} f)\varphi \, dx \\ &= \int_\Omega Du - f \bullet D\varphi + \operatorname{div}(Du - f)\varphi \, dx \\ &= \int_\Omega \operatorname{div}((Du - f)\varphi) \, dx \end{aligned}$$

for every $\varphi \in C^1(\overline{\Omega})$, and conclude by the Gauss–Green formulas that

$$\int_{\partial\Omega} \Big((Du - f) \bullet \nu_\Omega \Big) \varphi \, d\mathcal{H}^{n-1} = 0 \qquad \forall \varphi \in C^1(\overline{\Omega}),$$

that is,

$$\frac{du}{d\nu_\Omega} := Du \bullet \nu_\Omega = f \bullet \nu_\Omega \qquad \text{on } \partial\Omega,$$

$\frac{du}{d\nu}$ being the (external) normal derivative with respect to Ω.

In conclusion, (1.60) is a weak form of the *Neumann problem*

$$\begin{cases} -\Delta u + u = -f_0 + \operatorname{div} f & \text{in } \Omega, \\ \dfrac{du}{d\nu} = f \bullet \nu_\Omega & \text{on } \partial\Omega \end{cases} \tag{1.62}$$

since solutions of (1.60) solve (1.62) provided that the data f_0 and f *and the solution* are sufficiently regular. With this terminology, Proposition 1.78 then provides the unique weak solution of (1.62).

1.79 Approximation. The abstract Ritz and Faedo–Galerkin methods apply also to approximating the weak solution of the Neumann problem (1.60).

Assume that $\{u_n\}$ is an orthonormal complete system in $H^1(\Omega)$ with respect to the inner product of $H^1(\Omega)$

$$(u|v) := \int_\Omega (\, Du \bullet Dv \, + uv) \, dx.$$

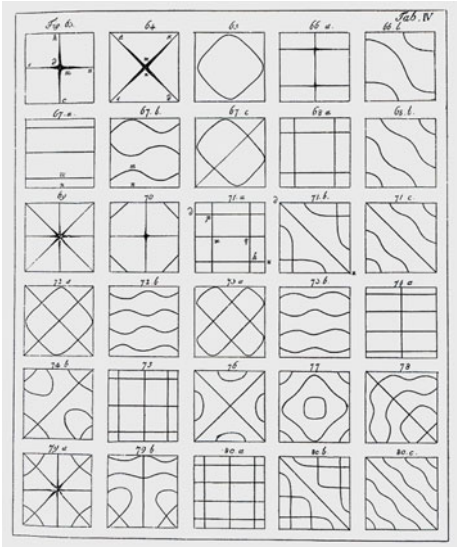


Figure 1.6. From E. F. Chadni *Die Akustik*: nodal lines obtained by spreading sand on a metal plate fixed in clamps and then applying a violin bow to the edges of such plates.

Then, for every $\varphi \in H^1(\Omega)$, we have $\varphi = \sum_{n=0}^{\infty} (\varphi|u_n)\, u_n$ in $H^1(\Omega)$ and the operator

$$L(\varphi) := -\int_\Omega \Big(f_0\varphi - f \bullet D\varphi \Big)\, dx, \qquad \varphi \in H^1(\Omega)$$

writes as

$$L(\varphi) = \sum_{n=0}^{\infty} L(u_n)\, (\varphi|u_n) \qquad \forall \varphi \in H^1(\Omega);$$

then the unique solution of (1.60) is

$$\overline{u} = \sum_{n=0}^{\infty} L(u_n)\, u_n \qquad \text{in } H^1(\Omega)$$

by the Ritz method, see [GM3].

Alternatively, select a sequence of finite vector spaces $V_n \subset H_0^1$ such that $\cup_n V_n$ is dense in $H^1(\Omega)$, and for every V_n, let $\{e_1^n, e_2^n, \ldots, e_{p(n)}^n\}$ be a basis of V_n. Then solve for $u_n \in V_n$ the finite-dimensional linear system

$$(u_n|e_i^n) = L(e_i^n), \qquad i = 1, \ldots, p(n).$$

The Faedo–Galerkin abstract result, see [GM3], implies that $\{u_n\}$ converges in $H^1(\Omega)$ to the unique solution $\overline{u}$ of (1.60).

f. Cauchy–Riemann equations

Let $f = u + iv : \Omega \subset \mathbb{C} \to \mathbb{C}$. The Cauchy–Riemann equations,

$$\frac{\partial f}{\partial y} = i\frac{\partial f}{\partial x} \qquad \text{in } \Omega,$$

take the weak form:

$$\int_\Omega f\,(\frac{\partial \varphi}{\partial y} - i\frac{\partial \varphi}{\partial x})\,dxdy = 0 \qquad \forall \varphi \in C_c^\infty(\Omega;\mathbb{C}). \tag{1.63}$$

Suppose $f \in L^1(\Omega)$ satisfies (1.63) and let ρ be a mollifying kernel in $\mathbb{R}^2$. For $\epsilon > 0$ set $\Omega_\epsilon := \{x \in \Omega \,|\, \operatorname{dist}(x, \partial\Omega > \epsilon)\}$ and let $f_\epsilon : \Omega_\epsilon \to \mathbb{R}^2$ be the ϵ-mollified of f, $f_\epsilon(x) := f * \rho_\epsilon(x)$. Of course, f_ϵ satisfies the weak Cauchy–Riemann equations and the classical one in Ω_ϵ, as it is regular. Therefore f_ϵ is holomorphic in Ω_ϵ. Using the Cauchy formula, one sees that $f_\epsilon \to f$ not only in L^1 but also uniformly on compact subsets of Ω, and, actually, all the complex derivatives of f_ϵ converge uniformly on compact sets in Ω to the corresponding complex derivative of f. Hence f is holomorphic in Ω.

1.80 ¶. Provide all of the details of the previous claims.

1.81 ¶. Let f be holomorphic in Ω and let f_ϵ be the ϵ-mollified of f via a spherical symmetric kernel $k(x) = k'(|x|)$. Show that

$$f_\epsilon(z) = f(z) \qquad \forall z \in \Omega_\epsilon := \Big\{ z \in \Omega \,\Big|\, \operatorname{dist}(z, \partial\Omega) > \epsilon \Big\}.$$

[*Hint.* Use Cauchy's formula.]

1.4.2 The alternative theorem

Let $H = H_0^1(\Omega)$, $\Omega \subset \mathbb{R}^n$ being bounded and open. We shall now discuss the existence of weak solutions of the Dirichlet problem for the Laplace operator

$$\begin{cases} -\Delta u + \lambda u = f_0 - \operatorname{div} f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

that is, of the existence of functions $u \in H_0^1(\Omega)$ that satisfy

$$\int_\Omega Du \bullet D\varphi \; dx + \lambda \int_\Omega u\varphi \, dx = -\int_O (f_0 - \, f \bullet D\varphi \,)\, dx \qquad \forall \varphi \in H_0^1(\Omega) \tag{1.64}$$

where $f_0, f \in L^2(\Omega)$. Observe the following.

(i) As we know,

$$a(u,v) := \int_\Omega Du \bullet Dv \; dx, \qquad u, v \in H_0^1(\Omega)$$

is an inner product in H_0^1.

(ii) For any $u \in L^2(\Omega)$, the map $\varphi \mapsto \int_\Omega u\varphi \, dx$ is a linear functional on $L^2(\Omega)$ and hence on $H_0^1(\Omega)$. Consequently, by Riesz's theorem, there exists a linear operator $K : L^2 \to H_0^1(\Omega)$ such that

$$\int_\Omega u\varphi \, dx = a(Ku, \varphi) \qquad \forall u \in L^2, \forall \varphi \in H_0^1(\Omega). \tag{1.65}$$

Moreover, the embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$ being compact by Rellich's theorem, the restriction of K to H_0^1 (into H_0^1) is compact; we also see from (1.65) that $K : H_0^1 \to H_0^1$ is self-adjoint.

(iii) The linear operator

$$L(\varphi) := -\int_O (f_0 - \, f \bullet D\varphi \,) \, dx \qquad \forall \varphi \in H_0^1(\Omega)$$

is a linear operator defined on H_0^1, $L \in H^{-1}$, and, by Riesz's theorem, there exists $g \in H_0^1(\Omega)$ such that $L(\varphi) = \alpha(g, \varphi)$ $\forall \varphi \in H_0^1(\Omega)$.

Consequently, (1.64) rewrites as

$$a(u + \lambda Ku, \varphi) = L(\varphi) = (g|\varphi) \qquad \forall \varphi \in H_0^1(\Omega)$$

or as the abstract linear equation

$$u + \lambda Ku = g \qquad \text{in } H_0^1(\Omega) \tag{1.66}$$

for the operator $Id + \lambda K$ which is, as we have seen, a compact self-adjoint perturbation of the identity. The abstract alternative theorem, see [GM3], yields then the following.

1.82 Theorem. *Let $f \in H_0^1(\Omega)$ and $\lambda \in \mathbb{R}$. The equation*

$$-\Delta u + \lambda u = f_0 - \operatorname{div} f \qquad \textit{in weak form,}$$

i.e., as we have seen, the abstract equation (1.64)*, has a solution in $H_0^1(\Omega)$ if and only if*

$$\int_\Omega (f_0 v + \, Df \bullet Dv \,) \, dx = 0$$

for all $v \in H_0^1(\Omega)$ that solve $-\Delta v + \lambda v = 0$ in the weak sense, i.e., that solve the abstract equation $v + \lambda Kv = 0$ in H_0^1.

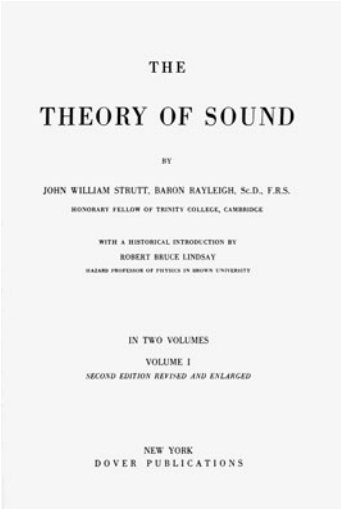


Figure 1.7. Heinrich Hertz (1857–1894) and the frontispiece of the *Theory of Sound* by Lord William Strutt Rayleigh (1842–1919).

1.83 Definition. *We say that $\lambda \in \mathbb{R}$ is an eigenvector for the Dirichlet problem for the Laplace operator*

$$\begin{cases} -\Delta u + \lambda u = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

if the equation

$$\int_\Omega Du \bullet D\varphi \, dx + \lambda \int_\Omega u\varphi \, dx = 0 \qquad \forall \varphi \in H_0^1(\Omega) \tag{1.67}$$

has a nonzero solution in $H_0^1(\Omega)$. If λ is an eigenvalue, the corresponding solutions of (1.67)*, called the* eigenfunctions of the Laplace operator in H_0^1 *relative to λ, form a vector space.*

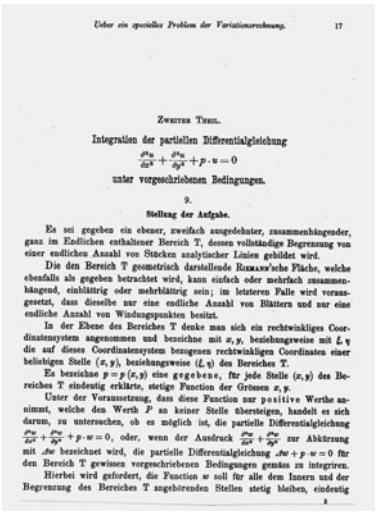
Since, as we have seen, (1.67) is equivalent to the abstract equation

$$u + \lambda K u = 0 \qquad \text{on } H_0^1(\Omega),$$

we may apply the Courant–Hilbert–Schmidt theory, see [GM3], to get the following.

1.84 Theorem. *The eigenvalues are denumerable, and we can form with them an increasing sequence $\{\lambda_n\}$ that converges to $+\infty$. Moreover:*

(i) *For all k, the space V_k of the eigenfunctions with eigenvalue λ_k is a finite-dimensional subspace of $H_0^1(\Omega)$.*



Über ein specielles Problem der Variationsrechnung. 17

ZWEITER THEIL.

Integration der partiellen Differentialgleichung

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+p\cdot u=0$$

unter vorgeschriebenen Bedingungen.

9.

Stellung der Aufgabe.

Es sei gegeben ein ebener, zweifach ausgedehnter, zusammenhängender, ganz im Endlichen enthaltener Bereich T, dessen vollständige Begrenzung von einer endlichen Anzahl von Stücken analytischer Linien gebildet wird.

Die den Bereich T geometrisch darstellende Riemann'sche Fläche, welche ebenfalls als gegeben betrachtet wird, kann einfach oder mehrfach zusammenhängend, einblättrig oder mehrblättrig sein; im letzteren Falle wird vorausgesetzt, dass dieselbe nur eine endliche Anzahl von Blättern und nur eine endliche Anzahl von Windungspunkten besitzt.

In der Ebene des Bereiches T denke man sich ein rechtwinkliges Coordinatensystem angenommen und bezeichne mit x, y, beziehungsweise mit ξ, η die auf dieses Coordinatensystem bezogenen rechtwinkligen Coordinaten einer beliebigen Stelle (x, y), beziehungsweise (ξ, η) des Bereiches T.

Es bezeichne $p = p(x, y)$ eine gegebene, für jede Stelle (x, y) des Bereiches T eindeutig erklärte, stetige Function der Grössen x, y.

Unter der Voraussetzung, dass diese Function nur positive Werthe annimmt, welche den Werth P an keiner Stelle übersteigen, handelt es sich darum, zu untersuchen, ob es möglich ist, die partielle Differentialgleichung $\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}+p\cdot w=0$, oder, wenn der Ausdruck $\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}$ zur Abkürzung mit Δw bezeichnet wird, die partielle Differentialgleichung $\Delta w + p\cdot w = 0$ für den Bereich T gewissen vorgeschriebenen Bedingungen gemäss zu integriren.

Hierbei wird gefordert, die Function w soll für alle dem Innern und der Begrenzung des Bereiches T angehörenden Stellen stetig bleiben, eindeutig

3

26 H. A. SCHWARZ.

Es bestehen also für jeden ganzzahligen Werth von k, welcher kleiner ist als n, die Gleichungen

$$(20.)\qquad \iint_T p\, w_i\, w_{\varkappa}\, dx\, dy = \iint_T p\, w_i\, w_{\varkappa-k}\, dx\, dy = W_{\varkappa},$$

$$\iint_T \left(\frac{\partial w_i}{\partial x}\frac{\partial w_{\varkappa}}{\partial x}+\frac{\partial w_i}{\partial y}\frac{\partial w_{\varkappa}}{\partial y}\right) dx\, dy = \iint_T \left(\frac{\partial w_{i+1}}{\partial x}\frac{\partial w_{\varkappa-1}}{\partial x}+\frac{\partial w_{i+1}}{\partial y}\frac{\partial w_{\varkappa-1}}{\partial y}\right) dx\, dy = W_{\varkappa}.$$

15.

Einführung der Constante c.

Der bekannte Satz, dass die Discriminante einer definiten binären quadratischen Form stets einen positiven von Null verschiedenen Werth besitzt, kann dazu angewendet werden, um eine Beziehung zwischen den absoluten Beträgen der über denselben Bereich T ausgedehnenden drei Doppelintegrale $A=\iint_T \varphi^2\, dx\, dy$, $B=\iint_T \varphi\chi\, dx\, dy$, $C=\iint_T \chi^2\, dx\, dy$ herzuleiten, eine Beziehung, deren Kenntnis für die folgende Untersuchung von Wichtigkeit ist.

Die Grössen φ, χ bedeuten zwei reelle, für alle Stellen (x, y) des Bereiches T eindeutig erklärte Functionen der beiden Argumente x, y, welche die Eigenschaft haben, erstens, dass die über den Bereich T ausgedehnten Doppelintegrale A, B, C unbedingt convergent sind, zweitens, dass der Quotient der beiden Functionen φ und χ nicht einer Constanten gleich ist.

Unter den angegebenen Voraussetzungen ist die binäre quadratische Form $\iint_T (\alpha\varphi+\beta\chi)^2\, dx\, dy = A\cdot\alpha^2+2B\cdot\alpha\beta+C\cdot\beta^2$ eine definite, weil das Doppelintegral, dessen Werth mit dem Werthe der quadratischen Form übereinstimmt, für kein von dem Werthepaare $\alpha=0, \beta=0$ verschiedenes Paar reeller Werthe der Grössen α, β gleich Null wird. Hieraus ergibt sich also die Beziehung

$$(21.)\qquad AC - B^2 > 0 \text{ oder } |B| < \sqrt{A}\cdot\sqrt{C}.$$

Wenn $\varphi = \sqrt{p}\cdot w_{\varkappa}$, $\chi = \sqrt{p}\cdot w_{\varkappa+1}$ gesetzt wird, so erhalten A, B, C beziehlich die Werthe $W_{2\varkappa}, W_{2\varkappa+1}, W_{2\varkappa+2}$. Es besteht demnach zwischen diesen drei Grössen die Beziehung

$$(22.)\qquad \frac{W_{2\varkappa+1}}{W_{2\varkappa}} < \frac{W_{2\varkappa+2}}{W_{2\varkappa+1}}.$$

Figure 1.8. Two pages from Hermann Schwarz (1843–1921), *Über ein die Flächen kleinsten Inhalts betreffendes Problem der Variationsrechnung*, 1885, where the eigenvalue problem for the Laplacian is studied.

(ii) *Different eigenspaces are orthogonal in H^1_0 and in $L^2(\Omega)$. Moreover, $\cup_k V_k$ is dense in $H^1_0(\Omega)$. Consequently, there exists a sequence of eigenvectors orthonormal in $L^2(\Omega)$ which is a complete system in $H^1_0(\Omega)$ (hence in $L^2(\Omega)$) such that for every $h, k \in \mathbb{N}$*

$$\int_\Omega Du_k \bullet Du_h \, dx = \lambda_k \delta_{hk},$$

where λ_k is the eigenvalue relative to u_k.

Finally, we may give a *variational characterization* of the eigenvalues, see [GM3], but we shall not insist on this.

1.4.3 The Sturm–Liouville theory

For any positive integer n, the function $\sin nt$, $t \in [0, \pi]$, solves the problem

$$\begin{cases} -u'' = n^2 u, \\ u(0) = u(\pi) = 0, \end{cases}$$

i.e., it may be regarded as an eigenfunction of the operator $-u''$ on H^1_0 with associated eigenvalue n^2. In this subsection we show that many properties of the sequence of functions $\{\sin nt\}$ and of their corresponding eigenvalues

$\{n^2\}$ are shared by the eigenfunctions and eigenvalues of a large class of second order ordinary differential operators.

Let us consider the general linear second order equation

$$a(x)u'' + b(x)u' + c(x)u = F(x), \qquad x \in]\alpha, \beta[\tag{1.68}$$

where a, b and c are continuous functions and $a > 0$. If we set

$$p(x) := \exp\left(\int_a^x \frac{b(\xi)}{a(\xi)}\, d\xi\right), \qquad q(x) := \frac{c(x)}{a(x)}p(x), \qquad f(x) := \frac{F(x)}{a(x)}p(x),$$

(1.68) multiplied by $p(x)/a(x)$ transforms into

$$\frac{d}{dx}\Big(p(x)\frac{du}{dx}\Big) + q(x)u = f(x). \tag{1.69}$$

We are interested in the *eigenvalue problem for the operator* $-(pu')' + qu$ with homogeneous Dirichlet data on $]\alpha, \beta[$, i.e., on the nonzero weak solvability of the problem

$$\begin{cases} -(pu')' + qu = \lambda u & \text{in }]\alpha, \beta[, \\ u(\alpha) = 0,\ u(\beta) = 0, \end{cases} \tag{1.70}$$

in dependence of λ, or, more explicitly, on the solvability in $H_0^1(]\alpha, \beta[)$ of

$$a(u, \varphi) = \lambda \int_\alpha^\beta u\varphi\, dx \qquad \forall \varphi \in H_0^1(]\alpha, \beta[) \tag{1.71}$$

where $a(v, \varphi)$ is the bilinear form

$$a(u, \varphi) := \int_\alpha^\beta \Big(p(x)u'(x)\varphi'(x) + q(x)u(x)\varphi(x)\Big)\, dx. \tag{1.72}$$

In the sequel we shall assume $p \in C^1([\alpha, \beta])$, $p > 0$ in $[\alpha, \beta]$, so that $0 < a \le p(x) \le b\ \forall x \in [\alpha, \beta]$ for suitable $b \ge a > 0$, and $q \in C^0([\alpha, \beta])$. Moreover, we may (modulus a translation of the values of λ) and do assume that $q(x) \ge 0$.

1.85 Proposition. *Let $p \in C^1([\alpha, \beta])$, $p > 0$ in $[\alpha, \beta]$ and let $q \in C^0([\alpha, \beta])$, $q \ge 0$. Every weak solution $u \in H_0^1(]\alpha, \beta[)$ of* (1.71) *is a smooth function of class $C^2([\alpha, \beta])$, hence a classical solution of* (1.70).

Proof. Recall that every $u \in H_0^1(]\alpha, \beta[)$ has a Hölder-continuous representative that we shall call u, $u \in C^{0,1/2}([\alpha, \beta])$. Set $Q(x) := \int_\alpha^x (q(t) - \lambda)u\, dt$. Clearly, Q is of class $C^1([\alpha, \beta])$ and we have

$$\int_\alpha^\beta (q - \lambda)u\varphi\, dx = -\int_\alpha^\beta Q\varphi'\, dx$$

and from (1.71)

$$\int_\alpha^\beta (pu' - Q)\varphi'\, dx = 0 \qquad \forall \varphi \in C_c^\infty([\alpha, \beta]).$$

Lemma 1.52 then yields $p(x)u'(x) - Q(x) = cost$ for a.e. $x \in]a, b[$, i.e., u' has a representative of class $C^1([\alpha, \beta])$, which yields $u \in C^2([\alpha, \beta])$. □

1.86 Proposition. *We have the following:*

(i) *Eigenfunctions $u, v \in H_0^1$ relative to distinct eigenvalues $\lambda \neq \mu$ are orthogonal in L^2.*
(ii) *Eigenfunctions u and v relative to the same eigenvalue are one a multiple of the other.*

Proof. (i) We have

$$a(u,v) = \int_\alpha^\beta (pu'v' + quv)\,dx = \lambda \int_\alpha^\beta uv\,dx,$$
$$a(v,u) = \int_\alpha^\beta (pv'u' + qvu)\,dx = \mu \int_\alpha^\beta vu\,dx,$$

hence

$$(\lambda - \mu)\int_\alpha^\beta uv\,dx = a(u,v) - a(v,u) = 0.$$

(ii) If $\gamma \in]\alpha,\beta[$, the function $\phi(x) = v'(\gamma)(u(x)-u(\gamma)) - u'(\gamma)(v(x)-v(\gamma))$ is of class C^2 and solves (1.70) with $\phi(\gamma) = \phi'(\gamma) = 0$. Hence $\phi(x) = 0\ \forall x$ according to the uniqueness of the Cauchy problem. □

Because of the assumptions on p and q ($a \le p(x) \le b$ and $0 \le q(x) \le c$) and Poincaré's inequality, the bilinear form (1.72) is an inner product in H_0^1 equivalent to the standard one. Therefore, as for the Laplace operator, in Section 1.4.2, the eigenvalue problem is subsumed by the Courant–Hilbert–Schmidt theory. Taking into account that the eigenspaces are all of dimension 1, see Proposition 1.86, we may therefore state the following.

1.87 Theorem. *Let $p \in C^1([\alpha,\beta])$, $p > 0$ in $[\alpha,\beta]$, $q \in C^0([\alpha,\beta])$, $q \ge 0$. The eigenvalues of* (1.71) *form an increasing sequence $\{\lambda_n\}$ that diverges to $+\infty$. For each n we can find an eigenfunction u_n relative to λ_n in such a way that*

(i) *the sequence $\{u_n\}$ is an orthonormal system in L^2,*
(ii) *$\{u_n\}$ is a complete system of eigenfunctions in H_0^1 (hence in L^2),*
(iii) *we have $a(u,u) = \lambda_n$ for every eigenfunction u relative to λ_n.*

The following variational characterization of the eigenvalues holds:

$$\lambda_1 = \min\Big\{ \frac{a(\varphi,\varphi)}{||\varphi||^2} \,\Big|\, \varphi \in H_0^1 \Big\}, \tag{1.73}$$

and for $k \ge 2$,

$$\lambda_k = \min\Big\{ \frac{a(\varphi,\varphi)}{||\varphi||_2} \,\Big|\, \varphi \in H_0^1,\ a(u,\varphi) = 0 \ \text{ for every eigenfunction } u \text{ relative to one of the eigenvalues } \lambda_1, \dots \lambda_{k-1} \Big\}. \tag{1.74}$$

The next theorem collects the main properties of the eigenvalues and eigenfunctions of problem (1.70).

1.88 Theorem (Sturm–Liouville). *Let $p \in C^1([\alpha,\beta])$, $0 < a \le p(x) \le b\ \forall x \in [\alpha,\beta]$, $q \in C^0([\alpha,\beta])$, $q \ge 0$, and let $\{\lambda_n\}$ be the increasing sequence of the eigenvalues of problem* (1.70)*. Then we have the following:*

(i) *Eigenfunctions relative to the first eigenvalue λ_1 never vanish in $]\alpha,\beta[$. Eigenfunctions relative to the other eigenvalues vanish at least once.*

(ii) (Separation theorem) *Let $\overline{u}$ and u be eigenfunctions relative to the eigenvalues $\overline{\lambda}$ and λ, respectively. Suppose that $\overline{\lambda} < \lambda$ and that $\overline{\alpha}$ and $\overline{\beta}$ are two consecutive zeros of u. Then $\overline{u}$ needs to vanish in at least a point in $]\overline{\alpha},\overline{\beta}[$.*

(iii) (Oscillation theorem) *The eigenfunctions relative to the kth eigenvalue λ_k have exactly $k-1$ zeros internal to $]\alpha,\beta[$.*

(iv) (Monotonic theorem) *When shortening the interval $]\alpha,\beta[$ or increasing p and q, the eigenvalues of* (1.70) *increase.*

Proof. Step 1. From (1.73) we infer that, if u_1 is an eigenfunction relative to λ_1, then also $|u_1|$ is an eigenfunction relative to λ_1. Since u_1 and $|u_1|$ are proportional by Proposition 1.86, we infer that either $u_1 \ge 0$ or $u_1 \le 0$ in $]\alpha,\beta[$. On the other hand, if $u_1(\gamma) = 0$ for some $\gamma \in]\alpha,\beta[$, also $u_1'(\gamma) = 0$ since u_1 is smooth. The uniqueness of the Cauchy problem would then give $u_1 = 0$ identically, a contradiction.

If v is an eigenfuction relative to an eigenvalue different from λ_1, then v and $u_1 > 0$ are orthogonal in L^2 by Proposition 1.86. It follows that v vanishes at least once. This concludes the proof of (i).

Step 2. We now prove (iv) for the first eigenvalue. Suppose $\alpha \le \overline{\alpha} < \overline{\beta} \le \beta$, $\overline{p} \ge p$ and $\overline{q} \ge q$, and let $\overline{u}_1$ be an eigenfunction relative to λ_1 for $-(\overline{p}u')' + \overline{q}u$ in $]\overline{\alpha},\overline{\beta}[$, i.e.,

$$\begin{cases} -(\overline{p}u')' + \overline{q}u = \lambda_1 u & \text{in }]\overline{\alpha},\overline{\beta}[, \\ u(\overline{\alpha}) = u(\overline{\beta}) = 0. \end{cases} \tag{1.75}$$

By (i) we may assume that $u_1 > 0$ in $]\overline{\alpha},\overline{\beta}[$. Consider the function

$$\widetilde{\phi}(x) = \begin{cases} \overline{u}_1(x) & \text{if } x \in]\overline{\alpha},\overline{\beta}[, \\ 0 & \text{otherwise,} \end{cases} \qquad x \in]\alpha,\beta[.$$

Of course, $\widetilde{\phi} \in H_0^1(]\alpha,\beta[)$ and

$$\begin{aligned} \lambda_1 :&= \min_{\phi(\alpha)=\phi(\beta)=0} \frac{\int_\alpha^\beta (p\phi'^2 + q\phi^2)\,dx}{\int_\alpha^\beta \phi^2\,dx} \le \frac{\int_\alpha^\beta (p(\widetilde{\phi}')^2 + q\widetilde{\phi}^2)\,dx}{\int_\alpha^\beta \widetilde{\phi}^2\,dx} \\ &= \frac{\int_{\overline{\alpha}}^{\overline{\beta}} (p\overline{u}_1'^{\,2} + q\overline{u}_1^2)\,dx}{\int_{\overline{\alpha}}^{\overline{\beta}} \overline{u}_1^2\,dx} \le \overline{\lambda_1}. \end{aligned}$$

Since by (i) $\widetilde{\phi}$ is not an eigenfunction relative to λ_1 for $-(pu')' + qu$ in $]\alpha,\beta[$, we conclude that

$$\lambda_1 < \overline{\lambda_1}. \tag{1.76}$$

Step 3. Notice that an eigenfunction of $-(pu')' + qu$ in $]\gamma,\delta[$ without zeros in $]\gamma,\delta[$ and vanishing at the extremal points needs to be an eigenfunction relative to the first eigenvalue of $-(pu')' + qu$ on $]\gamma,\delta[$.

We now prove (ii). Suppose that $\overline{u}$ has no zeros in $]\alpha,\beta[$. Then according to (i), $\overline{\lambda}$ is the first eigenvalue and $\overline{u}$ is a corresponding eigenfunction for $-(pu')'+qu$ in $]\overline{\alpha},\overline{\beta}[$; consequently, by (1.76) we ought to have $\overline{\lambda}>\lambda$, a contradiction.

Step 4. Let us prove (iii). Let u_k be an eigenfunction relative to λ_k for $-(pu')'+qu$ in $]\alpha,\beta[$. Since $\lambda_{k+1}>\lambda_k$, u_{k+1} needs to vanish at least more than u_k because of (ii); hence u_k has at least $k-1$ zeros in $]\alpha,\beta[$. Let $x_0=\alpha, x_1,\dots,x_{\ell-1},x_\ell=\beta$ be the zeros of u_k. For $m=1,\dots,\ell$, set

$$u^{(m)}(x):=\begin{cases}u_k(x) & \text{if } x_{m-1}\le x\le x_m,\\ 0 & \text{otherwise}\end{cases}$$

and

$$\phi(x):=\sum_{m=1}^{\ell} c_m u^{(m)}(x).$$

Choose c_1, $c_2,\dots,c_\ell$ in such a way that

$$\int_\alpha^\beta \phi u_j\,dx=0 \qquad \forall j=1,\dots,\ell-1$$

holds and then compute

$$\begin{aligned}\int_\alpha^\beta (p\phi'^2+q\phi^2)\,dx &= \sum_{m=1}^{\ell} c_m^2\int_{x_{m-1}}^{x_m}(pu_k'^2+qu_k^2)\,dx\\ &= \sum_{m=1}^{\ell} c_m^2\left(pu_ku_k'\Big|_{x_{m-1}}^{x_m}-\int_{x_{m-1}}^{x_m}u_k((pu_k')'-qu_k)\,dx\right)\\ &= \lambda_k\sum_{m=1}^{\ell} c_m^2\int_{x_{m-1}}^{x_m}u_k^2\,dx\\ &= \lambda_k\int_\alpha^\beta \phi^2\,dx\end{aligned}$$

to conclude from (1.74) that $\lambda_\ell\le\lambda_k$, i.e., $\ell\le k$. Hence, u_k has at most $k-1$ zeros in $]\alpha,\beta[$.

Step 5. We prove (iv). With the notations of Step 2, suppose by contradiction that $\overline{\lambda}_k\le\lambda_k$. If α_1 and β_1 are two consecutive zeros of u_k, then λ_k is the first eigenvalue of (1.70) in $]\alpha_1,\beta_1[$. It follows that $\overline{u_k}$ does not vanish more than twice in $]\alpha_1,\beta_1[$. Therefore there is at least one zero of u_k between two zeros of $\overline{u_k}$. Since $\overline{u}_k$ vanishes $k+1$ times in $[\overline{\alpha},\overline{\beta}]$, u_k needs to vanish at least k times in $]\alpha,\beta[$, but this contradicts Step 3. Therefore we have $\overline{\lambda}_k>\lambda_k$. □

Let $p\in C^1([\alpha,\beta])$, $0<a\le p(x)\le b\ \forall x\in[\alpha,\beta]$, $q\in C^0([\alpha,\beta])$, $q\ge 0$. Consider now the eigenvalue problem

$$\begin{cases}-(pv')'+qv=\mu v & \text{in }]\alpha,\beta[,\\ v(\alpha)=v'(\beta)=0,\end{cases}\tag{1.77}$$

i.e., the equation

$$a(u,\varphi)=\lambda\int_\alpha^\beta u\varphi\,dx,\qquad u,\varphi\in\mathcal{H},\tag{1.78}$$

where

$$a(u,\varphi) := \int_\alpha^\beta (pu'\varphi' + qu\varphi)\,dx$$

is defined in the space

$$\mathcal{H} := \{u \in H^1(]\alpha,\beta[) \,\Big|\, u(\alpha) = 0\}.$$

Observe that the bilinear form $a(u,v)$ is an inner product on the Hilbert space

$$\mathcal{H} := \{u \in H^1(]\alpha,\beta[) \,\Big|\, u(\alpha) = 0\}$$

equivalent to the standard ones. We may then apply the Courant–Hilbert–Schmidt theory to the equation

$$a(u,\varphi) = \lambda \int_\alpha^\beta u\varphi\,dx, \qquad u,\varphi \in \mathcal{H},$$

to get the theses of Theorem 1.87 for the corresponding eigenvalues μ_k and eigenfunctions, replacing H_0^1 with $\mathcal{H}$. In particular, we have the variational characterization

$$\mu_1 = \min\Big\{\frac{a(\varphi,\varphi)}{||\varphi||^2} \,\Big|\, \varphi \in \mathcal{H}\Big\}, \tag{1.79}$$

and, for $k \geq 2$,

$$\mu_k = \min\Big\{\frac{a(\varphi,\varphi)}{||\varphi||_2} \,\Big|\, \varphi \in \mathcal{H},\ a(u,\varphi) = 0 \ \text{ for every eigenfunction } u \text{ relative to one of the eigenvalues } \mu_1,\dots\mu_{k-1}\Big\}. \tag{1.80}$$

Moreover, we have the following.

1.89 Theorem (Separation theorem). *Let $p \in C^1([\alpha,\beta])$, $0 < a \leq p(x) \leq b\ \forall x \in [\alpha,\beta]$, $q \in C^0([\alpha,\beta])$, $q \geq 0$, and let $\{\mu_k\}$ be the sequence of the eigenvalues of* (1.78) *ordered increasingly. Then the following hold:*

(i) *The eigenvalues $\{\mu_k\}$ and the corresponding eigenfuncitons satisfy the properties stated in Theorem 1.88 for the sequence $\{\lambda_k\}$ of the eigenvalues of* (1.70) *and the relative eigenfunctions.*
(ii) *We have $\lambda_{k-1} \leq \mu_k \leq \lambda_k$.*

Proof. (i) follows by repeating the arguments in the proof of Theorem 1.88. (ii) remains to be proved. If v is a solution of (1.77) with $v(\alpha) = 0$ and $\lambda_{\ell-1} \leq \mu \leq \lambda_\ell$, because of (iii) and (iv) relative to the problem (1.77), $v(x)$ needs to have $\ell - 1$ zeros in $]\alpha,\beta[$; but, then, $\lambda_{k-1} < \mu_k < \lambda_k$. □

1.4.4 Convex functionals on H_0^1

As we have seen in [GM3], the Dirichlet abstract principle is a special case of the following theorem.

1.90 Theorem. *Let $\mathcal{F} : H \to \mathbb{R}$ be a functional defined on a Hilbert space H that is continuous, convex, bounded from below and coercive, i.e.,*

$$\mathcal{F}(u) > -\infty \qquad \text{and} \qquad \mathcal{F}(u) \to +\infty \text{ as } |u| \to +\infty.$$

Then $\mathcal{F}$ attains its minimum in H.

Consider now a continuous function $f = f(x,p) : \overline{\Omega} \times \mathbb{R}^n \to \mathbb{R}$ where Ω is a bounded and open set of $\mathbb{R}^n$. Suppose that

(i) $f(x,p)$ is convex in p for every x,
(ii) there are constants λ and Λ with $\Lambda \geq \lambda > 0$ such that

$$\lambda\,|p|^2 \leq f(x,p) \leq \Lambda\,|p|^2 \qquad \forall (x,p).$$

Then, for every $u \in H_0^1(\Omega)$, the function $x \to f(x, Du(x))$ is summable, and the functional $\mathcal{F} : H_0^1(\Omega) \to \mathbb{R}^N$ given by

$$\mathcal{F}(u) := \int_\Omega f(x, Du(x))\,dx$$

is well-defined. It is easily seen that $\mathcal{F}$ is convex, continuous and coercive in $H_0^1(\Omega)$. The following theorem then follows.

1.91 Theorem. *Under assumptions (i) and (ii) the functional $\mathcal{F}$ attains its minimum value in $H_0^1(\Omega)$.*

1.5 Exercises

1.92 ¶. Let $f \geq 0$ be a measurable function. Prove that the following claims are equivalent:
(i) $\int_E f(x)\,dx < \infty\ \forall E$ with $|E| < \infty$,
(ii) $\int_E f(x)\,dx \leq C\,(1 + |E|)\ \forall E$,
(iii) $f = g + h$ with $g \in L^1(\mathbb{R})$ and $h \in L^\infty(\mathbb{R})$.

1.93 ¶. Suppose that H is a subspace of $L^2(]0,1[)$ such that
(i) $f \in L^\infty(]0,1[)$ if $f \in H$,
(ii) there exists a constant $C > 0$ such that $||f||_\infty \leq C\,||f||_{L^2}\ \forall f \in H$.
Prove that H is finite-dimensional.
[*Hint.* If $(f_1, f_2, \ldots, f_n)$ is orthonormal, we have

$$\Big|\sum_{i=1}^n \alpha_i f_i(x)\Big|^2 \leq C^2 \sum_{i=1}^n \alpha_i^2 \qquad \text{for a.e. } x.]$$

1.94 ¶. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a function in $L^2(\mathbb{R})$ for which there exists a constant $C > 0$ such that

$$\int_{\mathbb{R}} |f(x+h) - f(x)|^2 \, dx \le C\,|h|^2 \qquad \forall h.$$

Show that for all $g \in C_c^1(\mathbb{R})$ we have

$$\left| \int_{\mathbb{R}} f g' \, dx \right| \le \sqrt{C} \Big(\int_{\mathbb{R}} g^2 \, dx \Big)^{1/2}.$$

1.95 ¶ Gravitational potential. Let $\mu : \mathbb{R}^3 \to \mathbb{R}_+$. The function

$$V(x) = \gamma \int_{\mathbb{R}^3} \frac{\mu(\xi)}{|\xi - x|} \, d\xi$$

is called the *gravitational potential* of the mass distribution μ.

(i) Show that $V(x)$ is well-defined if $\mu \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$.
(ii) If μ is radial, $\mu(\xi) := \nu(|\xi|)$, then V is radial, too.

1.96 ¶ Legendre's polynomials. We recall that by applying the orthonormalization process of Gram–Schmidt to the polynomials t^n in $]-1,1[$, we get the so-called Legendre's polynomials, see [GM3]. Show that they form a complete system in $L^2(]-1,1[)$.

1.97 ¶ Haar's basis. In $[0,1]$ consider the sequence of functions $\chi_n^{(k)}(t)$ defined by $\chi_0^{(0)}(t) = 1$, and, for $n \ge 0$ and $1 \le k \le 2^n$, by

$$\chi_n^{(k)}(t) := \begin{cases} 2^{n/2} & \text{in } \dfrac{k-1}{2^n} < t \le \dfrac{k-1/2}{2^n}, \\ -2^{n/2} & \text{in } \dfrac{k-1/2}{2^n} \le t < \dfrac{k}{2^n}, \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\{\chi_n^{(k)}\}$ is a complete orthonormal system in $L^2((0,1))$.
[*Hint.* Let ϕ be orthogonal to all elements of the sequence. Set $\psi(x) := \int_0^x \phi(t)\,dt$ and show by induction that $\psi(k/2^n) = 0$. For instance,

$$0 = (\psi | \chi_0^{(0)}) = \int_0^1 \phi(x)\,dx = \psi(1) - \psi(0),$$

hence $\psi(1) = \psi(0) = 0$.]

1.98 ¶. Solve the following problems:

$$\begin{cases} u_t - u_{xx} = 0 & \text{in } [0,\pi] \times \mathbb{R}_+, \\ u(0,t) = u(\pi,t) = 0 & t \in \mathbb{R}_+, \\ u(x,0) = \sin^3 x & \text{in } [0,\pi] \end{cases}$$

and

$$\begin{cases} \Delta u = 0 & \text{in } x^2 + y^2 < 1, \\ u = \cos^2\theta & \text{in } x^2 + y^2 = 1. \end{cases}$$

1.99 ¶. Let $u \in H_0^1((0,1))$. Show that (the continuous representative of) u satisfies $u(0) = u(1) = 0$.

1.100 ¶. By using the variational methods of Section 3.2, prove the following:

(i) Let $f \in L^2((0,1))$. Then there exists a solution to the problem

$$\min_{\substack{u \in H^1 \\ u(0)=\alpha,\ u(1)=\beta}} \Big\{ \frac{1}{2} \int_0^1 (u'^2 + u^2)\, dx - \int_0^1 f u\, dx \Big\};$$

the minimizer $\overline{u}$ is the unique weak solution to the Dirichlet boundary value problem

$$\begin{cases} -u'' + u = f & \text{in }]0,1[, \\ u(0) = \alpha,\ u(1) = \beta; \end{cases}$$

finally, show that $\overline{u} \in C^2([0,1])$ if $f \in C^0([0,1])$.

(ii) Let $f \in L^2((0,1))$. Then there exists a unique minimizer $\overline{u} \in H^1(]0,1[)$ of the functional

$$\frac{1}{2} \int_0^1 (u'^2 + u^2)\, dx - \int_0^1 f u\, dx;$$

$\overline{u}$ is the unique weak solution of the Neumann boundary value problem

$$\begin{cases} -u'' + u = f & \text{in }]0,1[, \\ u'(0) = 0,\ u'(1) = 0; \end{cases}$$

finally, show that $\overline{u} \in C^2([0,1])$ if $f \in C^0([0,1])$.

(iii) Let $f \in L^2((0,1))$ and $\alpha, \beta \in \mathbb{R}$. Show that there exists a unique minimizer in $H^1((0,1))$ of the functional

$$\frac{1}{2} \int_0^1 (u'^2 + u^2)\, dx - \int_0^1 f u\, dx + \alpha u(0) - \beta u(1);$$

$\overline{u}$ is the unique weak solution of the Neumann nonhomogeneous boundary value problem

$$\begin{cases} -u'' + u = f & \text{in }]0,1[, \\ u'(0) = \alpha,\ u'(1) = \beta; \end{cases}$$

finally, show that $u \in C^2([0,1])$ if $f \in C^0([\alpha, \beta])$.

1.101 ¶. Show that there exists a unique solution of the following:

(i) The mixed boundary value problem

$$\begin{cases} -u'' + u = f & \text{in }]0,1[, \\ u(0) = 0,\ u'(1) = 0. \end{cases}$$

(ii) The boundary value problem

$$\begin{cases} -u'' + u = f & \text{in }]0,1[, \\ u'(0) - k u(0) = 0,\ u(1) = 0. \end{cases}$$

(iii) The periodic boundary value problem

$$\begin{cases} -u'' + u = f & \text{in }]0,1[, \\ u(0) = u(1),\ u'(0) = u'(1). \end{cases}$$

2. Convex Sets and Convex Functions

We have encountered convex sets and convex functions on several occasions. Here we would like to discuss these notions in a more systematic way. Among nonlinear functions, the convex ones are the closest ones to the linear, in fact, functions that are convex and concave at the same time are just the linear affine functions.

Although convex figures appear since the beginning of mathematics — Archimedes, for instance, observed and made use of the fact that the perimeter of a convex figure K is larger than the perimeter of any other convex figure contained in K, more recently convexity played a relevant role in the study of the thermodynamic equilibrium by J. Willard Gibbs (1839–1903) — the systematic study of convexity began in the early years of the twentieth century with Hermann Minkowski (1864–1909), continued with the treatise of T. Bonnesen and Werner Fenchel (1905–1986) in 1934 and developed after 1950 both in finite and infinite dimensions due to its relevance in several branches of mathematics. Here we shall deal only with convexity in finite-dimensional spaces.

2.1 Convex Sets

a. Definitions

2.1 Definition. *A set $K \subset \mathbb{R}^n$ is said to be* convex *if either $K = \emptyset$ or, whenever we take two points in K, the segment that connects them is entirely contained in K, i.e.,*

$$\lambda x_1 + (1-\lambda)x_2 \in K \qquad \forall\ \lambda \in [0,1],\ \forall\ x_1, x_2 \in K.$$

The following properties, the proof of which we leave to the reader, follow easily from the definition.

2.2 ¶. Show the following:

(i) A linear subspace of $\mathbb{R}^n$ is convex.

M. Giaquinta and G. Modica, *Mathematical Analysis, Foundations and Advanced Techniques for Functions of Several Variables*, DOI 10.1007/978-0-8176-8310-8_2,

Figure 2.1. Hermann Minkowski (1864–1909) and the frontispiece of the treatise by T. Bonnesen and Werner Fenchel (1905–1986) on convexity.

(ii) Let $\ell : \mathbb{R}^n \to \mathbb{R}$ be linear and $\alpha \in \mathbb{R}$. Then the sets

$$\Big\{x \in \mathbb{R}^n \,\Big|\, \ell(x) < \alpha\Big\}, \qquad \Big\{x \in \mathbb{R}^n \,\Big|\, \ell(x) \le \alpha\Big\},$$
$$\Big\{x \in \mathbb{R}^n \,\Big|\, \ell(x) \ge \alpha\Big\}, \qquad \Big\{x \in \mathbb{R}^n \,\Big|\, \ell(x) > \alpha\Big\}$$

are convex.

(iii) The intersection of convex sets is convex; in particular, the intersection of any number of half-spaces is convex.

(iv) The interior and the closure of a convex set are convex.

(v) If K is convex, then $\mathrm{cl}(\mathrm{int}(K)) = \mathrm{cl}(K)$, $\mathrm{int}(\mathrm{cl}(K)) = \mathrm{int}(K)$.

(vi) If K is convex, then for $x_0 \in \mathbb{R}^n$ and $t \in \mathbb{R}$ the set

$$tx_0 + (1-t)K := \Big\{x \in \mathbb{R}^n \,\Big|\, x = tx_0 + (1-t)y,\ y \in K\Big\},$$

i.e., the *cone* with vertex x_0 generated by K, is convex.

A *linear combination* of points $(x_1, x_2, \ldots, x_k) \in \mathbb{R}^n$, $\sum_{i=1}^k \lambda_i x_i$, with coefficients $\lambda_1, \lambda_2, \ldots, \lambda_k$ such that $\sum_{i=1}^k \lambda_i = 1$ and $\lambda_i \ge 0\ \forall i$, is called a *convex combination* of $x_1, \ldots, x_k$. The coefficients $\lambda_1, \lambda_2, \ldots, \lambda_k$ are called the *barycentric coordinates* of $x := \sum_{i=1}^k \lambda_i x_i$.

Noticing that

$$\sum_{i=1}^{k} \lambda_i x_i = (1-\lambda_k) \sum_{i=1}^{k-1} \frac{\lambda_i}{1-\lambda_k} x_i + \lambda_k x_k,$$

whenever $0 < \lambda_k < 1$, we infer at once the following.

2.3 Proposition. *The set K is convex if and only if every convex combination of points in K is contained in K.*

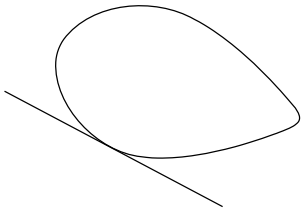


Figure 2.2. A support plane.

2.4 ¶. Show that the representation of a point x as convex combination of points $x_1, x_2, \ldots, x_k$ is unique if and only if the vectors $x_2 - x_1, x_3 - x_1, \ldots, x_k - x_1$ are linearly independent.

b. The support hyperplanes

We prove that every proper, nonempty, closed and convex subset of $\mathbb{R}^n$, $n \geq 2$, is the intersection of closed half-spaces. To do this, we first introduce the notions of *separating* and *supporting hyperplanes*.

2.5 Definition. *Let $\ell : \mathbb{R}^n \to \mathbb{R}$ be a linear function, $\alpha \in \mathbb{R}$ and $\mathcal{P}$ the hyperplane*

$$\mathcal{P} := \Big\{ x \in \mathbb{R}^n \;\Big|\; \ell(x) = \alpha \Big\},$$

and let

$$\mathcal{P}_+ := \Big\{ x \in \mathbb{R}^n \;\Big|\; \ell(x) \geq \alpha \Big\}, \qquad \mathcal{P}_- := \Big\{ x \in \mathbb{R}^n \;\Big|\; \ell(x) \leq \alpha \Big\}$$

be the corresponding half-spaces that are the closed convex sets of $\mathbb{R}^n$ for which $\mathcal{P}_+ \cup \mathcal{P}_- = \mathbb{R}^n$ and $\mathcal{P}_+ \cap \mathcal{P}_- = \mathcal{P}$. We say that

(i) *two nonempty sets $A, B \subset \mathbb{R}^n$ are* separated *by $\mathcal{P}$ if $A \subset \mathcal{P}_+$ and $B \subset \mathcal{P}_-$;*

(ii) *two nonempty sets $A, B \subset \mathbb{R}^n$ are* strongly separated *by $\mathcal{P}$ if there is $\epsilon > 0$ such that*

$$\ell(x) \leq \alpha - \epsilon \;\forall\; x \in A \qquad \text{and} \qquad \ell(x) \geq \alpha + \epsilon \;\forall\; x \in B.$$

(iii) *Let $K \subset \mathbb{R}^n$, $n \geq 2$. We say that $\mathcal{P}$ is a* supporting hyperplane *for K if $\mathcal{P} \cap \overline{K} \neq \emptyset$ and K is a subset of one of the two closed half-spaces $\mathcal{P}_+$ and $\mathcal{P}_-$ that is called a* supporting half-space *for K.*

2.6 Theorem. *Let K_1 and K_2 be two nonempty closed and disjoint convex sets. If either K_1 or K_2 is compact, then there exists a hyperplane that strongly separates K_1 and K_2.*

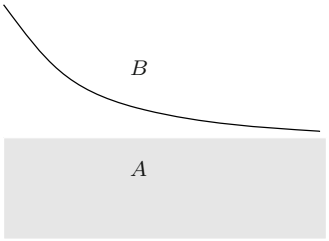


Figure 2.3. Two disjoint and closed convex sets that are not strongly separated.

Proof. Assume for instance that K_1 is compact and let $d := \inf\{|x-y| \,|\, x \in K_1,\ y \in K_2\}$. Clearly d is finite and, for R large,

$$d = \inf\Big\{|x-y| \;\Big|\; x \in K_1,\ y \in K_2 \cap \overline{B(0,R)}\Big\}.$$

The Weierstrass theorem then yields $x_0 \in K_1$ and $y_0 \in K_2 \cap \overline{B(0,R)}$ such that

$$d = |x_0 - y_0| > 0.$$

The hyperplane through x_0 and perpendicular to $y_0 - x_0$,

$$\mathcal{P}' := \Big\{x \in \mathbb{R}^n \;\Big|\; (x-x_0)\bullet(y_0-x_0) \;= 0\Big\},$$

is a supporting hyperplane for K_1. In fact, for $x \in K_1$, the function

$$\phi(\lambda) := |y_0 - (x_0 + \lambda(x-x_0))|^2, \qquad \lambda \in [0,1],$$

has a minimum at zero, hence

$$\phi'(0) = 2\ (y_0 - x_0)\bullet(x-x_0) \;\le 0. \tag{2.1}$$

Similarly, the hyperplane through y_0 and perpendicular to $x_0 - y_0$,

$$\mathcal{P}'' := \Big\{x \in \mathbb{R}^n \;\Big|\; (x-y_0)\bullet(x_0-y_0) \;= 0\Big\},$$

is a supporting hyperplane for K_2. The conclusion then follows since $\operatorname{dist}(\mathcal{P}', \mathcal{P}'') = d > 0$. □

2.7 Theorem. *We have the following:*

(i) *Every boundary point of a closed and convex set $K \subset \mathbb{R}^n$, $n \ge 2$, is contained in at least a supporting hyperplane.*
(ii) *Every closed convex set $K \ne \emptyset, \mathbb{R}^n$ of $\mathbb{R}^n$ is the intersection of all its supporting half-spaces.*
(iii) *Let $K \subset \mathbb{R}^n$ be a closed set with nonempty interior. Then K is convex if and only if at each of its boundary point there is a supporting hyperplane.*

Proof. (i) Assume $\partial K \ne \emptyset$, i.e., $K \ne \emptyset, \mathbb{R}^n$, let $x_0 \in \partial K$, and let $\{y_k\} \subset \mathbb{R}^n \setminus K$ be a sequence with $y_k \to x_0$ as $k \to \infty$. Let x_k be a point of K nearest to y_k and

$$e_k := \frac{y_k - x_k}{|y_k - x_k|}.$$

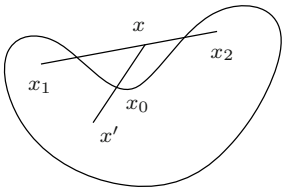


Figure 2.4. Illustration of the proof of (iii) Theorem 2.7.

Then $|e_k| = 1$, $x_k \to x_0$ as $k \to \infty$ and, as in the proof of Theorem 2.6, we see that the hyperplane through x_k and perpendicular to e_k is a supporting hyperplane for K,

$$K \subset \Big\{ x \in \mathbb{R}^n \;\Big|\; e_k \bullet (x - x_k) \le 0 \Big\}.$$

Possibly passing to a subsequence $\{e_k\}$ and $\{x_k\}$ converge, $e_k \to e$ and $x_k \to x_0$. It follows that

$$K \subset \Big\{ x \in \mathbb{R}^n \;\Big|\; e \bullet (x - x_0) \le 0 \Big\},$$

i.e., the hyperplane through x_0 perpendicular to e is a supporting hyperplane for K.

(ii) Since $K \neq \emptyset, \mathbb{R}^n$, the boundary of K is nonempty; in particular, the intersection K' of all its supporting half-spaces is closed, nonempty by (i), hence it contains K. Assume by contradiction that there is $x' \in K' \setminus K$. Since K is closed, there is a nearest point $x_0 \in K$ to x' and, as in the proof of Theorem 2.6,

$$K \subset S := \Big\{ x \in \mathbb{R}^n \;\Big|\; (x' - x_0) \bullet (x - x_0) \le 0 \Big\}.$$

On the other hand, from the definition of K', it follows that $K' \subset S$, hence $x' \in S$, which is a contradiction since $(x' - x_0) \bullet (x' - x_0) > 0$.

(iii) Let K be convex. By assumption $K \neq \emptyset$, if $K = \mathbb{R}^n$, we have $\partial K = \emptyset$ and nothing has to be proved. If $K \neq \mathbb{R}^n$, then through every boundary point there is a supporting hyperplane because of (i).

Conversely, suppose that K is not convex, in particular, $K \neq \emptyset, \mathbb{R}^n$ and $\partial K \neq \emptyset$. It suffices to show that through a point of ∂K there is no supporting hyperplane. Since K is not convex, there exists $x_1, x_2 \in K$ and x on the segment Σ connecting x_1 and x_2 with $x \notin K$. Let x' be a point in the interior of K and Σ' be the segment joining x with x'. At a point $x_0 \in \partial K \cap \Sigma'$ we claim that there is no supporting hyperplane. In fact, let π be such a hyperplane and let H be the corresponding supporting half-space. Since $x' \in \mathrm{int}(K)$, x' does not belong to π, thus Σ' is not contained in π. It follows that $x' \in \mathrm{int}(H)$ and $x \notin H$, hence x_1 and x_2 cannot both be in H since otherwise x also belongs to H. However, this contradicts the fact that H is a supporting half-space. □

2.8 ¶. In (iii) of Theorem 2.7 the assumption that $\mathrm{int}(K) \neq \emptyset$ is essential; think of a curve without inflection points in $\mathbb{R}^2$.

2.9 ¶. Let K be a closed, convex subset of $\mathbb{R}^n$ with $K \neq \emptyset, \mathbb{R}^n$.

(i) Prove that K is the intersection of at most a denumerable supporting half-spaces.

(ii) Moreover, if K is compact, then for any open set $A \supset K$ there exists finitely many supporting half-spaces such that

$$K \subset \bigcap_{k=1}^{N} H_k \subset A.$$

[*Hint.* Remember that $\mathbb{R}^n$ has a denumerable basis.]

2.10 ¶. Using Theorem 2.7, prove the following, compare Proposition 9.126 and Theorem 9.127 of [GM3].

Proposition. *Let $C \subset \mathbb{R}^n$ be an open convex subset and let $\overline{x} \notin C$. Then there exists a linear map $\ell : \mathbb{R}^n \to \mathbb{R}$ such that $\ell(x) < \ell(\overline{x})\ \forall x \in C$. In particular, C and $\overline{x}$ are separated by the hyperplane $\{x \,|\, \ell(x) = \ell(\overline{x})\}$.*

Consequently,

Theorem. *Let A and B be two nonempty disjoint convex sets. Suppose A is open. Then A and B can be separated by a hyperplane.*

2.11 Definition. *We say that K is* polyhedral *if it is the intersection of a finite number of closed half-spaces. A bounded polyhedral set is called a* polyhedron.

c. Convex hull

2.12 Definition. *The* convex hull *of a set $M \subset \mathbb{R}^n$, $\mathrm{co}(M)$, is the intersection of all convex subsets in $\mathbb{R}^n$ that contain M.*

2.13 Proposition. *The convex hull of $M \subset \mathbb{R}^n$ is convex, indeed the smallest convex set that contains M. Moreover, $\mathrm{co}(M)$ is the set of all convex combinations of points of M,*

$$\mathrm{co}(M) := \Big\{ x \in \mathbb{R}^n \,\Big|\, \exists x_1, x_2, \dots, x_N \in M \text{ such that } x = \sum_{i=1}^{N} \lambda_i x_i, \text{ for some } \lambda_1, \lambda_2, \dots, \lambda_N, \text{ where } \lambda_i \geq 0\ \forall i,\ \sum_{i=1}^{N} \lambda_i = 1 \Big\}.$$

2.14 ¶. Prove Proposition 2.13.

2.15 ¶. Prove that
(i) $\mathrm{co}(M)$ is open, if M is open,
(ii) $\mathrm{co}(M)$ is compact, if M is compact.

2.16 ¶. Give examples of sets $M \subset \mathbb{R}^2$ so that
(i) M is closed but $\mathrm{co}(M)$ is not,
(ii) $\mathrm{co}(\overline{M}) \neq \overline{\mathrm{co}(M)}$ although $\mathrm{co}(\overline{M}) \subset \overline{\mathrm{co}(M)}$.

If $M \subset \mathbb{R}^n$, then the convex combinations of at most $n+1$ points in M are sufficient to describe $\mathrm{co}(M)$. In fact, the following holds.

2.17 Theorem (Carathéodory). *Let $M \subset \mathbb{R}^n$. Then*

$$\mathrm{co}(M) := \Big\{ x \in \mathbb{R}^n \,\Big|\, x = \sum_{i=1}^{n+1} \lambda_i x_i,\ x_i \in M,\ \lambda_i \geq 0\ \forall i,\ \sum_{i=1}^{n+1} \lambda_i = 1 \Big\}.$$

Proof. Let x be a convex combination of m points $x_1, x_2, \dots, x_m$ of M with $m > n+1$,

$$x = \sum_{j=1}^{m} \lambda_j x_j, \qquad \sum_{j=1}^{m} \lambda_j = 1,\ \lambda_j > 0.$$

We want to show that x can be written as convex combinations of $m-1$ points of M.

Since $m - 1 > n$, there are numbers $c_1, c_2, \dots, c_{m-1}$ not all zero such that $\sum_{i=1}^{m-1} c_i(x_i - x_m) = 0$. If $c_m := -\sum_{i=1}^{m-1} c_i$, we have

$$\sum_{i=1}^{m} c_i x_i = 0 \qquad \text{and} \qquad \sum_{i=1}^{m} c_i = 0.$$

Since at least one of the c_i's is positive, we can find $t > 0$ and $k \in \{1, \dots, m\}$ such that

$$\frac{1}{t} = \max\Big(\frac{c_1}{\lambda_1}, \frac{c_2}{\lambda_2}, \dots, \frac{c_m}{\lambda_m}\Big) = \frac{c_k}{\lambda_k} > 0.$$

The point x is then a convex combination of $x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_m$; in fact, if

$$\gamma_j := \begin{cases} \lambda_j - t c_j & \text{if } j \neq k, \\ 0 & \text{if } j = k, \end{cases}$$

we have $\sum_{j\neq k} \gamma_j = \sum_{j=1}^{m} \gamma_j = \sum_{j=1}^{m} (\lambda_j - t c_j) = \sum_{j=1}^{m} \lambda_j = 1$ and

$$x = \sum_{j=1}^{m} \lambda_j x_j = \sum_{j=1}^{m} (\gamma_j + t c_j) x_j = \sum_{j\neq k} \gamma_j x_j.$$

We then conclude by backward induction on m. □

2.18 ¶. Prove the following:

(i) In Theorem 2.17 the number $n + 1$ is optimal.

(ii) If M is convex, then $\mathrm{co}(M) = M$ and every point in $\mathrm{co}(M)$ is a convex combination of itself.

(iii) If $M = M_1 \cup \cdots \cup M_k$, $k \leq n$, where $M_1, \dots, M_k$ are convex sets, then every point of $\mathrm{co}(M)$ is a convex combination of at most k points of M.

d. The distance function from a convex set

We conclude with a characterization of a convex set in terms of its distance function.

Let $C \subset \mathbb{R}^n$ be a nonempty closed set. For every $x \in \mathbb{R}^n$ we define

$$d_C(x) := \mathrm{dist}(x, C) := \inf\Big\{ |x - y| \,\Big|\, y \in C \Big\}.$$

It is easily seen that indeed the infimum is a minimum, i.e., there is (at least) a point $y \in C$ of least distance from x. Moreover, the function d_C is Lipschitz-continuous with Lipschitz constant 1,

$$|d_C(x) - d_C(y)| \leq |x - y| \qquad \forall x, y \in \mathbb{R}^n.$$

2.19 Lemma. *If $x \notin C$, then*

$$d_C(x+h) = d_C(x) + L(h;x) + o(|h|) \qquad \text{as } h \to 0, \tag{2.2}$$

where

$$L(h;x) := \min\Big\{ h \bullet \frac{x-z}{|x-z|} \;\Big|\; z \in C, |x-z| = d_C(x)\Big\}$$

is the minimum among the lengths of the projections of h into the lines connecting x to its nearest points $z \in C$. In particular, d_C is differentiable at x if and only if $h \to L(h;x)$ is linear, i.e., if and only if there is a unique minimum point $z \in C$ of least distance from x.

Proof. We prove (2.2), the rest easily follows. We may and do assume that $x = 0$. Moreover, we deal with the function

$$f(h) := d_C^2(h) = \min_{z \in C} |h - z|^2.$$

It suffices to show that

$$f(h) = f(0) + f'(h,0) + o(|h|), \qquad h \to 0, \tag{2.3}$$

where

$$f'(h;0) := \min\Big\{ -2\, h \bullet z \;\Big|\; z \in C, |z| = d_C(0)\Big\}.$$

First, we remark that the functions $q_\epsilon(h)$ defined for $\epsilon \geq 0$ as

$$q_\epsilon(h) := \inf\Big\{ -2\, h \bullet z \;\Big|\; |z| \leq f(0)^{1/2} + \epsilon \Big\}$$

are homogeneous of degree 1 and that $q_\epsilon \to q_0$ increasingly as $\epsilon \to 0$. By Dini's theorem, see [GM3], $\{q_\epsilon\}$ converges uniformly to q_0 in $B(0,1)$. Therefore, for every $\epsilon > 0$ there is c_ϵ such that

$$q_0(h) \geq q_\epsilon(h) \geq q_0(h) - c_\epsilon |h| \qquad \forall h \tag{2.4}$$

and $c_\epsilon \to 0$ as $\epsilon \to 0$.

Now, let us prove (2.3). Since $|y-z|^2 = |z|^2 - 2\, y \bullet z + |y|^2$, we have

$$\begin{aligned} f(h) &\leq \min_{\substack{z \in C \\ |z| = d_C(0)}} |h - z|^2 = |h|^2 + f(0) + \min_{\substack{z \in C \\ |z| = d_C(0)}} (-2\, h \bullet z\,) \\ &= f(0) + q_0(h) + |h|^2. \end{aligned} \tag{2.5}$$

On the other hand, if $|h| < \epsilon/2$, the minimum of $z \to |z-h|^2$, $z \in C$, is attained at points z_h such that $|z_h| \leq f(0)^{1/2} + \epsilon/2$, hence by (2.4)

$$\begin{aligned} f(h) = \min_{z \in C} |z-h|^2 &= \min_{\substack{z \in C \\ |z| < f(0)^{1/2} + \epsilon}} |z-h|^2 \\ &= \min_{\substack{z \in C \\ |z| < f(0)^{1/2} + \epsilon}} \Big(|z|^2 + |h|^2 - 2\, h \bullet z \Big) \\ &\geq f(0) + |h|^2 + q_\epsilon(h) \geq f(0) + q_0(h) - c_\epsilon |h| + |h|^2. \end{aligned}$$

Therefore

$$f(h) \geq f(0) + q_0(h) + o(|h|) \qquad \text{as } h \to 0,$$

which, together with (2.5), proves (2.3). □

2.20 ¶. Using (2.2), prove that, in general, if there are in C more than one nearest point to x, then

$$\lim_{t\to 0^{\pm}} \frac{d_C(x+th)-d_C(x)}{t} = \min\Big\{ h \bullet \frac{x-z}{|x-z|} \;\Big|\; z\in C, |x-z| = d_C(x)\Big\}.$$

2.21 Theorem (Motzkin). *Let $C \subset \mathbb{R}^n$ be a nonempty closed set. The following claims are equivalent:*

(i) *C is convex.*
(ii) *For all $x \notin C$ there is a unique nearest point in C to x.*
(iii) *d_C is differentiable at every point in $\mathbb{R}^n \setminus C$.*

Proof. The equivalence of (ii) and (iii) is the content of Lemma 2.19.

(i) $\Rightarrow$ (ii). If z is the nearest point in C to $x \notin C$, then $x - z - \epsilon(y-z) \in C$ if $y \in C$, therefore

$$|x-z|^2 \le |x-z-\epsilon(y-z)|^2 = |x-z|^2 - 2\,\epsilon\,(y-z)\bullet(x-z) + \epsilon^2|y-z|^2 \qquad (2.6)$$

for all $0 \le \epsilon \le 1$. For $\epsilon \to 0$ we get $(x-z)\bullet(y-z) \le 0$ and, because of (2.6) with $\epsilon = 1$

$$|x-y|^2 = |x-z|^2 - 2\,(x-z)\bullet(y-z) + |y-z|^2 > |x-z|^2 \qquad \forall y \in C.$$

(ii) $\Rightarrow$ (i). Suppose C is not convex. It suffices to show that there is a ball B such that $B \cap C = \emptyset$ and $\overline{B} \cap C$ has more than a point. Since C is not convex, there exist $x_1, x_2 \in C$, $x_1 \ne x_2$, such that the open segment connecting x_1 to x_2 is contained in $\mathbb{R}^n \setminus C$. We may suppose that the middle point of this segment is the origin, i.e., $x_2 = -x_1$, and let ρ be such that $\overline{B(0,\rho)} \cap C = \emptyset$. We now consider the family of balls $\{B(w,r)\}$ such that

$$B(w,r) \supset B(0,\rho), \qquad B(w,r)\cap C = \emptyset \qquad (2.7)$$

and claim that the corresponding set $\{(w,r)\} \subset \mathbb{R}^{n+1}$ is bounded and closed, hence compact. In fact, since $x_j \notin B(w,r)$, $j = 1, 2$, we have $r \ge |w| + \rho$ and $|w \pm x_1|^2 \ge r^2$, hence

$$(|w|+\rho)^2 \le r^2 \le \frac{1}{2}(|w+x_1|^2 + |w-x_1|^2) \le |w|^2 + r^2$$

from which we infer

$$|w| \le \frac{|x_1|^2 - \rho^2}{2\rho}, \qquad r \le \frac{(|x_1|^2+\rho^2)}{2\rho}.$$

Consider now a ball $B(w_0, r_0)$ with maximal radius r_0 among the family (2.7). We claim that $\partial B(w_0,r_0) \cap C$ contains at least two points. Assuming on the contrary that $\partial B(w_0,r_0)\cap C$ contains only one point y_1, for all θ such that $\theta\bullet(y_1 - w_0) < 0$ and for all $\epsilon > 0$ sufficiently small, $\overline{B(w_0+\theta\epsilon, r_0)} \cap C = \emptyset$, consequently, by maximality there exists y_ϵ such that

$$y_\epsilon \in \partial B(w_0+\epsilon\theta, r_0) \cap \partial B(0,\rho). \qquad (2.8)$$

From (2.8) we infer, as $\epsilon \to 0$, that there is a point $y_2 \in \partial B(w_0,r_0)\cap\partial B(0,\rho)$, which is unique since $r_0 > \rho$. However, if we choose $\overline{\theta} := y_2 - y_1$, we surely have $\partial B(w_0 + \epsilon\overline{\theta}, r_0)\cap\partial B(0,\rho) = \emptyset$, for sufficiently small ϵ. This contradicts (2.8). □

e. Extreme points

2.22 Definition. *Let $K \subset \mathbb{R}^n$ be a nonempty convex set. A point $x_0 \in K$ is said to be an* extreme point *for K if there are no $x_1, x_2 \in K$ and $\lambda \in]0,1[$ such that $x_0 = \lambda x_1 + (1-\lambda)x_2$.*

The extreme points of a cube are the vertices; the extreme points of a ball are all its boundary points. The extreme points of a set, if any, are boundary points; in particular, an open convex set has no extreme points. Additionally, a closed half-space has no extreme points.

2.23 Theorem. *Let $K \subset \mathbb{R}^n$ be nonempty closed and convex.*

(i) *If K does not contains lines, then K has extreme points.*
(ii) *If K is compact, then K is the convex hull of its extreme points.*

Proof. Let us prove (ii) by induction on the dimension of the smallest affine subspace containing K. We leave then to the reader the task of proving (i), still by induction. If $n = 1$, K is a segment and the claim is trivial. Suppose that the claim holds for convex sets contained in an affine subspace of dimension $n-1$. For $x_0 \in \partial K$, let $\mathcal{P}$ be a supporting hyperplane to K at x_0. The set $K \cap \mathcal{P}$ is compact and convex, hence by the inductive assumption, x_0 is a convex combination of extreme points of $K \cap \mathcal{P}$, that are also extreme points of K. If x_0 is an interior point of K, every line through x_0 cuts K into a segment of extremes x_1 and $x_2 \in \partial K$, hence x_0 is a convex combination of extreme points, since so are $x_1, x_2 \in \partial K$. □

2.2 Proper Convex Functions

a. Definitions

We have already introduced convex functions of one variable, discussed their properties and illustrated a few estimates related to the notion of convexity, see [GM1] and Section 5.3.7 of [GM4]. Here we shall discuss convex functions of several variables.

2.24 Definition. *A function $f : K \subset \mathbb{R}^n \to \mathbb{R}$ defined on a convex set K, is said to be* convex *in K if*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \qquad \forall x, y \in K, \quad \forall \lambda \in [0,1]. \tag{2.9}$$

The function f is said to be strictly convex *if the inequality in* (2.9) *for $x \neq y$ and $0 < \lambda < 1$ is strict.*

We say that $f : K \to \mathbb{R}$ is concave *if K is convex and $-f : K \to \mathbb{R}$ is convex.*

The convexity of K is needed to ensure that the segment $\{z \in \mathbb{R}^n \mid z = \lambda x + (1-\lambda)y,\ \lambda \in [0,1]\}$ belongs to the domain of definition of f. The geometric meaning of the definition is clear: The segment PQ connecting the point $P = (x, f(x))$ to $Q = (y, f(y))$ lies above the graph of the restriction of f to the segment with extreme points $x, y \in K$.

2.25 ¶. Prove the following.

(i) Linear functions are both convex and concave; in fact, they are the only functions that are at the same time convex and concave.
(ii) If f and g are convex, then $f+g, \alpha f,\ \alpha > 0$, $\max(f,g)$ and $\lambda f+(1-\lambda)g$, $\lambda \in [0,1]$, are convex.
(iii) If $f : K \to \mathbb{R}$ is convex and $g : I \supset f(K) \to \mathbb{R}$ is convex and not decreasing, then $g \circ f$ is convex.
(iv) The functions $|x|^p$, $(1+|x|^2)^{p/2}$, $p \geq 1$, $e^{\theta|x|}$, $\theta > 0$, and $x \log x - x$, $x > 0$, are convex.

b. A few characterizations of convexity

We recall that the *epigraph* of a function $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is the subset of $\mathbb{R}^n \times \mathbb{R}$ given by

$$\mathrm{Epi}(f) := \Big\{(x,z) \,\Big|\, x \in A,\ z \in \mathbb{R},\ z \geq f(x)\Big\}.$$

2.26 Proposition. *Let $f : K \subset \mathbb{R}^n \to \mathbb{R}$. The following claims are equivalent:*

(i) *K is convex, and $f : K \to \mathbb{R}$ is convex.*
(ii) *The epigraph of f is a convex set in $\mathbb{R}^{n+1}$.*
(iii) *For every $x_1, x_2 \in K$ the function $\varphi(\lambda) := f(\lambda x_1 + (1-\lambda)x_2)$, $\lambda \in [0,1]$, is well-defined and convex.*
(iv) (Jensen's inequality) *K is convex and for any choice of m points $x_1, x_2, \ldots, x_m \in K$, and nonnegative numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that $\sum_{i=1}^m \alpha_i = 1$, we have*

$$f\Big(\sum_{i=1}^m \alpha_i x_i\Big) \leq \sum_{i=1}^m \alpha_i f(x_i).$$

Proof. (i) $\Longrightarrow$ (ii) follows at once from the definition of convexity.

(ii) $\Longrightarrow$ (i). Let $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ be the projection map into the first factor, $\pi((x,t)) := x$. Since linear maps map convex sets into convex sets and $K = \pi(\mathrm{Epi}(f))$, we infer that K is a convex set, while the convexity of f follows just by definition.

(i)⇒(iii). For $\lambda, t, s \in [0,1]$ we have

$$\begin{aligned}\varphi(\lambda t + (1-\lambda)s) &= f\Big([\lambda t + (1-\lambda)s]x_1 + [1-\lambda t - (1-\lambda)s]x_2\Big)\\ &= f\Big(\lambda[tx_1 + (1-t)x_2] + (1-\lambda)[sx_1 + (1-s)x_2]\Big)\\ &\leq \lambda\varphi(t) + (1-\lambda)\varphi(s).\end{aligned}$$

(iii)⇒(i). We have

$$\begin{aligned}f(\lambda x_1 + (1-\lambda)x_2) &= \varphi(\lambda) = \varphi(\lambda\cdot 1 + (1-\lambda)\cdot 0)\\ &\leq \lambda\varphi(1) + (1-\lambda)\varphi(0) = \lambda f(x_1) + (1-\lambda)f(x_2).\end{aligned}$$

(iv)⇒(i). Trivial.

(i)⇒(iv). We proceed by induction on m. If $m = 1$, the claim is trivial. For $m > 1$, let $\alpha := \alpha_1 + \cdots + \alpha_{m-1}$, so that $\alpha_m = 1 - \alpha$. We have

$$\sum_{i=1}^{m} \alpha_i x_i = \alpha \sum_{i=1}^{m-1} \frac{\alpha_i}{\alpha} x_i + (1-\alpha) x_m,$$

with $0 \le \alpha_i/\alpha \le 1$ and $\sum_{i=1}^{m-1}(\alpha_i/\alpha) = 1$. Therefore we conclude, using the inductive assumption, that

$$\begin{aligned} f\Big(\sum_{i=1}^{m} \alpha_i x_i\Big) &\le \alpha f\Big(\sum_{i=1}^{m-1} \frac{\alpha_i}{\alpha} x_i\Big) + (1-\alpha) f(x_m) \\ &\le \alpha \sum_{i=1}^{m-1} \frac{\alpha_i}{\alpha} f(x_i) + (1-\alpha) f(x_m) = \sum_{i=1}^{m} \alpha_i x_i. \end{aligned}$$

□

From (ii) of Proposition 2.26 and Carathéodory's theorem, Theorem 2.17, we infer at once the following.

2.27 Corollary. *Let $K \subset \mathbb{R}^n$ be a convex set. The function $f : K \subset \mathbb{R}^n \to \mathbb{R}$ is convex if and only if*

$$f(x) := \inf\Big\{ \sum_{i=1}^{n+1} \lambda_i f(x_i) \,\Big|\, \forall x_1, x_2, \ldots, x_{n+1} \in K \text{ such that } x = \sum_{i=1}^{n+1} \lambda_i x_i, \text{ with } \lambda_i \ge 0,\ \sum_{i=1}^{n+1} \lambda_i = 1 \Big\}.$$

Of course, the level sets $\{x \in K \mid f(x) \le c\}$ and $\{x \in K \mid f(x) < c\}$ of a convex function $f : K \to \mathbb{R}$ are convex sets; however, there exist nonconvex functions whose level sets are convex; for instance, the function x^3, $x \in \mathbb{R}$, or, more generally, the composition of a convex function $f : K \to \mathbb{R}$ with a monotone function $\varphi : \mathbb{R} \to \mathbb{R}$.

2.28 Definition. *A function with convex level sets is called a* quasiconvex *function.*[1]

c. Support function

Let $f : K \subset \mathbb{R}^n \to \mathbb{R}$ be a function. We say that a linear function $\ell : \mathbb{R}^n \to \mathbb{R}$ is a *support function* for f at $x \in K$ if

$$f(y) \ge f(x) + \ell(y - x) \qquad \forall y \in K.$$

[1] We notice that "quasiconvex" is used with different meanings in different contexts.

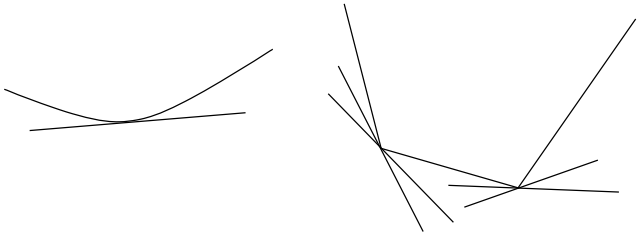


Figure 2.5. Convex functions and supporting affine hyperplanes.

2.29 Definition. *Let $f : K \to \mathbb{R}$ be a convex function. The set of linear maps $\ell : \mathbb{R}^n \to \mathbb{R}$ such that $y \to f(x) + \ell(y - x)$ is a support function for f at x is called the* subdifferential *of f at x and denoted by $\partial f(x)$.*

Trivially, if $\ell \in \partial f(x)$, then the graph of $y \mapsto f(x)+\ell(y-x)$ at $(x, f(x))$ is a supporting hyperplane for the epigraph of f at $(x, f(x))$. Conversely, on account of Proposition 2.30, every affine supporting hyperplane to $\mathrm{Epi}(f)$ is the graph of a linear map belonging to the subdifferential to f at x provided it contains no vertical vectors. This is the case if f is convex on an open set, as shown by the following proposition.

2.30 Proposition. *Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be a function, where Ω is convex and open. Then f is convex if and only if for every $x \in \Omega$ there is a linear support function for f at x.*

Proof. Let f be convex and $\overline{x} \in \Omega$. The epigraph of f is convex and its closure is convex; moreover, $(\overline{x}, f(\overline{x})) \in \partial \,\mathrm{Epi}(f)$. Consequently, there is a supporting hyperplane $\mathcal{P}$ of $\mathrm{Epi}(f)$ at $(\overline{x}, f(\overline{x}))$ that does not contain vertical vectors, otherwise it would divide Ω in two parts and, as a consequence, the epigraph of f. We then conclude that there exist a linear map $\varphi : \mathbb{R}^n \to \mathbb{R}$ and constants $\alpha, \beta \in \mathbb{R}$ such that $\mathcal{P} = \{(x, y) \,|\, \varphi(x) + \alpha y = \beta\}$ and

$$\varphi(x - \overline{x}) + \alpha(y - f(\overline{x})) \geq 0 \;\forall (x, y) \in \mathrm{Epi}(f), \qquad \alpha \neq 0. \tag{2.10}$$

Moreover, we have $\alpha \geq 0$ since in (2.10) we can choose y arbitrarily large. Thus, $\alpha > 0$ and, if we set $\ell(x) := -\varphi(x)/\alpha$, from (2.10) with $y = f(x)$, we infer

$$f(x) \geq f(\overline{x}) + \ell(x - \overline{x}) \qquad \forall x \in \Omega.$$

Conversely, let us prove that $f : \Omega \to \mathbb{R}$ is convex if it has at every point a linear support function. Let $x_1, x_2 \in \Omega$, $x_1 \neq x_2$, and $\lambda \in]0, 1[$, set $x_0 := \lambda x_1 + (1 - \lambda)x_2$, $h := x_1 - x_0$, so that $x_2 = x_0 - \frac{\lambda}{1-\lambda} h$. Let ℓ be the linear support function for f at x_0. We have

$$f(x_1) \geq f(x_0) + \ell(h), \qquad f(x_2) \geq f(x_0) - \frac{\lambda}{1 - \lambda}\ell(h).$$

Multiplying the first inequality by $\lambda/(1 - \lambda)$ and summing to the second, we get

$$\frac{\lambda}{1 - \lambda} f(x_1) + f(x_2) \geq \left(\frac{\lambda}{1 - \lambda} + 1\right) f(x_0),$$

i.e., $f(x_0) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)$. □

2.31 Remark. A consequence of the above is the following claim that complements Jensen's inequality. With the same notation of Proposition 2.26, *if f is* strictly convex *and $\alpha_i > 0$ $\forall i$, then the equality*

$$f\left(\sum_{i=1}^{m} \alpha_i x_i\right) = \sum_{i=1}^{m} \alpha_i f(x_i) \tag{2.11}$$

implies that $x_j = x_0$ $\forall j = 1, \dots, m$ where $x_0 := \sum_{i=1}^{m} \alpha_i x_i$. In fact, if $\ell(x) := f(x_0) + m \bullet (x - x_0)$ is a linear affine support function for f at x_0, the function

$$\psi(x) := f(x) - f(x_0) - m \bullet (x - x_0), \qquad x \in K$$

is nonnegative and, because of (2.11),

$$\sum_{i=1}^{m} \psi(x_i) = 0.$$

Hence $\psi(x_j) = 0$ $\forall j = 1 \dots, m$. Since ψ is strictly convex, we conclude that $x_j = x_0$ $\forall j = 1, \dots, m$.

d. Convex functions of class C^1 and C^2

We now present characterizations of smooth convex function in an open set.

2.32 Theorem. *Let Ω be an open and convex set in $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}$ be a function of class C^1. The following claims are equivalent:*

(i) *f is convex.*
(ii) *For all $x_0 \in \Omega$, the graph of f lies above the tangent plane to the graph of f at $(x_0, f(x_0))$,*

$$f(x) \geq f(x_0) + \nabla f x_0 \bullet (x - x_0) \qquad \forall x_0, x \in \Omega. \tag{2.12}$$

(iii) *The differential of f is a* monotone operator*, i.e.,*

$$\Big(\nabla f(y) - \nabla f(x)\Big) \bullet (y - x) \geq 0 \qquad \forall x, y \in \Omega. \tag{2.13}$$

Notice that in one dimension the fact that ∇f is monotone means simply that f' is increasing. Actually, we could deduce Theorem 2.32 from the analogous theorem in one dimension, see [GM1], but we prefer giving a self-contained proof.

Proof. (i)$\Longrightarrow$(ii). Let $x_0, x \in \Omega$ and $h := x - x_0$. The function $t \mapsto f(x_0 + th)$, $t \in [0,1]$, is convex, hence $f(x_0 + th) \le tf(x_0 + h) + (1-t)f(x_0)$, i.e.,

$$f(x_0 + th) - f(x_0) \le t[f(x_0 + h) - f(x_0)].$$

We infer

$$\frac{f(x_0 + th) - f(x_0)}{t} - \nabla f(x_0) \bullet h \le f(x_0 + h) - f(x_0) - \nabla f(x_0) \bullet h\,.$$

Since for $t \to 0^+$ the left-hand side converges to zero, we conclude that the right-hand side, which is independent from t, is nonnegative.

(ii)$\Longrightarrow$ (i). Let us repeat the argument in the proof of Proposition 2.30. For $x \in \Omega$ the map $h \to f(x) + \nabla f(x) \bullet h$ is a support function for f at x. Let $x_1, x_2 \in \Omega$, $x_1 \neq x_2$, and let $\lambda \in]0,1[$. We set $x_0 := \lambda x_1 + (1-\lambda)x_2$, $h := x_1 - x_0$, hence $x_2 = x_0 - \frac{\lambda}{1-\lambda}h$. From (2.12) we infer

$$f(x_1) \ge f(x_0) + \nabla f(x_0) \bullet h\,, \qquad f(x_2) \ge f(x_0) - \frac{\lambda}{1-\lambda}\nabla f(x_0) \bullet h\,.$$

Multiplying the first inequality by $\lambda/(1-\lambda)$ and summing to the second we get

$$\frac{\lambda}{1-\lambda} f(x_1) + f(x_2) \ge \left(\frac{\lambda}{1-\lambda} + 1\right) f(x_0),$$

i.e., $f(x_0) \le \lambda f(x_1) + (1-\lambda) f(x_2)$.

(ii)$\Rightarrow$(iii). Trivially, (2.12) yields

$$f(x) - f(y) \le \nabla f(x) \bullet (x - y)\,, \qquad f(x) - f(y) \ge \nabla f(y) \bullet (x - y)\,,$$

hence

$$\nabla f(y) \bullet (x - y) \le f(x) - f(y) \le \nabla f(x) \bullet (x - y)\,,$$

i.e., (2.13).

(iii)$\Rightarrow$(ii). Assume now that (2.13). For $x_0, x \in \Omega$ we have

$$f(x) - f(x_0) = \int_0^1 \frac{d}{dt} f(tx + (1-t)x_0)\, dt = \left(\int_0^1 \nabla f(tx + (1-t)y)\, dt\right) \bullet (x - x_0)$$

and

$$\Big(\nabla f(tx + (1-t)y)\Big) \bullet (x - x_0) \ge \nabla f(x_0) \bullet (x - x_0)\,,$$

hence

$$f(x) - f(x_0) \ge \left(\int_0^1 \nabla f(x_0)\, dt\right) \bullet (x - x_0) = \nabla f(x_0) \bullet (x - x_0)\,.$$

□

Let f belong to $C^2(\Omega)$. Because of (iii) of Proposition 2.26, $f : \Omega \to \mathbb{R}$ is convex if and only if for every $x_1, x_2 \in \Omega$ the function

$$\varphi(\lambda) := f((1-\lambda)x_1 + \lambda x_2) \qquad \lambda \in [0,1]$$

is convex and $C^2([0,1])$. By Theorem 2.32 φ is convex if and only if φ' is increasing in $[0,1]$, i.e., if and only if $\varphi'' \ge 0$. Since

$$\varphi''(0) = \Big(\mathbf{H} f(x_1)(x_2 - x_1)\Big) \bullet (x_2 - x_1)\,,$$

we conclude the following.

2.33 Theorem. *Let $\Omega \subset \mathbb{R}^n$ be an open and convex set of $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}$ be a function of class $C^2(\Omega)$. Then f is convex if and only if the Hessian matrix of f is nonnegative at every point in Ω,*

$$\mathbf{H}f(x)h \bullet h \geq 0 \qquad \forall x \in \Omega, \quad \forall h \in \mathbb{R}^n.$$

Similarly, one can prove that *f is strictly convex if the Hessian matrix of f is positive at every point in Ω.*

Notice that $f(x) = x^4$, $x \in \mathbb{R}$, is strictly convex, but $\mathbf{H}f(0) = 0$.

2.34 ¶. Let $f : K \subset \mathbb{R}^n \to \mathbb{R}$ be a convex function, K being convex and bounded. Prove the following:

(i) In general, f has no maximum points.

(ii) If f is not constant, then f has no interior maximum point; in other words, if f is not constant, then

$$f(x) < \sup_{y \in K} f(y) \qquad \forall x \in \operatorname{int}(K);$$

possible maximum points lie on ∂K if K is closed.

(iii) if K has extremal points, possible maximum points lie on the extremal points of K; in the case that K has finite many extremal points, then f has a maximum point and

$$\max_{x \in K} f(x) = \max_{i=1,N} f(x_i).$$

(iv) In general, f has no minimum points.

(v) The set of minimum points is convex and reduces to a point if f is strictly convex.

(vi) Local minimum points are global minimum points.

In particular, from (iii) it follows that if $f : Q \to \mathbb{R}$ is convex, Q being a closed cube in $\mathbb{R}^n$, then f has maximum and the maximum points lie on the vertices of Q.

e. Lipschitz continuity of convex functions

Let $f : Q \subset \mathbb{R}^n \to \mathbb{R}$ be a convex function defined on a closed cube Q. Then it is easy to see that $f(x) \leq \sup_{\partial Q} f$ for every $x \in Q$. Moreover, one sees by downward induction that f has maximum and the maximum points lie on the vertices of Q, see Exercise 2.34.

2.35 Theorem. *Let $\Omega \subset \mathbb{R}^n$ be an open and convex set and let $f : \Omega \to \mathbb{R}$ be convex. Then f is locally Lipschitz in Ω.*

Proof. Let $x_0 \in \Omega$ and let $Q(x_0, r)$ be a sufficiently small closed cube contained in Ω with sides of length $2r$ parallel to the axes. Since f is convex, $f_{|Q(x_0,r)}$ has maximum value at one of the vertices of $Q(x_0, r)$. If

$$L_r := \sup_{x \in \partial B(x_0,r)} f(x),$$

then $L_r < +\infty$ since $\partial B(x_0, r) \subset Q(x_0, r)$. Let us prove that

$$|f(x) - f(x_0)| \leq \frac{L_r - f(x_0)}{r} |x - x_0| \qquad \forall x \in B(x_0, r). \tag{2.14}$$

Without loss in generality, we may assume $x_0 = 0$ and $f(0) = 0$. Let $x \neq 0$ and let $x_1 := \frac{r}{|x|} x$ and $x_2 := -\frac{r}{|x|} x$. Since $x_1 \in \partial B(x_0, r)$ and $x = \lambda x_1 + (1 - \lambda)0$, $\lambda := |x|/r$, the convexity of f yields

$$f(x) \le \frac{|x|}{r} f(x_1) \le \frac{L_r}{r}|x|,$$

whereas, since $x_2 \in \partial B(x_0, r)$ and $0 = \lambda x + (1-\lambda)x_2$, $\lambda := 1/(1+|x|/r)$, we have $0 = f(0) \le \lambda f(x) + (1-\lambda) f(x_2) \le (1-\lambda) L_r$, i.e.,

$$-f(x) \le \frac{1-\lambda}{\lambda} L_r = \frac{L_r}{r}|x|.$$

Therefore, $|f(x)| \le (L_r/r)|x|$ for all $x \in B(0, r)$, and (2.14) is proved.

In particular, (2.14) tells that f *is continuous in* Ω.

Let K and K_1 be two compact sets in Ω with $K \subset\subset K_1 \subset \Omega$ and let $\delta := \operatorname{dist}(K, \partial K_1) > 0$. Let M_1 denote the oscillation of f in K_1,

$$M_1 := \sup_{x,y \in K_1} |f(x) - f(y)|,$$

which is finite by the Weierstrass theorem. For every $x_0 \in K$, the cube centered at x_0 with sides parallel to the axes of length $2r$, $r = \delta/\sqrt{n}$, is contained in K_1. It follows from (2.14) that

$$|f(x) - f(x_0)| \le \frac{L_r - f(x_0)}{r}|x - x_0| \le \frac{M_1}{r}|x - x_0| \qquad \forall x \in K \cap B(x_0, r).$$

On the other hand, for $x \in K \setminus B(x_0, r)$ we have $|x - x_0| \ge r$, hence

$$|f(x) - f(x_0)| \le M_1 \le \frac{M_1}{r}|x - x_0|.$$

In conclusion, for every $x \in K$

$$|f(x) - f(x_0)| \le \frac{M_1}{r}|x - x_0|$$

and, x_0 being arbitrary in K (and M_1 and r independent from r and x_0), we conclude that f is Lipschitz-continuous in K with Lipschitz constant smaller than M_1/r. □

Actually, the above argument shows more: *A locally equibounded family of convex functions is also locally equi-Lipschitz.*

f. Supporting planes and differentiability

2.36 Theorem. *Let $\Omega \subset \mathbb{R}^n$ be open and convex and let $f : \Omega \to \mathbb{R}$ be convex. Then f has a unique support function at x_0 if and only if f is differentiable at x_0.*

In this case, of course, the supporting function is the linear tangent map to f at x_0,

$$y \mapsto \nabla f(x_0) \bullet y\,.$$

As a first step, we prove the following proposition.

2.37 Proposition. *Let $\Omega \subset \mathbb{R}^n$ be open and convex, let $f : \Omega \to \mathbb{R}$ be convex and let $x_0 \in \Omega$. For every $v \in \mathbb{R}^n$ the right and left derivatives defined by*

$$\frac{\partial f}{\partial v^+}(x) := \lim_{t \to 0^+} \frac{f(x+tv) - f(v)}{t},$$

$$\frac{\partial f}{\partial v^-}(x) := \lim_{t \to 0^-} \frac{f(x+tv) - f(v)}{t},$$

exist and $\frac{\partial f}{\partial v^-}(x_0) \le \frac{\partial f}{\partial v^+}(x_0)$. *Moreover, for any* $m \in \mathbb{R}$ *such that* $\frac{\partial f}{\partial v^-}(x) \le m \le \frac{\partial f}{\partial v^-}(x)$, *there exists a linear map* $\ell : \mathbb{R}^n \to \mathbb{R}$ *such that* $f(x) \ge f(x_0) + \ell(x - x_0)\ \forall x \in \Omega$ *and* $\ell(v) = m$.

Proof. Without loss in generality we assume $x_0 = 0$ and $f(0) = 0$.

The function $\varphi(t) := f(tv)$ is convex in an interval around zero; thus, compare [GM1], φ has right-derivative $\varphi'_+(0)$ and left-derivative $\varphi'_-(0)$ and $\phi'_-(0) \le \varphi'_+(0)$. Since $\frac{\partial f}{\partial v^-}(0) = \varphi'_-(0)$ and $\frac{\partial f}{\partial v^+}(0) = \varphi'_+(0)$, the first part of the claim is proved.

(ii) If $\frac{\partial f}{\partial v^-}(0) \le m \le \frac{\partial f}{\partial v^+}(0)$, the graph of the linear map $t \to mt$ is a supporting line for $\mathrm{Epi}(\varphi)$ at $(0,0)$, i.e., for $\mathrm{Epi}(f) \cap V_0 \times \mathbb{R}$, $V_0 := \mathrm{Span}\{v\}$. We now show that the graph of the linear function $\ell_0 : V_0 \to \mathbb{R}$, $\ell_0(tv) := mt$, *extends to a supporting hyperplane to* $\mathrm{Epi}(f)$ *at* $(0, f(0))$, *which is in turn the graph of a linear map* $\ell : \mathbb{R}^n \to \mathbb{R}$.

Choose a vector $w \in \mathbb{R}^n$ with $w \notin V_0$, and remark that for $x, y \in V_0$ and $r, s > 0$ we have

$$\begin{aligned}\frac{r}{r+s}\ell_0(x) + \frac{s}{r+s}\ell_0(y) &= \ell_0\Big(\frac{r}{r+s}x + \frac{s}{r+s}y\Big)\\ &\le f\Big(\frac{r}{r+s}x + \frac{s}{r+s}y\Big) = f\Big(\frac{r}{r+s}(x - sw) + \frac{s}{r+s}(y + rw)\Big)\\ &\le \frac{r}{r+s}f(x - sw) + \frac{s}{r+s}f(y + rw);\end{aligned}$$

so that multiplying by $r + s$ we get

$$r\ell_0(x) + s\ell_0(y) \le rf(x - sw) + sf(x + rw),$$

i.e.,

$$g(x,s) := \frac{\ell_0(x) - f(x - sw)}{s} \le \frac{f(y + rw) - \ell_0(y)}{r} =: h(y,r).$$

For $\overline{x} \in V_0 \cap \Omega$ and $\overline{s}$ sufficiently small so that $\overline{x} + \overline{s}w$ and $\overline{x} - \overline{s}w$ ly in Ω, the values $g(\overline{x}, \overline{s})$ and $h(\overline{x}, \overline{s})$ are finite, hence

$$-\infty < g(\overline{x}, \overline{s}) \le \sup_{V_0 \times \mathbb{R}} g(x,s) \le \inf_{V_0 \times \mathbb{R}} h(x,s) \le h(\overline{x}, \overline{s}) < +\infty.$$

Consequently, there exists $\alpha \in \mathbb{R}$ such that

$$\frac{\ell_0(x) - f(x - sw)}{s} \le -\alpha \le \frac{f(x + rw) - \ell_0(x)}{r}$$

for all $x \in V_0$, $r, s \ge 0$ with $x - sw,\ x + rw \in \Omega$. This yields

$$\ell_0(x) + \alpha t \le f(x + tw) \qquad \forall x \in V_0,\ \forall t \in \mathbb{R} \text{ with } x + tw \in \Omega.$$

In conclusion, ℓ_0 has been extended to the linear function $\ell_1 : \mathrm{Span}\{v, w\} \to \mathbb{R}$ defined by $\ell_1(v) := \ell_0(v)$, $\ell_1(w) := \alpha$ for which $\ell_1(z) \le f(z)$ for all $z \in \mathrm{Span}\{v, w\}$. Of course, repeating the argument for finite many directions concludes the proof. □

Proof of Theorem 2.36. Without loss in generality, we assume $x_0 = 0$ and $f(0) = 0$.

Suppose that $\mathrm{Epi}(f)$ has a unique supporting hyperplane at 0. The restriction of f to any of the straight lines $\mathrm{Span}\, v$ through 0 has a unique support line since otherwise, as in Proposition 2.37, we could construct two different hyperplanes supporting $\mathrm{Epi}(f)$ at $(0,0)$. In particular, $\frac{\partial f}{\partial v^-}(0) = \frac{\partial f}{\partial v^+}(0)$, i.e., f is differentiable in the direction v at 0. Then, from Proposition 2.38, we conclude that f is differentiable at 0.

Conversely, suppose that f is differentiable in any direction and let $\ell : \mathbb{R}^n \to \mathbb{R}$ be a linear function, the graph of which is a supporting hyperplane for $\mathrm{Epi}(f)$ at $(0,0)$. Then $\ell(x) \le f(x)$ for all $x \in \Omega$ and, for every $v \in \mathbb{R}^n$ and $t > 0$ small,

$$\ell(v) = \frac{\ell(tv)}{t} \le \frac{f(tv)}{t}.$$

For $t \to 0^+$ we get $\ell(v) \le \frac{\partial f}{\partial v}(0)$; replacing v with $-v$ we also have $\ell(-v) \le \frac{\partial f}{\partial(-v)}(0)$, thus $\ell(v) = \frac{\partial f}{\partial v}(0)$, i.e., ℓ is uniquely defined. □

2.38 Proposition. *Let $\Omega \subset \mathbb{R}^n$ be open and convex and let $f : \Omega \to \mathbb{R}$ be convex. Then f is differentiable at $x_0 \in \Omega$ if and only if f has partial derivatives at x_0.*

Proof. We may and do assume that $x_0 = 0$ and $f(0) = 0$. Therefore, assume f is convex and has partial derivatives at 0. Additionally,

$$\phi(h) := f(h) - f(0) - \nabla f(0) \bullet h\,, \qquad h \in \Omega,$$

is convex and has zero partial derivatives at 0. Writing $h = \sum_{i=1}^n h^i e_i$, we have for every $i = 1, \dots, n$

$$\frac{\phi(nh^i e_i)}{nh^i} = o(1), \qquad h^i \to 0;$$

additionally, Jensen's inequality yields

$$\phi(h) = \phi\Big(\frac{1}{n}\sum_{i=1}^n h^i n e_i\Big) \le \frac{1}{n}\sum_{i=1}^n \phi(nh^i e_i).$$

Using Cauchy's inequality we then get

$$\phi(h) \le \sum_{i=1}^n h^i \frac{\phi(h^i n e_i)}{nh^i} \le |h| \Big(\sum_{i=1}^n \Big|\frac{\phi(h^i n e_i)}{h^i n}\Big|^2\Big)^{1/2} = |h|\epsilon(h)$$

where

$$\epsilon(h) := \Big(\sum_{i=1}^n \Big|\frac{\phi(h^i n e_i)}{h^i n}\Big|^2\Big)^{1/2}.$$

Notice that $\epsilon(h) \ge 0$, and $\epsilon(h) \to 0$ as $h \to 0$. Replacing h with $-h$ we also get

$$\phi(-h) \le |h|\epsilon(-h) \qquad \text{with } \epsilon(-h) \ge 0, \text{ and } \epsilon(-h) \to 0 \text{ as } h \to 0.$$

Since $\phi(h) \ge -\phi(-h)$ (in fact, $0 = \phi((h-h)/2) \le \phi(h)/2 + \phi(-h)/2$) we obtain

$$-|h|\epsilon(-h) \le \phi(-h) \le \phi(h) \le |h|\epsilon(h)$$

and conclude that

$$\Big|\frac{\phi(h)}{h}\Big| \le \max(\epsilon(h), \epsilon(-h)), \qquad \text{therefore} \qquad \lim_{h\to 0} \frac{\phi(h)}{|h|} = 0,$$

i.e., ϕ and, consequently, f, is differentiable at 0. □

2.39 ¶. For $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ and $v \in \mathbb{R}^n$ set

$$\frac{\partial f}{\partial v^+}(x) := \lim_{t \to 0^+} \frac{f(x+tv) - f(v)}{t}.$$

Assuming that Ω is open and convex and $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is convex, prove the following:

(i) For all $x \in \Omega$ and $v \in \mathbb{R}^n$, $\frac{\partial f}{\partial v^+}(x)$ exists.

(ii) $v \to \frac{\partial f}{\partial v^+}(x)$, $v \in \mathbb{R}^n$, is a convex and positively 1-homogeneous function.

(iii) $f(x+v) - f(x) \ge \frac{\partial f}{\partial v^+}(x)$ for all $x \in \Omega$ and all $v \in \mathbb{R}^n$.

(iv) $v \to \frac{\partial f}{\partial v^+}(x)$ is linear if and only if f is differentiable at x.

g. Extremal points of convex functions

The extremal points of convex functions have special features. In Exercise 2.34, for instance, we saw that a convex function $f : K \to \mathbb{R}$ need not have a minimum point even when K is compact; moreover, minimizers form a convex subset of K. We also saw that local minimizers are in fact global minimizers and that, assuming $f \in C^1(K)$ and x_0 interior to K, the point x_0 is a minimizer for f if and only if $Df(x_0) = 0$. When a minimizer x_0 is not necessarily an interior point, we have the following proposition.

2.40 Proposition. *Let Ω be an open set of $\mathbb{R}^n$, K a convex subset of Ω and $f : \Omega \to \mathbb{R}$ a convex function of class $C^1(\Omega)$. The following claims are equivalent:*

(i) *x_0 is a minimum point of f in K.*
(ii) $\mathbf{D}f(x_0) \bullet (x - x_0) \geq 0 \ \forall x \in K$.
(iii) $\mathbf{D}f(x) \bullet (x - x_0) \geq 0 \ \forall x \in K$.

Proof. (i) $\Leftrightarrow$ (ii). If x_0 is a minimizer in K, for all $x \in K$ and $\lambda \in]0, 1[$ we have

$$f(x_0) \leq f((1 - \lambda)x_0 + \lambda\, x),$$

hence

$$\frac{f(x_0 + \lambda\,(x - x_0)) - f(x_0)}{\lambda} \geq 0.$$

When $\lambda \to 0$, the left-hand side converges to $\mathbf{D}f(x_0) \bullet (x - x_0)$, hence (ii). Conversely, since f is convex and of class $C^1(\Omega)$ we have

$$f(x) \geq f(x_0) + \mathbf{D}f(x_0) \bullet (x - x_0) \geq f(x_0) \qquad \forall x \in K,$$

thus x_0 is a minimizer of f in K.

(ii) $\Leftrightarrow$ (iii). From Theorem 2.32 we know that $\mathbf{D}f$ is a monotone operator

$$(\mathbf{D}f(x) - \mathbf{D}f(x_0)) \bullet (x - x_0) \geq 0.$$

Thus (ii) implies (iii) trivially.

(iii) $\Leftrightarrow$ (ii). For any $x \in K$ and $\lambda \in]0, 1[$ (iii) yields

$$\mathbf{D}f(x_0 + \lambda(x - x_0)) \bullet (\lambda(x - x_0)) \geq 0,$$

hence for $\lambda > 0$

$$\mathbf{D}f(x_0 + \lambda(x - x_0)) \bullet (x - x_0) \geq 0.$$

Since $\lambda \to \mathbf{D}f(x_0 + \lambda(x - x_0)) \bullet (x - x_0)$ is continuous at 0, for $\lambda \to 0^+$ we get (ii). □

The analysis of maximum points is slightly more delicate. In the 1-dimensional case a convex function $f : [a, b] \to \mathbb{R}$ has a maximum point in a or b. However, in higher dimensions the situation is more complicated.

2.41 Example. The function

$$f(x, y) := \begin{cases} \dfrac{x^2}{y} & \text{if } y > 0, \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

is convex in $\{(x, y) \mid y > 0\} \cup \{(0, 0)\}$, as the reader can verify. Notice that f is discontinuous at $(0, 0)$ and $(0, 0)$ is a minimizer for f.

Consider the closed convex set

$$K_1 := \Big\{(x,y) \,\Big|\, x^4 \le y \le 1\Big\}.$$

We have $\sup_{\partial K_1} f(x,y) = +\infty$ since $f(x,x^4) = 1/x^2 \to \infty$ as $x \to 0$. Hence the function $f : K_1 \to \mathbb{R}$ is convex, K_1 is compact but f is unbounded on K_1.

Consider the compact and convex set

$$K_2 := \Big\{(x,y) \,\Big|\, x^2 + x^4 \le y \le 1\Big\}.$$

We have

$$f(x,y) \le \frac{x^2}{x^2+x^4} < 1 \quad \forall (x,y) \in K_2 \qquad \text{and} \qquad \sup_{(x,y)\in K_2} f(x,y) = 1.$$

Therefore, the function $f : K_2 \to \mathbb{R}$ is convex, defined on a compact convex set, bounded from above, but has no maximum point.

2.42 Proposition. *Let $K \subset \mathbb{R}^n$ be a convex and closed set that does not contain straight lines and let $f : K \to \mathbb{R}$ be a convex function.*

(i) *If f has a maximum point $\overline{x}$, then $\overline{x}$ is an extremal point of K.*
(ii) *If f is bounded from above and K is polyhedral, then f has a maximum point in K.*

Proof. The proof is by induction on the dimension. For $n = 1$, the unique closed convex subsets of $\mathbb{R}$ are the closed and bounded intervals $[a,b]$ or the closed half-lines, and in this case (i) and (ii) hold. We now proceed by induction on n.

(i) If f has a maximizer in K, then there exists $\overline{x} \in \partial K$ where f attains its maximum value. Denoting by L the supporting hyperplane of K at $\overline{x}$, then f attains its maximum in $L \cap K$ that is closed, convex and of dimension $n-1$. By the inductive assumption there exists $\widehat{x} \in L \cap K$ which is both an extremal point of $L \cap K$ and a maximizer for f. Since $\overline{x}$ needs to be also an extremal point for K, (i) holds in dimension n.

(ii) Let

$$M := \sup_{x\in K} f(x) = \sup_{x\in \partial K} f(x).$$

Since K is polyhedral, we have $\partial K = (K \cap L_1) \cup \cdots \cup (K \cap L_N)$, where $L_1, L_2, \ldots, L_N$ are the hyperplanes that define K. Hence

$$M = \sup_{x\in K\cap L_i} f(x) \qquad \text{for some } i.$$

However, $K \cap L_i$ is polyhedral and $\dim(K \cap L_i) < n$. It follows that there is $\widehat{x} \in K \cap L_i$ such that $f(\widehat{x}) = M$. □

2.3 Convex Duality

a. The polar set of a convex set

A basic construction when dealing with convexity is *convex duality*. Here we see it as the construction of the *polar set*.

Let $K \subset \mathbb{R}^n$ be an arbitrary set. The *polar* of K is defined as

$$K^* := \Big\{\xi \,\Big|\, \xi \bullet x \le 1 \ \forall x \in K\Big\}.$$

2.43 Example. (i) If $K = \{x\}$, $x \neq 0$, then its polar

$$K^* = \Big\{\xi \;\Big|\; \xi \bullet x \leq 1\Big\},$$

is the closed half-space delimited by the hyperplane $\xi \bullet x = 1$ and containing the origin. Notice that $\xi \bullet x = 1$ is one of the two hyperplanes orthogonal to x at distance $1/|x|$ from the origin.

(ii) If $K := \{0\}$, then trivially $K^* = \mathbb{R}^n$,

(iii) If $K = \overline{B(0,r)}$, then

$$K^* = \overline{B(0,1/r)}.$$

In fact, if $\xi \in B(0,1/r)$, then $\xi \bullet x \leq ||\xi||\,||x|| \leq \frac{1}{r} r = 1$, i.e., $B(0,1/r) \subset K^*$. On the other hand, $x \bullet y = ||x||\,||y||$ if and only if either $y = 0$ or x is a nonnegative multiple of y. For all $\xi \in K^*$, if $x := r\frac{\xi}{|\xi|} \in \overline{B(0,r)}$, we have $r\,||\xi|| = \xi \bullet x = ||x||\,||\xi|| \leq 1$; hence $K^* \subset \overline{B(0,1/r)}$.

Since the polar set is characterized by a family of linear inequalities, we infer the following.

2.44 Proposition. *We have the following:*

(i) *For every nonempty set K, the polar set K^* is convex, closed and contains the origin.*

(ii) *If $\{K_\alpha\}_{\alpha\in\mathcal{A}}$ is a family of nonempty sets of $\mathbb{R}^n$, then*

$$\Big(\bigcup_{\alpha\in\mathcal{A}} K_\alpha\Big)^* = \bigcap_{\alpha\in\mathcal{A}} K_\alpha^*.$$

(iii) *If $K_1 \subset K_2 \subset \mathbb{R}^n$, then $K_1^* \supset K_2^*$.*

(iv) *If $\lambda > 0$ and $K \subset \mathbb{R}^n$, then $(\lambda K)^* = \frac{1}{\lambda} K^*$.*

(v) *If $K \subset \mathbb{R}^n$, then $(\mathrm{co}(K))^* = K^*$.*

(vi) $(K \cup \{0\})^* = K^*$.

Proof. (i) By definition K^* is the intersection of a family of closed half-spaces containing 0, hence it is closed, convex and contains the origin.

(ii) From the definition

$$\Big(\bigcup_{\alpha\in\mathcal{A}} K_\alpha\Big)^* = \Big\{\xi \;\Big|\; \xi \bullet x \leq 1 \;\forall x \in \cup_{\alpha\in\mathcal{A}} K_\alpha\Big\}$$
$$= \bigcap_{\alpha\in\mathcal{A}} \Big\{\xi \;\Big|\; \xi \bullet x \leq 1 \;\forall x \in K_\alpha\Big\} = \bigcap_{\alpha\in\mathcal{A}} K_\alpha^*.$$

(iii) Writing $K_2 = K_1 \cup (K_2 \setminus K_1)$, it follows from (ii) that $K_2^* \subset K_1^* \cap (K_2 \setminus K_1)^* \subset K_1^*$.

(iv) $\xi \in (\lambda K)^*$ if and only if $\xi \bullet x \leq 1$ $\forall x \in \lambda K$, equivalently, if and only if $\xi \bullet \lambda x \leq 1$ $\forall x \in K$, i.e., $(\lambda\xi) \bullet x \leq 1$ $\forall x \in K$, that is, if and only if $\lambda\xi \in K^*$.

(v) It suffices to notice that ξ satisfies $\xi \bullet x_1 \leq 1$ and $\xi \bullet x_2 \leq 1$ if and only if $\xi \bullet x \leq 1$ for every x that is a convex combination of x_1 and x_2.

(vi) Trivial. □

2.45 Corollary. *Let $K \subset \mathbb{R}^n$. Then the following hold.*

(i) *If $0 \in \mathrm{int}(K)$, then K^* is closed, convex and compact.*

(ii) *If K is bounded, then $0 \in \operatorname{int}(K^*)$.*

Proof. If $0 \in \operatorname{int}(K)$, there is $B(0,r) \subset K$, hence, $K^* \subset B(0,r)^* = \overline{B(0,1/r)}$ and K is bounded. Similarly, if K is bounded, $K \subset B(0,M)$, then $\overline{B(0,1/M)} = B(0,M)^* \subset K^*$ and $0 \in \operatorname{int}(K^*)$. □

A compact convex set with interior points is called a *convex body*. From the above the polar set of a convex body K with $0 \in \operatorname{int}(K)$ is again a convex body with $0 \in \operatorname{int} K^*$.

The following important fact holds.

2.46 Theorem. *Let K be a closed convex set of $\mathbb{R}^n$ with $0 \in K$. Then $K^{**} = K$ where $K^{**} := (K^*)^*$.*

Proof. If $x \in K$, then $\xi \bullet x \leq 1 \ \forall \xi \in K^*$, hence $x \in K^{**}$ and $K \subset K^{**}$. Conversely, if $x_0 \notin K$, then there is a supporting hyperplane of K

$$\mathcal{P} = \left\{ x \;\middle|\; \eta \bullet x = 1 \right\}$$

that strongly separates K from x, see Theorem 2.6, and, since $0 \in K$,

$$\eta \bullet x < 1 \quad \forall x \in K \qquad \text{and} \qquad \eta \bullet x_0 > 1.$$

The first inequality states that $\eta \in K^*$, whereas the second states that $x_0 \notin K^*$. Consequently, $K^{**} \subset K$. □

Later, in Section 2.4, we shall see a few applications of polarity.

b. The Legendre transform for functions of one variable

In Paragraph a. we introduced the notion of *convex duality* for bodies. We now discuss a similar notion of duality for convex functions: the *Legendre transform*. We begin with functions of one real variable.

Let I be an interval of $\mathbb{R}$ and $f : I \to \mathbb{R}$ be a convex function. Suppose that f is of class C^2 and that $f'' > 0$ in I. Then $f' : I \to \mathbb{R}$ is strictly increasing and we may describe f in terms of the slope p by choosing for every $p \in f'(I)$ the unique $x \in I$ such that $f'(x) = p$ and defining the *Legendre transform* of f as

$$\mathcal{L}_f(p) := xp - f(x), \qquad x := x(p) = (f')^{-1}(p), \qquad p \in f'(I),$$

see Figure 2.6. In this way we have a description of f in terms of the variable p that we say is *dual* of the variable x. It is easy to prove that $\mathcal{L}_f(p)$ is of class C^2 as f and that $\mathcal{L}_f$ is strictly convex. In fact, writing $x = x(p)$ for $x = (f')^{-1}(p)$, we compute

$$(\mathcal{L}_f)'(p) = x(p) + px'(p) - f'(x(p))x'(p) = x(p), \tag{2.15}$$

$$(\mathcal{L}_f)''(p) = D(x(p)) = \frac{1}{D(f')(x(p))} = \frac{1}{f''(x(p))}. \tag{2.16}$$

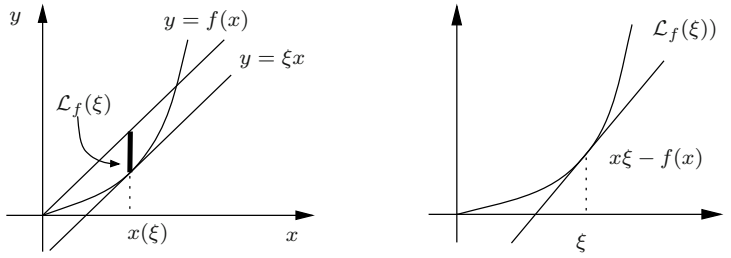


Figure 2.6. A geometric description of the Legendre transform.

c. The Legendre transform for functions of several variables

The previous construction extends to strictly convex functions of several variables giving rise to the *Legendre transform* that is relevant in several fields of mathematics and physics.

Let Ω be an open convex subset of $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}$ be a function of class C^s $s \geq 2$ with strictly positive Hessian matrix at every point $x \in \Omega$. Denote by $\mathbf{D}f : \Omega \to \mathbb{R}^n$ the *Jacobian map* of f, with $\Omega^* := \mathbf{D}f(\Omega) \subset \mathbb{R}^n$ and ξ the variable in Ω^*. The Jacobian map, or gradient map, is clearly of class C^{s-1}, and since

$$\det \mathbf{D}(\mathbf{D}f)(x) = \det \mathbf{H}f(x) > 0,$$

the implicit function theorem tells us that Ω^* is open and the gradient map is locally invertible. Actually, the gradient map is a diffeomorphism from Ω onto Ω^* of class C^{s-1}, since it is injective: In fact, if $x_1 \neq x_2 \in \Omega$ and $\gamma(t) := x_1 + tv$, $t \in [0, 1]$, $v := x_2 - x_1$, we have

$$\begin{aligned}(\mathbf{D}f(x_2) - \mathbf{D}f(x_1)) \bullet v \;&=\; \left(\int_0^1 \frac{d}{ds} (\mathbf{D}f(\gamma(s)))\, ds \right) \bullet v \\ &= \int_0^1 \mathbf{H}f(\gamma(s)) v \bullet v \; ds > 0,\end{aligned}$$

i.e., $\mathbf{D}f(x_1) \neq \mathbf{D}f(x_2)$.

Denote by $x(\xi) : \Omega^* \to \Omega$ the inverse of the gradient map

$$x(\xi) := [\mathbf{D}f]^{-1}(\xi) \quad \text{or} \quad \xi = \mathbf{D}f(x(\xi)) \qquad \forall \xi \in \Omega^*.$$

2.47 Definition. *The* Legendre transform *of f is the function $\mathcal{L}_f : \Omega^* \to \mathbb{R}$ given by*

$$\mathcal{L}_f(\xi) := \xi \bullet x(\xi) \; - f(x(\xi)), \qquad x(\xi) := (\mathbf{D}f)^{-1}(\xi). \tag{2.17}$$

2.48 Theorem. *$\mathcal{L}_f : \Omega^* \to \mathbb{R}$ is of class C^s, and the following formulas hold:*

$$\mathbf{D}\mathcal{L}_f(\xi) = x(\xi) = (\mathbf{D}f)^{-1}(\xi), \qquad \mathbf{H}\mathcal{L}_f(\xi) = \Big(\mathbf{H}f(x(\xi))\Big)^{-1}, \quad (2.18)$$

$$\mathcal{L}_f(\xi) = \xi \bullet x(\xi) - f(x(\xi)), \qquad x(\xi) := \mathbf{D}f^{-1}\xi = \mathbf{D}\mathcal{L}_f(\xi), \quad (2.19)$$

$$f(x) = \xi(x) \bullet x - \mathcal{L}_f(\xi(x)), \qquad \xi(x) = \mathbf{D}f(x). \quad (2.20)$$

In particular, if Ω^* is convex, the Legendre transform $f \to \mathcal{L}_f$ is involutive, i.e., $\mathcal{L}_{\mathcal{L}_f} = f$.

Proof. $\mathcal{L}_f$ is of class C^{s-1}, $s \geq 1$; let us prove that it is of class C^s as f. From $\xi = \mathbf{D}f(x(\xi))$ we infer

$$d\mathcal{L}_f(\xi) = x^\alpha(\xi)\, d\xi_\alpha + \xi_\alpha\, dx^\alpha - \frac{\partial f}{\partial x^\alpha}(x(\xi))\, dx^\alpha = x^\alpha(\xi)\, d\xi_\alpha,$$

i.e., $\frac{\partial \mathcal{L}_f}{\partial \xi_\alpha}(\xi) = x^\alpha(\xi)$. Since $x(\xi)$ is of class C^{s-1}, then $\mathcal{L}_f(\xi)$ is also of class C^s, and $\mathbf{D}\mathcal{L}_f(\xi) = x(\xi)$. Also from $\mathbf{D}f(x(\xi)) = \xi$ for all $\xi \in \Omega^*$ we infer $\mathbf{H}f(x(\xi))\mathbf{D}x(\xi) = \mathrm{Id}$, hence

$$\mathbf{H}\mathcal{L}_f(\xi) = \mathbf{D}x(\xi) = \Big(\mathbf{H}f(x(\xi))\Big)^{-1}.$$

In particular, the Hessian matrix of $\xi \to \mathcal{L}_f(\xi)$ is positive definite. The other claims now follow easily. □

If $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ has a positive definite Hessian matrix and Ω is convex, as previously, then f is strictly convex. However, if $n \geq 2$, the Legendre transform of f, $\mathcal{L}_f : \Omega^* \to \mathbb{R}$, need not be convex since its domain Ω^* in general may not be convex as for the Legendre transform of the function $\exp(|x|^2)$ defined on the unit cube $\Omega := \{x = (x_1, x_2, \ldots, x_n) \mid \max_i |x_i| \leq 1\}$. However, $\mathcal{L}_f$ has a strictly positive Hessian matrix, in particular, $\mathcal{L}_f$ is locally convex.

Finally, the following characterization of the Legendre transform holds.

2.49 Proposition. *Let $f \in C^s(\Omega)$, Ω be open and convex, $s \geq 2$, and $\mathbf{H}f > 0$ in Ω. Then*

$$\mathcal{L}_f(\xi) = \max\Big\{ \xi \bullet x - f(x) \,\Big|\, x \in \Omega \Big\}. \quad (2.21)$$

Proof. Fix $\xi \in \Omega^*$, and consider the concave function $g(x) := \xi \bullet x - f(x)$, $x \in \Omega$. The function $x \to \mathbf{D}g(x) := \xi - \mathbf{D}f(x)$ vanishes exactly at $\xi = \mathbf{D}f(x)$. It follows that $g(x)$ has an absolute maximum point at $x = \mathbf{D}f^{-1}(\xi)$ and the maximum value is $\mathcal{L}_f(\xi)$. □

Later we shall deal with (2.21).

2.4 Convexity at Work

2.4.1 Inequalities

a. Jensen inequality

Many inequalities find their natural context and can be conveniently treated in terms of convexity. We have already discussed in [GM1] and

Chapter 4 of [GM4] some inequalities as consequences of the convexity of suitable functions of one variable. We recall the *discrete Jensen's inequality.*

2.50 Proposition. *Let $\phi : [a,b] \to \mathbb{R}$ be a convex function, $x_1, \dots, x_m \in [a,b]$ and $\alpha_i \in [0,1]$ $\forall i = 1, \dots, m$ with $\sum_{i=1}^m \alpha_i = 1$. Then*

$$\phi\Big(\sum_{i=1}^m \alpha_i x_i\Big) \le \sum_{i=1}^m \alpha_i \phi(x_i).$$

Moreover, if ϕ is strictly convex *and $\alpha_i > 0$ $\forall i$, then $\phi\Big(\sum_{i=1}^m \alpha_i x_i\Big) = \sum_{i=1}^m \alpha_i \phi(x_i)$ if and only if $x_1 = \cdots = x_m$.*

We now list some consequences of Jensen's inequality:

(i) (Young inequality) If $p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, then

$$ab \le \frac{a^p}{p} + \frac{b^q}{q} \qquad \forall a, b \in \mathbb{R}^+$$

with equality if and only if $a^p = b^q$.

(ii) (Geometric and arithmetic means) If $x_1, x_2, \dots, x_n \ge 0$, then

$$\sqrt[n]{x_1 x_2 \dots x_n} \le \frac{1}{n} \sum_{i=1}^n x_i$$

with equality if and only if $x_1 = \cdots = x_n = \frac{1}{n}\sum_{i=1}^n x_i$.

(iii) (Hölder inequality) If $p, q > 1$ and $1/p + 1/q = 1$, then for all $x_1, x_2, \dots, x_n \ge 0$ and $y_1, y_2, \dots, y_n \ge 0$ we have

$$\sum_{i=1}^n x_i y_i \le \Big(\sum_{i=1}^n x_i^p\Big)^{1/p} \Big(\sum_{i=1}^n y_i^q\Big)^{1/q},$$

with equality if and only if either $x_i = \lambda y_i$ $\forall i$ for some $\lambda \ge 0$ or $y_1 = \cdots = y_n = 0$.

(iv) (Minkowski inequality) If $p, q > 1$ and $1/p + 1/q = 1$, then for all $x_1, x_2, \dots, x_n \ge 0$ and $y_1, y_2, \dots, y_n \ge 0$ we have

$$\Big(\sum_{i=1}^n (x_i + y_i)^p\Big)^{1/p} \le \Big(\sum_{i=1}^n x_i^p\Big)^{1/p} + \Big(\sum_{i=1}^n y_i^p\Big)^{1/p}$$

with equality if and only if either $x_i = \lambda\, y_i$ $\forall i$ for some $\lambda \ge 0$ or $y_1 = \cdots = y_n = 0$.

(v) (Entropy inequality) The function $f(p) := \sum_{i=1}^n p_i \log p_i$ defined on $K := \{p \in \mathbb{R}^n \,\big|\, p_i \ge 0, \sum_{i=1}^n p_i = 1\}$ has a unique strict minimum point at $\overline{p} = (1/n, \dots, 1/n)$.

(vi) (HADAMARD'S INEQUALITY) Since the determinant and the trace of a square matrix are respectively the product and the sum of the eigenvalues, the inequality between geometric and arithmetic means yields

$$\det \mathbf{A} \leq \Big(\frac{\operatorname{tr} \mathbf{A}}{n}\Big)^n$$

for every matrix $\mathbf{A}$ that is symmetric and with nonnegative eigenvalues. Moreover, equality holds if and only if $\mathbf{A}$ is a multiple of the identity matrix. A consequence is that for every $\mathbf{A} \in M_{n,n}(\mathbb{R})$ the following *Hadamard's inequality* holds:

$$(\det \mathbf{A})^2 \leq \prod_{i=1}^{n} |A_i|^2$$

where $A_1, A_2, \ldots, A_n$ are the columns of $\mathbf{A}$ and $|A_i|$ is the length of the column vector A_i; moreover, equality holds if and only if $\mathbf{A}$ is a multiple of an orthogonal matrix.

b. Inequalities for functions of matrices

Let $\mathbf{A} \in M_{n,n}(\mathbb{R})$ be symmetric and let $\mathbf{A}x = \sum_{i=1}^{n} \lambda_i (\, x \bullet u_i \,)\, u_i$ be its spectral decomposition. Recall that for $f : \mathbb{R} \to \mathbb{R}$, the matrix $f(\mathbf{A})$ is defined as the $n \times n$ symmetric matrix

$$f(\mathbf{A})(x) := \sum_{i=1}^{n} f(\lambda_i)(\, x \bullet u_i \,)\, u_i.$$

Notice that $\mathbf{A}$ and $f(\mathbf{A})$ have the same eigenvectors with corresponding eigenvalues λ and $f(\lambda)$, respectively.

2.51 Proposition. *Let $\mathbf{A} \in M_{n,n}(\mathbb{R})$ be symmetric and let $f : \mathbb{R} \to \mathbb{R}$ be convex. For all $x \in \mathbb{R}^n$ we have*

$$f(\, x \bullet \mathbf{A}x \,) \leq \, x \bullet f(\mathbf{A})x \,.$$

In particular, if $\{v_1, v_2, \ldots, v_n\}$ is an orthonormal basis of $\mathbb{R}^n$, we have

$$\sum_{j=1}^{n} f(\, v_j \bullet \mathbf{A}v_j \,) \leq \operatorname{tr}(f(\mathbf{A})).$$

Proof. Let $u_1, u_2, \ldots, u_n$ be an orthonormal basis of $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then

$$x \bullet \mathbf{A}x \, = \sum_{i=1}^{n} \lambda_i |\, x \bullet u_i \,|^2, \qquad x \bullet f(\mathbf{A})x \, = \sum_{i=1}^{n} f(\lambda_i) |\, x \bullet u_i \,|^2,$$

and, since $\sum_{i=1}^{n} |\, x \bullet u_i \,|^2 = |x|^2$, the discrete Jensen's inequality yields

$$f(\,x \bullet \mathbf{A}x\,) = f\Big(\sum_{i=1}^{n} \lambda_i |\, x \bullet u_i\,|^2\Big) \le \sum_{i=1}^{n} f(\lambda_i) |\, x \bullet u_i\,|^2 = \, x \bullet f(\mathbf{A}x)\,.$$

The second part of the claim then follows easily. In fact, from the first part of the claim,

$$\sum_{j=1}^{n} f\Big(\, v_j \bullet \mathbf{A}v_j\,\Big) \le \sum_{j=1}^{n} v_j \bullet f(\mathbf{A})v_j\,,$$

and, since $\{v_j\}$ is orthonormal, there exists an orthogonal matrix $\mathbf{R}$ such that $v_j = \mathbf{R}u_j$, and the spectral theorem yields

$$\sum_{j=1}^{n} v_j \bullet f(\mathbf{A})v_j \;= \sum_{j=1}^{n} u_j \bullet \mathbf{R}^T f(\mathbf{A})\mathbf{R}u_j \;= \sum_{j=1}^{n} f(\lambda_j) = \operatorname{tr} f(\mathbf{A}).$$

□

2.52 ¶. Show that

$$\frac{\Big(\prod_{i=1}^{N} x_i\Big)^{1/N} + \Big(\prod_{i=1}^{N} y_i\Big)^{1/N}}{\Big[\prod_{i=1}^{N}(x_i+y_i)\Big]^{1/N}} = \Big(\prod_{i=1}^{N} \frac{x_i}{x_i+y_i}\Big)^{1/N} + \Big(\prod_{i=1}^{N} \frac{y_i}{x_i+y_i}\Big)^{1/N}$$

$$\le \frac{1}{N}\sum_{i=1}^{N} \frac{x_i}{x_i+y_i} + \frac{1}{N}\sum_{i=1}^{N} \frac{y_i}{x_i+y_i} = 1.$$

2.53 ¶. Show that if $p, q > 1$, $1/p + 1/q = 1$, then for all $x_1, x_2, \ldots, x_n \ge 0$,

$$\Big(\sum_{i=1}^{n} x_i^p\Big)^{1/p} = \max\Big\{\sum_{i=1}^{n} x_i y_i \;\Big|\; y_i \ge 0,\ \sum_{i=1}^{n} y_i^q = 1\Big\}.$$

c. Doubly stochastic matrices

A matrix $\mathbf{A} = (a_{jk}) \in M_{n,n}(\mathbb{R})$ is said to be *doubly stochastic* if

$$a_{jk} \ge 0, \qquad \sum_{i=1}^{n} a_{ik} = 1, \qquad \sum_{i=1}^{n} a_{ji} = 1, \qquad \forall j,k = 1,\ldots,n. \tag{2.22}$$

Important examples are given by the matrix that in each row and in each column contains exactly an element equal to 1. They are characterized by a permutation σ of $\{1,\ldots,n\}$ such that $a_{jk} = 1$ if $k = \sigma(j)$ and $a_{jk} = 0$ if $k \ne \sigma(j)$; for this reason they are called *permutation matrices*. Clearly, if (a_{jk}) is a permutation matrix, then $a_{jk}x_k = x_{\sigma(j)}$.

Condition (2.22) defines the space Ω_n of doubly stochastic matrices as the intersection of closed half-spaces and affine hyperplanes of $\mathbb{R}^{n^2}$, hence as a closed convex subset of the space $M_{n,n}$ of $n \times n$ matrices.

2.54 Theorem (Birkhoff). *The set Ω_n of doubly stochastic matrices is a compact and convex subset of an affine subspace of dimension $(n-1)^2$, the extremal points of which are the permutation matrices. Consequently, every doubly stochastic matrix is the convex combination of at most $(n-1)^2+1$ permutation matrices.*

Proof. Since $a_{jk} \leq 1$, $\forall \mathbf{A} = (a_{jk}) \in \Omega_n$, the set Ω_n is bounded, hence compact being closed. Conditions (2.22) writes as $a_{ij} \geq 0$ and

$$\begin{cases} a_{nk} = 1 - \sum_{j<n} a_{jk} & k < n, \\ a_{jn} = 1 - \sum_{k<n} a_{jk} & j < n, \\ a_{nn} = 2 - n + \sum_{j,k<n} a_{jk}, \end{cases}$$

hence Ω_n is the image of the subset P defined by

$$\begin{cases} a_{jk} \geq 0 & j,k < n, \\ \sum_{j<n} a_{jk} \leq 1 & k < n, \\ \sum_{k<n} a_{jk} \leq 1 & j < n, \\ \sum_{ij} a_{jk} \geq n-2 \end{cases} \tag{2.23}$$

through an affine and *injective* map from $\mathbb{R}^{(n-1)^2}$ into $M_{n,n}$. Moreover, P has interior points as, for instance, $a_{jk} := A/(n-1)$, $1 \leq j,k < n$ with $(n-2)/(n-1) < A < 1$, hence Ω_n has dimension $(n-1)^2$.

Of course, the permutation matrices are extremal points of Ω_n. We now prove that they are the unique extremal points. We first observe that if $\mathbf{A} = (a_{jk})$ is an extremal point of Ω_n, then it has to satisfy at least $(n-1)^2$ equations of the n^2 conditions in (2.22). Otherwise we could find $\mathbf{B} := (b_{jk}) \neq 0$ such that $a_{jk} \pm \epsilon b_{jk}$, ϵ small, still satisfies (2.22); moreover, $a_{jk} = \frac{1}{2}(a_{jk} + \epsilon b_{jk}) + \frac{1}{2}(a_{jk} - \epsilon b_{jk})$ and $\mathbf{A}$ would not be an extremal point. This means that $\mathbf{A} = (a_{jk})$ has at most $n^2 - (n-1)^2 = 2n-1$ nonzero elements implying that at least one column has to have one nonzero element, hence 1, and, of course, the row corresponding to this 1 will have all other elements zero. Deleting this row and this column we still have an extremal point of Ω_{n-1}; by downward induction we then reduce to prove the claim for 2×2 matrices where it is trivially true. □

We shall now discuss an extension of Proposition 2.51.

2.55 Proposition. *Let* $\mathbf{A}$ *be an* $n \times n$ *symmetric matrix, let* $\{u_1, \ldots, u_n\}$ *be an orthonormal basis of eigenvectors of* $\mathbf{A}$ *with corresponding eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *and let* $v_1, v_2, \ldots, v_n$ *be any other orthonormal basis of* $\mathbb{R}^n$*. For* $\lambda \in \mathbb{R}^n$*, set*

$$K_\lambda := \left\{ x \in \mathbb{R}^n \;\middle|\; x = \mathbf{S}\lambda,\ \mathbf{S} \in \Omega_n \right\}.$$

Then K_λ *is convex and we have*

$$(\, v_1 \bullet \mathbf{A}v_1 \,,\, v_2 \bullet \mathbf{A}v_2 \,, \ldots,\, v_n \bullet \mathbf{A}v_n \,) \in K_\lambda.$$

Moreover, for any convex function $f : U \supset K_\lambda \to \mathbb{R}$ *the following inequality holds:*

$$f(\, \mathbf{A}v_1 \bullet v_1 \,, \ldots,\, \mathbf{A}v_n \bullet v_n \,) \leq f(\lambda_{\sigma_1}, \ldots \lambda_{\sigma_n})$$

for some permutation $\sigma \in \mathcal{P}_n$*.*

Proof. The matrix $\mathbf{S} = (s_{ij})$, $s_{ij} := |\, u_i \bullet v_j \,|^2$ is doubly stochastic. Moreover, on account of the spectral theorem, $v_j \bullet \mathbf{A}v_j = \sum_{i=1}^n \lambda_i |\, v_j \bullet u_i \,|^2$. Hence $\mathbf{A}v_j \bullet v_j = S_j \bullet \lambda$, where S_j is the jth column of the matrix $\mathbf{S}$. We then conclude that

$$(\, v_1 \bullet \mathbf{A}v_1 \,,\, v_2 \bullet \mathbf{A}v_2 \,, \ldots,\, v_n \bullet \mathbf{A}v_n \,) \in K_\lambda.$$

It is easily seen that $g(\mathbf{S}) := f(\mathbf{S}\lambda) : K_\lambda \to \mathbb{R}$ is convex. Therefore g attains its maximum value at the extremal points of K_λ, which are permutation matrices because of Birkhoff's theorem, Theorem 2.54. □

Different choices of f now lead to interesting inequalities.

(i) Choose $f(t_1, t_2, \dots, t_k) := \sum_{i=1}^k t_i$, so that both f and $-f$ are convex, and, as before, let $\mathbf{A}$ be a symmetric $n \times n$ matrix and let $\{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of $\mathbb{R}^n$. Then for $1 \le k \le n$ the following estimates for $\sum_{j=1}^k \mathbf{A}v_j \bullet v_j$ holds:

$$\sum_{j=1}^k \lambda_{n-j+1} \le \sum_{j=1}^k \mathbf{A}v_j \bullet v_j \le \sum_{j=1}^k \lambda_j, \tag{2.24}$$

$\lambda_1, \lambda_2, \dots, \lambda_n$ being the eigenvalues of $\mathbf{A}$ ordered so that $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$.

(ii) Choose $f(t) := (\prod_{i=1}^k t_i)^{1/k}$, $k \ge 1$, that is concave on $\{t \in \mathbb{R}^n \,|\, t \ge 0\}$, and let $\mathbf{A}$ be a symmetric positively semidefinite $n \times n$ matrix. Applying Proposition 2.55 to $-f$, for every orthonormal basis $\{v_1, v_2, \dots, v_n\}$ we find for every k, $1 \le k \le n$,

$$\Big(\prod_{i=1}^k \lambda_{n-i+1}\Big)^{1/k} \le \Big(\prod_{j=1}^k \mathbf{A}v_j \bullet v_j\Big)^{1/k} \tag{2.25}$$

$\lambda_1, \lambda_2, \dots, \lambda_n$ being the eigenvalues of $\mathbf{A}$ ordered so that $\lambda_1 \ge \lambda_2 \cdots \ge \lambda_n \ge 0$.

Using the inequality between the geometric and arithmetic means and (2.24) we also find

$$\Big(\prod_{j=1}^k \mathbf{A}v_j \bullet v_j\Big)^{1/k} \le \frac{1}{k}\sum_{j=1}^k \mathbf{A}v_j \bullet v_j \le \frac{1}{k}\sum_{j=1}^k \lambda_j. \tag{2.26}$$

When $k = n$ we find again

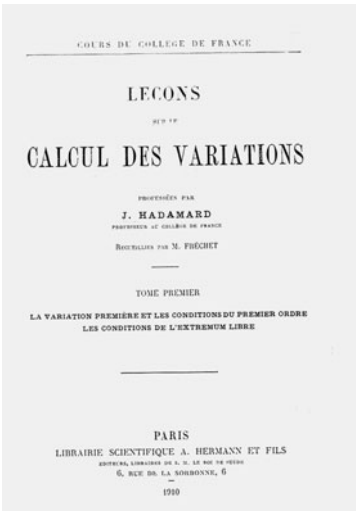
$$\det \mathbf{A} = \prod_{j=1}^n \lambda_j \le \prod_{j=1}^n \mathbf{A}v_j \bullet v_j \le \Big(\frac{\operatorname{tr}\mathbf{A}}{n}\Big)^n. \tag{2.27}$$

2.56 Theorem (Brunn–Minkowski). *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two symmetric and nonnegative matrices. Then*

$$\Big(\det(\mathbf{A}+\mathbf{B})\Big)^{1/n} \ge (\det \mathbf{A})^{1/n} + (\det \mathbf{B})^{1/n},$$
$$\det(\mathbf{A}+\mathbf{B}) \ge \det \mathbf{A} + \det \mathbf{B}.$$

Proof. Let $\{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of eigenvectors of $\mathbf{A}+\mathbf{B}$. Then

$$\begin{aligned}\Big(\det(\mathbf{A}+\mathbf{B})\Big)^{1/n} &= \Big(\prod_{i=1}^n (\mathbf{A}+\mathbf{B})v_j \bullet v_j\Big)^{1/n} \\ &\ge \Big(\prod_{j=1}^n \mathbf{A}v_j \bullet v_j\Big)^{1/n} + \Big(\prod_{j=1}^n \mathbf{B}v_j \bullet v_j\Big)^{1/n} \\ &\ge (\det \mathbf{A})^{1/n} + (\det \mathbf{B})^{1/n},\end{aligned}$$


COURS DU COLLÈGE DE FRANCE

LEÇONS

SUR LE

CALCUL DES VARIATIONS

PROFESSÉES PAR

J. HADAMARD

PROFESSEUR AU COLLÈGE DE FRANCE

RÉDIGÉES PAR M. FRÉCHET

TOME PREMIER

LA VARIATION PREMIÈRE ET LES CONDITIONS DU PREMIER ORDRE
LES CONDITIONS DE L'EXTREMUM LIBRE

PARIS
LIBRAIRIE SCIENTIFIQUE A. HERMANN ET FILS
ÉDITEURS, LIBRAIRES DE L. S. LE ROI DE SUÈDE
6, RUE DE LA SORBONNE, 6
1910


LEONIDA TONELLI

FONDAMENTI

DI

CALCOLO DELLE VARIAZIONI

VOLUME PRIMO

BOLOGNA
NICOLA ZANICHELLI
EDITORE

Figure 2.7. Frontispieces of two volumes about calculus of variations.

where we used Exercise 2.52 in the first estimate and (2.27) in the second one. The second inequality follows by taking the power n of the first. □

2.4.2 Dynamics: Action and energy

Legendre's transform has a central role in the dual description of the dynamics of mechanical systems, the *Lagrangian* and the *Hamiltonian* models.

According to the *Hamilton* or *minimal action principle*, see Chapter 3, a mechanical system is characterized by a function $L(t, x, v)$, $L : \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ called its *Lagrangian*, and its motion $t \to x(t) \in \mathbb{R}^N$ satisfies the following condition: If at times t_1 and t_2, $t_1 < t_2$, the system is at positions $x(t_1)$ and $x(t_2)$ respectively, then the motion in the interval of time $[t_1, t_2]$ happens in such a way as to make the *action*

$$\mathcal{A}(x) := \int_{t_1}^{t_2} L(t, x(t), x'(t))\, dt$$

stationary. More precisely, $x(t)$ is the actual motion from $x(t_1)$ to $x(t_2)$ if and only if for any arbitrary path $\gamma(t)$ with values in $\mathbb{R}^N$ such that $\gamma(t_1) = \gamma(t_2) = 0$, we have

$$0 = \frac{d}{d\epsilon}\mathcal{A}(x + \epsilon\gamma)\bigg|_{\epsilon=0} = \frac{d}{d\epsilon}\int_{t_1}^{t_2} L\Big(t, x(t) + \epsilon\gamma(t), x'(t) + \epsilon\gamma'(t)\Big)\, dt\bigg|_{\epsilon=0}.$$

Differentiating under the integral sign, we find

$$
\begin{aligned}
0 &= \int_{t_1}^{t_2} \sum_{i=1}^{N} \Big(L_{x^i}\gamma^i(t) + L_{v^i}\gamma^{i\prime}(t) \Big)\, dt \\
&= \int_{t_1}^{t_2} \sum_{i=1}^{N} \Big(L_{x^i} - \frac{d}{dt} L_{v^i} \Big) \gamma^i(t)\, dt + \sum_{i=1}^{N} L_{v^i}\gamma^i(t) \Big|_{t_1}^{t_2} \\
&= \int_{t_1}^{t_2} \sum_{i=1}^{N} \Big(L_{x^i} - \frac{d}{dt} L_{v^i} \Big) \gamma^i(t)\, dt
\end{aligned}
$$

for all $\gamma : [t_1, t_2] \to \mathbb{R}^N$, $\gamma(t_1) = \gamma(t_2) = 0$, where

$$
L_{x^i} := \frac{\partial L}{\partial x^i}(t, x(t), x'(t)), \qquad L_{v^i} := \frac{\partial L}{\partial v^i}(t, x(t), x'(t)).
$$

As a consequence of the fundamental lemma of the Calculus of Variations, see Lemma 1.51, the motion of the system is a solution of the *Euler–Lagrange* equations

$$
\frac{d}{dt} L_{v^i}(t, x(t), x'(t)) = L_{x^i}(t, x(t), x'(t)) \qquad \forall i = 1, \dots, N. \tag{2.28}
$$

This is an invariant way (with respect to changes of coordinates) of expressing Newton's law of dynamics. We notice that (2.28) are N ordinary differential equations of second order in the unknown $x(t)$.

There is another equivalent way of describing the law of dynamics at least when the Lagrangian L is of class C^2 and $\det \frac{\partial^2 L}{\partial v^2} > 0$, i.e., $L \in C^2(\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N)$ and $v \to L(t, x, v)$ is strictly convex for all (t, x). As we have seen, in this case the function

$$
v \;\longrightarrow\; p := L_v(t, x, v) = \frac{\partial}{\partial v} L(t, x, v)
$$

is globally invertible with inverse function $v = \psi(t, x, p)$ of class C^2 and we may form the Legendre transform of L with respect to v

$$
H(t, x, p) := p \bullet v - L(t, x, v), \qquad v := \psi(t, x, p),
$$

called the *Hamiltonian* or the *energy* of the system. For all (t, x, p) we have

$$
\begin{cases}
p = \dfrac{\partial L}{\partial v}(t, x, v), \\
L(t, x, v) + H(t, x, p) = p \bullet v\,, \\
H_t(t, x, p) + L_t(t, x, v) = 0, \\
H_x(t, x, p) + L_x(t, x, v) = 0,
\end{cases}
\qquad v = \psi(t, x, p)
$$

and, as we saw in (2.18),

$$
H_p(t, x, p) = v = \psi(t, x, p).
$$

For a curve $t \to x(t)$, if we set $v(t) = x'(t)$ and $p(t) := L_v(t, x(t), x'(t))$, we have

$$\begin{cases} v(t) = x'(t) = \psi(t, x(t), p(t)), \\ L(t, x(t), v(t)) + H(t, x(t), p(t)) = p(t) \bullet v(t), \\ H_t(t, x(t), p(t)) + L_t(t, x(t), v(t)) = 0, \\ H_x(t, x(t), p(t)) + L_x(t, x(t), v(t)) = 0. \end{cases}$$

Consequently, $t \to x(t)$ solves Euler–Lagrange equations (2.28), that can be written as

$$\begin{cases} \dfrac{dx}{dt} = v(t), \\ \dfrac{d}{dt} L_v(t, x(t), v(t)) = L_x(t, x(t), v(t)) \end{cases}$$

if and only if

$$\begin{cases} x'(t) = H_p(t, x(t), p(t)), \\ p'(t) = \dfrac{d}{dt} L_v(t, x(t), v(t)) = L_x(t, x(t), v(t)) = -H_x(t, x(t), p(t)). \end{cases}$$

Summing up, $t \to x(t)$ solves the Euler–Lagrange equations if and only if $t \to (x(t), p(t)) \in \mathbb{R}^{2N}$ solves the system of $2N$ first order differential equations, called the *canonical Hamilton system*

$$\begin{cases} x'(t) = H_p(t, x(t), p(t)), \\ p'(t) = -H_x(t, x(t), p(t)). \end{cases}$$

We emphasize the fact that, if the Hamiltonian does not depend explicitly on time (*autonomous Hamiltonians*), $H = H(x, p)$, then H is constant along the motion,

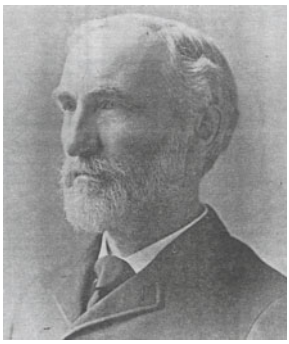
$$\frac{d}{dt} H(x(t), p(t)) = \frac{\partial H}{\partial x} \bullet x' + \frac{\partial H}{\partial p} \bullet p' = p' \bullet x' - x' \bullet p' = 0.$$

We shall return to the Lagrangian and Hamiltonian models of mechanics in Chapter 3.

2.4.3 The thermodynamic equilibrium

Here we briefly hint at the use of convexity in the discussion of the thermodynamic equilibrium by J. Willard Gibbs (1839–1903).

For the sake of simplicity we consider a quantity of N moles of a *simple fluid*, i.e., of a fluid in which equilibrium points may be described in terms of the following six thermodynamic variables:



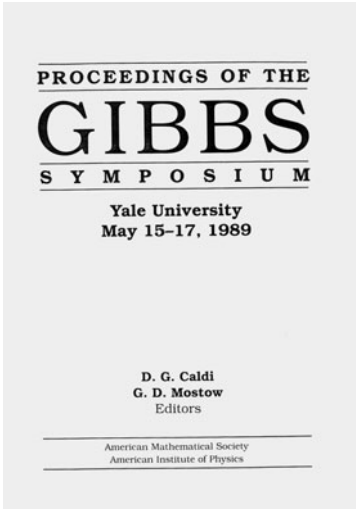


Figure 2.8. J. Willard Gibbs (1839–1903) and the frontispiece of Gibbs Symposium at Yale.

(i) V, the volume,
(ii) p, the pressure,
(iii) T, the absolute temperature,
(iv) U, the internal energy,
(v) S, the entropy,
(vi) μ, the chemical potential,
(vii) N, the number of moles.

For simple fluids, Gibbs provided a description of the thermodynamic equilibrium which is compatible with the thermodynamic laws established a few years earlier by Rudolf Clausius (1822–1888). In modern terms and freeing our presentation from experimental discussions, Gibbs assumed the following:

(i) The balance law, called the *fundamental equation*,

$$T\,dS = dU + p\,dV + \mu\,dN \tag{2.29}$$

in the variable domains $T > 0$, $V > 0$, $U > 0$, $p > 0$, $N > 0$, $\mu \in \mathbb{R}$ and $S \in \mathbb{R}$.

(ii) The equilibrium configurations can be parametrized either by the independent variables S, V and N or by the independent variables U, V and N, and, at equilibrium, the other thermodynamic quantities are functions of the chosen independent variables.

(iii) The entropy function $S = S(U, V, N)$ is of class C^1 and positively homogeneous of degree 1,

$$S(\lambda U, \lambda V, \lambda N) = \lambda S(U, V, N), \qquad \forall \lambda > 0.$$

(iv) The entropy function $S = S(U, V, N)$ is concave.
(v) The free energy function $U = U(S, V, N)$ is of class C^1, convex and positively homogeneous of degree 1.

A few comments on (i), (ii), ..., (v) are appropriate:

(i) The fundamental equation (2.29) contains the *first principle of thermodynamics*: *the elementary mechanic work done on a system plus the differential of the heat furnished to the system plus the variation of moles is an exact differential* $p\,dV - T\,dS + \mu\,dN = -dU$.
(ii) The homogeneity of S amounts, via (2.29), to the invariance at equilibrium of temperature, pressure and chemical potential when moles change.
(iii) The assumption of C^1-regularity of the entropy function, in addition to being useful, is essential in order to deduce the Gibbs necessary condition for the existence of coexisting phases.
(iv) If we choose as independent variables the internal energy U, the volume V and the number of moles N, then S, T and V are functions of (U, V, N). The fundamental equation then allows us to compute the absolute temperature and the chemical potential as partial derivatives of the entropy function $S = S(U, V, N)$, that thus describes the whole system, finding[2]

$$\frac{1}{T} = \Big(\frac{\partial S}{\partial U}\Big)_{V,N}, \qquad \frac{p}{T} = \Big(\frac{\partial S}{\partial V}\Big)_{U,N}, \qquad \frac{\mu}{T} = \Big(\frac{\partial S}{\partial N}\Big)_{U,V}. \quad (2.30)$$

(v) The function $U \to S(U, V, N)$ is strictly increasing. Therefore, we can replace the independent variables (U, V, N) with the variables (S, V, N) and obtain an equivalent description of the equilibrium of the fluid in terms of the internal energy function $U = U(S, V, N)$, concluding that

$$T = \Big(\frac{\partial U}{\partial S}\Big)_{V,N}, \qquad -p = \Big(\frac{\partial U}{\partial V}\Big)_{S,N}, \qquad \mu = \Big(\frac{\partial U}{\partial N}\Big)_{S,V}.$$

(vi) The concavity of the entropy function is a way to formulate the second principle of thermodynamics. Consider, in fact, two quantities of the same fluid with parameters at the equilibrium $x_1 := (U_1, V_1, N_1)$ and $x_2 := (U_2, V_2, N_2)$, and a quantity of $N_1 + N_2$ moles of the same fluid with volume $V_1 + V_2$ and internal energy $U_1 + U_2$. The second principle of thermodynamics states that the entropy has to increase

$$S(x_1 + x_2) \geq S(x_1) + S(x_2).$$

Because of the arbitrariness of x_1 and x_2 and the homogeneity of S, we may infer

[2] Here we use the symbolism of physicists. For instance, by $\left(\frac{\partial S}{\partial U}\right)_{V,N}$ we mean that the function S is seen as a function of the independent variables (U, V, N) and that it is differentiated with respect to U and, consequently, the resulting function is a function of (U, V, N).

$$S((1-\alpha)x_1+\alpha x_2) \geq (1-\alpha)S(x_1)+\alpha S(x_2) \qquad \forall x_1, x_2,\ \forall \alpha \in [0,1],$$

i.e., $S(x) = S(U,V,N)$ *is a concave function.*

(vii) Similar arguments may justify the homogeneity and convexity of the internal energy function.

Gibbs's conclusion is that a simple fluid is described by a 3-dimensional surface which is at the same time the graph of $S(x)$, $x = (U,V,N) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$ (concave, positively homogeneous of degree one and of class C^1) and the graph of the function $U(y)$, $y = (S,V,N) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$, convex, positively homogeneous of degree one and of class C^1.

Since S is positively homogeneous, it is determined by its values when restricted to a "section", i.e., by its values when the energy, the volume or the number of moles is prescribed. For instance, assuming $N = 1$ and denoting by (u,v) the internal energy and the volume per mole, the entropy function per mole

$$s(u,v) := S(u,v,1),$$

describes the equilibrium of a mole of the matter under scrutiny and from (2.30)

$$\frac{1}{T(u,v)} = \Big(\frac{\partial s}{\partial u}\Big)_v, \qquad \frac{p(u,v)}{T(u,v)} = \Big(\frac{\partial s}{\partial v}\Big)_u. \tag{2.31}$$

Clearly, $s(u,v)$ is concave and the entropy S for N moles by homogeneity is given by

$$S(U,V,N) = NS\Big(\frac{U}{N}, \frac{V}{N}, 1\Big) = N\, s\Big(\frac{U}{N}, \frac{V}{N}\Big).$$

In particular, differentiating we get

$$\begin{aligned}
\frac{1}{T(U,V,N)} &= \frac{\partial s}{\partial u}\Big(\frac{U}{N}, \frac{V}{N}\Big),\\
p(U,V,N) &= \frac{\partial s}{\partial v}\Big(\frac{U}{N}, \frac{V}{N}\Big),\\
\mu(U,V,N) &= s\Big(\frac{U}{N}, \frac{V}{N}\Big) - \frac{1}{T}\,\frac{U}{N} - p\,\frac{V}{N},
\end{aligned}$$

and (2.29) transforms into

$$T\, ds = du + p\, dv.$$

a. Pure and mixed phases

Gibbs also provided a description of the coexistence of different phases in terms of an analysis of the graph of a convex function. Let $s(x)$, $x \in \mathbb{R}_+ \times \mathbb{R}_+$, be a convex function in the variables $x := (u,v)$. We say that the phase x is *pure* for a liquid if $(x, s(x))$ is an extreme point of the epigraph of f. The other points are called points of *coexistent phases*: These are points x for which $(x, f(x))$ is a convex combination of the extreme points

$(x_i, f(x_i))$ of the epigraph $\mathrm{Epi}(f)$ of f. Since $\mathrm{Epi}(f)$ has dimension 3, Corollary 2.27 tells us that the boundary of $\mathrm{Epi}(f)$ splits into three sets

$$\Sigma_0 := \Big\{\text{extreme points of } \mathrm{Epi}(f)\Big\},$$
$$\Sigma_1 := \Big\{\text{linear combinations of two points in } \Sigma_0\Big\},$$
$$\Sigma_2 := \Big\{\text{linear combinations of three points of } \Sigma_0\Big\}$$

corresponding to equilibrium with pure phases, with two mixed phases and three mixed phases, respectively.

A typical situation is the one in which the pure phases are of three different types, as for water: solid, liquid and gaseous states. Then Σ_1 corresponds to the situation in which two states of the liquid coexist, and Σ_3 corresponds to states in which the three states are present at the same time.

2.57 Proposition. *Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be a convex function of class C^1 and let $x_1, x_2, \ldots, x_k$ be k points in Ω. A necessary and sufficient condition for the existence of $x \in \Omega$, $x \neq x_i$ $\forall i$, such that*

$$(x, f(x)) = \sum_{i=1}^{k} \alpha_i (x_i, f(x_i)) \qquad \text{with} \qquad \sum_{i=1}^{k} \alpha_i = 1,\ \alpha_i \in [0,1] \quad (2.32)$$

is that the supporting hyperplanes to f at the points $x_1, x_2, \ldots, x_k$ are the same plane. In particular, $Df(x)$ is then constant in the convex envelope of $x_1, x_2, \ldots, x_k$.

Proof. Let $M := \mathrm{co}(\{x_1, x_2, \ldots, x_k\})$. The convexity of $f(x)$ implies that f is linear affine in M,

$$(x, f(x)) = \sum_{i=1}^{k} \alpha_i (x_i, f(x_i)), \qquad \sum_{i=1}^{k} \alpha_i = 1,\ \alpha_i \in]0,1[,$$

for all $x \in M$ if and only if (2.32) holds. In this case the segment joining any two points $a, b \in M$ is contained in the support hyperplanes of f at a and at b. On the other hand, a support hyperplane to f at b that contains the segment joining $(a, f(a))$ with $(b, f(b))$ is also a supporting hyperplane to f at a. Since f is of class C^1, f has a unique support hyperplane at a, $z = \nabla f(a)(x-a) + f(a)$, hence the support hyperplanes to f at a and b must coincide, and $\nabla f(x)$ is constant in M. □

In the context of thermodynamics of simple fluids, the previous proposition when applied to the entropy function, see (2.31), yields the following statement.

2.58 Proposition (Gibbs). *In a simple fluid with entropy function of class C^1 two or three phases may coexists at the equilibrium only if they are at the same temperature and the same pressure.*

In principle, we may describe the thermodynamic equilibrium in terms of entropy function in the dual variables of the energy and volume, i.e., in terms of the absolute temperature and pressure. However, first we need to write $s = s(T,p)$ and $V = V(T,p)$. The Legendre duality formula turns out to be useful. In fact, starting from the internal energy $U := U(S,V,N)$ that can be obtained inverting the entropy function $S = S(U,V,N)$, we consider the internal energy per mole, $u(s,v) := U(s,v,1)$, for which we have

$$du = T\,ds - p\,dv.$$

The dual variables of (u,v) are then $(T,-p)$: the absolute temperature T and minus the pressure p. At this point, we introduce *Gibbs's energy* as

$$G(T,p) := \sup_{s,v}\Big\{u(s,v) + pv - Ts\Big\}$$

and observe that $G(-T,p)$ is the Legendre transform of the concave function $-u$,

$$G(T,p) = \mathcal{L}_u(T,-p).$$

Therefore, at least in the case where u is strictly convex, we infer

$$s = -\Big(\frac{\partial G}{\partial T}\Big)_p, \qquad v = \Big(\frac{\partial G}{\partial p}\Big)_T.$$

2.4.4 Polyhedral sets

a. Regular polyhedra

We recall that a set K is said to be *polyhedral* if it is the intersection of finitely many closed half-spaces. A bounded polyhedral set is called a *polyhedron*.

Consider a convex polygon K containing the origin with vertices $A_1, A_2, \dots, A_N$. The vertices are the extreme points of $K \subset \mathbb{R}^n$ and $K = \mathrm{co}(\{A_1, A_2, \dots, A_N\})$, hence, compare Proposition 2.44,

$$K^* = \{A_1, A_2, \dots, A_N\}^* = \bigcap_{i=1}^{N}\{A_i\}^*$$

and, compare Theorem 2.46, $K = (K^*)^*$. Accordingly, K^* is a polyhedron, the intersection of the N half-spaces containing the origin and delimited by the hyperplanes $\{\xi \mid \xi \bullet A_i = 1\}$ in $\mathbb{R}^n$, see Figure 2.9.

2.59 ¶. The reader is invited to compute the polar sets of various convex sets of the plane.

The construction works in the same way in all $\mathbb{R}^n$'s, $n \geq 2$. Though difficult to visualize, and cumbersome to check, in $\mathbb{R}^3$, the polar set of a regular tetrahedron centered at the origin is a regular tetrahedron centered at the origin, the polar set of a cube centered at the origin is an octahedron centered at the origin, and the polar set of a dodecahedron centered at the origin is an icosahedron centered at the origin.

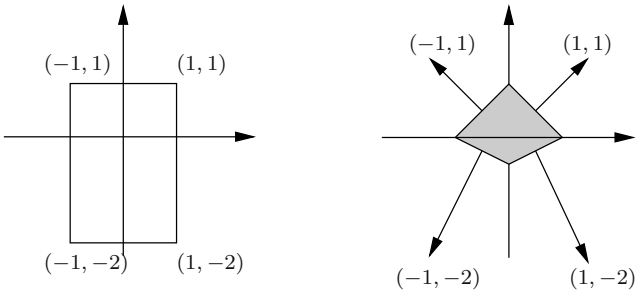


Figure 2.9. The polar set of a rectangle that contains the origin.

b. Implicit convex cones

Polyhedral sets that are cones play an important role. Let us start with cones defined implicitly by a matrix $\mathbf{A} \in M_{n,N}(\mathbb{R})$ and a vector $b \in \mathbb{R}^n$ as

$$K := \left\{ x \in \mathbb{R}^N \;\middle|\; x \geq 0,\ \mathbf{A}x = b \right\} \tag{2.33}$$

where if $x = (x^1, x^2, \ldots, x^N)$, $x \geq 0$ stands for $x^i \geq 0\ \forall i = 1, \ldots, N$. In this case, K is a convex polyhedral closed set of $\mathbb{R}^n$ that does not contain straight lines, hence, see Theorem 2.23, K does have extreme points. They are characterized as follows.

2.60 Definition. *Let K be as in* (2.33). *We say that $x \in K$ is a* base point *of K if either $x = 0$ (in this case $0 \in K$) or, if $\alpha_1, \alpha_2, \ldots, \alpha_k$ are the indices of the nonzero components of x, the columns $A_{\alpha_1}, \ldots A_{\alpha_k}$ of $\mathbf{A}$ are linearly independent.*

2.61 Theorem. *Let K be as in* (2.33). *Extreme points of K are all and only the base points of K.*

Proof. Clearly, if $0 \in K$, then 0 is an extreme point of K. Suppose that $x = (x^1 \ldots, x^k, 0, \ldots, 0) \in K$, $x^i > 0\ \forall i = 1, \ldots, k$, is a base point for K, and, contrary to the claim, x is not an extreme point for K. Then there are $y, z \in K$, $y \neq z$, such that $x = (y + z)/2$. Since $x, y, z \in K$, it would follow that $y = (y^1, y^2, \ldots, y^k, 0, \ldots, 0)$, $z = (z^1, z^2, \ldots, z^k, 0, \ldots, 0)$ and $b = \sum_{i=1}^k y^i A_i = \sum_{i=1}^k z^i A_i$. Since $A_1, A_2, \ldots, A_k$ are linearly independent, we would then have $y = z$, a contradiction.

Conversely, suppose that x is a nonzero extreme point of K and that $x = (x^1, x^2, \ldots, x^k, 0, \ldots, 0)$ with $x^i > 0\ \forall i = 1, \ldots, k$. Then

$$x^1 A_1 + \cdots + x^k A_k = b.$$

We now infer that $A_1, A_2, \ldots, A_k$ are linearly independent. Suppose they are not independent, i.e., there is a nonzero $y = (y^1, y^2, \ldots, y^k, 0, \ldots, 0)$ such that

$$y^1 A_1 + \cdots + y^k A_k = 0.$$

Now we choose $\theta > 0$ in such a way that $u := x + \theta\, y$ and $v := x - \theta\, y$ still have nonnegative coordinates and $u, v \in K$. Then $x = (u + v)/2$, $u \neq v$, and x would not be an extreme point. □

2.62 Remark. Actually, Theorem 2.61 provides us with an algorithm for computing the extreme points of a polyhedral convex set as base points. Since base points correspond to a choice of linearly independent columns, Theorem 2.61 shows that K has finitely many extreme points.

The next proposition shows the existence of a base point without any reference to the convex set theory. We include it for the reader's convenience.

2.63 Proposition. *Let $K \neq \emptyset$ be as in* (2.33)*. Then K has at least one base point.*

Proof. Of course, there is a point x with minimum, say k, nonzero components such that $\mathbf{A}x = b$ and no $x' \geq 0$ with $\mathbf{A}x' = b$ and number of components nonzero $< k$.

Let $\alpha_1, \ldots, \alpha_k$ be the indices of nonzero components of x. We now prove that *the columns $A_{\alpha_1}, \ldots, A_{\alpha_k}$ are linearly independent*, i.e., that x is a base point of K. Suppose they are not independent, i.e.,

$$\sum_{i=1}^{k} \theta_i A_{\alpha_i} = 0$$

where $\theta_1, \theta_2, \ldots, \theta_k$ are not all zero. We may assume that at least one of the θ_i is positive. Then

$$b = \sum_{i=1}^{k} A_{\alpha_i} x_i = \sum_{i=1}^{k} A_{\alpha_i} (x_i - \lambda\theta_i)$$

for all $\lambda \in \mathbb{R}$. However, for

$$\lambda := \min\Big\{ \frac{x_i}{\theta_i} \,\Big|\, \theta_i > 0 \Big\} =: \frac{x_{i_0}}{\theta_{i_0}}$$

we have $x_{i_0} - \lambda\theta_{i_0} = 0$. It follows that $x' := x - \lambda\theta \geq 0$, $b = \mathbf{A}'x'$ and x' has a number of nonzero components less than k, a contradiction. □

c. Parametrized convex cones

Particularly useful are the *finite cones*, i.e., cones generated by finitely many points, $A_1, A_2, \ldots, A_N \in \mathbb{R}^n$. They have the form

$$C := \Big\{ \sum_{i=1}^{N} x^i A_i \,\Big|\, x^i \geq 0,\ i = 1, \ldots, N \Big\}$$

and with the notation rows by columns, they can be written in a compact form as

$$C := \{ y \in \mathbb{R}^n \,\big|\, y = \mathbf{A}x,\ x \geq 0 \}$$

where $\mathbf{A} \in M_{n,N}$ is the $n \times N$ matrix

$$\mathbf{A} = [A_1 \,|\, A_2 \,|\, \ldots \,|\, A_N].$$

Trivially, a finite cone is a polyhedral set that does not contain straight lines, hence has extreme points. We say that a finite cone is a *base cone* if it is generated by linearly independent vectors.

2.64 Proposition. *Every finite cone C is convex, closed and contains the origin.*

Proof. Trivially, C is convex and contains the origin. so it remains to prove that C is closed. Let $\mathbf{A} \in M_{n,N}$ be such that $C = \Big\{y = \mathbf{A}x \,|\, x \geq 0\Big\}$. C is surely closed if $\mathbf{A}$ has linearly independent columns, i.e., if $\mathbf{A}$ is injective. In fact, in this case the map $x \to \mathbf{A}x$ has a linear inverse, hence it is a closed map and $C = \mathbf{A}(\{x \geq 0\})$. For the general case, consider the cones $C_1, \ldots, C_k$ associated to the submatrices of $\mathbf{A}$ that have linearly independent columns. As we have already remarked $C_1, \ldots, C_k$ are closed sets. We claim that

$$C = C_1 \cup C_2 \cup \cdots \cup C_k, \tag{2.34}$$

hence C is closed, too. In order to prove (2.34), observe that since every cone generated by a submatrix of $\mathbf{A}$ is contained in C, we have $C_i \subset C \ \forall i$. On the other hand, if $b \in C$, Proposition 2.63 yields a submatrix $\mathbf{A}'$ of $\mathbf{A}$ with linearly independent columns such that $b = \mathbf{A}'x'$ for some $x' \geq 0$, i.e., $b \in \cup_i C_i$. □

The following claims readily follow from the results of Paragraph a.

2.65 Corollary. *Let C_1 and C_2 be two finite cones in $\mathbb{R}^n$. Then*

(i) *if $C_1 \subset C_2$, then $C_2^* \subset C_1^*$,*
(ii) $C_1^* \cup C_2^* = (C_1 \cap C_2)^*$,
(iii) $C_1 = C_1^{**}$.

Finally, let us compute the polar set of a finite cone.

2.66 Proposition. *Let $C = \{\mathbf{A}x \,|\, x \geq 0\}$, $\mathbf{A} \in M_{n,N}(\mathbb{R})$. Then*

$$C^* = \Big\{\xi \,\Big|\, \mathbf{A}^T\xi \leq 0\Big\} \tag{2.35}$$

and

$$C^{**} := \Big\{x \,\Big|\, x \bullet \xi \ \leq 0 \ \forall \xi \text{ such that } \mathbf{A}^T\xi \leq 0\Big\}. \tag{2.36}$$

Proof. Since C is a cone, we have

$$C^* = \Big\{\xi \,\Big|\, \xi \bullet b \ \leq 1 \ \forall b \in C\Big\} = \Big\{\xi \,\Big|\, \xi \bullet b \ \leq 0 \ \forall b \in C\Big\}.$$

Consequently,

$$C^* = \Big\{\xi \,\Big|\, \xi \bullet \mathbf{A}x \ \leq 0 \ \forall x \geq 0\Big\} = \Big\{\xi \,\Big|\, \mathbf{A}^T\xi \bullet x \ \leq 0 \ \forall x \geq 0\Big\} = \Big\{\xi \,\Big|\, \mathbf{A}^T\xi \leq 0\Big\}$$

and

$$\begin{aligned} C^{**} &= \Big\{x \mid x \bullet \xi \ \leq 1 \ \forall \xi \in C^*\Big\} = \Big\{x \mid x \bullet \xi \ \leq 0 \ \forall \xi \in C^*\Big\} \\ &= \Big\{x \mid x \bullet \xi \ \leq 0 \ \forall \xi \text{ such that } \mathbf{A}^T\xi \leq 0\Big\}. \end{aligned}$$

□

d. Farkas–Minkowski's lemma

2.67 Theorem (Farkas–Minkowski). *Let $\mathbf{A} \in M_{n,N}(\mathbb{R})$ and $b \in \mathbb{R}^n$. One and only one of the following claims holds:*

(i) $\mathbf{A}x = b$ *has a nonnegative solution.*
(ii) *There exists a vector $y \in \mathbb{R}^n$ such that $\mathbf{A}^T y \geq 0$ and $y \bullet b < 0$.*

In other words, using the same notations as in Theorem 2.67, the claims

(i) x is a nonnegative solution of $\mathbf{A}x = b$,
(ii) if $\mathbf{A}^T y \leq 0$, then $y \bullet b \leq 0$

are equivalent.

Proof. The claim is a rewriting of the equality $C = C^{**}$ in the case of finite cones, and, ultimately, a direct consequence of the separation property of convex sets. Let $C := \{\mathbf{A}x \,|\, x > 0\}$. Claim (i) rewrites as $b \in C$, while, according to (2.36), claim (ii) rewrites as $b \notin C^{**}$. □

2.68 Example (Fredholm alternative theorem). The Farkas–Minkowski lemma, equivalently the equality $C = C^{**}$ for finite cones, can be also seen as a generalization of the Fredholm alternative theorem for linear maps: $\mathrm{Im}(\mathbf{A}) = (\ker \mathbf{A}^T)^\perp$. In fact, if $b = \mathbf{A}x$, $\mathbf{A} \in M_{n,N}$, and if we write $x = u - v$ with $u, v \geq 0$, the equation $\mathbf{A}x = b$ rewrites as

$$b = \begin{pmatrix} \boxed{\mathbf{A}} & \boxed{-\mathbf{A}} \end{pmatrix} \begin{pmatrix} u \\ \\ v \end{pmatrix}, \qquad u, v \geq 0.$$

Therefore, $b \in \mathrm{Im}\,\mathbf{A}$ if and only if the previous system has a nonnegative solution. This is equivalent to saying that the alternative provided by the Farkas lemma is not true; consequently,

$$\text{if } \begin{pmatrix} \boxed{\mathbf{A}^T} \\ \\ \boxed{-\mathbf{A}^T} \end{pmatrix} \xi \leq 0, \text{ then } b \bullet \xi \leq 0$$

i.e.,

$$b \bullet \xi \leq 0 \text{ for all } \xi \text{ such that } \mathbf{A}^T \xi = 0$$

and, in conclusion,

$$b \bullet \xi = 0 \text{ for all } \xi \text{ such that } \mathbf{A}^T \xi = 0,$$

i.e., $b \in (\ker \mathbf{A}^T)^\perp$.

2.69 ¶. Let $\mathbf{A} \in M_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$ and let K be the closed convex set

$$K := \{x \in \mathbb{R}^n \,\big|\, \mathbf{A}x \geq b,\ x \geq 0\}.$$

Characterize the extreme points of K.
[*Hint.* Introduce the new variables, called *slack variables* $x' \geq 0$, so that the constraints $\mathbf{A}x \geq b$ become

$$\mathbf{A}' \begin{pmatrix} x \\ x' \end{pmatrix} = b, \qquad \mathbf{A}' := \begin{pmatrix} \boxed{\mathbf{A}} & \boxed{-\mathrm{Id}} \end{pmatrix}.$$

Set $K' := \{z \,\big|\, \mathbf{A}'z \geq b,\ z \geq 0\}$. Show that x *is an extreme point for* K *if and only if* $z := (x, x')$ *with* $x' := \mathbf{A}x - b$ *is an extreme point for* K'.]

Figure 2.10. Gaspard Monge (1746–1818) and the frontispiece of the *Principes de la théorie des richesses* di Antoine Cournot (1801–1877).

2.70 ¶. Prove the following variants of the Farkas lemma.

Theorem. *Let* $\mathbf{A} \in M_{n,N}(\mathbb{R})$ *and* $b \in \mathbb{R}^n$. *One and only one of the following alternatives holds:*
- $\mathbf{A}x \geq b$ *has a solution* $x \geq 0$.
- *There exists* $y \leq 0$ *such that* $\mathbf{A}^T y \geq 0$ *and* $b \bullet y < 0$.

Theorem. *Let* $\mathbf{A} \in M_{n,N}(\mathbb{R})$ *and* $b \in \mathbb{R}^n$. *One and only one of the following alternatives holds:*
- $\mathbf{A}x \leq b$ *has a solution* $x \geq 0$.
- *There exists* $y \geq 0$ *such that* $\mathbf{A}^T y \geq 0$ *and* $b \bullet y < 0$.

[*Hint.* Introduce the slack variables, as in Example 2.68.]

2.4.5 Convex optimization

Let f and $\varphi^1, \varphi^2, \ldots, \varphi^m : \mathbb{R}^n \to \mathbb{R}$ be functions of class C^1. Here we discuss the constrained *minimum problem*

$$f(x) \to \min \qquad \text{in} \qquad \mathcal{F} := \left\{ x \in \mathbb{R}^n \,\middle|\, \varphi^j(x) \leq 0,\ j = 1, \ldots, m \right\} \quad (2.37)$$

and, in particular, we present necessary and sufficient conditions for its solvability, compare also Section 4.

Let $\varphi := (\varphi^1, \ldots, \varphi^m) : \mathbb{R}^n \to \mathbb{R}^m$ and let x_0 be a minimum point for f in $\mathcal{F}$. If $\varphi^j(x_0) < 0\ \forall j$, $\varphi(x_0) < 0$ for short, then x_0 is interior to $\mathcal{F}$ and Fermat's theorem implies $\mathbf{D}f(x_0) = 0$. If $\varphi(x_0) = 0$, then x_0 is a

minimum point constrained to $\partial\mathcal{F} := \{x \in \mathbb{R}^n \,|\, \varphi(x) = 0\}$. Consequently, if the Jacobian matrix $\mathbf{D}\varphi(x_0)$ has maximal rank so that $\partial\mathcal{F}$ is a regular submanifold in a neighborhood of x_0, we have

$$\mathbf{D}f(x_0)(v) = 0 \qquad \forall v \in \operatorname{Tan}_{x_0} \partial\mathcal{F},$$

i.e.,

$$\nabla f(x_0) \perp \operatorname{Tan}_{x_0} \partial\mathcal{F},$$

and, from Lagrange's multiplier theorem (or Fredholm's alternative theorem) we infer the existence of a vector $\lambda^0 = (\lambda_1^0, \ldots, \lambda_m^0) \in \mathbb{R}^m$ such that

$$\mathbf{D}f(x^0) = \sum_{j=1}^{m} \lambda_j^0 \mathbf{D}\varphi^j(x^0).$$

In general, it may happen that $\varphi^j(x^0) = 0$ for some j and $\varphi^j(x^0) < 0$ for the others. For $x \in \mathcal{F}$, denote by $J(x)$ the set of indices j such that $\varphi^j(x) = 0$. We say that the constraint φ^j is *active* at x if $j \in J(x)$.

2.71 Definition. *We say that a vector $h \in \mathbb{R}^n$ is an* admissible direction *for $\mathcal{F}$ at $x \in \mathcal{F}$ if there exists a sequence $\{x^k\} \subset \mathcal{F}$ such that*

$$x_k \neq x \ \forall k, \ x_k \to x \ \text{as} \ k \to \infty \qquad \text{and} \qquad \frac{x_k - x}{|x_k - x|} \to \frac{h}{|h|}.$$

The set of the admissible directions for $\mathcal{F}$ at x is denoted by $\Gamma(x)$. It is easily seen that $\Gamma(x)$ is a closed cone not necessarily convex. Additionally, it is easy to see that $\Gamma(x)$ is the set of directions $h \in \mathbb{R}^n$ for which there is a regular curve $r(t)$ in $\mathcal{F}$ with $r(0) = x$ and $r'(0) = h$.

Denote by $\widetilde{\Gamma}(x)$ the cone with vertex at zero, this time convex, of the directions that "point to $\mathcal{F}$",

$$\widetilde{\Gamma}(x) := \left\{ h \in \mathbb{R}^n \;\middle|\; \nabla\varphi^j(x) \bullet h \leq 0 \ \forall j \in J(x) \right\};$$

it is not difficult to prove that $\Gamma(x) \subset \widetilde{\Gamma}(x)$.

2.72 Definition. *We say that the constraints are* qualified *at $x \in \mathcal{F}$ if $\Gamma(x) = \widetilde{\Gamma}(x)$.*

Not always are the constraints qualified, see Example 2.76. The following proposition gives a sufficient condition which ensures that the constraints are qualified.

2.73 Proposition. *Let $\varphi = (\varphi^1, \varphi^2, \ldots, \varphi^m) : \mathbb{R}^n \to \mathbb{R}^m$ be of class C^1, $\mathcal{F} := \{x \in \mathbb{R}^n \,|\, \varphi(x) \leq 0\}$ and $x_0 \in \mathcal{F}$. If there exists $\overline{h} \in \mathbb{R}^n$ such that for all $j \in J(x_0)$ we have*

(i) *either* $\nabla\varphi^j(x_0) \bullet \overline{h} < 0$

(ii) *or* φ^j *is affine and* $\nabla\varphi^j(x_0) \bullet \overline{h} \le 0$,

then the constraints $\{\varphi^j\}$ *are qualified at* x_0. *Consequently, the constraints are qualified at* x_0 *if one of the following conditions holds:*

(i) *There exists* $\overline{x} \in \mathcal{F}$ *such that* $\forall j \in J(x_0)$, *either* φ^j *is convex and* $\varphi^j(\overline{x}) < 0$, *or* φ^j *is affine and* $\varphi^j(\overline{x}) \le 0$.
(ii) *The vectors* $\nabla\varphi^j(x_0)$, $j \in J(x_0)$, *are linearly independent.*

Proof. Step 1. Let us prove that $\widetilde{\Gamma}(x_0) \subset \Gamma(x_0)$. Let h be such that $\nabla\varphi^j(x_0) \bullet h \le 0$. We claim that for every $\delta > 0$ we have $h + \delta\overline{h} \in \Gamma(x_0)$, thus concluding that $h \in \Gamma(x_0)$, $\Gamma(x_0)$ being closed.

Choose a positive sequence $\{\epsilon_k\}$ such that $\epsilon_k \to 0$ and consider the sequence $\{x_k\}$ defined by $x_k := x_0 + \epsilon_k(h + \delta\overline{h})$. Trivially $x_k \to x_0$ and $\frac{x_k - x_0}{|x_k - x_0|} = \frac{h + \delta\overline{h}}{|h + \delta\overline{h}|}$, thus $h + \delta\overline{h} \in \Gamma(x_0)$ if we prove that $x_k \in \mathcal{F}$ for k large. Let $j \in J(x_0)$. If $\nabla\varphi^j(x_0) \bullet \overline{h} < 0$, then

$$\nabla\varphi^j(x_0) \bullet (h + \delta\overline{h}) < 0$$

and, since

$$\varphi^j(x_k) = \varphi^j(x_0) + \epsilon_k \nabla\varphi^j(x_0) \bullet (h + \delta\overline{h}) + o(\epsilon_k),$$

we conclude that $\varphi^j(x_k) < 0$ for k large. If φ^j is affine and $\nabla\varphi^j(x_0) \bullet \overline{h} \le 0$, then

$$\varphi^j(x_k) = \varphi^j(x_0) + \epsilon_k \nabla\varphi^j(x_0) \bullet h + \delta\overline{h} \le 0.$$

Step 2. Let us now prove the second part of the claim. Let $\overline{h} := \overline{x} - x_0$ and $j \in J(x_0)$. If φ^j is convex, we have

$$\nabla\varphi^j(x_0) \bullet \overline{h} \le \varphi(\overline{x}) < 0,$$

whereas if φ^j is affine, we have

$$\nabla\varphi^j(x_0) \bullet \overline{h} = \varphi(\overline{x}) \le 0.$$

Therefore, (i) follows from Step 1.

We now assume that $J(x_0) = \{1, 2 \dots, p\}$, $1 \le p \le n$, and let $\varphi := (\varphi^1, \dots, \varphi^p)$. Let $b := (-1, -1, \dots, -1) \in \mathbb{R}^p$. Then the linear system $\mathbf{D}\varphi(x_0)\overline{x} = b$, $x \in \mathbb{R}^n$ is solvable since $\operatorname{Rank} \mathbf{D}\varphi(x_0) = p$. If $\overline{h}$ is any such solution, then $\nabla\varphi^j(x_0) \bullet \overline{h} = \mathbf{D}\varphi(x_0)\overline{h} = -1$ for all $j \in J(x_0)$, and (ii) follows from Step 1. □

2.74 Theorem (Kuhn–Tucker). *Let* x_0 *be a solution of* (2.37). *Suppose that the constraints are qualified at* x_0. *Then the following* Kuhn–Tucker equilibrium condition *holds: For all* $j \in J(x_0)$ *there exists* $\lambda_j^0 \ge 0$ *such that*

$$\nabla f(x_0) + \sum_{j \in J(x_0)} \lambda_j^0 \nabla\varphi^j(x_0) = 0. \tag{2.38}$$

Theorem 2.74 is a simple application of the following version of the Farkas lemma.

2.75 Lemma (Farkas). *Let* v *and* $v_1, v_2, \dots, v_p$ *be vectors of* $\mathbb{R}^n$. *There exist* $\lambda_j \ge 0$ *such that*

$$v = \sum_{j=1}^{p} \lambda_j v_j \tag{2.39}$$

if and only if

$$\Big\{ h \in \mathbb{R}^n \;\Big|\; h \bullet v_j \le 0,\ \forall j = 1, \dots, p \Big\} \subset \Big\{ h \in \mathbb{R}^n \;\Big|\; h \bullet v \le 0 \Big\}. \tag{2.40}$$

Proof. In fact, if $\mathbf{A} := [v_1|v_2|\dots|v_n]$, (2.39) states that $\mathbf{A}\lambda = v$ has a nonnegative solution $\lambda \geq 0$. This is equivalent to saying that the second alternative of the Farkas lemma is false, i.e., $\forall h \in \mathbb{R}^n$ such that $\mathbf{A}^T h \geq 0$, we have $h \bullet v \geq 0$, that is, if $h \in \mathbb{R}^n$ satisfies $h \bullet v_j \geq 0$ for all j, then $h \bullet v \leq 0$. This is precisely (2.40). □

Proof of Theorem 2.74. For any $h \in \Gamma(x_0)$, let $r : [0,1] \to \mathcal{F}$ be a regular curve with $r(0) = x_0$ and $r'(0) = h$. Since 0 is a minimum point for $f(r(t))$, we have $\frac{d}{dt} f(r(t))_{|_{t=0}} \geq 0$, i.e.,

$$-\mathbf{D}f(x_0) \bullet h \leq 0 \qquad \forall h \in \Gamma(x_0),$$

i.e., $h \in \left\{ h \in \mathbb{R}^n \;\middle|\; h \bullet v \leq 0 \right\}$. Recalling the definition of $\Gamma(x_0)$, the claim follows by applying Lemma 2.75 with $v := -\nabla f(x^0)$ and $v_j = \nabla \varphi^j(x^0)$. □

2.76 Example. Let $\mathcal{P}$ be the problem of minimizing $-x_1$ with the constraints $x_1 \geq 0$ and $x_2 \geq 0$, $(1-x_1)^3 - x_2 \geq 0$. Clearly the unique solution is $x^0 = (1,0)$. Show that the constraints are not qualified at x^0 and that the Kuhn–Tucker theorem does not hold.

2.77 Remark. In analogy with Lagrange's multiplier theorem we may rewrite the Kuhn–Tucker equilibrium conditions (2.38) as

$$\begin{cases} \mathbf{D}f(x^0) + \sum_{j=1}^m \lambda_j^0 \mathbf{D}\varphi^j(x^0) = 0, \\ \lambda_j^0 \geq 0 \qquad \forall j = 1, \dots, m, \\ \sum_{j=1}^m \lambda_j^0 \varphi^j(x^0) = 0, \end{cases}$$

or, using the vectorial notation,

$$\begin{cases} \mathbf{D}f(x^0) + \lambda^0 \bullet \mathbf{D}\varphi(x_0) = 0, \\ \lambda^0 \geq 0, \\ \lambda^0 \bullet \varphi(x_0) = 0, \end{cases} \tag{2.41}$$

where $\lambda^0 = (\lambda_1^0, \dots, \lambda_m^0) \in \mathbb{R}^m$ and $\varphi = (\varphi^1, \dots, \varphi^m) : \mathbb{R}^n \to \mathbb{R}^m$. In fact, the equation $\sum_{j=1}^m \lambda_j^0 \varphi^j(x^0) = 0$ implies $\lambda_h^0 = 0$ if the corresponding constraint φ^h is not active. If (2.41) holds for some λ^0, we call it a *Lagrange multiplier* of (2.37) at x_0.

2.4.6 Stationary states for discrete-time Markov processes

Suppose that a system can be in one of n possible states, denote by $p_j^{(k)}$ the probability that it is in the state j at the discrete time k and set $p^{(k)} := (p_1^{(k)}, p_2^{(k)}, \dots, p_n^{(k)})$. A homogeneous Markov chain with values in a finite set is characterized by the fact that the probabilities of the states at time $k+1$ are a linear function of the probabilities at time k and that such a function does not depend on k, that is, there is a $n \times n$ matrix $\mathbf{P} \in M_{n,n}(\mathbb{R})$ such that

$$p^{(k+1)} = p^{(k)}\mathbf{P} \qquad \forall k, \tag{2.42}$$

where the product is the usual row by column product of linear algebra.

The matrix $\mathbf{P} = (p_{ij})$ is called the *transition matrix*, or *Markov matrix* of the system.

Since $\sum_{j=1}^{n} p_j^{(k)} = 1$ for every k, the matrix $\mathbf{P}$ has to be *stochastic* or *Markovian*, meaning that

$$\mathbf{P} = (p_{ij}), \qquad \sum_{j=1}^{n} p_{ij} = 1, \qquad p_{ij} \geq 0.$$

According to (2.42), the evolution of the system is then described by the powers of $\mathbf{P}$,

$$p^{(k)} = p^{(0)}\,\mathbf{P}^k \qquad \forall k. \tag{2.43}$$

A *stationary state* is a fixed point of $\mathbf{P}$ i.e., $x \in \mathbb{R}^n$ such that

$$x = \mathbf{P}^T x, \qquad \sum_{j=1}^{n} x^j = 1, \qquad x \geq 0. \tag{2.44}$$

The Perron–Frobenius theorem, see [GM3], ensures the existence of a stationary state.

2.78 Theorem (Perron–Frobenius). *Every Markov matrix has a stationary state.*

Proof. This is just a special case of the fact that every continuous map from a compact convex set into itself has a fixed point, see [GM3]. However, since here we deal with a linear map $x \to \mathbf{P}x$, we give a direct proof which uses compactness.

Let $S := \{x \in \mathbb{R}^n \,|\, x \geq 0, \sum_{j=1}^{n} x^j = 1\}$. S is a convex closed and bounded set of $\mathbb{R}^n$, and $\mathbf{P}$ maps S into S and is stochastic. Fix $x_0 \in S$ and consider the sequence $\{x_k\}$ given by

$$x_k := \frac{1}{k}\sum_{i=0}^{k-1} x_0\mathbf{P}^i.$$

x_k is a convex combination of points in S and therefore $x_k \in S$. The sequence $\{x_k\}$ is then bounded and, by the Bolzano–Weierstrass theorem, there exists a subsequence $\{x_{n_k}\}$ of $\{x_k\}$ and $x \in S$ such that $x_{n_k} \to x$. On the other hand, for any k we have

$$x_k - x_k\mathbf{P} = \frac{1}{k}\Big(\sum_{i=0}^{k-1} x_0\mathbf{P}^i - \sum_{i=0}^{k-1} x_0\mathbf{P}^{i+1}\Big) = \frac{1}{k}(x_0 - x_0\mathbf{P}^{k+1})$$

so that

$$|x_k - x_k\mathbf{P}| \leq \frac{1}{k}.$$

Passing to the limit along the subsequence $\{x_{k_n}\}$, we then get $x - x\mathbf{P} = 0$. □

Another proof of Theorem 2.78. We give another proof of this claim which uses only convexity arguments, in particular, the Farkas–Minkowski theorem. Let $\mathbf{P}$ be a stochastic $n \times n$ matrix. Define

$$u := (1, 1, \dots, 1) \in \mathbb{R}^n, \qquad b := (0, 0, \dots, 0, 1) \in \mathbb{R}^{n+1}$$

and

$$\mathbf{A} = \begin{pmatrix} \boxed{\mathbf{P}^T - \mathrm{Id}} \\ u^T \end{pmatrix} \qquad \text{in } M_{(n+1),n}(\mathbb{R}).$$

The existence of a stationary point x for $\mathbf{P}$ is then equivalent to

$$\mathbf{A}x = b \qquad \text{has a nonnegative solution } x \geq 0. \tag{2.45}$$

Now, we show that Farkas's alternative does not hold, i.e., the system $\mathbf{A}^T y \geq 0$, $b \bullet y < 0$ has no solution. Suppose it holds; then there is a y such that $b \bullet y = y_{n+1} < 0$. If we write y as $y = (z^1, z^2, \dots, z^n, -\lambda) =: (z, -\lambda)$, $\lambda > 0$, we then have

$$0 \leq \mathbf{A}^T y = y^T \mathbf{A} = (z, -\lambda) \begin{pmatrix} \boxed{\mathbf{P}^T - \mathrm{Id}} \\ u^T \end{pmatrix} = z(\mathbf{P}^T - \mathrm{Id}) - \lambda u^T,$$

i.e.,

$$z^T(\mathbf{P}^T - \mathrm{Id}) \geq \lambda u^T.$$

Thus

$$\sum_{j=1}^{n} z^j p_{ji} - z^i \geq \lambda > 0 \qquad \forall i = 1, \dots, n. \tag{2.46}$$

On the other hand, if m is the index such that $z^m = \max_j z^j$, we have

$$\sum_{j=1}^{n} z^j p_{jm} \leq \max_j z^j = z^m,$$

hence

$$\sum_{j=1}^{n} z^j p_j^m - z^m \leq 0,$$

and this contradicts (2.46). □

2.4.7 Linear programming

We shall begin by illustrating some classical examples.

2.79 Example (Investment management). A bank has 100 million dollars to invest: a part L in loans at a rate, say, of 10% and a part S in bonds, say at 5%, with the aim of maximizing its profits $0.1L + 0.05S$. Of course, the bank has trivial restrictions, $L \geq 0$, $S \geq 0$ and $L + S \leq 100$, but also needs some cash of at least 25% of the total amount, $S \geq 0.25(L + S)$, i.e., $3S \geq L$ and needs to satisfy requests for important clients which on average require 30 million dollars, i.e., $L \geq 30$. The problem is then

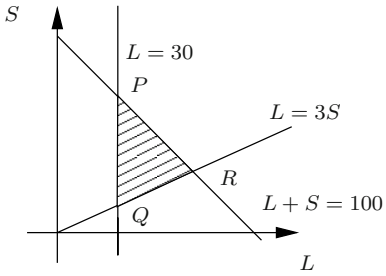


Figure 2.11. Illustration for Example 2.79.

$$\begin{cases} 0.10L + 0.05S \to \max, \\ L + S \le 100,\ L \le 3S,\ L \ge 30, \\ L \ge 0,\ S \ge 0. \end{cases}$$

With reference to Figure 2.11, the shaded triangle represent the admissible values (L, S); on the other hand, the gradient of the objective function $C = 0,1L + 0.05S$ is constant $\nabla C = (0.1, 0.05)$ and the level lines of C are straight lines. Consequently, the optimal portfolio is to be found among the extreme points P, Q and R of the triangle, and, as it is easy to verify, the optimal configuration is in R.

2.80 Example (The diet problem). The daily diet of a person is composed of a number of components $j = 1, \dots, n$. Suppose that component j has a unitary cost c_j and contains a quantity a_{ij} of the nourishing i, $i = 1, \dots, m$, that is required in a daily quantity b_i. We want to minimize the cost of the diet. With standard vectorial notation the problem is

$$c \bullet x \to \min \qquad \text{in} \qquad \left\{ x \,\middle|\, \mathbf{A}x \ge b, x \ge 0 \right\}.$$

2.81 Example (The transportation problem). Suppose that a product (say oil) is produced in quantity s_i at places $i = 1, 2, \dots, n$ (Arabia, Venezuela, Alaska, etc.) and is requested at the different markets j, $j = 1, 2, \dots, m$ (New York, Tokyo, etc.) in quantity d_j. If c_{ij} is the transportation cost from i to j, we want to minimize the cost of transportation taking into account the constraints. The problem is then finding $x = (x_{ij}) \in \mathbb{R}^{nm}$ such that

$$\begin{cases} \sum_{i,j} c_{ij} x_{ij} \to \min, \\ \sum_{i=1}^{n} x_{ij} = d_j,\ \forall j, \\ \sum_{j=1}^{m} x_{ij} \le s_i\ \forall i, \\ x \ge 0. \end{cases}$$

Here x is a vector with real-valued components, but for other products, for instance cars, the unknown would be a vector with integral components.

2.82 Example (Maximum profit). Suppose we are given $s_1, \dots, s_n$ quantities of basic products (resources) from which we may produce goods that sell at prices $p_1, p_2, \dots, p_m$. If a_{ij} is the quantity of product i, $i = 1, \dots, n$, to produce j, $j = 1, \dots, m$, our problem is finding the quantities x_j of goods j in order to maximize profits, i.e.,

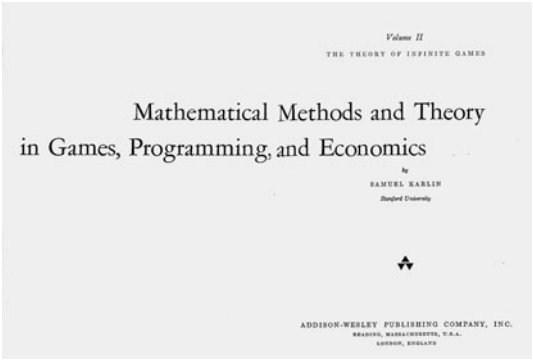


Figure 2.12. A classical textbook on linear programming and economics.

$$\begin{cases} \sum_{j=1}^{m} p_j x_j \to \max, \\ \sum_{j=1}^{n} a_{ij} x_j \le s_i, \\ x \ge 0. \end{cases}$$

In the previous examples, one wants to minimize or maximize a function, called the *objective function*, which is linear, in a set of *admissible* or *feasible* solutions, defined by a finite number of constraints defined by linear equalities or inequalities: This is the generic problem of *linear programming*. By possibly changing the sign of the objective function and/or of the inequalities constraints, observing that an equality constraint is equivalent to two inequalities constraints and replacing the variable x whose components are not necessarily nonnegative with $x = u - v$, $u, v \ge 0$, the linear programming problem can always be transformed into

$$f(x) := c \bullet x \to \min \qquad \text{in} \qquad \mathcal{P} := \Big\{ x \,\Big|\, \mathbf{A}x \ge b,\ x \ge 0 \Big\}, \tag{2.47}$$

where $c, x \in \mathbb{R}^n$, $\mathbf{A} \in M_{m,n}$ and $b \in \mathbb{R}^m$.

One of the following situations may, in principle, happen to hold:

(i) $\mathcal{P}$ is empty,
(ii) $\mathcal{P}$ is nonempty and the objective function in not bounded from below on $\mathcal{P}$,
(iii) $\mathcal{P}$ is nonempty and f is bounded from below.

In the last case, f has (at least) a minimizer and all the minimizers are extreme points of the convex set $\mathcal{P}$ by Proposition 2.42. We say that problem (2.47) has an *optimal solution*.

The problem transforms then into the problem of deciding in which of the previous cases we find ourselves and of possibly finding the optimal extreme points. In the real applications, where the number of constraints may be quite high, the effectiveness of the algorithm is also a further problem. Giving up efficiency, we approach the first two problems as follows.

We introduce the slack variables $x' := \mathbf{A}x - b \geq 0$ and transform the constraint $\mathbf{A}x \geq b$ into

$$\mathbf{A}' \begin{pmatrix} x \\ x' \end{pmatrix} = b, \qquad \mathbf{A}' := \left(\boxed{\mathbf{A}} \quad \boxed{-\mathrm{Id}} \right).$$

Writing $z = (x, x')$ and $F(z) := \sum_{i=1}^{n} c^i x^i + \sum_{i=1}^{m} 0 \cdot x'_i$, problem (2.47) transforms into

$$F(z) \to \min \qquad \text{in} \qquad \mathcal{F} := \Big\{ z \;\Big|\; \mathbf{A}'z = b,\ z \geq 0 \Big\}. \tag{2.48}$$

It is easily seen that $\mathcal{F}$ is nonempty if $\mathcal{P}$ is nonempty and that F is bounded from below on $\mathcal{F}$ if and only if f is bounded from below on $\mathcal{P}$. Therefore, F attains its minimum in one of the extreme points of $\mathcal{F}$ if and only if f has a minimizer in $\mathcal{P}$. All extreme points of $\mathcal{F}$ can be found by means of Theorem 2.61; the minimizers are then detected by comparison.

a. The primal and dual problem

Problem (2.47) is called the *primal* problem of linear programming, since one also introduces the *dual problem* of linear programming as

$$g(y) := b \bullet y \to \max \qquad \text{in} \qquad \mathcal{P}^* = \Big\{ y \;\Big|\; \mathbf{A}^T y \leq c,\ y \geq 0 \Big\}. \tag{2.49}$$

Of course, (2.49) can be rephrased as the minimum problem

$$h(y) := -b \bullet y \to \min \qquad \text{in} \qquad \mathcal{P}^* = \Big\{ y \;\Big|\; -\mathbf{A}^T y \geq -c,\ y \geq 0 \Big\} \tag{2.50}$$

which is similar to (2.47): Just exchange $-b$ and c, and replace $\mathbf{A}$ with $-\mathbf{A}^T$, and the following holds.

2.83 Proposition. *The dual problem of linear programming* (2.49) *has a solution if and only if $\mathcal{P}^* \neq \emptyset$ and g is bounded from above.*

The next theorem motivates the notation primal and dual problems of linear programming.

2.84 Theorem (Kuhn–Tucker equilibrium conditions). *Let f and $\mathcal{P}$ be as in* (2.47) *and let g and $\mathcal{P}^*$ be as in* (2.49)*. We have the following:*

(i) *$g(y) \leq f(x)$ for all $x \in \mathcal{P}$ and all $y \in \mathcal{P}^*$.*
(ii) *f has a minimizer $\overline{x} \in \mathcal{P}$ if and only if g has a maximizer $\overline{y} \in \mathcal{P}^*$ and, in this case, $f(\overline{x}) = g(\overline{y})$.*
(iii) *Let $x \in \mathcal{P}$ and $y \in \mathcal{P}^*$. The following claims are equivalent:*
  a) $(c - \mathbf{A}^T y) \bullet x = 0$.
  b) $(\mathbf{A}x - b) \bullet y = 0$.
  c) $f(x) = g(y)$.
  d) x is a minimizer for f and $y \in \mathcal{P}^*$ is a maximizer for g.

Proof. If $x \in \mathcal{P}$, then $x \geq 0$ and $\mathbf{A}x \geq b$. For $y \in \mathcal{P}^*$ we then get

$$f(x) = x \bullet c \geq x \bullet \mathbf{A}^T y = \mathbf{A}x \bullet y \geq b \bullet y = g(y),$$

i.e., (i).

(ii) Let $\overline{x}$ be a minimizer for the primal problem. Then f is bounded from below. We introduce the slack variables $x' = \mathbf{A}x - b \geq 0$ and set $z = (x, x')$. Then $\overline{x}$ is a solution of the primal problem (2.47) if and only if $\overline{z} := (\overline{x}, \overline{x'})^T$ minimizes

$$F(z) := c \bullet x \qquad \text{in} \qquad \mathcal{F} := \left\{ z \,\middle|\, \mathbf{A}' z = b,\ z \geq 0 \right\}$$

where

$$\mathbf{A}' := \left(\boxed{\mathbf{A}} \quad \boxed{\text{Id}} \right).$$

We may also assume that $\overline{z}$ is an extreme point of $\mathcal{F}$. As we saw in the proof of Theorem 2.61, if $\alpha_1, \alpha_2, \ldots, \alpha_k$ are the indices of the nonzero components of $\overline{z}$, the submatrix $\mathbf{B}$ of $\mathbf{A}'$ made of the columns of indices $\alpha_1, \alpha_2, \ldots, \alpha_k$ has maximal rank. If x_B denotes the vector with components the nonzero components of x, then $\mathbf{B}x_B = b$, and if we set $c_B := (c_{\alpha_1}, c_{\alpha_2}, \ldots, c_{\alpha_k})$ and choose $\overline{y}$ such that $\mathbf{B}^T \overline{y} = c_B$, we have

$$g(\overline{y}) = \overline{y} \bullet b = \overline{y} \bullet \mathbf{B}x_B = \overline{\mathbf{B}^T \overline{y}} x_B = c_B \bullet x_B = f(\overline{x}).$$

Then (i) yields that $\overline{y}$ is a maximizer of the dual problem.

(iii) (a) or (b) $\Rightarrow$ (c). If $(c - \mathbf{A}^T y) \bullet x = 0$ with $x \in \mathcal{P}$ and $y \in \mathcal{P}^*$, then

$$f(x) = c \bullet x = \mathbf{A}^T y \bullet x = y \bullet \mathbf{A}x \leq b \bullet y = g(y),$$

thus $f(x) = g(y)$ because of (i).

(c) $\Rightarrow$ (a) and (b). If $f(x) = g(y)$ and we set $\gamma := b - \mathbf{A}x$, we have

$$0 = f(x) - g(y) = c \bullet x - b \bullet y = c \bullet x - \mathbf{A}x \bullet y + \gamma \bullet y = (c - \mathbf{A}^T y) \bullet x + \gamma \bullet y\,.$$

Since the addenda are nonnegative, we conclude

$$(c - \mathbf{A}^T x) \bullet x = 0 \qquad \text{and} \qquad (\mathbf{A}x - b) \bullet y = 0.$$

(c) $\Rightarrow$ (d). If $f(x) = g(y)$, then (i) yields $f(x') \geq g(y) = f(x)$ for all $x' \in \mathcal{P}$, hence x is a minimizer of f. Similarly y is a maximizer of g in $\mathcal{P}^*$.

(d) $\Rightarrow$ (c). This follows trivially from (ii). □

A consequence of the previous theorem is the following *duality theorem of linear programming.*

2.85 Corollary (Duality theorem). *Let* (2.47) *and* (2.49) *be the primal and the dual problems of linear programming. One and only one of the following alternatives arises:*

(i) *There exist a minimizer $\overline{x} \in \mathcal{P}$ for f and a maximizer $\overline{y} \in \mathcal{P}^*$ for g and $f(\overline{x}) = g(\overline{y})$. This arises if and only if $\mathcal{P}$ and $\mathcal{P}^*$ are both nonempty.*
(ii) *$\mathcal{P} \neq \emptyset$ and f is not bounded from below in $\mathcal{P}$.*
(iii) *$\mathcal{P}^* \neq \emptyset$ and g is not bounded from above in $\mathcal{P}^*$.*
(iv) *$\mathcal{P}$ and $\mathcal{P}^*$ are both empty.*

Proof. Trivially, (iv) is inconsistent with any of (i), (ii) or (iii); (iii) is inconsistent with (ii) because of (i) of Theorem 2.84, and (iii) is inconsistent with (i). Similarly (ii) is inconsistent with (i). Therefore, the four alternatives are disjoint. If (ii), (iii) and (iv) do not hold, we therefore have

$$\begin{cases} \mathcal{P} = \emptyset \text{ or } (\mathcal{P} \neq \emptyset \text{ and } f \text{ is bounded from below}), \\ \mathcal{P}^* = \emptyset \text{ or } (\mathcal{P}^* \neq \emptyset \text{ and } g \text{ is bounded from above}), \\ \mathcal{P} \text{ or } \mathcal{P}^* \text{ are nonempty}, \end{cases}$$

that is, one of the following alternatives holds:

$\mathcal{P} \neq \emptyset$ and f is bounded from below,

$\mathcal{P}^* \neq \emptyset$ and g is bounded from above,

$\mathcal{P} \neq \emptyset,\ \mathcal{P}^* \neq \emptyset,\ f$ is bounded from below and g is bounded from above.

In any case, both the primal and the dual problem of linear programming have solutions and, according to (iii) of Theorem 2.84, the alternative (i) holds. □

Corollary 2.85 is actually a *convex duality* theorem: Here we supply a direct proof by duality, using Farkas's alternative.

A proof of Corollary 2.85 which uses convex duality. Set

$$\widehat{\mathbf{A}} := \begin{pmatrix} \boxed{-\mathbf{A}} & \boxed{0} \\ \boxed{0} & \boxed{\mathbf{A}^T} \\ c^T & -b \end{pmatrix}$$

and

$$\widehat{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \widehat{b} = \begin{pmatrix} -b \\ c \\ 0 \end{pmatrix}.$$

Then (i) is equivalent to

$$\widehat{\mathbf{A}}\widehat{x} \leq \widehat{b} \qquad \text{has a solution } \widehat{x} \geq 0.$$

Farkas's alternative then yields the following: If (i) does not hold, then there exists $\widehat{y} = (u, v, \lambda)$ such that

$$\begin{cases} \begin{pmatrix} u^T & v^T & \lambda \end{pmatrix} \begin{pmatrix} \boxed{-\mathbf{A}} & \boxed{0} \\ \boxed{0} & \boxed{\mathbf{A}^T} \\ c^T & -b^T \end{pmatrix} \geq 0, \\ \begin{pmatrix} u^T & v^T & \lambda \end{pmatrix} \begin{pmatrix} -b \\ c \\ 0 \end{pmatrix} < 0, \\ \begin{pmatrix} u^T & v^T & \lambda \end{pmatrix} \geq 0, \end{cases}$$

or, after a simple computation, the problem

$$\mathbf{A}u \geq \lambda\, b, \qquad \mathbf{A}^T v \leq \lambda\, c, \qquad c \bullet u \;\leq\; b \bullet v \tag{2.51}$$

has a solution (u, v, λ) with $u \geq 0$, $v \geq 0$ and $\lambda \geq 0$.

Now, we claim that $\lambda = 0$. In fact, if $\lambda \neq 0$, then $u/\lambda \in \mathcal{P}$, $v/\lambda \in \mathcal{P}^*$, consequently, $c \bullet u/\lambda \;<\; b \bullet v/\lambda$: a contradiction because of (i) of Theorem 2.84. Thus, (2.51) reduces to the following claim: The problem

$$\mathbf{A}u \geq 0, \qquad \mathbf{A}^T v \leq 0, \qquad c \bullet u \;<\; b \bullet v$$

has a solution (u, v) with $u \geq 0$ and $v \geq 0$.

We notice that the inequality $c \bullet u \;<\; b \bullet v$ implies that either $c \bullet u \;< 0$ or $b \bullet v \;> 0$ or both. In the case $c \bullet u \;< 0$, we have $\mathcal{P}^* = \emptyset$, since otherwise if $y \geq 0$ and $\mathbf{A}^T y \leq c$, then from $\mathbf{A}u \geq 0$, $u \geq 0$ we would infer $0 \leq\; y \bullet \mathbf{A}u \;=\; A^T y \bullet u \;\leq\; c \bullet u$, a contradiction. If, moreover, $\mathcal{P} = \emptyset$, the alternative (iv) holds; otherwise, if $x \in \mathcal{P}$, then $\mathbf{A}(x + \theta u) \geq b + \theta 0 = b$, $x + \theta u \geq 0$ for some $\theta \geq 0$, and $c \bullet x + \theta u \;=\; c \bullet x \;+ \theta\, c \bullet u \;\to\; -\infty$ as $\theta \to +\infty$, that is, the alternative (ii) holds.

In the case $b \bullet v \;> 0$, as in the case $c \bullet u \;< 0$, we see that $\mathcal{P} = \emptyset$. If also $\mathcal{P}^* = \emptyset$, then (iv) holds; while, if there exists $y \in \mathcal{P}^*$, then $v + \theta y \in \mathcal{P}^*$ and $v + \theta y \to +\infty$ as $\theta \to +\infty$, and (iii) holds. □

2.86 Example. Let us illustrate the above discussing the dual of the transportation problem. Suppose that crude oil is extracted in quantities s_i, $i = 1, \ldots, n$ in places $i = 1, \ldots, n$ and is requested in the markets $j = 1, \ldots, m$ in quantity d_j. Let c_{ij} be the transportation cost from i to j. The optimal transportation problem consists in determining the quantities of oil to be transported from i to j minimizing the overall transportation cost

$$\sum_{i,j} c_{ij} x_{ij} \to \min, \tag{2.52}$$

and satisfying the constraints, in our case, the markets requests and the capability of production

$$\begin{cases} \sum_{j=1}^{m} x_{ij} \leq s_i & \forall i, \\ \sum_{i=1}^{n} x_{ij} = d_j & \forall j, \\ x \geq 0. \end{cases} \tag{2.53}$$

Of course, a necessary condition for the solvability is that the production be larger than the markets requests

$$\sum_{j=1}^{m} d_j = \sum_{\substack{i=1,n \\ j=1,m}} x_{ij} \leq \sum_{i=1}^{n} s_i.$$

Introducing the matrix notation

$$\begin{cases} x := (x_{11}, \ldots, x_{1m}, x_{21}, \ldots, x_{2m}, \ldots, x_{n1}, \ldots, x_{nm}) \in \mathbb{R}^{nm}, \\ c := (c_{11}, \ldots, c_{1m}, c_{21}, \ldots, c_{2m}, \ldots, c_{n1}, \ldots, c_{nm}) \in \mathbb{R}^{nm}, \\ b := (s_1, s_2, \ldots, s_n, d_1, \ldots, d_m) \end{cases}$$

and setting $\mathbf{A} \in M_{n+m,nm}(\mathbb{R})$,

$$\mathbf{A} := \begin{pmatrix} u & 0 & 0 & \ldots & 0 \\ 0 & u & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & u \\ \ldots & & & & \\ e_1 & e_1 & e_1 & \ldots & e_1 \\ e_2 & e_2 & e_2 & \ldots & e_2 \\ \ldots & & & & \\ e_m & e_m & e_m & \ldots & e_m \end{pmatrix},$$

where $u := (1, 1, \dots, 1) \in \mathbb{R}^m$ and $0 = (0, 0, \dots, 0) \in \mathbb{R}^m$, we may formulate our problem as

$$\begin{cases} c \bullet x \ \to \min, \\ \mathbf{A}x \le b, \\ x \ge 0. \end{cases}$$

The dual problem is then

$$\begin{cases} b \bullet y \ \to \max, \\ \mathbf{A}^T y \le c, \\ y \ge 0, \end{cases}$$

that is, because of the form of $\mathbf{A}$ and setting

$$y := (u_1, u_2, \dots, u_n, v_1, v_2, \dots, v_m),$$

the maximum problem

$$\begin{cases} \sum_{i=1}^n s_i u_i + \sum_{j=1}^m d_j v_j \to \max, \\ u_i + v_j \le c_{ij} \qquad \forall i, j, \\ u \ge 0, \ v \ge 0. \end{cases}$$

If we interpret u_i as the toll at departure and v_i as the toll at the arrival requested by the shipping agent, the dual problem may be regarded as the problem of maximizing the profit of the shipping agent. Therefore, the quantities $\overline{u}_i$ and $\overline{v}_i$ which solve the dual problem represent the maximum tolls one may apply in order not to be out of the market.

2.87 Example. In the primal problem of linear programming one minimizes a linear function on a polyhedral set

$$\begin{cases} c \bullet x \ \to \min, \\ \mathbf{A}x \le b, \ x \ge 0, \end{cases}$$

or, equivalently,

$$\begin{cases} -c \bullet x \ \to \max, \\ \mathbf{A}x \le b, \ x \ge 0. \end{cases}$$

Since the constraint is qualified at all points, the primal problem has a minimum $x \ge 0$ if and only if the Kuhn–Tucker equilibrium condition holds, i.e., there exists $\lambda \ge 0$ such that

$$(c - \mathbf{A}^T \lambda)x = 0.$$

This way we find again the optimality conditions of linear programming.

2.4.8 Minimax theorems and the theory of games

The theory of games consists in mathematical models used in the study of processes of decisions that involve conflict or cooperation. The modern origin of the theory dates back to a famous paper by John von Neumann (1903–1957) published in German in 1928 with the title "On the Theory of Social Games"[3] and to the very well-known book by von Neumann and the

[3] J. von Neumann, Theorie der Gesellschaftsspiele, Math. Ann. **100** (1928) 295–320.


Figure 2.13. John von Neumann (1903–1957) and Oskar Morgenstern (1902–1976).

economist Oskar Morgenstern, *Theory of Games and Economic Behavior* published in 1944. There one can find several types of games with one or more players, with zero or nonzero sum, cooperative or non-cooperative, ... For its relevance in economy, social sciences or biology the theory has greatly developed[4]. Here we confine ourselves to illustrating only a few basic facts.

a. The minimax theorem of von Neumann

In a game with two players P and Q, each of them relays on a set of possible strategies, say respectively A and B; also, two *utility functions* $U_P(x, y)$ and $U_Q(x, y)$ are given, representing for each choice of the strategy $x \in A$ of P and $y \in B$ of Q the gain for P and Q resulting from the choices of the strategies x and y.

Let us consider the simplest case of a *zero sum game* in which the common value $K(x, y) := U_P(x, y) = -U_Q(x, y)$ is at the same time the gain for P and minus the gain for Q resulting from the choices of the strategies x and y.

[4] The interested reader is referred for classical literature to

- J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, 1944, that follows a work of Ernst Zermelo (1871–1951), *Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels*, 1913 and a work of Emile Borel (1871–1956) *La théorie du jeu et les équations intégrales à noyau symmétrique*, 1921.
- R. Luce, H. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Wiley, New York, 1957.
- S. Karlin, *Mathematical Methods and Theory in Games, Programming and Economics*, 2 vols., Addison–Wesley, Reading, MA, 1959.
- W. Lucas, An overview of the mathematical theory of games, Manage. Sci. **18** (1972), 3–19.
- M. Shubik, *Game Theory in the Social Sciences: Concepts and Solutions*, MIT Press, Boston, MA, 1982.

Each player tries to do his best against every strategy of the other player. In doing that, the *expected payoff* or, simply, *payoff*, i.e., the remuneration that P and Q can expect not taking into account the strategy of the other player, are

$$\begin{aligned}\text{Payoff}(P) &:= \inf_{y\in B}\sup_{x\in A} U_P(x,y) = \inf_{y\in B}\sup_{x\in A} K(x,y),\\ \text{Payoff}(Q) &:= \inf_{x\in A}\sup_{y\in B} U_Q(x,y) = \inf_{x\in A}\sup_{y\in B} -K(x,y) = -\sup_{x\in A}\inf_{y\in B} K(x,y).\end{aligned}$$

Although the game has zero sum, the payoffs of the two players are not related, in general, we trivially only have

$$\sup_{x\in A}\inf_{y\in B} K(x,y) \le \inf_{y\in B}\sup_{x\in A} K(x,y), \tag{2.54}$$

i.e.,

$$\text{Payoff}(P) + \text{Payoff}(Q) \ge 0.$$

Of course, if the previous inequality is strict, there are no choices of strategies that allow both players to reach their payoff.

The next proposition provides a condition for the existence of a couple of *optimal strategies*, i.e., of strategies that allow each players to reach their payoff.

2.88 Proposition. *Let A and B be arbitrary sets and $K : A \times B \to \mathbb{R}$. Define $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ respectively as*

$$f(x) := \inf_{y\in B} K(x,y), \qquad g(y) := \sup_{x\in A} K(x,y).$$

Then there exists $(\overline{x},\overline{y}) \in A \times B$ such that

$$K(x,\overline{y}) \le K(\overline{x},\overline{y}) \le K(\overline{x},y) \qquad \forall x,y \in A\times B \tag{2.55}$$

if and only if f attains its maximum in A, g attains its minimum in B and $\sup_{x\in A} f(x) = \inf_{y\in B} g(y)$. In this case,

$$\sup_{x\in A}\inf_{y\in B} K(x,y) = K(\overline{x},\overline{y}) = \inf_{y\in B}\sup_{x\in A} K(x,y).$$

Proof. If $(\overline{x},\overline{y})$ satisfies (2.55), then

$$\begin{aligned}K(\overline{x},\overline{y}) &= \inf_{y\in B} K(\overline{x},y) = f(\overline{x}) \le \sup_{x\in A} f(x),\\ K(\overline{x},\overline{y}) &= \sup_{x\in A} K(x,\overline{y}) = g(\overline{y}) \ge \inf_{y\in B} g(y),\end{aligned}$$

hence $\sup_{x\in A} f(x) = \inf_{y\in B} g(y)$ if we take into account (2.54). We leave the rest of the proof to the reader. □

A point $(\overline{x}, \overline{y})$ with property (2.55) is a *saddle point* for K. Therefore, in the context of games with zero sum, the saddle points of K yield couples of optimal strategies. The value of K on a couple of optimal strategies is called the *value of the play*. Answering the question of when there exists a saddle point is more difficult and is the content of the next theorem.

We recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *quasiconvex* if its sublevel sets are convex, and *quasiconcave* if $-f$ is quasiconvex.

2.89 Theorem (Minimax theorem of von Neumann). *Let $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^n$ be two compact convex sets and let $K : A \times B \to \mathbb{R}$ be a function such that*

(i) *$x \to K(x, y)$ is quasiconvex and lower semicontinuous $\forall y \in B$,*
(ii) *$y \to K(x, y)$ is quasiconcave and upper semicontinuous $\forall x \in A$.*

Then K has a saddle point in $A \times B$.

Proof. According to Proposition 2.88 it suffices to prove that numbers

$$a := \min_{x\in A} \max_{y\in B} K(x, y) \qquad \text{and} \qquad b := \max_{y\in B} \min_{x\in A} K(x, y)$$

exist and are equal. Fix $y \in B$, the function $x \to K(x, y)$ attains its minimum at $z(y) \in A$ being A compact, and $K(z(y), y) = \min_{x\in A} K(x, y)$. Set

$$h(y) := -K(z(y), y), \qquad y \in B.$$

We now show that h is quasiconvex and lower semicontinuous, thus there is

$$b := -\min_{y\in B} \Big(-\min_{x\in A} K(x, y)\Big) = \max_{y\in B} \min_{y\in A} K(x, y).$$

Similarly, one proves the existence of a.

Let us show that h is quasiconvex and lower semicontinuous, that is, that for all $t \in \mathbb{R}$ the set

$$H := \Big\{ y \in B \,\Big|\, h(y) \le t \Big\}$$

is convex and closed. First we will show that H is convex. For any $w \in B$, consider

$$G(w) := \Big\{ y \in B \,\Big|\, -K(z(w), y) \le t \Big\}.$$

Because of (ii), $G(w)$ is convex and closed; moreover, $H \subset G(w)\ \forall w$, since $K(z(y), y) \le K((z(w), y)\ \forall w, y \in B$. In particular, for $x, y \in H$ and $\lambda \in]0, 1[$ we have $u \in G(w)$ $\forall w \in B$ if $u := (1-\lambda)y+\lambda x$, hence $u \in G(u)$, i.e., $u \in H$. This proves that H is convex. Let us prove now that H is closed. Let $\{y_n\} \subset H$, $y_n \to y$ in B, then $y \in G(w)\ \forall w \in B$, in particular, $y \in G(y)$, i.e., $y \in H$. Therefore, H is closed.

Let us prove that $a = b$. Since $b \le a$ trivially, it remains to show that $a \le b$. Fix $\epsilon > 0$ and consider the function $T : A \times B \to \mathcal{P}(A \times B)$ given by

$$T(x, y) := \Big\{ (u, v) \in A \times B \,\Big|\, K(u, y) < b + \epsilon,\ K(x, v) > a - \epsilon \Big\}.$$

We have $T(x, y) \ne \emptyset$ since $\min_{u\in A} K(u, y) \le b$ and $\max_{v\in B} K(x, v) \ge a$; moreover, $T(x, y)$ is convex. Since

$$\begin{aligned} T^{-1}(\{(u, v)\}) :&= \Big\{ (x, y) \in A \times B \,\Big|\, (u, v) \in T(x, y) \Big\} \\ &= \Big\{ (x, y) \in A \times B \,\Big|\, K(u, y) < b + \epsilon,\ K(x, v) > a - \epsilon \Big\} \\ &= \Big\{ x \in A \,\Big|\, K(x, v) > a - \epsilon \Big\} \times \Big\{ y \in B \,\Big|\, K(u, y) < b - \epsilon \Big\}, \end{aligned}$$

$T^{-1}(\{(u,v)\})$ is also open. We now claim, compare Theorem 2.90, that there is a fixed point for T, i.e., that there exists $(\overline{x},\overline{y}) \in A \times B$ such that $(\overline{x},\overline{y}) \in T(\overline{x},\overline{y})$, i.e., $a - \epsilon < k(\overline{x},\overline{y}) < b + \epsilon$. ϵ being arbitrary, we conclude $a \le b$. □

For its relevance, we now state and prove the fixed point theorem we have used in the proof of the previous theorem.

2.90 Theorem (Kakutani). *Let K be a nonempty, convex and compact set, and let $F : K \to \mathcal{P}(K)$ be a function such that*

(i) *$F(x)$ is nonempty and convex for each $x \in K$,*
(ii) *$F^{-1}(y)$ is open in K for every $y \in \mathcal{P}(K)$.*

Then F has at least a fixed point, i.e., there exists $\overline{x}$ such that $\overline{x} \in F(\overline{x})$.

Proof. Clearly, the family of open sets $\{F^{-1}(y)\}_y$ is an open covering of K, consequently, there exist $y_1, y_2, \ldots, y_n \in \mathcal{P}(K)$ such that $K \subset \cup_{i=1}^n F^{-1}(y_i)$. Let $\{\varphi_i\}$ be a partition of unity associated to $\{F^{-1}(y_i)\}_{i=1,\ldots,n}$ and set

$$p(x) := \sum_{i=1}^{n} \varphi_i(x) y_i \qquad \forall x \in K_0 := \mathrm{co}(\{y_1, y_2, \ldots, y_n\}) \subset K.$$

Obviously, p is continuous and $p(K_0) \subset K_0$. According to Brouwer's theorem, see [GM3], p has a fixed point $\overline{x} \in K_0$. To conclude, we now prove that $p(x) \in F(x)\ \forall x \in K_0$, from which we infer that $\overline{x} = p(\overline{x}) \in F(\overline{x})$, i.e., $\overline{x}$ is a fixed point for F. Let $x \in K_0$. For each index j such that $\varphi_j(x) \neq 0$ we have trivially $x \in F^{-1}(y_j)$, thus $y_j \in F(x)$. Since $F(x)$ is convex, we see that

$$p(x) = \sum_{i=1}^{n} \varphi_i(x) y_i = \sum_{\{j \mid \varphi_j(x) \neq 0\}} \varphi_j(x) y_j,$$

hence $p(x) \in F(x)$. □

We now present a variant of Theorem 2.89.

2.91 Theorem. *Let $K : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $K = K(x,y)$, be a function convex in x for any fixed y and concave in y for any fixed x. Assume that there exist $\overline{x} \in \mathbb{R}^n$ and $\overline{y} \in \mathbb{R}^m$ such that*

$$K(x,\overline{y}) \to +\infty \qquad \text{as } x \to +\infty,$$
$$K(\overline{x},y) \to -\infty \qquad \text{as } y \to -\infty.$$

Then K has a saddle point (x_0, y_0).

Observe that $K(x,y)$ is continuous in each variable. Let us start with a special case of Theorem 2.89 for which we present a more direct proof.

2.92 Proposition. *Let A and B be compact subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and let $K : A \times B \to \mathbb{R}$, $K = K(x,y)$ be a function that is convex and lower semicontinuous in x for any fixed y and concave and upper semicontinuous in y for any fixed x. Then K has a a saddle point $(x_0, y_0) \in A \times B$.*

Proof. Step 1. Since $x \to K(x,y)$ is lower semicontinuous and A is compact, then for every $y \in B$ there exists at least one $x = x(y)$ such that

$$K(x(y),y) = \inf_{x\in A} K(x,y). \tag{2.56}$$

Let

$$g(y) := \inf_{x\in A} K(x,y) = K(x(y),y), \qquad y \in B. \tag{2.57}$$

The function g is upper semicontinuous, because $\forall y_0$ and $\forall \epsilon > 0$ there exists $\overline{x}$ such that

$$g(y_0) + \epsilon \geq K(\overline{x}, y_0) \geq \limsup_{y\to y_0} K(\overline{x}, y) \geq \limsup_{y\to y_0} g(y).$$

Consequently, there exists $y_0 \in B$ such that

$$g(y_0) := \max_{y\in B} g(y), \tag{2.58}$$

and, therefore,

$$g(y_0) \leq K(x, y_0) \qquad \forall x \in A. \tag{2.59}$$

Step 2. We now prove that for every $y \in B$ there exists $\widetilde{x}(y) \in A$ such that

$$K(\widetilde{x}(y), y) \leq g(y_0) \qquad \forall y \in B. \tag{2.60}$$

Fix $y \in B$. For $n = 1, 2, \ldots$, let $y_n := (1 - 1/n)y_0 + (1/n)y$. Denote by $x_n := x(y_n)$, a minimizer of $x \mapsto K(x, y_n)$, i.e., $K(x_n, y_n) = \min_{x\in A} K(x, y_n) = g(y_n)$. Since $y \mapsto K(x,y)$ is concave, by (2.58)

$$\Big(1 - \frac{1}{n}\Big)K(x_n, y_0) + \frac{1}{n}K(x_n, y) \leq K(x_n, y_n) = g(y_n) \leq g(y_0)$$

and, since $g(y_0) = K(x(y_0), y_0) \leq K(x_n, y_0)$, we conclude that

$$K(x(y_n), y) \leq g(y_0) \qquad \forall n,\ \forall y \in B. \tag{2.61}$$

Since A is compact, there exist $\widetilde{x}(y) \in A$ and a subsequence $\{k_n\}$ such that $x_{k_n} \to \widetilde{x}(y)$ and $K(\widetilde{x}(y), y) = \min_n K(x(y_n), y)$, and, in turn,

$$K(\widetilde{x}(y), y) \leq \liminf_{n\to\infty} K(x_{k_n}, y) \leq g(y_0) \qquad \forall y \in B.$$

Step 3. Let us prove that

$$K(\widetilde{x}(y), y_0) = g(y_0) \qquad \forall y \in B. \tag{2.62}$$

We need to prove that $K(\widetilde{x}(y), y_0) \leq g(y_0)$, as the opposite inequality is trivial. With the notations of Step 2, from the concavity of $y \mapsto K(x,y)$

$$\Big(1 - \frac{1}{n}\Big)K(x_n, y_0) + \frac{1}{n}K(x_n, y) \leq K(x_n, y_n) = g(y_n) \leq g(y_0).$$

Consequently,

$$K(\widetilde{x}(y), y_0) \leq \liminf_{n\to\infty} K(x(y_n), y_0) \leq g(y_0).$$

Step 4. Let us prove the claim when $x \to K(x,y)$ is strictly convex. By Step 3, $\widetilde{x}(y)$ is a minimizer of the map $x \to K(x, y_0)$ as x_0 is. Since $x \mapsto K(x, y_0)$ is strictly convex, the minimizer is *unique*, thus concluding $\widetilde{x}(y) = x_0\ \forall y \in B$. The claim then follows from (2.59), (2.60) and (2.62).

Step 5. In case $x \to K(x,y)$ is merely convex, we introduce for every $\epsilon > 0$ the perturbed Lagrangian K_ϵ

$$K_\epsilon(x,y) := K(x,y) + \epsilon||x||, \qquad x \in A,\ y \in B$$

which is strictly convex. From Step 4 we infer the existence of a saddle point (x_ϵ, y_ϵ) for K_ϵ, i.e.,

$$K(x_\epsilon, y) + \epsilon||x_e|| \leq K(x_\epsilon, y_\epsilon) + \epsilon||x_\epsilon|| \leq K(x, y_\epsilon) + \epsilon||x|| \qquad \forall x \in A,\ y \in B.$$

Passing to subsequences, $x_\epsilon \to x_0 \in A$, $y_\epsilon \to y_0 \in B$, and from the above

$$K(x_0, y) \leq K(x, y_0) \qquad \forall x \in A,\ y \in B,$$

that is, (x_0, y_0) is a saddle point for K. □

Proof of Theorem 2.91. For $k = 1, 2, \ldots$, let $A_k := \{x \,|\, |x| \le k\}$, $B_k := \{y \,|\, |y| \le k\}$. By Proposition 2.92, $K(x, y)$ has a saddle point (x_k, y_k) on $A_k \times B_k$, i.e.,

$$K(x_k, y) \le K(x_k, y_k) \le K(x, y_k) \qquad \forall x \in A_k,\ y \in B_k. \tag{2.63}$$

Choosing $x = \overline{x}$, $y := \overline{y}$ in (2.63) we then have

$$K(x_k, \overline{y}) \le K(x_k, y_k) \le K(\overline{x}, y_k) \qquad \forall k$$

which implies trivially that $\{x_k\}$ and $\{y_k\}$ are both bounded. Therefore, passing eventually to subsequences, $x_k \to x_0$, $y_k \to y_0$, and from (2.63)

$$K(x_0, y) \le K(x_0, y_0) \le K(x, y_0) \qquad \forall x \in A_k,\ y \in B_k.$$

Since k is arbitrary, (x_0, y_0) is a saddle point for K on the whole $\mathbb{R}^n \times \mathbb{R}^m$. □

b. Optimal mixed strategies

An interesting case in which the previous theory applies is the case of finite strategies. We assume that the game (with zero sum) is played many times and that players P and Q choose their strategies, which are finitely many, on the basis of the frequency of success or of the probability: If the strategies of P and Q are respectively $\{E_1, E_2, \ldots, E_m\}$ and $\{F_1, F_2, \ldots, F_n\}$ and if $U(E_i, F_j)$ is the utility function resulting from the choices of E_i by P and F_j by Q, we assume that P chooses E_i with probability x_i and Q chooses F_j with probability y_j. Define now

$$A := \{x \in \mathbb{R}^m \,\Big|\, 0 \le x_i \le 1,\ \sum_{i=1}^{m} x_i = 1\},$$

$$B := \{y \in \mathbb{R}^n \,\Big|\, 0 \le y_j \le 1,\ \sum_{j=1}^{n} y_j = 1\};$$

then the payoff functions of the two players are given by

$$U_P(x, y) = -U_Q(x, y) = K(x, y) := \sum_{i,j} U(E_i, E_j) x_i y_j. \tag{2.64}$$

Since $K(x, y)$ is a homogeneous polynomial of degree 2, von Neumann's theorem applies to get the following result.

2.93 Theorem. *In a game with zero sum, there exist optimal mixed strategies $(\overline{x}, \overline{y})$. They are given by saddle points of the expected payoff function* (2.64)*, and for them we have*

$$\max_{x \in A} \min_{y \in B} K(x, y) = K(\overline{x}, \overline{y}) = \min_{y \in B} \max_{x \in A} K(x, y).$$

2.94 A linear programming approach. Theorem 2.93, although ensuring the existence of optimal mixed strategies, gives no method to find them, which, of course, is quite important. Notice that A and B are compact and convex sets with the vectors of the standard basis $e_1, e_2, \ldots, e_m$

of $\mathbb{R}^m$ and $e_1, e_2, \dots, e_n$ of $\mathbb{R}^n$ as extreme points, respectively. Since $x \to K(x,y)$ and $y \to K(x,y)$ are linear, they attain their maximum and minimum at extreme points, hence

$$f(x) := \min_{y \in B} K(x,y) = \min_{1 \le j \le n} K(x, e_i),$$
$$g(y) := \max_{x \in A} K(x,y) = \max_{1 \le i \le m} K(e_i, y).$$

Notice that $f(x)$ and $g(y)$ are affine maps. Set $\mathbf{U} := (U_{ij})$, $U_{ij} := U(E_i, E_j)$; then maximizing f in A is equivalent to maximize a real number z subject to the constraints $z \le K(z, e_i)\ \forall i$ and $x \in A$, that is, to solve

$$\begin{cases} F(x,z) := z \to \max, \\ z \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \le \mathbf{U}x, \\ \sum_{i=1}^m x_i = 1, \\ x \ge 0. \end{cases}$$

Similarly, minimizing g in B is equivalent to solving

$$\begin{cases} G(y,w) := w \to \min, \\ w \ge \mathbf{U}^T y \le 0, \\ \sum_{i=1}^n y_i = 1, \\ y \ge 0. \end{cases}$$

These are two problems of linear programming, one the dual of the other, and they can be solved with the methods of linear programming, see Section 2.4.7.

c. Nash equilibria

2.95 Example (The prisoner dilemma). Two prisoners have to serve a one-year prison sentence for a minor crime, but they are suspected of a major crime. Each of them receives separately the following proposal: If he accuses the other of the major crime, he will not have to serve the one-year sentence for the minor crime and, if the other does not accuse him of the major crime (in which case he will have to serve the relative 5-year prison sentence), he will be freed. The possible strategies are two: (a) accusing the other and (b) not accusing the other; the corresponding utility functions for the two prisoners P and Q (in years of prison to serve, with negative sign, so that we have to maximize) are

$$U_P(a,a) = -5, \quad U_P(a,n) = 0, \quad U_P(n,a) = -6, \quad U_P(n,n) = -1,$$
$$U_Q(a,a) = -5, \quad U_Q(a,n) = -6, \quad U_Q(n,a) = 0, \quad U_Q(n,n) = -1.$$

We see at once that the strategy of accusing each other gives the worst result with respect to the choice of not accusing the other. Nevertheless, the choice of accusing the

EQUILIBRIUM POINTS IN N-PERSON GAMES

By John F. Nash, Jr.*

Princeton University

Communicated by S. Lefschetz, November 16, 1949

One may define a concept of an n-person game in which each player has a finite set of pure strategies and in which a definite set of payments to the n players corresponds to each n-tuple of pure strategies, one strategy being taken for each player. For mixed strategies, which are probability distributions over the pure strategies, the pay-off functions are the expectations of the players, thus becoming polylinear forms in the probabilities with which the various players play their various pure strategies.

Any n-tuple of strategies, one for each player, may be regarded as a point in the product space obtained by multiplying the n strategy spaces of the players. One such n-tuple counters another if the the strategy of each player in the countering n-tuple yields the highest obtainable expectation for its player against the $n-1$ strategies of the other players in the countered n-tuple. A self-countering n-tuple is called an equilibrium point.

The correspondence of each n-tuple with its set of countering n-tuples gives a one-to-many mapping of the product space into itself. From the definition of countering we see that the set of countering points of a point is convex. By using the continuity of the pay-off functions we see that the graph of the mapping is closed. The closedness is equivalent to saying: if $P_1, P_2, \ldots$ and $Q_1, Q_2, \ldots, Q_n, \ldots$ are sequences of points in the product space where $Q_n \to Q$, $P_n \to P$ and Q_n counters P_n, then Q counters P.

Since the graph is closed and since the image of each point under the mapping is convex, we infer from Kakutani's theorem[1] that the mapping has a fixed point (i.e., point contained in its image). Hence there is an equilibrium point.

In the two-person zero-sum case the "main theorem"[2] and the existence of an equilibrium point are equivalent. In this case any two equilibrium points lead to the same expectations for the players, but this need not occur in general.

* The author is indebted to Dr. David Gale for suggesting the use of Kakutani's theorem to simplify the proof and to the A. E. C. for financial support.

[1] Kakutani, S., *Duke Math. J.*, **8**, 457–459 (1941).

[2] Von Neumann, J., and Morgenstern, O., *The Theory of Games and Economic Behaviour*, Chap. 3, Princeton University Press, Princeton, 1947.

Annals of Mathematics
Vol. 54, No. 2, September, 1951

NON-COOPERATIVE GAMES

John Nash

(Received October 11, 1950)

Introduction

Von Neumann and Morgenstern have developed a very fruitful theory of two-person zero-sum games in their book *Theory of Games and Economic Behavior*. This book also contains a theory of n-person games of a type which we would call cooperative. This theory is based on an analysis of the interrelationships of the various coalitions which can be formed by the players of the game.

Our theory, in contradistinction, is based on the *absence* of coalitions in that it is assumed that each participant acts independently, without collaboration or communication with any of the others.

The notion of an *equilibrium point* is the basic ingredient in our theory. This notion yields a generalization of the concept of the solution of a two-person zero-sum game. It turns out that the set of equilibrium points of a two-person zero-sum game is simply the set of all pairs of opposing "good strategies."

In the immediately following sections we shall define equilibrium points and prove that a finite non-cooperative game always has at least one equilibrium point. We shall also introduce the notions of solvability and strong solvability of a non-cooperative game and prove a theorem on the geometrical structure of the set of equilibrium points of a solvable game.

As an example of the application of our theory we include a solution of a simplified three person poker game.

Formal Definitions and Terminology

In this section we define the basic concepts of this paper and set up standard terminology and notation. Important definitions will be preceded by a subtitle indicating the concept defined. The non-cooperative idea will be implicit, rather than explicit, below.

Finite Game:

For us an n-*person game* will be a set of n *players*, or *positions*, each with an associated finite set of *pure strategies*; and corresponding to each player, i, a *payoff function*, p_i, which maps the set of all n-tuples of pure strategies into the real numbers. When we use the term n-*tuple* we shall always mean a set of n items, with each item associated with a different player.

Mixed Strategy, s_i:

A *mixed strategy* of player i will be a collection of non-negative numbers which have unit sum and are in one to one correspondence with his pure strategies. We write $s_i = \sum_\alpha c_{i\alpha} \pi_{i\alpha}$ with $c_{i\alpha} \geqq 0$ and $\sum_\alpha c_{i\alpha} = 1$ to represent such a mixed strategy, where the $\pi_{i\alpha}$'s are the pure strategies of player i. We regard the s_i's as points in a simplex whose vertices are the $\pi_{i\alpha}$'s. This simplex may be re-

286

Figure 2.14. The initial pages of two papers by John Nash (1928–).

other brings the advantage of serving one year less in any case: The strategy of not accusing, which, from a *cooperative* point of view is the best, is not the *individual* point of view (even in the presence of a reciprocal agreement; in fact, neither of the two may ensure that the other will not accuse him). This paradox arises quite frequently.

The idea that individual rationality, typical of *noncooperative games* (in which there is no possibility of agreement among the players), precedes collective rationality is at the basis of the notion of the *Nash equilibrium*.

2.96 Definition. *Let A and B be two sets and let f and g be two maps from $A \times B$ into $\mathbb{R}$. The couple of points $(x_0, y_0) \in A \times B$ is called a* Nash point *for f and g if for all $(x, y) \in A \times B$ we have*

$$f(x_0, y_0) \geq f(x, y_0), \qquad g(x_0, y_0) \geq g(x_0, y).$$

In the prisoner's dilemma, the unique Nash point is the strategy of the reciprocal accusation. In a game with zero sum, i.e., $U_P(x, y) = -U_Q(x, y) =: K(x, y)$, clearly (x_0, y_0) is a Nash point if and only if (x_0, y_0) is a *saddle point* for K.

2.97 Theorem (of Nash for two players). *Let A and B be two non-empty, convex and compact sets. Let $f, g : A \times B \to \mathbb{R}$ be two continuous functions such that $x \to f(x, y)$ is concave for all $y \in B$ and $y \to g(x, y)$ is concave for all $x \in A$. Then there exists a Nash equilibrium point for f and g.*

Proof. Introduce the function $F : (A \times B) \times (A \times B) \to \mathbb{R}$ defined by

$$F(p, q) = f(p_1, q_2) + g(q_1, p_2), \qquad \forall p = (p_1, p_2),\ q = (q_1, q_2) \in A \times B.$$

Clearly, F is continuous and concave in p for every chosen q. We claim that there is $q_0 \in A \times B$ such that

$$\max_{p \in A \times B} F(p, q_0) = F(p_0, q_0). \tag{2.65}$$

Before proving the claim, let us complete the proof of the theorem on the basis of (2.65). If we set $(x_0, y_0) := q_0$, we have

$$f(x, y_0) + g(x_0, y) \le f(x_0, y_0) + g(x_0, y_0) \qquad \forall (x, y) \in A \times B.$$

Choosing $x = x_0$, we infer $g(x_0, y) \le g(x_0, y_0)\ \forall y \in B$, while, by choosing $y = y_0$, we find $f(x, y_0) \le f(x_0, y_0)\ \forall x \in A$, hence (x_0, y_0) is a Nash point.

Let us prove (2.65). Since the inequality $\ge$ is trivial, for all $q_0 \in A \times B$, we need to prove only the opposite inequality. By contradiction, suppose that $\forall q \in A \times B$ there is $p \in A \times B$ such that $F(p, q) > F(q, q)$ and, then, set

$$G_q := \Big\{ q \in A \times B \,\Big|\, F(p, q) > F(q, q) \Big\}, \qquad p \in A \times B.$$

The family $\{G_p\}_{p \in A \times B}$ is an open covering of $A \times B$; consequently, there are finitely many points $p_1, p_2, \ldots, p_k \in A \times B$ such that $A \times B \subset \cup_{i=1}^{k} G_{p_i}$. Set

$$\varphi_i(q) := \max \Big(F(p_i, q) - F(q, q), 0 \Big), \qquad q \in A \times B,\ i = 1, \ldots, k.$$

The functions $\{\varphi_i\}$ are continuous, nonnegative and, for every q, at least one of them does not vanish at q; we then set

$$\psi_i(q) := \frac{\varphi_i(q)}{\sum_{j=1}^{k} \varphi_j(q)}$$

and define the new map $\psi : A \times B \to A \times B$ by

$$\psi(q) := \sum_{i=1}^{k} \psi_i(q) p_i.$$

The map ψ maps the convex and compact set $A \times B$ into itself, consequently, it has a fixed point $q' \in A \times B$, $q' = \sum_i \psi(q') p_i$. F being concave,

$$F(q', q') = F\Big(\sum_i \psi_i(q') p_i, q' \Big) \ge \sum_{i=1}^{k} \psi_i(q') F(q_i, q').$$

On the other hand, $F(p_i, q') > F(q', q')$ if $\psi_i(q') > 0$, hence

$$F(q', q') \ge \sum_{i=1}^{k} \psi(q') F(q_i, q') > \sum_{i=1}^{k} \psi(q') F(q', q') = F(q', q'),$$

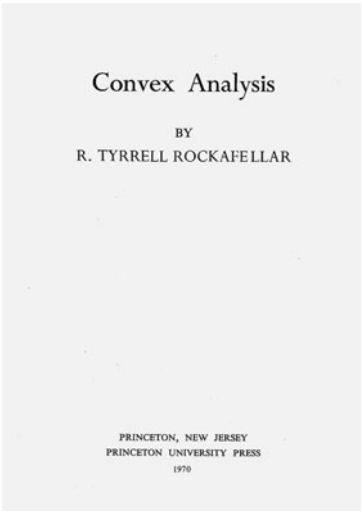
which is a contradiction. □

d. Convex duality

Let $f, \varphi^1, \ldots, \varphi^m : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be convex functions defined on a convex open set Ω. We assume for simplicity that f and $\varphi := (\varphi^1, \varphi^2, \ldots, \varphi^m)$ are differentiable. Let

$$\mathcal{F} := \Big\{ x \in \mathbb{R}^n \,\Big|\, \varphi^j(x) \le 0,\ \forall j = 1, \ldots, m \Big\}.$$

The *primal problem* of convex optimization is the *minimum problem*



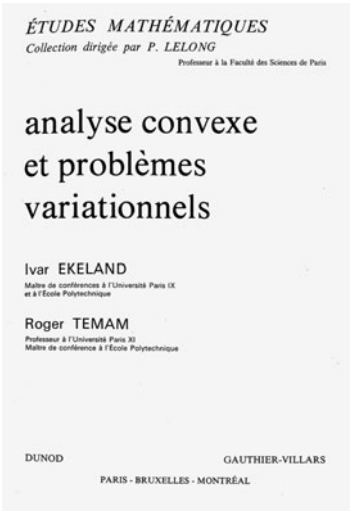


Figure 2.15. Two classical monographs on convexity.

$$\textit{Assuming } \mathcal{F} \neq \emptyset, \textit{ minimize } f \textit{ in } \mathcal{F}. \tag{2.66}$$

The associated *Lagrangian* $\mathcal{L} : \Omega \times \mathbb{R}^m_+$ to (2.66), defined by

$$\mathcal{L}(x, \lambda) := f(x) + \lambda \bullet \varphi(x)\,, \qquad x \in \Omega,\ \lambda \geq 0, \tag{2.67}$$

is convex in x for any fixed λ and linear in λ for every fixed x. Therefore, it is not surprising that the Kuhn–Tucker conditions (2.41)

$$\begin{cases} \mathbf{D}f(x^0) + \lambda^0 \bullet \mathbf{D}\varphi(x_0) = 0, \\ \lambda^0 \geq 0,\ x_0 \in \mathcal{F}, \\ \lambda^0 \bullet \varphi(x_0) = 0 \end{cases} \tag{2.68}$$

are also sufficient to characterize minimum points for f on $\mathcal{F}$. Actually, the Kuhn–Tucker equilibrium conditions (2.41) are strongly related to saddle points for the associated Lagrangian $\mathcal{L}(x, \lambda)$.

2.98 Theorem. *Consider the primal problem* (2.66)*. Then* (x_0, λ^0) *fulfills* (2.68) *if and only if* (x_0, λ^0) *is a saddle point for* $\mathcal{L}(x, \lambda)$ *on* $\Omega \times \mathbb{R}^m_+$*, i.e.,*

$$\mathcal{L}(x_0, \lambda) \leq \mathcal{L}(x_0, \lambda^0) \leq cL(x, \lambda^0)$$

for all $x \in \mathcal{F}$ *and* $\lambda \in \mathbb{R}^m$*,* $\lambda \geq 0$*. In particular, if the Kuhn–Tucker equilibrium conditions are satisfied at* $(x_0, \lambda^0) \in \mathcal{F} \times \mathbb{R}^m_+$*, then* x_0 *is a minimizer for* f *on* $\mathcal{F}$*.*

Proof. From the convexity of $x \mapsto \mathcal{L}(x, \lambda^0)$ and (2.41) we infer

$$\mathcal{L}(x, \lambda^0) \geq \mathcal{L}(x_0, \lambda_0) + \sum_{i=1}^{n} \Big(\nabla f(x_0) + scd\lambda^0 \mathbf{D}\varphi(x_0)\Big)^i (x - x_0)^i = \mathcal{L}(x_0, \lambda^0) = f(x_0)$$

for all $x \in \Omega$. In particular,

$$\begin{aligned} f(x) &\geq f(x) + \lambda^0 \bullet \varphi(x^0) = \mathcal{L}(x, \lambda^0) \geq f(x_0), \\ \mathcal{L}(x_0, \lambda^0) &\geq f(x_0) + \lambda \bullet \varphi(x_0) = \mathcal{L}(x, \lambda^0). \end{aligned}$$

Conversely, suppose that (x_0, λ^0) is a saddle point for $\mathcal{L}(x, \lambda)$ on $\Omega \times \mathbb{R}^m_+$, i.e.,

$$f(x_0) + \lambda \bullet \varphi(x_0) \leq f(x_0) + \lambda^0 \bullet \varphi(x_0) \leq f(x) + \lambda^0 \bullet \varphi(x)$$

for every $x \in \Omega$ and $\lambda \geq 0$. From the first inequality we infer

$$\lambda \bullet \varphi(x_0) \geq \lambda^0 \varphi(x_0) \tag{2.69}$$

for any $\lambda \geq 0$. This implies that $\varphi(x_0) \leq 0$ and, in turn, $\lambda_0 \bullet \varphi(x_0) \leq 0$. Using again (2.69) with $\lambda = 0$, we get the opposite inequality, thus concluding that $\lambda_0 \bullet \varphi(x_0) = 0$. Finally, from the first inequality, Fermat's theorem yields

$$\nabla f(x_0) + \lambda^0 \bullet \nabla \varphi(x_0) = 0.$$

□

Let us now introduce the *dual problem* of convex optimization. For $\lambda \in \mathbb{R}^m_+$, set

$$g(\lambda) := \inf_{x \in \mathcal{F}} \mathcal{L}(x, \lambda),$$

where $\mathcal{L}(x, \lambda)$ is the Lagrangian in (2.67).

Since $g(\lambda)$ is the infimum of a family of affine functions, $-g$ is convex and proper on

$$\mathcal{G} := \{\lambda \in \mathbb{R}^m \mid \lambda \geq 0, g(\lambda) > -\infty\}.$$

The dual problem of convex programming is

$$\textit{Assuming } \mathcal{G} \neq \emptyset, \textit{ maximize } g(\lambda) \textit{ on } \mathcal{G} \tag{2.70}$$

or, equivalently,

$$\textit{Assuming } \mathcal{G} \neq \emptyset, \textit{ maximize } g(\lambda) \textit{ on } \{\lambda \in \mathbb{R}^m \mid \lambda \geq 0\}. \tag{2.71}$$

2.99 Theorem. *If $(x_0, \lambda^0) \in \mathcal{F} \times \mathbb{R}^m$ satisfies the Kuhn–Tucker equilibrium conditions* (2.41)*, then x_0 maximizes the primal problem, λ_0 minimizes the dual problem and $f(x_0) = g(\lambda^0) = \mathcal{L}(x_0, \lambda^0)$.*

Proof. By definition, $g(\lambda) = \sup_{x \in \mathcal{F}} \mathcal{L}(x, \lambda)$, and, trivially, $f(x) := \inf_{\lambda \geq 0} \mathcal{L}(x, \lambda)$. Therefore $g(y) \leq f(x)$ for all $x \in \mathcal{F}$ and $\lambda \geq 0$, so that

$$\sup_{\lambda \geq 0} g(\lambda) \leq \inf_{x \in \mathcal{F}} f(x).$$

Since (x_0, λ^0) is a saddle point for $\mathcal{L}$ on $\Omega \times \mathbb{R}_+$, Proposition 2.88 yields the result. □

2.5 A General Approach to Convexity

As we have seen, every closed convex set K is the intersection of all closed half-spaces in which it is contained; in fact, K is the *envelope* of its *supporting hyperplanes*. In other words, a closed convex body is given in a dual way by the supporting hyperplanes. This remark, when applied to closed epigraphs of convex functions, yields a number of interesting correspondences. Here we discuss the so-called *polarity* correspondence.

a. Definitions

It is convenient to allow that convex functions take the value $+\infty$ with the convention $t + (+\infty) = +\infty$ for all $t \in \mathbb{R}$ and $t \cdot (+\infty) = +\infty$ for all $t > 0$. For technical reasons, it is also convenient to allow that convex functions take the value $-\infty$.

2.100 Definition. *$f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \qquad \forall x, y \in \mathbb{R}^n, \quad \forall \lambda \in [0,1]$$

unless $f(x) = -f(y) = \pm\infty$. The effective domain *of f is then defined by*

$$\operatorname{dom}(f) := \Big\{ x \in \mathbb{R}^n \;\Big|\; f(x) < \infty \Big\}.$$

We say that f is proper *if f is nowhere $-\infty$ and $\operatorname{dom}(f) \neq \emptyset$.*

Let $K \subset \mathbb{R}^n$ be a convex set and $f : K \subset \mathbb{R}^n \to \mathbb{R}$ be a convex function. It is readily seen that the function $\overline{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined as

$$\overline{f}(x) = \begin{cases} f(x) & \text{if } x \in K, \\ +\infty & \text{if } x \notin K \end{cases}$$

is convex according to Definition 2.100 with effective domain given by K.

One of the advantages of Definition 2.100 is that convex sets and convex functions are essentially the same object.

From one side, $K \subset \mathbb{R}^n$ is convex if and only if its *indicatrix function*

$$I_K(x) := \begin{cases} 0 & \text{if } x \in K, \\ +\infty & \text{if } x \notin K \end{cases} \tag{2.72}$$

is convex in the sense of Definition 2.100. On the other hand, $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if its *epigraph*, defined as usual by

$$\operatorname{Epi}(f) := \Big\{ (x,t) \in \mathbb{R}^n \times \mathbb{R} \;\Big|\; x \in \mathbb{R}^n,\ t \in \mathbb{R},\ t \ge f(x) \Big\}$$

is a convex set in $\mathbb{R}^n \times \mathbb{R}$.

Observe that the constrained minimization problem

$$\begin{cases} f(x) \to \min, \\ x \in K, \end{cases}$$

where f is a convex function and K is a convex set, transforms into the unconstrained minimization problem for the convex function

$$f(x) + I_K(x), \qquad x \in \mathbb{R}^n$$

which is defined by adding to f the indicatrix I_K of K as *penalty function.*

One easily verifies that

(i) f is convex if and only if its epigraph is convex,
(ii) the effective domain of a convex function is convex,
(iii) if f is convex, then $\mathrm{dom}(f) = \pi(\mathrm{Epi}(f))$ where $\pi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is the linear projection on the first factor.

We have also proved, compare Theorem 2.35, that *every proper convex function is locally Lipschitz in the interior of its effective domain.* However, in general, a convex function need not be continuous or semicontinuous at the boundary of its effective domain, as, for instance, for the functions f defined as $f(x) = 0$ if $x \in]-1, 1[$, $f(-1) = f(1) = 1$ and $f(x) = +\infty$ if $x \notin [0, 1]$.

b. Lower semicontinuous functions and closed epigraphs

We recall that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *lower semicontinuous*, see [GM3], in short *l.s.c.*, if $f(x) \leq \liminf_{y \to x} f(y)$. If $f(x) \in \mathbb{R}$, this means the following:

(i) For all $\epsilon > 0$ there is $\delta > 0$ such that for all $y \in B(x, \delta) \setminus \{x\}$ we have $f(x) - \epsilon \leq f(y)$.
(ii) There is a sequence $\{x_k\}$ with values in $\mathbb{R}^n \setminus \{x\}$ that converges to x such that $f(x_k) \to f(x)$.

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. We already know that f is l.s.c. if and only if for every $t \in \mathbb{R}$ the sublevel set $\{x \,|\, f(x) \leq t\}$ is closed. Moreover, the following holds.

2.101 Proposition. *The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed if and only if f is lower semicontinuous.*

Proof. Let f be l.s.c. and $\{(x_k, t_k)\} \subset \mathrm{Epi}(f)$ a sequence that converges to (x, t). Then $x_\kappa \to x$, $t_k \to t$ and $f(x_k) \leq t_k$. It follows that $f(x) \leq \liminf_{k\to\infty} f(x_k) \leq \liminf_{k\to\infty} t_k = t$, i.e., $(x, t) \in \mathrm{Epi}(f)$.

Conversely, suppose that $\mathrm{Epi}(f)$ is closed. Consider a sequence $\{x_k\}$ with $x_k \to x$ and let $L := \liminf_{k\to\infty} f(x_k)$. If $L = +\infty$, then $f(x) \leq L$. If $L < +\infty$, we find a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $f(x_{n_k}) \to L$. Since $(x_{n_k}, f(x_{n_k})) \in \mathrm{Epi}(f)$ and $L < +\infty$, we infer that $(x, L) \in \mathrm{Epi}(f)$, i.e., $f(x) \leq L = \liminf_{k\to\infty} f(x_k)$. Since the sequence $\{x_k\}$ is arbitrary, we finally conclude that $f(x) \leq \liminf_{y\to x} f(y)$. □

Finally, let us observe that if $f_\alpha : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\alpha \in \mathcal{A}$, is a family of l.s.c. functions, then

$$f(x) := \sup\Big\{ f_\alpha(x) \,\Big|\, \alpha \in \mathcal{A} \Big\}, \qquad x \in \mathbb{R}^n,$$

is lower semicontinuous.

2.102 Definition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. The* closure *of f or its* lower semicontinuous regularization*, in short its* l.s.c. regularization*, is the function*

$$\Gamma f(x) := \sup\Big\{ g(x) \,\Big|\, g : \mathbb{R}^n \to \overline{\mathbb{R}},\ g \text{ is l.s.c.},\ g(y) \le f(y)\ \forall y \Big\}.$$

Clearly, $\Gamma f(x) \le f(x)$ for every x, and, as the pointwise supremum of a family of l.s.c. functions, Γf is lower semicontinuous. Therefore, it is the *greatest lower semicontinuous minorant* of f.

2.103 Proposition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Then* $\mathrm{Epi}(\Gamma f) = \mathrm{cl}(\mathrm{Epi}(f))$ *and* $\Gamma f(x) = \liminf_{y \to x} f(y)$ *for every $x \in \mathbb{R}^n$.*

Consequently, $\Gamma f(x) = f(x)$ if and only if f is l.s.c. at x.

Proof. (i) First, let us prove that $\mathrm{cl}(\mathrm{Epi}(f))$ is the epigraph of a function $g \le f$, by proving that if $(x,t) \in \mathrm{cl}(\mathrm{Epi}(f))$, then for all $s > t$ we have $(x,s) \in \mathrm{cl}(\mathrm{Epi}(f))$. If $(x_k, t_k) \in \mathrm{Epi}(f)$ converges to (x,t) and $s > t$, for some large k we have $t_k < s$, hence $f(x_k) \le t_k < s$. It follows that definitively $(x_k, s) \in \mathrm{Epi}(f)$, hence $(x,s) \in \mathrm{cl}(\mathrm{Epi}(f))$.

By Proposition 2.101, g is l.s.c. and Γf is closed; therefore, we have $g \le \Gamma f$ and

$$\mathrm{Epi}(\Gamma f) \subset \mathrm{Epi}(g) = \mathrm{cl}(\mathrm{Epi}(f)) \subset \mathrm{Epi}(\Gamma f).$$

(ii) Let $x \in \mathbb{R}^n$. If $\Gamma f(x) = +\infty$, $\Gamma f = +\infty$ in a neighborhood of x, hence $\liminf_{y \to x} f(y) = +\infty$, too.

If $\Gamma f(x) < +\infty$, then for any $t \ge \mathbf{f}(x)$, $(x,t) \in \mathrm{Epi}(\Gamma f)$. (i) yields a sequence $\{(x_k, t_k)\} \subset \mathrm{Epi}(f)$ such that $x_k \to x$ and $y_k \to t$. Therefore

$$\liminf_{k \to \infty} f(x_k) \le \liminf_{k \to \infty} t_k = t,$$

hence

$$\liminf_{k \to \infty} f(x_k) \le \Gamma f(x).$$

On the other hand, since Γf is ls.c. and $\Gamma f \le f$,

$$\Gamma f(x) \le \liminf_{y \to x} \Gamma f(y) \le \liminf_{y \to x} f(y),$$

thus concluding that $\Gamma f(x) = \liminf_{y \to x} f(y)$. It is then easy to check that $f(x) = \Gamma f(x)$ if and only if f is l.s.c. at x. □

Since closed convex sets can be represented as intersections of their supporting half-spaces, of particular relevance are the convex functions with closed epigraphs. According to the above, we have the following.

2.104 Corollary. *$f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and l.s.c. if and only if its epigraph is convex and closed.*

The l.s.c. regularization Γf of a convex function f is a convex and l.s.c. function.

According to the above, $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is l.s.c. and convex if and only if its epigraph $\mathrm{Epi}(f)$ is closed and convex. In particular, $\mathrm{Epi}(f)$ is the intersection of all its supporting half-spaces. The next theorem states that $\mathrm{Epi}(f)$ is actually the intersection of all half-spaces associated to graphs of linear affine functions, i.e., to hyperplanes that do not contain vertical vectors.

We first state a proposition that contains the relevant property.

2.105 Proposition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and l.s.c. and let $\overline{x} \in \mathbb{R}^n$ be such that $f(\overline{x}) > -\infty$. Then the following hold:*

(i) *For every $\overline{y} < f(\overline{x})$ there exists an affine map $\ell : \mathbb{R}^n \to \mathbb{R}$ such that $f(x) > \ell(x)$ for every $x \in \mathbb{R}^n$ and $\overline{y} < \ell(\overline{x})$.*

(ii) *If $\overline{x} \in \mathrm{int}(\mathrm{dom}(f))$, then there exists an affine map $\ell : \mathbb{R}^n \to \mathbb{R}$ such that $f(x) > \ell(x)$ for every $x \in \mathbb{R}^n$ and $\ell(\overline{x}) = f(\overline{x})$.*

Proof. Since f is lower semicontinuous at $\overline{x}$, there exist $\epsilon > 0$ and $\delta > 0$ such that $\overline{y} \le f(x) - \epsilon \ \forall x \in B(\overline{x}, \delta)$, in particular, $(\overline{x}, \overline{y}) \notin \mathrm{cl}(\mathrm{Epi}(f))$. Therefore, there exists a hyperplane $\mathcal{P} \subset \mathbb{R}^{n+1}$ that strongly separates $\mathrm{Epi}(f)$ from $(\overline{x}, \overline{y})$, i.e., there are a linear map $m : \mathbb{R}^n \to \mathbb{R}$ and numbers $\alpha, \beta \in \mathbb{R}$ such that

$$\mathcal{P} := \Big\{ (x, y) \,\Big|\, m(x) + \alpha y = \beta \Big\} \tag{2.73}$$

with

$$m(x) + \alpha y > \beta \ \ \forall (x, y) \in \mathrm{Epi}(f) \qquad \text{and} \qquad m(\overline{x}) + \alpha \overline{y} < \beta. \tag{2.74}$$

Since y may be chosen arbitrarily large in the first inequality, we also have $\alpha \ge 0$. We now distinguish four cases.

(i) If $f(\overline{x}) < +\infty$, then $\alpha \neq 0$ since, otherwise, choosing $(\overline{x}, y)$ with $y > f(\overline{x})$ in (2.74), we get $m(\overline{x}) > \beta$ and $m(\overline{x}) < \beta$, a contradiction. By choosing ℓ as the linear affine map $\ell(x) := (\beta - m(x))/\alpha$, from the first of (2.74) with $y = f(x)$ it follows $\ell(x) < f(x)$ for all x, while from the second we get $\overline{y} \le \ell(\overline{x})$.

(ii) If $f(\overline{x}) = +\infty$ and the function takes value $+\infty$ everywhere, the claim is trivial.

(iii) If $f(\overline{x}) = +\infty$ and $\alpha > 0$ in (2.74), then one chooses ℓ as the linear affine map $\ell(x) := (\beta - m(x))/\alpha$, as in (i).

(iv) Let us consider the remaining case where $f(\overline{x}) = +\infty$. There exists x_0 such that $f(x_0) \in \mathbb{R}$ and $\alpha = 0$ in (2.74). By applying (i) at x_0, we find an affine linear map ϕ such that

$$f(x) \ge \phi(x) \qquad \forall x \in \mathbb{R}^n.$$

For all $c > 0$ the function $\ell(x) := \phi(x) + c(\beta - m(x))$ is then a linear affine minorant of $f(x)$ and, by choosing c sufficiently large, we can make $\ell(\overline{x}) = \phi(\overline{x}) + c(\beta - m(\overline{x})) > \overline{y}$. This concludes the proof of the first claim.

Let us now prove the last claim. Since $\overline{x} \in \mathrm{int}(\mathrm{dom}(f))$ and $f(\overline{x}) > -\infty$, a support hyperplane $\mathcal{P}'$ of $\mathrm{Epi}(f)$ at $(\overline{x}, f(\overline{x}))$ does not contain vertical vectors: otherwise none of the two subspaces associated to $\mathcal{P}'$ could contain $\mathrm{Epi}(f)$. Hence $\mathcal{P}' := \{(x, y) \,|\, m(x) + \alpha y = \beta\}$ for some linear map m and numbers $\alpha, \beta \in \mathbb{R}$ with

$$m(x) + \alpha y \ge \beta \ \ \forall (x, y) \in \mathrm{Epi}(f), \qquad m(\overline{x}) + \alpha f(\overline{x}) = \beta,$$

and $\alpha > 0$. If $\ell(x) := -m(x)/\alpha$, we see at once that

$$f(x) \ge f(\overline{x}) + \ell(x - \overline{x}) \qquad \forall x \in \mathbb{R}^n.$$

□

2.106 Remark. The previous proof yields the existence of a nontrivial lower affine minorant for f which is arbitrarily close to $f(\overline{x})$ at $\overline{x}$ when f is l.s.c. at $\overline{x} \in \mathbb{R}^n$, $f(\overline{x}) > -\infty$ and one the following conditions hold:

- $f(\overline{x}) \in \mathbb{R}$,
- $f = +\infty$ everywhere,
- $f(\overline{x}) = +\infty$ and there exists a further point $x_0 \in \mathbb{R}^n$ such that $f(x_0) \in \mathbb{R}$ and f is l.s.c. at x_0.

Notice also that if f is convex, then $f(x) > -\infty$ and $x \in \operatorname{int}(\operatorname{dom}(f))$ if and only if f is continuous at x, see Theorem 2.35.

2.107 Corollary. *If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and l.s.c. and $f(\overline{x}) > -\infty$ at some point $\overline{x}$, then $f > -\infty$ everywhere.*

2.108 Definition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. Its* linear l.s.c. envelope *$\Gamma\mathrm{L}\, f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined by*

$$\Gamma\mathrm{L}\, f(x) := \sup\Big\{ \ell(x) \;\Big|\; \ell : \mathbb{R}^n \to \mathbb{R},\ \ell \text{ affine}, \ell \le f \Big\}. \tag{2.75}$$

and, of course, $\Gamma\mathrm{L}\, f(x) = -\infty\ \forall x$ if no affine linear map ℓ below f exists.

2.109 Theorem. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$.*

(i) $\Gamma\mathrm{L}$ *is convex and l.s.c.*
(ii) *f is convex and l.s.c. if and only if $f(x) = \Gamma\mathrm{L}\, f(x)\ \forall x \in \mathbb{R}^n$.*
(iii) *Assume f is convex. If at some point $x \in \mathbb{R}^n$ we have $f(x) < +\infty$, then $f(x) = \Gamma\mathrm{L}\, f(x)$ if and only if f is l.s.c. at x.*
(iv) *If $\overline{x}$ is an interior point of the effective domain of f and $f(\overline{x}) > -\infty$, then the supremum in (2.75) is a maximum, i.e., there exists $\xi \in \mathbb{R}^n$ such that*

$$f(y) \ge f(\overline{x}) + \xi \bullet (y - \overline{x}) \qquad \forall y.$$

Proof. Since the supremum of a family of convex and l.s.c. functions is convex and l.s.c., (2.75) implies that $\Gamma\mathrm{L}\, f(x)$ is convex and l.s.c.. If $\Gamma\mathrm{L}\, f(x) = -\infty$ for all x, then, trivially, $\Gamma\mathrm{L}$ is convex and l.s.c.. This proves (i), (ii) and (iii) are trivial if f is identically $-\infty$, and easily follow from the above and (i) of Proposition 2.105, taking also into account Remark 2.106. Finally, (iv) rephrases (ii) of Proposition 2.105. □

The following observation is sometimes useful.

2.110 Proposition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and l.s.c. and let $r(t) = (1-t)x + t\overline{x}$, $t \in [0,1]$, be the segment joining x to $\overline{x}$. Suppose $f(\overline{x}) < +\infty$. Then*

$$f(x) = \lim_{t \to 0^+} f(r(t)).$$

Proof. Since $f(\overline{x}) < +\infty$,

$$f(x) \le \liminf_{t \to 0^+} f(t\overline{x} + (1-t)x) \le \lim_{t \to 0} \Big(t\, f(\overline{x}) + (1-t) f(x) \Big) = f(x).$$

□

c. The Fenchel transform

2.111 Definition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. The* polar *or* Fenchel transform *of f is the function $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by*

$$f^*(\xi) := \sup_{x\in\mathbb{R}^n} \Big(\xi \bullet x - f(x)\Big) = - \inf_{x\in\mathbb{R}^n} \Big(f(x) - \xi \bullet x\Big). \tag{2.76}$$

As we will see, the Fenchel transform rules the entire mechanism of convex duality.

2.112 Proposition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function and $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ its polar. Then we have the following:*

(i) $f(x) \geq \xi \bullet x - \eta \ \forall x$ *if and only if* $f^*(\xi) \leq \eta$;
(ii) $f^*(\xi) = -\infty$ *for some* ξ *if and only if* $f(x) = +\infty$ *for all* x;
(iii) *if* $f \leq g$, *then* $g^* \leq f^*$;
(iv) $f^*(0) = -\inf_{x\in\mathbb{R}^n} f(x)$;
(v) *the* Fenchel inequality *holds*

$$\xi \bullet x \leq f^*(\xi) + f(x) \qquad \forall x \in \mathbb{R}^n \ \forall \xi \in \mathbb{R}^n,$$

with equality at $(\overline{x}, \overline{\xi})$ if and only if $f(x) \geq f(\overline{x}) + \overline{\xi} \bullet (x - \overline{x})$.
(vi) f^* *is l.s.c. and convex.*

Proof. All of the claims follow immediately from the definition of f^*. □

The polar transform generalizes Legendre's transform.

2.113 Proposition. *Let Ω be an open set in $\mathbb{R}^n$, let $f : \Omega \to \mathbb{R}$ be a convex function of class C^2 with positive definite Hessian matrix and let $\Gamma\mathrm{L}\, f$ be the l.s.c linear envelope of f. Then*

$$\mathcal{L}_f(\xi) = (\Gamma\mathrm{L}\, f)^*(\xi) \qquad \forall \xi \in \mathbf{D}f(\Omega).$$

Proof. According to Theorem 2.109, $f(x) = \Gamma\mathrm{L}\, f(x)$ for all $x \in \Omega$, while Theorem 2.109 yields for all $\xi \in \mathbf{D}f(\Omega)$

$$\mathcal{L}_f(\xi) = \max_{\Omega}(x \bullet \xi - f(x)) \leq \sup_{x\in\mathbb{R}^n} (x \bullet \xi - \Gamma\mathrm{L}\, f(x)) = (\Gamma\mathrm{L}\, f)^*(\xi).$$

On the other hand,

$$(\Gamma\mathrm{L}\, f)^*(\xi) = \sup_{x\in\overline{\Omega}}(x \bullet \xi - \Gamma\mathrm{L}\, f(x)) =: L.$$

Given $\epsilon > 0$, let $x \in \overline{\Omega}$ be such that $L < \overline{x} \bullet \xi - \Gamma\mathrm{L}\, f(\overline{x}) + \epsilon$. There exists $\{x_k\} \subset \Omega$ such that $f(x_k) = \Gamma\mathrm{L}\, f(x_k) \to \Gamma\mathrm{L}\, f(\overline{x})$, hence for $k > \overline{k}$

$$L \leq x_k \bullet \xi - f(x_k) + 2\epsilon \leq \sup_{x\in\Omega}(x \bullet \xi - f(x)) + 2\epsilon.$$

Since $\epsilon > 0$ is arbitrary, $L \leq \sup_{x\in\Omega}(x \bullet \xi - f(x))$ and the proof is completed. □

The polar of a closed convex set is subsumed to the Fenchel transform, too. In fact, if K is a closed convex set, its indicatrix function, see (2.72), is l.s.c. and convex; hence

$$(I_K)^*(\xi) := \sup_{x\in\mathbb{R}^n} \Big(\xi \bullet x - I_K(x) \Big) = \sup_{x\in K} \xi \bullet x \,. \tag{2.77}$$

Therefore,

$$K^* = \Big\{ \xi \;\Big|\; x \bullet \xi \le 1 \;\forall x \in K \Big\} = \Big\{ \xi \;\Big|\; (I_K)^*(\xi) \le 1 \Big\}.$$

2.114 Definition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. Its* bipolar *is defined as the function $f^{**}(x) := (f^*)^*(x) : \mathbb{R}^n \to \mathbb{R}$, i.e.,*

$$f^{**}(x) := \sup\Big\{ \xi \bullet x - f^*(\xi) \;\Big|\; \forall \xi \in \mathbb{R}^n \Big\}.$$

2.115 ¶. Let $\ell(x) := \eta \bullet x + \beta$ be a linear affine map on $\mathbb{R}^n$. Prove that

$$\ell^*(\xi) = \begin{cases} +\infty & \text{if } \xi \ne \eta, \\ -\beta & \text{if } \xi = \eta, \end{cases}$$

and that $(\ell^*)^*(x) = \eta \bullet x + \beta = \ell(x)$.

2.116 Proposition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. Then*

(i) $f^{**} \le f$,
(ii) $f^{**} \le g^{**}$ *if* $f \le g$,
(iii) f^{**} *is the largest l.s.c. convex minorant of* f,

$$f^{**}(x) = \Gamma\mathrm{L}\, f(x) = \sup\Big\{ \ell(x) \;\Big|\; \ell : \mathbb{R}^n \to \mathbb{R},\ \ell \text{ affine},\ \ell \le f \Big\}.$$

Proof. (i) From the definition of f^* we have $\xi \bullet x - f^*(\xi) \le f(x)$, hence $f^{**}(x) = \sup_{\xi\in\mathbb{R}^n} (\xi \bullet x - f^*(\xi)) \le f(x)$.

(ii) if $f \le g$, then $g^* \le f^*$ hence $(f^*)^* \le (g^*)^*$.

(iii) f^{**} is convex and l.s.c. , hence $f^{**} = \Gamma\mathrm{L}\, f^{**}$. On the other hand, every linear affine minorant ℓ of f is also an affine linear minorant for f^{**}, since $\ell = \ell^{**} \le f^{**}$. Therefore $\Gamma\mathrm{L}\, f^{**} = \Gamma\mathrm{L}\, f$. □

The following theorem is an easy consequence of Proposition 2.116.

2.117 Theorem. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Then we have the following:*

(i) *f is convex and l.s.c. if and only if $f = f^{**}$.*
(ii) *Assume that f is convex and $f(x) < +\infty$ at some $x \in \mathbb{R}^n$. Then $f(x) = f^{**}(x)$ if and only if f is l.s.c. at x.*
(iii) *f^* is an involution on the class of proper, convex and l.s.c. functions.*

Proof. Since $f^{**}(x) = \Gamma\mathrm{L}\, f(x)$, (i) and (ii) are a rewriting of (ii) and (iii) of Theorem 2.109.

(iii) Let f be convex, l.s.c, and proper. By (ii) of Proposition 2.112 $f^*(\xi) > -\infty$ for every ξ if and only if $f(x) < +\infty$ at some x, and $f^{**}(x) > -\infty$ for every x if and only if $f^*(\xi) < +\infty$ at some ξ. Since $f^{**} = f$ by (i), we conclude that f^* is proper. Similarly one proves that $f = f^{**}$ is proper if f^* is convex, l.s.c and proper. □

d. Convex duality revisited

Fenchel duality resumes the mystery of convex duality. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function and consider the *primal problem*

$$(\mathcal{P}) \qquad f(x) \to \min$$

and let

$$p := \inf_x f(x).$$

Introduce a function $\phi(x, b) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ such that $\phi(x, 0) = f(x)$ and consider the *value function* of problem $(\mathcal{P})$ (associated to the "perturbation ϕ")

$$v(b) := \inf_x \phi(x, b). \tag{2.78}$$

We have $v(0) = p$.

Compute now the polar $v^*(\xi)$, $\xi \in \mathbb{R}^m$, of the *value function* $v(b)$. The *dual problem* of problem $(\mathcal{P})$ by means of the chosen perturbation $\phi(x, b)$ is the problem

$$(\mathcal{P}^*) \qquad -v^*(\xi) \to \max.$$

Let $d := \sup_\xi -v^*(\xi)$. Then $v^{**}(0) = d$, in fact,

$$v^{**}(0) = \sup_\xi \Big\{ 0 \bullet \xi \, - v^*(\xi) \Big\} = d.$$

The following theorem connects the existence of a maximizer of the dual problem $(\mathcal{P}^*)$ with the regularity properties of the value function v of the primal problem $(\mathcal{P})$. This is the true content of convex duality.

2.118 Theorem. *With the previous notations we have the following:*

(i) $p \geq d$.
(ii) *Assume v convex and $v(0) < +\infty$. Then $p = d$ if and only if v is l.s.c. at 0.*
(iii) *Assume v convex and $v(0) \in \mathbb{R}$. Then $v(b) \geq \eta \bullet b + v(0)$ $\forall b$ if and only if v is l.s.c. at 0 (equivalently $p = d$ by (ii)) and η is a maximizer for problem $(\mathcal{P}^*)$.*

In particular, if v is convex and continuous at 0, then $p = d$ and $(\mathcal{P}^*)$ has a maximizer.

Proof. (i) Since $v^{**} \leq v$ from Proposition 2.116, we get $d = v^{**}(0) \leq v(0) = p$.

(ii) Since $p = d$ means $v(0) = v^{**}(0)$, (ii) follows from (ii) of Theorem 2.117.

(iii) Assume v convex and $v(0) \in \mathbb{R}$. If $v(b) \geq \eta \bullet b + v(0)$ $\forall b$, we infer $v(0) = v^{**}(0)$, hence by (ii), we conclude that v is l.s.c. at 0. Moreover, the inequality $v(b) \geq \eta \bullet b + v(0)$ $\forall b$ is equivalent to $v(0) + v^*(\eta) = 0$ by the Fenchel inequality. Consequently, $-v^*(\eta) = v(0) = v^{**}(0) = d$, i.e., η is a maximizer for $(\mathcal{P}^*)$.

Conversely, if η maximizes $(\mathcal{P}^*)$ and v is l.s.c. at 0, then we have $-v^*(\eta) = d = v^{**}(0)$ and $v(0) = v^{**}(0)$ by (ii). Therefore $v(0) + v^*(\eta) = 0$, which is equivalent to $v(b) \geq \eta \bullet b + v(0)$ $\forall b$ by the Fenchel inequality. □

The following proposition yields a sufficient condition for applying Theorem 2.118.

2.119 Proposition. *With the previous notations, assume that ϕ is convex and that there exists x_0 such that $p \mapsto \phi(x_0, p)$ is continuous at 0. Then v is convex and $0 \in \operatorname{int}(\operatorname{dom}(v))$. If, moreover, $v(0) > -\infty$, then v is continuous at 0.*

Proof. Let us prove that v is convex since ϕ is convex. Choose $p, q \in \mathbb{R}^n$ and $\lambda \in [0,1]$. We have to prove that $v(\lambda p + (1-\lambda)q) \le \lambda v(p) + (1-\lambda)v(q)$. It is enough to assume $v(p), v(q) < +\infty$. For $a > v(p)$ and $b > v(q)$, let $\overline{x}$ and $\overline{y}$ be such that

$$v(p) \le \phi(\overline{x}, p) \le a, \qquad v(q) \le \phi(\overline{y}, q) \le b.$$

Then we have

$$\begin{aligned} v(\lambda p + (1-\lambda)q) &= \inf_z \phi(z, \lambda p + (1-\lambda)q) \le \phi(\lambda\overline{x} + (1-\lambda)\overline{y}, \lambda p + (1-\lambda)q) \\ &\le \lambda\phi(\overline{x}, p) + (1-\lambda)\phi(\overline{y}, q) \le \lambda a + (1-\lambda)b. \end{aligned}$$

Letting $a \to v(p)$ and $b \to v(q)$ we prove the convexity inequality.

(ii) Since $\phi(x_0, b)$ is continuous at 0, $\phi(x_0, b)$ is bounded near 0; i.e., for some $\delta, M > 0$, $\phi(x_0, b) \le M \ \forall b \in B(0, \delta)$. Therefore

$$v(b) = \inf_x \phi(x, b) \le M \qquad \forall b \in B(0, \delta),$$

i.e., $0 \in \operatorname{int}(\operatorname{dom}(v))$. If, moreover, $v(0) > -\infty$, then v is never $-\infty$. We then conclude that v takes only real values near 0, consequently, v is continuous at 0. □

A more symmetrical description of convex duality follows assuming that the perturbed functional $\phi(x, b)$ is convex and l.s.c.. In this case, we observe that

$$v^*(\xi) = \phi^*(0, \xi),$$

where $\phi^*(p, \xi)$ is the polar of ϕ on $\mathbb{R}^n \times \mathbb{R}^m$. In fact,

$$\begin{aligned} \phi^*(0,\xi) &= \sup_{x,b}\Big\{ 0 \bullet x \ + \ b \bullet \xi \ - \phi(x,b)\Big\} = \sup_{x,b}\Big\{ b \bullet \xi \ - \phi(x,b)\Big\} \\ &= \sup_b\Big\{ b \bullet \xi \ - \inf_x \phi(x,b)\Big\} = v^*(\xi). \end{aligned}$$

The dual problem $(\mathcal{P}^*)$ then rewrites as

$$(\mathcal{P}^*) \qquad -\phi^*(0,\xi) \to \max,$$

and the corresponding value function is then $-w(p)$, $p \in \mathbb{R}^m$,

$$w(p) := \inf_\xi \phi^*(p, \xi).$$

Since $\phi^{**} = \phi$, the dual problem of $(\mathcal{P}^*)$, namely

$$(\mathcal{P}^{**}) \qquad \phi^{**}(x, 0) \to \min$$

is again $(\mathcal{P})$. We say that $(\mathcal{P})$ and $(\mathcal{P}^*)$ are dual to each other. Therefore convex duality links the equality $\inf_x \phi(x,0) = \sup_\xi \phi^*(0,\xi)$ and the existence of solutions of one problem to the regularity properties of the value function of the dual problem.

There is also a connection between convex duality and min-max properties typical in game theory. Assume for simplicity that $\phi(x,b)$ is convex and l.s.c. The *Lagrangian* associated to ϕ is the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ defined by

$$-L(x,\xi) := \sup_{b\in\mathbb{R}^m} \Big\{ b \bullet \xi \, - \phi(x,b) \Big\},$$

i.e.,

$$L(x,\xi) = -\phi_x^*(\xi)$$

where $\phi_x(b) := \phi(x,b)$ for every x and b.

2.120 Proposition. *Let ϕ be convex. Then the following hold:*

(i) *For any $x \in \mathbb{R}^n$, $\xi \to L(x,\xi)$ is concave and upper semicontinuous.*
(ii) *For any $\xi \in \mathbb{R}^n$, $x \to L_x(x,\xi)$ is convex.*

Proof. (i) is trivial since $-L$ is the supremum of a family of linear affine functions. For (ii) observe that $L(x,\xi) = \inf_b\{\phi(x,b) - \xi \bullet b\}$. Let $u, v \in \mathbb{R}^n$ and let $\lambda \in [0,1]$. We want to prove that

$$L(\lambda u + (1-\lambda)v) \le \lambda L(u,\xi) + (1-\lambda)L(v,\xi). \tag{2.79}$$

It is enough to assume that $L(u,\xi) < +\infty$ and $L(v,\xi) < +\infty$. For $a > L(u,\xi)$ and $b > L(v,\xi)$ let $b, c \in \mathbb{R}^m$ be such that

$$\begin{aligned} L(u,\xi) &\le \phi(u,b) - \xi \bullet b \le \alpha, \\ L(v,\xi) &\le \phi(v,c) - \xi \bullet c \le \beta. \end{aligned}$$

Then we have

$$\begin{aligned} L(\lambda u + (1-\lambda)v, \xi) &\le \phi(\lambda u + (1-\lambda)v, \lambda b + (1-\lambda)c) - \xi \bullet \lambda b + (1-\lambda)c \\ &\le \lambda\phi(u,b) + (1-\lambda)\phi(v,c) - \lambda\,\xi \bullet b \, - (1-\lambda)\,\xi \bullet c \\ &\le \lambda\alpha + (1-\lambda)\beta. \end{aligned}$$

Letting $\alpha \downarrow L(u,b)$ and $\beta \downarrow L(v,c)$, (2.79) follows. □

Observe that

$$\begin{aligned} \phi^*(p,\xi) &= \sup_{x,b}\{\, p \bullet x \, + \, b \bullet \xi \, - \phi(x,b)\} \\ &= \sup_x\{\, p \bullet x \, + \sup_b\{\, b \bullet \xi \, - \phi(x,b)\}\} \\ &= \sup_x\{\, p \bullet x \, - L(x,\xi)\}. \end{aligned} \tag{2.80}$$

Consequently,

$$d = \sup_\xi -\phi^*(0,\xi) = \sup_\xi \inf_x L(x,\xi). \tag{2.81}$$

On the other hand, for every x, $b \to \phi_x(b)$ is convex and l.s.c., hence

$$\phi(x,b) = \phi_x(b) = \phi_x^{**}(b) = \sup_\xi\{\, b \bullet \xi \, - \phi_x^*(\xi)\} = \sup_\xi\{\, b \bullet \xi \, + L(x,\xi)\}.$$

Consequently,

$$p = \inf_x \phi(x,0) = \inf_x \sup_\xi L(x,\xi).$$

Therefore, the inequality $d \le p$ is a min-max inequality $\sup_\xi \inf_x L(x,\xi) \le \inf_x \sup_\xi L(x,\xi)$ for the Lagrangian, see Section 2.4.8. In particular, the existence of solutions for both $(\mathcal{P})$ and $(\mathcal{P}^*)$ is related to the existence of *saddle points* for the Lagrangian, see Proposition 2.88.

The above applies surprisingly well in quite a number of cases.

2.121 Example. Let φ be convex and l.s.c. Consider the perturbed function $\phi(x,b) := \varphi(x+b)$. The value function $v(b)$ is then constant, $v(b) = v(0)\ \forall b$, hence convex and l.s.c. Its polar is

$$v^*(\xi) := \sup_x\{\, \xi \bullet b \, - v(0)\} = \begin{cases} +\infty & \text{if } \xi \neq 0, \\ -v(0) & \text{if } \xi = 0. \end{cases}$$

The dual problem has then a maximum point at $\xi = 0$ with maximum value $d = v(0)$. Finally, we compute its Lagrangian: Changing variable $c := x + b$,

$$L(x,\xi) = -\sup_b\{\, \xi \bullet b \, - \varphi(x+b)\} = -\sup_c\{\, \xi \bullet c \, - \, \xi \bullet x \, - \varphi(c)\} = \xi \bullet x \, - \varphi^*(\xi).$$

Let $\varphi, \psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be two convex functions and consider the *primal problem*

$$\text{Minimize } \varphi(x) + \psi(x), \qquad x \in \mathbb{R}^n. \tag{2.82}$$

Introduce the perturbed function

$$\phi(x,b) = \varphi(x+b) + \psi(x), \qquad (x,b) \in \mathbb{R}^n \times \mathbb{R}^n, \tag{2.83}$$

for which $\phi(x,0) = \varphi(x) + \psi(x)$, and the corresponding *value function*

$$v(b) := \inf_x(\varphi(x) + \psi(x)). \tag{2.84}$$

Since ϕ is convex, then the value function v is convex, whereas the Lagrangian $L(x,\xi)$ is convex in x and concave in ξ. Let us first compute the Lagrangian. We have

$$-L(x,\xi) := \sup_b\{\, \xi \bullet b \, - \varphi(x+b) - \psi(x)\} = \varphi^*(\xi) - \, \xi \bullet x \, - \psi(x)$$

so that

$$L(x,\xi) = \psi(x) + \, \xi \bullet x \, - \varphi^*(\xi).$$

Now we compute the polar of ϕ. We have

$$\begin{aligned} \phi^*(p,\xi) &= \sup_x\{\, p \bullet x \, - L(x,\xi)\} = \sup_x\{\, p \bullet x \, - \, \xi \bullet x \, - \psi(x) + \varphi^*(\xi)\} \\ &= \sup_x\{\, (p-\xi) \bullet x \, - \psi(x)\} + \varphi^*(\xi) \\ &= \psi^*(p-\xi) + \varphi^*(\xi). \end{aligned}$$

Therefore, the polar of (2.84) is

$$v^*(\xi) = \phi^*(0, \xi) = \varphi^*(\xi) + \psi^*(-\xi) \qquad \forall \xi \in \mathbb{R}^n.$$

As an application of the above we have the following.

2.122 Theorem. *Let φ and ψ be as before, and let ϕ and v be defined by* (2.83) *and* (2.84). *Assume that we have φ continuous at x_0, $\psi(x_0) < +\infty$ at some point x_0 and that $v(0) > -\infty$. Let p and d be defined by the primal and dual optimization problems respectively, through $(x, b) \to \varphi(x + b) + \psi(x)$, given by*

$$p := \inf_x (\varphi(x) + \psi(x)), \tag{2.85}$$

$$d := \sup_\xi (-\varphi^*(\xi) - \psi^*(-\xi)). \tag{2.86}$$

Then $p = d \in \mathbb{R}$ and problem (2.86) *has a maximizer.*

Proof. $\phi(x, b) := \varphi(x + b) + \psi(x)$ is convex. Moreover, since φ is continuous at x_0, then $b \to \phi(x_0, b)$ is continuous at 0. From Proposition 2.119 we then infer that v is convex and continuous at 0. Then the conclusions follow from Theorem 2.118. □

2.123 Example. Let φ be convex. Choose as perturbed functional

$$\phi(x, b) = \varphi(x + b) + \varphi(x)$$

for which $\phi(x, 0) = 2\varphi(x)$. Then, by the above,

$$v^*(\xi) = \varphi^*(\xi) + \varphi^*(-\xi)$$

and the Lagrangian is

$$L(x, \xi) = \varphi(x) + \xi \bullet x - \varphi^*(\xi).$$

Let us consider the convex minimization problem already discussed in Paragraph d. Here we extend it a little further.

Let $f, g^1, \ldots, g^m : \mathbb{R}^n \subset \mathbb{R}^n \to \mathbb{R} \cup_{+\infty}$ be convex functions defined on $\mathbb{R}^n$. We assume for simplicity that either f or $g := (g^1, g^2, \ldots, g^m)$ are continuous. Consider the *primal* minimization problem

$$\text{Minimize } f(x) \text{ with the constraints } g(x) \leq 0. \tag{2.87}$$

Let I_K be the indicatrix of the closed convex set $K := \{x = (x_i) \in \mathbb{R}^n \mid x_i \leq 0 \ \forall i\}$. Problem (2.87) amounts to

$$(\mathcal{P}) \qquad f(x) + I_K(g(x)) \to \min.$$

Let us introduce the perturbed function

$$\phi(x, b) := f(x) + I_K(g(x) - b)$$

which is convex. Consequently, the associated value function

$$v(b) := \sup_x (f(x) + I_K(\varphi(x) - b)), \qquad b \in \mathbb{R}^m, \tag{2.88}$$

is convex by Proposition 2.119. Now, compute the polar of the value function. First we compute the polar of $I_K(y)$. We have

$$I_K^*(\xi) = \sup_b\{\, \xi \bullet b \, - I_K(b)\} = \begin{cases} 0 & \text{if } \xi \geq 0, \\ +\infty & \text{if } \xi < 0. \end{cases}$$

Therefore, changing variables, $c = g(x) - b$,

$$\begin{aligned} -L(x,\xi) &= \sup_b\{\, \xi \bullet b \, - f(x) - I_K(g(x) - b)\} \\ &= -f(x) + \sup_c\{\, \xi \bullet g(x) \, - \, \xi \bullet c \, - I_k(c)\} \\ &= -f(x) + \, g(x) \bullet \xi \, + (I_K)^*(-\xi), \end{aligned}$$

hence

$$L(x,\xi) = \begin{cases} f(x) - \, \xi \bullet g(x) & \text{if } \xi \leq 0, \\ -\infty & \text{if } \xi > 0. \end{cases}$$

Notice that $\sup_\xi L(x,\xi) = f(x) + I_K(g(x)) = \phi(x,0)$. Consequently,

$$\phi^*(p,\xi) = \inf_x \, p \bullet x \, - L(x,\xi) = \begin{cases} +\infty & \text{if } \xi > 0 \\ \sup_x\{\, p \bullet x \, - f(x) + \, g(x) \bullet \xi \,\} & \text{if } \xi \leq 0, \end{cases}$$

and the polar of the value function is

$$v^*(\xi) = \phi^*(0,\xi) = \sup_x\{\, g(x) \bullet \xi \, - f(x)\}.$$

Consequently, the dual problem through the perturbation ϕ is

$$(\mathcal{P}^*) \qquad -v^*(\xi) := \inf_x\{f(x) - \, \xi \bullet \varphi(x)\,\} \to \max \text{ on } \{\xi \geq 0\}.$$

2.124 Theorem. *Let $f, g^1, \ldots, g^m : \mathbb{R}^n \subset \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex functions defined on $\mathbb{R}^n$. Let p and d be defined by the primal and dual optimization problems*

$$p := \inf_x(f(x) + I_K(g(x))), \tag{2.89}$$

$$d := \sup_\xi \inf_x L(x,\xi). \tag{2.90}$$

Assume that $p > -\infty$ and that the Slater condition *holds (namely there exists $x_0 \in \mathbb{R}^n$ such that $f(x_0) < +\infty$, $g(x_0) < 0$ and g is continuous at x_0). Then the dual problem has a maximizer.*

Proof. The function $\phi(x,b) = f(x) + I_K(g(x) - b)$ is convex. Moreover the Slater condition implies that $\phi(x_0, b)$ is continuous at 0. We then infer from Proposition 2.119 that the value function v is convex and continuous at 0. The claims then follow from Theorem 2.118. □

2.6 Exercises

2.125 ¶. Prove that the *n-parallelepiped* of $\mathbb{R}^n$ generated by the vectors $e_1, \dots, e_n$ with vertex at 0,

$$K := \{x = \lambda_1 e_1 + \cdots + \lambda_n e_n \ \mid \ 0 \le \lambda_i \le 1,\ i = 1, \dots, n\},$$

is convex.

2.126 ¶. $K_1 + K_2,\ \alpha K_1,\ \lambda K_1 + (1-\lambda)K_2,\ \lambda \in [0,1]$, are all convex sets if K_1 and K_2 are convex.

2.127 ¶. Show that the convex hull of a closed set is not necessarily closed.

2.128 ¶. Find out which of the following functions is convex:

$$3x^2 + y^y - 4z^2, \qquad x + x^2 + y^2, \qquad (x+y+1)^p \text{ in } x+y+1 > 0,$$
$$\exp(xy), \qquad \log(1 + x^2 + y^2), \qquad \sin(x^2 + y^2).$$

2.129 ¶. Let K be a convex set. Prove that the following are convex functions:

(i) The *support function* $\delta(x) := \sup\{\, x \bullet y \mid y \in K\}$.
(ii) The *gauge function* $\gamma(x) := \inf\{\lambda \ge 0 \,|\, x \in \lambda K\}$.
(iii) The *distance function* $d(x) := \inf\{|x - y| \,|\, y \in K\}$.

2.130 ¶. Prove that $K \subset \mathbb{R}^n$ is a convex body with $0 \in \operatorname{int}(K)$ if and only if there is a gauge function $F : \mathbb{R}^n \to \mathbb{R}$ such that $K = \{x \in \mathbb{R}^n \mid F(x) \le 1\}$.

2.131 ¶. Let $K \subset \mathbb{R}^n$ with $0 \in K$, and for every $\xi \in \mathbb{R}^n$ set

$$d(\xi) := \inf\Big\{ d \in \mathbb{R} \ \Big|\ \xi \bullet x \ \le d \ \forall x \in K \Big\}.$$

Prove that if K is convex with $0 \in \operatorname{int}(K)$, then $d(\xi)$ is a gauge function, i.e.,

$$d(\xi) := \min\Big\{ \xi \bullet x \ \Big|\ x \in K \Big\}$$

and

$$K^* := \Big\{ \xi \in \mathbb{R}^n \ \Big|\ d(\xi) \le 1 \Big\}.$$

2.132 ¶. Let $f : \mathbb{R}_+ \to \mathbb{R}$ be strictly convex with $f(0) = 0$ and $f'(0) = 0$. Write $\alpha(s) := (f')^{-1}(s)$ and prove that

$$f(x) := \int_0^x f'(s)\, ds, \qquad \mathcal{L}_f(y) := \int_0^y \alpha(s)\, ds, \quad y \ge 0.$$

2.133 ¶. Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be a function which is continuous with respect to each variable separately. As we know, f need not be continuous. Prove that $f(t,x)$ is continuous if it is convex in x for every t.

2.134 ¶. Let $C \subset \mathbb{R}^n$ be a closed convex set. Prove that $x_0 \in C$ is an extreme point if and only if $C \setminus \{x_0\}$ is convex.

2.135 ¶. Let $C \subset \mathbb{R}^n$ be a closed convex set and let $f : C \to \mathbb{R}$ be a continuous, convex and bounded function. Prove that $\sup_C f = \sup_{\partial C} f$.

2.136 ¶. Let S be a set and $C = \mathrm{co}(S)$ its convex hull. Prove that $\sup_C f = \sup_S f$ if f is convex on C.

2.137 ¶. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let f_ϵ be its ϵ-mollified where k is a regularizing kernel. Prove that f_ϵ is convex.

2.138 ¶. Let $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi \geq 0$. Then $f(x,t) := \frac{\varphi(x)}{t}$ is convex in $\mathbb{R}\times]0,\infty[$ if and only if $\sqrt{\varphi}$ is convex.

2.139 ¶. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. Prove the following:

(i) If $\Gamma f(x) \neq f(x)$, then $x \in \partial\,\mathrm{dom}(f)$.
(ii) If $\mathrm{dom}(f)$ is closed and f is l.s.c. in $\mathrm{dom}(f)$, then $\Gamma f = f$ everywhere.
(iii) $\inf f = \inf \Gamma f$.
(iv) For all $\alpha \in \mathbb{R}$ we have $\{x \in \mathbb{R}^n \mid \Gamma f(x) \leq \alpha\} = \cap_{\beta>\alpha} \mathrm{cl}(\{x \in \mathbb{R}^n \mid f(x) \leq \beta\})$.
(v) If f_1 and f_2 are convex functions with $f_1 \leq f_2$, then $\Gamma f_1 \leq \Gamma f_2$.

2.140 ¶. Let f be a l.s.c. convex function and denote by $\mathcal{F}$ the class of affine functions $\ell : \mathbb{R}^n \to \mathbb{R}$ with $\ell(y) \leq f(y)\ \forall y \in \mathbb{R}^n$. From Theorem 2.109

$$f(x) = \sup\Big\{\ell(x) \;\Big|\; \ell \in \mathcal{F}\Big\}.$$

Prove that there exists an at most denumerable subfamily $\{\ell_n\} \subset \mathcal{F}$ such that $f(x) = \sup_n \ell_n(x)$.
[*Hint.* Recall that every covering has a denumerable subcovering.]

2.141 ¶. Let $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a function. Its *convex l.s.c. envelope* $\Gamma\mathrm{C}\, f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is defined by

$$\Gamma\mathrm{C}\, f(x) := \sup\Big\{g(x) \;\Big|\; g : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\},\ g \text{ convex and l.s.c., } g \leq f\Big\}.$$

Prove that $\Gamma\mathrm{C}\, f = \Gamma\mathrm{L}\, f$.
[*Hint.* Apply Theorem 2.109 to the convex and l.s.c. minorants of f.]

2.142 ¶. Prove the following: If $\{f_i\}_{i\in I}$ is a family of convex and l.s.c. functions $f_i : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, then

$$\Big(\inf_{i\in I} f_i\Big)^* = \sup_{i\in I} f_i^*, \qquad \Big(\sup_{i\in I} f_i\Big)^* \leq \inf_{i\in I} f_i^*.$$

2.143 ¶. Prove the following claims:

(i) Let $f(x) := \frac{1}{p}|x|^p$, $p > 1$. Then $f^*(\xi) = \frac{1}{q}|\xi|^q$, $1/p + 1/q = 1$.
(ii) Let $f(x) := |x|$, $x \in \mathbb{R}^n$. Then

$$f^*(\xi) = \begin{cases} 0 & \text{if } |\xi| \leq 1, \\ +\infty & \text{if } |\xi| > 1. \end{cases}$$

(iii) Let $f(t) := e^t$, $t \in \mathbb{R}$. Then

$$f^*(\xi) = \begin{cases} +\infty & \text{if } y \leq 0, \\ 0 & \text{if } y = 0, \\ \xi(\log \xi - 1) & \text{if } y > 0. \end{cases}$$

(iv) Let $f(x) := \sqrt{1+|x|^2}$. Then $\mathcal{L}_f$ is defined on $\Omega^* := \{\xi \,|\, |\xi| < 1\}$ and

$$\mathcal{L}_f(\xi) = -\sqrt{1-|\xi|^2},$$

consequently,

$$f^*(\xi) = \Gamma\mathrm{L}\,\mathcal{L}_f(\xi) = \begin{cases} -\sqrt{1-|\xi|^2} & \text{if } |\xi| \le 1, \\ +\infty & \text{if } |\xi| > 1. \end{cases}$$

(v) The function $f(x) = \frac{1}{2}|x|^2$ is the unique function for which $f^*(x) = f(x)$.

2.144 ¶. Show that the following computation rules hold.

Proposition. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. Then the following hold:*

(i) $(\lambda f)^*(\xi) = \lambda f^*(\xi/\lambda)$ $\forall \xi \in \mathbb{R}^n$ *and* $\forall \lambda > 0$.

(ii) *If we set $f_y(x) := f(x-y)$, then we have $f_y^*(\xi) = f^*(\xi) + \xi \bullet y$ $\forall \xi \in \mathbb{R}^n$ and $\forall y \in \mathbb{R}^n$.*

(iii) *Let $\mathbf{A} \in M_{N,n}(\mathbb{R})$, $N \le n$, be of maximal rank and let $g(x) := f(\mathbf{A}x)$. Then*

$$g^*(\xi) = \begin{cases} +\infty & \text{if } \xi \notin \ker \mathbf{A}^\perp, \\ f^*(\mathbf{A}^{-T}\xi) & \text{if } \xi \in \ker \mathbf{A}^\perp = \operatorname{Im} \mathbf{A}^T. \end{cases}$$

2.145 ¶. Let $A \subset \mathbb{R}^n$ and $I_A(x)$ be its indicatrix, see (2.72). Prove the following:

(i) If L is a linear subspace if $\mathbb{R}^n$, then $(I_L)^* = I_{L^\perp}$.

(ii) If C is a closed cone with the origin as vertex, then $(I_C)^*$ is the indicatrix function of the cone generated by the vectors through the origin that are orthogonal to C.

3. The Formalism of the Calculus of Variations

One of the most beautiful and widely spread paradigms of science and mathematics in particular is that of *minimum principles*. It is strongly related to the everyday principle of economy of means and to the research of optimal strategies to realize our goals. Therefore, it is not surprising that since the beginning minimum principles have been used to formulate *laws of nature*. We have already seen a few examples in Chapter 6 of [GM1] and in this volume.

In short we may state that the Calculus of Variations deals with the problem of finding optimal solutions and of describing their properties. Its development begins right after the introduction of calculus and is connected with the names of Gottfried W. von Leibniz (1646–1716), Johann Bernoulli (1667–1748), Sir Isaac Newton (1643–1727) and Christiaan Huygens (1629–1695). It becomes a sufficient flexible and efficient theory with Leonhard Euler (1707–1783), Joseph-Louis Lagrange (1736–1813) and with the subsequent contributions of Carl Jacobi (1804–1851), Karl Weierstrass (1815–1897) and Adrien-Marie Legendre (1752–1833).

In the fist 200 years of its history the *indirect approach* to minimum problems was prevailing. Its naive idea was that every minimum problem had a solution; the goal was therefore to find necessary conditions for minimality. These were expressed in terms of the so-called *Euler–Lagrange* equations, corresponding to the vanishing of the first variation, and then sufficient conditions to grant minimality corresponding to the positivity of the second variation.

Mainly due to the difficulty in solving even in principle Euler–Lagrange equations for multidimensional equations (that are partial differential equations), new methods, called *direct methods*, were developed. They consist in proving directly the existence of a minimizer (and, consequently, the existence of a solution of the Euler–Lagrange equations). This implies the necessity of extending the functional to be minimized to classes of generalized functions such as Sobolev classes (as we saw in the case of the *Dirichlet principle* in Chapter 1) and postpone the problem of the regularity of minima. This approach originated in the works of Carl Friedrich Gauss (1777–1855), Peter Lejeune Dirichlet (1805–1859) and Georg F. Bernhard Riemann (1826–1866), in the attempts to prove Dirichlet principle by Cesare Arzelà (1847–1912) and Beppo Levi (1875–1962) (it is in this context that the Ascoli–Arzelà theorem and the first ideas of Beppo Levi spaces

M. Giaquinta and G. Modica, *Mathematical Analysis, Foundations and Advanced Techniques for Functions of Several Variables*, DOI 10.1007/978-0-8176-8310-8_3,

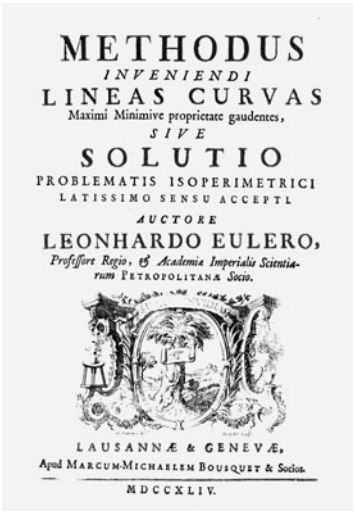

METHODUS
INVENIENDI
LINEAS CURVAS
Maximi Minimive proprietate gaudentes,
SIVE
SOLUTIO
PROBLEMATIS ISOPERIMETRICI
LATISSIMO SENSU ACCEPTI
AUCTORE
LEONHARDO EULERO,
Professore Regio, & Academiæ Imperialis Scientiarum Petropolitanæ Socio.
LAUSANNÆ & GENEVÆ,
Apud Marcum-Michaelem Bousquet & Socios.
MDCCXLIV.

Figure 3.1. Leonhard Euler (1707–1783) and the frontispiece of *Methodus Inveniendi Lineas Curvas Maximi Minimive Proprietate Gaudentes*, 1714.

and, subsequently, of Sobolev spaces appear) and, mainly, in the works of David Hilbert (1862–1943) and Henri Lebesgue (1875–1941). This approach finds its solid basis in the works of Leonida Tonelli (1885–1946), develops further with Charles Morrey (1907–1984) and still is a contemporary topic of research.

A contemporary research topic is also the *calculus of variations in the large*, in the terminology of Harald Marston Morse (1892–1977). It deals with the problem of critical points in terms of global structures. A typical example is Morse's theory of geodesics, and it distinguishes itself because of the prevalent use of topological methods.

The result is that the calculus of variations is a discipline made of general methods and specific problems relevant in geometry, physics and modelling. In fact, besides its beauty, it is fundamental in the formulation of classical mechanics and even in wave and quantum mechanics.

Many volumes have been dedicated to the calculus of variations both classical and modern, and it would be impossible to describe its content even partially. In this chapter we confine ourselves to providing a short introduction to its formalism with the aim to illustrate some of its connections to mechanics and geometrical optics, that is, of hinting at its foundational character and of discussing, although formally, some specific examples.

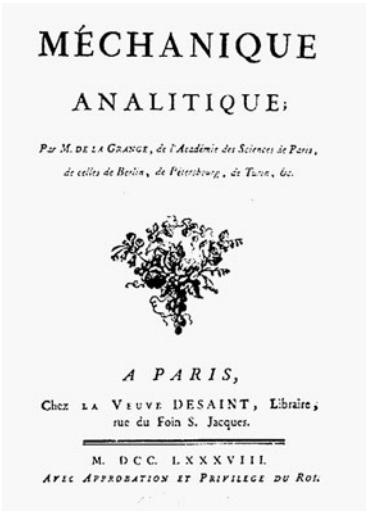


Figure 3.2. Joseph-Louis Lagrange (1736–1813) and the frontispiece of the first edition of his *Mécanique Analytique*, Paris, 1788.

3.1 Lagrangian Formalism

Mechanics and physics in general, geometry and modeling lead, for instance, to consider functionals of the type

$$\int_\Omega F(x, u(x), Du(x))\, dx$$

with *integrand* F depending on the independent variable x, the values of u and of its derivatives at x, or, more generally, upon x, $u(x)$ and the derivatives up to a fixed order $m \geq 1$ of u.

3.1.1 Euler–Lagrange equations

Let Ω be a bounded open and connected set of $\mathbb{R}^n$ with smooth boundary $\partial\Omega$, and let $F(x, u, p)$ be a function of class C^1 from $\overline{\Omega} \times \mathbb{R}^N \times \mathbb{R}^{nN}$ into $\mathbb{R}$. For any function $u : \Omega \to \mathbb{R}^N$ of class $C^1(\overline{\Omega})$ the *variational integral*

$$\mathcal{F}(u) := \int_\Omega F(x, u(x), Du(x))\, dx, \tag{3.1}$$

where $Du(x) = \{D_\alpha u^i(x)\}$, $\alpha = 1, \ldots, n$, $i = 1, \ldots, N$, is the Jacobian matrix of u, is well-defined.

In this subsection we state the equilibrium equations for minima of variational integrals known as *Euler–Lagrange equations*. Here we deal with unconstrained minimum problems; constrained minimum problems will be discussed in Section 3.1.3.

a. Dirichlet's problem

Given a function $\varphi : \partial\Omega \to \mathbb{R}^N$ of class $C^1(\partial\Omega)$, consider the class of maps

$$C^1_\varphi(\overline{\Omega}, \mathbb{R}^N) := \Big\{ v \in C^1(\overline{\Omega}, \mathbb{R}^N) \,\Big|\, v = \varphi \text{ on } \partial\Omega \Big\}$$

and suppose that $u \in C^1_\varphi(\overline{\Omega})$ is a minimum point for $\mathcal{F}$ in this class,

$$\mathcal{F}(u) = \min\Big\{ \mathcal{F}(v) \,\Big|\, v \in C^1_\varphi(\overline{\Omega}, \mathbb{R}^N) \Big\}. \tag{3.2}$$

Then for all $\psi \in C^1_c(\Omega, \mathbb{R}^N)$ the function

$$\epsilon \to \Psi(\epsilon) := \mathcal{F}(u + \epsilon\psi), \qquad \epsilon \in]-1, 1[,$$

is differentiable and has a minimum point at $\epsilon = 0$; hence, according to Fermat's theorem, the *first variation of* $\mathcal{F}$ *in the direction* ψ defined by

$$\delta\mathcal{F}(u, \psi) := \frac{d}{d\epsilon}\Psi(\epsilon)\Big|_{\epsilon=0} \tag{3.3}$$

vanishes, and we can state the following.

3.1 Theorem. *Let* u *be a minimizer of problem* (3.2)*. Then the following hold:*

(i) *For all* $\psi \in C^1_c(\Omega, \mathbb{R}^N)$ *the function* $\epsilon \to \mathcal{F}(u + \epsilon\varphi)$ *is differentiable and* u *is a solution of* Euler–Lagrange equations in the weak form

$$\sum_{i=1}^{N} \sum_{\alpha=1}^{n} \int_\Omega \Big(F_{u^i}(x, u, Du)\psi^i(x) + F_{p^i_\alpha}(x, u, Du) D_\alpha \psi^i(x) \Big)\, dx = 0 \tag{3.4}$$

$\forall \psi \in C^1_c(\Omega, \mathbb{R}^N)$.

(ii) *If* u *if of class* $C^2(\Omega)$*, then* (3.4) *is equivalent to the* Euler–Lagrange equations in the strong form

$$F_{u^i}(x, u, Du) - \sum_{\alpha=i}^{n} D_\alpha F_{p^i_\alpha}(x, u, Du) = 0 \qquad \text{in } \Omega, \qquad \forall i = 1, \dots, N. \tag{3.5}$$

3.2 Remark. Here and in the sequel we will deal with functions $F(x, u, p)$ defined in $\overline{\Omega} \times \mathbb{R}^n \times \mathbb{R}^{nN}$. We denote the relative variables by $x = (x^1, x^2, \dots, x^n)$, $u = (u^1, u^2, \dots, u^N)$ and $p = (p^i_\alpha)$, $\alpha = 1, \dots, n$, $i = 1, \dots, N$. The partial derivatives of F with respect to its variables will be denoted by F_{x^α}, F_{u^i} and $F_{p^\alpha_i}$, and we do not specify their arguments if it is not necessary. Moreover, we set

$$F_x := (F_{x^\alpha})^T, \qquad F_u := (F_{u^i})^T \qquad \text{and} \qquad F_p := [F_{p^i_\alpha}].$$

Often, for given $u = u(x, \epsilon)$ and $p = p(x, \epsilon)$, we consider

$$\phi(x,\epsilon) := F(x,u(x,\epsilon),p(x,\epsilon)).$$

We use the abbreviations

$$D_{x^\alpha}F = D_\alpha F := \frac{\partial\phi}{\partial x^\alpha}, \qquad \frac{\partial}{\partial\epsilon}F := \frac{\partial\phi}{\partial\epsilon},$$

possibly leaving out the point (x,ϵ) at which they are evaluated. We will often use the convention that couples of controvariant indices are understood as summed. Finally, we will use Greek indices to enumerate the independent variable $x = (x^\alpha)$, $\alpha = 1,\dots,n$ and Latin indices to enumerate the components of $u = (u^i)$, $i = 1,\dots,N$. For instance,

$$x^\alpha D_\alpha(F) := \sum_{\alpha=1}^{n} x^\alpha \frac{\partial}{\partial x^\alpha}\Big(F(x,u(x),Du(x))\Big).$$

With these agreements, the weak and strong forms of Euler–Lagrange equations, respectively (3.4) and (3.5), write as

$$\int_\Omega \Big(F_{u^i}\psi^i + F_{p^i_\alpha}D_\alpha\psi^i\Big)\,dx = 0 \qquad \forall\psi\in C^1_c(\Omega,\mathbb{R}^N),$$

and

$$F_{u^i} - D_\alpha F_{p^i_\alpha} = 0 \qquad \text{in } \Omega, \qquad i = 1,\dots,N.$$

Proof of Theorem 3.1. (i) Let $\psi\in C^1_c(\Omega,\mathbb{R}^N)$. Differentiating under the integral sign one easily sees that $\Psi(\epsilon) := \mathcal{F}(u+\epsilon\psi)$ is in fact differentiable and that

$$\delta\mathcal{F}(u,\psi) = \Psi'(0) = \int_\Omega \frac{\partial}{\partial\epsilon}F(x,u(x)+\epsilon\varphi(x),Du(x)+\epsilon D\varphi(x))\Big|_{\epsilon=0}dx.$$

The computation of the derivative then leads to (3.4).
(ii) For $\psi\in C^1_c(\Omega,\mathbb{R}^N)$, the functions $x\to F_{p^i_\alpha}(x,u(x),Du(x))\psi^i$ are of class $C^1_c(\Omega)$, hence Green's formulas yield

$$\int_\Omega F_{p^i_\alpha}(x,u(x),Du(x))D_\alpha\psi^i(x)\,dx = -\int_\Omega D_\alpha\Big(F_{p^i_\alpha}(x,u(x),Du(x))\Big)\psi^i(x)\,dx.$$

Equation (3.4) then becomes

$$\int_\Omega \sum_{i=1}^{N}\Big(F_{u^i}(x,u,Du) - \sum_{\alpha=1}^{n} D_\alpha F_{p^i_\alpha}(x,u,Du)\Big)\psi^i(x)\,dx = 0 \qquad \forall\psi\in C^1_c(\Omega,\mathbb{R}^N),$$

i.e., (3.5), if we take into account the fundamental lemma of the calculus of variations Lemma 1.51. □

3.3 Remark. Going through the derivation, we see at once that the weak and the strong forms of Euler–Lagrange equations are equivalent under the weaker condition that u and the functions

$$x\to F_{p^i_\alpha}(x,u(x),Du(x))$$

are of class $C^1(\Omega)$.

3.4 Definition. *A solution of the Euler–Lagrange equation in the weak form* (3.4) *is called an* extremal *of the functional* (3.1).

b. Natural boundary conditions

Often one does not want to prescribe the values of the competing maps along the entire boundary, but only on a part, or nowhere on the boundary. For instance, consider a point-mass m on a vertical plane that starting at $u(0) = 0$ slides along a curve under the action of gravity in such a way as to meet the vertical line $x = b$ in the shortest time. In this case the value of u at b is an unknown of the problem.

Let Ω be a bounded and connected open set of $\mathbb{R}^n$. Let $\Gamma \subset \partial\Omega$ and suppose that either $\Gamma = \emptyset$ or Γ is such that the separation surface between Γ and $\partial\Omega \setminus \Gamma$ is a smooth $(n-2)$-submanifold and let $\varphi : \Gamma \to \mathbb{R}^N$ be a smooth map. Define

$$C^1_{\varphi,\Gamma}(\overline{\Omega}, \mathbb{R}^N) := \Big\{ v \in C^1(\overline{\Omega}, \mathbb{R}^N) \,\Big|\, v = \varphi \text{ on } \Gamma \Big\},$$
$$C^1_{0,\Gamma}(\overline{\Omega}, \mathbb{R}^N) := \Big\{ v \in C^1(\overline{\Omega}, \mathbb{R}^N) \,\Big|\, v = 0 \text{ on } \Gamma \Big\}.$$

Suppose that $u \in C^1_{\varphi,\Gamma}(\overline{\Omega})$ is a minimizer of the functional

$$\mathcal{F}(u) := \int_\Omega F(x, u(x), Du(x))\, dx$$

in the class $C^1_{\varphi,\Gamma}(\overline{\Omega}, \mathbb{R}^N)$. In this case, the *admissible variations* are the functions $\psi \in C^1_{0,\Gamma}(\overline{\Omega}, \mathbb{R}^N)$. As in the case of Dirichlet's problem, we then conclude that the function $\epsilon \to \mathcal{F}(u + \epsilon\psi)$ is differentiable and that

$$\int_\Omega \Big(F_{u^i} \psi^i + F_{p^i_\alpha} D_\alpha \psi^i \Big)\, dx = 0 \qquad \forall \psi \in C^1_{0,\Gamma}(\overline{\Omega}, \mathbb{R}^N). \tag{3.6}$$

In particular, u is an extremal of $\mathcal{F}$ and, assuming $u \in C^2(\overline{\Omega}, \mathbb{R}^N)$, u solves Euler–Lagrange equations in the strong form

$$F_{u^i} - D_\alpha F_{p^i_\alpha} = 0 \qquad \text{in } \Omega,\ \forall i = 1, \dots, N. \tag{3.7}$$

Suppose now that u is of class C^2 up to the boundary. Then the functions $x \mapsto F_{p^i_\alpha}(x, u(x), Du(x))\psi^i(x)$ are of class $C^1(\overline{\Omega}, \mathbb{R}^N)$; hence, by choosing $\psi \in C^1_{0,\Gamma}(\overline{\Omega}, \mathbb{R}^N)$, $\psi = (\psi^i)$, summing on and integrating over Ω and subtracting (3.6), we conclude that

$$\int_\Omega D_\alpha (F_{p^i_\alpha} \psi^i)\, dx = \int_\Omega \Big(D_\alpha F_{p^i_\alpha} \psi^i + F_{p^i_\alpha} D_\alpha \psi^i \Big)\, dx = 0$$

i.e.,

$$\int_{\partial\Omega} F_{p^i_\alpha} \nu^\alpha_\Omega \psi^i \, d\mathcal{H}^{n-1} = 0,$$

$\nu_\Omega = (\nu^\alpha_\Omega)$ being the exterior normal vector to $\partial\Omega$. Since ψ is arbitrary in $\partial\Omega \setminus \Gamma$, we conclude that

$$F_{p^i_\alpha}(x, u(x), Du(x))\,\nu^\alpha_\Omega(x) = 0 \qquad \forall x \in \partial\Omega \setminus \Gamma,\ \forall i = 1, \dots, N. \tag{3.8}$$

These are called the *natural conditions* or the *vanishing of the co-normal derivative*. Summing up, if $u \in C^2(\overline{\Omega})$ is a minimizer of $\mathcal{F}$ on $C^1_{\psi,\Gamma}(\Omega)$, then u solves (3.6) and, actually, (3.6) is equivalent to the *Dirichlet–Neumann problem*.

$$\begin{cases} F_{u^i} - D_\alpha F_{p^i_\alpha} = 0 & \text{in } \Omega \qquad \forall i = 1, \dots, N, \\ u = \varphi & \text{on } \Gamma, \\ F_{p^i_\alpha}\,\nu^\alpha = 0 & \text{on } \partial\Omega \setminus \Gamma\ \forall i = 1, \dots, N. \end{cases} \tag{3.9}$$

If u is only of class $C^1_{\psi,\Gamma}$, then u satisfies in principle only (3.6), and we may interpret (3.6) as the weak form of (3.9).

3.5 Example. For the Dirichlet integral, the minimizer of

$$\frac{1}{2}\int_\Omega |Du|^2\,dx \to \min \qquad \text{in } C^1_{\varphi,\Gamma}(\overline{\Omega}, \mathbb{R}^N),$$

assuming that it exists, is regular and is unique, it is the solution of the weak form of the boundary value problem

$$\begin{cases} \Delta u = 0 & \text{in } \Omega, \\ u = \varphi & \text{su } \partial\Omega, \\ \frac{du}{d\nu} = 0 & \text{su } \partial\Omega \setminus \Gamma. \end{cases}$$

c. Examples

Let us illustrate a few examples and add some further remarks.

3.6 Graphs of prescribed curvature. Consider the functional

$$\mathcal{F}(u) := \int_{-1}^{1} \left(\sqrt{1 + u'(x)^2} + H(x, u(x))\right) dx,$$

where $H(x, u)$ is a given continuous function in $[-1, 1] \times \mathbb{R}$ and

$$J(u) := \int_{-1}^{1} \sqrt{1 + u'(x)^2}\,dx$$

is the length of the curve $x \to (x, u(x))$, which is the graph of $u : [-1, 1] \to \mathbb{R}$. The Euler–Lagrange equation in its strong form is then

$$\left(\frac{u'(x)}{\sqrt{1 + u'(x)^2}}\right)' = H_u(x, u(x)) \qquad \text{in }]-1, 1[; \tag{3.10}$$

that is, the graph of a C^2 extremal of $\mathcal{F}$ has at every point mean curvature $H_u(x, u(x))$. In particular, if $H(x, u) = Hu$, $H \in \mathbb{R}$, then the graph of u has constant mean curvature H. In this case we may explicitly integrate the equation and find that (3.10) has no solution if $|H| > 1$, whereas solutions are given by arcs of circles of radius $1/|H|$ if $|H| \le 1$.

3.7 Harmonic oscillator. The Euler–Lagrange equation of

$$\frac{1}{2}\int_a^b (u'(x)^2 - \omega^2 u^2(x))\,dx$$

is the harmonic oscillator equation

$$u'' + \omega^2 u = 0 \qquad \text{in }]a,b[.$$

3.8 Fermat's principle. Consider the following functional defined on curves $u : [t_1, t_2] \to \mathbb{R}^N$:

$$\mathcal{F}(u) := \int_{t_1}^{t_2} \omega(\tau, u(\tau))\sqrt{1 + |u'(\tau)|^2}\,d\tau.$$

Since $\sqrt{1 + |u'(\tau)|^2}\,d\tau$ is the length element of the curve $\tau \to (\tau, u(\tau))$, if we think of $\omega(\tau, u(\tau))$ as the inverse of a velocity, the quantity

$$\omega(\tau, u(\tau))\sqrt{1 + |u'(\tau)|^2}\,d\tau$$

has the dimension of time and we may think of $\mathcal{F}(u)$ as the time needed to travel along the curve $u(t)$ from $u(t_1)$ to $u(t_2)$.

The functional $\mathcal{F}$ may therefore describe the trajectory of a ray of light. In fact, according to Fermat, the refraction index of a medium at $P := (x, y)$ is the inverse of the velocity of light at P, and light travels along trajectories that minimize time.

More generally, denote by x the position of a point in a medium and, for any direction ξ, $|\xi| = 1$, denote by $F(x, \xi)$ the refraction index of the medium at x in the direction ξ and think of F as being extended to all vectors as the homogeneous function of degree 1

$$F(x, \xi) := |\xi| F\Big(x, \frac{\xi}{|\xi|}\Big) \qquad \forall \xi \neq 0.$$

If $s \to x(s)$ is a regular curve, the function

$$F(x(s), x'(s))\,ds = |x'(s)| F\Big(x(s), \frac{x'(s)}{|x'(s)|}\Big)\,ds$$

has the dimension of time and

$$\mathcal{F}(x) := \int_{s_1}^{s_2} F(x(s), x'(s))\,ds$$

is the time necessary to go from $x(s_1)$ to $x(s_2)$ along the trajectory of $x(s)$. Notice that $\mathcal{F}(x)$ does not depend on the parametrization chosen for $x(s)$.

Fermat's principle states that light moves in a medium characterized by F from P_1 to P_2 along a trajectory that minimizes the total time

$$\int_{s_1}^{s_2} F(x(s), x'(s))\,ds \to \min.$$

3.9 Hamilton's principle. Let $m_1, \ldots, m_n$, $m_j > 0$, be the masses of n points that move in time under the action of forces $F_1, F_2, \ldots, F_n$; denote by $X_j(t) := (x_j(t), y_j(t), z_j(t))$ the position vector at time t of the point-mass m_j and let $X(t) := (X_1(t), \ldots, X_n(t)) \in \mathbb{R}^{3n}$; finally, assume that $F_j = F_j(X)$. The motion of the n point-masses is ruled by Newton's law

$$mX_j''(t) = F_j(X(t)), \qquad j = 1, \ldots, n.$$

Assuming that the forces are conservative, i.e., that there is a function $V : \mathbb{R}^{3n} \to \mathbb{R}$ such that

$$F_j(X) = D_{X_j} V(X) := \Big(V_{x^j}(X), V_{y^j}(X), V_{z^j}(X) \Big),$$

it is easily seen that Newton's equations are Euler–Lagrange equations of the functional, called *action*,

$$\mathcal{L}(X) := \int_{t_1}^{t_2} \Big(\frac{1}{2} \sum_{j=1}^{n} m_j |X_j'(t)|^2 - V(X(t)) \Big)\, dt.$$

The function $L(X, Y) : \mathbb{R}^{3n} \times \mathbb{R}^{3n} \to \mathbb{R}$ defined by

$$L(X, Y) := \frac{1}{2} \sum_{j=1}^{n} m_j |Y_j|^2 - V(X),$$

where $Y = (Y_1, \ldots, Y_n)$, $Y_i \in \mathbb{R}^3$ $\forall i$, is called the *Lagrangian* of the system. The Lagrangian, computed at a trajectory X, i.e., $L(X(t), X'(t))$, is the difference between the *kinetic energy*

$$T(X'(t)) := \frac{1}{2} \sum_{j=1}^{n} m_j |X_j'(t)|^2$$

and the *potential energy*

$$V(X(t))$$

of the trajectory X at time t. $L(X(t), X'(t))$ is called the *free energy* of X at t, while $H(X(t), X'(t)) := T(X'(t)) + V(X(t))$, which is the sum of the kinetic energy and the potential energy, is called the *total energy* of the trajectory X at time t, see Section 3.2.

The previous remarks lead to the following.

Hamilton's principle of stationary action. The actual motion of a conservative system takes place in such a way that it makes the Lagrangian action stationary.

Notice that this way a conservative mechanical system is described by a single function, its Lagrangian; moreover, it clearly shows that Hamilton's principle is invariant with respect to the coordinates chosen to describe the system.

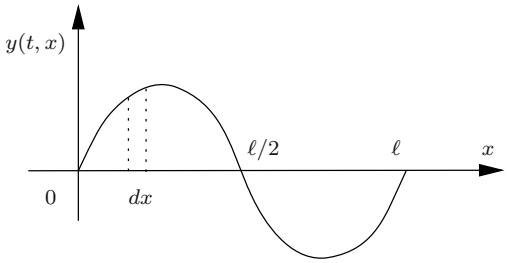


Figure 3.3. Wave equation.

3.10 Wave equations. Hamilton's principle allows us to deduce the wave equation in a variational way as the equation of motion of a vibrating string $y = y(t, x)$ with small transverse oscillations. Let ρ and τ denote the density and the tension of the string. The kinetic and potential energies are then given by

$$T := \frac{1}{2}\int_0^\ell \rho\Big(\frac{\partial y}{\partial t}\Big)^2\,dx, \qquad V := \frac{1}{2}\int_0^\ell \tau\Big(\frac{\partial y}{\partial x}\Big)^2\,dx,$$

respectively, so that the action of the system is

$$\mathcal{L}(y) := \frac{1}{2}\int_{t_0}^{t_1} dt \int_0^\ell \Big(\rho\Big(\frac{\partial y}{\partial t}\Big)^2 + \tau\Big(\frac{\partial y}{\partial x}\Big)^2\Big)\,dx.$$

When ρ and τ are independent of x, its Euler–Lagrange equation is

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}, \qquad c^2 := \frac{\tau}{\rho}.$$

3.11 Lagrangians of the type $F(x, p)$, $n = 1$. When $F = F(x, p)$, $n = 1$, Euler–Lagrange equations simplify to

$$\frac{d}{dx}F_p(x, u'(x)) = 0,$$

i.e.,

$$F_p(x, u'(x)) = a$$

where $a = (a_1, a_2, \ldots, a_N)$ is a constant vector. If $F_{pp}(x, u'(x)) \neq 0$, the implicit function theorem allows us to write, at least locally, u' as a function of (x, a),

$$u'(x) = g(x, a);$$

consequently,

$$u(x) = u(x_0) + \int_{x_0}^{x} g(t, a)\,dt.$$

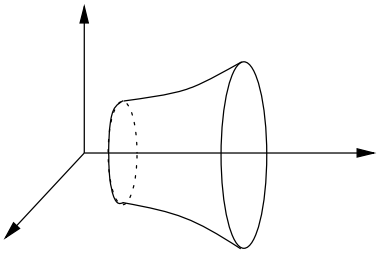


Figure 3.4. Rotationally symmetric minimal surfaces.

3.12 Lagrangians of the type $F(u,p)$, $n=N=1$. When $n = N = 1$, instead of looking for a solution of the Euler–Lagrange equation, that is a second order equation, we notice that the quantity

$$u'F_p - F$$

is constant in time along an extremal u: in fact,

$$\frac{d}{dt}(u'F_p - F) = u''F_p + u'\frac{d}{dt}F_p - F_u u' - F_p u'' = u'(F_u - \frac{d}{dt}F_p) = 0.$$

One also says that $u'\,F_p(u,u') - F(u,u')$ is a *first integral of the motion*. Later we shall discuss, see Theorem 3.60, how we may seek conservation laws or first integrals. Here we notice that the previous conservation law allows us to integrate Euler–Lagrange equation at least locally if $F_{pp} \neq 0$ since $n = N = 1$. In fact, according to the implicit function theorem, we may write $u'\,F_p(u,u') - F(u,u') = const = h$ as $u'(x) = \psi(u(x),h)$. Separation of variables then leads to

$$x = x_0 + \int_{u_0}^{u(x)} \frac{dz}{\psi(z,h)}$$

which, in principle, allows us to find $u(x)$.

3.13 Rotationally symmetric minimal surfaces. Let S be a surface in $\mathbb{R}^3$ obtained by rotating along the x-axis its meridian, $z = u(x)$, $a \leq x \leq b$, $u(x) > 0$ in $[a,b]$. The area of S is

$$A(S) := 2\pi \int_a^b u\sqrt{1+u'^2}\,dx \tag{3.11}$$

and the energy conservation law yields at once

$$u\sqrt{1+u'^2} - u\frac{u'^2}{\sqrt{1+u'^2}} = c.$$

We can infer

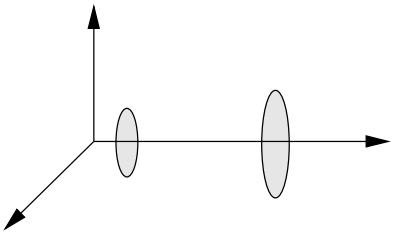


Figure 3.5. Goldsmith's degenerate rotational minimal surface.

$$u' = \sqrt{\frac{u^2 - c^2}{c^2}},$$

and, by separating variables,

$$x + c_1 = c \log \frac{u + \sqrt{u^2 - c^2}}{c},$$

i.e.,

$$u(x) = c \cosh \frac{x + c_1}{c}.$$

A rotational surface of minimal area is therefore generated by the *catenary* and the generated surface is called a *catenoid*; the values of constants c and c_1 are determined by $u(a)$ and $u(b)$.

Further inquiries, that we omit, would show that three cases are possible:

(i) There is a unique catenary through the given points; in this case, it solves the minimum problem.
(ii) There are two catenaries through the given points: One is a minimizer and the other just an extremal.
(iii) There is no catenary through the given points: The minimum problem (3.11) has no solution; in fact, minimizing sequences converge to Goldsmith's rotational surface in Figure 3.5.

3.14 The brachystochrone. One of the classical problems posed by Johann Bernoulli (1667–1748) in 1696 is that of the brachystochrone or of the quickest descent: For two points P_1 and P_2 on a vertical plane, find a line connecting them, on which a point descends from P_1 to P_2 under the influence of gravitation in the quickest possible way. By choosing coordinates as in Figure 3.6 with $P_1 = (0, 0)$ and $P_2 = (a, b)$, the conservation of energy tells us that

$$\frac{1}{2} v^2 + g\, z = 0.$$

On the other hand, $v = \frac{ds}{dt} = \sqrt{1 + u'^2}\, \frac{dx}{dt}$, hence

$$dt = \sqrt{\frac{1+|u'|^2}{-2gu}}$$

and we ought to minimize the functional

$$\mathcal{F}(u) := \int_0^a \sqrt{\frac{1+u'^2}{-2gu}}\, dx$$

with the conditions $u(0) = 0$ and $u(a) = b$. This is similar to the problem in Example 3.13, slightly more complicated due to the fact that u vanishes at the first boundary, so that the integrand is not regular in $[0, a] \times \mathbb{R} \times \mathbb{R}$. One can show, but we will not do it, that the problem has a unique solution; it is of class C^1 and it is given by an arc of cycloid.

3.15 The curvature functional. The condition of vanishing of the first variation extends easily to functionals with integrands depending in higher order derivatives. For instance, the equilibrium configuration of a *thin plate* is described by the minimizers of the integral

$$P(u) := \frac{1}{2}\int_\Omega |\Delta u|^2\, dx$$

under suitable boundary conditions. If u is a minimizer, u is an extremal, i.e.,

$$\frac{d}{d\epsilon} P(u + \epsilon\varphi)\Big|_{\epsilon=0} = 0 \qquad \forall \varphi \in C_c^1(\Omega),$$

and this yields, in turn, as Euler–Lagrange equations in strong form, the *biharmonic equation*

$$\Delta\Delta u = 0.$$

Here we confine ourselves to discussing with some details only the *curvature functional*. Let $s \to c(s)$, $s \in [0, L]$, $c'(s) \neq 0$, be a regular plane curve, parametrized by the arclength, and let $(\mathbf{t}(s), \mathbf{n}(s))$ be its moving frame, $\mathbf{t}(s) = c'(s)$ being its tangent unit vector and $\mathbf{n}(s)$ its normal unit

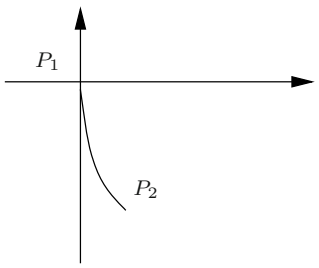


Figure 3.6. The brachystochrone.

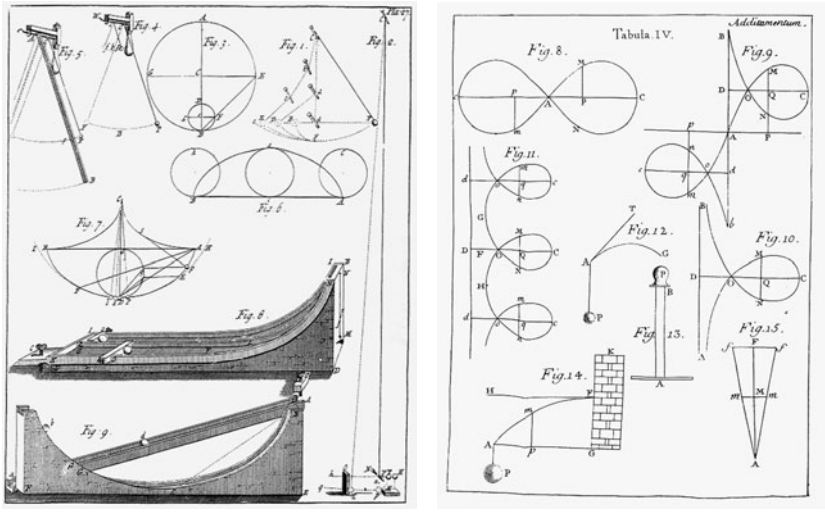


Figure 3.7. An illustration of the cycloid from *Course of Experimental Philosophy* by J. T. Desaguliers (1683–1744) and a table from *Methodus Inveniendi* by Leonhard Euler (1707–1783) on elastic curves.

vector oriented in such a way that $\det[\mathbf{t}(s)|\mathbf{n}(s)] = 1$. Recall that the signed curvature $k_c(s)$ of c at s is defined by

$$\mathbf{t}'(s) =: k_c(s)\,\mathbf{n}(s), \qquad \text{equivalently,} \qquad k_c(s) = \det[c'(s)\,|\,c''(s)],$$

and that

$$\mathbf{n}'(s) = -k_c(s)\,\mathbf{t}(s),$$

see [GM4].

Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function. We want to write the Euler–Lagrange equation of the integral

$$\mathcal{F}(c) := \int_0^L f(k_c(s))\,ds.$$

Let $\varphi \in C_c^\infty([0, L], \mathbb{R}^N)$. Since the curve $c(s)+\epsilon\varphi(s)$, $s \in [0, L]$, is no longer parametrized by the arc length when $\epsilon \neq 0$, it is convenient to rewrite the functional with respect to a generic parametrization $c : [a, b] \to \mathbb{R}^n$

$$\mathcal{F}(c) := \int_0^L f(k_c(t))|c'(t)|\,dt, \tag{3.12}$$

where, this time,

$$\mathbf{t}(t) = \frac{c'(t)}{|c'(t)|}, \qquad k_c(t) = \frac{1}{|c'(t)|^3}\det[c'(t)\,|\,c''(t)],$$

and

$$\mathbf{t}'(t) = k_c(t)\,\mathbf{n}(t), \qquad \mathbf{n}'(t) = -k_c(t)\,\mathbf{t}(t),$$

see [GM4], Section 5.4.1.

If $c_\epsilon(s) := c(s) + \epsilon\varphi(s)$, $s \in [0, L]$, differentiating under the integral sign the function $\epsilon \to \mathcal{F}(c_\epsilon)$, we get

$$\frac{d}{d\epsilon}\mathcal{F}(c_\epsilon) = \int_0^L \Big(f'(k_\epsilon(s))\Big(\frac{\partial}{\partial\epsilon}k_\epsilon(s)\Big)|c'_\epsilon(s)| + f(k_\epsilon(s))\frac{\partial}{\partial\epsilon}|c'_\epsilon(s)|\Big)\,dt,$$

where, for the sake of simplicity, we write $\mathbf{t}_\epsilon$, $\mathbf{n}_\epsilon$ and $k_\epsilon(s)$ for $\mathbf{t}_{c_\epsilon}$, $\mathbf{n}_{c_\epsilon}$ and $k_{c_\epsilon}(s)$ respectively, and where the prime denotes differentiation with respect to s. We now remark that

$$\frac{\partial}{\partial\epsilon}|c'_\epsilon(s)| = \frac{c'_\epsilon(s)}{|c_\epsilon(s)|} \bullet \frac{\partial}{\partial\epsilon}c'_\epsilon(s) = \mathbf{t}_\epsilon(s) \bullet \varphi'(s)\,,$$

hence

$$\frac{\partial}{\partial\epsilon}|c'_\epsilon(s)|\Big|_{\epsilon=0} = c'(s) \bullet \varphi'(s)\,.$$

Since

$$k_\epsilon(s) = \frac{1}{|c'_\epsilon(s)|^3}\det[c'_\epsilon(s)\,|\,c''_\epsilon(s)],$$

we have

$$\frac{\partial}{\partial\epsilon}k_\epsilon = \frac{1}{|c'_\epsilon|^3}\Big(\det\Big[\frac{\partial}{\partial\epsilon}c'_\epsilon\,\Big|\,c''_\epsilon\Big] + \det\Big[c'_\epsilon\,\Big|\,\frac{\partial}{\partial\epsilon}c''_\epsilon\Big]\Big) - 3\frac{1}{|c'_\epsilon|^4}\det\Big[c'_\epsilon\,\Big|\,c''_\epsilon\Big] - \frac{\partial}{\partial\epsilon}|c'_\epsilon|,$$

hence, if $k(s) := k_0(s) = k_c(s)$ and we recall that $c_0(s)$ is parametrized by the arc length,

$$\frac{\partial}{\partial\epsilon}k_\epsilon\Big|_{\epsilon=0} = \Big(\det[\varphi'\,|\,c''] + \det[c'\,|\,\varphi'']\Big) - 3k\,c' \bullet \varphi'$$

concluding that

$$\delta\mathcal{F}(c,\varphi) = \int_0^L \Big[f'(k)\Big(\det[c'\,|\,\varphi''] + \det[\varphi'\,|\,c'] - 3k\,c' \bullet \varphi'\Big) + f(k)\,c' \bullet \varphi'\Big]\,ds,$$

where, of course, $k = k(s)$, $c' = c'(s)$ and $\varphi' = \varphi'(s)$.

It is easily seen that for *tangential variations*, $\varphi(s) := \zeta(s)\mathbf{t}(s)$, we have $\delta\mathcal{F}(c,\varphi) = 0$, as it is clear from the invariance of $\mathcal{F}$ with respect to reparametrizations of c. Instead, for *normal variations*, $\varphi(s) = \zeta(s)\mathbf{n}(s)$, we find, after an integration by parts (ζ vanishes near the boundary)

$$\begin{aligned}\delta\mathcal{F}(c,\varphi) &= \int_0^L \Big(f'(k)\zeta'' + (f'(k)k^2 - f(k)k)\zeta\Big)\,ds \\ &= \int_0^L \Big(\frac{d^2}{ds^2}f'(k) + f'(k)k^2 - f(k)k\Big)\zeta\,ds \\ &= \int_0^L \Big(f'''(k)k'^2 + f''(k)k'' + f'(k)k^2 - f(k)k\Big)\zeta\,ds.\end{aligned}$$

Let c be an extremal of the functional (3.12), and suppose that it is sufficiently regular so that its curvature is of class C^2. Then the Euler–Lagrange equation in weak form is

$$\int_0^L \Big(f'''(k)k'^2 + f''(k)k'' + f'(k)k^2 - f(k)k \Big)\zeta \, ds = 0$$

for all $\zeta \in C_c^\infty(]0, L[, \mathbb{R})$ and the Euler–Lagrange equation in strong form for the functional (3.12) is

$$f''(k)k'' + f'''(k)k'^2 + k(kf'(k) - f(k)) = 0. \tag{3.13}$$

In principle, we can now find all extremals of the functional (3.12) by first integrating (3.13) and finding $k(s)$ and then $c(s)$ from $c'(s) = \mathbf{t}(s)$ after solving the linear ODE system with variable coefficients

$$\begin{cases} \mathbf{t}(s) = k(s)\mathbf{n}(s), \\ \mathbf{n}'(s) = -k(s)\mathbf{t}(s). \end{cases}$$

3.16 Example. Of interest is the simple integral

$$\mathcal{F}(c) := \int_0^L (k_c^2 + \lambda)\, ds$$

where λ is a real constant. It is related to the so-called *elastic lines* studied by Euler. From (3.13) its Euler–Lagrange equation is

$$2k'' + k^3 - \lambda k = 0.$$

By multiplying by k', we infer that there is a constant μ such that

$$k'^2 + \frac{1}{4}k^4 - \frac{\lambda}{2}k^2 = \mu;$$

it follows that $s = s(k)$ is given by the elliptic integral

$$s = \int \frac{dk}{\sqrt{\mu + \frac{\lambda}{2}k^2 - \frac{1}{4}k^4}}$$

that, passing to the inverse function, yields $k = k(s)$.

3.1.2 Some remarks on the existence and regularity of minimizers

As we have already remarked several times, the existence of a (even unique) critical point does not ensure the existence of a minimizer (even for critical points of functions of one variable).

The first question to be dealt with is, therefore, that of the existence of a minimizer for a functional of the type

$$\mathcal{F}(u) = \int_\Omega F(x, u, Du)\, dx$$

in suitable classes $\mathcal{C}$ of admissible functions. In the terminology of Lebesgue, this can be done in two different ways:

(i) By means of *indirect methods*: By finding one or more candidates, for instance the extremals of $\mathcal{F}$, and by means of methods that in suitable situations allow us to conclude that an extremal is actually a minimizer by a direct comparison of competing functions with our candidate.
(ii) By means of *direct methods*: Proving directly, by means of qualitative theorems such as Weierstrass's theorem, that under suitable conditions on the integrand and the class of competing functions there exists a minimizer; this requires the use of topological notions (such as convergence and compactness) for the class of competing functions and of continuity or semicontinuity information for the functional with respect to the introduced notions of convergence, and the choice of competing functions.

a. Existence

The indirect methods rely on the credibility that it is relatively easy to find extremals. This is plausible at least for equations and in dimension one, $n = N = 1$. In fact, in this case the Euler–Lagrange equation is

$$F_{pp}u'' + F_{pu}u' + F_{px} - F_u = 0,$$

that, under the assumption

$$F_{pp}(x, u, p) \neq 0 \qquad \forall (x, u, p),$$

rewrites as

$$u'' = f(x, u, u').$$

As we have seen in [GM3], a result that provides us with a critical point is then the following.

3.17 Theorem (Bernstein). [1] *Let $f(x, u, p)$ be a function of class C^1 in $[a, b] \times \mathbb{R}^N \times \mathbb{R}^N$. Suppose there exist two nonnegative functions $A(x, u)$ and $B(x, u)$ and a constant $k > 0$ such that*

$$f_u(x, u, p) > k, \qquad |f(x, u, p)| \leq A(x, u)|p|^2 + B(x, u).$$

Then, for all $\alpha, \beta \in \mathbb{R}^N$, the problem

$$\begin{cases} u''(x) = f(x, u(x), u'(x)) & \text{in }]a, b[, \\ u(a) = \alpha,\ u(b) = \beta \end{cases}$$

has a solution.

[1] S. N. Bernstein, *Sur les équations du calcul des variations*, Ann. Sci. Ec. Norm. Sup., Paris **29** (1912) 431–485.

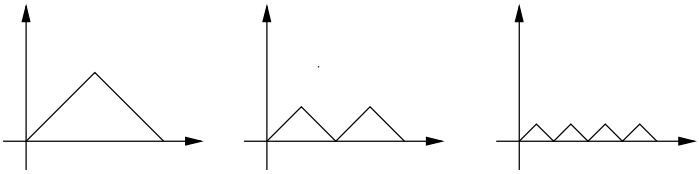


Figure 3.8. In the figure we have plotted the first three terms of a sequence of functions $\{u_n\}$ that converges uniformly to the function that is identically zero, but $\mathcal{F}(u_n) = 0$ $\forall n$ while $\mathcal{F}(0) = 1$, $\mathcal{F}$ being the functional of Example 3.18.

Having at our disposal one or more than one critical point, in order to decide whether one of them is a minimizer for $\mathcal{F}$ we may use the classical and elegant theory of *second variation* or of *conjugate points* of Carl Jacobi (1804–1851) or of *extremal fields* of Karl Weierstrass (1815–1897), that, incidentally, plays an important role in differential geometry. However, we do not pursue this here.

The indirect approach, the formalism of which is extremely important also for multiple integrals, $n > 1$, loses ground for multidimensional problems since, in this case, it is less clear how one can prove the existence of solutions of the Euler–Lagrange equation; on the contrary, a direct proof of the existence of a minimizer might instead provide a simple proof of the existence of critical points: The Dirichlet principle that we discuss in Chapter 1 is a prototype.

The direct methods of calculus of variations, that begin in the works of Leonida Tonelli (1885–1946) and have further development in the works of Charles Morrey (1907–1984), are still an important area of research. Their use requires the enlargement of the class of competing functions beyond the smooth ones among which we seek a minimizer and a related extension of the functional on the enlarged class, for instance, as we have seen, Sobolev classes for the Dirichlet principle. The so-called *regularity problem* of generalized solutions then naturally arises.

We have no chance here to illustrate those questions, and we refer the reader to the bibliographical appendix, confining ourselves to the discussion of a few simple examples.

3.18 Example (Lebesgue's sequence). Consider the problem

$$\mathcal{F}(u) := \int_0^1 (u'^2 - 1)^2 \, dx \to \min, \qquad u(0) = 0,\ u(1) = 0.$$

It is not difficult to see that

$$\inf\Big\{\mathcal{F}(u) \,\Big|\, u \in C^1([0,1]),\ u(0) = 0 = u(1) = 1\Big\} = 0;$$

however, the infimum is not taken, i.e., the above problem has no solution of class C^1. In fact, at functions with piecewise slope 1 or -1 we have $\mathcal{F}(u) = 0$, but $\mathcal{F}$ never vanishes at functions of class C^1 that vanish at 0 and 1. Moreover, minimizing sequences do not necessarily converge to a minimum u.

3.19 Example (Weierstrass's example). Consider the problem

$$\mathcal{F}(u) := \int_{-1}^{1} x^2 u'^2 \, dx \to \min, \qquad u(-1) = -1,\ u(1) = 1,$$

and, for $\epsilon > 0$,

$$u_\epsilon(x) := \frac{\arctan \frac{x}{\epsilon}}{\arctan \frac{1}{\epsilon}}$$

or

$$u_\epsilon(x) := \begin{cases} -1 & \text{if } x \le -\epsilon, \\ \dfrac{x}{\epsilon} & \text{if } |x| \le \epsilon, \\ 1 & \text{if } x \ge \epsilon. \end{cases}$$

The family $\{u_\epsilon\}$ is admissible and $\mathcal{F}(u_\epsilon) \to 0$, hence

$$\inf\Big\{\mathcal{F}(u) \,\Big|\, u \in C^1([-1,1]),\ u(-1) = -1,\ u(1) = 1\Big\} = 0,$$

but, clearly, there is no such admissible function with zero energy.

b. Regularity in the 1-dimensional case

We shall not discuss the regularity problem in the multidimensional case; here we only illustrate a regularity theorem in the 1-dimensional case.

In order to write the Euler–Lagrange equation in strong form, of course it suffices that the extremal be of class C^2, at least in the classical context. However, there is no reason why an extremal or even a minimizer of class C^1 should be of class C^2.

3.20 Example. The function $u(x) = x|x|$, that is of class C^1 but not of class C^2, clearly solves the problem

$$\mathcal{F}(u) := \int_{-1}^{1} (u' - 2|x|)^2 \, dx \to \min, \qquad u(-1) = -1,\ u(1) = 1.$$

3.21 Example. The function

$$u(x) := \begin{cases} 0 & \text{if } x \in [-1, 0], \\ x^2 & x \in [0, 1], \end{cases}$$

that is of class C^1 but not of class C^2, clearly solves the problem

$$\mathcal{F}(u) := \int_{-1}^{1} u^2(x)(u' - 2x)^2 \, dx \to \min, \qquad u(-1) = 0,\ u(1) = 1.$$

However, we have the following.

3.22 Proposition. *If $u \in C^1([a,b], \mathbb{R}^N)$ is an extremal of the functional*

$$\mathcal{F}(u) := \int_\Omega f(x, u(x), u'(x)) \, dx,$$

i.e., if


Figure 3.9. Eugenio Beltrami (1835–1899) and Paul du Bois–Reymond (1831–1889).

$$\int_a^b (F_p\varphi' + F_u\varphi)\,dx = 0$$

for all $\varphi \in C^1([a,b],\mathbb{R}^N)$ *with* $\varphi(a) = \varphi(b) = 0$, *then*

$$x \to F_p(x, u(x), u'(x))$$

is of class $C^1([a,b],\mathbb{R}^N)$.

Proof. Integrating by parts the second term of the Euler–Lagrange equation we get

$$\begin{aligned} 0 &= \int_a^b \left\{ F_p\varphi' + \frac{d}{dx}\left(\int_a^x F_u\,dt\ \varphi(x) \right) - \int_a^x F_u\,dt\ \varphi'(x) \right\} dx \\ &= \int_a^b \left(F_p - \int_a^x F_u\,dt \right) \varphi'(x)\,dx. \end{aligned}$$

Using *du Bois–Reymond lemma*, see Chapter 1, we infer

$$F_p(x, u(x), u'(x)) = \int_a^x F_u(t, u(t), u'(t))\,dt + const, \tag{3.14}$$

in $]a,b[$, hence in $[a,b]$ by continuity. The fundamental theorem of calculus then yields the result since $x \mapsto F_u(x, u(x), u'(x))$ is continuous in $[a,b]$. □

Notice that we are not allowed to compute the derivative of the function $x \mapsto F_p(x, u(x), u'(x))$ using the chain rule, since, in general, u' is not differentiable a priori. However, the following holds.

3.23 Theorem (Regularity). *Let* $F(x,u,p)$ *be an integrand of class* C^2 *and let* u *be an extremal of class* C^1 *of the functional*

$$\mathcal{F}(u) = \int_a^b F(x, u, u')\,dx.$$

If $\det F_{pp}(x, u(x), u'(x)) \neq 0\ \forall x \in]a,b[$, *then* $u \in C^2(]a,b[,\mathbb{R}^N)$.

Proof. Introduce the function $\phi :]a,b[\times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ given by

$$\phi(x,z,p,q) := F_p(x,z,p) - q.$$

Since $\det \frac{\partial \phi}{\partial p} = \det F_{pp} \neq 0$, the implicit function theorem applied to the equation $\phi(x,z,p,q) = 0$ tells us that for all $x_0 \in]a,b[$ there is a neighborhood U of (x_0, z_0, p_0, q_0), $z_0 = u(x_0)$, $q_0 = F_p(x_0, u(x_0), u'(x_0))$ and a map φ of class C^1 such that $\phi(x,z,\varphi(x,z,q),q) = 0$ $\forall (x,z,p,q) \in U$. Since for x close to x_0 we have $(x, u(x), F_p(x, u(x), u'(x))) \in U$, we infer that

$$u'(x) = \varphi(x, u(x), F_p(x, u(x), u'(x))).$$

This implies at once that u' is of class C^1, since φ, $u(x)$ and $x \to F_p(x, u(x), u'(x))$ are of class C^1, see Proposition 3.22. □

The previous theorem extends to extremals that are absolutely continuous.

3.24 Theorem (Regularity). *Let $F(x,u,p)$ be an integrand of class C^2 and let u be an extremal that is absolutely continuous of the functional*

$$\mathcal{F}(u) = \int_a^b F(x,u,u')\,dx$$

such that $F_p(x,u(x),u'(x))$ and $F_u(x,u(x),u'(x))$ are summable. If

$$\det F_{pp}(x,u(x),u'(x)) \neq 0 \qquad \text{for a.e. } x \in]a,b[,$$

then $u \in C^2(]a,b[, \mathbb{R}^N)$.

Proof. As in Proposition 3.22, the du Bois–Reymond lemma yields that (3.14) holds for a.e. x. Hence $x \to F_p(x,u(x),u'(x))$ is absolutely continuous. As in the proof of Theorem 3.23 we then conclude that

$$u'(x) = \varphi(x, u(x), F_p(x, u(x), u'(x))) \qquad \text{a.e. } x,$$

that implies that u' is absolutely continuous, in particular, u' agrees a.e. with a continuous function $v(x)$. Summing up, for $x_0 \in]a,b[$ we have

$$u(x) = u(x_0) + \int_{x_0}^x u'(s)\,ds = u(x_0) + \int_{x_0}^x v(s)\,ds,$$

hence $u \in C^1$ near x_0. Finally, Theorem 3.23 yields the conclusion. □

3.1.3 Constrained variational problems

In several instances one wants to minimize a functional among functions that ought to satisfy some constraints besides the boundary conditions. Here we write the vanishing of the first variation in two interesting cases: The so-called *Lagrange multipliers* appear in this context.

a. Isoperimetric constraints

3.25 Theorem. *Let u be a minimizer of the functional*

$$\mathcal{F}(u) = \int_a^b F(x, u, Du)\, dx$$

among the functions u that satisfy the constraints

$$\mathcal{G}_k(u) := \int_a^b G_k(x, u, Du)\, dx = c_k, \qquad k = 1, \dots, r$$

and suitable boundary conditions. Suppose that there exist functions ψ_1, $\dots$, ψ_r of class $C^1([a,b], \mathbb{R}^N)$ vanishing at a and b such that the matrix

$$\Big[\delta \mathcal{G}_k(u, \psi_\ell)\Big]$$

has maximal rank r. Then there exists $\lambda := (\lambda^1, \lambda^2, \dots, \lambda^r)$ such that u is an extremal of the functional

$$\mathcal{F}(u) + \sum_{k=1}^{r} \lambda^k \mathcal{G}_k(u).$$

Proof. For the sake of simplicity we confine ourselves to the case $r = 1$, setting $\mathcal{G} := \mathcal{G}_1$. Let $\varphi \in C^1([a,b], \mathbb{R}^N)$ vanish at a and b, and let $\psi \in C^1$ be such that $\delta\mathcal{G}(u, \psi) = 1$. For ϵ and t in a neighborhood of 0 define

$$\Phi(\epsilon, t) := \mathcal{F}(u + \epsilon\varphi + t\psi), \qquad \Psi(\epsilon, t) := \mathcal{G}(u + \epsilon\varphi + t\psi).$$

Trivially, $(0,0)$ is a minimum point of Φ with the constraint $\Psi(\epsilon, t) = c_1$. The Lagrange multiplier theorem, see Theorem 5.62 of [GM4], yields the existence of $\lambda \in \mathbb{R}$ such that

$$\begin{cases} \Phi_\epsilon(0,0) + \lambda \Psi_\epsilon(0,0) = 0, \\ \Phi_t(0,0) + \lambda \Psi_t(0,0) = 0 \end{cases}$$

that is,

$$\delta(\mathcal{F} + \lambda\mathcal{G})(u, \varphi) = 0, \qquad \delta(\mathcal{F} + \lambda\mathcal{G})(u, \psi) = 0,$$

i.e., u is an extremal of $\mathcal{F} + \lambda\mathcal{G}$ with $\lambda := -\delta\mathcal{F}(u, \psi)$. □

3.26 Isoperimetric problem. Among graphs in $[a,b] \times \mathbb{R}_+$ with prescribed boundary that enclose a given area

$$\int_a^b u(x)\, dx = A,$$

find the one of minimal length,

$$\int_a^b \sqrt{1 + u'^2}\, dx \to \min.$$

Assume that the minimum point u exists and is regular; then u solves

$$\left(\frac{u'}{\sqrt{1+u'^2}}\right)' = const$$

i.e., the possible solution has to have constant curvature. Similarly, the possible solutions of the problem of minimizing the area of graph of u

$$\int_\Omega \sqrt{1+|Du|^2}\,dx$$

among the functions $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ with prescribed boundary conditions and with prescribed integral

$$\int_\Omega u(x)\,dx = A,$$

solve the equation

$$\sum_{\alpha=1}^{n} D_\alpha \frac{D_\alpha u}{\sqrt{1+|Du|^2}} = const,$$

i.e., have graphs with constant mean curvature, see Chapter 5 of [GM4].

3.27 First eigenvalue of the Laplacian. The solution of the problem

$$\begin{cases} \frac{1}{2}\int_\Omega |Du|^2\,dx \to \min, \\ u = 0 \qquad\qquad \text{on } \partial\Omega, \\ \int_\Omega u^2\,dx = 1 \end{cases}$$

satisfies the equation

$$\Delta u + \lambda u = 0 \qquad \text{in } \Omega,$$

where λ is a constant, in fact, the first eigenvalue of the Laplacian.

3.28 Elastic lines. Elastic lines were studied by Robert Hooke (1635–1703), Gottfried W. von Leibniz (1646–1716) and Leonhard Euler (1707–1783) with the methods of the calculus of variations. In the formulation of Euler, the problem amounts to minimizing the curvature functional

$$\int_\gamma k^2\,ds \to \min$$

under the length constraint $\int_\gamma ds = L$. This leads to looking at the extremals of the functional

$$\int_\gamma (k^2+\lambda)\,ds,$$

the Euler–Lagrange equation we discussed in Example 3.15.

b. Holonomic constraints

Suppose we want to minimize the action

$$\int_a^b L(x, u, u')\, dx$$

of n material points among motions that take place on a surface, for instance, implicitly defined by the system of equations $G(z) = 0$.

More generally, suppose we want to minimize the integral

$$\int_\Omega F(x, u, Du)\, dx$$

among maps $u : \Omega \to \mathbb{R}^N$ that take values on a $(N - r)$-dimensional submanifold $\mathcal{Y}$ of a suitable $\mathbb{R}^N$, defined implicitly by the equation $G(z) = 0$, where $G : \mathbb{R}^n \to \mathbb{R}^r$, $1 \le r < N$, is of class C^1 with Jacobian matrix of maximal rank r in $u(\Omega)$.

Let $\psi(x, t)$ be a family of admissible variations for $u : \Omega \to \mathbb{R}^N$, $\psi(x, 0) = u(x)$. Then $G(\psi(x, t)) = 0\ \forall (x, t)$ and, differentiating in t and setting

$$\varphi(x) := \frac{\partial \psi}{\partial t}(x, 0),$$

we have $\mathbf{D}G(u(x))\varphi(x) = 0$, i.e.,

$$\varphi(x) \in \ker \mathbf{D}G(u(x)) = \mathrm{Tan}_{u(x)}\, \mathcal{Y},$$

as the tangent space $\mathrm{Tan}_z\, \mathcal{Y}$ at z to $\mathcal{Y} = \{G(z) = 0\}$ is $\ker \mathbf{D}G(z)$; we simply say that φ *is tangent to* $\mathcal{Y}$ *along* u.

3.29 Proposition. *Let $G : \mathbb{R}^n \to \mathbb{R}^r$, $1 \le r < N$, be a map of class C^1 with Jacobian matrix of maximal rank r at the points of $\mathcal{Y} := \{z \,|\, G(z) = 0\}$. If $u \in C^1(\Omega, \mathcal{Y})$ is a minimizer of the functional $\mathcal{F}(u)$ among maps with values in $\mathcal{Y}$, then u solves the Euler–Lagrange equations that in this case take the form*

$$\begin{cases} G(u(x)) = 0, \\ \displaystyle\int_\Omega (F_{u^i} - D_\alpha F_{p^i_\alpha})\psi^i\, dx = 0 \quad \forall \psi \in C^1_c(\Omega, \mathbb{R}^N) \text{ tangent to } \mathcal{Y} \text{ along } u. \end{cases} \tag{3.15}$$

Moreover, if u is of class C^2, then u satisfies (3.15) *if and only if*

$$\begin{cases} G(u(x)) = 0, \\ \left(F_{u^i} - D_\alpha F_{p^i_\alpha}\right) \perp \mathrm{Tan}_{u(x)}\, \mathcal{Y}, \end{cases} \qquad \forall x \in \Omega. \tag{3.16}$$

Proof. We prove the proposition under the extra assumption that $\mathcal{Y}$ be of class $C^2(\Omega)$.[2]

Recall that every submanifold $\mathcal{Y} \subset \mathbb{R}^N$ of class C^2 has a neighborhood U with a projection $\pi : U \to \mathcal{Y}$ that maps a point $z \in U$ uniquely into $\pi(z) \in U$, the foot of the perpendicular through z to $\mathcal{Y}$. The map π is of class C^1, its tangent map $d\pi(z)$ has $\mathrm{Tan}_{\pi(z)}\,\mathcal{Y}$ as image and $\mathrm{Im}\,\mathbf{D}\pi(z) = \mathrm{Tan}_{\pi(z)}\,\mathcal{Y}$, see Chapter 5 of [GM4].

Let $\zeta \in C^1_c(\Omega, \mathbb{R}^N)$. Since the support of ζ is compact, there is $\epsilon_0 > 0$ such that for $|\epsilon| < \epsilon_0$ we have $u(x) + \epsilon\zeta(x) \in U\ \forall x \in U$. The function

$$\psi(x, \epsilon) := \pi(u(x) + \epsilon\zeta(x)) \tag{3.17}$$

is then an admissible variation, i.e., $G(\psi(x, \epsilon)) = 0$ and $\psi(x, 0) = u(x)$, and

$$\frac{\partial \psi}{\partial \epsilon}(x, 0) = \mathbf{D}\pi(u(x))\zeta(x) \qquad \forall x \in \Omega.$$

In particular, $\varphi(x) := \frac{\partial \psi}{\partial \epsilon}(x, 0) \in \mathrm{Tan}_{u(x)}\,\mathcal{Y}$, i.e., φ is tangent to $\mathcal{Y}$ along u; moreover, $\varphi(x) = \zeta(x)$ if ζ is tangent to $\mathcal{Y}$ along u.

(i) Let u be a minimizer of $\mathcal{F}$ constrained to $\mathcal{Y}$, ζ tangent to $\mathcal{Y}$ along u and $\psi(x, \epsilon)$ be defined by (3.17), then, as we have seen, $\frac{\partial \psi}{\partial \epsilon} = \zeta$. The function

$$\epsilon \to \mathcal{F}(\psi(\cdot, \epsilon))$$

has a minimizer at $\epsilon = 0$. Differentiating under the integral sign, Fermat's theorem yields

$$0 = \frac{d}{d\epsilon}\mathcal{F}(\psi(\cdot, \epsilon))\Big|_{\epsilon=0} = \int_\Omega \left(F_{u^i}\zeta^i + F_{p^i_\alpha} D_\alpha \zeta^i\right) dx.$$

(ii) If, moreover, $u \in C^2(\Omega)$, an integration by parts yields

$$\int_\Omega \left(F_{u^i} - D_\alpha F_{p^i_\alpha}\right)\zeta^i\, dx = 0$$

for all ζ tangent to $\mathcal{Y}$ along u. In particular,

$$\int_\Omega \left(F_{u^i} - D_\alpha F_{p^i_\alpha}\right) D_j \pi^i(u(x))\varphi^j(x)\, dx = 0$$

for all $\varphi \in C^1_c(\Omega, \mathbb{R}^N)$. From the fundamental lemma of calculus of variations we infer

$$\left(F_{u^i} - D_\alpha F_{p^i_\alpha}\right) D_j \pi^i(u(x)) = 0,$$

i.e., the vector $(F_{u^i} - D_\alpha F_{p^i_\alpha})$ is perpendicular to $\mathrm{Im}\,\mathbf{D}\pi(u(x)) = \mathrm{Tan}_{u(x)}\,\mathcal{Y}$. □

More generally, we can state the following.

[2] The same proof works if $\mathcal{Y}$ is of class C^1 using a deep result of Hassler Whitney (1907–1989): *If $\mathcal{Y}$ is a submanifold of class C^1, then there is an open set $U \supset \mathcal{Y}$ and a map of retraction $\pi : U \to \mathcal{Y}$, i.e., such that $\pi(U) = \mathcal{Y}$ and $\pi_{|\mathcal{Y}} = \mathrm{Id}_{|\mathcal{Y}}$ of class C^1.* Actually, the following holds, see H. Federer, *Geometric Measure Theory*, Springer–Verlag, New York, 1969, Theorem 3.1.20,

Theorem (Federer–Whitney). *Let $B \subset \mathbb{R}^n$ be connected and let $k \geq 1$. B is a submanifold of class C^k if and only if there are an open set U and a map π of class C^k that retracts U onto B.*

3.30 Theorem. *Let u be an extremal of class C^1 of the functional*

$$\mathcal{F}(u) := \int_\Omega F(x, u(x), Du(x))\, dx$$

constrained to

$$\mathcal{C} := \{u : \Omega \to \mathbb{R}^N \,\Big|\, G(x, u(x)) = 0,\ x \in \Omega\},$$

where $G(x,u) : \Omega \times \mathbb{R}^N \to \mathbb{R}^r$ denotes a map of class C^1, with Jacobian matrix with respect to u of maximal rank r on $\{(x, u(x)),\ x \in \Omega\}$. Then there exist continuous functions $\lambda^1(x), \dots, \lambda^r(x)$ such that u is an extremal of the functional

$$\int_\Omega \Big(F(x, u(x), Du(x)) + \sum_{k=1}^{r} \lambda^k(x) G^k(x, u(x))\Big)\, dx.$$

3.31 Example (Force of constraints). In the case of the action

$$\int_{t_1}^{t_2} \Big(\frac{1}{2} m X'^2 - V(X)\Big)\, dt, \qquad G^j(t, X) = 0,\ j = 1, \dots, r,$$

we infer from Theorem 3.30 that Euler–Lagrange equations are

$$mX''(t) = \nabla V(X(t)) + \sum_{k=1}^{r} \lambda^k(t) \nabla G^k(t, X(t)),$$

where the last term is physically interpreted as the force exerted by the constraint. The multipliers can be computed from $G(t, X(t)) = 0$ differentiating twice in t as in Examples 3.32 and 3.33.

3.32 Example (Harmonic maps into S^{N-1}). The extremal $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}^N$ of the problem

$$\frac{1}{2} \int_\Omega |Du|^2\, dx \to \min, \qquad |u|^2 = 1,$$

are the solutions of the equations

$$-\Delta u + \mu(x) u = 0, \qquad |u|^2 = 1, \tag{3.18}$$

and we can easily compute that Lagrange's multiplier μ is given by $-|Du|^2$. In fact, from $|u|^2 = 1$ we find by differentiation $\sum_{i=1}^N u^i Du^i = 0$ and, differentiating again, $u \bullet \Delta u + |Du|^2 = 0$. Comparing the last equation with (3.18), we conclude that $\mu(x) = -|Du|^2$.

In conclusion, the extremals of the Dirichlet integral for maps with values in a sphere solve the equations

$$-\Delta u = u |Du|^2.$$

3.33 Example (Harmonic maps into a manifold). The extremals $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}^N$ of the problem

$$\frac{1}{2} \int_\Omega |Du|^2\, dx \to \min, \qquad G(u(x)) = 0,$$

where $G : \mathbb{R}^N \to \mathbb{R}^r$ is a smooth map with Jacobian matrix of maximal rank on $\{y \,|\, G(y) = 0\}$, solve the equations

$$-\Delta u^i + \sum_{k=1}^{r} \lambda^k(x) D_i G^k = 0, \qquad G(u(x)) = 0, \tag{3.19}$$

where $\lambda^1, \lambda^2, \dots, \lambda^r$ are suitable functions that can be computed as follows. Differentiating twice in x the relation $G(u(x)) = 0$ we get

$$D_i G^k(u(x))\Delta u^i + E^k(x) = 0 \qquad k = 1, \dots, r,$$

where

$$E^k(x) := \sum_{i,j=1}^{N} \sum_{\alpha=1}^{n} \frac{\partial^2 G^k}{\partial y^i \partial y^j}(u(x)) D_\alpha u^i D_\alpha u^j.$$

Comparing with (3.19) and writing in vector notation $E(x) := (E^1, E^2, \dots, E^r)$, $\lambda := (\lambda^1, \lambda^2, \dots, \lambda^r)$ and $\Delta u := (\Delta u^1, \dots, \Delta u^N)$, we find

$$E = -\mathbf{D}G\Delta u = \mathbf{D}G\mathbf{D}G^T\lambda,$$

i.e.,

$$\lambda(x) = \left(\mathbf{D}G(u(x))\mathbf{D}G(u(x))^T\right)^{-1} E(x).$$

3.34 The general Dirichlet integral. Let $\mathcal{X}$ and $\mathcal{Y}$ be two submanifolds of respectively two Euclidean spaces $\mathbb{R}^m$ and $\mathbb{R}^N$, and let $U : \mathcal{X} \to \mathcal{Y}$ be a smooth map thought of as a map with values in $\mathbb{R}^N$ with image into $\mathcal{Y}$. The differential of U at x maps the tangent space $T_x\mathcal{X}$ of $\mathcal{X}$ at x into the tangent space $T_{U(x)}\mathcal{Y}$ of $\mathcal{Y}$ at $U(x)$. Since $T_x\mathcal{X}$ and $T_y\mathcal{Y}$ are naturally endowed with the inner products of the ambient spaces, respectively $\mathbb{R}^m$ and $\mathbb{R}^N$, the adjoint map $dU_x^* : T_{U(x)}\mathcal{Y} \to T_x\mathcal{X}$ is well-defined. We can therefore define an inner product and a norm on the space of linear maps from $T_x\mathcal{X}$ into $T_{U(x)}\mathcal{Y}$ as

$$\mathbf{A} \bullet \mathbf{B} := \operatorname{tr} \mathbf{B}^*\mathbf{A} \qquad \text{and} \qquad |\mathbf{A}| := \sqrt{\mathbf{A}^*\mathbf{A}}.$$

The *energy density of* U is then

$$e(U)(x) := \frac{1}{2}|dU_x|^2$$

and, by means of the volume element of $\mathcal{X}$, we define the *generalized Dirichlet energy* as

$$\mathcal{D}(U, \mathcal{X}) = \mathcal{D}(U) := \int_{\mathcal{X}} e(U)\, d\mathcal{H}^n.$$

In this way the Dirichlet energy is defined independently of the chosen coordinates. Notice that in orthonormal coordinates, for maps $U : \Omega \subset \mathbb{R}^n \to \mathcal{Y} \subset \mathbb{R}^N$ we have

$$e(U)(x) = \frac{1}{2}|DU(x)|^2, \qquad \mathcal{D}(U, \Omega) = \frac{1}{2}\int_\Omega |DU|^2\, dx.$$

Since the following considerations are local and U is continuous, it is not restrictive to choose local coordinates in $\mathcal{X}$ and $\mathcal{Y}$ respectively, i.e., diffeomorphisms $\varphi : B(0,1) \subset \mathbb{R}^n \to \mathcal{X}$ and $\psi : B(0,1) \subset \mathbb{R}^N \to \mathcal{Y}$.

Denote the local coordinates by $x = (x^1, x^2, \dots, x^n)$ in $\mathbb{R}^n$ and by $y = (y^1, y^2, \dots, y^N)$ in $\mathbb{R}^N$. Set $u := \psi^{-1} \circ U \circ \varphi$ and let

$$\Gamma = (\gamma_{\alpha\beta}(x)) := \frac{\partial \varphi}{\partial x^\alpha} \bullet \frac{\partial \varphi}{\partial x^\beta}, \qquad G = (g_{ij}(y)) := \frac{\partial \psi}{\partial y^i} \bullet \frac{\partial \psi}{\partial y^j}$$

be respectively the metric tensors in $\mathrm{Tan}_x \mathcal{X}$ and $\mathrm{Tan}_y \mathcal{Y}$. Then dU_x is represented by the matrix $\mathbf{D}u(x)$ and dU_x^* is represented by the matrix

$$\Gamma^{-1}(x)\mathbf{D}u(x)^T G(u(x)),$$

consequently,

$$e(U) = \frac{1}{2}\gamma^{\alpha\beta} g_{ij}(u(x)) \frac{\partial u^i}{\partial x^\alpha} \frac{\partial u^j}{\partial x^\beta}$$

where $(\gamma^{\alpha\beta}) = (\gamma_{\alpha\beta})^{-1}$. Since $d\mathcal{H}^n = \sqrt{\gamma}\, dx$, $\gamma := \det(\gamma_{\alpha\beta})$, we conclude that in local coordinates the energy takes the form

$$\begin{aligned} \mathcal{D}(U, \mathcal{X}) &= \mathcal{D}(u) \\ &:= \frac{1}{2} \int_{B(0,1)} \gamma^{\alpha\beta}(x) g_{ij}(u(x)) \frac{\partial u^i}{\partial x^\alpha}(x) \frac{\partial u^j}{\partial x^\beta}(x) \sqrt{\gamma(x)}\, dx. \end{aligned} \tag{3.20}$$

Notice that in the standard case, $\mathcal{X} = \Omega \subset \mathbb{R}^n$ and $\mathcal{Y} = \mathbb{R}^N$, choosing orthonormal coordinates to write u, then $u = U$ and $\Gamma = \Gamma^{-1} = \mathrm{Id}$; hence $\mathcal{D}(u)$ is the standard Dirichlet energy.

3.35 ¶. Write the Dirichlet integral for maps $u : B(0,1) \subset \mathbb{R}^2 \to \mathbb{R}$ in polar coordinates and for maps $u : B(0,1) \subset \mathbb{R}^3 \to \mathbb{R}$ in spherical coordinates.
[*Hint.* Use (3.20).]

Let us write Euler–Lagrange equations for the functional in (3.20). For every $\varphi : B(0,1) \to \mathbb{R}^N$ of class C^1, that vanishes on the boundary of $B(0,1)$ and for every $t \in \mathbb{R}$, the maps $u + t\varphi$ are admissible. Thus, differentiating under the integral sign, one easily deduces

$$\begin{aligned} &\delta\mathcal{D}(u, \varphi) \\ &\quad = \int_{B(0,1)} \Big(\gamma^{\alpha\beta} g_{ij} D_\alpha u^i D_\beta \varphi^j\, dx + \frac{1}{2} \gamma^{\alpha\beta} g_{ij,\ell}(u) D_\alpha u^i D_\beta u^j \varphi^\ell \Big) \sqrt{\gamma}\, dx, \end{aligned}$$

and, integrating by parts (assuming u to be of class C^2), one gets that the minimizers of the Dirichlet integral written in local coordinates solve the system of PDE's

$$\begin{aligned} &\frac{1}{\sqrt{\gamma}} D_\beta (\gamma^{\alpha\beta} \sqrt{\gamma} g_{ki} D_\alpha u^i) \\ &\qquad - \frac{1}{2} \gamma^{\alpha\beta} g_{ij,k}(u) D_\alpha u^i D_\beta u^j = 0 \qquad \forall k = 1, \dots, N. \end{aligned} \tag{3.21}$$

System (3.21) can be written equivalently as

$$g_{ik}\frac{1}{\sqrt{\gamma}}D_\beta(\gamma^{\alpha\beta}\sqrt{\gamma}D_\alpha u^i) + \gamma^{\alpha\beta}\Big(g_{ik,j}(u) - \frac{1}{2}g_{ij,k}(u)\Big)D_\alpha u^i D_\beta u^j = 0 \qquad \forall k = 1,\dots,N. \tag{3.22}$$

Notice that the following hold:

(i) The first term in (3.21) is Laplace–Beltrami's operator on $\mathcal{X}$ applied to u, see Chapter 5 of [GM4],

$$\frac{1}{\sqrt{\gamma}}D_\beta(\gamma^{\alpha\beta}\sqrt{\gamma}g_{ki}D_\alpha u^i) = \operatorname{div}_{\mathcal{X}}(\nabla_{\mathcal{X}}u^k) = \Delta_{\mathcal{X}}u^k.$$

(ii) The second term in (3.21) vanishes if the metric of $\mathcal{Y}$ is constant.

(iii) Introducing *Christoffel symbols of the first kind* of $\mathcal{Y}$

$$\Gamma_{ikj}(u) := \frac{1}{2}\Big(g_{kj,i}(u) - g_{ij,k}(u) + g_{ik,j}(u)\Big),$$

(3.22) becomes

$$g_{ik}\Delta_{\mathcal{X}}u^i + \gamma^{\alpha\beta}\Gamma_{ikj}(u)D_\alpha u^i D_\beta u^j = 0 \qquad \forall k = 1,\dots,N,$$

or, since (g_{ij}) is invertible,

$$\Delta_{\mathcal{X}}u^\ell + \gamma^{\alpha\beta}\Gamma^\ell_{ij}D_\alpha u^i D_\beta u^j = 0 \qquad \forall \ell = 1,\dots,N, \tag{3.23}$$

where $(g^{ij}) := (g_{ij})^{-1}$ and $\Gamma^\ell_{ij} := g^{\ell k}\Gamma_{ikj}$ denote the *Christoffel symbols of the second kind* of the metric g on $\mathcal{Y}$.

3.36 Example. If $\mathcal{X}$ is an interval of the real axis, system (3.23) takes the form

$$\frac{d^2}{dt^2}u^\ell + \Gamma^\ell_{ij}\frac{du^i}{dt}\frac{du^j}{dt} = 0,$$

that are the *geodesic equations*.

3.1.4 Noether's theorem

Let us return to the discussion of unconstrained problems. It turns out that Euler–Lagrange equations are only some of the stationary conditions for a minimizer. Here we state stationary conditions for arbitrary variations. As a consequence we state *Noether's theorem*, which yields, in particular, the stationary conditions in the presence of symmetries.


Figure 3.10. William R. Hamilton (1805–1865) and Emmy Noether (1882–1935).

a. General variations

Let Ω be a bounded open set of $\mathbb{R}^n$. For $|\epsilon| < \epsilon_0$ we consider a family of domains Ω_ϵ, that are small variation of Ω; more precisely, we consider a bounded open set U that strictly contains $\overline{\Omega}$ and, for every ϵ with $-\epsilon_0 < \epsilon < \epsilon_0$, a map

$$\eta(x,\epsilon) : U \times]-\epsilon_0, \epsilon_0[\to \mathbb{R}^N$$

with $\eta(x,0) = x$ such that $\eta_\epsilon(x) := \eta(x,\epsilon)$ is a diffeomorphism from U onto its image; then

$$\Omega_\epsilon := \eta_\epsilon(\Omega) \subset U, \qquad \overline{\Omega} \subset U_\epsilon$$

for ϵ sufficiently small. The *infinitesimal generator* of $\epsilon \to \eta(\cdot,\epsilon)$ is the function

$$\mu(x) := \frac{\partial \eta}{\partial \epsilon}(x,0), \qquad x \in U;$$

clearly, when $\epsilon \to 0$ we have

$$\eta_\epsilon(x) = x + \epsilon\mu(x) + o(\epsilon), \ \forall x \in U,$$
$$(\eta_\epsilon)^{-1}(y) = y - \epsilon\mu(y) + o(\epsilon), \ \forall y \in \overline{\Omega}.$$

Notice that, if $\lambda : U \to \mathbb{R}^n$ is a smooth map with a bounded difference quotient, then the maps $\eta_\epsilon(x) := x + \epsilon\lambda(x)$ are diffeomorphisms onto their images for ϵ sufficiently small (as one easily infers from the implicit function theorem); thus, they are variations of the identity in U with infinitesimal generator λ.

Let $u \in C^1(U)$. We consider the C^1 perturbation

$$v(y,\epsilon) : U \times]-\epsilon_0, \epsilon_0[\to \mathbb{R}^N$$

with $v(x,0) = u(x)$ given by $v(y,\epsilon) := u(\eta(x,\epsilon))$. Since for $y \in U$ we have $\eta(y,\epsilon) \in U$ for ϵ small, $|\epsilon| < \epsilon(y)$, the infinitesimal generator of the family $\{v_\epsilon\}$, $v_\epsilon(y) := v(y,\epsilon)$, is well-defined

$$\varphi(x) := \frac{\partial v}{\partial \epsilon}(x, 0)$$

and, of course, we have $v(x, \epsilon) = u(x) + \epsilon\varphi(x) + o(\epsilon)$ as $\epsilon \to 0$.

Finally, we consider an integrand $F(x, u, p)$ defined in $U \times \mathbb{R}^N \times \mathbb{R}^{nN}$ such that for $|\epsilon|$ sufficiently small the function

$$\phi(\epsilon) := \int_{\Omega_\epsilon} F(y, v(y, \epsilon), Dv(y, \epsilon))\, dy$$

is well-defined. We want to compute $\phi'(0)$.

3.37 Proposition. *With the previous notation,*

$$\begin{aligned}\phi'(0) &= \int_\Omega \Big(F_{u^i}\varphi^i + F_{p^i_\alpha} D_\alpha \varphi^i + D_\alpha(F\mu^\alpha)\Big)\, dx \\ &= \delta\mathcal{F}(u, \varphi) + \int_\Omega D_\alpha(F\mu^\alpha)\, dx \\ &= \int_\Omega \Big(- D_\alpha F_{p^i_\alpha} + F_{u^i}\Big)\varphi^i\, dx + \int_\Omega D_\alpha\Big(F\mu^\alpha + F_{p^i_\alpha}\varphi^i\Big)\, dx.\end{aligned}$$

Proof. We notice that $\det D_x\eta(x, 0) = 1$, $\det D_x\eta(x, \epsilon) > 0$, $\frac{\partial}{\partial\epsilon} \det D_x\eta(x, \epsilon)|_{\epsilon=0} = \operatorname{div}\mu(x)$ and that

$$\phi(\epsilon) = \int_\Omega F\Big(\eta(x, \epsilon), v(\eta(x, \epsilon)), D_y v(\eta(x, \epsilon))\Big) \det D_x\eta(x, \epsilon)\, dx.$$

The result then easily follows differentiating under the integral sign. □

We see, therefore, that $\phi'(0)$ depends only on the infinitesimal generators $\mu(x)$ and $\varphi(x)$ of the perturbations $\eta(x, \epsilon)$ and $v(y, \epsilon)$; $\phi'(0)$ is called the *variation in the direction* (φ, μ) of the functional

$$\mathcal{F}(u) := \int_\Omega F(x, u(x), Du(x))\, dx \tag{3.24}$$

at the point u.

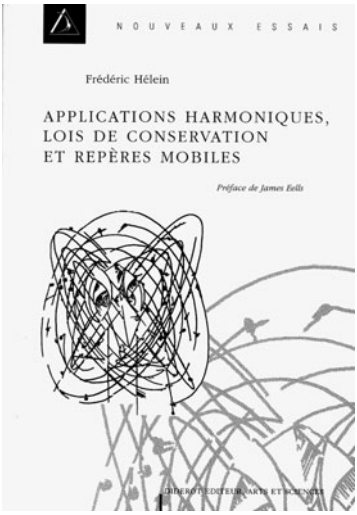
b. Inner variations

When $\mu(x) = 0$ and $\varphi \in C^1_c(\Omega, \mathbb{R}^N)$, so that $\eta(x, \epsilon) = x\ \forall\epsilon$ and $v(x, \epsilon) = u(x) + \epsilon\varphi(x)$, Proposition 3.37 is simply the computation of the first variation, as in this case

$$\phi'(0) = \delta\mathcal{F}(u, \varphi).$$

Instead, if we choose $\mu \in C^1(U, \mathbb{R}^N)$, $\eta(x, \epsilon) := x + \epsilon\mu(x)$ and $v(y, \epsilon) = u(\eta(y, \epsilon)^{-1})$, then $u(x) = v(\eta(x, \epsilon), \epsilon)$, and differentiating in ϵ gives

$$0 = D_y v(x, 0)\mu(x) + \varphi(x), \qquad \text{i.e.,} \qquad \varphi(x) := -Du(x)\mu(x).$$



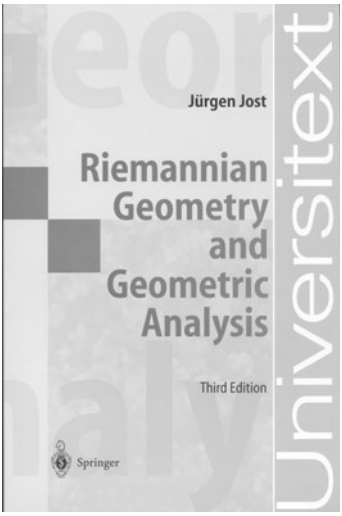


Figure 3.11. Frontispieces of two volumes on "geometric analysis".

It follows from Proposition 3.37 that for all $\lambda \in C^1(U, \mathbb{R}^N)$ the variation at u in the direction $(Du(x)\lambda(x), -\lambda(x))$, called the *inner variation* of $\mathcal{F}$ at u with respect to λ, is given by

$$\begin{aligned}\partial\mathcal{F}(u,\lambda) &:= \phi'(0)\\ &:= \int_\Omega \Big(F_{u^i} D_\alpha u^i \lambda^\alpha + F_{p^i_\beta} D_\beta (D_\alpha u^i \lambda^\alpha) - F \operatorname{div} \lambda - D_\alpha F\, \lambda^\alpha \Big)\, dx\\ &= \int_\Omega \Big(- F_{x^\alpha} \lambda^\alpha - F \operatorname{div} \lambda + F_{p^i_\beta} D_\alpha u^i D_\beta \lambda^\alpha \Big)\, dx.\end{aligned}$$

This can be written in a more compact way introducing the *Hamilton tensor*, called also *Hamilton–Eshelby* or *energy-momentum tensor*,

$$T^\beta_\alpha := p^i_\alpha F_{p^i_\beta} - \delta^\beta_a F;$$

(3.25) becoming

$$\partial\mathcal{F}(u,\lambda) = \int_\Omega \Big(T^\beta_\alpha \lambda^\alpha_{x^\beta} - F_{x^\alpha} \lambda^\alpha \Big)\, dx. \tag{3.25}$$

3.38 Definition. *We say that an extremal of $\mathcal{F}$, that is a solution of $\delta\mathcal{F}(u,\varphi) = 0\ \forall \varphi \in C^1_c(\Omega, \mathbb{R}^N)$, is* stationary *(respectively* strongly stationary*) if*

$$\partial\mathcal{F}(u,\lambda) = 0 \qquad \forall \lambda \in C^1_c(\Omega, \mathbb{R}^N)$$

(respectively $\forall \lambda \in C^1(\overline{\Omega}, \mathbb{R}^N)$).

If $u \in C^2(\Omega)$ (respectively $u \in C^2(\overline{\Omega})$), the relative Hamilton tensor is of class $C^1(\Omega)$ (respectively $C^1(\overline{\Omega})$), hence an integration by parts yields for all $\lambda \in C^1_c(\Omega, \mathbb{R}^n)$ (respectively $C^1(\overline{\Omega}, \mathbb{R}^n)$)

$$\partial \mathcal{F}(u, \lambda) = \int_\Omega (-D_\beta T^\beta_\alpha - F_{x_\alpha})\lambda^\alpha \, dx + \int_{\partial\Omega} \nu_\beta T^\beta_\alpha \lambda^\alpha \, d\mathcal{H}^{n-1}, \qquad (3.26)$$

where $\nu = (\nu_\beta)$ is the exterior unit normal vector to $\partial\Omega$. Consequently, an extremal $u \in C^2(\Omega)$ is stationary if and only if

$$D_\beta T^\beta_\alpha + F_{x^\alpha} = 0 \qquad \text{in } \Omega,$$

whereas $u \in C^2(\overline{\Omega})$ is a strong stationary extremal if and only if

$$\begin{cases} D_\beta T^\beta_\alpha + F_{x^\alpha} = 0 & \text{in } \Omega, \\ T^\beta_\alpha \nu_\beta = 0 & \text{su } \partial\Omega. \end{cases}$$

Trivially, (3.26) implies the following.

3.39 Corollary. *Minimizers of class C^1 are stationary extremals.*

There are examples of extremals of class C^1 that are not stationary. However, the following holds.

3.40 Proposition. *All extremals of class C^2 are stationary extremals.*

Proof. In fact, if u is an extremal of class C^2, we have $D_\beta F_{p^i_\beta} = F_{u^i}$ and, for $\alpha = 1, \dots, n$, we compute

$$\begin{aligned} D_\beta T^\beta_\alpha + F_{x^\alpha} &= D_{\beta\alpha} u^i F_{p^i_\beta} + D_\alpha u^i D_\beta (F_{p^i_\beta}) - D_\alpha F + F_{x^\alpha} \\ &= D_{\alpha\beta} u^i F_{p^i_\beta} + F_{u^i} D_\alpha u^i + F_{x^\alpha} - D_\alpha F = D_\alpha F - D_\alpha F = 0. \end{aligned}$$

□

3.41 Example (Conservation of energy). Minimizers of Lagrangians ot the type $F = F(u, p)$ satisfy the law of conservation of energy

$$F(u, u') - u' F_p(u, u') = const.$$

c. Curves of minimal energy and curves of minimal length

Let $U : [0, 1] :\to \mathbb{R}^N$ be a curve parametrized with constant velocity. Then

$$\mathcal{D}(U) = \frac{1}{2} \int_0^1 |U'(t)|^2 \, dt = \frac{1}{2} L(U)^2,$$

where $L(U)$ is the length of the curve $U(t)$. Consequently, if $U : [0, 1] \to \mathcal{Y} \subset \mathbb{R}^N$ is a curve with velocity of constant modulus, with extreme points p and q and of minimal length among the curves in $\mathcal{Y}$ with extreme points p and q, then for all $v : [0, 1] \to \mathcal{Y}$ with $v(0) = p$ and $v(1) = q$,

$$\mathcal{D}(U) = \frac{1}{2}L(U)^2 \le \frac{1}{2}L(v)^2 = \frac{1}{2}\left(\int_0^1 |v'|\,dt\right)^2 \le \frac{1}{2}\int_0^1 |v'|^2\,dt = \mathcal{D}(v),$$

i.e., U minimizes the Dirichlet integral among all curves in $\mathcal{Y}$ with extreme points p and q.

Conversely, consider a minimizer $U : [0,1] \to \mathcal{Y} \subset \mathbb{R}^N$ of the Dirichlet integral among smooth curves in $\mathcal{Y}$ with prescribed boundary values p and q,

$$\mathcal{D}(U) \to \min, \qquad U(x) \in \mathcal{Y},\ U(0) = p,\ U(1) = q,$$

and let $u : [0,1] \to \mathbb{R}^m$ be a representation of U in local coordinates of $\mathcal{Y}$. Then, see (3.20),

$$\mathcal{D}(U) = \mathcal{D}(u) = \frac{1}{2}\int_0^1 |U'|^2\,dt = \int_0^1 g_{ij}(u(t))u^{i\prime}(t)u^{j\prime}(t)\,dt$$

and u is a minimizer of the functional

$$\int_0^1 F(t, u(t), u'(t))\,dt$$

with integrand $F(t,u,p) = \sum_{i,j=1}^N g_{ij}(u)p^i p^j$. The corresponding Hamilton tensor is

$$T = p^i F_{p_i} - F = g_{ij}p^i p^j - \frac{1}{2}g_{ij}p^i p^j = \frac{1}{2}g_{ij}p^i p^j = \frac{1}{2}|U'|^2,$$

hence, since the interior variation has to vanish according to Corollary 3.39, we have

$$0 = \partial\mathcal{D}(u,\lambda) = \frac{1}{2}\int_0^1 |U'(t)|^2\lambda'\,dt \qquad \forall\lambda \in C_c^1(]0,1[) \tag{3.27}$$

or, in other words, U has velocity of constant modulus. In particular, $2\mathcal{D}(U) = L(U)^2$.

Now, if c is any regular curve with extreme points p and q with values in $\mathcal{Y}$ and v is a reparametrization of it with velocity of constant modulus, then

$$L^2(U) = \left(\int_0^1 |U'|\,dx\right)^2 = 2\mathcal{D}(U) \le 2\mathcal{D}(v) = L^2(v) = L^2(c).$$

We can therefore state the following.

3.42 Theorem. *A curve γ in $\mathcal{Y}$ has minimal Dirichlet energy if and only if it has minimal length and it is reparametrized with velocity of constant modulus.*

This claim makes clear that a rubber band on a surface withdraws to a curve of minimal length.

d. Surfaces of minimal energy and surfaces of minimal area

Let $B(0,1)$ be the unit ball of $\mathbb{R}^2$, $\mathcal{Y}$ a m-dimensional submanifold of $\mathbb{R}^n$, and $u : B(0,1) \to \mathbb{R}^m$ a representation in local coordinates of a map $U : B(0,1) \to \mathcal{Y}$. The Hamilton tensor for the map u relative to the Dirichlet integral can be written as

$$\begin{pmatrix} T_1^1 & T_2^1 \\ T_1^2 & T_2^2 \end{pmatrix} = \begin{pmatrix} a & b \\ b & -a \end{pmatrix},$$

where

$$\begin{aligned} a &:= \frac{1}{2}\Big(|D_1U|^2 - |D_2U|^2\Big) = \frac{1}{2}(g_{ik}(u)u^i_{x^1}u^k_{x^1} - g_{ik}(u)u^i_{x^2}u^k_{x^2}), \\ b &:= D_1u \bullet D_2u = g_{ik}(u)u^i_{x^1}u^k_{x^2}. \end{aligned}$$

3.43 Theorem. *A strong stationary extremal $U : B(0,1) \subset \mathbb{R}^2 \to \mathcal{Y}$ of class $C^2(\overline{B(0,1)})$ of the Dirichlet integral satisfies the* conformality relations

$$|D_1U|^2 = |D_2U|^2, \qquad D_1U \bullet D_2U = 0.$$

Proof. Since U is a strong stationary extremal of the Dirichlet integral, we have

$$\int_{B(0,1)} \Big(a(D_1\lambda^1 - D_2\lambda^2) + b(D_2\lambda^1 + D_1\lambda^2)\Big)\,dx = 0 \qquad \forall \lambda \in C^1(\overline{B(0,1)}, \mathbb{R}^2). \quad (3.28)$$

Restricting ourselves to fields $\lambda \in C^1_c(B(0,1), \mathbb{R}^2)$ and integrating by parts, we infer at once from (3.28) that

$$\phi(z) := a(x^1, x^2) - ib(x^1, x^2)$$

is a holomorphic function of $z = x^1 + ix^2$ in $B(0,1)$. Moreover, testing with arbitrary fields $\lambda \in C^1(\overline{B(0,1)}, \mathbb{R}^2)$, we find

$$\nu_\beta T^\beta_\alpha = 0 \qquad \text{on } \partial B(0,1),$$

where ν is the unit exterior normal to $B(0,1)$, see (3.26). In particular, $T^1_1((1,0)) = 0$ and $T^1_2((1,0)) = 0$. Since the Dirichlet integral is invariant under rotations R of the plane $\mathbb{R}^2$, hence $u \circ R$ are extremals, and actually strong stationary extremals. It follows that $a = \Re\phi$ and $b = \Im\phi$ vanish on $\partial B(0,1)$, and, by the Cauchy formula ($\phi \in C^1(\overline{B(0,1)})$) that $\phi(z) = 0$ in $B(0,1)$. □

Finally, recall that the area of $U(B(0,1))$ in $\mathcal{Y}$ is given by

$$\mathcal{A}(U, B(0,1)) := \mathcal{H}^2(U(B(0,1))) = \int_{B(0,1)} J_U(x)\,dx,$$

where J_U is the Jacobian of U, $J_U := \sqrt{\det DU^* DU}$, and that

$$|J_U(x)| \le \frac{1}{2}|DU(x)|^2 \qquad \forall x \in B(0,1)$$

with equality if and only if U satisfies the conformality relations.

Therefore, if U is conformal and U_t, $t \in]-1,1[$ is a variation of $U := U_0$, we then have

$$\mathcal{A}(U_t, B(0,1)) \leq \mathcal{D}(U_t, B(0,1))$$

for all $t \in]-1,1[$. It follows that the first variation of the area functional at U in a direction v vanishes if and only if the first variation of the Dirichlet integral in the direction v vanishes. Maps which have minimal area or just critical points of the area functional are called *minimal surfaces*. Accordingly, harmonic maps which are conformal are called *parametric minimal surfaces*. As we have seen, parametric minimal surfaces are minimal surfaces and strong inner extremals of the Dirichlet integral are conformal, hence parametric minimal surfaces.

e. Noether theorem

Finally, we consider a family of smooth transformations from $\mathbb{R}^n \times \mathbb{R}^N$ into $\mathbb{R}^n \times \mathbb{R}^N$ that depend smoothly from a parameter ϵ

$$\begin{cases} y = Y(x,z,\epsilon), \\ w = W(x,z,\epsilon), \end{cases} \tag{3.29}$$

and that at $\epsilon = 0$ are the identity, $Y(x,z,0) = x$, $W(x,z,0) = z\ \forall (x,z)$. Their infinitesimal generators are denoted by

$$\mu(x,z) := \frac{\partial Y}{\partial \epsilon}(x,z,0), \qquad \omega(x,z) := \frac{\partial W}{\partial \epsilon}(x,z,0).$$

For a given $u(x) \in C^1(U)$, we compose (3.29) with $(x, u(x))$ to get the transformations

$$\eta(x,\epsilon) := Y(x,u(x),\epsilon), \qquad w(x,\epsilon) := W(x,u(x),\epsilon)$$

with infinitesimal generators

$$\frac{\partial \eta}{\partial \epsilon}(x,0) = \mu(x,u(x)), \qquad \frac{\partial w}{\partial \epsilon}(x,0) = \omega(x,u(x)).$$

Of course, $\mu(x,u(x))$ is of class $C^1(U)$ and with bounded differential quotient in an open set $V \subset\subset U$. Since $\eta(x,0) = x$, we have $\eta(x,\epsilon) = x + \epsilon\mu(x,u(x)) + o(\epsilon)$ as $\epsilon \to 0$. It follows that for ϵ small, $\eta(\cdot,\epsilon)$ is a diffeomorphism from V into its image with inverse $\xi(\cdot,\epsilon)$ with Taylor expansion $\xi(y,\epsilon) = y - \epsilon\mu(y,u(y)) + o(\epsilon)$ as $\epsilon \to 0$.

Finally, we consider the perturbation of $u(x)$

$$v(y,\epsilon) := w(\xi(y,\epsilon),\epsilon)$$

with infinitesimal generator

$$\varphi(x) := \frac{\partial v}{\partial \epsilon}(x,0)$$

that we can compute as follows. Since $w(x,\epsilon) = v(\eta(x,\epsilon),\epsilon)$, differentiating at 0 we find

$$\omega(x,u(x)) = \frac{\partial w}{\partial \epsilon}(x,0) = D_y v(x,0)\mu(x,u(x)) + \varphi(x)$$

and, since $D_y v(x,0) = Du(x)$, we conclude

$$\varphi(x) = \omega(x,u(x)) - Du(x)\mu(x,u(x)).$$

Proposition 3.37 then yields at once the following.

3.44 Theorem (Noether). *Assume that the functional $\mathcal{F}(u)$ is invariant with respect to the family of transformations in* (3.29) *with infinitesimal generators μ and ω. Then the extremals u of class $C^2(\Omega)$ of $\mathcal{F}$ satisfy the* conservation law

$$\sum_{\alpha=1}^{n} D_\alpha (F_{p_\alpha^i}\omega^i - T_\beta^\alpha \mu^\beta) = 0.$$

Proof. Due to the invariance of $\mathcal{F}$, its variation at u in the direction $(\omega(x,u(x)) - Du(x)\mu(x,u(x)), \mu(x,u(x))$ vanishes. The result then follows easily from Proposition 3.37. □

3.45 Example (Conservation of energy). Let $F = F(u,p)$. The functional

$$\mathcal{F}(u) := \int_\Omega F(u, Du)\, dx$$

is invariant with respect to translation in x, $\eta(x,\epsilon) = x + \epsilon$, $w(x,\epsilon) = u(x)$ that have as infinitesimal generators $\mu = 1$ and $\omega = 0$.

It follows in the case $n = N = 1$

$$F(u,u') - u' F_p(u,u') = const.$$

3.46 Example (Newton's gravitation law). Let $m_1, m_2, \ldots, m_n$ be the masses of n point-masses in $x_1(t), x_2(t), \ldots, x_n(t)$ at time t. According to Newton's gravitation law they attract each other with forces given by

$$F_{ik} := K\frac{m_i m_k}{r_{ik}^3}(x_k - x_i), \qquad i \neq k,\ r_{ik} = |x_i - x_k|,$$

according to Hamilton's principle, the actual motion, $x(t) = (x_1(t), x_2(t), \ldots, x_n(t)) \in \mathbb{R}^{3n}$ is an extremal of the *action*

$$\mathcal{F}(x) := \int_{t_1}^{t_2} F(x,x')\, dt,$$

where $F(x,x') := T(x') - V(x)$ with

$$T(x') = \frac{1}{2}\sum_{j=1}^{n} m_j {x_j'}^2, \qquad V(x) = -\sum_{i<k} K\frac{m_i m_k}{r_{ik}}.$$

In this case $F_{p^i} = m_i p_i$, and the Hamilton tensor is the function

$$T(x,x') = x_i' F_{p^i}(x,x') - F(x,x') = \frac{1}{2}\sum_{i=1}^{n} m_i |x_i'|^2 + V(x),$$

i.e., the total energy of the system. It is readily seen that $\mathcal{F}$ is invariant

(i) with respect to *translations of time*

$$\begin{cases} t^* = t + \epsilon, \\ x^* = x, \end{cases}$$

with generators $\mu = 1$ and $\omega = 0$. We find as a consequence of Noether's theorem the law of conservation of energy

$$\Big(\frac{1}{2}\sum_{i=1}^{n} m_i |x_i'|^2 + V(x)\Big)' = 0,$$

at least as far as no collision arises;

(ii) with respect to *translations in space*, for instance in the direction e_1,

$$\begin{cases} t^* = t, \\ x^* = x + e_1, \end{cases}$$

with generators $\mu = 0$ and $\omega = (e_1, e_1, \dots, e_1)$, hence

$$F_{p_1} e_1 + \cdots + F_{p_n} e_1 = 0$$

or, equivalently,

$$e_1 \bullet \sum_{i=1}^{n} m_i x_i'(t) \ = const,$$

that expresses the *conservation of momentum*, in absence of collisions;

(iii) with respect to *rotations in the space*

$$\begin{cases} t^* = t, \\ x_i^* = x_i \cos\epsilon + y_i \sin\epsilon, \\ y_i^* = -x_i \sin\epsilon + y_i \cos\epsilon, \\ z_i^* = z_i, \end{cases}$$

with infinitesimal generators $\mu = 0$ and $\omega = (a_1, \dots, a_n)$, $a_j := (y_j, -x_j, 0)$, hence

$$F_{p_1} a_1 + \cdots + F_{p_n} a_n = const,$$

equivalently,

$$m_1(y_1 x_1' - x_1 y_1') + \cdots + m_n(y_n x_n' - x_n y_n') = const,$$

i.e.,

$$\sum_{j=1}^{n} m_j x_j \wedge x_j' = const,$$

that expresses the *conservation of the angular momentum*, again in absence of collisions.

3.1.5 The eikonal and the Huygens principle

In this subsection we illustrate an approach to sufficient conditions for optimality, discussed by Carathéodory, the ideas of which date back to Johann Bernoulli (1667–1748). This will lead us, in particular, to the formulation of the Huygens principle, see Paragraph d.. For the sake of simplicity we shall consider only simple integrals.

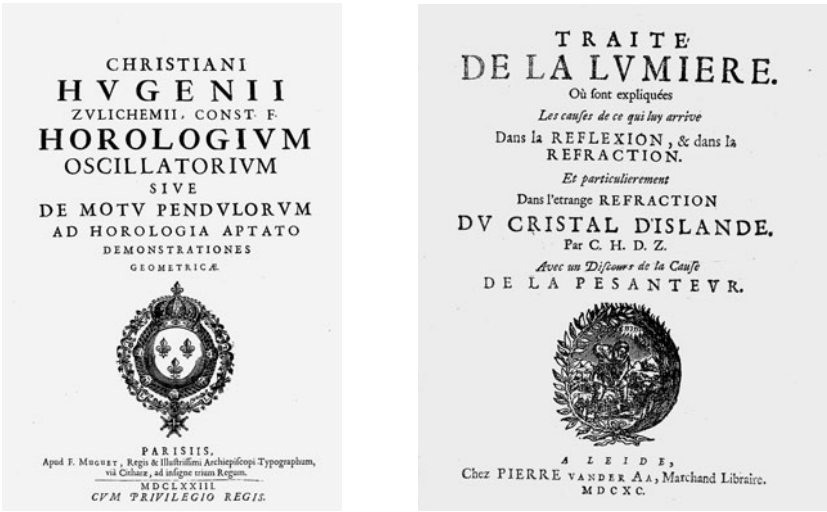


Figure 3.12. Frontispieces of *Horologium oscillatorium*, Paris, 1673, and of *Traité de la Lumière*, Leiden, 1690 by Christiaan Huygens (1629–1695).

a. Calibrations and fields of extremals

Let

$$\mathcal{F}(u) := \int_a^b F(t, u, u')\, dt$$

be a variational integral defined in the class of admissible functions

$$\mathcal{C} := \Big\{ u \in C^1([a,b], \mathbb{R}^N) \,\Big|\, u(a) = \alpha,\ u(b) = \beta \Big\}.$$

Given $u_0 \in \mathcal{C}$, suppose we may find a functional $\mathcal{M}(u)$ such that

$$\begin{cases} \mathcal{F}(u) \geq \mathcal{M}(u) & \forall u \in \mathcal{C}, \\ \mathcal{F}(u_0) = \mathcal{M}(u_0), \end{cases} \tag{3.30}$$

then, clearly, if u_0 is a minimizer for $\mathcal{M}$, it is a minimizer of $\mathcal{F}$. A functional

$$\mathcal{M}(u) := \int_a^b M(t, u, u')\, dt \tag{3.31}$$

has the property (3.30) if we assume that

$$\begin{cases} M(t, u, p) \leq F(t, u, p) & \forall (t, u, p), \\ M(t, u_0(t), u_0'(t)) = F(t, u_0(t), u_0'(t)) & \forall t \in [a, b]. \end{cases} \tag{3.32}$$

Although at first it is hard to believe, integrals (3.31) and (3.32) for which $\mathcal{M}(u) = const$ for all $u \in \mathcal{C}$ are particularly interesting; they are called *calibrations* for $(\mathcal{F}, u_0, \mathcal{C})$. Before continuing, it is convenient to introduce two new notions.

3.47 Null Lagrangians. Suppose $\mathcal{M}$ is a calibration for $\{\mathcal{F}, u_0, \mathcal{C}\}$ with integrand $M(t,u,p)$, then for all φ with compact support $\delta\mathcal{M}(u,\varphi) = 0$, i.e.,

$$M_u(t,u(t),u'(t)) - \frac{d}{dt}M_p(t,u(t),u'(t)) = 0 \qquad \text{in }]a,b[\tag{3.33}$$

for all $u \in C^1(]a,b[\times\mathbb{R}^N) \cap \mathcal{C}$. We say that M is a *null Lagrangian* if (3.33) holds for all u. By testing with functions of the type $u(t) = (u^1(t), \dots, u^N(t))$ where

$$u^i(t) = z^i + (t-t_0)p^i + \frac{1}{2}(t-t_0)^2 q^i, \qquad i = 1, \dots, N,$$

it is not difficult to see that M is a null Lagrangian if and only if there is a scalar function $S(t,u)$ such that

$$M(t,u,p) = S_t(x,u) + S_u(t,u) \bullet p\,;$$

in this case, we have

$$\mathcal{M}(u) = \int_a^b \frac{d}{dt}S(t,u(t))\,dt = S(b,u(b)) - S(a,u(a))$$

for all functions u of class $C^1([a,b])$.

3.48 Fields of extremals. Let $\Omega \subset \mathbb{R}^{N+1}$ be a simply connected domain that is foliated by a family of graphs that do not intersect. More precisely, consider a simply connected domain in $\mathbb{R} \times \mathbb{R}^N$ that is normal to the axis t,

$$\Gamma := \left\{ (t,c) \,\middle|\, c \in A \subset \mathbb{R}^N,\ t_1(c) \le t \le t_2(c) \right\}$$

and a diffeomorphism $f : \Gamma \to \Omega$ of the form $f(t,c) := (t, \varphi(t,c))$, $\varphi : \Gamma \to \mathbb{R}^N$ or, as it is called, a *field* in Ω. The graphs of the functions $t \to \varphi(t,c)$ are called the *lines of the field*. Since for every $(t,u) \in \Omega$ there exists a unique $c \in A$, $c = c(t,u)$ such that $\varphi(t,c) = u$, a *slope field* $\mathcal{P} : \Omega \to \mathbb{R}^N$ is well-defined as the derivative in t of the single line of the field through (t,u)

$$\mathcal{P}(t,u) := \frac{\partial\varphi}{\partial t}(t, c(t,u)).$$

In particular, every line of the field is the graph of a function $u(t)$ that solves the first order system

$$u'(t) = \mathcal{P}(t,u(t)).$$

In this way the slope field fully characterizes the field.

b. Mayer fields

We introduce an important class of fields associated to a functional

$$\mathcal{F}(u) := \int_a^b F(t, u, u')\, dt,$$

where $F(t, u, p) : [a, b] \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$.

3.49 Definition. *A field with slope field $\mathcal{P}(t, u)$ defined on $\Omega \subset \mathbb{R}^{N+1}$ is called a* Mayer field *for $\mathcal{F}$ if the following holds:*

(i) *The lines of the field are extremals of $\mathcal{F}(u)$.*
(ii) *There is a function $S : \Omega \to \mathbb{R}$ called the* eikonal of the field*, that satisfies* Carathéodory equations*,*

$$\begin{cases} F^*(t, u, \mathcal{P}(t, u)) = 0, \\ F^*_p(t, u, \mathcal{P}(t, u)) = 0 \end{cases} \qquad \forall (t, u) \in \Omega,$$

where $F^*(t, u, p) := F(t, u, p) - S_t(t, u) - S_u(t, u) \bullet p$.

3.50 Remark. We notice the following:

(i) $M(t, u, p) := S_t(t, u) - S_u(t, u) \bullet p$ is a null Lagrangian.
(ii) Carathéodory equations can be written in various ways; first, it is easily seen that as equations for S and $\mathcal{P}$, they are equivalent to the system of PDE's

$$\begin{cases} S_u(t, u) = F_p(t, u, \mathcal{P}(t, u)), \\ S_t(t, u) = (F(t, u, \mathcal{P}(t, u)) - F_p(t, u, \mathcal{P}(t, u)) \bullet \mathcal{P}(t, u)\,. \end{cases}$$

(iii) By introducing the differential 1-form, called the *Beltrami* 1*-form*,

$$\gamma(t, u, p) := (F - p \bullet F_p)\, dx - \sum_{i=1}^{N} F_{p^i}\, du^i$$

in $[a, b] \times \mathbb{R}^N \times \mathbb{R}^N$ and the map $\mathfrak{p}(t, u) := (t, u, \mathcal{P}(t, u))$, the eikonal fulfills Carathéodory equations if and only if

$$dS = \mathfrak{p}^{\#}\gamma,$$

i.e., S is a primitive of $\mathfrak{p}^{\#}\gamma$. If the domain is simply connected, this happens if and only if $d\,\mathfrak{p}^{\#}\gamma = 0$.

3.51 Optimal fields. A field of extremals in $\Omega \subset \mathbb{R}^{N+1}$ for $\mathcal{F}$ is said to be *optimal* if there is $S(t,u)$, $(t,u) \in \Omega$, such that the integrand

$$F^*(t,u,p) := F(t,u,p) - S_t(t,u) - S_u(t,u) \bullet p\,, \qquad \forall (t,u,p) \tag{3.34}$$

fulfills

$$F^* \geq 0, \qquad F^*(t,u,\mathcal{P}(t,u)) = 0 \qquad \forall (t,u) \in \Omega. \tag{3.35}$$

Optimal fields for $\mathcal{F}$ are also Mayer fields. In fact, its slope field fulfills (3.35), therefore, for all (t,u) the function $p \to F^*(t,u,p)$ has a minimum at $p = \mathcal{P}(t,u)$, hence $F^*_p(t,u,\mathcal{P}(t,u)) = 0$. By definition, the functional

$$\mathcal{M}(u) := \int_a^b (S_t(t,u(t)) + S_u(t,u(t)) \bullet u'(t)\,)\,dt$$

is a *calibration* for $(\mathcal{F}, u_0, \mathcal{C})$ for any line u_0 of the field.

It is readily seen that the lines $u(x) := \varphi(x,c)$ of an optimal field are actually minimizers for $\mathcal{F}(u)$ in the class of smooth functions with prescribed boundary values. In fact, $F^*(t,u,\mathcal{P}(t,u)) = 0$ implies

$$\begin{aligned}\mathcal{F}(u) &= \int_a^b F(t,u(t),\mathcal{P}(t,u(t)))\,dt \\ &= \int_a^b \Big(S_t(t,u(t)) + S_u(t,u(t)) \bullet u'(t)\Big)\,dt \\ &= \int_a^b \frac{d}{dt} S(t,u(t))\,dt = S(b,u(b)) - S(a,u(a)).\end{aligned}$$

On the other hand, since $F^*(x,u,p) \geq 0$ for all $v \in C^1([a,b])$ with the same boundary values of $u(t)$, we have

$$\begin{aligned}\mathcal{F}(v) &= \int_a^b F(t,v(t),v'(t))\,dt \geq \int_a^b \Big(S_t(t,v(t)) + S_u(t,v(t)) \bullet v'(t)\Big)\,dt \\ &= \int_a^b \frac{d}{dt} S(t,v(t))\,dt = S(b,v(b)) - S(a,v(a)) \\ &= S(b,u(b)) - S(a,u(a)) = \mathcal{F}(u).\end{aligned}$$

c. The Weierstrass representation formula

3.52 Definition. *The Weierstrass excess function is defined by*

$$\mathcal{E}_F(t,u,p,q) := F(t,u,p) - F(t,u,q) - (p-q) \bullet F_p(t,u,q)\,.$$

3.53 Theorem (Weierstrass formula). *Given a Mayer field with slope $\mathcal{P}(t,u)$ and eikonal $S(t,u)$, then we have*

$$F(t,u,p) - S_t(t,u) - S_u(t,u) \bullet p = \mathcal{E}_F(t,u,p,\mathcal{P}(t,u)),$$

consequently, for all u we have

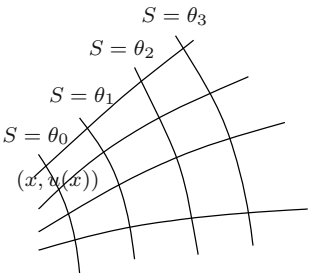


Figure 3.13. Light rays and wave fronts.

$$\int_a^b F(t,u,u')\,dt \\ = S(b,u(b)) - S(a,u(a)) + \int_a^b \mathcal{E}_F(t,u(t),u'(t),\mathcal{P}(t,u(t)))\,dt \tag{3.36}$$

and, trivially,

$$\int_a^b F(t,u(t),u'(t))\,dt = S(b,u(b)) - S(a,u(a))$$

if the graph of $u(t)$ is a line of the field.

Notice that $\mathcal{E}_F(t,u,p,q) \geq 0$ if $F(t,u,p)$ is convex in p for every fixed (t,u). We may therefore state the following.

3.54 Theorem. *Let u be an extremal of $\mathcal{F}(u)$ with integrand $F(t,u,p)$ convex in p for every fixed (t,u). If we can embed $u(t)$ in a Mayer field, then u is a minimizer of $\mathcal{F}$ among functions with the same boundary values as u.*

We are now left with the problem of analyzing when an extremal can be embedded in a Mayer field. In the *elliptic case*, $F_{pp} > 0$, one can show that this is always possible, at least locally, although there exist situations in which this cannot be done globally. Conditions under which the global embedding is possible are done in terms of conjugate points or eigenvalues of Jacobi operator, but we will not pursue the argument any further. We only remark that a connected piece of maximal circle on the unit sphere of length less than π has minimal length among the curves on the sphere connecting its extreme points, whereas its complement is an extremal but not a minimizer.

d. Huygens principle

Formula (3.36) is connected to the Huygens principle. This would require introducing *parametric integrals*; nevertheless, we conclude this subsection with some remarks.

As we have seen, the slope and the eikonal of a Mayer field are related by Carathéodory equations

$$\nabla S(x,z) = (F - p \bullet F_p\,, F_p)(\mathfrak{p}(x,z)), \qquad \mathfrak{p}(x,z) := (x,z,\mathcal{P}(x,z)),$$

hence ∇S is not perpendicular to the lines of the field if F and F_p do not vanish simultaneously. Since ∇S is orthogonal to the level lines of S, we may and do conclude that the lines of the fields are *transversal* (i.e., they are not tangent) to the level lines of the eikonal S.

Given an optimal Mayer field with slope $\mathcal{P}(x,z)$ and eikonal $S(x,z)$ in $\Omega \subset \mathbb{R} \times \mathbb{R}^N$, consider two level sets $\Sigma_1 = \{(x,z) \,|\, S(x,z) = \theta_1\}$ and $\Sigma_2 = \{(x,z) \,|\, S(x,z) = \theta_1\}$ of the eikonal. Then

$$\mathcal{F}(u) \geq \theta_1 - \theta_0$$

for every curve with extreme point on Σ_1 and Σ_0 and the equality holds if and only if u is a line of the field. Now, if $P_1 := (x_1, z_1)$ and $P_2 := (x_2, z_2)$ are two points, it is easily seen that (we assume $F > 0$)

$$d_F(P_1, P_2) := \inf\Big\{\mathcal{F}(u) \,\Big|\, u(x_1) = z_1,\ u(x_2) = z_2\Big\}$$

is a distance in Ω. Trivially, the surfaces Σ_0 and Σ_1 are d_F-equidistant and, if $P_1 \in \Sigma_1 \cap \Omega$ and $P_2 \in \Sigma_2 \cap \Omega$, then $d_F(P_1, P_2) \geq \theta_2 - \theta_1$ with equality if and only if P_1 and P_2 are on the same line. Given a point $P \in \Sigma_{\theta_1}$, we consider the *geodesic ball* with center at P and radius θ

$$B_F(P,\theta) = \Big\{Q \in \Omega \,\Big|\, d_F(P,Q) < \theta\Big\}.$$

If θ is small, then $B_F(P,\theta) \subset \Omega$ and, from the above, $B_F(P,\theta)$ is "tangent" to $\Sigma_{\theta_1+\theta}$. In fact, $\Sigma_{\theta_1+\theta}$ *is the envelope of the geodetic spheres* $\partial B_F(P,\theta)$ *with center* $P \in \Sigma_{\theta_1}$ when P moves in Σ_{θ_1}.

We may interpret the lines of an optimal Mayer field as the trajectories of a system of particles or as a bundle of light rays, the functional as the propagation time needed by a particle to move along $(x, u(x))$ from $(x_1, u(x_1))$ to $(x_2, u(x_2))$, the level lines of the eikonal as the *wave fronts* of the field, i.e., equidistant surfaces with respect to the time of propagation of the particles or of the light, and $\partial B(P,\theta)$ as the wave fronts of a bundle of particles or of rays of light emanating from P. We can therefore state the following.

Huygens principle. Consider every point P of the wave front Σ_{θ_0} at time θ_0 as source of new wave fronts $\partial B_F(P,\theta)$ propagating with time θ. Then the wave front $\Sigma_{\theta_0+\theta}$, $\theta > 0$ is the envelope of the elementary waves $\partial B_F(P,\theta)$ with center P on Σ_{θ_0}, see Figure 3.14.

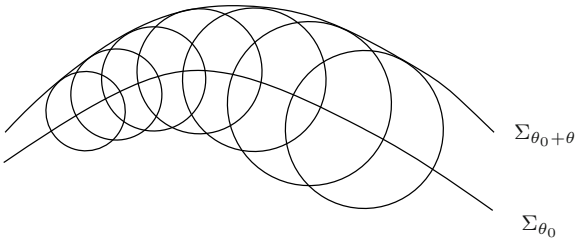


Figure 3.14. Huygens principle.

3.2 The Classical Hamiltonian Formalism

In this section we present a short introduction to Hamiltonian formalism in the 1-dimensional case as its extension to many dimensions is more complex.

3.2.1 The canonical equations of Hamilton and Hamilton–Jacobi

Let $F(t,u,v) : [a,b] \times \mathbb{R}^N \times \mathbb{R}^N$ be a Lagrangian of class C^s, $s \geq 2$, with matrix F_{vv} positive definite so that the transformation $\mathcal{L}_F$ given by

$$(t,u,v) \to (t,u,p), \qquad p := F_v(t,u,v)$$

is a diffeomorphism. The Legendre transform of F with respect to v is then well-defined by

$$H(t,u,p) := p \bullet v - F(t,u,v) \qquad (t,u,v) = \mathcal{L}_F^{-1}(t,u,p), \tag{3.37}$$

and is called the *Hamiltonian* corresponding to F. As we saw in Chapter 2, we have

$$F(t,u,v) + H(t,u,p) = p \bullet v\,, \qquad p = F_v(t,u,v), \tag{3.38}$$

$$v = H_p(t,u,p), \qquad \mathcal{L}_F^{-1} = \mathcal{L}_H, \tag{3.39}$$

and, trivially,

$$F_t(t,u,v) + H_t(t,u,p) = 0, \qquad F_u(t,u,v) + H_u(t,u,p) = 0. \tag{3.40}$$

a. Hamilton equations

If $u : [a, b] \to \mathbb{R}^N$ is an extremal of

$$\mathcal{F}(u) := \int_a^b F(t, u(t), u'(t))\, dt,$$

then the curve $e(t) := (t, u(t), v(t))$, $v(t) := u'(t)$, in the phase space $[a, b] \times \mathbb{R}^N \times \mathbb{R}^N$ solves the Euler–Lagrange system

$$\frac{du}{dt}(t) = v(t), \qquad \frac{d}{dt}F_v(e(t)) = F_u(e(t)). \tag{3.41}$$

If $\pi(t) := \mathcal{L}_F(e(t))$ is the image of the curve $e(t)$ in the *co-phase space*

$$\pi(t) := (t, u(t), p(t)), \qquad p(t) := F_v(t, u(t), v(t)),$$

or, equivalently,

$$e(t) = \mathcal{L}_F^{-1}(\pi(t)) = \mathcal{L}_H(\pi(t)),$$

then $e(t)$ solves the Euler–Lagrange system if and only if $\pi(t)$ solves the system of $2N$ first order differential equations

$$u' = H_p(t, u, p), \qquad p' = -H_u(t, u, p), \tag{3.42}$$

called the *Hamilton equations*. When the Legendre transform is invertible we therefore obtain an equivalent description of extremals. In particular, every mechanical or optical system may be described equivalently by the Lagrangian or by the Hamiltonian formalism (at least when the Legendre transform is a diffeomorphism).

In terms of Hamiltonian equations, more than on a single solution, the emphasis is on discussing the family of solutions, that is, the *Hamiltonian flux* and its geometric features. This naturally leads to topics such as *symplectic geometry*, *first integrals*, *complete integrable systems* and *ergodic theory*. Of course, we shall not discuss these topics and we confine ourselves to presenting some of the simple and old formalism.

For instance, notice that

$$\frac{d}{dt}H(t, u(t), p(t)) = H_t + H_u \bullet u' + H_p \bullet p' = \frac{\partial}{\partial t}H(t, u(t), u'(t))$$

if $(u(t), p(t))$ solves Hamilton equations. For *autonomous systems*, i.e., when F does not depend explicitly on time, the quantity $H(u(t), p(t))$ is constant along the trajectories of the motion or, in other words, the trajectory $(u(t), p(t))$ is on a level surface $H(u, p) = const$, and one says that $H(u, p)$ is a *first integral* of the motion.

b. Liouville's theorem

By setting $z(t) := (u(t), p(t))^T \in \mathbb{R}^{2N}$, Hamilton's system takes the form $z'(t) = X(t, z(t))$ with $X(t, z) := (H_p(t, z), -H_u(t, z))^T$. Since

$$\operatorname{div} X(t, z) = \frac{\partial}{\partial u} H_p(t, z) + \frac{\partial}{\partial p}(-H_u)(t, z) = 0.$$

We infer from Paragraph 2.6.h of [GM4] the following.

3.55 Theorem (Liouville). *The Hamiltonian flux preserves the volumes.*

A first consequence is that the Hamiltonian flux cannot be asymptotically stable. A second consequence is that the following *recurrence theorem of Poincaré* applies to the flux of an autonomous Hamiltonian system for which only a bounded region is accessible.

3.56 Theorem. *Let g be a one-to-one continuous map that preserves the volumes. Suppose that $g(\Omega) \subset \Omega$ for a bounded domain Ω. Then for every neighborhood U of any point in Ω there is a point $x \in U$ that returns into U after some time. More precisely, $g^n x \in U$ for some $n > 0$; here $g^n := g \circ g \circ \cdots \circ g$.*

Proof. Since $U, g(U), g^2(U), \ldots, g^n(U) := g \circ g \circ \cdots \circ g(U)$ have the same volume and $|\Omega| < \infty$, there are k and ℓ, $k > \ell$, such that

$$g^k(U) \cap g^\ell(U) \neq \emptyset, \qquad \text{hence} \qquad g^{k-\ell}(U) \cap U \neq \emptyset,$$

i.e., $x \in U$ and $g^{k-\ell} x \in U$. □

c. Hamilton–Jacobi equation

Consider a Mayer field, i.e., a field of extremals with N degrees of freedom with slope field $\mathcal{P}(x, u)$ and eikonal $S(t, u)$ solving Carathéodory's equations

$$\begin{cases} S_t(t, u) = (F - v \bullet F_v)(t, u, \mathcal{P}(t, u)), \\ S_u(t, u) = F_u(x, u, \mathcal{P}(x, u)). \end{cases} \tag{3.43}$$

By introducing the *dual slope field*

$$\psi(t, u) := F_v(t, u, \mathcal{P}(t, u)),$$

we have $\mathcal{P}(t, u) = H_p(t, u, \psi(t, u))$ and Carathéodory's equations can be written in terms of the dual slope field

$$\begin{cases} S_t(t, u) = -H(t, u, \psi(t, u)), \\ S_u(t, u) = \psi(t, u), \end{cases} \tag{3.44}$$

i.e., the eikonal solves the first order partial differential equation, called *Hamilton–Jacobi* equation,

$$S_t + H(t, u, S_u) = 0. \tag{3.45}$$

Actually, *S is the eikonal of a Mayer field if and only if it solves the Hamilton–Jacobi equation.* In fact, assuming that S solves (3.45) and setting $\psi(t,u) := S_u(t,u)$, the couple (S,ψ) solves (3.44) and, if $\mathcal{P}(t,v) := H_p(x,u,\psi(x,u))$, then S solves (3.43).

d. Poincaré–Cartan integral

Hamilton's system is the Euler–Lagrange equation of an integral; more precisely, Hamilton's canonical equations are the Euler–Lagrange equations of the *Poincaré–Cartan integral*

$$\mathcal{G}(u,p) = \int_a^b \Big(p(t)\frac{du(t)}{dt} - H(t,u(t),p(t))\Big)\,dt.$$

In fact, $\mathcal{G}((u,p)) := \int_a^b G(t,(u,p),(u',p'))\,dt$ with integrand

$$G(t,(u,p),(v,q)) := pv - H(t,u,p),$$

and its Euler–Lagrange equations are

$$\begin{cases} ((G_v)' - G_u)(t,u,p,u',p') = 0, \\ ((G_q)' - G_p)(t,u,p,u',p') = 0, \end{cases}$$

i.e.,

$$\begin{cases} u'(t) = H_p(t,u,p), \\ p'(t) = -H_u(t,u,p). \end{cases}$$

3.57 Remark. To bring some light on the origin of the integral $\mathcal{G}$, we introduce the differential 1-form of Cartan

$$\mathcal{K}_H := p_i\,du^i - H(t,u,p)\,dt.$$

If $F(t,u,v)$ and $H(t,u,p)$ are the Legendre transform of each other and

$$\gamma_F := (F - v \bullet F_v)\,dt + L_{v^i}\,du^i$$

is the Beltrami 1-form associated to F, we have

$$\gamma_F = \mathcal{L}_L^{\#}\mathcal{K}_H \qquad \text{equivalently} \qquad \mathcal{K}_H = \mathcal{L}_H^{\#}\gamma_L.$$

If $h(t) := (t,u(t),p(t))$ is a curve in the co-phase space and $e(t) = (t,u(t),\pi(t))$ is its Legendre transform via H, $e(t) := \mathcal{L}_F^{-1}(h(t)) = \mathcal{L}_H(h(t)) = (t,u(t),v(t))$, $v(t) = H_p(t,u(t),p(t))$, we have

$$h^{\#}\mathcal{K}_H = e^{\#}(\mathcal{L}_F^{\#}\mathcal{K}_H) = e^{\#}\gamma_F,$$

hence

$$\int_I e^{\#}\gamma_F = \int_I h^{\#}\mathcal{K}_H \qquad \text{or} \qquad \int_e \gamma_L = \int_h \mathcal{K}_H.$$

In general,

$$\int_e \gamma_L \neq \int_I F(t, u(t), u'(t))\, dt,$$

but, if $v = u'$, or, equivalently, if

$$e^{\#}\omega^i = 0, \qquad \omega^i := du^i - v^i dt, \qquad i = 1, \ldots, N,$$

then

$$\int_I F(t, u(t), u'(t))\, dt = \int_e \gamma_L = \int_h \mathcal{K}_H = \int_I \Big(p(t)u'(t) - H(t, u(t), p(t))\Big)\, dt.$$

e. Cyclic variables

A variable u^i is said to be *ignorable* or *cyclic* for $H(t, u, p)$ if $H_{u^i}(t, u, p) = 0$. In this case $p_i' = 0$, i.e., $p_i(t) = const$ and the integration of the Hamilton system (3.42) is reduced to the integration of a system of $2N-1$ equations, and, actually, to the integration of a system of $2N-2$ equations plus the subsequent integration of a first order ordinary differential equation.

We can therefore reduce the global and/or explicit integrability of a Hamiltonian system to the search of cyclic variables.

3.58 ¶. If $H = H(u, p)$ and all variables are cyclic, $H_{u^i} = 0$, i.e., $H = H(p)$, then $p' = -H_u = 0$ and $p_i(t) = J_i\ \forall i = 1, \ldots, N$, where J_i are constants; from $u' = H_p(p)$ we then conclude

$$u^i(t) = \omega_i t + \beta_i, \qquad \omega^i(t) := H_{p_i}(J) \tag{3.46}$$

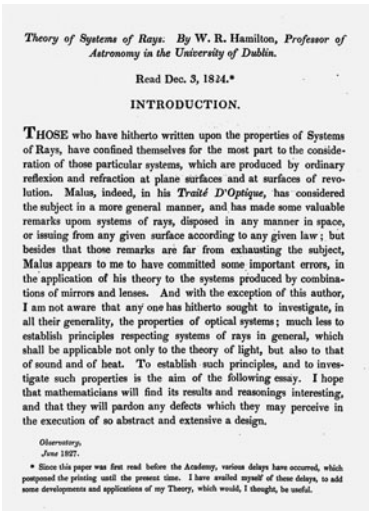
for suitable constants β_i and $J = (J_1, J_2, \ldots, J_N)$.

When the u^i's are angular variables, i.e., for instance, the motion in Cartesian coordinates has the form $x^i(t) = (\cos(u^i(t)), \sin(u^i(t)))$, then (3.46) describes periodic motions with angular velocities $\omega = (\omega_1, \omega_2, \ldots, \omega_N)$. The variables (u^i) are called *angular variables* and the J_i's the *action variables*. Of course, the action-angle variables are useful when treating periodic motions or perturbations of periodic motions. .

In principle, cyclic variables correspond to symmetries of the system and, in fact, they arise in a sort of dual formulation of Noether's theorem. If F and H are connected by the Legendre transform, the variable u^k is cyclic for H if and only if $F_{u^k}(t, u, u') = 0$ for all curves u. In this case, using the Euler–Lagrange equation

$$\frac{d}{dt} F_{v^k}(t, u(t), u'(t)) = F_{u^k}(t, u(t), u'(t)),$$

we find that $\frac{d}{dt} F_{v^k}(t, u(t), u'(t))$ vanishes for all extremals, consequently, $F_{v_k}(t, u(t), u'(t))$ is constant in t for all extremals, i.e., is a *first integral*.



Theory of Systems of Rays. By W. R. Hamilton, *Professor of Astronomy in the University of Dublin.*

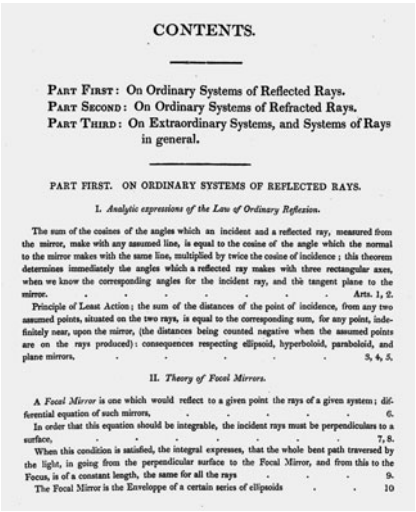
Read Dec. 3, 1824.*

INTRODUCTION.

THOSE who have hitherto written upon the properties of Systems of Rays, have confined themselves for the most part to the consideration of those particular systems, which are produced by ordinary reflexion and refraction at plane surfaces and at surfaces of revolution. Malus, indeed, in his *Traité D'Optique*, has considered the subject in a more general manner, and has made some valuable remarks upon systems of rays, disposed in any manner in space, or issuing from any given surface according to any given law; but besides that those remarks are far from exhausting the subject, Malus appears to me to have committed some important errors, in the application of his theory to the systems produced by combinations of mirrors and lenses. And with the exception of this author, I am not aware that any one has hitherto sought to investigate, in all their generality, the properties of optical systems; much less to establish principles respecting systems of rays in general, which shall be applicable not only to the theory of light, but also to that of sound and of heat. To establish such principles, and to investigate such properties is the aim of the following essay. I hope that mathematicians will find its results and reasonings interesting, and that they will pardon any defects which they may perceive in the execution of so abstract and extensive a design.

Observatory,
June 1827.

* Since this paper was first read before the Academy, various delays have occurred, which postponed the printing until the present time. I have availed myself of these delays, to add some developments and applications of my Theory, which would, I thought, be useful.



CONTENTS.

PART FIRST. ON ORDINARY SYSTEMS OF REFLECTED RAYS.

I. *Analytic expressions of the Law of Ordinary Reflexion.*

The sum of the cosines of the angles which an incident and a reflected ray, measured from the mirror, make with any assumed line, is equal to the cosine of the angle which the normal to the mirror makes with the same line, multiplied by twice the cosine of incidence; this theorem determines immediately the angles which a reflected ray makes with three rectangular axes, when we know the corresponding angles for the incident ray, and the tangent plane to the mirror. . . . . Arts. 1, 2.

Principle of Least Action; the sum of the distances of the point of incidence, from any two assumed points, situated on the two rays, is equal to the corresponding sum, for any point, indefinitely near, upon the mirror, (the distances being counted negative when the assumed points are on the rays produced): consequences respecting ellipsoid, hyperboloid, paraboloid, and plane mirrors. . . . . 3, 4, 5.

II. *Theory of Focal Mirrors.*

A *Focal Mirror* is one which would reflect to a given point the rays of a given system; differential equation of such mirrors, ... 6.

In order that this equation should be integrable, the incident rays must be perpendiculars to a surface, ... 7, 8.

When this condition is satisfied, the integral expresses, that the whole bent path traversed by the light, in going from the perpendicular surface to the Focal Mirror, and from this to the Focus, is of a constant length, the same for all the rays . . . . . 9.

The Focal Mirror is the Enveloppe of a certain series of ellipsoids . . . 10

Figure 3.15. Two pages from *Theory of Systems of Rays* by William R. Hamilton (1805–1865) published in *Transactions of the Irish Academy.*

f. Hamilton's approach to geometrical optics

First, it is convenient to hint briefly at Hamilton's approach in the context of optical instruments, that was transferred to mechanics mainly by Jacobi.

Given a nonnegative Lagrangian $F(t,u,p) : [a,b] \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$, for every couple of points $\overline{P} = (\overline{t}, \overline{x})$ and $P = (t,x)$, $\overline{t} < t$, in $[a,b] \times \mathbb{R}$, we define the function

$$\operatorname{dist}_F(\overline{P}, P) := \inf\Big\{ \int_{\overline{t}}^{t} F(t,\zeta,\zeta')\,dt \;\Big|\; \zeta : [\overline{t}, t] \to \mathbb{R}^N, \\ (\overline{t}, \zeta(\overline{t})) = \overline{P},\ (t, \zeta(t)) = P \Big\},$$

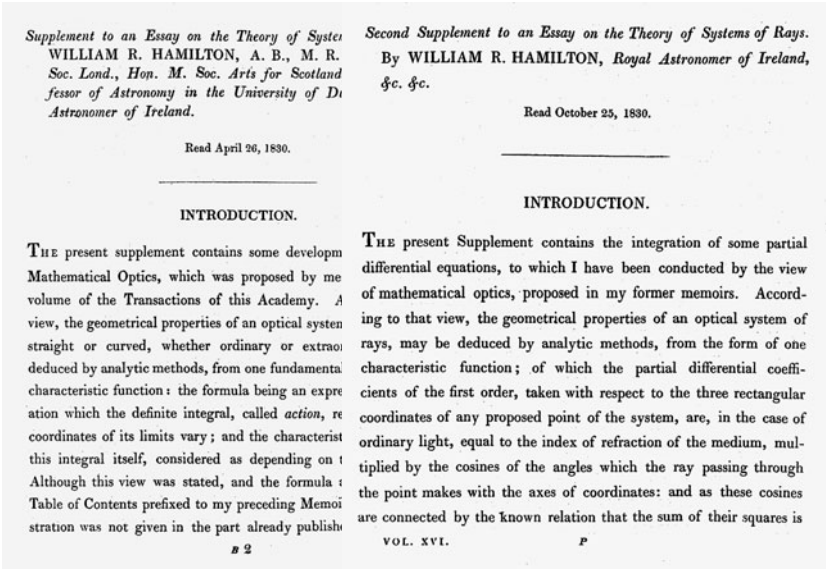
called the *principal function of Hamilton.*

The principal function of Hamilton

$$W(\overline{t}, \overline{x}, t, x) := \operatorname{dist}_F((\overline{t}, \overline{x}), (t,x)),$$

defined for $(\overline{t}, \overline{x})$, $(t,x) \in [a,b] \times \mathbb{R}^N$, $\overline{t} \le t$, may be regarded as the *value function* or *action* of the path of a particle in a mechanical system described by the Lagrangian L, or, in case the Lagrangian describes an optical system, as the time needed by a ray to go from $\overline{P}$ to P. Hamilton was the first to realize that the system associated to L is fully described by W and that all solutions can be obtained from W. In fact, he showed that the following equations hold for W: Let

$$y := L_v(t, x, x'(t)), \qquad \overline{y} := L_v(\overline{t}, \overline{x}, x'(\overline{t}))$$



Supplement to an Essay on the Theory of Syste
WILLIAM R. HAMILTON, A. B., M. R.
Soc. Lond., Hon. M. Soc. Arts for Scotland
fessor of Astronomy in the University of Di
Astronomer of Ireland.

Read April 26, 1830.

INTRODUCTION.

THE present supplement contains some developm
Mathematical Optics, which was proposed by me
volume of the Transactions of this Academy. A
view, the geometrical properties of an optical systen
straight or curved, whether ordinary or extraoi
deduced by analytic methods, from one fundamenta
characteristic function: the formula being an expre
ation which the definite integral, called *action*, re
coordinates of its limits vary; and the characterist
this integral itself, considered as depending on t
Although this view was stated, and the formula
Table of Contents prefixed to my preceding Memoi
stration was not given in the part already publish

B 2

Second Supplement to an Essay on the Theory of Systems of Rays.
By WILLIAM R. HAMILTON, *Royal Astronomer of Ireland,*
&c. &c.

Read October 25, 1830.

INTRODUCTION.

THE present Supplement contains the integration of some partial differential equations, to which I have been conducted by the view of mathematical optics, proposed in my former memoirs. According to that view, the geometrical properties of an optical system of rays, may be deduced by analytic methods, from the form of one characteristic function; of which the partial differential coefficients of the first order, taken with respect to the three rectangular coordinates of any proposed point of the system, are, in the case of ordinary light, equal to the index of refraction of the medium, multiplied by the cosines of the angles which the ray passing through the point makes with the axes of coordinates: and as these cosines are connected by the known relation that the sum of their squares is

VOL. XVI. P

Figure 3.16. Two pages from *Supplements to the Essay on the Theory of Systems of Rays* by William R. Hamilton (1805–1865).

be the canonical moments of the extremal $x(t)$ at times t and $\overline{t}$, respectively. Then we have

$$\begin{aligned} y &= W_x(\overline{t},\overline{x},t,x), & H(t,x,y) &= -W_t(\overline{t},\overline{x},t,x),\\ \overline{y} &= -W_{\overline{x}}(\overline{t},\overline{x},t,x), & H(\overline{t},\overline{x},\overline{y}) &= -W_{\overline{t}}(\overline{t},\overline{x},t,x). \end{aligned} \tag{3.47}$$

Assuming that the extremal through $(\overline{t},\overline{x})$ and (t,x) is embedded in a Mayer field with eikonal S, we can easily prove (3.47). In fact, in this case

$$W(\overline{t},\overline{x},t,x) = S(t,x) - S(\overline{t},\overline{x}),$$

S solves the Hamilton–Jacobi equation and $S_x(t,x)$ is the dual slope field,

$$\begin{cases} S_t(t,x) = -H(t,x(t),S_x(t,x(t))),\\ S_x(t,x(t)) = L_v(t,x(t),x'(t)), \end{cases}$$

see (3.44). Consequently, for $\overline{t}$ and $\overline{x}$ fixed and varying (t,x), we get

$$y = S_x(t,x) = W_x(\overline{t},\overline{x},t,x), \qquad W_t = S_t = -H(t,x,S_x)$$

and similarly for the other equations in (3.47), by keeping (t,x) fixed and varying $(\overline{t},\overline{x})$.

The equations in (3.47) can be used in two different ways. Suppose we know at time $\overline{t}$ the position a and we specify at the same instant

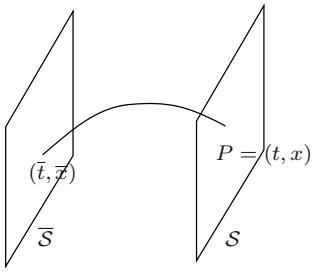


Figure 3.17. The scheme of William R. Hamilton (1805–1865).

the velocity v of an extremal or, equivalently, the canonical momentum, $b := L_v(\bar{t}, a, v)$. Then the equation $b = -W_{\overline{x}}(\bar{t}, \overline{a}, t, x)$ allows us to find x as a function of (t, a, b), $x = x(t, a, b)$ and then to compute the canonical momentum $(t, x(t, a, b))$. Inserting $x = x(t, a, b)$ in $y = W_x(\bar{t}, \overline{a}, t, x)$, we see that

$$\begin{cases} x(t) = x(t, a, b), \\ y(t) = W_x(\bar{t}, a, t, x(t)) \end{cases} \tag{3.48}$$

is a solution of Hamilton's equations. In other words, the system associated to the Lagrangian F or the Hamilton H, or, even better, the solution of Hamilton's equations can be obtained from W in an algebraic way and by differentiation.

Another way of looking at (3.47) is as a couple $(x(t, a, b), y(t, a, b))$ that for every t yield a family of transformations $\varphi^t(a, b) = (x(t, a, b), y(t, a, b))$ that describe the Hamiltonian flux. Actually, the system is fully described if we are able to reconstruct *all* solutions of Hamilton's equations, i.e., a family depending on $2N$ parameters of solutions of Hamilton's equations. To do this, we need a nondegeneracy condition first emphasized by Carl Jacobi (1804–1851). Observe that for a fixed $\bar{t}$, $E(t, x, a) := W(\bar{t}, a, t, x) = S(t, x) - S(\bar{t}, a)$ is a family of eikonals with N parameters, hence a family of solutions of the Hamilton–Jacobi equations,

$$y = E_x(t, x, a), \qquad b = -E_a(t, x, a).$$

3.59 Definition. *We say that a function φ of class C^2 $\varphi(t, x, a)$, where $x = (x^1, x^2, \dots, x^N)$, $a = (a_1, a_2, \dots, a_N)$ and $\lambda \in \mathbb{R}$, defined on an open set of $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$, is a* complete integral *of the Hamilton–Jacobi equation with Hamiltonian H if for each a, the function $(t, x) \to \varphi(t, x, a)$ is a solution of the Hamilton-Jacobi equations*

$$\varphi_t + H(t, x, \varphi_x) = 0 \tag{3.49}$$

and

$$\det \left[\frac{\partial^2 \varphi}{\partial x^i \partial a_j} \right] \neq 0. \tag{3.50}$$

3.60 Theorem (Jacobi). *Let φ be a complete integral of Hamilton–Jacobi equation* (3.50)*. Then we can find a family of independent solutions of the canonical equations*

$$\begin{cases} x' = H_p(t, x, p), \\ p' = -H_x(t, x, p) \end{cases}$$

depending on $2N$ parameters $a = (a_1, \dots, a_N)$ and $b = (b^1, \dots, b^N)$ by solving in $(x(t), p(t))$ the system

$$\begin{cases} p(t) = \varphi_x(t, x, a), \\ b = -\varphi_a(t, x, a). \end{cases} \tag{3.51}$$

Proof. First, we notice that because of the assumptions we can solve $\varphi_a(t, x, a) = -b$. We need to prove that $(x(t), p(t))$ solves the Hamilton equations. It is so since for each a, the function $\varphi(t, x, a)$ solves the Hamilton–Jacobi equation and φ_x is the dual of the slope field of a Mayer field. Or, more directly, differentiating with respect to x^i, we find

$$\varphi_{tx^i} + H_{p_k}\varphi_{x^k x^i} + H_{x^i} = 0, \tag{3.52}$$

hence

$$\begin{aligned} 0 &= H_{x^i} x^i_{a_j} + \varphi_{tx^i} x^i_{a_j} + H_{p_k}\varphi_{x^k x^i} x^i_{a_j} \\ &= H_{x^i} x^i_{a_j} + \frac{d}{dt}\varphi_{a_j} + H_{p_k}\frac{d}{dt}\varphi_{x^k a_j} = H_{x^i} x^i_{a_j}. \end{aligned}$$

Next, we differentiate (3.49) with respect to a_i to find

$$\varphi_{ta_i} + H_{p_k}\varphi_{x^k a_i} = 0$$

and the second of (3.51) with respect to t to find

$$\varphi_{ta_i} + \varphi_{a_i x^k}(x^k)' = 0.$$

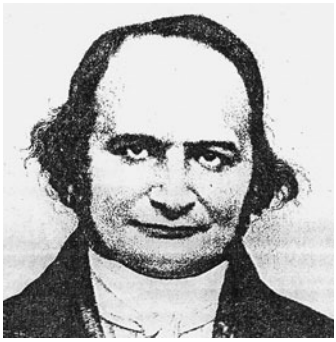
The last equations yield $H_{p_i} = (x^i)'$ because of (3.50). Finally, from the first of (3.51) we get

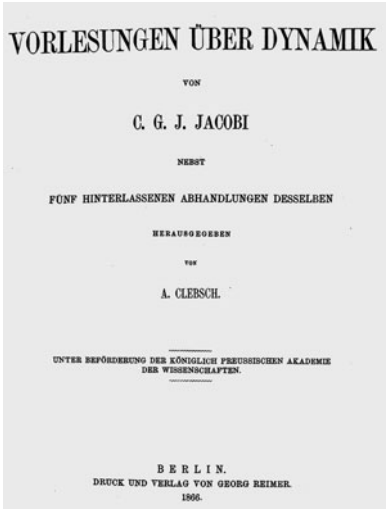
$$p' = \varphi_{x^i t} + \varphi_{x^i x^k}(x^k)' = \varphi_{x^i t} + H_{p_k}\varphi_{x^i x^k}$$

that, together with (3.52), yields $p'_i = -H_{x^i}$. □

3.2.2 Canonical transformations

We now illustrate, although briefly, a fundamental component of Hamiltonian formalism that finds its natural evolution in the symplectic geometry.




VORLESUNGEN ÜBER DYNAMIK

VON

C. G. J. JACOBI

NEBST

FÜNF HINTERLASSENEN ABHANDLUNGEN DESSELBEN

HERAUSGEGEBEN

VON

A. CLEBSCH.

UNTER BEFÖRDERUNG DER KÖNIGLICH PREUSSISCHEN AKADEMIE DER WISSENSCHAFTEN.

BERLIN.
DRUCK UND VERLAG VON GEORG REIMER.
1866.

Figure 3.18. Carl Jacobi (1804–1851) and the frontispiece of his *Vorlesungen über Dynamik*.

a. Canonical transformations

Consider an autonomous Hamiltonian system

$$x' = H_y(x, y), \qquad y' = -H_x(x, y). \tag{3.53}$$

By introducing the column vector $z := (x, y)^T$ and the *special symplectic matrix*

$$\mathbf{J} := \begin{pmatrix} 0 & \boxed{\mathrm{Id}_N} \\ \boxed{-\mathrm{Id}_N} & 0 \end{pmatrix},$$

and denoting by $H_z = \nabla H(z) = \mathbf{D}H(z)^T$ the gradient of the function H, the equations in (3.53) can be written as

$$z' = \mathbf{J}\mathbf{D}H(z)^T = \mathbf{J}H_z(z). \tag{3.54}$$

First, we want to find sufficient conditions so that a transformation $u : \mathbb{R}^{2N} \to \mathbb{R}^{2N}$, $\xi \to u(\xi)$, maps a Hamiltonian system into the transformed Hamiltonian system $H \circ u$. If

$$z(t) = u(\zeta(t)), \qquad K(\zeta) := H(u(\zeta)),$$

we have

$$z' = \mathbf{D}u(\zeta)\zeta', \qquad K_\zeta(\zeta) = \mathbf{D}u(\zeta)^T H_z(u(\zeta)).$$

Since $\mathbf{J}^2 = -\mathrm{Id}$, we write (3.54) as $-\mathbf{J}z' = H_z(z)$ and conclude that $z = u(\zeta(t))$ is a solution of (3.54) if and only if $-\mathbf{J}\mathbf{D}u(\zeta)\zeta' = H_z(u(\zeta))$ or, equivalently,

$$-\mathbf{D}u(\zeta)^T\mathbf{J}\mathbf{D}u(\zeta)\zeta' = K_\zeta(u).$$

In conclusion,

$$\zeta' = \mathbf{J}K_\zeta(\zeta)$$

if and only if the transformation fulfills

$$\mathbf{D}u^T\mathbf{J}\mathbf{D}u = \mathbf{J}.$$

3.61 Definition. *A* $(2N)\times(2N)$ *matrix* $\mathbf{A}$ *is called* symplectic *if*

$$\mathbf{A}^T\mathbf{J}\mathbf{A} = \mathbf{J}. \tag{3.55}$$

A transformation $u : \mathbb{R}^{2N} \to u(z) \in \mathbb{R}^{2N}$ *is called a* canonical transformation *if its Jacobian matrix is symplectic,*

$$\mathbf{D}u(z)^T\mathbf{J}\mathbf{D}u(z) = \mathbf{J} \qquad \forall z \in \mathbb{R}^{2N}. \tag{3.56}$$

3.62 Remark. We notice the following:

(i) The $(2N)\times(2N)$ symplectic matrices have determinant ± 1 and form a subgroup of the linear group $GL(2N,\mathbb{R})$, called the *symplectic group* and denoted by $Sp(N,\mathbb{R})$; moreover, one sees that the transpose and the inverse of a symplectic matrix are symplectic.

(ii) The canonical transformations are local diffeomorphisms, but, in general, they are not global diffeomorphisms; for instance, the map $(\xi,\eta) \in \mathbb{R}^4 \to (x,y) \in \mathbb{R}^4$

$$x^1 := \frac{1}{2}(\xi^1\xi^1 - \xi^2\xi^2), \qquad x^2 := \xi^1\xi^2,$$
$$y_1 := \frac{1}{|\xi|^{-2}}(\xi^1\eta_1 - \xi^2\eta_2), \qquad y_2 := \frac{1}{|\xi|^{-2}}(\xi^1\eta_2 - \xi^2\eta_1)$$

is canonical but is not a global diffeomorphism.

(iii) One can show that the transformations that preserve the Hamiltonian structure of an autonomous system are the *generalized canonical transformations* characterized by the condition

$$\mathbf{D}u(z)^T\mathbf{J}\mathbf{D}u(z) = \lambda\,\mathbf{J} \qquad \forall z \in \mathbb{R}^{2N},$$

where λ is a nonzero constant.

The canonical transformations can be characterized in several equivalent ways that we briefly present omitting the proofs of equivalence.

First, we introduce the differential 1-form of Poincaré $\theta = \theta(t, x, y)$ on $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$

$$\theta := y_i dx^i \qquad \text{in } \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N,$$

with differential

$$\omega := d\theta = dy_i \wedge dx^i,$$

see Chapter 4, called the *symplectic form*, and the 1-*form of Cartan*

$$K_H := y_i dx^i - H\, dt = \theta - H\, dt \qquad \text{in } \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$$

with differential

$$dK_H = \omega - dH \wedge dt.$$

One can show that u is a canonical transformation if and only if

$$u^{\#}\omega = \omega. \tag{3.57}$$

In other words, all canonical transformations are the only transformations that preserve the symplectic form,

$$dY_i \wedge dX^i = dy_i \wedge dx^i \qquad \text{if } (x, y) = u(X, Y).$$

3.63 ¶. Using the Stokes formula, see Chapter 4, infer from (3.57) that the canonical transformations preserve the surface integral $\int_S \omega$ for all 2-surfaces S.

3.64 Exact canonical transformations. In terms of the Poincaré 1-form θ we have $\omega = d\theta$, hence (3.57) amounts to

$$d\theta = \omega = u^{\#}\omega = u^{\#} d\theta = du^{\#}\theta,$$

that is to

$$d(\theta - u^{\#}\theta) = 0.$$

We can therefore state that, in a simply connected domain, u is canonical if and only if there exists ψ such that

$$u^{\#}\theta = \theta + d\psi. \tag{3.58}$$

Transformations for which (3.58) holds are called *exact canonical transformations with generator* ψ. They preserve the line integral $\int_\gamma \theta$ for all closed lines γ.

3.65 Canonical maps parametrized by time. Let u be a canonical map, $\overline{H}$ a Hamiltonian function, and set $U(t, z) := (t, u(z))$ and $H := U^{\#}\overline{H}$. It is easily seen that

$$U^{\#} K_{\overline{H}} = K_H + d\psi,$$

K_H and $K_{\overline{H}}$ being the Cartan forms associated to H and $\overline{H}$, respectively.

It is convenient to consider 1-parameter families $\{u^t\}$ of exact canonical maps with generators $\{\psi^t\}$, i.e., such that

$$(u^t)^{\#}\theta = \theta + d\psi^t \qquad \forall t.$$

In this case, if

$$U(t,z) := (t, u(z)) = (t, X(t,x,y), Y(t,x,y)), \qquad \psi(t,z) := \psi^t(z),$$

one sees that

$$U^{\#}\mathcal{K}_{\overline{H}} = \mathcal{K}_H + d\psi \tag{3.59}$$

and that every Hamiltonian system

$$\overline{x}' = \overline{H}_{\overline{y}}(t,\overline{x},\overline{y}), \qquad \overline{y}' = -\overline{H}_{\overline{x}}(t,\overline{x},\overline{y})$$

pulls back to a Hamiltonian system

$$x' = H_y(t,x,y), \qquad y' = -H_x(t,x,y)$$

where

$$H := U^{\#}\overline{H} + \frac{\partial\psi}{\partial t} - Y \bullet X_t\,; \tag{3.60}$$

moreover, a generic transformation U is a canonical transformation if and only if (3.59) holds for some ψ, H and $\overline{H}$ being related by (3.60).

3.66 Lagrange's parentheses. The *Lagrange parentheses* of a transformation $u(x,y) = (X(x,y), Y(x,y))$ are defined by

$$\begin{aligned}
&[x^i, x^k] := Y_{x^i} X_{x^k} - Y_{x^k} X_{x^i},\\
&[y_i, y_k] := Y_{y_i} X_{y_k} - Y_{y_k} X_{y_i},\\
&[y_i, x^k] = -[x^k, y_i] := Y_{y_i} X_{x^k} - Y_{x^k} X_{y_i}.
\end{aligned}$$

One proves that the transformation u is a canonical transformation if and only if

$$[x^i, x^k] = 0, \qquad [y_i, y_k] = 0, \qquad [y_i, x^k] = \delta_i^k. \tag{3.61}$$

3.67 Hamilton flows. Consider the autonomous Hamiltonian system

$$X' = H_y(X,Y), \qquad Y' = -H_x(X,Y)$$

with initial conditions $X(0,x,y) = x$ and $Y(0,x,y) = y$ and denote by

$$u^t(x,y) := (X(t,x,y), Y(t,x,y))$$

its integral flux. It is easily seen that Lagrange's parentheses of u^t are constant in time. Since the flux is the identity for $t = 0$, we deduce

$$[x^i, x^k] = 0, \qquad [y_i, y_k] = 0, \qquad [y_i, x^k] = \delta_i^k$$

for $t = 0$, hence for all $t \geq 0$; consequently, the Hamiltonian flux $\{u^t\}$ defines a 1-parameter family of canonical maps. The converse is also true: *Every 1-parameter family of canonical maps is the Hamiltonian flux of a suitable autonomous Hamiltonian.*

3.68 Poisson's parentheses. The *Poisson parenthesis* of two functions $f(t,z)$ and $g(t,z)$, $z \in \mathbb{R}^{2N}$, is the symplectic product of the gradients of the two functions

$$\{f,g\} := \Big(\frac{\partial f}{\partial z}\Big)^T \mathbf{J}\Big(\frac{\partial g}{\partial z}\Big).$$

If $z = (x,y) \in \mathbb{R}^{2N}$, then

$$\{f,g\} = \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y_i} - \frac{\partial g}{\partial x^i}\frac{\partial f}{\partial y_i}.$$

In terms of Poisson's parentheses, Hamilton's equations can be written as

$$y_i' = \{y_i, H\}, \qquad (x^i)' = \{x^i, H\}.$$

More generally, for every solution $z(t)$ of $z' = \mathbf{J}H_z(z)$ and every observable $F : \mathbb{R}^{2N} \to \mathbb{R}$, i.e., every function $F : \mathbb{R}^{2N} \to \mathbb{R}$, we have

$$\frac{d}{dt}F(z(t)) = \{F, H\};$$

consequently, an observable is a first integral, i.e., is constant along the Hamiltonian flux if and only if its Poisson's parenthesis with the Hamiltonian vanishes. Moreover, the identities

$$\{x^i, x^k\} = \{y_i, y_k\} = 0, \qquad \{x^i, y_j\} = -\{y_j, x^i\} = \delta^i_j$$

hold for the *fundamental parentheses* $\{x^i, x^k\}$, $\{y_i, y_k\}$ and $\{x^i, y_k\}$. Notice that, in general, for $F : \mathbb{R} \times \mathbb{R}^{2N} \to \mathbb{R}$ we have

$$\frac{d}{dt}F(t,z(t)) = \frac{\partial F}{\partial t}(t,z(t)) + \{F,H\}.$$

Canonical maps can be characterized in terms of Poisson's parentheses. One proves that the following claims are equivalent:

(i) The transformation $\zeta \to u^t(\zeta)$ is canonical.
(ii) For every couple of functions f and g we have $\{f,g\} \circ u^t = \{f \circ u^t, g \circ u^t\}$.
(iii) The fundamental parentheses are preserved by u^t.

Finally, it is easy to show that for Poisson's parentheses the following Jacobi's identity holds

$$\{f,\{g,h\}\} + \{g,\{h,f\}\} + \{h,\{f,g\}\} = 0.$$

This yields, in particular, the following theorem.

3.69 Theorem (Poisson). *The parenthesis $\{f,g\}$ of two first integrals of an autonomous Hamiltonian system is again a first integral.*

3.70 Remark. Poisson's theorem might suggest that we can easily produce as many first integrals as we want, but this is not the case: Completely integrable systems are quite rare. The torus of completely integrable systems may disappear for Hamiltonian small perturbations. The celebrated Kolmogorov–Arnold–Moser theory gives sufficient conditions in order for an invariant torus to be preserved for small perturbations.

3.71 Jacobi's method. Finally, we shall show how Jacobi's theorem can be reinterpreted in terms of canonical maps and how canonical maps can be generated from a given function.

In the proof of Jacobi's theorem we saw that, given a function $S(t,x,a)$ with $\det S_{ax}(t,x,a) \neq 0$, we can define, at least locally, a map $(t,a,b) \to U(t,a,b)$,

$$U(t,a,b) := (t, X(t,a,b), Y(t,a,b)),$$

by choosing $x = X(t,a,b)$ so that

$$S_a(t, X(t), a) = -b$$

and then computing

$$Y(t,a,b) := S_x(t, X(t,a,b), a).$$

If

$$\Psi(t,a,b) := S(t, X(t,a,b), a),$$

then for fixed t we have

$$Y_i\, DX^i - b_i da^i = d\Psi.$$

According to the above, $U^t(x,y) := U(t,x,y)$ is an exact canonical map and each Hamiltonian $H(t,x,y)$ is transformed into the new Hamiltonian $K(t,a,b)$ given by

$$K = H(t,X,Y) + \Psi_t + Y \bullet X_t$$

and every solution $(a(t), b(t))$ of

$$a' = K_b(t,a,b), \qquad b' = -K_a(t,a,b)$$

is mapped into a solution $(x(t), y(t))$ of

$$x' = H_y(t,x,y), \qquad y' = -H_x(t,x,y)$$

and vice versa. If for each a the function $(t,x) \to S(t,x,a)$ is a solution of the Hamilton–Jacobi equation

$$S_t + H(t,x,S_x) = 0,$$

then

$$\psi_t = S_t(t,X,a) + S_{x^i}(t,X,a)X^{i'} = -H(t,X,Y) + Y^i X^{i'},$$

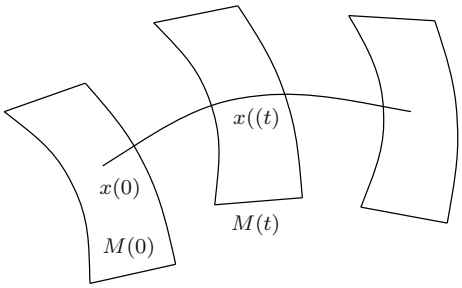


Figure 3.19. Wave fronts.

hence $H(t, X, Y)$ is transformed into the Hamiltonian $K(t, a, b) = 0$. Therefore $U^t(X, Y)$ transforms the initial Hamiltonian system into the Hamiltonian system

$$a' = 0, \qquad b' = 0.$$

The proof of Jacobi's theorem consists, therefore, in finding a canonical map that rectifies the Hamiltonian flux.

As a consequence, if we can find N first integrals, i.e., a complete integral of the Hamilton–Jacobi equation, the corresponding Hamiltonian system is completely integrable. One shows that this is always possible in a neighborhood of a nonsingular point (via a canonical map), but not globally.

b. Analytic mechanics and Schrödinger equations

Let $H(x, y)$ be a Hamiltonian with N degrees of freedom and S be the eikonal of one of its Mayer fields. Since H is independent of time, $S(t, x)$ has the form

$$S(t, x) = W(x) - Et,$$

where $W : \mathbb{R}^n \to \mathbb{R}$ and $E \in \mathbb{R}$, see Exercise 3.81. As we have seen, the level lines of $S(t, x)$,

$$M_{t,\alpha} := \Big\{ x \,\Big|\, S(t, x) = \alpha \Big\},$$

are the "fronts of propagation" of the extremals of the field and, for fixed α, the velocity of the surfaces $M_t := M_{t,\alpha}$ measured along the normal to the surface is

$$v(x) := \frac{E}{|W_x(x)|^2} W_x(x).$$

In fact, if $x(t)$ is a curve with $x(t) \in M_t$, i.e., $W(x(t)) - Et = \alpha$ and with $x'(t) \perp M_t$, then

$$\begin{cases} W_x(x(t)) \bullet x'(t) = E, \\ x'(t) = \lambda(x(t)) W_x(x(t)), \end{cases}$$

hence $|x'(t)| = E/|W_x(x)|$.

Suppose now that we see the front of propagation of the eikonal as surfaces of constant phase of waves of the type

$$\phi = \psi_0 e^{A(x)+i(L(x)-ct)}, \tag{3.62}$$

possibly with variable amplitude $\psi_0 e^{A(x)}$, that move with the same landscape and velocity. Since the surfaces of constant phase are

$$N_t = \Big\{ x \in \mathbb{R}^N \;\Big|\; L(x) - ct = \beta \Big\},$$

from $M_t = N_t$ for all t we infer that $L(x) - ct$ is a function of $W(x) - Et$, see Chapter 5 of [GM4]. Since $\{M_t\}$ and $\{N_t\}$ move with the same normal velocity, $L(x) - ct$ and $W(x) - Et$ have to be proportional,

$$L(x) - ct = \frac{1}{h}(W(x) - Et), \qquad h \in \mathbb{R},\ h > 0.$$

Therefore, we conclude that the surfaces with constant phase of waves of the type

$$\phi = \phi_0 e^{A(x)+i\frac{1}{h}(W(x)-E\,t)} \tag{3.63}$$

with arbitrary $A(x)$ and $h > 0$ describe the trajectories of a Hamiltonian system.

The waves in (3.63) solve various partial differential equations. We can compute, for example,

$$\begin{aligned} \Delta\phi &= \Big(\frac{1}{h^2}|W_x|^2 \quad + \frac{i}{h}(2\, A_x \bullet W_x \, - \Delta W) + (|A_x|^2 + \Delta A) \Big)\, \phi, \\ \frac{\partial \phi}{\partial t} &= -\frac{E}{h}\phi, \\ \frac{\partial^2 \phi}{\partial t^2} &= -\frac{E^2}{h^2}\phi, \end{aligned} \tag{3.64}$$

by eliminating, globally or partially, the terms containing ϕ.

For example, we have

$$\Delta\phi + \frac{1}{E^2}\Big(|W_x|^2 + 2ih(\, W_x \bullet A_x \, + \Delta W) + h^2(\Delta A + |A_x|^2) \Big)\phi_{tt} = 0.$$

When $h \to 0$, this converges to the linear wave equation with velocity equal to the velocity of propagation of the wave front

$$\Delta\phi - \frac{1}{n^2}\phi_{tt} = 0, \qquad n(x) := \frac{E}{|W_x(x)|}.$$

In this way, for $h \to 0$ we find the geometric optics.

3.72 Example. For a material point of mass m with Hamiltonian $H(x,y) := \frac{|y|^2}{2m} + V(x)$, the reduced Hamilton–Jacobi equation, see Exercise 3.81, is

$$|W_x(x)| = \sqrt{2m}\sqrt{E - V(x)},$$

thus (3.64) becomes, if we neglect the terms in $h\phi$ and $h^2\phi$, the *Schrödinger equation* of ondulatory mechanics

$$-\frac{h^2}{2m}\Delta\phi + V(x)\phi = E\,\phi,$$

or, noticing that $\phi_t = -i\frac{E}{h}\phi$, the *Schrödinger equation*

$$-\frac{h^2}{2m}\Delta\phi + V(x)\phi = ih\frac{\partial\phi}{\partial t}.$$

3.3 Exercises

3.73 ¶. Study the equation

$$\left(\frac{u'}{\sqrt{1+(u')^2}}\right)' = H \qquad \text{in }]-1,1[,\ H \in \mathbb{R}.$$

3.74 ¶. Consider the functional

$$\mathcal{F}(\theta) := \int_{\varphi_1}^{\varphi_2} \sqrt{(\theta'(\varphi))^2 + \sin^2\varphi}\; d\varphi,$$

that represents the length of a curve $\theta = \theta(\varphi)$, $\varphi_1 \le \varphi \le \varphi_2$, on the unit sphere of $\mathbb{R}^3$ in polar coordinates (φ,θ). Show that for $\theta(\varphi) := \arcsin\varphi$, the conservation law of energy holds, although $\theta(\varphi)$ is not an extremal.

3.75 ¶. Let $(x(s),y(s))$ be the arc length parametrization of a curve through the origin and symmetric with respect to the x-axis and with $y(s) > 0$ for $0 < s < \ell/2$, ℓ being its total length. The area enclosed by the curve is

$$A := 2\int_0^{\ell/2} y(1-(y')^2)^{1/2}\,ds.$$

Using the law of conservation of energy, prove that A is maximal if the curve is the circle $(x - \ell/(2\pi))^2 + y^2 = \ell^2/(4\pi)$. Find the same result in polar coordinates.

3.76 ¶ Obstacle problems. Suppose we want to minimize the area of the graph of a function in $\Omega \times \mathbb{R}$ with prescribed boundary and constrained to stay above an obstacle given by the graph of a function ψ, i.e.,

$$\int_\Omega \sqrt{1+|Du|^2}\,dx, \qquad u|_{\partial\Omega} = \varphi,\ u \ge \psi \text{ in } \Omega.$$

Prove that the Euler–Lagrange equilibrium equation is replaced by the inequality

$$\int_\Omega \frac{DuD(v-u)}{\sqrt{1+|Du|^2}}\,dx \ge 0 \qquad \forall v \text{ with } v = 0 \text{ on } \partial\Omega \text{ and } v \ge \psi \text{ in } \Omega.$$

3.77 ¶ Variational inequalities. Let $\mathbb{K}$ be a convex set in the space $C^1_\varphi(\Omega, \mathbb{R}^N)$. Show that the condition of vanishing of the first variation for

$$\int_\Omega F(x, u, Du)\, dx \to \min, \qquad u \in \mathbb{K},$$

is replaced by

$$\int_\Omega \Big(F_p(x, u, Du) D(v - u) + F_u(x, u, Du)(v - u) \Big)\, dx \geq 0 \qquad \forall v \in \mathbb{K}.$$

3.78 ¶. Show that every extremal of $\int_a^b F(x, u')\, dx$ with $F(x, p)$ convex in p for every fixed x is a minimizer (among the functions with the same boundary values).

3.79 ¶. Show the following:

(i) $X(x, y) = y$, $Y(x, y) = -x$ is an exact canonical map with generator $\psi(x, y) := x \bullet y$, whereas for $N = 1$, the map $(x, y) \to (y, x)$ is not a canonical transformation.

(ii) (Poincaré transform) For $N = 1$ the map

$$X(x, y) := \sqrt{x} \cos(2y), \qquad Y(x, y) := \sqrt{x} \sin(2y)$$

is an exact canonical transformation with generator $\psi(x, y) = \frac{1}{4} x(\sin(4y) - 4y)$.

(iii) (Levi–Civita transformation) For $N = 3$ the map

$$X(x, y) := |y|^2 x - 2\, x \bullet y\,, \qquad Y(x, y) := \frac{y}{|y|^2}$$

is an exact canonical transformation with generator $\psi(x, y) := -2\, x \bullet y$.

3.80 ¶ Harmonic oscillator. The equation of the harmonic oscillator is $x'' + \omega^2 x = 0$. It is the Euler–Lagrange equation for the Lagrangian $L(t, x, v) := \frac{v^2}{2\omega} - \frac{\omega x^2}{2}$ and Hamiltonian $H(x, y) := y \bullet v - L(t, x, v)$, $y = L_v(t, x, v)$, i.e.,

$$H(x, y) = \frac{\omega}{2}(x^2 + y^2),$$

and the corresponding Hamiltonian system is

$$\begin{cases} x' = \omega y, \\ y' = -\omega x. \end{cases}$$

Show that the canonical Poincaré transformation $x = \sqrt{2\tau} \cos\varphi$, $y = \sqrt{2\tau} \sin\varphi$ transforms the system into

$$\tau' = 0, \qquad \varphi' = -\omega.$$

3.81 ¶ Reduced Hamilton–Jacobi equation. Consider an autonomous Hamiltonian $H(q, p)$ and let $S(t, q)$ be a solution of the associated Hamilton–Jacobi equation. Show that $S(t, q) = W(q) + Et$, where $E \in \mathbb{R}$ and W satisfies the *reduced Hamilton–Jacobi equation*

$$H\Big(q, \frac{\partial W}{\partial q}\Big) = E.$$

[*Hint.* Since H is independent of t, the conservation of energy holds:

$$H\Big(q, \frac{\partial S}{\partial q})\Big) = E,$$

i.e., $S(t, q) = S(t, q, E)$ and $H(q, \frac{\partial S}{\partial q}(t, q, E)) = E$.]

3.82 ¶. Consider a material point of mass m, moving (subject to a conservative field with potential $V(x)$) on a line. The Hamiltonian of the system is

$$H(x,p) := \frac{p^2}{2m} + V(x)$$

and its reduced Hamilton–Jacobi equation is

$$\frac{1}{2m}\Big(\frac{\partial W}{\partial x}\Big)^2 + V(x) = E$$

that can be easily integrated:

$$W(x,E) = \sqrt{2m}\int_{x_0}^{x} \sqrt{E - V(\xi)}\, d\xi.$$

The generated canonical map is then

$$\begin{cases} p = \dfrac{\partial W}{\partial x} = \sqrt{2m(E - V(x))}, \\ \beta = \dfrac{\partial W}{\partial E} = \sqrt{\dfrac{m}{2}}\displaystyle\int_{x_0}^{x} \frac{d\xi}{E - V(\xi)} \end{cases}$$

where $\beta = t - t_0$. We thus find

$$dt = \pm \frac{dx}{\frac{2}{m}(E - V(x))},$$

see Section 6.3 of [GM1].

3.83 ¶. Show that if $x' = \mathbf{J}x$, then $|x| = const.$

4. Differential Forms

In this chapter we present Stokes's theorem and Poincaré's lemma in the general setting of differential forms and illustrate some of the relevant applications of the theory of differential k-forms.

4.1 Multivectors and Covectors

4.1.1 The exterior algebra

We begin by illustrating some basic elements of the so-called *exterior algebra* over a vector space.

a. Alternating bilinear maps, antisymmetric matrices and 2-vectors

Let V and Z be two vector spaces over $\mathbb{R}$. A map $f : V \times V \to Z$, $f = f(x, y)$, that is linear on each factor is called a *bilinear map*. If $(e_1, e_2, \dots, e_n)$ is a basis of V, then f is uniquely determined by the n^2 values $\{f(e_i, e_j)\}$. In fact, if $x = \sum_{i=1}^n x^i e_i$ and $y = \sum_{i=1}^n y^i e_i$, we have

$$f(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x^i y^j, \qquad a_{ij} := f(e_i, e_j). \tag{4.1}$$

We say that the bilinear map $f(x, y)$ is *alternating* if $f(x, y) = -f(y, x)$ $\forall x, y \in V$ or, equivalently, if the matrix $\mathbf{A} := (a_{ij})$, $a_{ij} := f(e_i, e_j)$, is antisymmetric, i.e., if $\mathbf{A}^T = -\mathbf{A}$. In this case the map f is identified by the values a_{ij} with $i < j$ since, for $x = \sum_{i=1}^n x^i e_i$ and $y = \sum_{i=1}^n y^i e_i$, we have

$$\begin{aligned} f(x, y) &= \sum_{i,j=1,n} a_{ij} x^i y^j \\ &= \sum_{i<j} a_{ij} x^i y^j + \sum_{i>j} a_{ij} x^i y^j = \sum_{i<j} a_{ij} (x^i y^j - x^j y^i). \end{aligned} \tag{4.2}$$

Formula (4.2) hints at the following three facts:

M. Giaquinta and G. Modica, *Mathematical Analysis, Foundations and Advanced Techniques for Functions of Several Variables*, DOI 10.1007/978-0-8176-8310-8_4,

(i) An alternating bilinear map defines in a natural way a *linear map* on a space of dimension

$$\#\Big\{(i,j)\,|\,i<j\Big\}=\frac{n(n-1)}{2}=\binom{n}{2}.$$

(ii) Given two vectors x and y, we may consider a new object, called a 2-*vector* denoted by $x\wedge y$ with coordinates $\{(x^iy^j-x^jy^i)\}_{i<j}$, so that all alternating bilinear maps act linearly on these objects.

(iii) The map $(x,y)\to x\wedge y:=\{x^iy^j-x^jy^i\}_{i<j}$ from $\mathbb{R}^n$ into $\mathbb{R}^q$, $q:=\binom{n}{2}$, is alternating, i.e., $x\wedge y=-y\wedge x$.

4.1 Definition. *Let V be a vector space of dimension n. The vector space Λ_2V of* 2-vectors *of V is a vector space of dimension $\binom{n}{2}$, together with an alternating bilinear map $\cdot\wedge\cdot: V\times V\to\Lambda_2V$, the image of which generates Λ_2V. For $x,y\in V$ we call $x\wedge y$ the* exterior product *of x and y.*

Let $(e_1, e_2, \dots, e_n)$ be a basis of V. Since the 2-vectors

$$\Big\{e_i\wedge e_j\,\Big|\,i,j=1\dots,n,\ i<j\Big\}$$

span Λ_2V and are $\binom{n}{2}$, they form a basis of Λ_2V, i.e.,

$$\Lambda_2V:=\Big\{\xi=\sum_{i<j}c^{ij}e_i\wedge e_j\,\Big|\,c^{ij}\in\mathbb{R}\Big\}.$$

Clearly, if $g:\Lambda_2V\to Z$ is a linear map, then the map $(x,y)\to g(x\wedge y)$ from $V\times V$ into Z is bilinear and alternating. The converse is also true.

4.2 Theorem. *Let V be a vector space of dimension n. For every alternating bilinear map $f: V\times V\to Z$ there is a unique linear map $g:\Lambda_2V\to Z$ such that*

$$f(x,y)=g(x\wedge y)\qquad\forall x,y\in V.$$

Proof. Since $e_i\wedge e_j$, $i<j$, form a basis of Λ_2V whenever $(e_1, e_2, \dots, e_n)$ is a basis of V, it suffices to specify the values of $g:\Lambda_2V\to Z$ on them

$$g(e_i\wedge e_j):=f(e_i,e_j),\qquad i<j. \tag{4.3}$$

Then $f(x,y)=g(x\wedge y)\ \forall x,y\in V$ because f is alternating, compare (4.2), and $x\wedge y=\sum_{i<j}(x^iy^j-x^jy^i)e_i\wedge e_j$ if $x=\sum_{i=1}^n x^ie_i$ and $y=\sum_{i=1}^n y^ie_i$. □

At first sight, the definition of Λ_2V appears as ambiguous since the definition of exterior product is not uniquely specified. However, from Theorem 4.2 we get the following.

4.3 Proposition. *Let V be a vector space of dimension n and let $\pi_1,\pi_2: V\times V\to\Lambda_2V$ be two alternating bilinear maps on V. Then $\pi_2=\phi\circ\pi_1$ for some linear isomorphism ϕ of Λ_2V.*

Proof. According to Theorem 4.2 there exist $g_1, g_2 : \Lambda_2 V \to \Lambda_2 V$ such that

$$\pi_2 = g \circ \pi_1, \qquad \pi_1 = h \circ \pi_2 \qquad \text{on } V \times V,$$

hence

$$\pi_2 = g \circ h \circ \pi_2, \qquad \pi_1 = h \circ g \circ \pi_1 \qquad \text{on } V \times V.$$

Since the images of π_1 and π_2 span $\Lambda_2 V$, it follows that $g \circ h = h \circ g = \mathrm{Id}_{\Lambda_2 V}$, i.e., $g = h^{-1}$, that is, g is a linear isomorphism. □

Consequently, not specifying the exterior product on V is equivalent to defining $\Lambda_2 V$ apart from a linear isomorphism. In terms of bases, Proposition 4.3 tells us that, if $(e_1, e_2, \dots, e_n)$ is a basis of V, choosing an exterior product is equivalent to choosing a basis in $\Lambda_2 V$ that we name $(e_i \wedge e_j)_{i<j}$.

Summing up, the 2-vectors in $\Lambda_2 \mathbb{R}^3$ have $\binom{3}{2} = 3$ components and can be written as

$$\xi = a\, e_y \wedge e_z + b\, e_x \wedge e_z + c\, e_x \wedge e_y, \qquad a, b, c \in \mathbb{R},$$

where

$$e_y \wedge e_z = -e_z \wedge e_y, \qquad e_x \wedge e_z = -e_z \wedge e_x, \qquad e_x \wedge e_y = -e_y \wedge e_x,$$

whereas every 2-alternating map, $f : \mathbb{R}^3 \times \mathbb{R}^3 \to Z$, uniquely defines a linear map $g : \Lambda_2 \mathbb{R}^3 \to \mathbb{R}$ such that $g(x \wedge y) = f(x, y)\ \forall x, y \in \mathbb{R}^3$.

b. *k*-alternating maps

We may generalize the previous construction to alternating k-linear maps. First, let us describe k-linear alternating maps in coordinates.

4.4 Permutations and signature. Let S be a finite set. A permutation of S is a bijective map of S onto itself. Let us denote the set of all permutations of S by $\mathcal{P}(S)$. Then $\#\mathcal{P}(S) = k!$ if S has k elements.

A *transposition* of $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ is a permutation that interchanges two distinct elements and leaves the others fixed. It is easy to see that every permutation σ of S may be obtained as compositions of successive transpositions of S. Although there are several ways of doing this, the *parity* of the number n of transpositions that compose σ depends only on σ. The *signature* of σ is then defined as $(-1)^n$ and, with some impropriety of language, is denoted by $(-1)^\sigma$.

4.5 Example. The signature of each transposition is clearly -1. If $S = \{1, 2, 3\}$, then the signature of the permutation $(1, 2, 3) \to (3, 2, 1)$ is -1, whereas the signature of the permutation $(1, 2, 3) \to (3, 1, 2)$ is $+1$.

4.6 Multiindices. Let S be a subset of $\{1, 2, \dots, n\}$. Denote by $|S|$ the cardinality of S. A *multiindex* of length k, $k \geq 2$, in $\{1, \dots, n\}$ or, simply, a *k-multiindex*, is an increasing k-tuple of numbers between 1 and n,

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k), \qquad 1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_k \leq n,$$

and we denote by $S(\alpha) := \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ the set of values of α. Similarly, a 1-multiindex is a number between 1 and n that, however, we denote by i instead of (i); moreover, let us introduce a unique 0-multiindex of length 0 denoted 0. Finally, let us notice that there is no k-multiindex in $\{1, \dots, n\}$ if $k > n$.

For $k \geq 0$, we denote the set of k-multiindices by $I(k, n)$. Since $I(k, n)$ is in a one-to-one correspondence to the subsets of $\{1, \dots, n\}$ with k elements, we have

$$\#I(k, n) = \binom{n}{k} \qquad \forall k,\ n.$$

In particular, $\#I(0, n) = 1$, $\#I(k, n) = 0$ if $k > n$.

Notice that a k-tuple $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ of $\{1, \dots, n\}$ either has at least two coincident elements or all elements are distinct. The last case happens if and only if we may order in increasing order the components of β, that is, there is $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k) \in I(k, n)$ and a permutation σ of α (or better of the set $S(\alpha) := \{\alpha_1, \alpha_2, \dots, \alpha_k\}$) such that $\beta = \sigma(\alpha)$.

Given $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k) \in I(k, n)$, we denote by $\overline{\alpha}$ the unique $(n-k)$-multiindex $\beta \in I(n-k, n)$ such that $S(\beta) = \{1, \dots, n\} \setminus S(\alpha)$. In particular, $\overline{\overline{\alpha}} = \alpha$, $\overline{0} = (1, 2, \dots, n)$, and

$$\overline{i} := (1, 2 \dots, i-1, i+1, \dots, n) \qquad \text{if } 1 \leq i \leq n.$$

Finally, we denote by $\sigma(\alpha, \overline{\alpha})$ the signature of the permutation that reorders the n-tuple $(\alpha, \overline{\alpha})$ to $(1, 2, \dots, n)$.

Let V and Z be vector spaces. A map $f : V^k := V \times V \times \cdots \times V \to Z$, $f = f(x_1, x_2, \dots, x_k)$, is said to be *$k$-linear* if it is linear on each factor. Clearly, if V has dimension n and $(e_1, e_2, \dots, e_n)$ is a basis of V, then f is completely defined by its values on the n^k k-tuples $(e_{i_1}, e_{i_2}, \dots, e_{i_k})$, $i_j \in \{1, \dots, n\}$, $j = 1, \dots, k$.

4.7 Definition. *A map $f : V^k \to Z$, $f = f(x_1, x_2, \dots, x_k)$, $k \geq 2$, is said to be* k-alternating *if it is k-linear and* alternating, *i.e.,*

$$f(x_1, \dots, x_i, \dots, x_j, \dots, x_k) = -f(x_1, \dots, x_j, \dots, x_i, \dots, x_k) \tag{4.4}$$

for all $x_1, x_2, \dots, x_k \in V$.

4.8 Proposition. *Let $f : V^k \to Z$ be k-linear. The following claims are equivalent:*

(i) *f is alternating.*
(ii) *$f(x_1, \dots, x_k) = 0$ for every k-tuple $(x_1, x_2, \dots, x_k)$ with $x_i = x_j$ for some $i \neq j$.*
(iii) *$f(x_1, \dots, x_k) = 0$ for every k-tuple of vectors $(x_1, x_2, \dots, x_k)$ that are linearly dependent.*

In particular, there are no nonzero k-alternating maps on a vector space of dimension n when $k > n$.

Proof. (i) ⇒ (ii). The proof is trivial. (ii) ⇒ (iii). Assume, for instance, that $x_1 = \sum_{i=2}^{k} a^k x_k$. Then

$$f(\sum_{i=2}^{n} a^i x_i, x_2, \dots, x_n) = \sum_{i=2}^{n} \alpha^i f(x_i, x_2, \dots, x_n) = 0.$$

(iii) ⇒ (i). Let $i \neq j$. The vectors

$$x_1, \dots, x_i + x_j, \dots, x_i + x_j, \dots, x_k$$

are linearly dependent, hence

$$\begin{aligned} 0 &= f(x_1, \dots, x_i + x_j, \dots, x_i + x_j, \dots, x_k) \\ &= f(x_1, \dots, x_i, \dots, x_i, \dots, x_k) + f(x_1, \dots, x_j, \dots, x_j, \dots, x_k) \\ &\quad + f(x_1, \dots, x_i, \dots, x_j, \dots, x_k) + f(x_1, \dots, x_j, \dots, x_i, \dots, x_k) \\ &= f(x_1, \dots, x_i, \dots, x_j, \dots, x_k) + f(x_1, \dots, x_j, \dots, x_i, \dots, x_k). \end{aligned}$$

□

Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k) \in I(k, n)$ and let σ be a permutation of α. If $\beta = \sigma(\alpha) = (\beta_1, \beta_2, \dots, \beta_k)$, it follows from (4.4) that

$$f(x_{\beta_1}, \dots, x_{\beta_k}) = (-1)^\sigma f(x_{\alpha_1}, \dots, x_{\alpha_k}) \tag{4.5}$$

if $f : V^k \to Z$ is k-alternating. In particular, if $(e_1, e_2, \dots, e_n)$ is a basis of V, every k-alternating map on V is uniquely defined by its values on the k-tuples

$$(e_{\alpha_1}, e_{\alpha_2}, \dots, e_{\alpha_k}), \qquad \alpha := (\alpha_1, \alpha_2, \dots, \alpha_k) \in I(k, n).$$

4.9 Determinant. k-alternating maps are strongly related to the notion of determinant of a matrix, as we shall soon see in detail. Let us recall that the *determinant* of a matrix $\mathbf{A} = (A^i_j) \in M_{n,n}$ is the number

$$\det \mathbf{A} := \sum_{\sigma \in \mathcal{P}(\{1,\dots,n\})} (-1)^\sigma A_1^{\sigma_1} A_2^{\sigma_2} \cdots A_n^{\sigma_n}, \tag{4.6}$$

and that, as it is easily seen,

(i) the determinant is an n-alternating map of the columns of $\mathbf{A}$,
(ii) we have

$$\det \mathbf{A}^T = \det \mathbf{A}.$$

c. k-vectors

4.10 Definition. *Let V be a vector space of dimension n and let $2 \leq k$. The vector space V of k-vectors is a vector space $\Lambda_k V$ of dimension $\binom{n}{k}$ together with a k-alternating map*

$$(v_1, v_2, \dots, v_k) \longrightarrow v_1 \wedge v_2 \wedge \cdots \wedge v_k$$

from V^k to $\Lambda_k V$ with image that spans $\Lambda_k V$, called the exterior product *of k vectors.*

As for 2-vectors, the following holds.

4.11 Theorem. *Let V be a vector space of dimension n. For every k-alternating map $f : V^k \to Z$ with values in a vector space Z there exists a unique linear map $g : \Lambda_k V \to Z$ such that $f(v_1, v_2, \dots, v_k) = g(v_1 \wedge v_2 \wedge \cdots \wedge v_k)$.*

4.12 Proposition. *Let V be a vector space of dimension n. If $\pi_1, \pi_2 : V^k \to \Lambda_k V$ are two exterior products on V, then there exists an isomorphism $g : \Lambda_k V \to \Lambda_k V$ such that $\pi_2 = \gamma \circ \pi_1$.*

Thus, not specifying the exterior product of k-vectors is equivalent to defining $\Lambda_k V$ apart from an isomorphism. This makes Definition 4.10 consistent. This can be summed up in the following theorem.

4.13 Theorem (Universal property). *Let V be a finite-dimensional vector space. There exist (and are unique up to isomorphisms) a vector space $\Lambda_k V$ and a k-alternating map $\cdot \wedge \cdots \wedge \cdot : V^k \to \Lambda_k V$ with the following property: For every k-alternating map $f : V^k \to Z$ there is a unique linear map $g : \Lambda_k V \to Z$ such that $f(x_1, x_2, \dots, x_k) = g(x_1 \wedge x_2 \wedge \cdots \wedge x_k)$ $\forall x_1, x_2, \dots, x_k \in V$.*

We conclude with the following proposition.

4.14 Proposition. *Let $\{v_1, v_2, \dots, v_k\}$ be vectors in V. Then $v_1 \wedge v_2 \wedge \cdots \wedge v_k = 0$ if and only if $v_1, v_2, \dots, v_k$ are linearly dependent.*

Proof. Since $\wedge$ is alternating, $v_1 \wedge v_2 \wedge \cdots \wedge v_k = 0$ if $v_1, v_2, \dots, v_k$ are linearly dependent by Proposition 4.8. Let us prove the converse by contradiction. Suppose that $v_1, v_2, \dots, v_k$ are linearly independent and let $(e_1, e_2, \dots, e_n)$ be a basis of V such that $e_i = v_i$ $\forall i = 1, \dots, k$. Since the k-vectors

$$e_\alpha := e_{\alpha_1} \wedge \cdots \wedge e_{\alpha_k}, \qquad \alpha \in I(k, n),$$

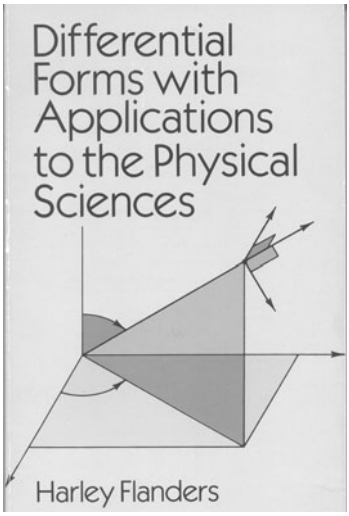
form a basis of $\Lambda_k V$, we would have $v_1 \wedge v_2 \wedge \cdots \wedge v_k = e_1 \wedge e_2 \wedge \cdots \wedge e_k \neq 0$, a contradiction. □

d. *k*-vectors in coordinates

Let us compute in coordinates the exterior product of k-vectors. Let V be a vector space of dimension n, let $(e_1, e_2, \dots, e_n)$ be a basis of V and let $v_1, v_2, \dots, v_n$ be n vectors of V. Let $\mathbf{A} = (a_i^j)$ be the $n \times n$ matrix with the coordinates of $v_1, v_2, \dots, v_n$ in the basis $e_1, e_2, \dots, e_n$ as columns, $v_i = \sum_{j=1}^n a_i^j e_j$. For all $\alpha, \beta \in I(k, n)$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ and $\beta = (\beta_1, \beta_2, \dots, \beta_k)$, we denote by

$$M_\alpha^\beta(\mathbf{A})$$

the determinant of the $k \times k$ submatrix of $\mathbf{A}$ of the elements of $\mathbf{A}$ that are in the rows $\beta_1, \beta_2, \dots, \beta_k$ and in the columns $\alpha_1, \dots, \alpha_k$ of $\mathbf{A}$. Of course, $M_{\overline{0}}^{\overline{0}}(\mathbf{A}) = \det \mathbf{A}$. For convenience, we also set $M_0^0(\mathbf{A}) = 1$.



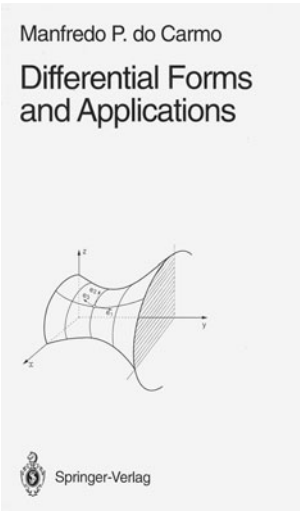


Figure 4.1. Two introductory books on the theory of differential forms.

4.15 Proposition. *We have*

$$v_{\alpha_1} \wedge \cdots \wedge v_{\alpha_k} = \sum_{\beta \in I(k,n)} M_\alpha^\beta(\mathbf{A})\, e_\beta \tag{4.7}$$

for every $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k) \in I(k,n)$ *where*

$$e_\beta := e_{\beta_1} \wedge e_{\beta_2} \wedge \cdots \wedge e_{\beta_k}.$$

Proof. From the linearity,

$$v_{\alpha_1} \wedge \cdots \wedge v_{\alpha_k} = \sum_{j_1,\ldots,j_k=1,\ldots,n} a_{\alpha_1}^{j_1} a_{\alpha_2}^{j_2} \ldots a_{\alpha_k}^{j_k}\, e_{j_1} \wedge \cdots \wedge e_{j_k}. \tag{4.8}$$

Consider the k-tuple of indices $(j_1, \ldots, j_k)$. If two indices agree, then $e_{j_1} \wedge \cdots \wedge e_{j_k} = 0$. Otherwise $j_1, \ldots, j_k$ are distinct. Let $\beta \in I(k,n)$ be the increasing reordering of $(j_1, \ldots, j_k)$ and let σ be the permutation such that $(j_1, \ldots, j_k) = \sigma(\beta)$, then, by (4.5),

$$e_{j_1} \wedge \cdots \wedge e_{j_k} = (-1)^\sigma e_\beta.$$

Thus, on account of (4.6) and (4.8), we conclude that

$$v_{\alpha_1} \wedge \cdots \wedge v_{\alpha_k} = \sum_{\beta \in I(k,n)} \Big(\sum_{\sigma \in \mathcal{P}(S(\beta))} (-1)^\sigma a_{\alpha_1}^{\sigma_1} \ldots a_{\alpha_n}^{\sigma_n} \Big) e_\beta = \sum_{\beta \in I(k,n)} M_\alpha^\beta(\mathbf{A})\, e_\beta.$$

□

e. The exterior algebra and the exterior product

Let V be a vector space of dimension n and let us consider the vector spaces $\Lambda_k V$, $k \geq 0$, defined by

$$\Lambda_0 V := \mathbb{R}, \qquad \Lambda_1 V := V, \qquad \Lambda_k V := \{0\} \text{ for } k > n,$$

where for $2 \leq k \leq n$, $\Lambda_k V$ is the space of k-vectors of V.

We extend the family of exterior products of $2, 3, \ldots$ vectors of V to a family of bilinear maps

$$\Lambda_h V \times \Lambda_k V \to \Lambda_{h+k} V, \qquad (\alpha, \beta) \to \alpha \wedge \beta, \tag{4.9}$$

called *the exterior product of multivectors* defined for $0 \leq h, k \leq n$ as follows. If h or $k = 0$, then $\lambda \wedge \alpha = \alpha \wedge \lambda := \lambda\alpha$; if $h, k \geq 1$, $\cdot \wedge \cdot : \Lambda_h V \times \Lambda_k V \to \mathbb{R}$ is the unique bilinear map characterized by

$$(\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_h) \wedge (\beta_1 \wedge \beta_2 \wedge \cdots \wedge \beta_k) = \alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_h \wedge \beta_1 \wedge \beta_2 \wedge \cdots \wedge \beta_k.$$

It is not difficult to see that the exterior product of multivectors

(i) is bilinear,
(ii) is associative,
(iii) $\alpha \wedge \beta = (-1)^{hk} \beta \wedge \alpha$ if $\alpha \in I(h, n)$, $\beta \in I(k, n)$, that is, is anticommutative if h and k are both odd and commutative otherwise.

The family of vector spaces $\{\Lambda_k V\}_{k \geq 0}$ and the corresponding family of exterior products $\cdot \wedge \cdot : \Lambda_h V \times \Lambda_k V \to \Lambda_{h+k} V$ form the *exterior algebra* of the vector space V.

f. *k*-covectors

Let V be a finite-dimensional vector space and V^* its dual. Recall that, if $(e_1, e_2, \ldots, e_n)$ is a basis of V and $(x^1, x^2, \ldots, x^n)$ are the coordinates of a point $x \in V$ with respect to $(e_1, e_2, \ldots, e_n)$, then the coordinate functions $V \to \mathbb{R}$, $x \mapsto x^1, \ldots, x \mapsto x^n$, denoted also as $dx^1, \ldots, dx^n$, are linear maps that span V^*. Since

$$dx^i(e_j) = \delta^i_j \qquad \forall i, j = 1, \ldots, n, \tag{4.10}$$

$(dx^1, \ldots, dx^n)$ is the basis of V^*, called the *dual basis* of $(e_1, e_2, \ldots, e_n)$. In particular, the map $< , >: V^* \times V \to \mathbb{R}$ given by

$$< \ell, v > := \ell(v) \tag{4.11}$$

is a *duality* between V and V^*, i.e., the following hold:

(i) $< \ell, v >$ is bilinear.
(ii) $< \ell, v >= 0$ $\forall v$ implies $\ell = 0$.
(iii) $< \ell, v >= 0$ $\forall \ell$ implies $v = 0$.

Furthermore, if $x = \sum_{i=1}^n x^i e_i$ and $\ell = \sum_{i=1}^n \ell_i dx^i$, then $< \ell, x >:= \ell(x) = \sum_{i=1}^n \ell_i x^i$.

Finally, let us recall Einstein's index convention, the indices of the components of a vector are up and the indices of the components of a linear map (with respect to the dual basis) are down. It is convenient to arrange the components of a vector as columns and the components of a linear map as a row vector, so that if $\ell = (\ell_1, \ell_2, \dots, \ell_n) \in V^*$ and $x = (x^1, x^2, \dots, x^n)^T \in V$ then $< \ell, x >= \ell x$ where the product ℓx is row by column.

4.16 Definition. *Let V be a finite-dimensional vector space, let V^* be its dual and $2 \le k \le n := \dim V$. A k-vector of V^* is called a k-covector of V. The vector space of k-covectors of V is denoted by*

$$\Lambda^k V := \Lambda_k V^*.$$

We also set $\Lambda^0 V = \mathbb{R}$, $\Lambda^1 V = V^*$ and, for $k > n$, $\Lambda_k V = \{0\}$.

4.17 Duality between $\Lambda_k V$ and $\Lambda^k V$. The map $V^{*k} \times V^k \to \mathbb{R}$ given by

$$((\omega^1, \omega^2, \dots, \omega^k), (v_1, v_2, \dots, v_k)) := \det < \omega^i, v_j >,$$

where $< , >$ is the duality between V and V^*, is $(2k)$-linear and alternating on each factor, therefore, it induces a bilinear map $< , >: \Lambda^k V \times \Lambda_k V \to \mathbb{R}$ characterized by

$$< \omega^1 \wedge \omega^2 \wedge \cdots \wedge \omega^k, v_1 \wedge v_2 \wedge \cdots \wedge v_k >:= \det(< \omega^i, v_j >) \qquad (4.12)$$

if $\omega^1 \wedge \omega^2 \wedge \cdots \wedge \omega^k \in \Lambda^k V$ and $v_1 \wedge v_2 \wedge \cdots \wedge v_k \in \Lambda_k V$.

If $(dx^1, \dots, dx^n)$ is the dual basis in V^* of the basis $(e_1, e_2, \dots, e_n)$ of V, the covectors $dx^\alpha = dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_k}$, $\alpha \in I(k,n)$, form a basis of $\Lambda^k V$; moreover, $\forall \alpha, \beta \in I(k,n)$, we have

$$< dx^\alpha, e_\beta >= \det \Big(< dx^{\alpha_i}, e_{\beta_j} > \Big)_{i,j} = \delta^\alpha_\beta.$$

Consequently, (4.12) *is a duality between $\Lambda_k V$ and $\Lambda^k V$*, and if $\omega = \sum_{\alpha \in I(k,n)} \omega_\alpha dx^\alpha$ and $v = \sum_{\alpha \in I(k,n)} v^\alpha e_\alpha$, we have

$$< \omega, v >= \sum_{\alpha, \beta \in I(k,n)} \omega_\alpha v^\beta < dx^\alpha, e_\beta >= \sum_{\alpha \in I(k,n)} \omega_\alpha v^\alpha.$$

4.18 k-covectors in coordinates. Let $(e_1, \dots, e_n)$ be a basis of a vector space V and $(dx^1, \dots, dx^n)$ be its dual basis in V^*. Let $\omega^1, \dots, \omega^n$ be 1-covectors in V^* and let $\mathbf{A} = (a^i_j)$ be the $n \times n$ matrix with rows the coordinates of $\omega^1, \omega^2, \dots, \omega^n$, in the basis $(dx^1, \dots dx^n)$, i.e., $\omega_i = \sum_{j=1}^n a^j_i dx^j = \sum_{j=1}^n (\mathbf{A}^T)^i_j dx^j$. Then, compare (4.7),

$$\omega^{\alpha_1} \wedge \cdots \wedge \omega^{\alpha_k} = \sum_{\beta \in I(k,n)} M^\alpha_\beta(\mathbf{A}^T)\, dx^\beta \qquad \forall \alpha \in I(k,n). \qquad (4.13)$$

g. Linear transformations

4.19 Exterior power. Let V and W be vector spaces of dimension n and N, respectively, and let k be an integer with $k \geq 1$. For every linear map $\phi : V \to W$, the transformation

$$(v_1, v_2, \ldots, v_k) \to \phi(v_1) \wedge \phi(v_2) \wedge \cdots \wedge \phi(v_k)$$

defines a k-alternating map from V^k into $\Lambda_k W$, consequently, a linear map $\Lambda_k\phi : \Lambda_k V \to \Lambda_k W$ characterized by

$$\Lambda_k\phi(v_1 \wedge v_2 \wedge \cdots \wedge v_k) := \phi(v_1) \wedge \cdots \wedge \phi(v_k) \qquad \forall v_1, v_2, \ldots, v_k \in V. \tag{4.14}$$

Of course, $\Lambda_1\phi = \phi$ and $\Lambda_k\phi = 0$ for $k > n$. If we set $\Lambda_0\phi = \mathrm{Id}$, we can then write

$$\Lambda_{h+k}\phi(\xi \wedge \eta) = \Lambda_h\phi(\xi) \wedge \Lambda_k\phi(\eta) \tag{4.15}$$

$\forall \xi \in \Lambda_h V$, $\forall \eta \in \Lambda_k V$, $\forall h, k \geq 0$.

It is easily seen that for linear maps between finite-dimensional spaces $\phi : V \to W$ and $\psi : W \to Z$ and for $k \geq 0$ we have

$$\Lambda_k(\psi \circ \phi) = \Lambda_k\psi \circ \Lambda_k\phi. \tag{4.16}$$

4.20 ¶. Let $\phi : V \to W$ be a linear map and $\{v_1, v_2, \ldots, v_k\} \subset V$. Then $\Lambda_k\phi(v_1 \wedge v_2 \wedge \cdots \wedge v_k) = 0$ if and only if the vectors $\phi(v_1), \ldots, \phi(v_k)$ are linearly dependent.

4.21 Adjoint map. Recall that for every linear map $\phi : V \to W$ between finite-dimensional spaces the *formal adjoint* is defined as the map $\phi^* : W^* \to V^*$ given by

$$< \phi^*(w^*), v > := < w^*, \phi(v) > \qquad \forall v \in V,\ \forall w^* \in W^*$$

where $< , >$ denotes accordingly the dualities between V and V^* and W and W^*, see (4.11).

If $\mathbf{A}$ is the matrix associated to ϕ with respect to bases $(e_1, e_2, \ldots, e_n)$ in V and $(f_1, f_2, \ldots, f_N)$ in W, then the matrix $\mathbf{B}$ associated to ϕ^* in the corresponding dual bases $(dx^1, \ldots, dx^n)$ of V^* and $(dy^1, \ldots, dy^N)$ of W^* is $\mathbf{A}^T$. In fact, the elements of the jth column of $\mathbf{B}$ are the coordinates of $\phi^*(dx^j)$, therefore

$$\mathbf{B}^i_j = < \phi^*(dx^j), e_i > = < dx^j, \phi(e_i) > = \mathbf{A}^j_i.$$

4.22 Exterior power of the adjoint. If $\phi : V \to W$ is linear and $k \geq 1$, the map

$$(\omega^1, \omega^2, \ldots, \omega^k) \to \phi^*(\omega^1) \wedge \cdots \wedge \phi^*(\omega^k)$$

is k-alternating on W^{*k} with values in $\Lambda^k V$, consequently, it defines a unique linear map $\Lambda^k\phi : \Lambda^k W \to \Lambda^k V$ characterized by

$$\Lambda^k\phi(\omega^1 \wedge \omega^2 \wedge \cdots \wedge \omega^k) = \phi^*(\omega^1) \wedge \cdots \wedge \phi^*(\omega^k).$$

If we set $\Lambda^0\phi = \mathrm{Id}$ in $\mathbb{R}$, for all $h, k \geq 0$ we have

$$\Lambda^{h+k}\phi(\omega\wedge\zeta)=\Lambda^h\phi(\omega)\wedge\Lambda^k\phi(\zeta)\qquad\forall\omega\in\Lambda^hV,\ \forall\zeta\in\Lambda^kV \tag{4.17}$$

and for linear maps $\phi:\mathbb{R}^n\to\mathbb{R}^N$ and $\psi:\mathbb{R}^N\to\mathbb{R}^M$,

$$\Lambda^k(\psi\circ\phi)=\Lambda^k\phi\circ\Lambda^k\psi. \tag{4.18}$$

Moreover, the map $\Lambda^k\phi:\Lambda^kW\to\Lambda^kV$ is the formal adjoint of $\Lambda_k\phi:\Lambda_kV\to\Lambda_kW$ with respect to the induced dualities between Λ_kV and Λ^kV and between Λ_kW and Λ^kW given by (4.12). In fact,

$$\begin{aligned}
&<\Lambda^k\phi(\omega^1\wedge\omega^2\wedge\cdots\wedge\omega^k),k_1\wedge k_2\wedge\cdots\wedge k_>\\
&\qquad=<\phi^*(\omega^1)\wedge\cdots\wedge\phi^*(\omega^k),v_1\wedge v_2\wedge\cdots\wedge v_k>\\
&\qquad=\det<\phi^*(\omega^i),v_j>=\det<\omega^i,\phi(v_j)>\\
&\qquad=<\omega^1\wedge\omega^2\wedge\cdots\wedge\omega^k,\phi(v_1)\wedge\cdots\wedge f(v_k)>\\
&\qquad=<\omega^1\wedge\omega^2\wedge\cdots\wedge\omega^k,\Lambda_k\phi(v_1\wedge v_2\wedge\cdots\wedge v_k)>.
\end{aligned}$$

4.23 Exterior power in coordinates. Let V and W be two finite-dimensional vector spaces, $(e_1,\,e_2,\ldots,\,e_n)$ a basis in V and $(f_1,\,f_2,\ldots,\,f_N)$ a basis in W. Let $(dx^1,\ldots,dx^n)$ and $(dy^1,\ldots,dy^N)$ be the corresponding dual bases in V^* and W^*. If $\mathbf{A}\in M_{N,n}$ is the matrix associated to the linear map $\phi:V\to W$, then

$$\begin{aligned}
\phi(e_i)&:=\sum_{j=1}^N\mathbf{A}_i^jf_j,\\
\phi^*(dy^i)&=\sum_{j=1}^n\mathbf{A}_j^idx^j=\sum_{j=1}^n(\mathbf{A}^T)_i^jdx^j.
\end{aligned}$$

It follows from (4.7) that

$$\begin{aligned}
\Lambda_kf(e_\alpha)&=\sum_{\beta\in I(k,N)}M_\alpha^\beta(\mathbf{A})f_\beta.\\
\Lambda^kf(dy^\alpha)&=\sum_{\beta\in I(k,n)}M_\alpha^\beta(\mathbf{A}^T)\,dx^\beta.
\end{aligned} \tag{4.19}$$

4.24 Example. Let $\phi:\mathbb{R}^2\to\mathbb{R}^3$ be a linear map and let

$$\mathbf{A}:=\begin{pmatrix}a&d\\b&e\\c&f\end{pmatrix}$$

be the matrix associated to ϕ in the bases (e_1,e_2) of $\mathbb{R}^2$ and (f_1,f_2,f_3) of $\mathbb{R}^3$. Then

$$\begin{aligned}
&\Lambda_1\phi(e_1):=a\,f_1+b\,f_2+c\,f_3,\qquad\Lambda_1\phi(e_2):=d\,f_1+e\,f_2+f\,f_3,\\
&\Lambda_2\phi(e_1\wedge e_2)=(ae-bd)\,f_1\wedge f_2+(af-dc)\,f_1\wedge f_3+(bf-ce)\,f_2\wedge f_3.
\end{aligned}$$

h. The determinant

The formalism we have introduced is useful for dealing with the properties of the determinant avoiding indices and permutations. Let us begin with an *intrinsic* definition of *determinant of a linear map* $\phi : V \to V$.

If $\dim V = n$, we have $\dim \Lambda_n V = 1$, hence there is a unique number, called the *determinant of* ϕ and denoted by $\det \phi$ such that

$$\Lambda_n \phi(v_1 \wedge v_2 \wedge \cdots \wedge v_n) = (\det \phi)\, v_1 \wedge v_2 \wedge \cdots \wedge v_n \tag{4.20}$$

for every $v_1, v_2, \ldots, v_n \in V$. Clearly, $\det \phi$ does not depend either on the coordinates or the choice of the exterior product: It is a characteristic of the linear map ϕ. Moreover, according to (4.19), if $\phi : \mathbb{R}^n \to \mathbb{R}^n$, $\phi(x) := \mathbf{A}x$, then

$$\Lambda_n \phi(v_1 \wedge v_2 \wedge \cdots \wedge v_n) = (\det \mathbf{A})\, v_1 \wedge v_2 \wedge \cdots \wedge v_n, \tag{4.21}$$

i.e., $\det \phi = \det \mathbf{A}$, where $\mathbf{A}$ is the matrix associated to ϕ once the basis of V is chosen.

4.25 Binet's formula. Let $\mathbf{A}$ and $\mathbf{B}$ be two $n \times n$ matrices and let $\phi(x) = \mathbf{A}x$ and $\psi(y) := \mathbf{B}y$. Then $\psi \circ \phi(x) = \mathbf{B}\mathbf{A}x$ and, therefore, the equality

$$\Lambda_n(\psi \circ \phi) = \Lambda_n \psi \circ \Lambda_n \phi$$

in (4.18) rewrites in terms of matrices, according to (4.21), as *Binet's formula*

$$\det(\mathbf{A}\mathbf{B}) = \det \mathbf{A}\, \det \mathbf{B}.$$

The associative property of the exterior product

$$\Lambda_n \phi(v_1 \wedge v_2 \wedge \cdots \wedge v_n) = \Lambda_k(v_1, v_2, \ldots, v_k) \wedge \Lambda_{n-k}(v_{k+1} \wedge \cdots \wedge v_n) \tag{4.22}$$

$\forall v_1, v_2, \ldots, v_n \in V$, yields Laplace's formula for the determinant of a matrix:

4.26 Theorem (Laplace's formula). *Let* $\mathbf{A} \in M_{n,n}(\mathbb{R})$. *For* $\beta, \gamma \in I(k,n)$ *we have*

$$\delta_{\beta\gamma} \det \mathbf{A} = \sigma(\beta, \overline{\beta}) \sum_{\alpha \in I(k,n)} \sigma(\alpha, \overline{\alpha}) M_\beta^\alpha(\mathbf{A})\, M_{\overline{\gamma}}^{\overline{\alpha}}(\mathbf{A}). \tag{4.23}$$

Proof. According to (4.19),

$$\Lambda_k \phi(e_\beta) = \sum_{\alpha \in I(k,n)} M_\beta^\alpha(\mathbf{A})\, e_\alpha,$$

$$\Lambda_{n-k} \phi(e_{\overline{\gamma}}) = \sum_{\tau \in I(n-k,n)} M_{\overline{\gamma}}^\tau(\mathbf{A}) e_\tau,$$

hence

$$\begin{aligned}\Lambda_n\phi(e_\beta \wedge e_{\overline{\gamma}}) &= \Lambda_k\phi(e_\beta) \wedge \Lambda_{n-k}\phi(e_{\overline{\gamma}}) \\ &= \sum_{\alpha\in I(k,n)} \sum_{\tau\in I(n-k,n)} M^\alpha_\beta(\mathbf{A}) M^\tau_{\overline{\gamma}}(\mathbf{A})\, e_\alpha \wedge e_\tau \\ &= \sum_{\alpha\in I(k,n)} M^\alpha_\beta(\mathbf{A}) M^{\overline{\alpha}}_{\overline{\gamma}}(\mathbf{A})\, e_\alpha \wedge e_{\overline{\alpha}} \\ &= \Big(\sum_{\alpha\in I(k,n)} \sigma(\alpha,\overline{\alpha}) M^\alpha_\beta(\mathbf{A}) M^{\overline{\alpha}}_{\overline{\gamma}}(\mathbf{A}) \Big)\, e_1 \wedge e_2 \wedge \cdots \wedge e_n.\end{aligned}$$

On the other hand $e_\beta \wedge e_{\overline{\gamma}} = 0$ if $b \neq \gamma$, and

$$\Lambda_n(e_\beta \wedge e_{\overline{\beta}}) = \sigma(\beta,\overline{\beta})\Lambda_n(w_1 \wedge w_2 \wedge \cdots \wedge w_n) = \det \mathbf{A}\, e_1 \wedge e_2 \wedge \cdots \wedge e_n.$$

□

4.27 Remark. When $|\beta| = 1$, say $\beta = j \in \{1, \ldots, n\}$, Laplace's formula yields the development of the determinant in terms of the *cofactors* along the column j,

$$\delta^i_h \det \mathbf{A} = \sum_{\alpha\in I(k,n)} (-1)^{i+j} \mathbf{A}^i_j M^{\overline{i}}_{\overline{j}}(\mathbf{A}) = (\mathbf{A}\,\mathrm{cof}\,\mathbf{A})^i_h,$$

if $(\mathrm{cof}\,\mathbf{A})^j_i := (-1)^{i+j} M^{\overline{i}}_{\overline{j}}(\mathbf{A})$.

4.28 Remark ($\det \mathbf{A} = \det \mathbf{A}^T$). Finally, from (4.21) we infer that the equality $\det \mathbf{A} = \det \mathbf{A}^T$ is equivalent to the fact that $\Lambda^n\phi$ *is the formal adjoint of* $\Lambda_n\phi$ for the map $\phi(x) := \mathbf{A}x$. In fact, from (4.19)

$$\begin{aligned}\Lambda_n\phi(e_1 \wedge e_2 \wedge \cdots \wedge e_n) &= \phi(e_1) \wedge \cdots \wedge \phi(e_n) \\ &= (\det \mathbf{A})\, e_1 \wedge e_2 \wedge \cdots \wedge e_n, \\ \Lambda^n\phi(dx^1 \wedge \cdots \wedge dx^n) &= \phi^*(dx^1) \wedge \cdots \wedge \phi^*(dx^n) \\ &= (\det \mathbf{A}^T)\, dx^1 \wedge \cdots \wedge dx^n.\end{aligned}$$

i. Inner product of multivectors

Let V be a finite-dimensional space with an inner product $(\ |\ )$. Then the Riesz isomorphism $R : V \to V^*$ given by $R(v)(w) := (v|w)$ is well-defined, and we may set

$$\begin{aligned}\xi \bullet \eta &:= < \Lambda_k R(\xi), \eta > \qquad \forall \xi, \eta \in \Lambda_k V, \\ \omega \bullet \zeta &:= < \omega, \Lambda^k R(\zeta) > \qquad \forall \omega, \zeta \in \Lambda^k V.\end{aligned} \tag{4.24}$$

From (4.12) and (4.14), recalling also that the $k \times k$ matrix $\mathbf{G} = (g_{ij})$, $g_{ij} := v_i \bullet v_j$, is positive definite if and only if $(v_1, v_2, \ldots, v_k)$ are linearly independent, we then infer the following proposition.

4.29 Proposition. *The bilinear maps in* (4.24) *define two inner products respectively in* $\Lambda_k V$ *and* $\Lambda^k V$*, and we have*

$$\begin{aligned}(v_1 \wedge v_2 \wedge \cdots \wedge v_k) \bullet (w_1 \wedge w_2 \wedge \cdots \wedge w_k) &= \det(\, v_i \bullet w_j \,),\\ (\omega^1 \wedge \omega^2 \wedge \cdots \wedge \omega^k) \bullet (\eta^1 \wedge \eta^2 \wedge \cdots \wedge \eta^k) &= \det(\, \omega^i \bullet \eta^j \,).\end{aligned} \tag{4.25}$$

In particular, if $(e_1, e_2, \ldots, e_n)$ is an orthonormal basis of V, then the dual basis $(dx^1, \ldots, dx^n)$ is orthonormal in V^* and the bases $(e_\alpha)_{\alpha \in I(k,n)}$ in $\Lambda_k V$ and $(dx^\alpha)_{\alpha \in I(k,n)}$ in $\Lambda^k V$ are orthonormal. It follows that

$$\xi = \sum_{\alpha \in I(k,n)} (\, \xi \bullet e_\alpha \,)\, e_\alpha, \qquad \omega = \sum_{\alpha \in I(k,n)} (\, \omega \bullet dx^\alpha \,)\, dx^\alpha$$

for all $\xi \in \Lambda_k V$ and $\eta \in \Lambda^k V$, and

$$\begin{aligned}|\xi|^2 := \xi \bullet \xi &= \sum_{\alpha \in I(k,n)} |\, \xi \bullet e_\alpha \,|^2,\\ |\omega|^2 := \omega \bullet \omega &= \sum_{\alpha \in I(k,n)} |\, \omega \bullet dx^a \,|^2.\end{aligned}$$

j. The Jacobian and the Cauchy–Binet formula

Let us give a simple and interesting geometric interpretation of (4.25). Let us consider $\mathbb{R}^n$ and $\mathbb{R}^N$, $N \geq n$, with the standard inner products and orthonormal bases $(e_1, e_2, \ldots, e_n)$ and $(f_1, f_2, \ldots, f_N)$, respectively. Let $\mathbf{T}$ be a $N \times n$ matrix and denote by $v_j \in \mathbb{R}^N$ the jth column of $\mathbf{T}$, $1 \leq j \leq n$. Since

$$v_i \bullet v_j = (\mathbf{T}^T \mathbf{T})_{ij} \qquad \forall i, j = 1, \ldots, n,$$

(4.25) yields

$$\begin{aligned}|\Lambda_n \mathbf{T}|^2 &= \Big| v_1 \wedge v_2 \wedge \cdots \wedge v_n \Big|^2\\ &= (v_1 \wedge v_2 \wedge \cdots \wedge v_n) \bullet (v_1 \wedge v_2 \wedge \cdots \wedge v_n) = \det(\mathbf{T}^T \mathbf{T}).\end{aligned} \tag{4.26}$$

In particular, $\ker \mathbf{T} = \{0\}$ if and only if $\Lambda_n \mathbf{T} = 0$. On the other hand, (4.19) yields

$$\begin{aligned}v_1 \wedge v_2 \wedge \cdots \wedge v_k &= \mathbf{T}(e_1) \wedge \cdots \wedge \mathbf{T}(e_k)\\ &= \Lambda_k \mathbf{T}(e_1 \wedge e_2 \wedge \cdots \wedge e_k) = \sum_{\alpha \in I(k,N)} M^\alpha_{(1,\ldots,k)}(\mathbf{T}) f_\beta\end{aligned}$$

and, since $f_1, f_2, \ldots, f_N$ are orthonormal,

$$\Big| v_1 \wedge v_2 \wedge \cdots \wedge v_k \Big|^2 = \sum_{\alpha \in I(k,N)} \Big| M^\alpha_{(1,\ldots,k)}(\mathbf{T}) \Big|^2. \tag{4.27}$$

By a comparison of (4.26) and (4.27) with $k = n$ we then infer the *Cauchy–Binet formula*

$$\det(\mathbf{T}^T\mathbf{T}) = \sum_{\alpha\in I(n,N)} \left|M^{\alpha}_{(1,\dots,n)}(\mathbf{T})\right|^2. \tag{4.28}$$

4.30 Example. Let $\mathbf{A} \in M_{3,2}(\mathbb{R})$,

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix},$$

and let

$$u := \begin{pmatrix} a \\ c \\ e \end{pmatrix} \qquad v := \begin{pmatrix} b \\ d \\ f \end{pmatrix}$$

be the two columns of $\mathbf{A}$. Set

$$\begin{cases} E := |u|^2 = a^2 + b^2 + c^2, \\ F := (u|v) = ab + cd + ef, \\ G := |v|^2 = b^2 + d^2 + f^2. \end{cases}$$

Then (4.26) yields

$$\det(\mathbf{A}^T\mathbf{A}) = \det \begin{pmatrix} E & F \\ F & G \end{pmatrix} = EG - F^2.$$

On the other hand, $I(2,3) = \{(1,2),(2,3),(1,3)\}$, thus, if we set

$$\mathbf{B} := \mathbf{A}^{(1,2)}_{(1,2)} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \mathbf{C} := \mathbf{A}^{(2,3)}_{(1,2)} = \begin{pmatrix} c & d \\ e & f \end{pmatrix}, \qquad \mathbf{D} := \mathbf{A}^{(1,3)}_{(1,2)} = \begin{pmatrix} a & b \\ e & f \end{pmatrix},$$

the Cauchy–Binet formula states that

$$\det(\mathbf{A}^T\mathbf{A}) = (\det \mathbf{B})^2 + (\det \mathbf{C})^2 + (\det \mathbf{D})^2.$$

4.1.2 Subspaces and k-vectors

a. Simple vectors

Let V be a vector space of dimension n and $k \le n$. A k-vector $\xi \in \Lambda_k V$ is said to be *simple* if there exist $v_1, v_2, \dots, v_k \in V$ such that $\xi = v_1 \wedge v_2 \wedge \cdots \wedge v_k$. Since $\lambda\xi$, $\lambda \in \mathbb{R}$, is simple if ξ is simple, the simple vectors form a cone of $\Lambda_n V$.

Clearly, 0-vectors, 1-vectors and n-vectors in $\Lambda_n V$ are simple. Moreover, the following holds, see (iii) of Proposition 4.34,

4.31 Proposition. *If* $\dim V = n$, *then all* $(n-1)$*-vectors are simple.*

The following example shows that not all k-vectors are simple.

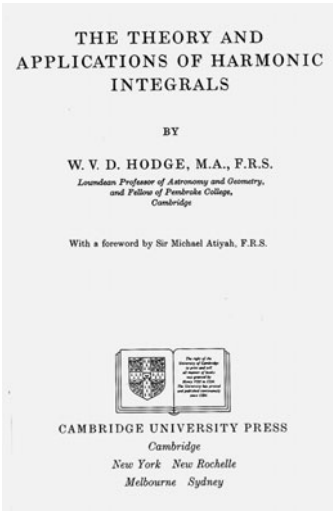


Figure 4.2. Frontispieces of two celebrated monographs.

4.32 Example. Let (e_1, e_2, e_3, e_4) be the standard basis of $\mathbb{R}^4$. The 2-vector $\xi := e_1 \wedge e_2 + e_3 \wedge e_4 \in \Lambda_2 \mathbb{R}^4$ is not simple. Otherwise $\xi = v_1 \wedge v_2$ for some $v_1, v_2 \in V$. Since $\xi \wedge \xi = 0$, we would then have

$$2\, e_1 \wedge e_2 \wedge e_3 \wedge e_4 = 0,$$

a contradiction.

Actually, one can show that there exist nonsimple k-vectors in $\Lambda_k V$ if and only if $\dim V \geq 4$ and $2 \leq k \leq n-2$.

b. Simple vectors and k-subspaces

In addition to being useful in describing the properties of the determinant, the k-vectors of a vector space V are useful in dealing with subspaces of dimension k of V.

Let V be a vector space of dimension n and let $v_1, v_2, \ldots, v_k \in V$, $k \leq n$. As we noticed in Proposition 4.14, $v_1 \wedge v_2 \wedge \cdots \wedge v_k$ is nonzero if and only if $v_1, v_2, \ldots, v_k$ are linearly independent. Consequently, to every simple and nonzero k-vector $\xi = v_1 \wedge v_2 \wedge \cdots \wedge v_k$, we may and do associate the k-dimensional subspace of V

$$\operatorname{Span}(\xi) := \operatorname{Span}\Big\{v_1, v_2, \ldots, v_k\Big\}$$

and we have

$$\operatorname{Span}(\xi) = \Big\{v \in V \;\Big|\; \xi \wedge v = 0\Big\}. \tag{4.29}$$

Furthermore, if $(v_1, v_2, \ldots, v_k)$ and $(w_1, w_2, \ldots, w_k)$ are two k-tuples of linearly independent vectors of V, we have $\operatorname{Span}\{v_1, v_2, \ldots, v_k\} = \operatorname{Span}\{w_1, w_2, \ldots, w_k\}$ if and only if

$$v_1 \wedge v_2 \wedge \cdots \wedge v_k = \lambda\, w_1 \wedge w_2 \wedge \cdots \wedge w_k \qquad \text{for some } \lambda \neq 0. \tag{4.30}$$

In fact, if $V' := \operatorname{Span}\{v_1, v_2, \ldots, v_k\} = \operatorname{Span}\{w_1, w_2, \ldots, w_k\}$, then $(v_1, v_2, \ldots, v_k)$ and $(w_1, w_2, \ldots, w_k)$ are two bases for V', and, because of (4.20), we have

$$v_1 \wedge v_2 \wedge \cdots \wedge v_k = \lambda\, w_1 \wedge w_2 \wedge \cdots \wedge w_k,$$

where $\lambda \neq 0$ is the determinant of the matrix that changes the basis from $(w_1, w_2, \ldots, w_k)$ to $(v_1, v_2, \ldots, v_k)$. Conversely, if $v_1 \wedge v_2 \wedge \cdots \wedge v_k = \lambda\, w_1 \wedge w_2 \wedge \cdots \wedge w_k$ for some $\lambda \neq 0$, then $\operatorname{Span}\{v_1, v_2, \ldots, v_k\} = \operatorname{Span}\{w_1, w_2, \ldots, w_k\}$ according to (4.29).

Summing up, *the linear subspaces of dimension k of V are in a one-to-one correspondence with the lines through the origin in $\Lambda_k\mathbb{R}^n$ generated by the nonzero simple vectors $\xi = v_1 \wedge v_2 \wedge \cdots \wedge v_k$.*

c. Orientation and simple k-vectors

Let $\phi : V \to V$ be a linear isomorphism. We say that ϕ *preserves the orientation* if $\det\phi > 0$ and *reverses the orientation* if $\det\phi < 0$. Recalling that a basis is an ordered set of vectors, we say that two bases $(v_1, v_2, \ldots, v_k)$ and $(w_1, w_2, \ldots, w_k)$ of V have the same *orientation* (respectively the opposite orientation) if the map that changes coordinates from one basis to the other preserves (respectively reverses) the orientation. Accordingly, the choice of an ordered basis $(v_1, v_2, \ldots, v_k)$ of V fixes the orientation, and any other basis $(w_1, w_2, \ldots, w_k)$ has either the orientation of $(v_1, v_2, \ldots, v_k)$ or the opposite orientation.

Since according to (4.30),

$$w_1 \wedge w_2 \wedge \cdots \wedge w_k = \lambda\, v_1 \wedge v_2 \wedge \cdots \wedge v_k,$$

where λ is the determinant of the map that maps the basis $(v_1, v_2, \ldots, v_k)$ on $(w_1, w_2, \ldots, w_k)$, we may identify an *oriented k-plane*, i.e., a k-plane with a selected ordered basis $(v_1, v_2, \ldots, v_k)$ on it, with one of the two half-lines in $\Lambda_k V$ through 0 generated by the simple vector $w_1 \wedge w_2 \wedge \cdots \wedge w_k$.

If V is a Euclidean space, we have a norm on $\Lambda_k V$, hence the two multiples of $\xi = w_1 \wedge w_2 \wedge \cdots \wedge w_k$ of unit norm 1 describe the two possible orientations of $\operatorname{Span}(v_1, v_2, \ldots, v_k)$. This motivates the following definition.

4.33 Definition. *Let V be a vector space of dimension n endowed with an inner product and let P be a k-subspace of V and $\xi \neq 0$ a simple k-vector that spans P,*

$$P = \Big\{ v \in V \;\Big|\; v \wedge \xi = 0 \Big\}.$$

Then the two k-vectors $\pm\xi/|\xi|$ are the orientations *of P. If $(v_1, v_2, \ldots, v_k)$ is a basis of P, then the orientation fixed by $(v_1, v_2, \ldots, v_k)$ is the k-vector*

$$\frac{v_1 \wedge v_2 \wedge \cdots \wedge v_k}{|v_1 \wedge v_2 \wedge \cdots \wedge v_k|}.$$

We conclude noticing that the set of oriented, k-dimensional subspaces of V is in a one-to-one correspondence with the k-dimensional *Grassmannian* of V defined by

$$G_k(V) := \Big\{ \xi \in \Lambda_k V \,\Big|\, \xi \text{ simple},\ |\xi| = 1 \Big\}.$$

According to the above, $G_k(V)$ is a proper subset of (the unit sphere of) $\Lambda_k V$ if $\dim V = n \geq 4$ and $2 \leq k \leq n-2$.

4.1.3 Vector product and Hodge operator

a. Hodge operator

Let V be a vector space of dimension n, endowed with an inner product and oriented by the choice of an n-vector $\mu \in \Lambda_n V$ with $|\mu| = 1$, and let $0 \leq k \leq n$. Since $\dim \Lambda_n V = 1$, the map $t \to t\,\mu$ is an isomorphism between $\mathbb{R}$ and $\Lambda_n V$. The bilinear map $(\xi, \eta) \to \xi \wedge \eta$ from $\Lambda_k V \times \Lambda_{n-k} V$ into $\Lambda_n V$ then yields a bilinear map $a : \Lambda_k V \times \Lambda_{n-k} V \to \mathbb{R}$ defined by

$$\xi \wedge \eta =: a(\xi, \eta)\, \mu.$$

By Riesz's theorem, there is a linear map $* : \Lambda_k V \to \Lambda_{n-k} V$, called the *Hodge operator*, such that $a(\xi, \eta) = (*\xi) \bullet \eta$, i.e.,

$$\xi \wedge \eta =: \big((*\xi) \bullet \eta \big)\, \mu \qquad \forall \xi \in \Lambda_k V,\ \forall \eta \in \Lambda_{n-k} V. \tag{4.31}$$

Of course, Hodge's operator depends on the choices of the inner product and of the orientation μ of V ($*$ changes sign if we change the orientation of V).

By replacing V with V^* and μ with the dual n-form $\Omega \in \Lambda^n V$, we may define the Hodge operator on covectors $* : \Lambda^k V \to \Lambda^{n-k} V$. By definition

$$\begin{aligned} \xi \wedge \eta &=: \big(*\xi \bullet \eta \big)\, \mu \qquad && \forall \xi \in \Lambda_k V,\ \eta \in \Lambda_{n-k} V, \\ \omega \wedge \zeta &=: \big(*\omega \bullet \zeta \big)\, \Omega \qquad && \forall \omega \in \Lambda^k V,\ \eta \in \Lambda^{n-k} V. \end{aligned} \tag{4.32}$$

Before computing a few formulas, it is convenient to make the action of Hodge's operator on an orthonormal basis $(e_1, e_2, \ldots, e_n)$ of V explicit. For all $\alpha \in I(k, n)$ and $\beta \in I(n-k, n)$ we have

$$\big(*e_\alpha \bullet e_\beta \big)\, \mu = e_\alpha \wedge e_\beta = \begin{cases} 0 & \text{if } \beta \neq \overline{\alpha}, \\ \sigma(\alpha, \overline{\alpha})\, e_1 \wedge e_2 \wedge \cdots \wedge e_n & \text{if } \beta = \overline{\alpha}, \end{cases}$$

$$\big(*dx^\alpha \bullet dx^\beta \big)\, \Omega = dx^\alpha \wedge dx^\beta = \begin{cases} 0 & \text{if } \beta \neq \overline{\alpha}, \\ \sigma(\alpha, \overline{\alpha})\, dx^1 \wedge \cdots \wedge dx^n & \text{if } \beta = \overline{\alpha} \end{cases}$$

hence

$$\begin{aligned} *\, e_\alpha &= \sigma(\alpha,\overline{\alpha})\, e_{\overline{\alpha}} < e_1 \wedge \cdots \wedge e_n, \mu >, \\ *\, dx^\alpha &= \sigma(\alpha,\overline{\alpha})\, dx^{\overline{\alpha}} < dx^1 \wedge \cdots \wedge dx^n, \Omega > . \end{aligned} \tag{4.33}$$

In particular, if we choose as orientation the one induced by the orthonormal basis $(e_1, e_2, \ldots, e_n)$, i.e., $\mu := e_1 \wedge e_2 \wedge \cdots \wedge e_n$, then $\Omega = dx^1 \wedge \cdots \wedge dx^n$ and

$$*e_\alpha = \sigma(\alpha,\overline{\alpha})\, e_{\overline{\alpha}}, \qquad *dx^\alpha = \sigma(\alpha,\overline{\alpha})\, dx^{\overline{\alpha}}. \tag{4.34}$$

4.34 Proposition. *We have the following:*

(i) $*1 = \mu$, $*\mu = 1$.

(ii) *The Hodge operators* $* : \Lambda_k V \to \Lambda_{n-k} V$ *and* $* : \Lambda^k V \to \Lambda^{n-k} V$, $0 \le k \le n$, *are isomorphisms*

$$\begin{aligned} *\,(*\xi) &= (-1)^{k(n-k)} \xi \qquad \forall \xi \in \Lambda_k V, \\ *\,(*\omega) &= (-1)^{k(n-k)} \omega \qquad \forall \omega \in \Lambda^k V. \end{aligned} \tag{4.35}$$

(iii) $\xi \in \Lambda_k V$ *is simple if and only if* $*\xi \in \Lambda_{n-k} V$ *is simple.*

(iv) *If* $\xi \in \Lambda_k V$ *is simple and nonzero, then* $\mathrm{Span}(\xi)$ *and* $\mathrm{Span}(*\xi)$ *are perpendicular, actually supplementary.*

Proof. (i) is trivial.

(ii) If $(e_1, e_2, \ldots, e_n)$ is an orthonormal basis of V, we have

$$\begin{aligned} *(*e_\alpha) &= \sigma(\alpha,\overline{\alpha}) * e_{\overline{\alpha}} < e_1 \wedge e_2 \wedge \cdots \wedge e_n, \mu > \\ &= \sigma(\alpha,\overline{\alpha})\, \sigma(\overline{\alpha},\alpha)\, e_\alpha < e_1 \wedge e_2 \wedge \cdots \wedge e_n, \mu >^2 \\ &= \sigma(\alpha,\overline{\alpha}) \sigma(\overline{\alpha},\alpha)\, e_\alpha. \end{aligned}$$

We need $k(n-k)$ transpositions to permute $(\alpha,\overline{\alpha})$ into $(\overline{\alpha},\alpha)$, hence $\sigma(\overline{\alpha},\alpha) = \sigma(\alpha,\overline{\alpha})(-1)^{k(n-k)}$. Similarly we show that $**\omega = (-1)^{k(n-k)}\omega$.

(iii) and (iv). Let $\xi \in \Lambda_k V$ be simple and nonzero. By choosing an orthonormal basis $(e_1, e_2, \ldots, e_n)$ of V such that $\xi = \lambda\, e_1 \wedge e_2 \wedge \cdots \wedge e_k$ for some $\lambda \neq 0$, we have

$$*\xi = \lambda * (e_1 \wedge e_2 \wedge \cdots \wedge e_k) = e_{k+1} \wedge \cdots \wedge e_n < e_1 \wedge e_2 \wedge \cdots \wedge e_n, \mu >$$

and this proves (iii) and (iv). □

From (ii) of Proposition 4.34 we infer for all $\xi, \eta \in \Lambda_k V$, $\forall \omega, \zeta \in \Lambda^k V$

$$\begin{aligned} &\eta \wedge *\xi = (-1)^{k(n-k)} * \xi \wedge \eta = (\, \xi \bullet \eta\,)\, \mu, \\ &* \omega \wedge \zeta = (-1)^{k(n-k)} \zeta \wedge *\omega = (\, \omega \bullet \zeta\,)\, \Omega, \\ &\xi \bullet \eta \;=\; (*\xi) \bullet (*\eta)\,, \\ &\omega \bullet \zeta \;=\; (*\omega) \bullet (*\zeta)\,, \\ &< \omega, *\xi > = (-1)^{k(n-k)} < *\omega, \xi > \end{aligned} \tag{4.36}$$

and

$$\begin{aligned} &\xi \wedge *\xi = |\xi|^2\, \mu = |*\xi|^2 \mu, && \omega \wedge *\omega = |\omega|^2\, \Omega = |*\omega|^2\, \Omega, \\ &|\xi| = |*\xi|, && |\omega| = |*\omega|, \\ &|\xi \wedge *\xi| = |\xi|^2, && |\omega \wedge *\omega| = |\omega|^2. \end{aligned}$$

4.35 Example. Consider $\mathbb{R}^3$ with the standard orientation induced by (e_x, e_y, e_z). Then $** = (-1)^{k(3-k)} = +1\ \forall k = 0, 1, 2, 3$ and

$$\begin{aligned} &* e_x = e_y \wedge ez, && *e_y = -e_x \wedge e_z, && * e_z = e_x \wedge e_y, \\ &* dx = dy \wedge dz, && *dy = -dx \wedge dz, && * dz = dx \wedge dy. \end{aligned}$$

b. Vector product

In terms of Hodge's operator we may define the *vector product* in $\mathbb{R}^3$.

4.36 Definition. *Let V be a vector space of dimension n endowed with an inner product and oriented by $\mu \in \Lambda_n V$ with $|\mu| = 1$, and let $*$ be the associated Hodge operator on $\Lambda_{n-1}V$. The* vector product *of $(n-1)$ vectors $v_1, v_2, \ldots, v_{n-1} \in V$ denoted by $v_1 \times \cdots \times v_{n-1}$ is defined to be the* 1*-vector*

$$v_1 \times \cdots \times v_{n-1} := *(v_1 \wedge v_2 \wedge \cdots \wedge v_{n-1}).$$

We readily infer the following from (4.36):

(i) $|v_1 \times \cdots \times v_{n-1}| = |v_1 \wedge v_2 \wedge \cdots \wedge v_{n-1}|$.
(ii) $v_1 \times \cdots \times v_{n-1}$ is nonzero if and only if $v_1, v_2, \ldots, v_{n-1}$ are linearly independent and in this case $v_1 \times \cdots \times v_{n-1}$ is perpendicular to $\mathrm{Span}\{v_1, v_2, \ldots, v_{n-1}\}$.
(iii) We have

$$v_1 \wedge v_2 \wedge \cdots \wedge v_{n-1} \wedge (v_1 \times \cdots \times v_{n-1}) = |v_1 \wedge v_2 \wedge \cdots \wedge v_{n-1}|^2 \mu.$$

In particular, if $v_1, v_2, \ldots, v_{n-1}$ are linearly independent, the basis

$$(v_1, \ldots, v_{n-1}, v_1 \times \cdots \times v_{n-1})$$

has the same orientation of $(e_1, e_2, \ldots, e_n)$.

4.37 Example. Let $V = \mathbb{R}^3$, (e_1, e_2, e_3) be the standard basis of $\mathbb{R}^3$, and $\mu = e_1 \wedge e_2 \wedge e_3$. Then

$$\begin{aligned} e_1 \times e_2 &:= *(e_1 \wedge e_2) = e_3, \\ e_1 \times e_3 &:= *(e_1 \wedge e_3) = -e_2, \\ e_2 \times e_3 &:= *(e_2 \wedge e_3) = e_1. \end{aligned}$$

4.38 ¶. Let $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^{n-1}$ be a linear map and $\mathbf{L}^T = \Big[L_1 | L_2 | \ldots | L_n\Big]$. Show that $\mathrm{Rank}\, \mathbf{L} = n-1$ if and only if $L_1 \times L_2 \times \cdots \times L_n \neq 0$ and, in this case, $L_1 \times L_2 \times \cdots \times L_n$ spans $\ker \mathbf{L}$.

4.2 Integration of Differential k-Forms

4.2.1 Differential k-forms

Let $\Omega \subset \mathbb{R}^n$. A *differential k-form* (for short, a *k-form*) ω in $\Omega \subset \mathbb{R}^n$ is a map $\omega : \Omega \to \Lambda^k \mathbb{R}^n$. Thus, if $k = 0$, ω is a function $\omega : \Omega \to \mathbb{R}$; if $k > n$ trivially $\omega(x) = 0\ \forall x$, and for $1 \le k \le n$, with respect to the basis of $\Lambda^k \mathbb{R}^n$ we can write

$$\omega(x) = \sum_{\alpha \in I(k,n)} \omega_\alpha(x)\, dx^\alpha, \qquad \forall x \in \Omega,$$

$(\omega_\alpha)_\alpha$ being the components of ω with respect to the chosen basis $(dx^\alpha)_\alpha$. We say that ω is of class C^s if its components are functions of class C^s.

a. Exterior differential

The special structure of $\Lambda_k \mathbb{R}^n$ and the notion of *exterior differential* distinguish k-forms from maps into a vector space of dimension $\binom{n}{k}$.

Let $\Omega \subset \mathbb{R}^n$ be an open set. The *exterior differential* of a k-form ω of class C^1 is defined as the $(k+1)$-form $d\omega$ of class C^0 given by

$$d\omega(x) := \sum_{\alpha \in I(k,n)} \sum_{i=1}^{n} \frac{\partial \omega_\alpha}{\partial x^i}(x)\, dx^i \wedge dx^\alpha.$$

4.39 Example. In $\mathbb{R}^3$ the following hold:

(i) If $k = 0$ and $\omega(x) : \Omega \to \Lambda^0 \mathbb{R} = \mathbb{R}$, then

$$d\omega(x) = \omega_x\, dx + \omega_y\, dy + \omega_z\, dz,$$

i.e, $d\omega(x)$ is the differential of the function ω at x.

(ii) If $k = 1$ and $\omega(x) = a\, dx + b\, dy + c\, dz$, then

$$d\omega(x) = (c_y - b_z)\, dy \wedge dz + (c_x - a_z)\, dx \wedge dz + (b_x - a_y)\, dx \wedge dy.$$

(iii) If $k = 2$ and $\omega = A\, dx \wedge dy + B\, dx \wedge dz + C\, dy \wedge dz$, then

$$d\omega = \Big(\frac{\partial A}{\partial z} - \frac{\partial B}{\partial y} + \frac{\partial C}{\partial x}\Big)\, dx \wedge dy \wedge dz.$$

(iv) For every 3-form ω, we have $d\omega = 0$.

4.40 Proposition. *We have the following:*

(i) *If ω is a 0-form of class C^1, then $d\omega(x) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(x)\, dx^i$.*
(ii) *If $k \ge n$ and ω is a k-form of class C^1, then $d\omega(x) = 0\ \forall x$.*
(iii) *d is linear: For k-forms ω and η of class C^1, $k \ge 0$, and $\lambda, \mu \in \mathbb{R}$, we have $d(\lambda\omega + \mu\eta) = \lambda\, d\omega + \mu\, d\eta$.*
(iv) *If ω and η are respectively a h-form and a k-form of class C^1, $h, k \ge 0$, then $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^h \omega \wedge d\eta$.*

(v) *If ω is of class C^2, then $d(d\omega) = 0$.*

Proof. (i), (ii) and (iii) are trivial.
(iv) If $\omega = \sum_{\alpha\in I(h,n)} \omega_\alpha dx^\alpha$ and $\eta = \sum_{\beta\in I(k,n)} \eta_\beta dx^\beta$, then

$$\begin{aligned} d(\omega\wedge\eta) &= \sum_{i=1}^n \sum_{\alpha\in I(h,n)} \sum_{\beta\in I(k,n)} (\omega_\alpha\eta_\beta)_{x^i}\, dx^i\wedge dx^\alpha\wedge dx^\beta \\ &= \sum_{i=1}^n \sum_{\alpha\in I(h,n)} \sum_{\beta\in I(k,n)} (\omega_\alpha)_{x^i}\eta_\beta\, dx^i\wedge dx^\alpha\wedge dx^\beta \\ &\quad + (-1)^h \sum_{i=1}^n \sum_{\alpha\in I(h,n)} \sum_{\beta\in I(k,n)} \omega_\alpha(\eta_\beta)_{x^i}\, dx^\alpha\wedge dx^i\wedge dx^\beta \\ &= d\omega\wedge d\eta + (-1)^h\omega\wedge d\eta. \end{aligned}$$

(v) If $\omega = \sum_{\alpha\in I(h,n)} \omega_\alpha dx^\alpha$, then

$$\begin{aligned} d(d\omega) &= d\Big(\sum_{i=1}^n \sum_{\alpha\in I(h,n)} (\omega_\alpha)_{x^i} dx^i\wedge dx^\alpha\Big) \\ &= \sum_{\alpha\in I(h,n)} \sum_{\substack{i,j=1,n\\ i<j}} \Big(\frac{\partial^2\omega_\alpha}{\partial x^i\partial x^j} - \frac{\partial^2\omega_\alpha}{\partial x^j\partial x^i}\Big) dx^i\wedge dx^j\wedge dx^\alpha \\ &= 0 \end{aligned}$$

by Schwarz's theorem, see [GM4]. □

4.41 Remark. As seen, property (iv), $d(d\omega) = 0$, amounts to the equality of the mixed second derivatives of the components of ω. It is a source of several "integrability conditions" in the theory of PDE's and in differential geometry, see e.g., the proof of Poincaré's lemma, Theorem 4.75.

b. Pull-back of differential forms

The structure of exterior algebra on the spaces $\{\Lambda^k\mathbb{R}^n\}$ explicitly shows up in several issues, as for instance when dealing with the *inverse image* of a k-form.

Let $U\subset\mathbb{R}^n$ and $\Omega\subset\mathbb{R}^N$ be open sets and let $\phi : U\to\Omega$, $\phi = (\phi^1, \phi^2, \dots, \phi^N)$, a map of class C^1. As usual, denote by $d\phi(u) : \mathbb{R}^n\to\mathbb{R}^N$ the linear tangent map to ϕ at u. Given a differential k-form ω of class C^1 on $\Omega\subset\mathbb{R}^N$,

$$\omega = \sum_{\alpha\in I(k,N)} \omega_\alpha(x) dx^\alpha,$$

the *pull-back* or *inverse image* of ω is the differential k-form in $U\subset\mathbb{R}^n$ defined for every $u\in U$ by

$$\phi^\#\omega(u) := \begin{cases} \omega(\phi(u)) & \text{if } k = 0, \\ 0 & \text{if } k > \min(n, N), \\ \sum_{\alpha\in I(k,N)} \omega_\alpha(\phi(u))\, d\phi^{\alpha_1}(u)\wedge\cdots\wedge d\phi^{\alpha_k}(u) \end{cases}$$

if $1 \le k \le \min(n, N)$. If we introduce the exterior power of $d\phi(u)$, that is for $1 \le k \le n$,

$$\Lambda^k(d\phi(u))(dx^\alpha) := d\phi^{\alpha_1}(u) \wedge \cdots \wedge d\phi^{\alpha_k}(u),$$

and $\Lambda^0(d\phi(u)) = \mathrm{Id}$, we have

$$\begin{aligned} \phi^\# \omega(u) &= \sum_{\alpha \in I(k,N)} \omega_\alpha(\phi(u)) \Lambda^k(d\phi(u))(dx^\alpha) \\ &= \Lambda^k(d\phi(u)) \Big(\omega(\phi(u)) \Big) \end{aligned} \tag{4.37}$$

i.e., $\phi^\#\omega(u)$ is the k-covector in $\Lambda_k \mathbb{R}^n$ image of $\omega(\phi(u))$ through $\Lambda^k(d\phi_u)$.

We can compute explicitly the components of $\phi^\#\omega$ by means of (4.19): If $\mathbf{D}\phi(u)$ is the Jacobian of ϕ at u in a given basis, then

$$\begin{aligned} \phi^\# \omega(u) &= \sum_{\alpha \in I(k,n)} \omega_\alpha(\phi(u)) \Lambda^k(d\phi_u)(dx^\alpha) \\ &= \sum_{\beta \in I(k,n)} \Bigg(\sum_{\alpha \in I(k,N)} \omega_\alpha(\phi(u)) M^\alpha_\beta(\mathbf{D}\phi(u)) \Bigg) du^\beta. \end{aligned}$$

4.42 Remark. Notice that for a k-form ω of class C^r, $r \ge 1$, and a map ϕ of class C^s, $s \ge 1$, $\phi^\#\omega$ is of class $\min(r, s-1)$ if $k > 0$ and of class $\min(r, s)$ if $k = 0$.

4.43 Proposition. *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^N$ be open sets and let $\phi : U \to V$ be a map of class C^1. Then the following hold:*

(i) *$\phi^\#$ is linear.*
(ii) *For a k-form ω and an h-form η on V with continuous coefficients, we have $\phi^\#(\omega \wedge \eta) = \phi^\#\omega \wedge \phi^\#\eta$.*
(iii) *If ω is a k-form of class C^2, $k > 0$, and ϕ is of class C^2, then $d\phi^\#\omega$ and $\phi^\#(d\omega)$ have continuous coefficients and $d\phi^\#\omega = \phi^\# d\omega$.*

Proof. (i) is trivial. (ii) follows from (4.17) and (4.37).

(iii). Since $d^2 = 0$, we have

$$d(d\phi^{\alpha_1} \wedge \cdots \wedge d\phi^{\alpha_k}) = 0,$$

consequently, for

$$\omega = \sum_{\alpha \in I(k,N)} \omega_\alpha(x) dx^\alpha$$

we have

$$
\begin{aligned}
d\phi^{\#}\omega &= \sum_{\alpha\in I(k,N)} \sum_{i=1}^{n} \frac{\partial \omega_\alpha(\phi(u))}{\partial u^i}\, du^i \wedge d\phi^{\alpha_1} \wedge \cdots \wedge d\phi^{\alpha_k} + 0 \\
&= \sum_{\alpha\in I(k,N)} \sum_{i=1}^{n} \sum_{j=1}^{N} \frac{\partial \omega_\alpha}{\partial x^j}(\phi(u)) \frac{\partial \phi^j}{\partial u^i}(u)\, du^i \wedge d\phi^{\alpha_1} \wedge \cdots \wedge d\phi^{\alpha_k} \\
&= \sum_{\alpha\in I(k,N)} \sum_{j=1}^{N} \frac{\partial \omega_\alpha}{\partial x^j}(\phi(u))\, d\phi^j \wedge d\phi^{\alpha_1} \wedge \cdots \wedge d\phi_{\alpha_k} \\
&= \phi^{\#}(d\omega).
\end{aligned}
$$

□

4.2.2 The area formula on submanifolds

In connection with the definition and the properties of the integral of a differential k-form on a k-submanifold or, more generally, on the injective image of a k-submanifold, the *area formula* or the *change of variable formula* (that we discussed in [GM4] and will be proved in Theorem 5.100, and that we restate for the reader's convenience) plays an important role.

We recall that a nonempty set $\mathcal{X} \subset \mathbb{R}^n$ is a *k-dimensional submanifold* of $\mathbb{R}^n$ if $\mathcal{X}$ is locally diffeomorphic to an open set of $\mathbb{R}^k$. More precisely, $\mathcal{X}$ is a k-submanifold of $\mathbb{R}^n$ if for every point $x \in \mathcal{X}$ there exist an open set $\Omega_x \subset \mathbb{R}^n$, an open set $U_x \subset \mathbb{R}^k$ and a diffeomorphism $\varphi_x : U_x \to \Omega_x \cap \mathcal{X}$, see [GM4]. Of course, we may refine the open covering $\{\Omega_x\}_{x\in\mathcal{X}}$ to a denumerable subcovering, indeed a denumerable and locally finite subcovering[1]. Consequently, we may associate to the system of local charts a *decomposition of unity* $\{\alpha_i\}$

(i) $0 \le \alpha_i \le 1\ \forall i$,
(ii) $\alpha_i \in C_c^\infty(\Omega_i)$,
(iii) $\sum_i \alpha_i = 1$ on $\mathcal{X}$.

The k-submanifolds we have just defined are usually called *submanifolds without boundary*; but since we shall not discuss manifolds with boundary in details, we stick to with our submanifold notation.

a. The area formula

Let $\phi : U \subset \mathbb{R}^k \to \mathbb{R}^n$, $n \ge k$, be a map of class C^1 defined on the open set U. Choose orthonormal coordinates in $\mathbb{R}^k$ and $\mathbb{R}^n$ and let $(e_1, e_2, \ldots, e_k)$ be the chosen basis in $\mathbb{R}^k$. Set, for $u \in U$

$$
J(\mathbf{D}\phi(u)) := \det(\mathbf{D}\phi(u)^T \mathbf{D}\phi(u))^{1/2} = |\Lambda_k(\mathbf{D}\phi(u))(e_1 \wedge e_2 \wedge \cdots \wedge e_k)|
$$

and

[1] The claim is trivial if $\mathcal{X}$ is compact. For the general case, see e.g., M. Berger, B. Gaustiaux, *Géometrie différentielle: variétés, courbes et surfaces*, Presses Universitaires de France, Paris, **1992**, p. 117.

$$R := \Big\{ u \in U \;\Big|\; \operatorname{Rank} \mathbf{D}\phi(u) = k \Big\} = \Big\{ u \in U \;\Big|\; J(\mathbf{D}\phi(u)) \neq 0 \Big\}.$$

4.44 Theorem (Change of variable formula in $\mathbb{R}^n$). *Let $f : U \to \mathbb{R}$ be $\mathcal{L}^k$-measurable. Then the following hold:*

(i) *The function*
$$F(x) := \sum_{\substack{u \in R \\ \phi(u) = x}} f(u)$$
is $\mathcal{H}^k$-measurable in $\mathbb{R}^n$.

(ii) *F is $\mathcal{H}^k$-summable if and only if $u \mapsto f(u) J(\mathbf{D}\phi(u))$ is summable in U.*

(iii) *The change of variable formula holds:*
$$\int_U f(u) J(\mathbf{D}\phi(u))\, du = \int_{\phi(U)} F(x)\, d\mathcal{H}^k(x).$$

In particular, it follows that

$$\int_U g(\phi(u)) J(\mathbf{D}\phi(u))\, du = \int_{\mathbb{R}^n} g(x)\, d\mathcal{H}^k(x) \tag{4.38}$$

for every $\mathcal{H}^k$-measurable function in $\mathbb{R}^N$, and, taking as g the characteristic function of $\phi(U \setminus R)$, we get that the set

$$\phi(U \setminus R) = \phi\Big(\Big\{ u \in U \;\Big|\; J(\mathbf{D}\phi(u)) = 0 \Big\} \Big)$$

is a null set.

b. The area formula on submanifolds

Let $\mathcal{X} \subset \mathbb{R}^n$ be a k-submanifold of $\mathbb{R}^n$, $k \leq n$, and let $\phi : \mathcal{X} \to \mathbb{R}^N$, $N \geq k$, be a map of class C^1. Let $\xi : \mathcal{X} \to \Lambda_k \mathbb{R}^n$ be a $\mathcal{H}^k$-measurable field of unit k-vectors on $\mathcal{X}$ that spans $\operatorname{Tan}_x \mathcal{X}$ for $\mathcal{H}^k$-a.e. x. Introduce the *Jacobian* of ϕ at $x \in \mathcal{X}$

$$J(\mathbf{D}\phi(x)) := \Big| \Lambda_n \mathbf{D}\phi(x)(\xi(x)) \Big|$$

and consider the *regular points of* ϕ,

$$R := \Big\{ x \in \mathcal{X} \;\Big|\; J(\mathbf{D}\phi(x)) \neq 0 \Big\},$$

and for $v : \mathcal{X} \to \mathbb{R}$ set

$$V(y) := \sum_{\substack{x \in R \\ \phi(x) = y}} v(x), \qquad y \in \mathbb{R}^N.$$

4.45 Theorem (Change of variables formula). *Let $v : \mathcal{X} \to \mathbb{R}$ be a $\mathcal{H}^k$-measurable function. Then V is $\mathcal{H}^k$-measurable in $\mathbb{R}^N$ and V is $\mathcal{H}^k$-summable if and only if $x \mapsto v(x)|\Lambda_n \mathbf{D}\phi(x)(\xi(x))|$ is $\mathcal{H}^k$-summable in $\mathcal{X}$. Moreover,*

$$\int_{\mathcal{X}} v(x) J(\mathbf{D}\phi(x))\, d\mathcal{H}^k(x) = \int_{\phi(\mathcal{X})} V(y)\, d\mathcal{H}^k(y). \tag{4.39}$$

Proof. The claim follows from Theorem 4.44. using local coordinates and a partition of unity.

Choose a locally finite open cover $\{\Omega_i\}$, open sets $U_i \subset \mathbb{R}^k$ and diffeomorphisms $\varphi_i : U_i \to \Omega_i \cap \mathcal{X}$. Choose an orthonormal basis $(e_1, e_2, \dots, e_k)$ in each U_i and let $\psi_i := \phi \circ \varphi_i : U_i \to \mathbb{R}^N$. Then

$$\begin{aligned}
&\Lambda_k D\varphi_i(u)(e_1 \wedge e_2 \wedge \cdots \wedge e_k) = \lambda(u)\xi(x),\\
&\Lambda_k D\psi_i(u)(e_1 \wedge e_2 \wedge \cdots \wedge e_k) = \Lambda_k D\phi(x)(\lambda(u)\xi(x)),
\end{aligned}$$

where $x := \varphi_i(u)$, hence

$$\begin{aligned}
J(\mathbf{D}\psi_i(u)) &= |\Lambda_k D\psi_i(u)(e_1 \wedge e_2 \wedge \cdots \wedge e_k)|\\
&= |\Lambda_k D\phi(x)(\xi(x))|\, |\Lambda_k (D\varphi_i(u)(e_1 \wedge e_2 \wedge \cdots \wedge e_k)|\\
&= J(\mathbf{D}\phi(x))\, J(\mathbf{D}\varphi_i(u)).
\end{aligned} \tag{4.40}$$

Let $v_i(x) := \alpha_i(x)v(x)$ and

$$V_i(y) = \sum_{\substack{x \in R\\ \phi(x)=y}} v_i(x) = \sum_{\substack{u \in R_i\\ \psi(u)=y}} v_i(\varphi_i(u))$$

where $R_i := \varphi_i^{-1}(R)$. The area formula (4.38) applied to the maps $v_i(x)J(\mathbf{D}\phi(x))$ and $V_i(x)$ yields that $v_i(\varphi_i(u))J(\mathbf{D}\phi(\varphi_i(u)))J(\mathbf{D}\varphi_i(u))$ is $\mathcal{L}^k$-measurable if and only if $v_i(x)J(\mathbf{D}\phi(x))$ is $\mathcal{H}^k$-measurable if and only if V_i is $\mathcal{H}^k$-measurable and

$$\int_{U_i} v_i(\varphi_i(u))J(\mathbf{D}\phi(\varphi_i(u)))J(\mathbf{D}\varphi_i(u))\, d\mathcal{L}^k(u) = \int_{\Omega_i \cap \mathcal{X}} v_i(x)J(\mathbf{D}\phi(x))\, d\mathcal{H}^k(x), \tag{4.41}$$

and

$$\int_{U_i} v_i(\varphi_i(u))J(\mathbf{D}\psi_i(u))\, d\mathcal{L}^k(u) = \int_{\mathbb{R}^N} V_i(y)\, d\mathcal{H}^k(y), \tag{4.42}$$

hence, because of (4.40)

$$\int_{\Omega_i \cap \mathcal{X}} v_i(x)J(\mathbf{D}\phi(x))\, d\mathcal{H}^k(x) = \int_{\mathbb{R}^N} V_i(y)\, d\mathcal{H}^k(y).$$

Summing on i, the claim follows. □

As a consequence of the area formula, we get a Sard-type theorem, see Theorem 5.55 of [GM4]: *The image of the set of nonregular points of a C^1 map defined on a C^1 k-submanifold of $\mathbb{R}^n$ is a null set*:

$$\mathcal{H}^k(\phi(\mathcal{X} \setminus R)) = 0.$$

4.2.3 The oriented integral

In this subsection we introduce the *oriented integral* of a differential k-form over suitable k-dimensional sets of $\mathbb{R}^n$, $k \leq n$, such as *oriented (sub)manifolds* and C^1 *injective images of oriented manifolds*.

Essentially, the *oriented integral* of a k-form is defined once we are given

(i) an $\mathcal{H}^k$-measurable set $S \subset \mathbb{R}^n$,
(ii) a field $\xi : S \to \Lambda_k \mathbb{R}^n$ of $\mathcal{H}^k$-measurable k-vectors with $|\xi(x)| = 1$ for $\mathcal{H}^k$ a.e. $x \in S$.

Indeed, if ω is a k-form with $\mathcal{H}^k$-summable coefficients on S, $\int_S |\omega| \, d\mathcal{H}^k < +\infty$, then the *integral of* ω *over* S *in the direction* ξ is

$$\tau(S, \xi)(\omega) := \int_S < \omega(x), \xi(x) > \, d\mathcal{H}^k(x).$$

Of course, the integral depends on the chosen direction field. Sometimes the ambiguous notation

$$\int_S \omega := \tau(S, \xi)(\omega) = \int_S < \omega(x), \xi(x) > \, d\mathcal{H}^k(x) \tag{4.43}$$

is used specifying the direction field in the context in which the notation appears. As stated, ω is summable on S if

$$\int_S |\omega(x)| \, d\mathcal{H}^k(x) < +\infty;$$

we say that it is *summable on* S *along* ξ if

$$\int_S | < \omega(x), \xi(x) > | \, d\mathcal{H}^k(x) < +\infty,$$

i.e., if the component of ω in the direction ξ is summable on S. Notice the following:

(i) We have

$$| < \omega(x), \xi(x) > | \leq |\omega(x)| \, |\xi(x)| \leq |\omega(x)| \qquad \forall x \in S,$$

hence

$$\int_S | < \omega(x), \xi(x) > | \, d\mathcal{H}^k(x) \leq \int_S |\omega(x)| \, d\mathcal{H}^s(x),$$

i.e., ω is $\mathcal{H}^k$-summable on S along any direction if the coefficients of ω are summable on S.
(ii) Every continuous k-form ω on an open set U is bounded on S if $S \subset\subset U$ and bounded, and $\mathcal{H}^k$-summable on S if $S \subset\subset U$ and $\mathcal{H}^k(S) < +\infty$.

Let us discuss some interesting cases of the natural and tacitly understood choice of the direction field ξ.

a. Oriented open sets in $\mathbb{R}^k$

Consider $\mathbb{R}^k$ oriented by an orthonormal basis $(e_1, e_2, \dots, e_k)$. Choose the constant n-vector

$$e_1 \wedge e_2 \wedge \cdots \wedge e_k.$$

Since $\mathcal{L}^k = \mathcal{H}^k$, see Theorem 6.75, (4.43) defines the *oriented integral* over an $\mathcal{L}^k$-measurable set $U \subset \mathbb{R}^k$ of a k-form as

$$\int_U \omega := \int_U < \omega(x), e_1 \wedge e_2 \wedge \cdots \wedge e_k > \, d\mathcal{L}^k(x).$$

Of course, the symbol $\int_U \omega$ is ambiguous since its value depends on the orientation of $\mathbb{R}^k$.

b. Oriented k-submanifolds of $\mathbb{R}^n$

4.46 Definition. *A k-submanifold $\mathcal{X}$ of $\mathbb{R}^n$ is said to be* orientable *if there is a* continuous *field $\xi : \mathcal{X} \to \Lambda_k\mathbb{R}^n$ of unit k-vectors such that $\xi(x)$ orients* $\mathrm{Tan}_x \mathcal{X}$ *$\forall x \in \mathcal{X}$, i.e., $\xi(x)$ is simple, $|\xi(x)| = 1$ and* $\mathrm{Span}(\xi(x)) = \mathrm{Tan}_x \mathcal{X}$ *$\forall x \in \mathcal{X}$. We say that $\xi : \mathcal{X} \to \Lambda_k\mathbb{R}^n$* orients *$\mathcal{X}$.*

We notice the following:

(i) If $\mathbb{R}^n$ is oriented by an orthonormal basis $(e_1, e_2, \dots, e_n)$, then every open set $\Omega \subset \mathbb{R}^n$ is a n-submanifold of $\mathbb{R}^n$ oriented by $e_1 \wedge e_2 \wedge \cdots \wedge e_n$.
(ii) If $\mathcal{X}$ is orientable and connected, there are exactly two possible orientations of $\mathcal{X}$, one opposite the other.
(iii) There exist nonorientable submanifolds, as for instance, the Möbius strip.

Let $\mathcal{X}$ be a k-submanifold of class C^1 oriented by $\xi : \mathcal{X} \to \Lambda_k\mathbb{R}^n$. Since $\mathcal{X}$ is a denumerable union of compacts, $\mathcal{X}$ is $\mathcal{H}^k$-measurable. Therefore, (4.43) defines the *oriented integral* of a k-form ω with summable coefficients over an oriented k-submanifold $\mathcal{X}$ oriented by ξ as

$$\int_{\mathcal{X}} \omega := \int_{\mathcal{X}} < \omega(x), \xi(x) > \, d\mathcal{H}^k(x).$$

Notice the ambiguity of the symbol $\int_{\mathcal{X}} \omega$ that does not specify the dependence on the orienting field ξ on $\mathcal{X}$.

c. Admissible open sets

Recall, see Chapter 2 of [GM4], that a bounded open set $\Omega \subset \mathbb{R}^n$ is said to be *admissible* if its boundary is $\mathcal{H}^{n-1}$-measurable with finite $\mathcal{H}^{n-1}$-measure and decomposes as $\partial\Omega = R \cup N$, where R is a $(n-1)$-submanifold and N is a closed set with $\mathcal{H}^{n-1}(N) = 0$.

We now define an "orientation" on $\partial\Omega$. First, recall that if $R \neq \emptyset$, we may consider the *field of exterior unit normal vectors to* Ω $\nu_R(x)$ at $x \in R$ and notice that $\nu_R(x)$ is continuous on R and $\mathcal{H}^{n-1}$-measurable.

Moreover, such a field is in fact uniquely defined $\mathcal{H}^{n-1}$-a.e. on $\partial\Omega$. In fact, if $\partial\Omega = R \cup N = R_1 \cup N_1$ with $\mathcal{H}^{n-1}(N) = \mathcal{H}^{n-1}(N_1) = 0$ and R and R_1 being submanifolds of $\mathbb{R}^n$, then $\nu_R(x) = \nu_{R_1}(x)$ for all $x \in R \cap R_1$ since $R \cap R_1$ is open both in R and R_1 and $\mathcal{H}^{n-1}(\partial\Omega \setminus R \cap R_1) = 0$. Therefore, the orientation of Ω uniquely defines the *field of exterior unit normal vectors to* $\partial\Omega$,

$$\nu_\Omega(x) := \nu_R(x) \qquad \mathcal{H}^{n-1} \text{ a.e. } x \in \partial\Omega. \tag{4.44}$$

Now, if $*$ is the Hodge operator associated to the orientation of $\mathbb{R}^n$, again (4.43) allows us to define the *oriented* (by the exterior normal) *integral* of an $(n-1)$-form with summable coefficients on $\partial\Omega$,

$$\int_{\partial\Omega} \omega := \int_{\partial\Omega} < \omega(x), *\nu_\Omega(x) > \, d\mathcal{H}^{n-1}(x),$$

where ν_Ω is given by (4.44). Of course, the symbol $\int_{\partial\Omega} \omega$ depends only on ω, $\partial\Omega$ and implicitly on the orientation of $\mathbb{R}^n$.

4.47 . Let us compute in local coordinates the oriented integral of an $(n-1)$-form on the boundary of an admissible domain. Let $(e_1, e_2, \ldots, e_n)$ be an orthonormal basis that orients $\mathbb{R}^n$. If $\nu(x) = \sum_{i=1}^n \nu^i e_i$ is the exterior vector field and $\omega := \sum_{i=1}^n (-1)^{i-1} \omega_i(x) \widehat{dx^i}$, [2] we have

$$*\nu = \sum_{i=1}^n (-1)^{i-1} \nu^i(x) \widehat{e_i}, \qquad *\omega = \sum_{i=1}^n (-1)^{n-i} \omega_i(x) dx^i$$

and $**\nu = (-1)^{1(n-1)} = (-1)^{n-1}$. Therefore, we compute

$$\begin{aligned} \int_{\partial\Omega} \omega :&= \int_{\partial\Omega} < \omega(x), *\nu(x) > \, d\mathcal{H}^{n-1}(x) \qquad (4.45) \\ &= \int_{\partial\Omega} < *\omega(x), *(*\nu(x)) > \, d\mathcal{H}^{n-1}(x) \\ &= \int_{\partial\Omega} \sum_{i=1}^n \omega_i(x) \nu^i(x) \, d\mathcal{H}^{n-1}(x). \qquad (4.46) \end{aligned}$$

d. Immersions and C^1 images of an open set

Let $\varphi : U \subset \mathbb{R}^k \to \mathbb{R}^n$, $n \geq k$, be an injective map of class C^1, U open. Fix an orientation on U by choosing a basis $(e_1, e_2, \ldots, e_k)$ in $\mathbb{R}^k$.

Since U is the denumerable union of compact sets, $\varphi(U)$ is $\mathcal{H}^k$-measurable.

We now define an orientation on $\phi(U)$ as follows. Let $\mathbb{R}^n$ be oriented by the choice of a basis, and let $\mathbf{D}\varphi(u)$ be the Jacobian matrix of φ at $u \in U$ and

[2] $\widehat{dx^i} = dx^{\bar{i}}$ and $\widehat{e_i} = e_{\bar{i}}$.

ERICH KÄHLER
1962
Rendiconti di Matematica
(3-4) Vol. 21, pp. 425-523.

Der innere Differentialkalkül(*)

Dem Andenken
Wilhelm Blaschkes
gewidmet

von ERICH KÄHLER (in Berlin)

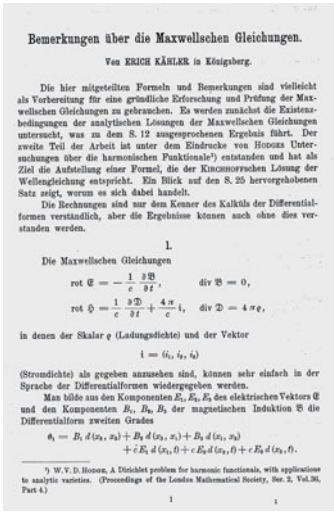
Wenn eine Metrik vorliegt, wie es in der Physik der Fall ist und in der Funktionentheorie sich als fruchtbare Voraussetzung erwiesen hat, gewinnt der äussere Differentialkalkül ein rechnerisch nahezu ebenso einfaches Spiegelbild, den inneren Differentialkalkül.

Was ich darüber bei einem Lehrgang des Centro Internazionale Matematico Estivo im September 1960 in Vallombrosa vorgetragen hatte, findet in der vorliegenden Abhandlung ausführliche und reifere Darstellung. Die grosse Verzögerung dieser Veröffentlichung erklärt sich aus der erfreulichen Tatsache, dass jeder Versuch einer Ausarbeitung jener Vorträge neue Vereinfachungen ergab mit dem Erfolg, dass insbesondere meine in den Abhandlungen der Berliner Akademie veröffentlichte Untersuchung der Dirac-Gleichung nunmehr als überholt anzusehen ist.

Da der innere Differentialkalkül seine Bewährungsprobe in der Quanten- und Relativitätstheorie zu bestehen hat, muss er dem Physiker zugänglich sein, weshalb es mir zweckmässig schien, auch über den äusseren Differentialkalkül mehr zu sagen, als zur Vorbereitung des inneren Kalküls notwendig gewesen wäre.

Gern hätte ich, wenn mehr Zeit geblieben wäre ein viertes Kapitel der funktionentheoretischen Seite des inneren Kalküls gewidmet, was nunmehr einer anderen Arbeit vorbehalten bleiben muss.

(*) Corso di otto lezioni avute nel Ciclo del CIME (Centro Internazionale Matematico Estivo) su *Forme Differenziali e loro integrali*, tenuto al Saltino di Vallombrosa (Firenze) dal 23 al 31 agosto 1960.



Bemerkungen über die Maxwellschen Gleichungen.

Von ERICH KÄHLER in Königsberg.

Die hier mitgeteilten Formeln und Bemerkungen sind vielleicht als Vorbereitung für eine gründliche Erforschung und Prüfung der Maxwellschen Gleichungen zu gebrauchen. Es werden zunächst die Existenzbedingungen der analytischen Lösungen der Maxwellschen Gleichungen untersucht, was zu dem S. 12 ausgesprochenen Ergebnis führt. Der zweite Teil der Arbeit ist unter dem Eindrucke von Hodges Untersuchungen über die harmonischen Funktionale[1]) entstanden und hat als Ziel die Aufstellung einer Formel, die der Kirchhoffschen Lösung der Wellengleichung entspricht. Ein Blick auf den S. 25 hervorgehobenen Satz zeigt, worum es sich dabei handelt.

Die Rechnungen sind nur dem Kenner des Kalküls der Differentialformen verständlich, aber die Ergebnisse können auch ohne dies verstanden werden.

1.

Die Maxwellschen Gleichungen

$$\text{rot } \mathfrak{E} = -\frac{1}{c}\frac{\partial \mathfrak{B}}{\partial t}, \qquad \text{div } \mathfrak{B} = 0,$$

$$\text{rot } \mathfrak{H} = \frac{1}{c}\frac{\partial \mathfrak{D}}{\partial t} + \frac{4\pi}{c}\mathfrak{i}, \qquad \text{div } \mathfrak{D} = 4\pi\varrho,$$

in denen der Skalar ϱ (Ladungsdichte) und der Vektor

$$\mathfrak{i} = (i_1, i_2, i_3)$$

(Stromdichte) als gegeben anzusehen sind, können sehr einfach in der Sprache der Differentialformen wiedergegeben werden.

Man bilde aus den Komponenten E_1, E_2, E_3 des elektrischen Vektors $\mathfrak{E}$ und den Komponenten B_1, B_2, B_3 der magnetischen Induktion $\mathfrak{B}$ die Differentialform zweiten Grades

$$\vartheta_1 = B_1\, d(x_2, x_3) + B_2\, d(x_3, x_1) + B_3\, d(x_1, x_2) + \tilde{c} E_1\, d(x_1, t) + c E_2\, d(x_3, t) + c E_3\, d(x_2, t).$$

[1]) W. V. D. Hodge, A Dirichlet problem for harmonic functionals, with applications to analytic varieties. (Proceedings of the London Mathematical Society, Ser. 2, Vol.36, Part 4.)

1 1

Figure 4.3. Two pages from two papers by Erich Kähler (1906–2000).

$$J(\mathbf{D}\varphi(u)) = (\det(\mathbf{D}\varphi(u)^T \mathbf{D}\varphi(u)))^{1/2}.$$

We have $J(\mathbf{D}\varphi)(u) \neq 0$ if and only if $\operatorname{Rank} \mathbf{D}\varphi(u) = k$ and

$$\begin{aligned} |J(\mathbf{D}\varphi(x))| &= \det(\mathbf{D}\varphi(x)^T \mathbf{D}\varphi(x)) \\ &= |\mathbf{D}\varphi(e_1) \wedge \cdots \wedge \mathbf{D}\varphi(e_k)|^2 \\ &= |\Lambda_k(\mathbf{D}\varphi(x))(e_1 \wedge e_2 \wedge \cdots \wedge e_n)|^2. \end{aligned} \tag{4.47}$$

Now, we set

$$R := \Big\{ u \in U \;\Big|\; \operatorname{Rank} \mathbf{D}\varphi(u) = k \Big\},$$

so that $\varphi_{|R}$ is an injective immersion. Then $\varphi_{|R}$ is a local homeomorphism, $\varphi^{-1} : \varphi(R) \to R$ is continuous, and $\varphi(R)$ has a tangent plane $\operatorname{Tan}_x \varphi(R)$ at every $x \in \varphi(R)$ defined as

$$\operatorname{Tan}_x \varphi(R) := \operatorname{Im} \mathbf{D}\varphi(u), \qquad u = \varphi^{-1}(x),$$

of dimension k. The field $\alpha : \varphi(R) \to \Lambda_k \mathbb{R}^n$,

$$\alpha(x) := \Lambda_k(\mathbf{D}\varphi(u))(e_1 \wedge e_2 \wedge \cdots \wedge e_k) = \frac{\partial \varphi}{\partial u^1}(u) \wedge \cdots \wedge \frac{\partial \varphi}{\partial u^k}(u),$$

is nonzero, continuous and simple and $\operatorname{Span}(\alpha(x)) = \operatorname{Tan}_x \varphi(R)$ since the vectors

$$\frac{\partial \varphi}{\partial u^1}(u), \;\; \frac{\partial \varphi}{\partial u^2}(u), \;\; \ldots, \;\; \frac{\partial \varphi}{\partial u^k}(u)$$

form a basis of $\operatorname{Im} \mathbf{D}\varphi(u)$. Consequently, the field of k-vectors

$$\xi(x) := \frac{\alpha(x)}{|\alpha(x)|} \tag{4.48}$$

orients $\mathrm{Tan}_x\, \varphi(R)$. Notice that ξ depends on the chosen orientation of $\mathbb{R}^k$.

On the other hand, from the Sard type theorem, see Theorem 5.55 of [GM4], (or from the area formula that implies it, compare Chapter 2 of [GM4] and Theorem 4.44), we infer

$$\mathcal{H}^k(\varphi(U) \setminus \varphi(R)) = \mathcal{H}^k(\varphi(U \setminus R)) = 0.$$

Therefore, ξ is $\mathcal{H}^k$-measurable and $\mathcal{H}^k$-a.e. defined on $\varphi(U)$: ξ is by definition the *orientation on* $\varphi(U)$ induced by U.

Then, again by means of (4.43), we define the *oriented integral over* $\varphi(U)$ oriented by the orientation induced by U of a $\mathcal{H}^k$-summable k-form ω as

$$\int_{\varphi(U)} \omega := \int_{\varphi(U)} < \omega(x), \xi(x) > \, d\mathcal{H}^k(x), \tag{4.49}$$

where ξ is given by (4.48). Notice that the integral depends on the orientation of U.

e. C^1 images of oriented submanifolds

Let Ω be an open set of $\mathbb{R}^n$ and $\phi : \Omega \to \mathbb{R}^N$ be an injective map of class C^1; also, let $k \leq \min(n, N)$ and

$$S = \mathcal{X} \cup \mathcal{N} \subset\subset \Omega,$$

where $\mathcal{X}$ is a k-submanifold of $\mathbb{R}^n$ oriented by a continuous field $\eta : \mathcal{X} \to \Lambda_k \mathbb{R}^n$ of k-vectors and $\mathcal{H}^k(\mathcal{N}) = 0$. Examples of S are given by open sets of $\mathbb{R}^n$ for $k = n$ and boundaries of admissible open sets for $k = n - 1$.

Since $\mathcal{H}^k(\mathcal{N}) = 0$, the unit n-vector field η that orients $\mathcal{X}$ is defined $\mathcal{H}^k$-a.e. on S and is $\mathcal{H}^k$-measurable. Moreover, we may think of η as the unit vector that orients S since it does not depend on the decomposition of S for $\mathcal{H}^k$-a.e $x \in S$. In fact, if $S = \mathcal{X}_1 \cup \mathcal{N}_1 = \mathcal{X}_2 \cup \mathcal{N}_2$ with $\mathcal{H}^k(\mathcal{N}_1) = \mathcal{H}^k(\mathcal{N}_2) = 0$ and $\mathcal{X}_1$ and $\mathcal{X}_2$ are k-submanifolds of $\mathbb{R}^n$ oriented in such a way that $\mathcal{X}_1$ and $\mathcal{X}_2$ have the same orientation on $\mathcal{X}_1 \cap \mathcal{X}_2$, then $\mathcal{H}^k(S \setminus (\mathcal{X}_1 \cap \mathcal{X}_2)) = 0$.

Since ϕ maps compact sets into compact sets and $\mathcal{H}^k$-null sets into $\mathcal{H}^k$-null sets, $\phi(S)$ is $\mathcal{H}^k$-measurable and $\mathcal{H}^k(\phi(\mathcal{N})) = 0$.

We now define an orientation on $\phi(S)$ as follows. Introduce the tangential Jacobian to $\mathcal{X}$ at $u \in \mathcal{X}$,

$$J(\mathbf{D}\phi(u)) := |\Lambda_k(\mathbf{D}\phi(u))(\eta(u))|,$$

and consider the *regular values of* ϕ,

$$R := \Big\{ u \in \mathcal{X} \,\Big|\, J(\mathbf{D}\phi(u)) \neq 0 \Big\}.$$

For each $u \in \mathcal{X}$, choose an orthonormal basis $(v_1, v_2, \ldots, v_k)$ of $\mathrm{Tan}_u\, \mathcal{X}$ that orients $\mathrm{Tan}_u\, \mathcal{X}$ as η, i.e., $\eta(u) = v_1 \wedge v_2 \wedge \cdots \wedge v_k$. Then for $x = \varphi(u)$

$$\alpha(x) := \Lambda_k(\mathbf{D}\phi(u))(\eta(u)) = \frac{\partial\phi}{\partial v^1} \wedge \cdots \wedge \frac{\partial\phi}{\partial v^k}(u)$$

is nonzero if and only if $u \in R$; consequently,

$$\xi(x) := \frac{\alpha(x)}{|\alpha(x)|} \tag{4.50}$$

orients $\mathrm{Tan}_x\, \phi(R)$ $\forall x \in \phi(R)$. Furthermore, the unit field $x \to \xi(x)$ is continuous on $\phi(R)$, in fact, the map $x \mapsto u := \phi^{-1}(u)$ is continuous in $\phi(R)$ since ϕ is continuous and injective in the compact set $\overline{S} \supset R$. In particular, $\xi(x)$ is $\mathcal{H}^k$-measurable.

Finally, Theorem 4.45 implies the Sard-type property

$$\mathcal{H}^k(\phi(S) \setminus \phi(R)) = \mathcal{H}^k(\phi(\mathcal{X} \setminus R)) = 0.$$

Therefore, the field ξ in (4.50) is well-defined $\mathcal{H}^k$-a.e. on $\phi(S)$. It is referred to as the *orientation on $\phi(S)$ induced by the orientation on S*.

Again, by means of (4.43), we define the *oriented integral* of a differential k-form with summable coefficients on $\phi(S)$ oriented by the orientation induced by the orientation of S as

$$\int_{\phi(S)} \omega := \int_{\phi(S)} < \omega(x), \xi(x) > \, d\mathcal{H}^k(x),$$

where for $\mathcal{H}^k$-a.e. x, $\xi(x)$ is defined by (4.50). As previously, the symbol $\int_{\phi(S)} \omega$ is ambiguous since it does not explicitly show the dependence on the orientation of $\mathcal{X}$.

4.2.4 Integration and pull-back

A rewriting of the area formula yields also the interplay of integration and pull-back of differential forms.

4.48 Proposition. *Let U be an oriented open set of $\mathbb{R}^k$, $\phi : U \to \mathbb{R}^n$ be an injective C^1 map and let $\phi(U)$ be oriented by the orientation induced by the orientation of U. Then for every differential k-form that is summable on $\phi(U)$, the pull-back $\phi^{\#}\omega$ is summable in U and*

$$\int_{\phi(U)} \omega = \int_U \phi^{\#}\omega.$$

Proof. Let $(e_1, e_2, \ldots, e_k)$ be a basis that orients $\mathbb{R}^k$, $\alpha(u) := \Lambda_k(\mathbf{D}\phi(u))(e_1 \wedge e_2 \wedge \cdots \wedge e_k)$, and $J(\mathbf{D}\phi(u)) = |\alpha(u)|$ and

$$R := \Big\{ u \in U \,\Big|\, J(\mathbf{D}\phi(u)) \neq 0 \Big\}.$$

Then $\xi(y) = \frac{\alpha(u)}{|\alpha(u)|}$, $u = \phi^{-1}(y) \in R$, is the induced orientation on $\phi(S)$. By applying the area formula to

$$f(u) := \begin{cases} < \omega(\phi(u)), \frac{\alpha(u)}{|\alpha(u)|} > & \text{if } u \in R, \\ 0 & \text{otherwise} \end{cases}$$

we find the following:

(i) If ω is $\mathcal{H}^k$-summable on $\phi(U)$, then $\omega(\phi(u))J(\mathbf{D}\phi(u))$ is $\mathcal{H}^k$-summable in U; hence $\phi^{\#}\omega$ is summable as

$$|\phi^{\#}(u)| = |\Lambda^k(\mathbf{D}\phi(u))\omega(\phi(u))| \leq J(\mathbf{D}\phi(u))\,|\omega(\phi(u))|.$$

(ii) Since

$$\begin{aligned} < \phi^{\#}\omega(u), e_1 \wedge e_2 \wedge \cdots \wedge e_k > &= < \Lambda^k(\mathbf{D}\phi(u))(\omega(\phi(u))), e_1 \wedge e_2 \wedge \cdots \wedge e_k > \\ &= < \omega(\phi(u)), \alpha(u) >, \end{aligned}$$

we deduce that $\phi^{\#}\omega(u) = 0$ if $u \notin R$ and

$$\begin{aligned} \int_U \phi^{\#}\omega &= \int_U < \phi^{\#}\omega(u), e_1 \wedge e_2 \wedge \cdots \wedge e_k > du = \int_R < \omega(\phi(u)), \alpha(u) > du \\ &= \int_R < \omega(\phi(u)), \frac{\alpha(u)}{|\alpha(u)|} > J(\mathbf{D}\phi(u))\, du \\ &= \int_{\phi(U)} < \omega(y), \xi(y) > d\mathcal{H}^k(y) = \int_{\phi(U)} \omega. \end{aligned}$$

□

Similarly, using the change of variable formula on submanifolds, we get the following theorem.

4.49 Theorem. *Let Ω be an open set of $\mathbb{R}^n$ and $\phi : \Omega \to \mathbb{R}^N$ be a C^1 map. Suppose that $S \subset\subset \Omega$ is such that $S := \mathcal{X} \cup \mathcal{N}$ where $\mathcal{H}^k(\mathcal{N}) = 0$ and $\mathcal{X}$ is an oriented k-submanifold of $\mathbb{R}^n$. Assume that $\phi(S)$ has the induced orientation. Then for every $\mathcal{H}^k$-summable k-form ω on $\phi(S)$, $\phi^{\#}\omega$ is $\mathcal{H}^k$-summable on S and*

$$\int_{\phi(\mathcal{S})} \omega = \int_{\mathcal{S}} \phi^{\#}\omega.$$

4.50 ¶. A differential k-form ω defined in $\mathbb{R}^n \setminus \{0\}$ is said to be *radial* if $\mathbf{R}^{\#}\omega = \omega$ for every orthogonal $\mathbf{R}$ with $\det \mathbf{R} = 1$. Show the following:

(i) A 1-form is radial if and only if $\omega(x) = f(r)\, dr$, $r = |x|$, i.e., if and only if ω is the pull-back of a 1-form on $\mathbb{R}_+$ by means of the map $x \to |x|$.

(ii) ω is radial if and only if $*\omega$ is radial.

(iii) A $(n-1)$-form is radial if and only if it has the form

$$\omega = f(|x|) \sum_{i=1}^{n} (-1)^{i-1} x^i \widehat{dx^i}. \tag{4.51}$$

[*Hint.* (i) Let $\omega := \sum_{i=1}^n \omega_i(x)\, dx^i$ in $\mathbb{R}^n \setminus \{0\}$. Notice that ω is radial if and only if for the field $\Omega := (\omega_i(x))$, we have $\Omega(\mathbf{R}x) = \mathbf{R}\Omega(x)$ $\forall x$ for all matrices $\mathbf{R}$ with $\det \mathbf{R} = 1$. Infer that $\Omega(x) \bullet x = f(|x|)$ and that the derivatives of Ω in the tangential directions to the unit sphere vanish, concluding that $\omega = f(r)\, dr$. Conversely, if $\omega = f(r)\, dr$, compute $\mathbf{R}^{\#}\omega$ using for instance Laplace's formulas for the determinant.]

4.51 Example (Volume form on $S^{n-1} \subset \mathbb{R}^n$). Let ω be the $(n-1)$-form in $\mathbb{R}^n \setminus \{0\}$ defined by

$$\omega(x) := \sum_{i=1}^{n} (-1)^{i-1} \frac{x^i}{|x|^n} \widehat{dx^i}.$$

Show that

(i) $d\omega = 0$ in $\mathbb{R}^n \setminus \{0\}$,
(ii) $\int_{S^{n-1}} \omega = \mathcal{H}^{n-1}(S^{n-1})$,
(iii) if $\pi(x) := x/|x|$ is the retraction on S^{n-1}, then $\pi^{\#}\omega = \omega$ in $\mathbb{R}^n \setminus \{0\}$.

[*Hint.* In fact, (i) follows by computing

$$d\omega = \sum_{i=1}^{n} D_i \Big(\frac{x^i}{|x|^n} \Big)\, dx^1 \wedge \cdots \wedge dx^n = 0.$$

Moreover, the $(n-1)$-vector that orients S^{n-1} at $x \in S^{n-1}$ is $\xi(x) := *x$. Therefore, see (4.46),

$$< \omega(x), \xi(x) > = < \sum_{i=1}^{n} x^i dx^i, \sum_{i=1}^{n} x^i e_i > = |x|^2 = 1,$$

hence

$$\int_{S^{n-1}} \omega = \int_{S^{n-1}} < \omega, \xi > \, d\mathcal{H}^{n-1} = \mathcal{H}^{n-1}(S^{n-1}),$$

i.e. (ii).

In order to prove (iii), notice that ω is radial, see Exercise 4.50. Then, since the retraction π on S^{n-1} commutes with the rotations, infer that $\pi^{\#}(*\omega)$ is also radial and, by Exercise 4.50, that

$$\pi^{\#}(*\omega) = \frac{f(r)}{r^n} \sum_{i=1}^{n} (-1)^{i-1} x^i \widehat{dx^i}.$$

Differentiating,

$$0 = \pi^{\#} d\omega = d(\pi^{\#}\omega) = \sum_{i=1}^{n} D_i \Big(\frac{f(r)}{r^n} x^i \Big)\, dx^1 \wedge \cdots \wedge dx^n = \frac{f'(r)}{r^{n-1}}\, dx^1 \wedge \cdots \wedge dx^n$$

from which $f'(r) = 0$ $\forall r$, i.e., $f(r) = C$ or, equivalently, $\pi^{\#}\omega = C\,\omega$. Finally, since π is the identity on S^{n-1}, conclude that $C = 1$.]

4.52 . We may compute the integral of a differential k-form on $\mathcal{X}$ by means of local coordinates. In order to do so, we choose a locally finite open cover $\{\Omega_i\}$, open sets $U_i \subset \mathbb{R}^k$ and diffeomorphisms $\varphi_i : U_i \to \Omega_i \cap \mathcal{X}$. Choose on each U_i the orientation ξ induced by $\mathcal{X} \cap \Omega_i$ and let $\{\alpha_i\}$ be a partition of unity relative to the covering $\{\Omega_i\}$. From Proposition 4.48 we infer that every $\mathcal{H}^k$-summable differential k-form on $\mathcal{X}$

$$\int_{\mathcal{X}} \omega = \sum_i \int_{\mathcal{X}} \alpha_i \omega = \sum_i \int_{\Omega_i \cap \mathcal{X}} \alpha_i \omega = \sum_i \int_{U_i} \varphi_i^{\#}(\alpha_i \omega). \tag{4.52}$$

4.3 Stokes's Theorem

4.3.1 The theorem

Stokes's theorem is the version of the fundamental theorem of calculus for differential forms.

4.53 Theorem (Stokes, I). *Let U be an admissible (in particular, U is bounded and $\mathcal{H}^{n-1}(\partial U) < \infty$) open set of $\mathbb{R}^k$ (thought as oriented by $\mathbb{R}^k$). and let ∂U be the boundary of U oriented by the exterior normal vector to U and the orientation of $\mathbb{R}^k$. Then*

$$\int_U d\omega = \int_{\partial U} \omega \tag{4.53}$$

for every $(k-1)$-form of class C^1 in an open neighborhood of $\overline{U}$. Moreover, if $\phi : \overline{U} \to \mathbb{R}^N$ is an injective map of class C^1 in a neighborhood of U, and the images $\phi(U)$ and $\phi(\partial U)$ take the orientations induced by the orientations of U and ∂U, respectively, then

$$\int_{\phi(U)} d\omega = \int_{\partial \phi(U)} \omega \tag{4.54}$$

for any $(k-1)$-form ω of class C^1 in a neighborhood of $\overline{\phi(U)}$.

Proof. Let $(e_1, e_2, \dots, e_k)$ be an orthonormal basis in $\mathbb{R}^k$ so that $\xi := e_1 \wedge e_2 \wedge \cdots \wedge e_k$ is the orientation of $\mathbb{R}^k$ and let $\nu = (\nu^1, \nu^2, \dots, \nu^k)$ be the exterior unit normal field to ∂U. Every $(n-1)$-form on $\overline{U}$ writes as $\omega = \sum_{i=1}^k (-1)^{i-1} \omega_i \widehat{dx^i}$ where $\omega_i \in C^1(\overline{U})$, hence, see (4.46),

$$\int_{\partial U} \omega = \int_{\partial U} \sum_{i=1}^k \omega_i \nu^i \, d\mathcal{H}^{n-1}.$$

On the other hand,

$$\begin{aligned}
\int_U d\omega &= \int_U < d\omega, e_1 \wedge e_2 \wedge \cdots \wedge e_k > \, d\mathcal{L}^k \\
&= \int_U \sum_{i=1}^n \frac{\partial \omega_i}{\partial x^i} < dx^1 \wedge \cdots \wedge dx^k, e_1 \wedge e_2 \wedge \cdots \wedge e_k > \, dx \\
&= \int_U \sum_{i=1}^n \frac{\partial \omega_i}{\partial x^i} \, dx.
\end{aligned}$$

Claim (4.53) then follows, or rather is equivalent to the Gauss-Green formulas.

Let us prove (4.54). If ϕ is of class C^2 in an open neighborhood of $\overline{\phi(U)}$, then $\phi^\# \omega$ is of class C^1 in an open neighborhood of U and $d(\phi^\# \omega) = \phi^\#(d\omega)$; (4.53) then yields

$$\int_U \phi^\# d\omega = \int_U d(\phi^\# \omega) = \int_{\partial U} \phi^\# \omega. \tag{4.55}$$

If ϕ is only of class C^1, we proceed by approximation. Let $\{\phi_\epsilon\}$, $\phi_\epsilon : \mathbb{R}^k \to \mathbb{R}^N$, be a family of mollifying of ϕ converging to ϕ in C^1 norm. Since the pull-back involves only the first derivatives of ϕ, we have

$$\phi_\epsilon^\# \omega \to \phi^\# \omega, \qquad \phi_\epsilon^\# d\omega \to \phi^\# d\omega$$

uniformly on $\overline{U}$. If we now write (4.55) for ϕ_ϵ and pass to the limit as $\epsilon \to 0$, we conclude that (4.55) holds for ϕ.

Finally, ω is $\mathcal{H}^k$-summable on $\phi(U)$ and $d\omega$ is $\mathcal{H}^{k-1}$-summable on $\phi(\partial U)$ since $\phi(U)$ and $\phi(\partial U)$ are bounded and of finite measure. Thus, Theorem 4.49 and (4.55) then yield

$$\int_{\phi(U)} d\omega = \int_U \phi^\#(d\omega) = \int_{\partial U} \phi^\# \omega = \int_{\partial \phi(U)} \omega.$$

□

4.54 Theorem (Stokes, II). *Let $\mathcal{X}$ be a compact and oriented k-submanifold of class C^1 in $\mathbb{R}^N$ with $\mathcal{H}^k(\mathcal{X}) < \infty$. For every $(k-1)$-form ω of class C^1 in an open neighborhood of $\mathcal{X}$ we have*

$$\int_{\mathcal{X}} d\omega = 0. \tag{4.56}$$

Moreover, if $\phi : \mathcal{X} \to \mathbb{R}^N$ is of class C^1, then for every $(n-1)$-form η of class C^1 in a neighborhood of $\phi(\mathcal{X})$, we have

$$\int_{\phi(\mathcal{X})} d\eta = 0.$$

Proof. Let A be an open bounded set containing $\mathcal{X}$ and on which ω is of class C^1. Let $\{\Omega_i\}$ be a finite covering of $\mathcal{X}$ with open connected sets $\Omega_i \subset A$, and for every i, let B_i be a ball in $\mathbb{R}^k$ and let $\varphi_i : B_i \subset \mathbb{R}^n \to \Omega_i \cap \mathcal{X}$ be a diffeomorphism. We orient each B^i in such a way that the induced orientation on $\mathcal{X} \cap \Omega_i$ is the orientation of $\mathcal{X}$. Let $\{\alpha_i\}$ be a partition of unity associated to $\{\Omega_i\}$. Then $\sum_i \alpha_i = 1$, hence $\sum_i d\alpha_i = d(\sum_i \alpha_i) = 0$ in $\mathcal{X}$. Consequently,

$$\sum_i \alpha_i d\omega = \sum_i d(\alpha_i \omega) \qquad \text{in } \mathcal{X}.$$

By integration

$$\int_{\mathcal{X}} d\omega = \int_{\mathcal{X}} \sum_i \alpha_i d\omega = \int_{\mathcal{X}} \sum_i d(\alpha_i \omega)$$

and, since the $\{\alpha_i\}$ are finitely many, from (4.54) we infer

$$\int_{\mathcal{X}} d\omega = \int_{\mathcal{X}} \sum_i d(\alpha_i \omega) = \sum_i \int_{\mathcal{X}} d(\alpha_i \omega) = \sum_i \int_{\varphi_i(B_i)} d(\alpha_i \omega) = \sum_i \int_{\partial \varphi_i(B_i)} \alpha_i \omega = 0,$$

each α_i vanishing near the boundary $\partial(\Omega_i \cap \mathcal{X}) = \partial \varphi_i(B_i)$.

If ϕ is of class C^2 in an open neighborhood of $\mathcal{X}$, then $\phi^\# \eta$ is of class C^1 in an open neighborhood of $\mathcal{X}$ and $\phi^\#(d\eta) = d(\phi^\# \eta)$. Consequently, from (4.56) we have

$$\int_{\mathcal{X}} \phi^\#(d\eta) = \int_{\mathcal{X}} d(\phi^\# \omega) = 0. \tag{4.57}$$

If ϕ is only of class C^1, we proceed by approximation as in the proof of Theorem 4.53 to get (4.57). The claim then follows from (4.57) and Theorem 4.49. □

4.3.2 Some applications

a. Piola's identities

Let $\mathbf{A} \in M_{n,n}(\mathbb{R})$ be a matrix; for $i, j = 1, \dots, n$ denote by $M(\mathbf{A})^{\overline{i}}_{\overline{j}}$ the determinant of the submatrix of $\mathbf{A}$ obtained by deleting row i and column j. We arrange them in a matrix, the *matrix of cofactors* $\operatorname{cof}(\mathbf{A})$ defined by

$$\operatorname{cof}(\mathbf{A})^i_j := (-1)^{i+j} M(\mathbf{A})^{\overline{j}}_{\overline{i}}$$

so that Laplace's formulas read as

$$\mathbf{A}\, \operatorname{cof}(\mathbf{A}) = \det \mathbf{A}\ \mathrm{Id}.$$

4.55 Proposition (Piola's identities). *Let Ω be an open set in $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}^n$ be a map of class $C^1(\Omega)$. Then $\forall i = 1, \dots, n$*

$$\sum_{j=1}^{n} D_j (\operatorname{cof} \mathbf{D}f)^j_i = 0 \qquad \text{in } \Omega.$$

Proof. Let $(e_1, e_2, \dots, e_n)$ be an orthonormal basis in $\mathbb{R}^n$. Reordering the components of $f = (f^1, \dots, f^n)$ we may assume $i = 1$. An integration by parts yields for all $\varphi \in C^\infty_c(\Omega)$,

$$\begin{aligned} -\int_\Omega \sum_{j=1}^{n} D_j (\operatorname{cof} \mathbf{D}f)^j_1 \varphi \, dx &= \int_\Omega \sum_{j=1}^{n} (\operatorname{cof} \mathbf{D}f)^j_1 D_j \varphi \, dx = \int_\Omega \sum_{j=1}^{n} \mathbf{A}^1_j \operatorname{cof}(\mathbf{A})^j_1 \, dx \\ &= \int_\Omega \det \mathbf{A} \, dx \\ &= \int_\Omega < \det \mathbf{A}\, dx^1 \wedge \cdots \wedge dx^n, e_1 \wedge e_2 \wedge \cdots \wedge e_n > dx \end{aligned}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{D}\varphi \\ \mathbf{D}f^2 \\ \cdots \\ \mathbf{D}f^n \end{pmatrix}.$$

On the other hand,

$$\det \mathbf{A}\, dx^1 \wedge \cdots \wedge dx^n = d\varphi \wedge df^2 \wedge \cdots \wedge df^n = d(\varphi \wedge df^2 \wedge \cdots \wedge df^n),$$

hence by Stokes's theorem

$$\int_\Omega < \det \mathbf{A}\, dx^1 \wedge \cdots \wedge dx^n, e_1 \wedge e_2 \wedge \cdots \wedge e_n > dx = \int_{\partial\Omega} \varphi \wedge df^2 \wedge \cdots \wedge df^n = 0.$$

Consequently,

$$\int_\Omega \sum_{j=1}^{n} D_j (\operatorname{cof} \mathbf{D}f)^j_1 \varphi \, dx = 0$$

and the claim follows since $\varphi \in C^\infty_c(\Omega)$ is arbitrary. □

b. Brouwer's fixed point theorem

We discussed Brouwer's fixed point theorem in [GM3]. Here we deduce it (in one of its equivalent forms) from Stokes's theorem.

4.56 Theorem. *There is no continuous map $f : B(0,1) \subset \mathbb{R}^n \to \partial B(0,1)$ such that $f(x) = x \ \forall x \in \partial B(0,1)$.*

Proof. Suppose to the contrary that there is such a map f. Smoothing the continuous function

$$\widehat{f}(x) := \begin{cases} (1-\epsilon) f\Big(\frac{x}{1-\epsilon}\Big) & \text{if } |x| \le 1-\epsilon, \\ x & \text{if } |x| > 1-\epsilon, \end{cases} \qquad \epsilon > 0,$$

we find then a function $g : B(0,2) \to \partial B(0,1)$ of class C^∞ such that $g(x) = x$ on $\partial B(0,1)$. The vectors $\frac{\partial g}{\partial x^1}(x), \dots, \frac{\partial g}{\partial x^n}(x)$ are linearly dependent since they belong to $\mathrm{Tan}_{g(x)}\, \partial B(0,1)$ that is $(n-1)$-dimensional. Consequently, $\Lambda^n \mathbf{D} g(x)) = 0$ and $g^\# \eta = 0$ for all n-forms η. Let $\omega(x) = \sum_{i=1}^n (-1)^{i-1} \frac{x^i}{|x|^n} \widehat{dx^i}$ be the volume $(n-1)$-form on $\partial B(0,1)$, From Example 4.51 and Stokes's theorem we then have

$$\begin{aligned} 0 \neq \mathcal{H}^{n-1}(\partial B(0,1)) &= \int_{\partial B(0,1)} \omega = \int_{\partial B(0,1)} g^\# \omega = \int_{B(0,1)} d(g^\# \omega) \\ &= \int_{B(0,1)} g^\# (d\omega) = \int_{B(0,1)} 0 \, dx = 0, \end{aligned}$$

a contradiction. □

c. Brouwer's degree

Let $\mathcal{X}$ and $\mathcal{Y}$ be two connected and oriented submanifolds of $\mathbb{R}^N$ and let $f : \mathcal{X} \to \mathcal{Y}$ be a map of class C^1. Let ξ and η be the orientations of $\mathcal{X}$ and $\mathcal{Y}$, respectively, let $\alpha(x) := \Lambda_n \mathbf{D} f(x)(\xi(x))$, and let $R := \{x \in \mathcal{X} \,|\, \alpha(x) \neq 0\}$ be the set of regular points of f. Then, as we saw in Section 4.2.3, for all $x \in R$

$$\zeta(x) := \frac{\alpha(x)}{|\alpha(x)|}$$

orients $\mathrm{Tan}_{f(x)}\, f(R)$. Since $\mathrm{Tan}_{f(x)}\, f(R) = \mathrm{Tan}_{f(x)}\, \mathcal{Y}$, we then infer

$$\zeta(x) = \epsilon(x)\, \eta(f(x)) \qquad \text{with } \epsilon(x) = \pm 1.$$

4.57 Definition. *With the previous notations, the* degree *of the map f at $y \in \mathcal{Y}$ is the integer*

$$\begin{aligned} \deg(f,y) :&= \#\Big\{x \in R \,\Big|\, f(x) = y,\ \alpha(x) = \eta(f(x))\Big\} \\ &\quad - \#\Big\{x \in R \,\Big|\, f(x) = y,\ \alpha(x) = -\eta(f(x))\Big\} \\ &= \sum_{\substack{x \in R \\ f(x) = y}} \epsilon(x). \end{aligned} \tag{4.58}$$

In other words, we count the inverse images x of y in R with sign $+1$ if $Df(x)$ preserves the orientation of $\mathcal{Y}$ and -1 if $Df(x)$ reverses the orientation of $\mathcal{Y}$.

It is readily seen that

(i) $\deg(f, y)$ is integer, zero outside $f(R)$,
(ii) $\deg(f, y) = +1$ on $f(R)$ if f is injective and the orientation of $f(R)$ induced by the orientation of $\mathcal{X}$ is that of $\mathcal{Y}$.

4.58 Proposition. *The function $y \mapsto \deg(f, y)$ is $\mathcal{H}^k$-measurable. Moreover, for every n-form ω on $\mathcal{Y}$ of class C^1, the map $y \mapsto \deg(f, y)\omega(y)$ is $\mathcal{H}^k$-summable on $\mathcal{Y}$ if and only if $x \mapsto < f^{\#}\omega(x), \xi(x) >$ is summable on $\mathcal{X}$ and*

$$\int_{\mathcal{X}} f^{\#}\omega = \int_{\mathcal{Y}} \deg(f, y)\, \omega(y). \tag{4.59}$$

Proof. Set

$$v(x) = \begin{cases} < \omega(f(x)), \zeta(x) > & \text{if } x \in R, \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$< f^{\#}\omega(x), \xi(x) > = < \omega(f(x)), \alpha(x) > = v(x)|\Lambda_n(\mathbf{D}f(x))(\xi(x))| \qquad \forall x \in R$$

and for all $y \in f(R)$

$$\begin{aligned} V(y) &= \sum_{\substack{f(x)=y \\ x \in R}} v(x) = \sum_{\substack{x \in R \\ f(x)=y}} < \omega(f(x)), \alpha(x) > \\ &= \Big\langle \omega(y), \sum_{\substack{x \in R \\ f(x)=y}} \epsilon(x)\eta(y) \Big\rangle = \deg(f, y) < \omega(y), \eta(y) > . \end{aligned}$$

The claim is then a trivial application of Theorem 4.45 since

$$\int_{\mathcal{X}} < f^{\#}\omega(x), \xi(x) > \, d\mathcal{H}^k(x) = \int_{\mathcal{X}} f^{\#}\omega,$$

and

$$\int_{\mathcal{Y}} V(y)\, d\mathcal{H}^k(y) = \int_{\mathcal{Y}} \deg(f, y)\, \omega(y).$$

□

4.59 Theorem (Brouwer's degree). *Let $\mathcal{X}$ and $\mathcal{Y}$ be two oriented, connected and compact n-submanifolds of $\mathbb{R}^N$. Let $f : \mathcal{X} \to \mathcal{Y}$ be of class C^1. Then $y \mapsto \deg(f, y)$ is constant $\mathcal{H}^n$-a.e. $y \in \mathcal{Y}$. In other words, there exists an integer $\deg(f) \in \mathbb{Z}$ such that for every summable n-form in $\mathcal{Y}$ we have*

$$\int_{\mathcal{X}} f^{\#}\omega = \deg(f) \int_{\mathcal{Y}} \omega. \tag{4.60}$$

Proof. It suffices to prove that $\deg(f,y)$ is locally constant for $\mathcal{H}^k$-a.e. $y \in \mathcal{Y}$. Since $\mathcal{Y}$ is connected and $\deg(f,y)$ is an integer, it follows that $\deg(f,y)$ is constant. Then (4.60) follows from (4.59).

Let $\varphi : U \to \Omega$ be a diffeomorphism from an open set $U \subset \mathbb{R}^n$ onto a local chart Ω of $\mathcal{Y}$ and choose in U an orthonormal basis so that the induced orientation on $\mathcal{Y}$ is the one of $\mathcal{Y}$. By choosing a continuous n-form in $\mathcal{Y}$ that is nonzero in Ω, the area formula yields that $\deg(f,y)$ is locally summable in Ω. If $g(z) := \deg(f, \varphi(z))$, $z \in U$, then g is locally summable in U.

For all $\phi \in C^2_c(U)$ and $1 \le i \le n$, we consider the $(n-1)$-form $\eta := \phi(z)(-1)^{i-1}\widehat{dz^i}$ and using (4.59) and Stokes's theorem, we compute

$$\begin{aligned}\int_U g(z) D_i\phi(z)\,dz &= \int_U g\,d\eta = \int_\Omega (\varphi^{-1})^\#(g\,d\eta) = \int_{\mathcal{Y}} \deg(f,y)(\varphi^{-1})^\#(d\eta)\\ &= \int_{\mathcal{X}} f^\#(\varphi^{-1})^\#(d\eta) = \int_{\mathcal{X}} (\varphi^{-1}\circ f)^\#(d\eta)\\ &= \int_{\mathcal{X}} d((\varphi^{-1}\circ f)^\#\eta) = 0.\end{aligned}$$

From the DuBois-Reymond lemma, Lemma 1.52, we then infer that $g(y)$ is constant in U, hence $\deg(f,y)$ is constant in Ω. □

The integer $\deg(f)$ defined by (4.58) and satisfying (4.60) is called the *degree* of the map $f : \mathcal{X} \to \mathcal{Y}$. Of course, it depends on $\mathcal{X}$ and $\mathcal{Y}$ and its sign depends on the chosen orientations of $\mathcal{X}$ and $\mathcal{Y}$.

We list some trivial consequences of (4.60) keeping the previous notations.

(i) If $\deg(f) \neq 0$, then the set of regular values of f has zero measure in $\mathcal{Y}$, i.e., $\mathcal{H}^n(\mathcal{Y} \setminus f(R)) = 0$.
(ii) If $\deg(f) \neq 0$ and $R = \mathcal{X}$, then for all $y \in \mathcal{Y}$ the equation $f(x) = y$ has at least a solution $x \in \mathcal{X}$.

Moreover, see Proposition 4.79, *the degree of two homotopic maps is the same.*

When $\mathcal{X} = \mathcal{Y} = \partial B(0,1) \subset \mathbb{R}^n$, the degree of f is simply Brouwer's degree, compare [GM3], for which we have, for instance, the following:

(i) Two maps from a sphere into itself are homotopic if and only if they have the same degree.
(ii) $f : S^n \to S^n$ has a continuous extension to all of $\overline{B(0,1)} \subset \mathbb{R}^{n+1}$ if and only if $\deg(f) = 0$.

d. Gauss-Bonnet's theorem

Let $\mathcal{X}$ be a 2-dimensional submanifold in $\mathbb{R}^3$ and $\nu : \mathcal{X} \to \mathbb{R}^3$ denote the field of unit normal vectors that orients $\mathcal{X}$. Since $|\nu| = 1$, ν takes values in $S^2 = \{y \,|\, |y| = 1\}$ and the map $\nu : \mathcal{X} \to S^2$ that orients $\mathcal{X}$ takes the name of *Gauss map*. Since $|\nu(x)| = 1$ $\forall x$, for every tangent direction $a \in \mathrm{Tan}_x\,\mathcal{X}$ we find

$$\sum_{i=1}^{3} \frac{\partial \nu}{\partial a}(x) \bullet \nu(x) \;= 0 \qquad \forall x \in \mathcal{X},$$

i.e., $\frac{\partial \nu}{\partial a}(x) \in \operatorname{Tan}_{\nu(x)} S^2$. Since $+\nu(x) \in \Lambda_2 \operatorname{Tan}_{\nu(x)} S^2$,

$$\Lambda_2(D\nu(x))(*\nu(x)) = k(x) * \nu(x).$$

The proportionality factor $k(x)$ is called the *Gaussian curvature of* $\mathcal{X}$ *at* x.

If ω denotes the volume 2-form of S^2,we have $< \omega(\nu(x)), *\nu(x) >= 1$, hence

$$\begin{aligned} < \nu^{\#}(\omega)(x), *\nu(x) > &=< \omega(\nu(x)), \Lambda_2(D\nu(x))(*\nu(x)) > \\ &=< \omega(\nu(x)), k(x) * \nu(x) > \\ &= k(x). \end{aligned}$$

From the constancy of the degree, we then infer the following theorem.

4.60 Theorem (Gauss–Bonnet). *The integral of the Gaussian curvature is an integer multiple of* 4π*, more precisely,* 4π *times the degree of the Gauss map,*

$$\frac{1}{4\pi} \int_{\mathcal{X}} k(x)\, d\mathcal{H}^2(x) = \deg(\nu).$$

Proof. In fact, from the above,

$$\begin{aligned} \int_{\mathcal{X}} k(x)\, d\mathcal{H}^2(x) = \int_{\mathcal{X}} < \nu^{\#}(\omega)(x), *\nu(x) >\, d\mathcal{H}^2(x) &= \int_{\mathcal{X}} \nu^{\#}\omega \\ = \deg(\nu) \int_{S^2} \omega &= 4\pi \deg(\nu). \end{aligned}$$

□

e. Linking number

Let $\mathcal{X}$ and $\mathcal{Y}$ be two boundaryless, compact and nonintersecting oriented submanifolds in $\mathbb{R}^n$ of dimension k and $n-k-1$, respectively (for instance, two regular, closed curves without intersections in $\mathbb{R}^3$). Consider the product submanifold $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^{2n}$ oriented by $\xi(x) \wedge \eta(y)$, where ξ and η are the fields of k-vectors and $(n-k-1)$-vectors that orients $\mathcal{X}$ and $\mathcal{Y}$, respectively, and the map $f : \mathcal{X} \times \mathcal{Y} \to S^{n-1}$ given by

$$f(x, y) := \frac{y - x}{|y - x|}.$$

The map f is smooth in a neighborhood of $\mathcal{X} \times \mathcal{Y}$ and its degree is called the *linking number* of $\mathcal{X}$ and $\mathcal{Y}$ and is denoted by

$$\operatorname{link}(\mathcal{X}, \mathcal{Y}) := \deg(f).$$

It follows from (4.60) that

$$\operatorname{link}(\mathcal{X}, \mathcal{Y}) = \frac{1}{\mathcal{H}^{n-1}(S^{n-1})} \int_{\mathcal{X} \times \mathcal{Y}} f^{\#}(\omega)$$

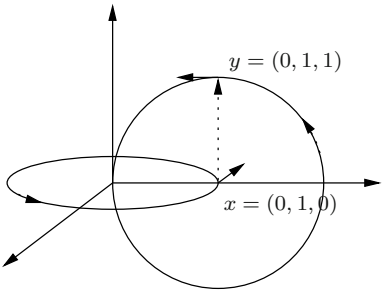


Figure 4.4. Linking of two curves in $\mathbb{R}^3$.

where ω is the volume $(n-1)$-form of S^{n-1}. Notice that the pointwise definition of the degree in (4.58) yields an explicit formula for computing the linking number.

4.61 Example. Compute the linking number of the two curves in $\mathbb{R}^3$:

$$\gamma(t) := (\cos t, \sin t, 0),\ t \in [0, 2\pi], \qquad \delta(s) := (0, 1 + \cos s, \sin s),\ s \in [0, 2\pi].$$

Let $\mathcal{X}$ and $\mathcal{Y}$ be the trajectories oriented by the direction of movement. We need to compute the degree of the map $f : \mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^3 \times \mathbb{R}^3 \to S^2 \subset \mathbb{R}^3$ given by $f(x, y) := \frac{y-x}{|y-x|}$. In order to do it, we use (4.58).

Consider the point $(0, 0, 1)^T \in S^2 \subset \mathbb{R}^3$ whose unit normal vector that yields the standard orientation of S^2 is $(0, 0, 1)^T$.

Step 1. First, we look for points $(x, y) \in \mathbb{R}^3 \times \mathbb{R}^3$ such that

$$x \in \mathcal{X}, \qquad y \in \mathcal{Y} \qquad \text{and} \qquad \frac{y-x}{|y-x|} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

We need to find $t, s \in [0, 2\pi[$ such that $\delta(s) - \gamma(t) = (0, 0, \lambda)^T$ with $\lambda > 0$, i.e.,

$$\begin{pmatrix} -\cos t \\ 1 + \cos s - \sin t \\ \sin s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \lambda \end{pmatrix}, \qquad \lambda > 0.$$

This system of equations has solutions if and only if $t = \pi/2$ and $s = \pi/2$, hence the couple

$$\overline{x} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \overline{y} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

is the unique solution of $f(x, y) = (0, 0, 1)^T$.

Step 2. The unit tangent vector to $\gamma(t)$ at $\overline{x} = (0, 1, 0)^T$ is $(-1, 0, 0)^T = -e_{x^1}$ and the unit tangent vector to $\delta(s)$ at $\overline{y} = (0, 1, 1)^T$ is $(0, -1, 0)^T = -e_{y^2}$. Therefore, the unit tangent 2-vector to $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^3 \times \mathbb{R}^3$ is $e_{x^1} \wedge e_{y^2}$. The transformation matrix of $g(x, y) := y - x$ is the 3×6 matrix

$$\mathbf{D}g(x, y) = \left(\boxed{-\operatorname{Id}} \quad \boxed{+\operatorname{Id}} \right),$$

and, since $\overline{y} - \overline{x} = (0,0,1)^T$,

$$\mathbf{D}f(\overline{x},\overline{y}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \left(\boxed{-\mathrm{Id}} \quad \boxed{+\mathrm{Id}} \right) = \begin{pmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since $\Lambda_2\mathbf{D}f(e_{x^1} \wedge e_{y^2})$ is the exterior product of columns 1 and 5 of $\mathbf{D}f$, we find

$$\Lambda_2\mathbf{D}f(e_{x^1} \wedge e_{y^2}) = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \wedge \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = - * \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Hence $\Lambda_2\mathbf{D}f(e_{x^1} \wedge e_{y^2}) \neq 0$, i.e., $(0,1,1)^T$ is a regular value for f and f reverses the orientation of the sphere S^2. In conclusion,

$$\mathrm{link}(\mathcal{X},\mathcal{Y}) = \deg(f) = \deg(f,(0,0,1)^T) = -1.$$

Notice that the result is in accord with the intuition: We choose a smooth surface S with $\partial S = \mathcal{Y}$ oriented in such a way that ∂S and $\mathcal{Y}$ have the same orientation and such that S intersects $\mathcal{X}$ transversally. If n_S is the unit normal to S, we count the intersections with $\mathcal{X}$ positively if $\delta' \bullet n_s > 0$ and negatively if $\delta' \bullet n_s < 0$, δ representing $\mathcal{X}$. The sum is the linking number of $\mathcal{X}$ and $\mathcal{Y}$.

Another example is given in Example 4.70.

4.4 Vector Calculus

In this section we develop some calculus for forms, in particular, we see that the classical differential operators div and rot are suitable combinations of Hodge's operators $*$ and of the exterior differentiation operator d.

4.4.1 Codifferential

Consider $\mathbb{R}^n$ oriented by the standard basis $(e_1, e_2, \ldots, e_n)$ and endowed with the standard inner product $x \bullet y := \sum_{i=1}^m x^i y^i$. Let us denote by $\mathcal{E}^k(U)$ the space of k-forms with smooth coefficients on U.

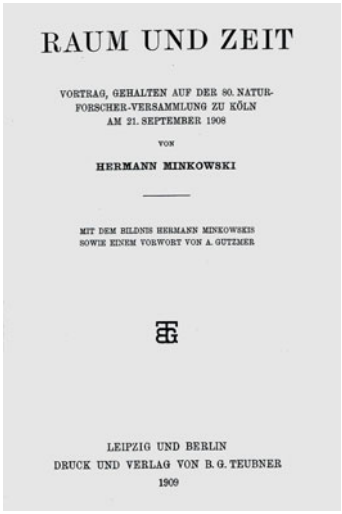
In addition to the operator of *exterior differentiation*,

$$\omega \in \mathcal{E}^k(U) \;\rightarrow\; d\omega \in \mathcal{E}^{k+1}(U),$$

we introduce the *operator of codifferentiation*, or *codifferential*, $\delta : \mathcal{E}^k(U) \rightarrow \mathcal{E}^{k-1}(U)$ by

$$\delta := (-1)^{n(k+1)} * d * \omega.$$

Notice that $\delta\omega = 0$ if ω is a 0-form and that δ does not depend on the orientation because in its composition the $*$ operator appears twice.


RAUM UND ZEIT

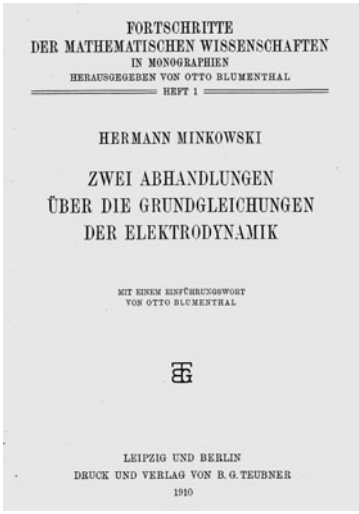
VORTRAG, GEHALTEN AUF DER 80. NATURFORSCHER-VERSAMMLUNG ZU KÖLN AM 21. SEPTEMBER 1908

VON

HERMANN MINKOWSKI

MIT DEM BILDNIS HERMANN MINKOWSKIS SOWIE EINEM VORWORT VON A. GUTZMER

LEIPZIG UND BERLIN
DRUCK UND VERLAG VON B. G. TEUBNER
1909


FORTSCHRITTE
DER MATHEMATISCHEN WISSENSCHAFTEN
IN MONOGRAPHIEN
HERAUSGEGEBEN VON OTTO BLUMENTHAL
HEFT 1

HERMANN MINKOWSKI

ZWEI ABHANDLUNGEN
ÜBER DIE GRUNDGLEICHUNGEN
DER ELEKTRODYNAMIK

MIT EINEM EINFÜHRUNGSWORT
VON OTTO BLUMENTHAL

LEIPZIG UND BERLIN
DRUCK UND VERLAG VON B. G. TEUBNER
1910

Figure 4.5. Frontispieces of two monographs by Hermann Minkowski (1864–1909).

4.62 Example. Let $\omega := \sum_{i=1}^n \omega_i\, dx^i$ be a 1-form. Then $*\omega = \sum_{i=1}^n (-1)^{i-1}\omega_i\, \widehat{dx^i}$ and

$$d(*\omega) = \Big(\sum_{i=1}^n D_i\omega_i\Big)\, dx^1 \wedge \ldots dx^n,$$

$$\delta\omega = (-1)^{2n} * d * \omega = \sum_{i=1}^n D_i\omega_i.$$

The operator δ is called the codifferential operator since d and δ are adjoint to one another. In fact, the following holds.

4.63 Proposition. *Let U be an admissible set in $\mathbb{R}^n$ and let ω and η be a $(k-1)$-form and a k-form in U, respectively, with coefficients of class C^1 in a neighborhood of U. Then*

$$\int_U d\omega \bullet \eta\; dx + \int_U \omega \bullet \delta\eta\; dx = \int_{\partial U} \omega \wedge (*\eta).$$

Proof. In fact, compare the formulas in (4.36), we compute

$$\begin{aligned} d(\omega \wedge *\eta) &= d\omega \wedge *\eta + (-1)^{k-1}\omega \wedge d * \eta \\ &= d\omega \wedge *\eta + (-1)^{k-1}(-1)^{(n-k+1)(k-1)}\omega \wedge * * d * \eta \\ &= d\omega \wedge *\eta + (-1)^{n(k+1)}\omega \wedge *\delta\eta \\ &= (\, d\omega \bullet \eta\,)\, dx^1 \wedge \ldots dx^n + (\, \omega \bullet \delta\eta\,)\, dx^1 \wedge \ldots dx^n. \end{aligned}$$

Integrating on U and applying Stokes's theorem, this yields the result at once. □

Of course, if the coefficients of ω and η have compact support in U, the boundary term vanishes. Let us discuss more closely the vanishing of

the boundary term. Let $\nu(x)$ denote the unit exterior normal vector to U at x. Every k-vector decomposes uniquely as

$$\xi = \xi_1 + \xi_2 \wedge \nu(x), \qquad \xi_1 \in \Lambda_k \mathbb{R}^n,\ \xi_2 \in \Lambda_{k-1}\mathbb{R}^n.$$

Correspondingly, every k-form ω in $\overline{U}$ determines two functions $\mathfrak{t}\omega$ and $\mathfrak{n}\omega$ from ∂U into $\Lambda_k \mathbb{R}^n$ by

$$< \mathfrak{t}\omega, \xi > := < \omega, \xi_1 >, \qquad < \mathfrak{n}\omega, \xi > := < \omega, \xi_2 \wedge \nu >,$$

called the *tangential* and the *normal part* of ω on ∂U, respectively. It is easy to see that

(i) $\omega = \mathfrak{t}\omega + \mathfrak{n}\omega$ on ∂U,
(ii) the integral on ∂U of a differential form depends exclusively on its tangential part

$$\int_{\partial U} \omega = \int_{\partial U} \mathfrak{t}\omega, \qquad \int_{\partial U} \mathfrak{n}\omega = 0, \tag{4.61}$$

(iii) $d(\mathfrak{t}\omega) = \mathfrak{t}(d\omega)$,
(iv) $*\mathfrak{t}\omega = \mathfrak{n}(*\omega)$,
(v) $*\mathfrak{n}\omega = \mathfrak{t}(*\omega)$,
(vi) $\delta(\mathfrak{n}\omega) = \mathfrak{n}(\delta\omega)$,
(vii) $\mathfrak{t}(\omega \wedge \eta) = \mathfrak{t}(\omega) \wedge \mathfrak{t}(\eta)$,
(viii) $|\omega|^2 = |\mathfrak{t}\omega|^2 + |\mathfrak{n}\omega|^2$ su ∂U.

In particular,

$$\omega \wedge (*\eta) = \mathfrak{t}(\omega \wedge (*\eta)) = \mathfrak{t}(\omega) \wedge \mathfrak{t}(*\eta) = \mathfrak{t}(\omega) \wedge *(\mathfrak{n}\,\eta).$$

As a consequence of Proposition 4.63 the following holds.

4.64 Corollary. *Let ω and η be a $(k-1)$-form and a k-form of class C^1, respectively, such that either $\mathfrak{t}\omega(x) = 0$ or $\mathfrak{n}\,\eta(x) = 0$ at every point $x \in \partial U$. Then*

$$\int_U d\omega \bullet \eta + \int_U \omega \bullet \delta\eta = 0.$$

4.4.2 Laplace's operator on forms

By means of the first order operators d and δ we define *Laplace's operator* for every k-form ω of class C^2 as

$$\Delta\omega := (d\delta + \delta d)\omega.$$

Notice that for 0-forms, $\omega = f \in C^2(U)$, since $\delta f = 0$ we have

$$\Delta f = \delta(df) = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial {x^i}^2},$$

i.e., the ordinary Laplace's operator for functions. Proposition 4.63 yields

$$\int_U (-\Delta\omega \bullet \omega)\, dx = \int_U (|d\omega|^2 + |\delta\omega|^2)\, dx + \int_{\partial U} (\delta\omega \wedge *\omega + \omega \wedge *d\omega) \quad (4.62)$$

for every k-form of class C^2 and, when one of the following three conditions holds:

- $d\omega = 0$ and $\delta\omega = 0$ on ∂U,
- $\mathfrak{t}\,\omega = 0$ on ∂U,
- $\mathfrak{n}\,\omega = 0$ on ∂U,

we have

$$\int_{\partial U} (\delta\omega \wedge *\omega + \omega \wedge d\omega) = 0.$$

This is trivial when the first condition holds. When the second or the third condition holds, it suffices to note that

$$\begin{aligned}\mathfrak{t}\,(\delta\omega \wedge *\omega + \omega \wedge *d\omega) &= \mathfrak{t}\,(\omega \wedge *\delta\omega + d\omega \wedge *\omega)\\ &= \mathfrak{t}\,(\omega) \wedge *\delta(\mathfrak{n}\,\omega) + d(\mathfrak{t}\,\omega) \wedge (\mathfrak{n}\,\omega).\end{aligned}$$

Under one of the previous boundary conditions, (4.62) then yields

$$\int_U (-\Delta\omega \bullet \omega)\, dx = \int_U (|d\omega|^2 + |\delta\omega|^2)\, dx. \quad (4.63)$$

4.65 Definition. *We say that a k-form is a solution of the* self-dual *equations in U if it solves the first order differential system*

$$\begin{cases} d\omega = 0, \\ \delta\omega = 0 \end{cases} \qquad \text{in } U.$$

We say that ω is harmonic *if $\Delta\omega = 0$ in U.*

By applying (4.62) to domains $V \subset\subset U$, one shows that ω *is harmonic if and only if ω is of class C^2 and ω solves the self-dual equations $d\omega = \delta\omega = 0$.*

By means of a computation that we omit, one also shows that if $\omega := \sum_{\alpha\in I(k,n)} \omega_\alpha dx^\alpha$, then

$$\Delta\omega = \sum_{\alpha\in I(k,n)} \Delta\omega_\alpha\, dx^\alpha.$$

It follows that

$$\begin{aligned}\int_U (-\Delta\omega \bullet \omega)\, dx &= -\int_U \sum_\alpha \omega_\alpha \Delta\omega_\alpha\, dx\\ &= \int_U \sum_\alpha |\nabla\omega_\alpha|^2\, dx - \int_{\partial U} \sum_{i=1}^n \sum_\alpha \omega_\alpha \frac{\partial\omega_\alpha}{\partial\nu}\, d\mathcal{H}^{n-1}.\end{aligned}$$

A comparison with (4.62) then shows that

$$\int_U (|d\omega|^2 + |\delta\omega|^2)\, dx \qquad \text{and} \qquad \int_U \sum_\alpha |\nabla \omega_\alpha|^2\, dx$$

differ by a boundary integral and, more precisely,

$$\int_U (|d\omega|^2 + |\delta\omega|^2)\, dx = \int_U \sum_\alpha |\nabla \omega_\alpha|^2\, dx - \int_{\partial U} \Big(\delta\omega \wedge *\omega + \omega \wedge *d\omega + \sum_{i=1}^{n} (-1)^{i-1} \frac{1}{2} \frac{\partial |\omega|^2}{\partial x^i}\, \widehat{dx^i} \Big).$$

4.4.3 Vector calculus in two dimensions

Consider $\mathbb{R}^2$ with its inner product and oriented by its standard basis (e_x, e_y), where $e_x := (1,0)$, $e_y := (0,1)$. We recall that for a plane field $E = (P, Q)$ we define

$$\begin{aligned} \operatorname{div} E = \nabla \bullet E &:= \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}, \\ \operatorname{rot} E = \nabla \times E &:= \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}. \end{aligned}$$

Hodge's operator $*$ acts on the dual basis (dx, dy) of (e_x, e_y) as

$$*dx = dy, \qquad *dy = -dx.$$

Finally, the isomorphism between vectors and linear maps is represented in our coordinate system as the identity, i.e., we may identify the vector

$$E(x,y) = P(x,y)\, e_x + Q(x,y)\, e_y$$

with components $P(x,y)$ and $Q(x,y)$ with the 1-form

$$\omega := P(x,y)\, dx + Q(x,y)\, dy$$

so that $< \omega, v >= E \bullet v \ \ \forall v \in \mathbb{R}^2$. We then have

$$\begin{cases} \omega = P(x,y)\, dx + Q(x,y)\, dy, \\ *\omega = P(x,y)\, dy - Q(x,y)\, dx \end{cases}$$

and

$$\begin{cases} \delta\omega = *d*\omega = \dfrac{\partial P}{\partial x} + \dfrac{\partial Q}{\partial y} = \operatorname{div} E, \\ d\omega = \Big(-\dfrac{\partial P}{\partial y} + \dfrac{\partial Q}{\partial x} \Big) dx \wedge dy = \operatorname{rot} E\, dx \wedge dy. \end{cases}$$

Notice that the self-dual equations

$$\begin{cases} \delta\omega = 0, \\ d\omega = 0, \end{cases} \qquad \text{equivalently} \qquad \begin{cases} \operatorname{div} E = 0, \\ \operatorname{rot} E = 0, \end{cases}$$

are the Cauchy–Riemann equations for the complex function

$$f(z) := P(x,y) - i\,Q(x,y), \qquad z = x + iy,$$

and that ω and $*\omega$ are the real and imaginary part, respectively, of the differential $f(z)\,dz$, i.e., $f(z)\,dz = \omega + i(*\omega)$. In conclusion, we may state that the following facts are equivalent:

(i) f is holomorphic.
(ii) $f(z)\,dz$ is a holomorphic differential.
(iii) the real part $\omega := \Re(f(z)\,dz)$ of $f(z)\,dz$ is a solution of the self-dual equations.
(iv) the imaginary part $*\omega := \Im(f(z)\,dz)$ of $f(z)\,dz$ is a solution of the self-dual equations.
(v) $\omega := \Re(f(z)\,dz)$ is a harmonic form.
(vi) $*\omega := \Im(f(z)\,dz)$ is a harmonic form.
(vii) P and $-Q$ are two conjugate harmonic functions.

4.4.4 Vector calculus in three dimensions

Consider $\mathbb{R}^3$ with its standard inner product oriented by its standard basis (e_x, e_y, e_z), where $e_x := (1,0,0)$, $e_y := (0,1,0)$ and $e_z = (0,0,1)$. Recall that for a field $E = (E^1, E^2, E^3)$, we define the divergence and the curl of the field E, respectively, by

$$\operatorname{div} E = \nabla \bullet E := \frac{\partial E^1}{\partial x} + \frac{\partial E^2}{\partial y} + \frac{\partial E^3}{\partial z},$$

and

$$\begin{aligned} \operatorname{rot} E = \nabla \times E : &= \begin{pmatrix} e_x & ey & ez \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E^1 & E^2 & E^3 \end{pmatrix} \\ &= (E^3_y - E^2_z)\,e_x + (E^3_x - E^1_z)\,e_y + (E^2_y - E^1_x)\,e_z. \end{aligned}$$

The Hodge $*$ operator acts on the dual basis (dx, dy, dz) of (e_x, e_y, e_z) as

$$\begin{aligned} &*\,dx = dy \wedge dz, && *\,dy = -dx \wedge dz, && *\,dz = dx \wedge dy, \\ &*\,dy \wedge dz = dx, && *\,dx \wedge dz = -dy, && *\,dx \wedge dy = dz \end{aligned}$$

(i) Some formulas from vector calculus

$$\begin{aligned}
&\nabla(f+g) = \nabla f + \nabla g,\\
&\nabla(fg) = f\nabla g + g\nabla f,\\
&\operatorname{div}(F+G) = \operatorname{div} F + \operatorname{div} G,\\
&\operatorname{div}(fF) = f\operatorname{div} F + \nabla f \bullet F,\\
&\operatorname{rot}(F+G) = \operatorname{rot} F + \operatorname{rot} G,\\
&\operatorname{rot}(fF) = f\operatorname{rot} F + \nabla f \times F,\\
&\nabla(F\times G) = G\bullet \operatorname{rot} F - F\bullet \operatorname{rot} G,\\
&\nabla(F\bullet G) = (F\bullet\nabla)G + (G\bullet\nabla)F + F\times\operatorname{rot} G + G\times\operatorname{rot} F,\\
&\operatorname{rot}(F\times G) = F\operatorname{div} G - G\operatorname{div} F + (G\bullet\nabla)F - (F\bullet\nabla)G,\\
&\operatorname{rot}(\nabla f) = 0,\\
&\operatorname{div}\operatorname{rot} F = 0,\\
&\Delta f = \operatorname{div}\nabla f,\\
&\Delta F = \nabla\operatorname{div} F - \operatorname{rot}(\operatorname{rot} F),\\
&\nabla(|F|^2) = 2(F\bullet\nabla)F + 2F\times\operatorname{rot} F,\\
&\operatorname{div}(\nabla f\times\nabla g) = 0,\\
&\Delta(fg) = f\Delta g + g\Delta f + 2\,\nabla f\bullet\nabla g,\\
&H\bullet F\times G = G\bullet H\times F = F\bullet G\times H,\\
&H\bullet((F\times\nabla)\times G) = ((H\bullet\nabla)G)\bullet F - (H\bullet\nabla)(\nabla\bullet G),\\
&F\times(G\times H) = (F\bullet H)G - (F\bullet G)H.
\end{aligned}$$

(ii) Let $E:\Omega\subset\mathbb{R}^3\to\mathbb{R}^3$ be a field of class C^1 and Ω be simply connected. Then $E=\nabla f$ if $\operatorname{rot} E = 0$.

(iii) Let $E:\Omega\subset\mathbb{R}^3\to\mathbb{R}^3$ be a field of class C^1 and Ω be è contractible. Then $E=\operatorname{rot} H$ if $\operatorname{div} E = 0$.

Figure 4.6. Some useful formulas from vector calculus in $\mathbb{R}^3$: f and g are functions and E, F, G and H are fields.

and

$$** = (-1)^{k(3-k)} = +1 \qquad \forall k = 0,1,2,3.$$

Since the isomorphism between vectors and covectors in $\mathbb{R}^3$ in our coordinates is the identity, we may identify the field

$$E = E^1(x,y,z)\,e_x + E^2(x,y,z)\,e_y + E^3(x,y,z)\,e_z$$

of components (E^1, E^2, E^3) with the 1-form

$$\omega := E^1(x,y,z)\,dx + E^2(x,y,z)\,dy + E^3(x,y,z)\,dz.$$

Now,

$$\begin{aligned}
\omega &= E^1\,dx + E^2\,dy + E^3\,dz,\\
*\,\omega &= E^1\,dy\wedge dz - E^2\,dx\wedge dz + E^3\,dx\wedge dy,
\end{aligned} \tag{4.64}$$

thus

$$d(*\omega) = \Big(\frac{\partial E^1}{\partial x} + \frac{\partial E^2}{\partial y} + \frac{\partial E^3}{\partial z}\Big)\, dx \wedge dy \wedge dz = \operatorname{div} E\, dx \wedge dy \wedge dz,$$

$$\delta\omega = *d * \omega = \sum_{i=1}^{3} \frac{\partial E^i}{\partial x^i} = \operatorname{div} E,$$

$$d\omega = \Big(E^3_y - E^2_z\Big)\, dy \wedge dz + \Big(E^3_x - E^1_z\Big)\, dx \wedge dz + \Big(E^2_y - E^1_x\Big)\, dx \wedge dy,$$

$$* d\omega = \Big(E^3_y - E^2_z\Big)\, dx - \Big(E^3_x - E^1_z\Big)\, dy + \Big(E^2_y - E^1_x\Big)\, dz = \operatorname{rot} E.$$

Moreover, the components E^1, E^2 and E^3 of E are harmonic functions if and only if $\Delta\omega = 0$, or, equivalently,

$$\begin{cases} d\omega = 0, \\ \delta\omega = 0, \end{cases} \qquad \text{or also} \qquad \begin{cases} \operatorname{div} E = 0, \\ \operatorname{rot} E = 0. \end{cases}$$

4.66 ¶. Let a, b and c be three vectors in $\mathbb{R}^3$. Show that $*(a \times b) \bullet c = a \wedge b \wedge c$. The scalar $(a \times b) \bullet c$ is called the *triple product of* a, b *and* c, and its modulus is the volume of the parallelepiped of sides a, b and c. One sets $vol(a, b, c) := (a \times b) \bullet c$.

4.67 ¶. Prove some of the formulas in Figure 4.6.

a. Stokes's theorem in $\mathbb{R}^3$

Let U be an admissible open set in $\mathbb{R}^2$ and Ω an open set in $\mathbb{R}^3$, let $\phi : \overline{U} \to \Omega$ be a map of class C^1 in an open neighborhood of $\overline{U}$, injective on $\overline{U}$ with Jacobian matrix of maximal rank, and let $S = \phi(U)$. Consider in Ω a field $E = (E_1, E_2, E_3) : \Omega \to \mathbb{R}^3$ of class $C^1(\Omega)$ and the associated 1-form $\omega := E_1\, dx^1 + E_2\, dx^2 + E_3\, dx^3$. If $\gamma : [0, 1] \to \mathbb{R}^2$ is a simple closed curve that travels ∂U anticlockwise, i.e., is injective in $[0, 1[$ and its tangent vector $\mathbf{t}(x)$ orients $\operatorname{Tan}_x \partial U$, then

$$\begin{aligned} \int_{\phi(\partial U)} \omega &= \int_0^1 < \gamma^{\#}\phi^{\#}\omega, (1, 0) > \, dt = \int_0^1 < \phi^{\#}\omega(\gamma(t)), \gamma'(t) > d\theta \\ &= \mathcal{L}(\phi^{\#}\omega, \gamma) = \mathcal{L}(\omega, \phi \circ \gamma), \end{aligned}$$

i.e., $\int_{\phi(\partial U)} \omega$ is the *work* done by ω along the curve $\phi \circ \gamma$ that travels ∂S.

On the other hand, since ϕ has maximal rank

$$\Lambda_2 \mathbf{D}\phi(u)(e_1 \wedge e_2) \neq 0 \qquad \forall x \in U,$$

the 2-vector

$$\xi(x) := \frac{\Lambda_2 \mathbf{D}\phi(u)(e_1 \wedge e_2)}{|\Lambda_2 \mathbf{D}\phi(u)(e_1 \wedge e_2)|}$$

orients $S = \phi(U)$ with the orientation induced by U via ϕ. If

$$\nu_S(x) := * \xi(x) = \frac{\phi_{u^1} \times \phi_{u^2}}{|\phi_{u^1} \times \phi_{u^2}|}, \qquad \phi(u) = x,$$

then $*\nu_S(x) = **\xi(x) = \xi(x)$ and, compare (4.36),

$$< d\omega(x), \xi(x) >=< *d\omega(x), *\xi(x) >= \operatorname{rot} E(x) \bullet \nu_S(x)\,.$$

Stokes's theorem,

$$\int_{\phi(\partial U)} \omega = \int_{\phi(\partial U)} \omega = \int_{\phi(U)} d\omega = \int_S < d\omega, \xi > \, d\mathcal{H}^k,$$

then reads as *circulation* or *rotation* or *curl theorem*.

$$\mathcal{L}(E, \phi \circ \gamma) = \int_S \operatorname{rot} E \bullet \nu_S \; d\mathcal{H}^2. \tag{4.65}$$

4.68 The curl as circulation of a field. Let E be a vector field of class $C^1(\Omega)$, $x_0 \in \Omega$, Ω being open, and let ν be a unit vector. Consider the plane perpendicular to ν through x_0 and oriented by ν; for every $\epsilon > 0$ denote by $\delta_\epsilon : [0,1] \to P$ the circle $B(x_0, \epsilon) \cap P$ oriented by $\partial B(z_0, \epsilon)$, i.e, so that

$$\det[(x - x_0) \,|\, t(x) \,|\, \nu] > 0.$$

Then Stokes's theorem yields

$$\int_{B(x_0,\epsilon)\cap P} (\operatorname{rot} E \bullet \nu)\, d\mathcal{H}^2 = \mathcal{L}(E, \delta_\epsilon),$$

and, according to the integral mean theorem,

$$\operatorname{rot} E(x_0) \bullet \nu \; = \lim_{\epsilon \to 0} \frac{1}{\pi \epsilon^2} \mathcal{L}(E, \delta_\epsilon)$$

or, equivalently,

$$\mathcal{L}(E, \delta_\epsilon) = \pi \, (\operatorname{rot} E(x_0) \bullet \nu)\, \epsilon^2 + o(\epsilon^2).$$

In other words, $(\operatorname{rot} E(x_0) \bullet \nu)\, \pi \epsilon^2$ is the measure at the first order of the work done by E at the circuit δ_ϵ. Of course, the work depends on the orientation of the circuit δ and takes its maximum (at first order) when $\nu = \operatorname{rot} E(x_0)/|\operatorname{rot} E(x_0)|$ where it is given by $|\operatorname{rot} E(x_0)|\, \pi \epsilon^2$.

4.69 Example. Let us illustrate a new computation of the linking number of two simple and closed curves γ and δ in $\mathbb{R}^3$. Let $\omega := \sum_{i=1}^3 (-1)^{i-1} \frac{z^i}{|z|^3} \widehat{dz^i}$ be the volume 2-form of S^2, $g(x,y) := y - x$, $x, y \in \mathbb{R}^3$ and $f(x,y) := \frac{y-x}{|y-x|}$. Then $f(x,y) = \pi \circ g$, $\pi(z) := z/|z|$, hence $f^\# \omega = \gamma^\# \pi^\# \omega = g^\# \omega$, compare Example 4.51. We now compute

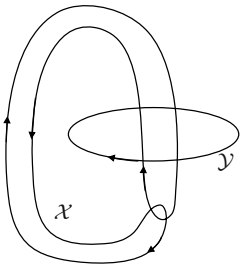


Figure 4.7. The linking number of the two curves in the figure is zero although the two curves cannot be unlinked.

$$\begin{aligned} f^{\#}\omega = \gamma^{\#}\omega &= \sum_{i=1}^{3}(-1)^{i-1}\frac{(\delta(s)-\gamma(t))^i}{|\delta(s)-\gamma(t)|^3}d(\widehat{\delta(s)-\gamma}(t))^i \\ &= \sum_{i=1}^{3}(-1)^{i-1}\frac{(\delta(s)-\gamma(t))^i}{|\delta(s)-\gamma(t)|^3}M(\mathbf{D}(\delta(s)-\gamma(t)))^{\bar{i}}_{(1,2)}\,dt\wedge ds \\ &= -\sum_{i=1}^{3}\frac{(\delta(s)-\gamma(t))^i}{|\delta(s)-\gamma(t)|^3}(\gamma'(t)-\delta'(s))^i\,dt\wedge ds \\ &= -\frac{\mathrm{vol}\Big(\delta(s)-\gamma(t)),\gamma'(t),\delta'(s)\Big)}{|\delta(s)-\gamma(t)|^3}\,dt\wedge ds, \end{aligned}$$

where $\mathrm{vol}\,(a,b,c)$ is the triple product of the vectors a, b and c, compare Exercise 4.66. Since $\int_{S^2}\omega = \mathcal{H}^2(S^2) = 4\pi$, we infer

$$4\pi\ \mathrm{link}(\mathcal{X},\mathcal{Y}) = \int_0^1\int_0^1 \frac{\mathrm{vol}\,(\gamma(s)-\delta(t),\gamma'(s),\delta'(t))}{|\gamma(s)-\delta(t)|^3}\,ds\,dt.$$

4.70 Example (Ampère's law). The linking number is strongly related to Ampère's law on the circulation of the magnetic field. Suppose that an electric current travels along a closed regular curve $\gamma(t)$, $t\in[0,1]$. A magnetic field is generated and, according to the *Biot–Savart law*, the magnetic field H at x due to the current traveling the infinitesimal oriented arc dy is proportional to

$$\frac{dy\times(x-y)}{|y-x|^3},$$

hence the total magnetic flux at x due to the circulation of the electric current in γ is proportional to

$$B(x) := \int_0^1 \frac{(\gamma(s)-x)}{|\gamma(s)-x|^3}\times\gamma'(s)\,ds.$$

Consequently, the work $\mathcal{L}$ done along a curve $\delta:[0,1]\to\mathbb{R}^3$, with disjoint trajectory from γ is proportional to

$$\int_0^1 B(\delta(t))\bullet\delta'(t)\ dt = \int_0^1 dt\int_0^1 \frac{\mathrm{vol}\,(\gamma(s)-\delta(t),\gamma'(s),\delta'(t))}{|\gamma(s)-\delta(t)|^3}\,ds,$$

i.e., compare Example 4.69, $\mathcal{L}$ is proportional to the linking number of γ and δ: this is the *Ampère law*.

One sees that two "unlinked" curves have zero linking, but it is not true in general that two curves with zero linking number can be unlinked, see Figure 4.7. This shows that the linking number is more related to work than to the intuitive notion of geometric link.

4.71 Example. Let us compute the divergence and the curl of the field

$$B(x) := \int_0^1 \frac{\gamma(s)-x}{|\gamma(s)-x|^3} \times \gamma'(s)\, ds, \qquad x \in \mathbb{R}^3 \setminus \operatorname{Im}\gamma.$$

For $x, y \in \mathbb{R}^3$ we set $F(y,x) := \frac{y-x}{|y-x|^3}$.

Computing the divergence. We have

$$\nabla \bullet B\,(x) = \nabla \bullet \int_0^1 F(\gamma(s),x) \times \gamma'(s)\, ds = \int_0^1 \nabla \bullet (F(\gamma(s),x) \times \gamma'(s))\, ds.$$

Now,

$$\nabla \bullet a \times b = *d*(*(a \wedge b)) = *d(a \wedge b) = *(da \wedge b - a \wedge db).$$

In our case, $a = a(x) = f(\gamma(s),x)$ and $b = \gamma'(s)$. Consequently, $db = 0$ but also $da = 0$. Therefore, $\nabla \bullet (F(\gamma(s),x) \times \gamma'(s)) = 0\ \forall s$ and, by integration, we conclude $\operatorname{div} B = 0$ in $\mathbb{R}^3 \setminus \operatorname{Im}\gamma$.

Computing the curl. We have

$$(\nabla \times B)(x) = \int_0^1 \nabla \times (F(\gamma(s),x) \times \gamma'(s))\, ds$$

and recall that

$$\nabla \times (F \times G) = F \operatorname{div} G - G \operatorname{div} F + (\,G \bullet \nabla\,)F - (\,F \bullet \nabla\,)G.$$

Since in our case $F = F(\gamma(s),x)$ and $G = \gamma'(s)$, this yields

$$\nabla_x \times (F(\gamma(s),x) \times \gamma'(s)) = \gamma'(s) \operatorname{div}_x F(\gamma(s),x) + (\,\gamma'(s) \bullet \nabla_x\,)F(\gamma(s),x).$$

On the other hand, it is easily seen that

$$\operatorname{div}_x\Big(\frac{\gamma(s)-x}{|\gamma(s)-x|^3}\Big) = 0 \qquad \text{and} \qquad \nabla_x F(y,x) = -\nabla_y F(y,x);$$

therefore,

$$\begin{aligned}\nabla_x \times (F(\gamma(s),x) \times \gamma'(s)) &= \gamma'(s) \bullet \nabla_x\, F(x,\gamma(s)) \\ &= -\,\gamma'(s) \bullet \nabla_y F(x,\gamma(s)) = -\frac{d}{ds} F(\gamma(s),x).\end{aligned}$$

Hence, we conclude that

$$\nabla \times B(x) = -\int_0^1 \frac{d}{ds} F(\gamma(s),x)\, ds = F(\gamma(1),x) - F(\gamma(0),x);$$

that is, $\operatorname{rot} B = 0$ in $\mathbb{R}^3 \setminus \operatorname{Im}\gamma$ because γ is closed.

4.5 Closed and Exact Forms

The question of whether a closed differential form is exact is deeply connected and actually is one of the formulations of the difficult question of deciding whether a $(k-1)$-dimensional submanifold Ω is or is not the boundary of a k-submanifold. However, it is simpler as it involves only oriented integrals, i.e., mean properties.

4.5.1 Poincaré's lemma

4.72 Definition. *Let Ω be an open set in $\mathbb{R}^n$. A k-form ω of class C^1 is said to be* closed *in Ω if $d\omega = 0$. A k-form with continuous coefficients is said to be* exact *in Ω if there exists a $(k-1)$-form α of class $C^1(\Omega)$ such that $\omega(x) = d\alpha(x)$ $\forall x \in \Omega$. In this case we say that α is a* primitive *of ω in Ω.*

Since $d^2\omega = 0$, every exact form of class C^2 is closed. As we saw in [GM4] Chapter 3, there are closed 1-forms that are not exact, although as we saw there, if Ω is simply connected, then all closed 1-forms in Ω are exact in Ω. We now partially extend the previous result to k-forms.

We state a few facts.

4.73 Definition. *An open set $\Omega \subset \mathbb{R}^n$ is said to be* contractible *if there exist a continuous map $H : [0,1] \times \Omega \to \Omega$ and a point $x_0 \in \Omega$ such that $H(1,x) = x$ and $H(0,x) = x_0$ $\forall x \in \Omega$.*

In other words, Ω is contractible if and only if the identity map in Ω is homotopic with values in Ω to a constant. We observe that when Ω is contractible, by a suitable procedure of regularization that we do not discuss here, we may assume that the homotopy $H : \mathbb{R} \times \Omega \to \Omega$ is a map defined in $\mathbb{R} \times \Omega$ and of class $C^\infty(\mathbb{R} \times \Omega)$.

Let Ω be an open set of $\mathbb{R}^n$ and denote by $(e_t, e_1, e_2, \dots, e_n)$ the standard basis of $\mathbb{R} \times \mathbb{R}^n$ and by $(dt, dx^1, \dots, dx^n)$ its dual basis. Every k-form ω in $\mathbb{R} \times \Omega$ may be written uniquely as

$$\omega = \sum_{\alpha \in I(k,n)} \omega_\alpha dx^\alpha + \sum_{\beta \in I(k-1,n)} \omega_{(\beta,t)} dt \wedge dx^\beta =: \omega_1 + dt \wedge \eta_\omega. \quad (4.66)$$

Trivially $< \omega_1, v_1 \wedge v_2 \wedge \cdots \wedge v_k >= 0$ and $< \eta_\omega, v_1 \wedge v_2 \wedge \cdots \wedge v_{k-1} >= 0$ if one of the $\{v_i\}$ is a multiple of $e_t = (1, 0, \dots, 0)$. We introduce the *homotopy map*

$$K : k\text{-forms in } \mathbb{R} \times \Omega \ \to \ (k-1)\text{-forms in } \Omega$$

that maps a k-form ω in $\Omega \times \mathbb{R}$ into the $(k-1)$-form $K(\omega)$ in Ω obtained by *integrating along the fiber* the form η_ω in the decomposition (4.66),

$$< K(\omega)(x), \xi >:= \int_0^1 < \eta_\omega(t,x), \xi > \, dt, \qquad \forall \xi \in \Lambda_k \mathbb{R}^n. \quad (4.67)$$

Clearly, $K(\omega)$ is of class C^s if ω is of class C^s for every $s \geq 0$. For all $t \in \mathbb{R}$ denote by i_t the map $x \to (t,x)$ from Ω into $\mathbb{R} \times \Omega$.

4.74 Proposition. *For every k-form ω of class C^s, $s \geq 1$, we have*

$$i_1^\# \omega - i_0^\# \omega = dK(\omega) + K(d\omega).$$

Proof. Since the formula is linear in ω, it suffices to prove it for forms of the type

$$\omega := f(x,t)dx^\alpha,\ |\alpha| = k, \qquad \text{and} \qquad \omega = dt \wedge g(x,t)dx^\beta,\ |\beta| = k-1,$$

where f and g are of class C^1. In the first case,

$$d\omega = dt \wedge \frac{\partial f}{\partial t}(x,t)\, dx^\alpha + \text{ terms that do not involve } dt,$$

hence for every $\xi \in \Lambda_k \mathbb{R}^n$

$$\begin{aligned} < K(d\omega), \xi > &= \int_0^1 \frac{\partial f}{\partial t}(t,x) dt < dx^\alpha, \xi > \\ &= (f(1,x) - f(0,x)) < dx^\alpha, \xi > = < i_1^\# \omega - i_0^\# \omega, \xi >, \end{aligned} \tag{4.68}$$

and the claim follows since trivially $K(\omega)$ vanishes. In the second case,

$$d\omega = \sum_{i=1}^n dt \wedge g_{x^i} dx^i \wedge dx^\beta,$$

hence

$$< K(d\omega), \xi > = \sum_{i=1}^n \int_0^1 \frac{\partial g}{\partial x^i}(t,x)\, dt < dx^i \wedge dx^\beta, \xi > .$$

On the other hand, differentiating under the integral sign,

$$d(K\omega) = d\Big(\int_0^1 g(t,x)\, dt \Big) \wedge dx^\beta = -\sum_{i=1}^n \Big(\int_0^1 \frac{\partial g}{\partial x^i}(t,x) dt \Big) dx^i \wedge dx^\beta.$$

It follows that $d(K\omega) + K(d\omega) = 0$. As $i_1^\# \omega = i_0^\# \omega = 0$, we again see that our formula holds. □

4.75 Theorem (Poincaré's lemma). *Let $\Omega \subset \mathbb{R}^n$ be contractible with a smooth homotopy H. Then every closed k-form in Ω is exact in Ω. More precisely, for every closed k-form ω of class C^s, $s \geq 1$, the $(k-1)$-form $K(H^\# \omega)$ is of class C^s and*

$$dK(H^\# \omega) = \omega \qquad \text{in } \Omega. \tag{4.69}$$

Proof of Theorem 4.75. Let ω be a closed k-form in Ω that we think of as being extended to $\mathbb{R} \times \Omega$ by extending its coefficients as constant in t. Let $H : \Omega \times \mathbb{R} \to \Omega$ be the contraction map and $i_1, i_0 : \Omega \to \Omega \times \mathbb{R}$ given by $i_1(x) = (1,x)$, $i_0(x) = (0,x)$. Then

$$H \circ i_1 = \text{Id on } \Omega, \qquad H \circ i_0 = x_0 \text{ on } \Omega,$$

hence

$$\omega = (H \circ i_1)^\# \omega = i_1^\# H^\# \omega, \qquad 0 = (H \circ i_0)^\# \omega = i_0^\# H^\# \omega.$$

Since $d\omega = 0$, we find $dH^\# \omega = H^\#(d\omega) = 0$, thus Proposition 4.74 yields $\omega = i_1^\# H^\# \omega = d(K(H^\# \omega))$ and $K(H^\# \omega)$ is a primitive of ω. It is easily seen that the primitive is of class C^s if ω is of class C^s and H is of class C^{s+1}. □

Notice that (4.69) provides an explicit formula for computing the primitive of an exact form.

4.5.2 The homotopy formula

As a consequence of Proposition 4.74 we have the following.

4.76 Theorem (Homotopy formula). *Let $U \subset \mathbb{R}^n$ and $\Omega \subset \mathbb{R}^N$ be open sets and $\phi, \psi : U \to \Omega$ maps of class $C^2(U)$ that are homotopic, with the homotopy map $H : [0,1] \times U \to \Omega$ of class $C^2(\mathbb{R} \times \Omega)$. Then for every k-form ω of class $C^1(\Omega)$ we have*

$$\phi^\# \omega - \psi^\# \omega = (d \circ K + K \circ d) H^\# \omega.$$

4.77 Corollary. *Let U be an open set in $\mathbb{R}^k$, V an open neighborhood of $[0,1] \times \overline{U}$ in $\mathbb{R}^{k+1}$, Ω an open set in $\mathbb{R}^n$ and $\phi, \psi : \overline{U} \to \Omega$ two maps with $\phi = \psi$ on ∂U and homotopic by the homotopy $H : V \to \Omega$ of class $C^2(V)$ such that*

$$H(1,x) = \phi(x), \qquad H(0,x) = \psi(x) \qquad \forall x \in U$$

and $H(t,x) = H(0,x)$ for every $t \in [0,1]$ and $x \in \partial U$. Then for every closed k-form ω of class $C^1(\Omega)$

$$\int_U \phi^\# \omega = \int_U \psi^\# \omega.$$

Proof. In fact, from the homotopy formula, $\phi^\# \omega - \psi^\# \omega = d(KH^\# \omega)$ in U with $KH^\# \omega$ of class $C^1(V)$. Stokes's formula then yields

$$\int_{\phi(U)} \omega - \int_{\psi(U)} \omega = \int_U (\phi^\# \omega - \psi^\# \omega) = \int_U d(KH^\# \omega) = \int_{\partial U} KH^\# \omega = 0.$$

In fact, the last integral vanishes because $H(t,x) = H(0,x)$ is constant in t for all $x \in \partial U$, hence the components of the form $H^\# \omega$ relative to the differentials containing dt vanish on ∂U. □

4.78 Proposition. *Let Ω be an open set of $\mathbb{R}^N$, $\mathcal{X}$ a smooth, compact, boundaryless, oriented k-submanifold of finite measure in $\mathbb{R}^n$, and $f, g : \mathcal{X} \to \Omega$ two homotopic maps of class C^2 with homotopy $H :]a,b[\times \mathcal{X} \to \Omega$, $]a,b[\supset\supset [0,1]$, of class C^2. Then*

$$\int_{\mathcal{X}} f^\# \omega = \int_{\mathcal{X}} g^\# \omega$$

for every closed k-form of class $C^1(\Omega)$.

Proof. By joining the orthogonal projection onto $\mathcal{X}$, see [GM4], with the homotopy we may extend H to a C^2 homotopy to an open neighborhood $U \subset \mathbb{R} \times \mathbb{R}^n$ of $[0,1] \times \mathcal{X}$ into Ω in such a way that again $H(1,x) = f(x)$ $\forall x \in \mathcal{X}$ and $H(0,x) = g(x)$ $\forall x \in \mathcal{X}$. From Proposition 4.74, $KH^\# \omega$ is of class $C^1(U)$ and

$$f^\# \omega - g^\# \omega = dKH^\# \omega \qquad \text{in } U,$$

hence, integrating and using Stokes's formula,

$$\int_{\mathcal{X}} f^\# \omega = \int_{\mathcal{X}} g^\# \omega + \int_{\mathcal{X}} dKH^\# \omega = \int_{\mathcal{X}} g^\# \omega.$$

□

4.79 Proposition (Degree homotopic invariance). *Let $\mathcal{X}$ and $\mathcal{Y}$ be two smooth, compact, boundaryless, oriented k-dimensional submanifolds of finite measure of $\mathbb{R}^n$ and let $f, g : \mathcal{X} \to \mathcal{Y}$ be two homotopic maps of class C^2 in an open neighborhood of $[0,1] \times \mathcal{X}$ into $\mathcal{Y}$. Then $\deg(f) = \deg(g)$.*

Proof. The k-covector field ω on $\mathcal{Y}$ dual of the field of k-vectors that orients $\mathcal{Y}$ has a nonzero integral according to the area formula. On the other hand, by means of the orthogonal projection onto $\mathcal{Y}$, ω extends to a tubular neighborhood of $\mathcal{Y}$ as a smooth k-form. It then follows from the definition of degree and from Proposition 4.78 that

$$\deg(f) \int_{\mathcal{Y}} \omega = \int_{\mathcal{X}} f^{\#}\omega = \int_{\mathcal{X}} g^{\#}\omega = \deg(g) \int_{\mathcal{Y}} \omega,$$

which yields the result at once. □

The invariance of $\mathrm{link}(\mathcal{X}, \mathcal{Y})$ with respect to homotopic transformations of $\mathcal{X}$ and $\mathcal{Y}$ that maintain empty intersection now follows at once.

4.80 Proposition. *Let $\mathcal{X}$ and $\mathcal{X}'$ be two k-submanifolds, $\mathcal{Y}$ and $\mathcal{Y}'$ be two $(n-k-1)$-submanifolds and $f : \mathcal{X} \to \mathcal{X}'$ and $g : \mathcal{Y} \to \mathcal{Y}'$ be two injective maps of class C^2. Let $\Delta := \{(x,y) \in \Omega \times \Omega \,|\, y = x\}$. Suppose there is a map $H : V \to ((\Omega \times \Omega) \setminus \Delta)$ of class $C^2(V)$ where V is an open neighborhood of $[0,1] \times ((\Omega \times \Omega) \setminus \Delta)$ such that $H(1,x,y) = x - y$ and $H(0,x,y) = f(x) - g(y)$. Then $\mathrm{link}(\mathcal{X}, \mathcal{Y}) = \mathrm{link}(\mathcal{X}', \mathcal{Y}')$.*

4.81 Definition. *Let Ω be an open set in $\mathbb{R}^n$. A k-submanifold $\mathcal{X} \subset \Omega$ is said to be* contractible to a set of zero $\mathcal{H}^k$-measure in Ω *if there exists a map $H : V \to \Omega$ of class C^2, V being an open neighborhood of $[0,1] \times \mathcal{X}$, such that $H(1,x) = x \ \forall x \in \mathcal{X}$ and $H(\{0\} \times \mathcal{X})$ has zero $\mathcal{H}^k$-measure.*

4.82 Proposition. *If an oriented k-submanifold $\mathcal{X}$ of finite $\mathcal{H}^k$-measure is contractible in Ω to a set of zero $\mathcal{H}^k$-measure, then*

$$\int_{\mathcal{X}} \omega = 0$$

for all k-forms ω of class $C^1(\Omega)$.

4.83 Remark. By approximation the results of this subsection extend to homotopies and maps of class C^1. We do not insist on this point and refer the reader to Chapter 3 of [GM4].

4.5.3 A theorem by de Rham

We saw in Chapter 3 of [GM4] that the simple connectedness of a domain Ω, a condition that is weaker than the contractibility of Ω, is a necessary and sufficient condition for the exactness of a closed form or for an irrotational field to have a potential in Ω. However, the simple connectedness of the domain does not suffice for a 2-form to be exact.

4.84 Example. In $\mathbb{R}^n$, $n \geq 2$, consider the field

$$E(x) := \frac{x}{|x|^n}$$

defined in $\mathbb{R}^n \setminus \{0\}$ and regular there and the corresponding $(n-1)$-form

$$\omega := *E := \sum_{i=1}^{n} (-1)^{i-1} \frac{x^i}{|x|^n} \widehat{dx^i}.$$

One readily sees that $d\omega = 0$ in $\mathbb{R}^n \setminus \{0\}$. However, ω is not exact. Suppose on the contrary that $\omega = d\alpha$ for some α, then Stokes's theorem yields

$$\int_{S^{n-1}} \omega = \int_{S^{n-1}} d\alpha = 0$$

and, as we have computed several times,

$$\int_{S^{n-1}} \omega = \int_{S^{n-1}} *\omega \bullet * \nu = \int_{S^{n-1}} \frac{1}{|x|^{n-1}} \, d\mathcal{H}^{n-1} = \mathcal{H}^{n-1}(S^{n-1}) \neq 0.$$

If we take into account Poincaré's lemma, this shows, in particular, that $\mathbb{R}^n \setminus \{0\}$ is not contractible if $n \geq 2$, whereas it is simply connected if $n \geq 3$.

Poincaré's lemma provides us only a sufficient condition for a closed form to be exact. A characterization involves the geometry of the domain and has an integral equivalent[3]. For instance the following result, that we do not prove, holds.

4.85 Theorem (de Rham). *Let Ω be a bounded domain in $\mathbb{R}^n$ with a smooth boundary. A closed k-form is exact in Ω if and only if for every smooth, compact, boundaryless, oriented k-submanifold $\mathcal{Y}$ in Ω with finite measure we have*

$$\int_{\mathcal{Y}} \omega = 0.$$

The following then follows at once.

4.86 Corollary. *Let Ω be an open set of $\mathbb{R}^n$. Suppose that every smooth, compact, boundaryless, oriented k-submanifold of Ω is contractible in Ω to a set of zero $\mathcal{H}^k$-measure. Then all closed k-forms in Ω are exact in Ω, more precisely, all closed k-forms in Ω of class C^∞ have a potential of class $C^\infty(\Omega)$.*

4.87 Example (Vector potential). Let $\Omega \subset \mathbb{R}^3$ be a star-shaped domain with respect to the origin so that $H(t,x) := tx$ is a contraction of Ω to $\{0\}$, and let $\omega = \sum_{i=1}^{3} a_i(x)\, dx^i$ be a closed 1-form in Ω, equivalently, the relations

$$a_{i,x^j}(x) = a_{j,x^i}(x)$$

hold for all i and j at every $x \in \Omega$. We have

[3] Also delicate regularity considerations are involved that we do not want to deal with. For this reason, all forms in the rest of this chapter are assumed to be smooth, meaning C^∞, if not otherwise stated.

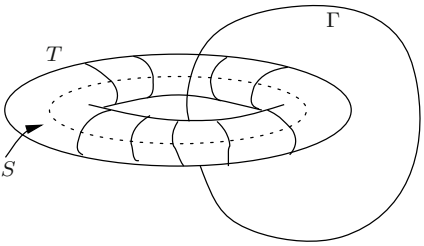


Figure 4.8. In $\mathbb{R}^3 \setminus \Gamma$, every closed and boundaryless 2-submanifold can be deformed without touching Γ to a set of zero 2-dimensional measure. In the figure, T collapses to S without touching Γ.

$$H^{\#}\omega(x) = \sum_{i=1}^{3} a_i(tx)\, d(tx^i) = \sum_{i=1}^{3} a_i(tx) x^i\, dt + \sum_{i=1}^{3} a_i(tx) t\, dx^i,$$

hence

$$f(x) := K(H^{\#}\omega)(x) := \sum_{i=1}^{3} \int_0^1 a_i(tx) x^i\, dt$$

is a primitive of ω, $df = \omega$, according to (4.69). This can also be proved directly, in fact

$$\begin{aligned} D_j f(x) &= \int_0^1 \Big(\sum_{i=1}^{3} a_{i,x^j}(tx) t x^i + a_j(tx) \Big)\, dt = \int_0^1 \Big(\sum_{i=1}^{3} a_{j,x^i}(tx) t x^i + a_j(tx) \Big)\, dt \\ &= \int_0^1 \frac{d}{dt}(t a_j(tx))\, dt = a_j(x). \end{aligned}$$

Similarly, if $E = E_1\, dx^1 + E_2\, dx^2 + E_3\, dx^3$ has zero divergence, i.e., compare (4.64), if

$$\omega = *E = E_1\, dx^2 \wedge dx^3 - E_2\, dx^1 \wedge dx^3 + E_3\, dx^1 \wedge dx^2$$

is a closed 2-form in Ω, then $\alpha := KH^{\#}\omega$ is a primitive of ω. We compute

$$\begin{aligned} H^{\#}\omega &= E_1(tx)\, d(tx^2) \wedge d(tx^3) - E_2(tx)\, d(tx^1) \wedge d(tx^3) + E_3(tx)\, d(tx^1) \wedge d(tx^2) \\ &= E_1(tx)(x^2 dt + t dx^2) \wedge (dt x^3 + t dx^3) + \cdots \\ &= dt \wedge \Big(E_1(tx) t\, (x^2\, dx^3 - x^3\, dx^2) - E_2(tx) t\, (x^1\, dx^3 - x^3\, dx^1) \\ &\qquad + E_3(tx)\, t (x^1\, dx^2 - x^2\, dx^1) \Big) + \text{ terms that do not involve } dt. \end{aligned}$$

If we set $F := (F_1, F_2, F_3)$ with

$$\begin{aligned} F_1(x) &:= \int_0^1 E_1(tx) t\, dt, \\ F_2(x) &:= \int_0^1 E_2(tx) t\, dt, \\ F_3(x) &:= \int_0^1 E_3(tx) t\, dt, \end{aligned}$$

it follows from (4.69) that the 1-form

$$\begin{aligned}\alpha(x) := KH^{\#}\omega(x) &= F_1\,(x^2\,dx^3 - x^3\,dx^2) - F_2\,(x^1\,dx^3 - x^3\,dx^1)\\ &\quad + F_3\,(x^1\,dx^2 - x^2\,dx^1)\\ &= (F_2x^3 - F^3x^2)\,dx^1 - (F^1x^3 - F^3x^1)\,dx^2 + (F_1x^2 - F_2x^1)\,dx^3\\ &= \sum_{i=1}^{3}(F\times x)^i dx^i\end{aligned}$$

is a primitive of ω, $d\alpha = \omega$. In other words, the field $H(x) = (H_1, H_2, H_3)$ given by

$$H(x) = \int_0^1 (E(tx)\times x)\,t\,dt, \tag{4.70}$$

solves the equation $\operatorname{rot} H = E$.

4.88 Proposition. *Let $\Omega \subset \mathbb{R}^3$ be contractible or, more generally, assume that every boundaryless 2-submanifold of it is contractible to a set of zero $\mathcal{H}^2$-measure. Then the equation $\operatorname{rot} H = E$ has a solution in Ω of class C^∞ if and only if E is of class C^∞ and $\operatorname{div} E = 0$ in Ω.*

Proof. Consider $\mathbb{R}^3$ endowed with the standard orthonormal basis so that we can identify vectors and forms. If $E = (E_1, E_2, E_3)$ is a vector field and $\omega = E_1\,dx^1 + E_2\,dx^2 + E_3\,dx^3$ the corresponding 1-form, as we have seen

$$\operatorname{div} E = *d*\omega, \qquad \operatorname{rot} E = *d\omega.$$

If $\operatorname{rot} H = E$ and α is the corresponding differential 1-form to H, then $*d\alpha = \omega$, i.e., $d\alpha = *\omega$, hence $0 = d^2\alpha = d(*\omega)$. Consequently $\operatorname{div} E = *d(*\omega) = 0$. Conversely, if $\operatorname{div} E = 0$ in Ω, then $*\omega$ is a closed 2-form in Ω, consequently, by de Rham's theorem (or Poincaré's lemma if Ω is contractible), there exists a 1-form α such that $d\alpha = *\omega$, equivalently, $*d\alpha = **\omega = \omega$. In terms of the associated vector field H associated to α, i.e., $H = (H_1, H_2, H_3)$ with $\alpha = H_1dx^1 + H_2dx^2 + H_3dx^3$, the relation $d\alpha = *\omega$ just amounts to $\operatorname{rot} H = E$. □

We emphasize that (4.70) in Example 4.87 yields an explicit formula for H when Ω is star-shaped with respect to the origin.

4.89 Example. Let $\mathcal{Y}$ be the image of a regular smooth curve in $\mathbb{R}^3$. Every compact, boundaryless, oriented 2-submanifold $\mathcal{X}$ that does not intersect $\mathcal{Y}$ is contractible to one or more lines. Consequently, for every divergence-free field B in $\mathbb{R}^3 \setminus \mathcal{Y}$, there is a field A such that $\operatorname{rot} A = B$ in $\mathbb{R}^3 \setminus \mathcal{Y}$.

4.90 Remark. In terms of PDE's, $\operatorname{rot} H = E$ is the following system of first order PDE's

$$\begin{cases} \dfrac{\partial H_3}{\partial x^2} - \dfrac{\partial H_2}{\partial x^3} = E_1,\\[2ex] \dfrac{\partial H_1}{\partial x^3} - \dfrac{\partial H_3}{\partial x^1} = E_2, \qquad \text{in } \Omega.\\[2ex] \dfrac{\partial H_2}{\partial x^1} - \dfrac{\partial H_1}{\partial x^2} = E_3 \end{cases} \tag{4.71}$$

Proposition 4.88 shows that (4.71) has a solution in a contractible open set Ω if and only if $\operatorname{div} E = 0$ in Ω and, in this case, the solution is found by integration.


Figure 4.9. Hermann von Helmholtz (1821–1894) and Hermann Weyl (1885–1955).

4.91 Example. de Rham's theorem provides a necessary and sufficient condition for *every* divergence-free field to have a vector potential. For specific fields, it is often easier to write down a vector potential.

For instance, in Example 4.71 we showed that the vector field

$$B(x) := -\frac{1}{4\pi}\int_0^1 \frac{\gamma(s)-x}{|\gamma(s)-x|^3}\times\gamma'(s)\,ds, \qquad x\in\mathbb{R}^3\setminus \operatorname{Im}\gamma,$$

where $\gamma(t)$, $t\in[0,1]$, is a regular curve, has zero divergence. Proposition 4.88 grants us a vector potential on $\mathbb{R}^3\setminus\operatorname{Im}\gamma$ of B, i.e., a field A such that $\operatorname{rot} A = B$. However, without taking into account de Rham's theorem, we may observe that if a is a constant vector and φ is a scalar function, then $\nabla\times(\varphi(x)a) = (\nabla\varphi)\times a$ and that

$$\frac{\gamma(s)-x}{|\gamma(s)-x|^3} = \nabla_x \frac{1}{|\gamma(s)-x|}.$$

Therefore,

$$\frac{\gamma(s)-x}{|\gamma(s)-x|^3}\times\gamma'(s) = \nabla_x\frac{1}{|\gamma(s)-x|}\times\gamma'(s) = \nabla\times\Big(\frac{1}{|\gamma(s)-x|}\times\gamma'(s)\Big),$$

and, integrating,

$$B(x) = \int_0^1 \nabla\frac{1}{|\gamma(s)-x|}\times\gamma'(s)\,ds == \int_0^1 \nabla\times\Big(\frac{1}{|\gamma(s)-x|}\gamma'(s)\Big)\,ds.$$

Consequently, if

$$A(x) := \int_0^1 \frac{1}{|\gamma(s)-x|}\gamma'(s)\,ds,$$

we have $\operatorname{rot} A = \nabla\times A = B$ in $\mathbb{R}^3\setminus\operatorname{Im}\gamma$.

Finally, we state, as a consequence of Proposition 4.88, the *decomposition formula for fields* due to Hermann von Helmholtz (1821–1894).

4.92 Theorem (Helmholtz). *Let $E:\Omega\subset\mathbb{R}^3\to\mathbb{R}^3$ be a smooth, say $C^\infty(\Omega)$, field and Ω a contractible bounded open set with smooth boundary. There exist $f\in C^\infty(\Omega,\mathbb{R})$ and a field $H\in C^\infty(\Omega,\mathbb{R}^3)$ such that*

$$E = \nabla f + \operatorname{rot} H.$$

Proof. In fact, let $u : \Omega \to \mathbb{R}$ be the solution of the Neumann problem

$$\begin{cases} \Delta u = \operatorname{div} E & \text{in } \Omega, \\ \dfrac{du}{dn} = E \bullet n & \text{on } \partial\Omega, \end{cases}$$

then $\operatorname{div}(E - \nabla u) = 0$ in Ω, which one can prove to be of class $C^\infty(\Omega)$. Because of Proposition 4.88, there is a field $H : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^3$ of class $C^\infty(\Omega)$ such that

$$E - \nabla u = \operatorname{rot} H.$$

□

4.5.4 Hodge's decomposition formula

For the sake of completeness we state another result of decomposition for forms in bounded domains with smooth boundaries, not necessarily contractible. Its proof uses variational techniques, but it is not simple, thus we refer the reader to the specialized literature.

Let U be a bounded open set in $\mathbb{R}^n$ with smooth boundary. As usual $\mathfrak{t}\,\omega$ and $\mathfrak{n}\,\omega$ denote the tangential and normal components, respectively, of a k-form ω along ∂U, and we denote by $L := L^2(U, \Lambda_k\mathbb{R}^n)$ the Hilbert space of k-forms endowed with the inner product

$$(\omega|\eta)_{L^2} := \int_U \sum_\alpha \omega_\alpha \eta_\alpha \, dx.$$

Define the following subspaces of smooth forms in L:

$$\begin{aligned} \mathbf{H}_T^k &:= \Big\{ \omega \,\Big|\, \Delta\omega = 0,\ \mathfrak{t}\,(\omega) = 0 \Big\}, \\ \mathbf{H}_N^k &:= \Big\{ \omega \,\Big|\, , \Delta\omega = 0,\ \mathfrak{n}\,(\omega) = 0 \Big\}, \\ \operatorname{Im} d_T &:= \Big\{ \omega = d\alpha \,\Big|\, \mathfrak{t}\,\alpha = 0 \Big\}, \\ \operatorname{Im} d_N &:= \Big\{ \omega = d\alpha \,\Big|\, \mathfrak{n}\,\alpha = 0 \Big\}, \\ \operatorname{Im} \delta_T &:= \Big\{ \omega = \delta\alpha \,\Big|\, \mathfrak{t}\,\alpha = 0 \Big\}, \\ \operatorname{Im} \delta_N &:= \Big\{ \omega = \delta\alpha \,\Big|\, \mathfrak{n}\,\alpha = 0 \Big\}. \end{aligned}$$

4.93 Theorem (Hodge–Morrey decomposition theorem). *Let U be a bounded open set with boundary of class $C^\infty(U)$. Then the following hold:*

(i) $\mathbf{H}_T^k$ *is finite-dimensional.*
(ii) $\mathbf{H}_T^k$, $\operatorname{Im} d_T$ *and* $\operatorname{Im} \delta_T$ *are orthogonal, closed and supplementary on* L.
(iii) $(\operatorname{Im} d_T)^\perp = \{\alpha \,|\, \delta\alpha = 0,\ \mathfrak{t}\,\alpha = 0\}$ *in* L.
(iv) $(\operatorname{Im} \delta_T)^\perp = \{\alpha \,|\, d\alpha = 0,\ \mathfrak{t}\,\alpha = 0\}$ *in* L.

Similarly,

(i) $\mathbf{H}_N^k$ *is finite-dimensional.*
(ii) $\mathbf{H}_N^k$, $\operatorname{Im} d_N$ *e* $\operatorname{Im} \delta_N$ *are orthogonal, closed and supplementary in* L.
(iii) $(\operatorname{Im} d_N)^\perp = \{\alpha \,|\, \delta\alpha = 0,\ \mathfrak{n}\,\alpha = 0\}$ *in* L.
(iv) $(\operatorname{Im} \delta_N)^\perp = \{\alpha \,|\, d\alpha = 0,\ \mathfrak{n}\,\alpha = 0\}$ *in* L.

4.94 Remark. Let us state a few comments.

(i) One actually shows that the dimensions of $\mathbf{H}_T^k$ and $\mathbf{H}_N^k$ depend in a very precise way on the geometry of the domain, which we omit to illustrate. In particular, $\mathbf{H}_T^k = \mathbf{H}_N^k = \{0\}$ if U is contractible.
(ii) Denote by $\mathcal{E}^k(U)$ the space of k-forms of class C^∞. Hodge's decomposition theorem with some extra work implies that every differential form $\omega \in \mathcal{E}^k(U)$ with $\mathfrak{t}\,(\omega) = 0$ on ∂U uniquely decomposes in three orthogonal (in L) components as

$$\omega = H + d\alpha + \delta\beta,$$

where H is harmonic, $\alpha \in \mathcal{E}^{k-1}(U)$ and $\beta \in \mathcal{E}^{k+1}(U)$, and $\mathfrak{t}\,(H) = 0$, $\mathfrak{t}\,(\alpha) = 0$ and $\mathfrak{t}\,(\beta) = 0$ on ∂U. Similarly, every form $\omega \in \mathcal{E}^k(U)$ with $\mathfrak{n}\,(\omega) = 0$ on ∂U uniquely decomposes as

$$\omega = H + d\alpha + \delta\beta,$$

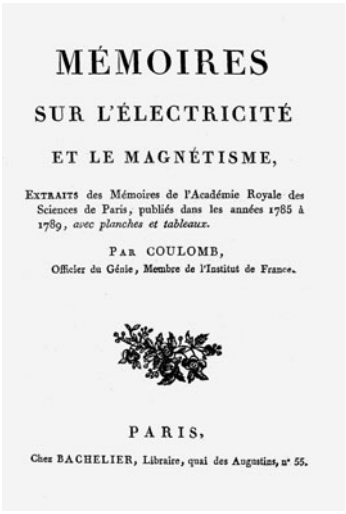
where H is harmonic, $\alpha \in \mathcal{E}^{k-1}(U)$ and $\beta \in \mathcal{E}^{k+1}(U)$, and $\mathfrak{n}\,(H) = 0$, $\mathfrak{n}\,(\alpha) = 0$ and $\mathfrak{n}\,(\beta) = 0$ on ∂U.
(iii) Suppose $d\omega = 0$ and $\mathfrak{t}\,\omega{=}0$. Then $\omega \in (\operatorname{Im} \delta_T)^\perp$ and ω uniquely decomposes as

$$\omega = H + d\alpha,$$

where H is harmonic and $\mathfrak{t}\,(H) = 0$ and $\mathfrak{t}\,(\alpha) = 0$ on ∂U. It is remarkable that the obstruction to the exactness of ω is the existence of harmonic forms (and definitively are finitely many!) Similarly, every closed form ω with $\mathfrak{n}\,\omega = 0$ on ∂U uniquely decomposes as $\omega = H + d\alpha$, where H is harmonic and this time $\mathfrak{n}\,H = 0$ and $\mathfrak{n}\,\alpha = 0$ on ∂U.

4.5.5 Maxwell equations

The equations of the electromagnetic field in $\mathbb{R}^3$ involve quite a number of physical quantities: the electric field $\mathbf{E}$, the magnetic field $\mathbf{H}$, the magnetic inductance $\mathbf{B}$, the dielectric displacement $\mathbf{D}$, the charge density ρ and the field of current density $\mathbf{j}$. These quantities are connected by *constitutive relations*, relations that describe the behavior of a material (for instance $\mathbf{E} = \mathbf{D}$ and $\mathbf{B} = \mathbf{H}$ in empty space), by boundary conditions on the


MÉMOIRES

SUR L'ÉLECTRICITÉ

ET LE MAGNÉTISME,

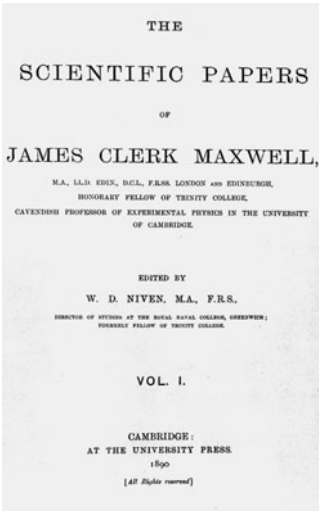
EXTRAITS des Mémoires de l'Académie Royale des Sciences de Paris, publiés dans les années 1785 à 1789, *avec planches et tableaux.*

PAR COULOMB,

Officier du Génie, Membre de l'Institut de France.

PARIS,

Chez BACHELIER, Libraire, quai des Augustins, n° 55.


THE

SCIENTIFIC PAPERS

OF

JAMES CLERK MAXWELL,

M.A., LL.D. EDIN., D.C.L., F.R.SS. LONDON AND EDINBURGH,

HONORARY FELLOW OF TRINITY COLLEGE,

CAVENDISH PROFESSOR OF EXPERIMENTAL PHYSICS IN THE UNIVERSITY OF CAMBRIDGE.

EDITED BY

W. D. NIVEN, M.A., F.R.S.,

DIRECTOR OF STUDIES AT THE ROYAL NAVAL COLLEGE, GREENWICH; FORMERLY FELLOW OF TRINITY COLLEGE.

VOL. I.

CAMBRIDGE:

AT THE UNIVERSITY PRESS.

1890

[*All Rights reserved*]

Figure 4.10. Frontispieces of two collections of works of Charles Coulomb (1736–1806) and James Clerk Maxwell (1831–1879).

boundary of the domain where these fields are defined and mainly by *Maxwell equations*

$$\begin{aligned}
&\text{(i)} && \operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} && \text{Faraday's induction law,}\\
&\text{(ii)} && \operatorname{rot}\mathbf{H} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} && \text{Ampère's law,}\\
&\text{(iii)} && \operatorname{div}\mathbf{D} = 4\pi\rho && \text{continuity equation,}\\
&\text{(iv)} && \operatorname{div}\mathbf{B} = 0 && \text{no magnetic charges.}
\end{aligned} \tag{4.72}$$

A convenient way to formulate Maxwell equations is in terms of differential forms in $\mathbb{R}^3$, thinking of time as of a parameter. Let us introduce the following five differential forms:

$$\begin{aligned}
\omega_1 &:= E_1 dx^1 + E_2 dx^2 + E_3 dx^3\\
\omega_2 &:= B_1 dx^2 \wedge dx^3 - B_2 dx^1 \wedge dx^3 + B_3 dx^1 \wedge dx^2,\\
\omega_3 &:= H_1 dx^1 + H_2 dx^2 + H_3 dx^3\\
\omega_4 &:= D_1 dx^2 \wedge dx^3 - D_2 dx^1 \wedge dx^3 + D_3 dx^1 \wedge dx^2),\\
\omega_5 &:= (j_1 dx^2 \wedge dx^3 - j_2 dx^1 \wedge dx^3 + j_3 dx^1 \wedge dx^2) \wedge dt,
\end{aligned} \tag{4.73}$$

and, if $\omega = \sum_\alpha \omega_a(x,t)dx^\alpha$, set

$$\frac{\partial \omega}{\partial t} := \sum_\alpha \frac{\partial \omega_\alpha}{\partial t}\, dx^\alpha.$$

Maxwell equations then rewrite as

$$\begin{cases} d\omega_1 = -\dfrac{1}{c}\dfrac{\partial \omega_2}{\partial t}, \\ d\omega_3 = \dfrac{4\pi}{c}\omega_5 + \dfrac{1}{c}\dfrac{\partial \omega_4}{\partial t}, \\ d\omega_2 = 0, \\ d\omega_4 = 4\pi\rho\, dx^1 \wedge dx^2 \wedge dx^3. \end{cases} \tag{4.74}$$

The constitutive relations in the empty space, $\mathbf{E} = \mathbf{D}$, $\mathbf{H} = \mathbf{B}$, become $\omega_4 = *\omega_1$, $\omega_2 = *\omega_3$, and, in absence of charges and currents and for time independent fields, Maxwell equations write as

$$\begin{cases} d\omega_1 = 0, \\ \delta\omega_1 = 0, \\ d\omega_3 = 0, \\ \delta\omega_3 = 0. \end{cases}$$

Assume, for the sake of simplicity, that Ω is contractible. Since $\operatorname{rot}\mathbf{E} = 0$, $\mathbf{E}$ has a potential in Ω, $\mathbf{E} = -\nabla\phi$ and the first two equations of the electric field become a single second order equation for the potential

$$-\Delta\phi = 4\pi\rho.$$

Moreover, from $\operatorname{div}\mathbf{B} = 0$ we also deduce the existence of a vector potential $\mathbf{A}$, $\operatorname{rot}\mathbf{A} = \mathbf{B}$. The equations for the electric potential and the magnetic potential $\mathbf{A}$ are then

$$\begin{cases} -\Delta\phi = 4\pi\rho, \\ -\Delta\mathbf{A} + \nabla\operatorname{div}\mathbf{A} = \operatorname{rot}\operatorname{rot}\mathbf{A} = \dfrac{4\pi}{c}\mathbf{j}. \end{cases} \tag{4.75}$$

In the nonstationary case, when fields are time dependent, we may use differential forms in the space-time $\mathbb{R}^4$ as follows. Define the 2- and 3-forms in $\mathbb{R}^4$

$$\begin{aligned} \alpha &:= (E_1 dx^1 + E_2 dx^2 + E_3 dx^3) \wedge c\, dt \\ &\quad + (B_1 dx^2 \wedge dx^3 - B_2 dx^1 \wedge dx^3 + B_3 dx^1 \wedge dx^2), \\ \beta &:= -(H_1 dx^1 + H_2 dx^2 + H_3 dx^3) \wedge c\, dt \\ &\quad + (D_1 dx^2 \wedge dx^3 - D_2 dx^1 \wedge dx^3 + D_3 dx^1 \wedge dx^2), \\ \gamma &:= (j_1 dx^2 \wedge dx^3 - j_2 dx^1 \wedge dx^3 + j_3 dx^1 \wedge dx^2) \wedge dt - \rho\, dx^1 \wedge dx^2 \wedge dx^3; \end{aligned}$$

then Maxwell equations become a system of two equations corresponding to (i), (iv) and (ii), (iii), respectively,

$$\begin{cases} d\alpha = 0, \\ d\beta + 4\pi\gamma = 0. \end{cases} \tag{4.76}$$

4.95 Maxwell equations in empty space are self-dual. In empty space $\mathbf{E} = \mathbf{D}$, $\mathbf{H} = \mathbf{B}$, $\mathbf{j} = 0$, $\rho = 0$ and $\gamma = 0$. The differential forms

$$\begin{aligned}\alpha &:= (E_1 dx^1 + E_2 dx^2 + E_3 dx^3) \wedge c\,dt \\ &\quad + (H_1 dx^2 \wedge dx^3 - H_2 dx^1 \wedge dx^3 + H_3 dx^1 \wedge dx^2), \\ \beta &:= -(H_1 dx^1 + H_2 dx^2 + H_3 dx^3) \wedge c\,dt \\ &\quad + (E_1 dx^2 \wedge dx^3 - E_2 dx^1 \wedge dx^3 + E_3 dx^1 \wedge dx^2).\end{aligned}$$

are quite similar. If we introduce *Lorentz metrics* in $\mathbb{R}^4$,

$$dx^2 + dy^2 + dz^2 - c^2 dt^2,$$

i.e., the nondegenerate bilinear form $a : \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$ given by

$$a((x,t),(y,s)) := x^1 y^1 + x^2 y^2 + x^3 y^3 - c^2 ts,$$

we may define the Hodge operator relative to the Lorentz metric by

$$\omega \wedge \eta =: a(*\omega, \eta)\, dx \wedge dy \wedge dz \wedge c\,dt.$$

Now, Hodge's operator acts on forms as follows:

$$\begin{cases} *\widehat{dx^i} = -dx^i \wedge c\,dt, \\ *(dx^i \wedge c\,dt) = \widehat{dx^i}, \\ ** = -\operatorname{Id}; \end{cases}$$

and it is easily seen that $\beta = *\alpha$. If we set $\delta\alpha := - * d * \alpha$, (4.76) rewrites in empty space as self-dual equations with respect to the Lorentz metrics

$$\begin{cases} d\alpha = 0, \\ \delta\alpha = 0. \end{cases}$$

Let us make a few remarks on some consequences of (4.76). The equation $d\gamma = 0$ that simply follows differentiating the second of (4.76) is, in terms of fields, the continuity equation involving currents and charges,

$$\operatorname{div} \mathbf{j} + \frac{\partial \rho}{\partial t} = 0.$$

The separation of the equations relative to the electric and magnetic fields may be performed by means of potentials when we operate in a domain Ω of the space-time that is contractible. In fact, in this case, there is a 1-form λ such that $d\lambda = \alpha$. If

$$\lambda = A_1\, dx^2 + A_2\, dx^2 + A_3\, dx^3 + \phi\, c\, dt,$$

$\mathbf{A} = (A_1, A_2, A_3)$ is called the *vector potential* and ϕ is called the *scalar potential*, The equation $d\lambda = \alpha$ that replaces the first of (4.76) corresponds, in terms of fields, to the equations


Figure 4.11. Michael Faraday (1791–1867) and James Clerk Maxwell (1831–1879).

$$\begin{cases} \operatorname{rot} \mathbf{A} = \mathbf{B}, \\ \nabla\phi - \dfrac{1}{c}\dfrac{\partial \mathbf{A}}{\partial t} = \mathbf{E}. \end{cases}$$

Replacing in (4.75), we conclude with a system of two second order equations that are equivalent to Maxwell equations

$$\begin{cases} \Delta\phi + \dfrac{1}{c}\dfrac{\partial}{\partial t}\operatorname{div} \mathbf{A} = -4\pi\rho, \\ \Delta\mathbf{A} - \dfrac{1}{c^2}\dfrac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla\Big(\operatorname{div} \mathbf{A} + \dfrac{1}{c}\dfrac{\partial\phi}{\partial t}\Big) = -\dfrac{4\pi}{c}\mathbf{j}. \end{cases} \tag{4.77}$$

However, the equations in (4.77) are still coupled, but we have a certain freedom in choosing the potential λ. In fact, if $d\lambda_1 = d\lambda_2 = \alpha$, then $d(\lambda_1 - \lambda_2) = 0$ and, in a simply connected region, $\lambda_1 - \lambda_2 = df$. Replacing $\mathbf{A}$ with $\overline{\mathbf{A}} := \mathbf{A} + \nabla f$, and ϕ with $\overline{\phi} := \phi - \frac{1}{c}\frac{\partial f}{\partial t}$, where f is an arbitrary function, (4.77) still hold. If we now choose f as a solution of the wave equation

$$\Delta f - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} = -\operatorname{div} \mathbf{A} - \frac{1}{c}\frac{\partial\phi}{\partial t}$$

and we set $\overline{\mathbf{A}} := \mathbf{A} + \nabla f$ and $\overline{\phi} := \phi - \frac{1}{c}\frac{\partial f}{\partial t}$, then

$$\operatorname{div} \overline{\mathbf{A}} + \frac{1}{c}\frac{\partial\overline{\phi}}{\partial t} = 0,$$

and the equations in (4.77) decouples into wave equations with velocity of propagation c:

$$\begin{cases} \Delta\overline{\phi} - \dfrac{1}{c^2}\dfrac{\partial^2\overline{\phi}}{\partial t^2} = -4\pi\rho, \\ \Delta\overline{\mathbf{A}} - \dfrac{1}{c^2}\dfrac{\partial^2\overline{\mathbf{A}}}{\partial t^2} = -\dfrac{4\pi}{c}\mathbf{j}. \end{cases}$$

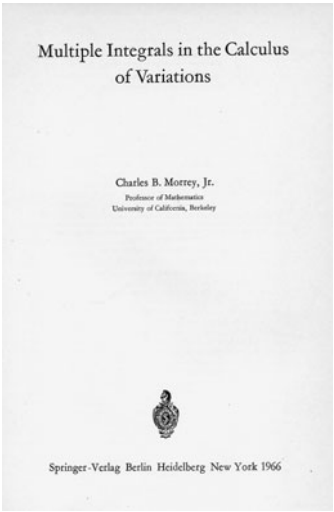


Figure 4.12. William Hodge (1903–1975) and the frontispiece of the monograph on Calculus of Variations of Charles Morrey (1907–1984).

4.96 Poynting flux-energy. The *Poynting flux-energy* vector is defined by

$$\mathbf{S} := \frac{c}{4\pi}\mathbf{E} \times \mathbf{H}$$

or, in terms of forms, as the field $\mathbf{S} = (S_1, S_2, S_3)$ such that

$$S_1\, dx^2 \wedge dx^3 - S_2\, dx^1 \wedge dx^3 + S_3\, dx^1 \wedge dx^2 := \frac{c}{4\pi}\omega_1 \wedge \omega_3.$$

Proposition (Poynting). *The following conservative law holds:*

$$\frac{1}{4\pi}\frac{\partial \mathbf{B}}{\partial t} \bullet \mathbf{H} + \mathbf{E} \bullet \mathbf{j} + \frac{1}{4\pi}\mathbf{E} \bullet \frac{\partial \mathbf{D}}{\partial t} + \operatorname{div} \mathbf{S} = 0.$$

Proof. Taking into account (4.73) and (4.74), it suffices to remark that $\operatorname{div} \mathbf{S}\, dx^1 \wedge dx^2 \wedge dx^3 = \frac{c}{4\pi} d(\omega^1 \wedge \omega_3)$ and compute

$$\begin{aligned} d(\omega_1 \wedge \omega_3) = d\omega_1 \wedge \omega_3 - \omega_1 \wedge d\omega_3 &= \Big(-\frac{1}{c}\frac{\partial \omega_2}{\partial t}\Big) \wedge \omega_3 - \omega_1 \wedge \Big(\frac{4\pi}{c}\omega_5 + \frac{1}{c}\frac{\partial \omega_4}{\partial t}\Big) \\ &= -\frac{1}{c}\frac{\partial \omega_2}{\partial t} \wedge \omega_3 - \frac{4\pi}{c}\omega_1 \wedge \omega_5 - \frac{1}{c}\omega_1 \wedge \frac{\partial \omega_4}{\partial t}. \end{aligned}$$

□

Finally, if $\mathbf{D} = k\,\mathbf{E}$ and $\mathbf{B} = \mu\,\mathbf{H}$, where k and μ are constant, called the dielectric and magnetic permeability of the means, respectively, then the Poynting theorem takes the form of a continuity equation

$$\frac{\partial u}{\partial t} + \operatorname{div} \mathbf{S} = -\,\mathbf{E} \bullet \mathbf{j}\,,$$

where

$$u := \frac{1}{8\pi}(k|\mathbf{E}|^2 + \mu|\mathbf{H}|^2)$$

is the *density of energy* of the field.

4.6 Exercises

4.97 ¶. Let $\mathbf{A}, \mathbf{B} \in M_{n,n}(\mathbb{R})$. Prove that

$$\det(\mathbf{A}+\mathbf{B}) = \sum_{\substack{\alpha,\beta \\ |\alpha|=|\beta|}} \sigma(\alpha,\overline{\alpha})\sigma(\beta,\overline{\beta})M_\alpha^\beta(\mathbf{A})M_{\overline{\alpha}}^{\overline{\beta}}(\mathbf{B}).$$

As a special case, apply the result to compute the characteristic polynomial of a matrix $\mathbf{A}$.

4.98 ¶ Binet's formula. Let $\mathbf{A}, \mathbf{B} \in M_{n,n}(\mathbb{R})$. Show that for every couple of multi-indices α and β with $1 \le |\alpha| = |\beta| \le b$, we have

$$M_\alpha^\beta(\mathbf{BA}) = \sum_{|\beta'|=|\beta|} M_{\beta'}^\beta(\mathbf{B})M_\alpha^{\beta'}(\mathbf{A}).$$

4.99 ¶. Let ω_1 and ω_2 be two differential forms in $\Omega \subset \mathbb{R}^n$, ω_1 being closed and ω_2 being exact. Show that $\omega_1 \wedge \omega_2$ is exact in Ω.

4.100 ¶ Volume form of a hypersurface. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of class C^1 and

$$\Gamma := \{x \,|\, f(x) = 0\}.$$

Suppose that $\mathbf{D}f(x) \neq 0$ on Γ so that Γ is oriented by $\nabla f/|\nabla f|$. Show that

$$\mathcal{H}^{n-1}(\Gamma) = \int_\Gamma \omega,$$

where ω is the $(n-1)$-form

$$\omega := \sum_{i=1}^n (-1)^{i-1} \frac{\partial f}{\partial x^i} \widehat{dx^i}.$$

4.101 ¶. Let $\omega(x) := \sum_{i=1}^n \frac{x^i}{|x|^n}\, dx^i$. Show that ω is a solution of the self-dual equations $d\omega = 0$ and $\delta\omega = 0$ in $\mathbb{R}^n \setminus \{0\}$.

4.102 ¶. Let $E(x) : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ be a central field, i.e., $E(x) = \varphi(x)\frac{x}{|x|}$, where $\varphi : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$. Show that E is conservative if and only if E is radial, i.e. $\varphi(x) = f(|x|)$.

4.103 ¶. Show that

$$|\mathbf{D}u|^2 = \operatorname{tr}(\mathbf{D}u)^2 + |\operatorname{rot} u|^2,$$
$$|\operatorname{rot} u|^2 = |\, u \bullet \operatorname{rot} u\,|^2 + |u \times \operatorname{rot} u|^2,$$
$$\operatorname{tr}(\mathbf{D}u)^2 - (\operatorname{div} u)^2 = \operatorname{div}\left((\nabla u)u - (\operatorname{div} u)u\right).$$

4.104 ¶. Let U be an admissible open set in $\mathbb{R}^n$ and η an $(n-2)$-form with C^2 coefficients in an open neighborhood of $\overline{U}$. Show that

$$\int_{\partial U} d\eta = 0.$$

5. Measures and Integration

In this chapter we deal with the construction and the properties of Lebesgue measure and Lebesgue integral and, more generally, with the abstract measure and integration theory, providing proofs and details that we avoided in Chapter 2 of [GM4].

The chapter is organized as follows. In Section 5.1 we present a detailed construction of Lebesgue's measure (including measurable sets, Cantor-type sets and nonmeasurable sets) and Vitali's characterization of Riemann integrable functions. We then extend the analysis of the process of construction of Lebesgue's measure in view of the discussion of general measures, starting from Carathéodory's characterization of measurable sets.

In Section 5.2 we deal with integration with respect to a measure and the two fundamental theorems for the calculus of integral for functions in several variables, Fubini theorem and the theorem of change of variables are discussed in Sections 5.3 and 5.4.

5.1 Measures

5.1.1 Set functions and measures

Let us begin with a few definitions. Here X will denote a generic set. We recall that for a generic subset E of X, $E^c := X \setminus E$ denotes the *complement* of E in X and $\mathcal{P}(X)$ denotes the family of all subsets of X. A family $\mathcal{E}$ of subsets of X is then a subset of $\mathcal{P}(X)$, $\mathcal{E} \subset \mathcal{P}(X)$.

A *set function* on X is a couple $(\mathcal{E}, \mu)$ of a family of subsets $\mathcal{E}$ of X that contains the empty set and of a nonnegative function $\mu : \mathcal{E} \to \overline{\mathbb{R}}_+$ such that $\mu(\emptyset) = 0$. A set function $(\mathcal{E}, \mu)$ on X is said to be

(i) *monotone* if for all $E, F \in \mathcal{E}$ with $E \subset F$, we have $\mu(E) \leq \mu(F)$,
(ii) *additive* if for every finite family of pairwise disjoint sets $E_1, \dots E_N \in \mathcal{E}$ with $\cup_k E_k \in \mathcal{E}$, we have $\mu(\cup_k E_k) = \sum_{k=1}^N \mu(E_k)$,
(iii) *σ-additive* or *countably additive* if for every sequence of pairwise disjoint subsets $\{E_k\} \subset \mathcal{E}$ with $\cup_k E_k \in \mathcal{E}$, we have $\mu(\cup_k E_k) = \sum_{k=1}^\infty \mu(E_k)$,

M. Giaquinta and G. Modica, *Mathematical Analysis, Foundations and Advanced Techniques for Functions of Several Variables*, DOI 10.1007/978-0-8176-8310-8_5,

(iv) *σ-subadditive* or *countably subadditive* if for every family of subsets $\{E_k\} \subset \mathcal{E}$ with $\cup_k E_k \in \mathcal{E}$, we have $\mu(\cup_k E_k) \leq \sum_{k=1}^{\infty} \mu(E_k)$.

The symbol $\sum_{k=1}^{\infty} \mu(E_k)$ is the sum of the series of positive terms $\{\mu(E_k)\}$. Recall, see [GM2], that the sum of a series of positive terms exists finite or $+\infty$ and is invariant under reordering.

5.1 Definition. *An* outer measure μ^* *on a set X is a set function $(\mathcal{P}(X), \mu^*)$ that is monotone and σ-subadditive.*

It is convenient to set also a language to denote families of subsets of a set, that are stable under union, intersection or under passage to complement. We say that a family $\mathcal{A} \subset \mathcal{P}(X)$ of subsets of a set X is an *algebra* if $\emptyset, X \in \mathcal{A}$ and $E \cup F$, $E \cap F$ and $E^c \in \mathcal{A}$ whenever $E, F \in \mathcal{A}$.

We say that $\mathcal{A}$ is a *σ-algebra* if $\mathcal{A}$ is an algebra and for every sequence of subsets $\{E_k\} \subset \mathcal{A}$, we also have $\cup_k E_k$ and $\cap_k F_k \in \mathcal{A}$. In other words, if we operate on sets of a σ-algebra with differences, countable unions or intersections, we get sets of the same σ-algebra; we also say that a σ-algebra is *closed* with respect to differences, countable unions and intersections.

Let $\mathcal{A} \subset \mathcal{P}(X)$ be a family of subsets of X. It is readily seen that the class

$$\mathcal{B} := \bigcap \Big\{ \mathcal{C} \;\Big|\; \mathcal{C} \supset \mathcal{A},\ \mathcal{C} \text{ is a } \sigma\text{-algebra} \Big\}$$

is again a σ-algebra, hence *the smallest σ-algebra containing $\mathcal{A}$*. We say that $\mathcal{B}$ is the σ-algebra *generated* by $\mathcal{A}$.

The smallest σ-algebra $\mathcal{B} \subset \mathcal{P}(\mathbb{R}^n)$ containing the open sets of $\mathbb{R}^n$ is called the σ-algebra of *Borel sets*.

5.2 ¶. Let $\mathcal{A} \subset \mathcal{P}(X)$ be a family of subsets of X and $\mathcal{B}$ the σ-algebra generated by $\mathcal{A}$. Show that

$$\mathcal{B} = \Big\{ B \;\Big|\; B \in \mathcal{S} \text{ for all } \sigma\text{-algebra } \mathcal{S} \supset \mathcal{A} \Big\}.$$

5.3 Definition. *A* measure *on a set X is the couple $(\mathcal{E}, \mu)$ of a σ-algebra $\mathcal{E} \subset \mathcal{P}(X)$ of subsets of X and of a σ-additive set function $\mu : \mathcal{E} \to \overline{\mathbb{R}}_+$.*

We also say that μ is a measure on the measurable space $(X, \mathcal{E})$, or that $(X, \mathcal{E}, \mu)$ is a measure space.

We explicitly state that, if $(\mathcal{E}, \mu)$ is a measure, then we have the following:

(i) $\mathcal{E}$ is a σ-algebra.
(ii) $\mu(\emptyset) = 0$.
(iii) For $E, F \in \mathcal{E}$ with $E \subset F$ we have $\mu(E) \leq \mu(F)$.
(iv) If $\{E_k\} \subset \mathcal{E}$, then $\mu(\cup_k E_k) \leq \sum_{k=1}^{\infty} \mu(E_k)$.
(v) If $\{E_k\} \subset \mathcal{E}$ is a disjoint sequence, then $\mu(\cup_k E_k) = \sum_{k=1}^{\infty} \mu(E_k)$.

a. Continuity properties of measures

A sequence of subsets $\{E_k\}$ of X is said to be *increasing* if $E_k \subset E_{k+1}$ $\forall k$ and *decreasing* if $E_k \supset E_{k+1}$ $\forall k$. The following theorem of monotone convergence for measures is a consequence of the countable σ-additivity.

5.4 Theorem (Monotone convergence for measures). *Let $(\mathcal{E}, \mu)$ be a measure on X. We have the following:*

(i) *If $\{E_k\}$ is an increasing sequence of sets in $\mathcal{E}$, then $\lim_{k\to\infty} \mu(E_k) = \mu(\cup_k E_k)$.*
(ii) *If $\{E_k\}$ is a decreasing sequence of sets in $\mathcal{E}$ and $\mu(E_1) < +\infty$, then $\lim_{k\to\infty} \mu(E_k) = \mu(\cap_k E_k)$.*

Proof. (i) From $\mu(E_k) \le \mu(\cup_k E_k)$ for every k, the claim is trivial if $\mu(E_k) = +\infty$ for some k. We may therefore assume $\mu(E_k) < \infty$ for all k. We set $E := \cup_k E_k$ and decompose E as

$$E = E_1 \bigcup \Big(\bigcup_{k=2}^{\infty} (E_k \setminus E_{k-1}) \Big).$$

The sets E_1 and $E_k \setminus E_{k-1}$, $k \ge 1$, are, of course, in $\mathcal{E}$ and pairwise disjoint. Because of the σ-additivity of μ, we then have

$$\mu(E) = \mu(E_1) + \sum_{k=2}^{\infty} \mu(E_k \setminus E_{k-1}) = \mu(E_1) + \sum_{k=2}^{\infty} (\mu(E_k) - \mu(E_{k-1})) = \lim_{k\to\infty} \mu(E_k).$$

(ii) Since $\mu(E_1) < +\infty$ and $E_k \subset E_1$, we have $\mu(E_k) = \mu(E_1) - \mu(E_1 \setminus E_k)$ for all k. Since $\{E_1 \setminus E_k\}$ is an increasing sequence of sets, we deduce from (i) that

$$\mu(E_1) - \lim_{k\to\infty} \mu(E_k) = \lim_{k\to\infty} \mu(E_1 \setminus E_k) = \mu\Big(\bigcup_k (E_1 \setminus E_k) \Big) = \mu(E_1) - \mu\Big(\bigcap_k E_k \Big).$$

□

5.1.2 Lebesgue's measure

a. Lebesgue's outer measure

An n-dimensional interval of extreme points $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ in $\mathbb{R}^n$, $a_i < b_i$ $\forall i$, is the semiclosed plurirectangle of $\mathbb{R}^n$

$$I = I(a, b) := \Big\{ x \in \mathbb{R}^n \;\Big|\; a_i < x \le b_i,\ i = 1, \dots, n \Big\} = \prod_{i=1}^{n}]a_i, b_i].$$

We denote its volume by $|I|$, and, of course,

$$|I| = \prod_{i=1}^{n} (b_i - a_i).$$

It is convenient to consider semiclosed intervals since they stack nicely; we can put them along without intersections to form new figures and cover,

for instance, the whole of $\mathbb{R}^n$. The family of intervals of $\mathbb{R}^n$ is denoted by $\mathcal{I}$.

Given any subset $E \subset \mathbb{R}^n$, we cover E by intervals; the sum of the elementary volumes of these intervals gives us an estimate from above of the "measure of E". Allowing denumerable coverings, we then set the following.

5.5 Definition. *The* Lebesgue's outer measure *of a subset $E \subset \mathbb{R}^n$ is defined as*

$$\mathcal{L}^{n*}(E) := \inf\Big\{\sum_{k=1}^{\infty} |I_k| \;\Big|\; I_k \in \mathcal{I}, E \subset \bigcup_{k=1}^{\infty} I_k\Big\}.$$

5.6 Theorem. *$(\mathcal{P}(\mathbb{R}^n), \mathcal{L}^{n*})$ is actually an outer measure in $\mathbb{R}^n$. Moreover, $\mathcal{L}^{n*}(I)$ is the elementary measure $|I|$ of I, if I is an interval.*

Proof. According to the properties of the infimum, clearly $\mathcal{L}^{n*}$ is a monotone set function defined on $\mathcal{P}(X)$. Let us prove that $\mathcal{L}^{n*}$ is σ-subadditive. Consider a family $\{E_j\} \subset \mathcal{P}(X)$; it is not restrictive to assume $\mathcal{L}^{n*}(E_j) < \infty$ for all j. For any given $\epsilon > 0$, for every j we consider a denumerable covering $\{I_k^{(j)}\}$ of E_j made by intervals such that $\sum_{k=1}^{\infty} \mathcal{L}^{n*}(I_k^{(j)}) \le \mathcal{L}^{n*}(E_j) + \epsilon 2^{-j}$. Then $\{I_k^{(j)}\}_{k,j}$ is a denumerable covering of $\cup_j E_j$ and we have

$$\sum_{k,j} |I_k^{(j)}| \le \sum_{j=1}^{\infty} \Big(\sum_{k=1}^{\infty} |I_k^{(j)}| \Big) \le \sum_{j=1}^{\infty} \Big(\mathcal{L}^{n*}(E_j) + \epsilon 2^{-j} \Big) = \sum_{j=1}^{\infty} \mathcal{L}^{n*}(E_j) + \epsilon,$$

i.e., $\mathcal{L}^{n*}(\cup_j E_j) \le \sum_{j=1}^{\infty} \mathcal{L}^{n*}(E_j) + \epsilon$. This proves the σ-additivity of $\mathcal{L}^{n*}$, ϵ being arbitrary. Therefore, $(\mathcal{P}(\mathbb{R}^n), \mathcal{L}^{n*})$ is an outer measure.

Let I be an interval. From the definition of $\mathcal{L}^{n*}$ we have $\mathcal{L}^{n*}(I) \le |I|$. Given $\epsilon > 0$, consider a denumerable covering $\{I_k\}$ of Imade by intervals, $\cup_{k=1}^{\infty} I_k \supset I$ such that $\sum_{k=1}^{\infty} |I_k| < \mathcal{L}^{n*}(I) + \epsilon$. For each k we choose an interval J_k that contains strictly I_k such that $|J_k| \le |I_k| + \epsilon\, 2^{-k}$. The family $\{\mathrm{int}(J_k)\}_k$ is an open covering of the compact set $\overline{I}$, hence there are finitely many indices $k_1, k_2, \dots k_N$ such that $I \subset \cup_{i=1}^{N} \mathrm{int}(J_{k_i})$. Since, of course, $|I| \le \sum_{i=1}^{N} |I_{k_i}|$, we have

$$|I| \le \sum_{k=1}^{\infty} |J_k| \le \sum_{k=1}^{\infty} (|I_k| + \epsilon 2^{-k}) = \sum_{k=1}^{\infty} |I_k| + \epsilon,$$

i.e., $|I| \le \mathcal{L}^{n*}(I)$. □

5.7 ¶. Prove that a point has zero Lebesgue outer measure. Since the denumerable union of sets of zero measure has zero outer measure, conclude that the set of rationals in $\mathbb{R}$, $\mathbb{Q} \subset \mathbb{R}$, has zero Lebesgue outer measure in $\mathbb{R}$.

5.8 ¶. Prove that the boundary of an interval in $\mathbb{R}^n$ has zero $\mathcal{L}^{n*}$ outer measure.

b. On the additivity of $\mathcal{L}^{n*}$

The Lebesgue outer measure has the advantage of being defined on *all* subsets of $\mathbb{R}^n$. However, it has the disadvantage of *not* being additive in general: There are *disjoint* sets E and F such that $\mathcal{L}^{n*}(E \cup F) < \mathcal{L}^{n*}(E) + \mathcal{L}^{n*}(F)$, see Corollary 5.22 and Proposition 5.23 below, although, in general, the following holds.

5.9 Proposition (Test of Carathéodory). *If for $E, F \subset \mathbb{R}^n$ we have*

$$\operatorname{dist}(E,F) := \inf\Big\{ |x-y| \;\Big|\; x \in E,\ y \in F \Big\} > 0,$$

then $\mathcal{L}^{n*}(E \cup F) = \mathcal{L}^{n*}(E) + \mathcal{L}^{n*}(F)$.

5.10 ¶. Prove Proposition 5.9. [*Hint.* Use the definition of $\mathcal{L}^{n*}$ after noticing that if $E = \cup_k I_k$ and $\delta > 0$, one can find disjoint intervals I'_j such that $E = \cup_j I'_j$, $\operatorname{diam}(I'_j) \le \delta$ $\forall j$ and $\sum_{j=1}^{\infty} |I'_j| \le \sum_{k=1}^{\infty} |I_k|$.]

c. Approximation by denumerable unions of intervals: Measurable sets

5.11 Definition. *A subset $E \subset \mathbb{R}^n$ is said to be* Lebesgue measurable *or* $\mathcal{L}^n$-measurable *or simply* measurable *if for all $\epsilon > 0$ there is a denumerable union of intervals P such that*

$$P \supset E \qquad \text{and} \qquad \mathcal{L}^{n*}(P \setminus E) < \epsilon.$$

The class of Lebesgue measurable sets is denoted by $\mathcal{M}$. For a measurable set E we write $\mathcal{L}^n(E)$ or $|E|$ instead of $\mathcal{L}^{n*}(E)$.

The family of measurable sets is quite wide. In fact, sets of zero measure, to which we shall refer to as *zero sets*, are measurable. Intervals and denumerable unions of intervals are also trivially measurable; since every open set is the union of denumerable many intervals, open sets are measurable, too.

However, there are nonmeasurable sets, see Theorem 5.21. In this respect, we notice that measurability is rather tricky. For instance, for every set $E \subset \mathbb{R}^n$ with $\mathcal{L}^{n*}(E) < +\infty$ and for every $\epsilon > 0$ we find a denumerable union of intervals $P = \cup_k I_k$, such that

$$\mathcal{L}^{n*}(P) = \sum_{k=1}^{\infty} |I_k| \le \mathcal{L}^{n*}(E) + \epsilon.$$

Since the outer measure is subadditive, we also know that $\mathcal{L}^{n*}(P) \le \mathcal{L}^{n*}(E) + \mathcal{L}^{n*}(P \setminus E)$; we have no easy way to infer that $\mathcal{L}^{n*}(P \setminus E) < \epsilon$.

Finally, notice that Definition 5.11 is quite natural in view of the following claim.

5.12 Proposition. *A subset $E \subset \mathbb{R}^n$ with $\mathcal{L}^{n*}(E) < +\infty$ is measurable if and only if for all $\epsilon > 0$ there is a finite union of intervals F such that $\mathcal{L}^{n*}(F \Delta E) < \epsilon$.*[1]

The point is that, in general, F does not contain E and is not contained in it.

5.13 ¶. Prove Proposition 5.12.

[1] Recall that the *symmetric difference* of two sets $A, B \subset \mathbb{R}^n$ is defined by $A \Delta B := (A \setminus B) \cup (B \setminus A)$.

d. Measurable sets and additivity

The next theorem together with the simple corollary that follows illustrate the structural properties of measurable sets and the behavior of the Lebesgue outer measure on the family of measurable sets.

5.14 Theorem. *The class of Lebesgue measurable sets $\mathcal{M}$ is a σ-algebra of subsets of $\mathbb{R}^n$ and $(\mathcal{M}, \mathcal{L}^{n*})$ is a measure in $\mathbb{R}^n$. Moreover, E is measurable if and only if for all $\epsilon > 0$ there are a closed set F and an open set A with $F \subset E \subset A$ such that $\mathcal{L}^n(A \setminus F) < \epsilon$.*

5.15 Corollary. *Let $E \subset \mathbb{R}^n$. The following claims are equivalent:*

(i) *E is Lebesgue measurable.*
(ii) *For all $\epsilon > 0$ there is an open set A with $A \supset E$ such that $\mathcal{L}^{n*}(A \setminus E) < \epsilon$.*
(iii) *For all $\epsilon > 0$ there is a closed set F with $F \subset E \subset A$ such that $\mathcal{L}^{n*}(E \setminus F) < \epsilon$.*
(iv) *There exists a decreasing sequence of open sets $\{A_k\}$ containing E and a zero set N such that $E = (\cap_k A_k) \setminus N$.*
(v) *There exists an increasing sequence of closed sets $\{F_k\}$ contained in E and a zero set N such that $E = (\cup_k F_h) \cup N$.*

5.16 ¶. Prove that E is measurable if and only if there is a Borel set B such that $B \supset E$ and $\mathcal{L}^{n*}(B \setminus E) = 0$, or if and only if there is a Borel set C such that $C \subset E$ and $\mathcal{L}^{n*}(E \setminus C) = 0$.

Deduce that $\mathcal{M}$ is the σ-algebra generated by the open sets (or the intervals) and the null sets for $\mathcal{L}^{n*}$.

Proof of Theorem 5.14. We outline the proof and leave to the reader the task of completing it specifying all needed details. We proceed by steps, noticing that we already know the first three steps.

(i) *Sets of zero outer measure are measurable.*

(ii) *Open sets are measurable.*

(iii) *E is measurable if and only if for all $\epsilon > 0$ there is an open set $A \supset E$ with $\mathcal{L}^{n*}(A \setminus E) < \epsilon$.*

(iv) *The denumerable union of measurable sets is measurable.* Given $\epsilon > 0$ for each E_k, we choose an open set $A_k \supset E_k$ with $\mathcal{L}^{n*}(A_k \setminus E_k) < \epsilon 2^{-k}$. Then $\cup_k A_k$ is open, $\cup_k A_k \supset \cup_k E_k$ and $\mathcal{L}^{n*}(\cup_k A_k \setminus \cup_k E_k) = \mathcal{L}^{n*}(\cup_k (A_k \setminus E_k)) \leq \epsilon$.

(v) *Let $\{I_k\}$ be a finite number of disjoint intervals. We have $\mathcal{L}^{n*}(\cup_{k=1}^N I_k) = \sum_{k=1}^N \mathcal{L}^{n*}(I_k)$.* For each k, let J_k be an interval that is strictly contained in I_k with $\mathcal{L}^{n*}(I_k) = |I_k| \leq |J_k| + \epsilon 2^{-k} = \mathcal{L}^{n*}(J_k) + \epsilon 2^{-k}$ so that the intervals J_k have a positive distance from each other. Proposition 5.9 then yields

$$\sum_{k=1}^N \mathcal{L}^{n*}(I_k) = \sum_{k=1}^N |I_k| \leq \sum_{k=1}^N |J_k| + \epsilon = \mathcal{L}^{n*}\Big(\bigcup_k J_k\Big) + \epsilon \leq \mathcal{L}^{n*}\Big(\bigcup_k I_k\Big) + \epsilon,$$

and, ϵ being arbitrary and $\mathcal{L}^{n*}$ subadditive, we conclude

$$\sum_{k=1}^N \mathcal{L}^{n*}(I_k) = \mathcal{L}^{n*}\Big(\bigcup_{k=1}^N I_k\Big).$$

(vi) *Closed sets are measurable.* Let F be a compact set and, for $\epsilon > 0$, let A be an open set with $\mathcal{L}^{n*}(A) \leq \mathcal{L}^{n*}(F) + \epsilon$. Since $A \setminus F$ is open, we can write $A \setminus F = \cup_k I_k$, where I_k are intervals that do not overlap and $\mathcal{L}^{n*}(A \setminus F) \leq \sum_k |I_k|$. In order to prove that F is measurable, it suffices to show that $\sum_k |I_k| < \epsilon$. We have

$$A = F \cup \Big(\bigcup_{i=1}^{\infty} I_k \Big) \supset F \bigcup \Big(\bigcup_{i=1}^{N} I_k \Big).$$

Since F and $\cup_{i=1}^{N} I_k$ are disjoint and compact, we infer $\operatorname{dist}(F, \cup_1^N I_k) > 0$, hence, by Proposition 5.9,

$$\mathcal{L}^{n*}(A) \geq \mathcal{L}^{n*}\Big(F \bigcup \Big(\bigcup_{1}^{N} I_k \Big)\Big) = \mathcal{L}^{n*}(F) + \mathcal{L}^{n*}\Big(\bigcup_{1}^{N} I_k \Big)$$

and, because of (iv),

$$\sum_{k=1}^{N} |I_k| = \mathcal{L}^{n*}\Big(\bigcup_{1}^{N} I_k \Big) \leq \mathcal{L}^{n*}(A) - \mathcal{L}^{n*}(F) \leq \epsilon \qquad \forall N.$$

This shows that compact sets are measurable. Since every closed set F is the denumerable union of compact sets, $F = \cup_k F_k$ with $F_k := \{x \in F \,\big|\, |x| \leq k\}$, we conclude that $F = \cup_k F_k$ is measurable by (ii).

(vii) *The complement of a measurable set is measurable.* Assume E is measurable and for every integer k choose an open set A_k with $E \subset A_k$ and $\mathcal{L}^{n*}(A_k \setminus E) < 1/k$. Then

$$E^c = \Big(\bigcup_k A_k^c \Big) \bigcup \Big(E^c \setminus \bigcup_k A_k^c \Big).$$

From (iv) and (vi) we infer that $\cup_k A_k^c$ is measurable. On the other hand, we have

$$E^c \setminus \Big(\bigcup_k A_k^c \Big) \subset E^c \setminus A_k^c = A_k \setminus E \qquad \forall k,$$

hence

$$\mathcal{L}^{n*}\Big(E^c \setminus \bigcup_k A_k^c \Big) = 0,$$

i.e., E^c is the union of a measurable and a zero set. This shows that E^c is measurable. Of course, claims (iv) and (vii) prove that $\mathcal{M}$ is a σ-algebra.

(viii) *E is measurable if and only if for all $\epsilon > 0$ there is a closed set F contained in E such that $\mathcal{L}^{n*}(E \setminus F) < \epsilon$.* This follows from (iii) applied to E^c and proves the last part of the claim.

(ix) *The outer measure is σ-additive on $\mathcal{M}$.* Let $\{E_k\}$ be a family of disjoint sets. Assume that they are also bounded. For all k we find an open set A_k and a closed set F_k with $F_k \subset E_k \subset A_k$ and $\mathcal{L}^{n*}(A_k \setminus F_k) < \epsilon\, 2^{-}k$. Moreover, the sets F_k are compact and disjoint, thus with positive distance. Proposition 5.9 then implies

$$\mathcal{L}^{n*}\Big(\bigcup_{k=1}^{N} F_k \Big) = \sum_{k=1}^{N} \mathcal{L}^{n*}(F_k) \qquad \forall N,$$

consequently,

$$\mathcal{L}^{n*}\Big(\bigcup_k E_k \Big) \geq \sum_{k=1}^{\infty} \mathcal{L}^{n*}(F_k) \geq \sum_{k=1}^{\infty} (\mathcal{L}^{n*}(E_k) - \epsilon\, 2^{-k}) = \sum_{k=1}^{\infty} \mathcal{L}^{n*}(E_k) - \epsilon,$$

from which we easily get

$$\mathcal{L}^{n*}\Big(\bigcup_{k=1}^{\infty} E_k\Big) = \sum_{k=1}^{\infty} \mathcal{L}^{n*}(E_k).$$

In the case that the E_k's are not necessarily bounded, it suffices to consider

$$E_{k,j} := E_k \cap (\overline{B_j} \setminus B_{j-1}),$$

where B_j is the open ball centered at the origin with radius j and to compute

$$\begin{aligned}\mathcal{L}^{n*}\Big(\bigcup_k E_k\Big) &= \mathcal{L}^{n*}\Big(\bigcup_{k,j} E_{k,j}\Big) = \sum_{k,j} \mathcal{L}^{n*}(E_{k,j}) \\ &= \sum_{k=1}^{\infty}\Big(\sum_{j=1}^{\infty} \mathcal{L}^{n*}(E_{k,j})\Big) = \sum_{k=1}^{\infty} \mathcal{L}^{n*}(E_k).\end{aligned}$$

□

5.17 Definition. *The measure $(\mathcal{M}, \mathcal{L}^{n*})$ is called the n-dimensional* Lebesgue measure *in $\mathbb{R}^n$. For $E \in \mathcal{M}$ we write $\mathcal{L}^n(E)$, instead of $\mathcal{L}^{n*}(E)$ or even $|E|$ when the dimension n is clear from the context.*

For the reader's convenience we collect in the next proposition some of the properties of the Lebesgue measure and Lebesgue measurable sets that are simple consequences of Theorem 5.14.

5.18 Proposition. *We have the following:*

(i) $\emptyset$, $\mathbb{R}^n$ *are measurable sets.*
(ii) *If $E, F \in \mathcal{M}$, then $E \cup F$, $E \setminus F$, $E \cap F \in \mathcal{M}$ and $|E \cup F| + |E \cap F| = |E| + |F|$.*
(iii) *If $E, F \in \mathcal{M}$ and $E \subset F$, then $|E| \leq |F|$.*
(iv) *If $\{E_k\} \subset \mathcal{M}$ is a denumerable family of measurable sets, then $\cup_k E_k$, $\cap_k E_k \in \mathcal{M}$ and, if moreover the E_k's are pairwise disjoint, then*

$$\Big|\bigcup_{k=1}^{\infty} E_k\Big| = \sum_{k=1}^{\infty} |E_k|.$$

(v) *Intervals, denumerable unions of intervals, open and closed sets are measurable.*
(vi) *$E \in \mathcal{M}$ if and only if for all $\epsilon > 0$ there is an open set $A \supset E$ with $\mathcal{L}^{n*}(A \setminus E) < \epsilon$.*
(vii) *$E \in \mathcal{M}$ if and only if for all $\epsilon > 0$ there is a closed set $F \subset E$ with $\mathcal{L}^{n*}(E \setminus F) < \epsilon$.*
(viii) *If $\{E_k\} \subset \mathcal{M}$ is increasing, $E_k \subset E_{k+1}$ $\forall k$, then $\lim_{k\to\infty} |E_k| = |\bigcup_{k=1}^{\infty} E_k|$.*
(ix) *If $\{E_k\} \in \mathcal{M}$ is decreasing and $|E_1| < +\infty$, then $\lim_{k\to\infty} |E_k| = |\bigcap_{k=1}^{\infty} E_k|$.*

5.1.3 A few complements

Here we add a few comments about measurability according to Riemann and Lebesgue.

a. A Riemann nonintegrable function

We begin by selecting a disjoint and denumerable family of open intervals in $[0,1]$ as follows. Fix $0 < \sigma \leq 1/3$ and set $C_0 := [0,1]$. At step 0, we cut in the center of $[0,1]$ an open interval $A_{0,1}$ of length $\sigma < 1/3$; at step 1, in the center of each two remaining closed intervals, we cut two open intervals $A_{1,0}$ and $A_{1,1}$ of length $\sigma/3$. By induction, at step k, in the center of each of the 2^k closed intervals that remain from step k we cut 2^k open intervals $A_{k,j}$, $j = 0, \ldots, 2^k - 1$ of length $\sigma/3^k$. Of course, the open intervals $A_{k,j}$ are pairwise disjoint,

$$A_\sigma := \bigcup_{k=0}^{\infty} \bigcup_{j=0}^{2^k-1} A_{k,j}$$

is open and, furthermore, A_σ is dense in $[0,1]$. Finally,

$$\mathcal{L}^1(A_\sigma) = \sum_{k=0}^{\infty} \sum_{j=0}^{2^k-1} |A_{k,j}| = \sum_{k=0}^{\infty} 2^k \frac{\sigma}{3^k} = \frac{\sigma}{1-2/3} = 3\sigma.$$

The set $C_\sigma := [0,1] \setminus A_\sigma$ is clearly compact, measures $1 - 3\sigma$ according to Lebesgue's measure and does not contain intervals. Actually, C_σ is *perfect* (i.e., all of its points are accumulation points for C_σ) and it is not denumerable.

Denote by $f : \mathbb{R} \to \mathbb{R}$ the characteristic function of A_σ,

$$f(x) := \begin{cases} 1 & \text{if } x \in A_\sigma, \\ 0 & \text{if } x \notin A_\sigma. \end{cases}$$

We claim that f is not Riemann integrable if $\sigma < 1/3$. In fact, since $f(x) = 1$ on A_σ that is dense in $[0,1]$, the upper Riemann integral of f is 1. On the other hand, since every minorant of f may differ from zero only on A_σ, the lower Riemann integral of f cannot be larger than and, actually, equal to

$$\sum_{k=0}^{\infty} \sum_{j=0}^{2^k-1} |A_{k,j}| = \mathcal{L}^1(A_\sigma) = 3\sigma.$$

Consequently, f is not Riemann integrable in $[0,1]$. We emphasize that f is the characteristic function of the union of denumerable open intervals that are pairwise disjoint.

b. Cantor set

We have already discussed the autosimilarity of the Cantor set, see [GM2]. For the reader's convenience we review the construction.

Let $0 < \delta < 1/2$ and $E_0 := [0,1]$. At step 0, we cut in the center of $[0,1]$ the open interval A_0 of size $(1 - 2\delta)$ and set $E_1 := [0,1] \setminus A_0$. At step 1,

we cut at the center of the two remaining intervals two open intervals $A_{1,0}$ and $A_{1,1}$ of length $\delta(1-2\delta)$ and set $E_2 := E_1 \setminus (A_{1,0} \cup A_{1,1})$. By induction, at step k, we cut at the center of the 2^k remaining closed intervals 2^k open intervals $A_{k,j}$ of length $\delta^k(1-2\delta)$ and set $E_{k+1} := E_k \setminus (\cup_{k=0}^{2^k-1} A_{k,j})$. The *Cantor set* C_δ is then defined by

$$C_\delta := \bigcap_k E_k$$

or, equivalently, as

$$C_\delta = [0,1] \setminus A_\delta \qquad \text{where} \qquad A_\delta := \bigcup_{k=0}^{\infty} \bigcup_{j=0}^{2^k-1} A_{k,j}.$$

By construction, A_δ is an open dense set in $[0,1]$ made by a disjoint union of denumerable open intervals, hence

$$|A_\delta| = \sum_{k=0}^{\infty} 2^k \delta^k (1-2\delta) = 1.$$

Therefore, the Cantor set C_δ is nonempty, compact with no interior points and zero measure, $|C_\delta| = 0$.

c. Cantor ternary set

In the case $\delta = 1/3$, C_δ has another description. Represent the numbers in [0,1] in basis 3, i.e., with digits 0, 1 and 2. It is readily seen that $x \in A_k = \cup_j A_{k,j}$ if and only if the $k+1$ digit of x is 1, hence

$$C_{1/3} = \Big\{ x = \sum_{k=1}^{\infty} \frac{\alpha_k}{3^k} \;\Big|\; a_k \in \{0,2\} \Big\}.$$

In particular, see [GM2], $C_{1/3}$ is not countable.

5.19 ¶. Show that there are Lebesgue measurable sets that are not Borel sets. [*Hint.* The class of Borel sets is generated from the intervals with rational extreme points, hence it has the power of reals. Parts of the Cantor set $C_{1/3}$ are measurable and form a set with power larger than the power of $\mathbb{R}$, since the cardinality of $C_{1/3}$ equals the cardinality of $\mathbb{R}$.]

d. Cantor–Vitali function

The Cantor–Vitali function is in some sense naturally associated to the ternary set $C := C_{1/3}$. With reference to the notation used up to now we set $A_k := \cup_{j=0}^{2^k-1} A_{k,j}$, i.e., the union of the intervals cut at step k and $E_0 := [0,1]$, $E_{k+1} := E_k \setminus A_k$. We define $f_k : [0,1] \to [0,1]$ as the continuous function

$$f_k(0) = 0,\ f_k(1) = 1,$$
$$f_k(x) = \frac{j+1}{2^{h+1}} \text{ if } x \in A_{h,j},\ j = 0, \dots, 2^h - 1, h = 1, \dots, k;$$

f_k is linear on each interval of $[0,1] \setminus A_k$. In formula,

$$f_k(x) = \left(\frac{3}{2}\right)^k \int_0^x \chi_{E_k}(t)\, dt.$$

By construction, f_k is piecewise linear, increasing, $f_{k+1} = f_k$ on each $A_{k,j}$ and $|f_k - f_{k+1}| \le 2^{-k}$. In particular, the series $\sum_{k=1}^{\infty}(f_k - f_{k-1})$ converges uniformly in $[0,1]$. If we denote by f the uniform limit of $\{f_k\}$,

$$f(x) := \lim_{k\to\infty} f_k(x),$$

we then have $f(0) = 0$, $f(1) = 1$, f is continuous and increasing in $[0,1]$ and f is constant on each interval $A_{k,j}$ of $[0,1] \setminus C$. Since each $A_{k,j}$ is open, f has zero derivative in $[0,1] \setminus C$. Notice that the image of C via f covers $[0,1]$ except for the denumerable set $\{j/2^k \,|\, j = 0, \dots, 2^k,\ k \ge 1\}$.

Actually, the approximating functions $\{f_k\}$ are equi-Hölder-continuous with exponent $\alpha := \log 2/\log 3$. To prove this, it suffices to show that f_k is Hölder-continuous with exponent α in $[0, 3^{-k}]$. Since in $[0, 3^{-k}]$, f_k is linear with slope $(3/2)^k$, we compute for $0 \le x < y \le 3^{-k}$

$$\begin{aligned} |f_k(x) - f_k(y)| &\le \left(\frac{3}{2}\right)^k |x-y| = \left(\frac{3}{2}\right)^k |x-y|^{1-\alpha}|x-y|^{\alpha} \\ &\le \left(\frac{3}{2}\right)^k \left(\frac{1}{3}\right)^{k(1-\alpha)} |x-y|^{\alpha} = |x-y|^{\alpha}, \end{aligned}$$

as $3^{-\alpha} = 1/2$. A consequence is that *the Cantor–Vitali function is Hölder-continuous with exponent* α.

e. Lebesgue nonmeasurable sets

In spite of the generality of the notion of Lebesgue measurable sets, there still exist nonmeasurable sets in the sense of Lebesgue. Examples are not constructive and involve the *axiom of choice*, see [GM2]. This is not fortuitous; in fact, Robert Solovay (1938–) has proved that the existence of a nonmeasurable set implies the validity of the axiom of choice.

First we state the following.

5.20 Lemma. *Let $E \subset \mathbb{R}$ be a measurable set with $|E| > 0$. Then the set of differences $\{x - y \,|\, x, y \in E\}$ contains an open interval $]-\delta, \delta[$, $\delta > 0$.*

Proof. Of course, we may assume that $|E| < +\infty$. Fix $\epsilon > 0$ and let A be an open set with $E \subset A$ and $|A| < (1+\epsilon)|E|$. We may also assume that A is a denumerable union of intervals (that are left-open and right-closed) that are disjoint, $A = \cup_k I_k$. We then set $E_k := E \cap I_k$. Of course, since E is measurable, $|E| = \sum_{k=1}^{\infty} |E_k|$, and, since

SULL'INTEGRAZIONE PER SERIE.

Memoria di G. Vitali (Genova).

Vari lavori recenti hanno messo in evidenza l'importanza che nello studio delle funzioni ha la trasformazione proposta dal signor Henri Lebesgue dei concetti noti sotto il nome *di misura di un gruppo di punti* e *di integrale di una funzione* *).

I nuovi concetti si sono mostrati indispensabili per lo studio delle questioni più fondamentali del Calcolo, da tanto tempo già poste, come quelle che riguardano la costruzione delle funzioni primitive e l'esistenza delle derivate.

Gli integrali di Lebesgue si presentano come una naturale generalizzazione dell'integrale di Riemann, non solo per la maggior vastità del campo delle funzioni che risultano integrabili **), ma anche per la maggior varietà dei campi a cui si può estendere una integrazione.

Invero l'integrazione di Lebesgue estesa ad un segmento si può considerare come un caso specialissimo dell'integrazione estesa ad un gruppo misurabile di punti.

In particolare il problema dell'integrazione per serie non può più essere considerato soltanto dal punto di vista speciale dell'integrazione estesa a dei segmenti, ma si deve studiare da quello più generale dell'integrazione estesa a qualunque gruppo misurabile di punti.

*) *Leçons sur l'intégration et la recherche des fonctions primitives*, par Henri Lebesgue (Paris, Gauthier-Villars, 1904) e G. Vitali, *Sui gruppi di punti* (questi Rendiconti, t. XVIII (1904), pp. 116-126).

**) Veramente una funzione illimitata può avere un integrale nel senso comune della parola e non un integrale di Lebesgue. Ma la comune definizione degli integrali di tali funzioni filmbare non è meno arbitraria di quel che sarebbe il definire la somma di una serie non convergente colla somma della serie che si ottiene aggruppando in un modo stabilito una volta per sempre i termini della data. Del resto potremmo in modo analogo estendere il significato di integrale di Lebesgue profittando una successione di gruppi

$$G_1, G_2, \ldots G_n, \ldots$$

in cui la funzione è sommabile, ciascuno contenuto nel successivo, e tendente al campo di integrazione, ed assumere per valore dell'integrale il limite (quando esiste) dell'integrale esteso a G_n per $n = \infty$. Ma ciò non sembra opportuno, perchè dal punto di vista generale non è preferibile una successione piuttosto che un'altra, e cambiando la scelta della successione può venire a cambiare il valore che si attribuisce all'integrale.

Rend. Circ. Matem. Palermo, t. XXIII (1907). — Stampato il 3 Ottobre 1907.

LEÇONS

SUR LA

THÉORIE DES FONCTIONS

PAR

ÉMILE BOREL,

MAÎTRE DE CONFÉRENCES À L'ÉCOLE NORMALE SUPÉRIEURE.

PARIS,

GAUTHIER-VILLARS ET FILS, IMPRIMEURS-LIBRAIRES

DE L'OBSERVATOIRE DE PARIS ET DU BUREAU DES LONGITUDES,

Quai des Grands-Augustins, 55.

1908

Figure 5.1. A page from a paper of Giuseppe Vitali (1875–1932) and the frontispiece of *Théorie des Fonctions* by Emile Borel (1871–1956).

$$0 < (1+\epsilon)|E| - |A| = \sum_{k=1}^{\infty} ((1+\epsilon)|E_k| - |I_k|),$$

we can find k_ϵ such that $|E_{k_\epsilon}|(1+\epsilon) > |I_{k_\epsilon}|$. If we choose, for instance, $\epsilon := 1/3$, $I := I_{k_\epsilon}$ and $F := E_{k_\epsilon}$, we then have

$$F \subset I, \qquad F \subset E \qquad \text{and} \qquad |I| \le \frac{4}{3}\,|F|.$$

Finally, we choose $\delta := |I|/2$ and prove that *the translated* F^d *of* F *by any number* d *with* $|d| < \delta$ *has points in common with* F. In fact, suppose that for some d, $0 < |d| < \delta$, F and F^d were disjoint, then

$$2\,|F| = |F| + |F^d| = |F \cup F^d| \le |I| + |d| \le |I| + \delta = \frac{3}{2}|I|,$$

which contradicts $|I| < 4/3|F|$. In other words, we have proved that for all d with $|d| < \delta$ there exist $x, y \in F$ with $|x - y| = d$, and this is just the claim of the lemma. □

5.21 Theorem (Vitali). *There exist Lebesgue nonmeasurable sets in* $\mathbb{R}$.

Proof. We say that $x, y \in \mathbb{R}$ are *equivalent* if $x - y$ is rational and denote by $\mathcal{E}$ the set of equivalence classes defined this way. Of course, equivalent classes are disjoint: One is the class of all rationals and the others have a form such as $\{x = \sqrt{2} + r \,|\, r \in \mathbb{Q}\}$. Each class is denumerable, while $\mathcal{E}$ has the cardinality of $\mathbb{R}$. According to Zermelo's axiom of choice, we may consider a set $E \subset \mathbb{R}$ consisting of exactly one element from each distinct equivalence class. Since any two points of E must differ by an irrational, the numbers in the set $\Delta := \{x - y \,|\, x, y \in E\}$ cannot contain an interval; thus, according to Lemma 5.20, either E is not measurable or $|E| = 0$. Since the union of the rational translates of E is all of $\mathbb{R}$, it is excluded that $|E| = 0$ (as this would imply $|\mathbb{R}| = 0$). We conclude that E is not measurable. □

5.22 Corollary. *Any set $A \subset \mathbb{R}$ with $\mathcal{L}^{1*}(A) > 0$ contains a nonmeasurable set.*

Proof. Let E be Vitali's nonmeasurable set in the proof of Theorem 5.21 and E^r be the translated of E by r. As in the proof of Theorem 5.21, either $|A \cap E^r| = 0$ or $A \cap E^r$ is nonmeasurable since $\Delta = \{x - y \,|\, x, y \in A \cap E^r\}$ cannot contain an interval. On the other hand, A clearly decomposes as the denumerable union of disjoint sets

$$A = \bigcup_{r \in \mathbb{Q}} (A \bigcap E^r).$$

Since $\mathcal{L}^{1*}(A) > 0$, $A \cap E^r$ need to be not measurable for some r. □

Other examples of nonmeasurable sets are available. One that is close to Vitali's example is the following: The basis of $\mathbb{R}$ as vector space over $\mathbb{Q}$ is nonmeasurable. Zermelo's axiom is used in order to establish the existence of a basis of $\mathbb{R}$ over $\mathbb{Q}$.

The celebrated *Banach–Tarski paradox* states that we may split the unit ball of $\mathbb{R}^3$ in three disjoint pieces A, B and C, each congruent to the other (superimposable by means of a rotation and a translation) and each with (outer) measure equal to the measure of the entire ball. Of course, the three parts are nonmeasurable.

5.1.4 Abstract measures

a. Measurability according to Carathéodory

A first attempt to define a class $\mathcal{M}$ on which an outer measure is additive is to select all sets E for which the outer measure is indeed additive, i.e., the sets E such that

$$\mathcal{L}^{n*}(\Omega \cup E) = \mathcal{L}^{n*}(E) + \mathcal{L}^{n*}(\Omega)$$

for *all* subsets Ω that are disjoint from E, or, equivalently,

$$\mathcal{L}^{n*}(A) = \mathcal{L}^{n*}(A \cap E) + \mathcal{L}^{n*}(A \cap E^c) \qquad \forall A,\ A \supset E,$$

where $E^c := X \setminus E$. However, we would like $\mathcal{M}$ to be a σ-algebra. As proved by Carathéodory, a localization of the previous condition suffices to characterize Lebesgue measurable sets.

5.23 Proposition (Carathéodory). *E is Lebesgue measurable if and only if*

$$\mathcal{L}^{n*}(A \cap E) + \mathcal{L}^{n*}(A \cap E^c) = \mathcal{L}^{n*}(A) \qquad \forall A \subset \mathbb{R}^n. \tag{5.1}$$

In other words, Lebesgue measurable sets are those which split every set into pieces that are additive with respect to the outer measure.

Proof. Suppose E is measurable. For every measurable set I we, of course, have

$$\mathcal{L}^{n*}(I \cap E) + \mathcal{L}^{n*}(I \cap E^c) = \mathcal{L}^{n*}(I).$$

In view of the subadditivity of $\mathcal{L}^{n*}$, it suffices to prove

$$\mathcal{L}^{n*}(A \cap E) + \mathcal{L}^{n*}(A \cap E^c) \leq \mathcal{L}^{n*}(A) \qquad \forall A \subset \mathbb{R}^n,$$

in order to prove (5.1). It is not restrictive to assume, furthermore, that $\mathcal{L}^{n*}(A) < \infty$. For every $\epsilon > 0$, we choose a measurable set I such that $I \supset A$ and $\mathcal{L}^{n*}(I) \leq \mathcal{L}^{n*}(A)+\epsilon$. Then

$$\mathcal{L}^{n*}(A \cap E) + \mathcal{L}^{n*}(A \cap E^c) \leq \mathcal{L}^{n*}(I \cap E) + \mathcal{L}^{n*}(I \cap E^c) \leq \mathcal{L}^{n*}(I) \leq \mathcal{L}^{n*}(A) + \epsilon$$

and (5.1) is proved.

Conversely, suppose that (5.1) holds and, moreover, that E is bounded. For every $\epsilon > 0$ choose A open with $A \supset E$ and $\mathcal{L}^{n*}(A) < \mathcal{L}^{n*}(E) + \epsilon$; then (5.1) yields

$$\mathcal{L}^{n*}(A \setminus E) = \mathcal{L}^{n*}(A) - \mathcal{L}^{n*}(E) < \epsilon,$$

which implies that E is measurable. We leave the case of E unbounded to the reader as an exercise. □

Carathéodory's criterion in (5.1) is important because it characterizes measurable sets merely in terms of the outer measures and, therefore, suggests itself as a suitable definition of measurable sets for an arbitrary outer measure.

5.24 Definition (Carathéodory). *Let μ^* be an outer measure on a set X. We say that $E \subset X$ is* μ^*-measurable *if*

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c) \qquad \forall A \subset \mathbb{R}^n.$$

We denote by $\mathcal{M}_{\mu^*}$ the class of μ^*-measurable sets.

5.25 Theorem. *For an outer measure μ^* on X the class $\mathcal{M}_{\mu^*}$ of μ^*-measurable subsets of X is a σ-algebra and $(\mathcal{M}_{\mu^*}, \mu^*)$ is a measure on X.*

Proof. (i) First we observe that from Carathéodory's definition of measurable subsets we may easily infer that

a. μ^* null sets are μ^*-measurable,
b. E is μ^*-measurable if and only if E^c is μ^*-measurable,
c. E is μ^*-measurable if and only if

$$\mu^*(A \cap E) + \mu^*(A \cap E^c) \leq \mu^*(A)$$

for all $A \subset X$ with $\mu^*(A) < \infty$,
d. if E is μ^*-measurable and $A \supset E$ and $\mu^*(A \setminus E) < \infty$, then $\mu^*(E) = \mu^*(A) - \mu^*(A \setminus E)$.

(ii) Let us prove that *if $E, F \in \mathcal{M}_{\mu^*}$, then $E \cup F$, $E \cap F$, $E \setminus F \in \mathcal{M}_{\mu^*}$ and $\mu^*(E \cup F) = \mu^*(E) + \mu^*(F)$ if, moreover, E and F are disjoint.*

Let $A \subset X$ with $\mu^*(A) < \infty$. We have

$$\begin{aligned}\mu^*(A \cap (E \cup F)) + \mu^*(A \cap (E \cup F)^c) &\\ \leq \mu^*(A \cap E) + \mu^*(A \cap E^c \cap F) + \mu^*(A \cap E^c \cap F^c) &\\ = \mu^*(A \cap E) + \mu^*(A \cap E^c) = \mu^*(A), &\end{aligned}$$

hence $E \cup F \in \mathcal{M}_{\mu^*}$ according to (c) of (i). Since $E \cap F = (E^c \cup F^c)^c$ and $E \setminus F = E \cap F^c$, we also infer that $E \cap F$ and $E \setminus F \in \mathcal{M}_{\mu^*}$. Finally, the addition formula $\mu^*(E \cup F) = \mu^*(E) + \mu^*(F)$ follows from the μ^*-measurability of E.

(iii) If $E_1, E_2 \in \mathcal{M}_{\mu^*}$ are disjoint, for $\Omega := A \cap (E_1 \cup E_2)$ we have $\Omega \cap E_1 = A \cap E_1$ and $\Omega \cap E_1^c = A \cap E_2$. The measurability of E_1 then yields

$$\mu^*(A \cap E_1) + \mu^*(A \cap E_2) = \mu^*(A \cap (E_1 \cup E_2)) \qquad \forall A \subset X.$$

(iv) By induction from (ii) and (iii), we conclude that if $\{E_k\}$ is a disjoint family of μ^*-measurable sets, then for all integer N $\cup_{k=1}^N E_k \in \mathcal{M}_{\mu^*}$ and

$$\mu^*\Big(A \cap \bigcup_{k=1}^N E_k\Big) = \sum_{k=1}^N \mu^*(A \cap E_k) \qquad \forall A \subset X.$$

(v) Let us prove that $\cup_{k=1}^\infty E_k \in \mathcal{M}_{\mu^*}$ *if* $E_k \in \mathcal{M}_{\mu^*}$ $\forall k$. We may write $\cup_k E_k$ as the union of measurable and disjoint sets

$$\bigcup_k E_k = E_1 \cup \bigcup_{k=1}^\infty (E_{k+1} \setminus E_k).$$

Consequently, we may assume that the E_k's are measurable and pairwise disjoint. For all integers p and any $A \subset X$ (iv) yields

$$\mu^*\Big(A \cap \bigcup_{k=1}^p E_k\Big) + \mu^*\Big(A \cap \Big(\bigcup_{k=1}^p E_k\Big)^c\Big) \le \mu^*(A),$$

and

$$\mu^*\Big(A \cap \Big(\bigcup_{k=1}^p E_k\Big)\Big) = \mu^*\Big(\bigcup_{k=1}^p (A \cap E_k)\Big) = \sum_{k=1}^p \mu^*(A \cap E_k),$$

whereas

$$A \cap \Big(\bigcup_{k=1}^\infty E_k\Big)^c \subset A \cap \Big(\bigcup_{k=1}^p E_k\Big)^c \qquad \forall p.$$

Therefore,

$$\sum_{k=1}^\infty \mu^*(A \cap E_k) + \mu^*\Big(A \cap \Big(\bigcup_{k=1}^\infty E_k\Big)^c\Big) \le \mu^*(A)$$

and the σ-subadditivity finally yields

$$\mu^*\Big(A \cap \Big(\bigcup_{k=1}^\infty E_k\Big)\Big) + \mu^*\Big(A \cap \Big(\bigcup_{k=1}^\infty E_k\Big)^c\Big) \le \sum_{k=1}^\infty \mu^*(A \cap E_k) + \mu^*\Big(A \cap \Big(\bigcup_{k=1}^\infty E_k\Big)^c\Big) \le \mu^*(A).$$

(vi) Let us prove that μ^* *is* σ*-additive on* $\mathcal{M}_{\mu^*}$. Let $\{E_k\}$ be a disjoint family of measurable sets and let p be an integer. According to (iii),

$$\sum_{k=1}^p \mu^*(E_k) = \mu^*\Big(\bigcup_{k=1}^p E_k\Big) \le \mu^*\Big(\bigcup_{k=1}^\infty E_k\Big),$$

hence $\sum_{k=1}^\infty \mu^*(E_k) \le \mu^*(\cup_{k=1}^\infty E_k)$. The proof is then concluded since the opposite inequality follows from the σ-subadditivity of μ^*. □

5.26 ¶. Give an explicit characterization of $\mathcal{M}_\mu$ when μ is defined by
- $\mu(A) =$ # points (possibly ∞) in A,
- $\mu(A) = 1$ if $A \neq \emptyset$ and $\mu(\emptyset) = 0$.


Figure 5.2. Emile Borel (1871–1956) and Constantin Carathéodory (1873–1950).

b. Construction of measures: Method I

The process that led us to Lebesgue's measure extends to a more general setting. Let $\mathcal{I} \subset \mathcal{P}(X)$ be any family of subsets of X that contains the empty set and let $\alpha : \mathcal{I} \to \overline{\mathbb{R}}_+$ be any set function (i.e., any map with $\alpha(\emptyset) = 0$). Define a new set function $\mu^* : \mathcal{P}(X) \to \overline{\mathbb{R}}_+$ by setting for all $E \subset X$

$$\mu^*(E) := \inf\Big\{\sum_{i=1}^{\infty} \alpha(I_i) \,\Big|\, \cup_i I_i \supset E,\ I_i \in \mathcal{I}\Big\} \tag{5.2}$$

(we understand $\mu^*(E) = +\infty$ if there is no sequence $\{I_i\} \subset \mathcal{I}$ such that $\cup_i I_i \supset E$). It is not difficult to prove the following.

5.27 Proposition. *$\mu^* : \mathcal{P}(X) \to \overline{\mathbb{R}}$ is an outer measure on X.*

Consequently, by restricting μ^* to the μ^*-measurable sets, we define a measure $(\mathcal{M}_{\mu^*}, \mu^*)$ in X, see Theorem 5.25. However, in general, we have the following:

(i) The class of μ^*-measurable sets may reduce to $\{\emptyset, X\}$.
(ii) μ^* might not be in turn an extension of α.
(iii) The sets in $\mathcal{I}$ need not be μ^*-measurable.

However, in the following important case the previous irregularities do not occur.

5.28 Definition. *A family $\mathcal{I} \subset \mathcal{P}(X)$ is called a* semiring *if $\emptyset \in \mathcal{I}$, for all E and F in $\mathcal{I}$ we have $E \cap F \in \mathcal{I}$ and we can decompose $E \setminus F$ as $E \setminus F = \cup_{j=1}^{N} I_j$ with pairwise disjoint $I_j \in \mathcal{I}$.*

Examples of semirings are trivially the family of intervals in $\mathbb{R}^n$ or any algebra of subsets of a set X. Notice that if $E, F \in \mathcal{I}$, where $\mathcal{I}$ is a semiring, then $E \cup F = \cup_j I_j$ with pairwise disjoint $I_j \in \mathcal{I}$.

5.29 Theorem. *Let $\alpha : \mathcal{I} \to \overline{\mathbb{R}}_+$ be a σ-additive set function defined on a semiring $\mathcal{I} \subset \mathcal{P}(X)$. Then α is monotone and σ-subadditive on $\mathcal{I}$. Moreover, let μ^* be the outer measure defined by (5.2), $\mathcal{M}_{\mu^*}$ the corresponding class of μ^*-measurable sets and $(\mathcal{M}_{\mu^*}, \mu^*)$ the measure associated to μ^*. We have the following:*

(i) *μ^* extends α.*
(ii) *E is μ^*-measurable if and only if we have*

$$\mu^*(I \cap E) + \mu^*(I \cap E^c) \le \mu^*(I) \tag{5.3}$$

for all $I \in \mathcal{I}$ with $\mu^*(I) < +\infty$.
(iii) *$\mathcal{I} \subset \mathcal{M}_{\mu^*}$.*
(iv) *For all $E \subset X$ with $\mu^*(E) < +\infty$ there is a decreasing sequence of sets $\{F_k\}$ each being μ^*-measurable with $\mu^*(F_k) < \infty$ and the union of elements of $\mathcal{I}$, $F_k = \cup_j I_j^{(k)}$, such that $E \subset \cap_k F_k$ and $\mu^*(\cap_k F_k) = \mu^*(E)$.*

Proof. Monotonicity of α is trivial. Let us prove the σ-subadditivity of α on $\mathcal{I}$. In order to do it, for $I, I_k \in \mathcal{I}$, $I \subset \cup_k I_k$, we write $I = \cup_k H_k$, $H_k := I_k \cap I \in \mathcal{I}$ and decompose $\cup_k H_k$ as the disjoint union of

$$H_1, H_2 \setminus H_1, \dots, H_{p+1} \setminus \Big(\bigcup_{k=1}^{p} H_k \Big), \dots .$$

Since $\mathcal{I}$ is a semiring, each piece of the previous set is a finite union of disjoint sets in I. Hence $I = \cup_k H_k$ is the union of disjoint sets $\{J_j\} \subset \mathcal{I}$, where each J_j is contained in at least one of the $H_k = I \cap I_k$. The σ-additivity of α on $\mathcal{I}$ yields

$$\alpha(I) = \alpha\Big(\bigcup_j J_j \Big) = \sum_{j=1}^{\infty} \alpha(J_j) = \sum_{k=1}^{\infty} \Big(\sum_{J_i \subset I \cap I_k} \alpha(J_i) \Big) \le \sum_{k=1}^{\infty} \alpha(I_k \cap I) \le \sum_{k=1}^{\infty} \alpha(I_k).$$

Let now us discuss the properties of μ^*.

(i) $\mu^*(I) \le \alpha(I)$ for all $I \in \mathcal{I}$ by definition of μ^*; the σ-subadditivity of α yields $\alpha(I) \le \sum_{k=1}^{\infty} \alpha(I_k)$ whenever $\{I_k\} \subset \mathcal{I}$ and $I \subset \cup_k I_k$. Hence $\alpha(I) \le \mu^*(I)$ and $\alpha(I) = \mu^*(I)$ for $I \in \mathcal{I}$.

(ii) We need to prove that E is measurable if (5.3) holds. Let $A \subset X$ with $\mu^*(A) < \infty$ and, for $\epsilon > 0$ let $\{I_k\}$ be a sequence in $\mathcal{I}$ with $\cup_k I_k \supset A$ and $\sum_{k=1}^{\infty} \alpha(I_k) \le \mu^*(A) + \epsilon$. We compute, because of the hypotheses and (i)

$$\begin{aligned} \mu^*(A \cap E) + \mu^*(A \cap E^c) &\le \mu^*\Big(\bigcup_k I_k \cap E \Big) + \mu^*\Big(\bigcup_k I_k \cap E^c \Big) \\ &\le \sum_{k=1}^{\infty} \Big(\mu^*(I_k \cap E) + \mu^*(I_k \cap E^c) \Big) \\ &\le \sum_{k=1}^{\infty} \mu^*(I_k) = \sum_{k=1}^{\infty} \alpha(I_k) \le \mu^*(A) + \epsilon. \end{aligned}$$

Hence E is μ^*-measurable.

(iii) Let $E \in \mathcal{I}$. For any $I \in \mathcal{I}$ the sets $I \cap E \in \mathcal{I}$ and $I \setminus E$ are unions of finitely many disjoint sets in $\mathcal{I}$, consequently (because of the additivity of α), $\alpha(I \cap E) + \alpha(I \cap E^c) = \alpha(I)$ and, because of (i),

$$\mu^*(I \cap E) + \mu^*(I \cap E^c) = \alpha(I \cap E) + \alpha(I \cap E^c) = \alpha(I) = \mu^*(I).$$

From (ii) we then conclude that E is μ^*-measurable.

(iv) Let $E \subset X$ with $\mu^*(E) < +\infty$. For $k = 1, 2, \ldots$, let $\{I_j^{(k)}\}$ be sequences of elements of $\mathcal{I}$ with $\cup_j I_j^{(k)} \supset E$ and $\sum_{j=1}^{\infty} \alpha(I_j^{(k)}) \leq \mu^*(E) + 2^{-k}$. For $F_k := \cup_j I_j^{(k)}$, taking into account the σ-subadditivity of μ^* and (i), we get

$$\mu^*(F_k) \leq \sum_{j=1}^{\infty} \mu^*(I_j^{(k)}) = \sum_{j=1}^{\infty} \alpha(I_j^{(k)}) \leq \mu^*(E) + 2^{-k}.$$

Therefore, $\mu^*(F_k) < \infty$, $\mu^*(\cap_k F_k) = \mu^*(E)$, while from (iii) we infer that F_k is μ^*-measurable. □

5.30 Corollary (Structure of measurable sets). *Let $\alpha : \mathcal{I} \to \overline{\mathbb{R}}_+$ be a σ-additive set function defined on a semiring $\mathcal{I}$, let μ^* be the outer measure associated to α via* (5.2) *and let $\mathcal{M}_{\mu^*}$ denote the family of μ^*-measurable sets. If $E \subset X$ has a finite μ^* measure, then E is μ^*-measurable if and only if*

$$E = \Big(\bigcap_k F_k\Big) \setminus N, \tag{5.4}$$

where $\mu^*(N) = 0$ and $\{F_k\}$ is a decreasing family of sets that are union of elements of $\mathcal{I}$, i.e., $F_{k+1} \subset F_k$ and $F_k = \cup_j I_j^{(k)}$, $I_j^{(k)} \in \mathcal{I}$.

Notice that the approximation property (5.4) holds for any set $E \subset X$ that is μ^*-measurable with $\mu^*(E) < +\infty$. Finally, notice that the last restriction is natural, see Exercise 5.31.

5.31 ¶. Let $\mathcal{I}$ be the family of finite subsets of $\mathbb{R}$ and let $\alpha : \mathcal{I} \to \overline{\mathbb{R}}_+$ be defined by $\alpha(A) := \#A$. Then $\mu^* : \mathcal{P}(\mathbb{R}) \to \overline{\mathbb{R}}_+$ is the *counting measure*, $\mu^*(A) = \#A$ and $\mathcal{M}_{\mu^*} = \mathcal{P}(\mathbb{R})$. Show that sets E for which (5.4) holds are at most denumerable.

In fact, the class of sets for which (5.4) holds is slightly larger.

5.32 Definition. *Let μ^* be an outer measure on X. We say that $E \subset X$ is μ^* σ-*finite *if E is the union of an increasing sequence $\{E_k\}$ of sets with finite measure. If X, and consequently each of its subsets, is μ^* σ-finite, we say that the outer measure μ^* is σ-finite.*

5.33 ¶. Under the hypotheses of Corollary 5.30, extend the characterization of measurability in terms of the approximation property (5.4) to any $E \subset X$ that is μ^* σ-finite.

Deduce that if μ^* is σ-finite, then $\mathcal{M}_{\mu^*}$ is the σ-algebra generated by $\mathcal{I}$ and the null sets of μ^*.

5.34 ¶. Let $(\mathcal{E}, \alpha)$ be a measure and let μ^* and $\mathcal{M}_{\mu^*}$ be the measure constructed starting from $(\mathcal{E}, \alpha)$ by Method I. Show that $\mu^*(N) = 0$ if and only if there exists $E \in \mathcal{E}$ such that $N \subset E$ and $\alpha(E) = 0$.

Later we shall apply Method I to construct interesting new measures. Here we notice that it yields an alternative approach to Lebesgue's measure with respect to the one in Section 5.1.2.

5.35 Example (Lebesgue's measure). Denote by $\mathcal{I}$ the class of left-closed intervals of $\mathbb{R}^n$ and by $|\ |:\mathcal{I}\to\overline{\mathbb{R}}_+$ the elementary measure of intervals. Then the definition of Lebesgue's outer measure in Definition 5.5 is exactly (5.2), and the following holds.

Proposition. *$\mathcal{I}$ is a semiring and $|\ |:\mathcal{I}\to\mathbb{R}_+$ is a σ-additive set function on $\mathcal{I}$.*

Proof. It is easily seen that $\mathcal{I}$ is a semiring and that the elementary measure $|\ |$ is finitely additive. Let us prove that it is σ-subadditive. For that, let I and I_k be intervals with $I=\cup_k I_k$ and, for $\epsilon>0$ and any k denote by J_k an interval centered as I_k that contains strictly I_k with $|J_k|\le|I_k|+\epsilon\,2^{-k}$. The family of open sets $\{\mathrm{int}(J_k)\}_k$ covers the compact set $\overline{I}$, hence we can select $k_1,k_2,\dots k_N$ such that $I\subset\cup_{i=1}^N\mathrm{int}(J_{k_i})$ concluding

$$|I|\le\sum_{i=1}^N|J_{k_i}|\le\sum_{k=1}^\infty|J_k|\le\sum_{k=1}^\infty(|I_k|+\epsilon 2^{-k})\le\sum_{k=1}^\infty|I_k|+2\epsilon,$$

i.e., that $|\ |$ is σ-subadditive on $\mathcal{I}$.

Suppose now that $I=\cup_k I_k$, where the $\{I_k\}$'s are pairwise disjoint. Of course, by the σ-subadditivity property of $|\ |$, $|I|\le\sum_{k=1}^\infty|I_k|$. On the other hand, $\cup_{k=0}^N I_k\subset I$ for any integer N. Finite additivity then yields

$$\sum_{k=0}^N|I_k|=\Big|\bigcup_{k=0}^N I_k\Big|\le|I|\qquad\forall N$$

and, as $N\to\infty$, also the opposite inequality $\sum_{k=0}^\infty|I_k|\le|I|$. □

Theorem 5.29 then tells us that Lebesgue's outer measure $\mathcal{L}^{n*}$ is an extension of the elementary volume measure and that intervals are measurable. As clearly $\mathbb{R}^n$ is $\mathcal{L}^{n*}$ σ-finite, (5.4) characterizes the measurable sets.

c. Construction of measures: Method II

Here we are interested in constructing measures on a metric space X for which open sets will be measurable. Denote by $\mathcal{B}(X)$ the smallest σ-algebra of subsets of X that contains all open sets. Its elements are called *Borel sets* of X. Of course, $\mathcal{B}(X)$ is also the smallest σ-algebra containing all closed sets of X.

An outer measure μ^* on a metric space X is said to be a *Borel measure* if Borel sets are μ^*-measurable, and we say that μ^* is *Borel-regular* if μ^* is a Borel measure and for any $A\subset X$ there is a Borel set B with $B\supset A$ and $\mu^*(B)=\mu^*(A)$. Notice that not necessarily $\mu^*(B\setminus A)=0$ except when A is measurable and $\mu^*(A)<+\infty$.

Borel measures will be discussed in Chapter 6; here we present a method for their construction. We begin with Carathéodory's characterization of Borel measures.

5.36 Theorem (Carathéodory). *Let X be a metric space with distance d. An outer measure μ^* in X is a Borel measure if and only if*

$$\mu^*(A)+\mu^*(B)=\mu^*(A\cup B)\qquad\forall A,B\subset X\ \textit{with}\ d(A,B)>0.$$

Proof. Suppose that μ^* is a Borel measure. If $A, B \subset X$ have positive distance, there is an open set Ω such that $A \subset \Omega$, $B \subset \Omega^c$ and, since Ω is measurable,

$$\mu^*(A) + \mu^*(B) = \mu^*((A \cup B) \cap \Omega) + \mu^*((A \cup B) \cap \Omega^c) = \mu^*(A \cup B).$$

Proving the converse is slightly more complicated. Since $\mathcal{M}_{\mu^*}$ is a σ-algebra, it suffices to prove that closed sets are μ^* measurable, i.e.,

$$\mu^*(A \cap C) + \mu^*(A \cap C^c) \leq \mu^*(A)$$

for every closed set C and every set A with $\mu^*(A) < +\infty$. Let $C_k := \{x \in X \,|\, d(x, C) \leq 1/k\}$. Since $d(A \cap C_k^c, A \cap C) > 0$, we know that

$$\mu^*(A \cap C) + \mu^*(A \cap C_k^c) = \mu^*(A \cap (C \cup C_k^c)) \leq \mu^*(A).$$

Therefore the conclusion follows if $\mu^*(A \cap C_k^c) \to \mu^*(A \cap C^c)$. In order to prove this we notice that, since C is closed,

$$A \cap C^c = A \cap C_k^c \cup \bigcup_{j=k}^{\infty} (A \cap R_j)$$

where $R_j := \{x \,|\, 1/(j+1) < d(x, C) \leq 1/j\}$. The subadditivity of μ^* then yields

$$\mu^*(A \cap C_k^c) \leq \mu^*(A \cap C^c) \leq \mu^*(A \cap C_k^c) + \sum_{j=k}^{\infty} \mu^*(A \cap R_j)$$

and the proof is completed if $\sum_{j=1}^{\infty} \mu^*(A \cap R_j) < +\infty$. To prove the last inequality, we notice that $d(R_i, R_j) > 0$ $\forall i, j$, $i \geq j+2$, hence, applying inductively the assumption, we compute

$$\sum_{j=1}^{N} \mu^*(A \cap R_{2j}) = \mu^*(A \cap \cup_{j=1}^{N} R_{2j}) \leq \mu^*(A) < +\infty,$$

$$\sum_{j=1}^{N} \mu^*(A \cap R_{2j+1}) = \mu^*(A \cap \cup_{j=1}^{N} R_{2j+1}) \leq \mu^*(A) < +\infty.$$

□

Let X be a metric space and $(\mathcal{I}, \alpha)$ a set function. For $\delta > 0$ denote by μ_δ^* the outer measure defined for all $E \subset X$ by

$$\mu_\delta^*(E) := \inf\Big\{ \sum_{k=1}^{\infty} \alpha(I_k) \;\Big|\; \cup_k I_k \supset E,\ I_k \in \mathcal{I},\ \operatorname{diam}(I_k) < \delta \Big\}.$$

Since $\delta \to \mu_\delta^*(E)$ is nondecreasing, we may and do define a new outer measure $\mu^* : \mathcal{P}(X) \to \overline{\mathbb{R}}$ given by

$$\mu^*(E) := \lim_{\delta \to 0^+} \mu_\delta^*(E).$$

We say that μ^* *is the outer measure constructed from* α *by Method II.*

5.37 Proposition. *Let X be a metric space and $(\mathcal{I}, \alpha)$ a set function in X. Let $\mu^* : \mathcal{P}(X) \to \overline{\mathbb{R}}$ be the outer measure constructed from $(\mathcal{I}, \alpha)$ by Method II. Then μ^* is a Borel measure. Moreover, if $\mathcal{I} \subset \mathcal{B}(X)$, then μ^* is Borel-regular.*

Proof. To prove that μ^* is a Borel measure we use Caratéodory's test in Theorem 5.36. Let $A, B \subset X$ with $\text{dist}(A, B) > 0$ and $\mu^*(A \cup B) < \infty$. Suppose that $\text{dist}(A, B) > 2\delta > 0$. Since $\mu^*_\sigma(A \cup B) < +\infty \ \forall \sigma$, for any $\epsilon > 0$ there is a covering $\{I_k\}$ of $A \cup B$ with sets in $\mathcal{I}$ of diameter at most σ. $\{I_k\}$ then splits into a covering $\{I'_k\}$ of A and a covering $\{I''_k\}$ of B and we may compute

$$\mu^*_\sigma(A \cup B) \geq \sum_{k=1}^{\infty} \alpha(I_k) - \epsilon \geq \sum_{k=1}^{\infty} \alpha(I'_k) + \sum_{k=1}^{\infty} \alpha(I''_k) - \epsilon \geq \mu^*_\sigma(A) + \mu^*_\sigma(B) - \epsilon.$$

Letting first $\epsilon \to 0$ and then $\sigma \to 0$, we conclude $\mu^*(A \cup B) \geq \mu^*(A) + \mu^*(B)$.

Assume now $\mathcal{I} \subset \mathcal{B}(X)$ and let $A \subset X$. Without restrictions we may assume that $\mu^*(A) < +\infty$. Then for all $\delta > 0$ $\mu^*_\delta(A) < +\infty$ and by the definition of μ^*_δ, there is a Borel set $B_\delta \supset A$ with $\mu^*_\delta(A) = \mu^*_\delta(B_\delta)$. We may also arrange things in such a way that $B_\delta \subset B_\sigma$ if $\sigma \leq \delta$. By choosing a decreasing sequence $\delta_n \to 0$, we find an increasing sequence of Borel sets $B_n := B_{\delta_n}$ such that $\mu^*_{\delta_n}(A) = \mu^*_{\delta_n}(B_n)$. Passing to the limit in n, we then get $\mu^*(A) = \mu^*(B)$ with $B = \cup_n B_n$ that is a Borel set. □

5.2 Measurable Functions and the Integral

Given a measure $(\mathcal{E}, \mu)$ on a set X, for instance, the Lebesgue measure $(\mathcal{L}^n, \mathcal{M})$ in $\mathbb{R}^n$, we may construct a corresponding integral with respect to that measure. To do this we first introduce the notion of $\mathcal{E}$-measurable functions.

5.2.1 Measurable functions

Let X be a set and $\mathcal{E}$ a σ-algebra of subsets of X.

5.38 Definition. *We say that* $f : X \to \overline{\mathbb{R}}$ *is* $\mathcal{E}$-measurable *if for all* $t \in \mathbb{R}$ *the set*

$$E_{f,t} := f^{-1}(]t, +\infty[) = \left\{ x \in X \,\middle|\, f(x) > t \right\}$$

belongs to $\mathcal{E}$*. The class of* $\mathcal{E}$*-measurable functions is denoted by* $\mathcal{M}_\mathcal{E}$ *or by* $\mathcal{M}$ *when no confusion may arise.*

If $\mathcal{E}$ is the σ-algebra $\mathcal{M}_{\mu^*}$ of the measurable sets of an outer measure μ^* and μ denotes the restriction of μ^* to $\mathcal{M}_{\mu^*}$ we say that f is *μ-measurable* (or simply *measurable* if μ is clear from the context) instead of f is $\mathcal{M}_{\mu^*}$-measurable. If $\mathcal{E} = \mathcal{B}(X)$, the σ-algebra of Borel sets in a metric space X, then $\mathcal{E}$-measurable functions are called *Borel functions*. In particular, $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *Lebesgue measurable* or *$\mathcal{L}^n$-measurable* (respectively Borel measurable) if for all $t \in \mathbb{R}$ the set

$$\Big\{x \in \mathbb{R}^n \,\Big|\, f(x) > t\Big\}$$

is $\mathcal{L}^n$-measurable (respectively is a Borel set).

Let $E \in \mathcal{E}$. A function $f : E \subset \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *$\mathcal{E}$-measurable* in E if $E \in \mathcal{E}$ and $\{x \in E \,|\, f(x) > t\} \in \mathcal{E}$ for all $t \in \mathbb{R}$. Notice that f is $\mathcal{E}$-measurable in E if and only if f is $\mathcal{F}$-measurable in the set space E with respect to the σ-algebra

$$\mathcal{F} := \Big\{A \,|\, A \in \mathcal{E}, A \subset E\Big\}.$$

5.39 ¶. Show that f is $\mathcal{E}$-measurable in $E \in \mathcal{E}$ if and only if its extension to $-\infty$ or to a constant in all of X is $\mathcal{E}$-measurable.

The next proposition collects several equivalent ways of saying that a function $f : X \to \mathbb{R}$ is $\mathcal{E}$-measurable.

5.40 Proposition. *Let $\mathcal{E}$ be a σ-algebra of subsets of a set X and $f : X \to \overline{\mathbb{R}}$ a function. The following claims are equivalent:*

(i) *f is $\mathcal{E}$-measurable, i.e., $E_{f,t} \in \mathcal{E}$ for all t.*
(ii) *$\{x \in X \,|\, f(x) \geq t\} \in \mathcal{E}$ for all t.*
(iii) *$\{x \in X \,|\, f(x) \leq t\} \in \mathcal{E}$ for all t.*
(iv) *$\{x \in X \,|\, f(x) < t\} \in \mathcal{E}$ for all t.*
(v) *For every open set $A \subset \mathbb{R}$ we have $f^{-1}(A) \in \mathcal{E}$.*
(vi) *For every closed set $F \subset \mathbb{R}$ we have $f^{-1}(F) \in \mathcal{E}$.*
(vii) *For every Borel set $B \subset \mathbb{R}$ we have $f^{-1}(B) \in \mathcal{E}$.*

Moreover, in (i), (ii), (iii) and (iv) we can replace "for all t" with "for all t in a dense subset of $\mathbb{R}$".

Proof. (i) $\Rightarrow$ (ii). Notice that

$$\Big\{x \in X \,\Big|\, f(x) \geq t\Big\} = \bigcap_{n=1}^{\infty} \Big\{x \in X \,\Big|\, f(x) > t - \frac{1}{n}\Big\};$$

thus, $\{x \in \mathbb{R}^n \,\Big|\, f(x) \geq t\} \in \mathcal{E}$ as intersection of the $\mathcal{E}$-measurable sets $\{x \in X \,|\, f(x) > t - 1/n\} \in \mathcal{E}$.

The other implications are proved similarly. We leave them to the reader as exercises. We only prove that

(i) $\Rightarrow$ (vii). For that, first notice that

$$f^{-1}(]a, b]) = \Big\{x \,\Big|\, a < f(x) \leq b\Big\} = \Big\{f(x) > a\Big\} \setminus \Big\{f(x) > b\Big\},$$

hence $f^{-1}(I) \in \mathcal{E}$ for all intervals I. Since the Borel sets form a σ-algebra and for any family of sets

$$f^{-1}(\cup_i A_i) = \bigcup_i f^{-1}(A_i), \qquad f^{-1}(\cap_i A_i) = \bigcap_i f^{-1}(A_i),$$
$$f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B),$$

we infer that

$$\mathcal{F} = \left\{ f^{-1}(B) \,\middle|\, B \subset \mathbb{R} \text{ is a Borel set} \right\}$$

is a σ-algebra contained in $\mathcal{E}$.

Finally, if $Z \subset \mathbb{R}$ is dense in $\mathbb{R}$ and $E_{f,t} \in \mathcal{E}$ $\forall t \in Z$, then for all $t \in \mathbb{R}$ we choose $\{t_n\} \subset Z$ such that $t_n \downarrow t$. Since

$$E_{f,t} = \bigcup_n E_{f,t_n},$$

we conclude that $E_{f,t}$ is in $\mathcal{E}$, too. □

5.41 ¶. Show that there are Lebesgue measurable functions $f : \mathbb{R} \to \mathbb{R}$ and Lebesgue measurable sets $E \subset \mathbb{R}$ such that $f^{-1}(E)$ is not Lebesgue measurable.
[*Hint.* Consider the inverse of $g(x) := x + f(x) : [0,1] \to [0,2]$, f being the Cantor–Vitali function. Notice that g is invertible with continuous inverse and maps the Cantor set, which is a zero set, onto a set of positive measure. Recall also that sets of positive measure always contain nonmeasurable sets.]

a. Families of measurable functions

5.42 Lemma. *Let $\mathcal{E}$ be a σ-algebra of subsets of a set X and $f, g : X \to \overline{\mathbb{R}}$ be two $\mathcal{E}$-measurable functions. Then $\{x \in X \,|\, f(x) > g(x)\} \in \mathcal{E}$.*

Proof. For every rational $r \in \mathbb{Q}$, the set $A_r := \{x \in E \,|\, f(x) > r,\ g(x) < r\} \in \mathcal{E}$ is the intersection of two sets in $\mathcal{E}$. On the other hand,

$$\left\{ x \in E \,\middle|\, f(x) > g(x) \right\} = \bigcup_{r \in \mathbb{Q}} A_r,$$

hence $\{x \in E \,|\, f(x) > g(x)\} \in \mathcal{E}$ is a denumerable union of sets in $\mathcal{E}$. □

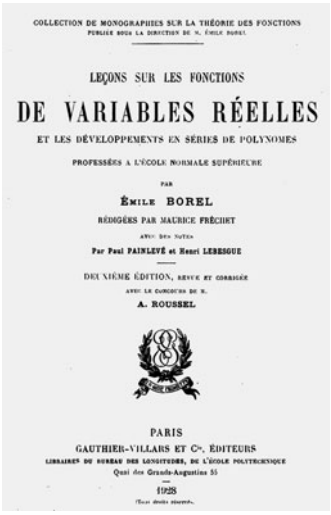
5.43 Proposition. *Let X be a set, $\mathcal{E}$ a σ-algebra of subsets of X and $\mathcal{M}$ the class of all $\mathcal{E}$-measurable functions. We have the following:*

(i) *Constant functions are $\mathcal{E}$-measurable; moreover, $E \in \mathcal{E}$ if and only if $\chi_E \in \mathcal{M}$.*
(ii) *Let $f, g \in \mathcal{M}$. Then $\max(f,g), \min(f,g) \in \mathcal{M}$; in particular, $f_+ = \max(f, 0)$, $f_- = \max(-f, 0)$, $|f| \in \mathcal{M}$.*
(iii) *Let $f, g \in \mathcal{M}$ and $\alpha, \beta \in \mathbb{R}$. Then $\alpha f + \beta g \in \mathcal{M}$.*
(iv) *Let $f, g \in \mathcal{M}$. Then $1/f \in \mathcal{M}$, $f^2 \in \mathcal{M}$ and $fg \in \mathcal{M}$; in particular, if f is $\mathcal{E}$-measurable in E, then f is $\mathcal{E}$-measurable in any $F \in \mathcal{E}$, $F \subset E$.*
(v) *If $\{f_k\} \subset \mathcal{M}$, then $\inf_k f_k(x)$ and $\sup_k f_k(x) \in \mathcal{M}$.*
(vi) *If $\{f_k\} \subset \mathcal{M}$ and $f_k(x) \to f(x)$ pointwise, then $f \in \mathcal{M}$.*
(vii) *If $\{f_k\} \subset \mathcal{M}$, then $\liminf_{k\to\infty} f_k(x)$ and $\limsup_{k\to\infty} f_k(x) \in \mathcal{M}$.*
(viii) *If $f \in \mathcal{M}$ and $\phi : \mathbb{R} \to \mathbb{R}$ is a continuous function, then $\phi \circ f \in \mathcal{M}$; in particular, $|f|$ and $|f|^p$, $\forall p \in \mathbb{R}$, and e^f belong to $\mathcal{M}$.*

Proof. (i) is trivial.

(ii) For all $t \in \mathbb{R}$ we have

$$E_{\max(f,g),t} = E_{f,t} \bigcup E_{g,t}, \qquad E_{\min(f,g),t} = E_{f,t} \bigcap E_{g,t}.$$


COLLECTION DE MONOGRAPHIES SUR LA THÉORIE DES FONCTIONS
PUBLIÉE SOUS LA DIRECTION DE M. ÉMILE BOREL

LEÇONS SUR LES FONCTIONS
DE VARIABLES RÉELLES
ET LES DÉVELOPPEMENTS EN SÉRIES DE POLYNOMES
PROFESSÉES A L'ÉCOLE NORMALE SUPÉRIEURE
PAR
ÉMILE BOREL
RÉDIGÉES PAR MAURICE FRÉCHET
AVEC DES NOTES
Par Paul PAINLEVÉ et Henri LEBESGUE
DEUXIÈME ÉDITION, REVUE ET CORRIGÉE
AVEC LE CONCOURS DE M.
A. ROUSSEL
PARIS
GAUTHIER-VILLARS ET Cie, ÉDITEURS
LIBRAIRES DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE
Quai des Grands-Augustins 55
1928
(Tous droits réservés.)

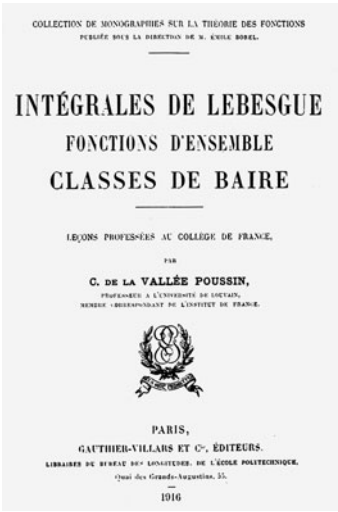

COLLECTION DE MONOGRAPHIES SUR LA THÉORIE DES FONCTIONS
PUBLIÉE SOUS LA DIRECTION DE M. ÉMILE BOREL.
INTÉGRALES DE LEBESGUE
FONCTIONS D'ENSEMBLE
CLASSES DE BAIRE
LEÇONS PROFESSÉES AU COLLÈGE DE FRANCE,
PAR
C. DE LA VALLÉE POUSSIN,
PROFESSEUR A L'UNIVERSITÉ DE LOUVAIN,
MEMBRE CORRESPONDANT DE L'INSTITUT DE FRANCE.
PARIS,
GAUTHIER-VILLARS ET Cie, ÉDITEURS.
LIBRAIRES DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE,
Quai des Grands-Augustins, 55.
1916

Figure 5.3. Frontispieces of two works by Emile Borel (1871–1956) and Charles de la Vallée–Poussin (1866–1962), respectively.

(iii) Since for $h : X \to \overline{\mathbb{R}}$ we have $E_{-h,t} = \{x \,|\, h(x) < -t\}$, Proposition 5.40 yields that $-h \in \mathcal{M}$ if and only if $h \in \mathcal{M}$. Similarly, if $\beta \in \mathbb{R}$ and $h \in \mathcal{M}$, then $\beta h \in \mathcal{M}$. In fact,

$$E_{\beta h,t} = \begin{cases} E_{h,t/\beta} & \text{if } \beta > 0, \\ X & \text{if } \beta = 0 \text{ and } t < 0, \\ \emptyset & \text{if } \beta = 0 \text{ and } t \geq 0, \\ \{x \,|\, h(x) < t/\beta\} & \text{if } \beta < 0. \end{cases}$$

Additionally, notice that for $h \in \mathcal{M}$ we have $h(x) + c \in \mathcal{M}$ since $E_{g+c,t} = E_{g,t-c}\ \forall t$. Hence $\alpha f,\ t - \beta g \in \mathcal{M}$ and

$$E_{\alpha f+\beta g,t} = \{x \,|\, \alpha f(x) > t - \beta g(x)\}$$

for all $t \in \mathbb{R}$. The thesis follows from Lemma 5.42.

(iv) We have

$$E_{1/f,t} = \begin{cases} \{x \,|\, 0 < f(x) < 1/t\} & \text{if } t > 0, \\ \{x \,|\, f(x) > 0\} & \text{if } t = 0, \\ \{x \,|\, f(x) > 0\} \cup \{f(x) < 1/t\} & \text{if } t < 0. \end{cases}$$

In all cases $E_{1/f,t} \in \mathcal{E}$, hence $1/f \in \mathcal{M}$. Moreover, $f^2 \in \mathcal{M}$ since

$$E_{f^2,t} = \begin{cases} X & \text{if } t \leq 0, \\ \{x \,|\, f(x) > \sqrt{t}\} \cup \{x \,|\, f(x) < -\sqrt{t}\} & \text{if } t > 0. \end{cases}$$

It then follows that $fg \in \mathcal{M}$ since $fg = \frac{1}{2}((f+g)^2 - f^2 - g^2)$.

(v) First we prove that

$$\Big\{x \,\Big|\, f(x) > t\Big\} = \bigcup_{n=1}^{\infty} \bigcup_{\overline{k}=1}^{\infty} \bigcap_{k=\overline{k}}^{\infty} \Big\{x \,\Big|\, f_k(x) > t + 1/n\Big\}. \tag{5.5}$$

If $f_k(x) \to f(x)$ and $f(x) > t$, then there exist $n \in \mathbb{N}$ and $\overline{k} \in \mathbb{N}$ such that for all $k \geq \overline{k}$ $f_k(x) > t + \frac{1}{n}$. In other words, there exist n and $\overline{k}$ such that

$$x \in \bigcap_{k=\overline{k}}^{\infty} \Big\{ x \,\Big|\, f_k(x) > t + \frac{1}{n} \Big\},$$

hence

$$\Big\{ x \,\Big|\, f(x) > t \Big\} \subset \bigcup_{n=1}^{\infty} \bigcup_{\overline{k}=1}^{\infty} \bigcap_{k=\overline{k}}^{\infty} \Big\{ x \,\Big|\, f_k(x) > t + 1/n \Big\}.$$

Conversely, if x belongs to the set on the right of (5.5), then there exist n and $\overline{k}$ such that for $k > \overline{k}$ we have $f_k(x) > t + \frac{1}{n}$. Letting $k \to \infty$ we find $f(x) \geq t + \frac{1}{n} > t$, hence $x \in \{f(x) > t\}$.

In words, (5.5) says that for all t's the set $\{f(x) > t\}$ is a denumerable union of intersections of sets in $\mathcal{E}$; therefore, $\{f(x) > t\} \in \mathcal{E}$ i.e., f is $\mathcal{E}$-measurable.

(vi) In fact, we have

$$E_{\sup_k f_k, t} = \bigcup_k E_{f_k, t}, \qquad E_{\inf_k f_k, t} = \bigcap_k E_{f_k, t}.$$

(vii) By definition

$$\liminf_{k\to\infty} f_k(x) = \lim_{j\to\infty} \inf_{k\geq j} f_k(x), \qquad \limsup_{k\to\infty} f_k(x) = \lim_{j\to\infty} \sup_{k\geq j} f_k(x);$$

the claim then follows from (v) and (vi).

(viii) If $A \subset \mathbb{R}$ is open, then $\phi^{-1}(A)$ is open; Proposition 5.40 then yields $(\phi \circ f)^{-1}(A) = f^{-1}(\phi^{-1}(A)) \in \mathcal{E}$. □

5.44 ¶. Let $\mathcal{E}$ be a σ-algebra of subsets of a set X. If $f, g : X \to \overline{\mathbb{R}}$ are $\mathcal{E}$-measurable, then

$$h(x) := \begin{cases} f(x) & \text{if } x \in E, \\ g(x) & \text{if } x \in E^c \end{cases}$$

is $\mathcal{E}$-measurable.

b. Approximation by simple functions

Let X be a set and $\mathcal{E}$ a σ-algebra of subsets of X. We recall that the characteristic function of $A \subset X$ is

$$\chi_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

We say that $\varphi : X \to \overline{\mathbb{R}}$ is a *simple function* if it takes finitely many finite distinct values $a_1, a_2, \ldots, a_N$. The class of all $\mathcal{E}$-measurable simple functions is denoted by $\mathcal{S}$. Clearly, $\varphi \in \mathcal{S}$ takes distinct values $a_1, a_2, \ldots, a_N$ if and only if

$$\varphi(x) = \sum_{k=1}^{N} a_k \chi_{E_k},$$

where $E_k := \{x \in X \,|\, \varphi(x) = a_k\}$. Moreover, φ is $\mathcal{E}$-measurable if and only if $E_k \in \mathcal{E}$ for all k.

5.45 Lemma. *Let X be a set and $\mathcal{E}$ a σ-algebra of subsets of X. A nonnegative function $f : X \to \overline{\mathbb{R}}_+$ is $\mathcal{E}$-measurable if and only if there exists a nondecreasing sequence of $\mathcal{E}$-measurable simple functions $\{\varphi_k\}$ such that $\varphi_k(x) \to f(x)$ pointwise.*

Proof. Let us construct the sequence $\{\varphi_k\}$.

Let $f : X \to \mathbb{R}$ be a nonnegative function. For $k = 1, 2, 3, \dots$ set $E_k := \{x \,|\, f(x) > 2^k\}$ and for $h = 0, 1, , \dots, 4^k - 1$ set $E_{k,h} := \{x \,|\, h/2^k < f(x) \le (h+1)/2^k\}$, and, finally, define $\varphi_k : \mathbb{R}^n \to \mathbb{R}$ by

$$\varphi_k(x) = \begin{cases} 2^k & \text{if } x \in E_k, \\ \frac{h}{2^k} & \text{if } x \in E_{k,h}. \end{cases} \tag{5.6}$$

Trivially $\varphi_k(x)$ takes finitely many values. It is not difficult to show that $\varphi_{k+1}(x) \ge \varphi_k(x)$ and that $\varphi_k(x) \le f(x)$ for all x and

$$\varphi_k(x) = \sum_{h=0}^{4^k-1} \frac{h}{2^k} \chi_{E_{k,h}}(x) + 2^k \chi_{E_k}(x) \qquad \forall x \in X.$$

Moreover, $\varphi_k(x) \to f(x)$ pointwise. In fact, if $f(x) = +\infty$, then $\varphi_k(x) = 2^k \ \forall k$, hence $\varphi_k(x) \to +\infty = f(x)$. On the other hand, if $f(x) \in \mathbb{R}$ for sufficiently large k, we have $f(x) < 2^k$, hence $x \in E_{k,h}$ for some $h = 0, 1, \dots, 4^k - 1$. Therefore,

$$f(x) - \varphi_k(x) \le \frac{h+1}{2^k} - \frac{h}{2^k} = \frac{1}{2^k},$$

and again $\varphi_k(x) \to f(x)$ as $k \to \infty$.

To conclude, it remains to show that φ_k is $\mathcal{E}$-measurable. If f is $\mathcal{E}$-measurable, then $E_k \in \mathcal{E}$ and $E_{k,h} \in \mathcal{E}$ thus, $\{\varphi_k\}$ is $\mathcal{E}$-measurable. □

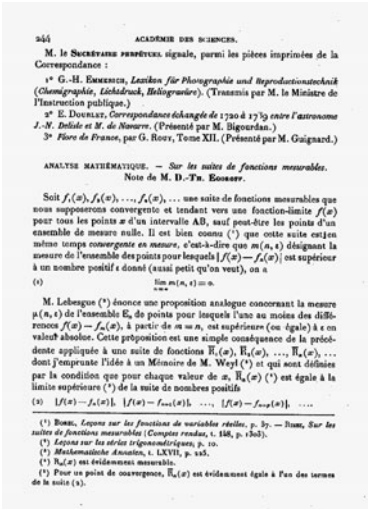
c. Null sets

Let $(\mathcal{E}, \mu)$ be a measure on a set X and $E \in \mathcal{E}$. We say that a function $f : E \to \mathbb{R}$ vanishes *μ-almost everywhere* if for the set $N = \{x \in E \,|\, f(x) \ne 0\} \in \mathcal{E}$ we have $\mu(N) = 0$.

5.46 Definition. *Let $(\mathcal{E}, \mu)$ be a measure on a set X and $E \in \mathcal{E}$. A predicate $p(x)$ depending on $x \in E$ is said to hold μ-almost everywhere in E or for μ-almost every $x \in E$ and we write $p(x)$ is true for μ-a.e. $x \in E$, if $\{x \in E \,|\, p(x)$ is false$\} \subset N \in \mathcal{E}$ with $\mu(N) = 0$.*

5.47 ¶. Let $(\mathcal{E}, \mu)$ be a measure on a set X, let $E \in \mathcal{E}$ and let $f, g : E \to \overline{\mathbb{R}}$ be two functions such that $f(x) = g(x)$ for μ-a.e. x. Show that f is $\mathcal{E}$-measurable if and only if g is $\mathcal{E}$-measurable. [*Hint.* $E_{g,t} \Delta E_{f,t} \subset N$. Therefore $E_{g,t} = (E_{f,t} \setminus N_1) \cup N_2$ with $\mu(N_1) = \mu(N_2) = 0$.]

5.48 ¶. Let $(\mathcal{E}, \mu)$ be a measure on a set X, let $E \in \mathcal{E}$ and $f : X \to \overline{\mathbb{R}}$ an $\mathcal{E}$-measurable function. Show that $\mu(\{x \in E \,|\, f(x) = t\}) = 0$ except for at most a dense denumerable set of t's.



244 ACADÉMIE DES SCIENCES.

M. le SECRÉTAIRE PERPÉTUEL signale, parmi les pièces imprimées de la Correspondance :

1° G.-H. EMMERICH, *Lexikon für Photographie und Reproductionstechnik (Chemigraphie, Lichtdruck, Heliogravüre)*. (Transmis par M. le Ministre de l'Instruction publique.)

2° E. DOUBLET, *Correspondance échangée de 1730 à 1759 entre l'astronome J.-N. Delisle et M. de Maraldi*. (Présenté par M. Bigourdan.)

3° *Flore de France*, par G. ROUY, Tome XII. (Présenté par M. Guignard.)

ANALYSE MATHÉMATIQUE. — *Sur les suites de fonctions mesurables.* Note de M. D.-TH. EGOROFF.

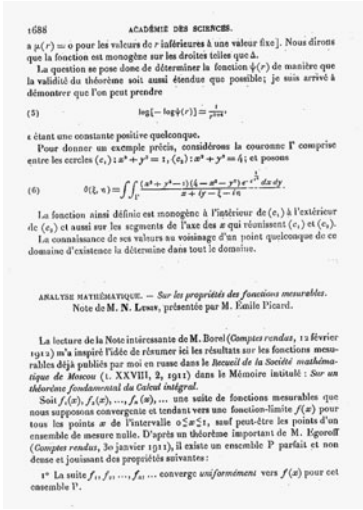
Soit $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ une suite de fonctions mesurables que nous supposerons convergente et tendant vers une fonction-limite $f(x)$ pour tous les points x d'un intervalle AB, sauf peut-être les points d'un ensemble de mesure nulle. Il est bien connu (1) que cette suite est en même temps *convergente en mesure*, c'est-à-dire que $m(n, \varepsilon)$ désignant la mesure de l'ensemble des points pour lesquels $|f(x) - f_n(x)|$ est supérieur à un nombre positif ε donné (aussi petit qu'on veut), on a

$$\lim_{n=\infty} m(n, \varepsilon) = 0. \qquad (1)$$

M. Lebesgue (2) énonce une proposition analogue concernant la mesure $\mu(n, \varepsilon)$ de l'ensemble E_n de points pour lesquels l'une au moins des différences $f(x) - f_m(x)$, à partir de $m = n$, est supérieure (ou égale) à ε en valeur absolue. Cette proposition est une simple conséquence de la précédente appliquée à une suite de fonctions $\overline{\Pi}_1(x), \overline{\Pi}_2(x), \ldots, \overline{\Pi}_n(x), \ldots$ dont j'emprunte l'idée à un Mémoire de M. Weyl (3) et qui sont définies par la condition que pour chaque valeur de x, $\overline{\Pi}_n(x)$ (4) est égale à la limite supérieure (5) de la suite de nombres positifs

$$(2) \quad |f(x) - f_n(x)|, \quad |f(x) - f_{n+1}(x)|, \quad \ldots, \quad |f(x) - f_{n+p}(x)|, \quad \ldots$$

(1) BOREL, *Leçons sur les fonctions de variables réelles*, p. 37. — RIESZ, *Sur les suites de fonctions mesurables* (*Comptes rendus*, t. 148, p. 1303).
(2) *Leçons sur les séries trigonométriques*, p. 10.
(3) *Mathematische Annalen*, t. LXVII, p. 225.
(4) $\overline{\Pi}_n(x)$ est évidemment mesurable.
(5) Pour un point de convergence, $\overline{\Pi}_n(x)$ est évidemment égale à l'un des termes de la suite (2).



1688 ACADÉMIE DES SCIENCES.

a $\mu(r) = 0$ pour les valeurs de r inférieures à une valeur fixe]. Nous dirons que la fonction est monogène sur les droites telles que Δ.

La question se pose donc de déterminer la fonction $\psi(r)$ de manière que la validité du théorème soit aussi étendue que possible; je suis arrivé à démontrer que l'on peut prendre

$$(5) \qquad \log[-\log\psi(r)] = \frac{1}{r^{2+\varepsilon}},$$

ε étant une constante positive quelconque.

Pour donner un exemple précis, considérons la couronne Γ comprise entre les cercles $(c_1) : x^2 + y^2 = 1$, $(c_2) : x^2 + y^2 = 4$; et posons

$$(6) \qquad \theta(\xi, \eta) = \iint_{\Gamma} \frac{(x^2 + y^2 - 1)(4 - x^2 - y^2)\, e^{-\frac{\lambda}{x+iy-\xi-i\eta}}\, dx\, dy}{x + iy - \xi - i\eta}.$$

La fonction ainsi définie est monogène à l'intérieur de (c_1), à l'extérieur de (c_2) et aussi sur les segments de l'axe des x qui réunissent (c_1) et (c_2). La connaissance de ses valeurs au voisinage d'un point quelconque de ce domaine d'existence la détermine dans tout le domaine.

ANALYSE MATHÉMATIQUE. — *Sur les propriétés des fonctions mesurables.* Note de M. N. LUSIN, présentée par M. Émile Picard.

La lecture de la Note intéressante de M. Borel (*Comptes rendus*, 12 février 1912) m'a inspiré l'idée de résumer ici les résultats sur les fonctions mesurables déjà publiés par moi en russe dans le *Recueil de la Société mathématique de Moscou* (t. XXVIII, 2, 1911) dans le Mémoire intitulé : *Sur un théorème fondamental du Calcul intégral*.

Soit $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ une suite de fonctions mesurables que nous supposons convergente et tendant vers une fonction-limite $f(x)$ pour tous les points x de l'intervalle $0 \leq x \leq 1$, sauf peut-être les points d'un ensemble de mesure nulle. D'après un théorème important de M. Egoroff (*Comptes rendus*, 30 janvier 1911), il existe un ensemble P parfait et non dense et jouissant des propriétés suivantes :

1° La suite $f_1, f_2, \ldots, f_n, \ldots$ converge *uniformément* vers $f(x)$ pour cet ensemble P.

Figure 5.4. The beginning of two papers by Dimitri Egorov (1869–1931) and Nikolai Lusin (1883–1950) in which Egorov and Lusin theorems are proved.

d. Lebesgue measurable functions

We now restrict ourselves to $\mathcal{L}^n$-measurable functions in $\mathbb{R}^n$.

We know that open and closed sets in $\mathbb{R}^n$ are $\mathcal{L}^n$-measurable. Therefore, Borel sets and Borel functions are $\mathcal{L}^n$-measurable. One then easily infers that all elementary functions are measurable.

(i) *Continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ are Borel functions.* In fact, for any t the set $E_{f,t}$ is open, hence Borel. We have more.

(ii) *If $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is continuous in A and A is measurable, then f is measurable in A.* In fact, for any t the set $E_{f,t}$ is open in A, i.e., $E_{f,t} = \Omega_t \cap A$, Ω_t being an open set of $\mathbb{R}^n$. As intersection of measurable sets, $E_{f,t}$ is measurable.

(iii) *Lower or upper semicontinuous functions are Borel functions*, since $E_{f,t}$ or, respectively, $\{f(x) < t\}$ are open sets.

We conclude this section by characterizing Lebesgue measurable functions as continuous functions except on small sets.

5.49 Theorem (Lusin). *Let $f : E \to \mathbb{R}$ be a function defined on a measurable set $E \subset \mathbb{R}^n$. f is $\mathcal{L}^n$-measurable in E if and only if for any $\epsilon > 0$ there exists a closed set $F_\epsilon \subset E$ such that the restriction of f to F_ϵ is continuous and $|E \setminus F_\epsilon| < \epsilon$.*

Proof. Assume that f is $\mathcal{E}$-measurable in E. First, assume $|E| = \mathcal{L}^n(E) < +\infty$. For any integer k, we divide the real line into the denumerable union of intervals $I_{k,j}$ of length $1/k$. The inverse images of these intervals, $A_{k,j} := f^{-1}(I_{k,j}) \subset E$, are measurable and form a partition of $\mathcal{E}$. According to Corollary 5.15, given $\epsilon > 0$ for each j there is a compact set $E_{k,j}$ such that

$$E_{k,j} \subset A_{k,j}, \qquad |A_{k,j} \setminus E_{k,j}| < \epsilon\, 2^{-k}\, 2^{-j},$$

so that $|E \setminus \cup_{j=1}^{\infty} E_{k,j}| \le \epsilon\, 2^{-k}$. Consequently, we find an integer N such that the set $F_k := \cup_{i=1}^{N} E_{k,j}$ is closed and

$$|E \setminus F_k| < \epsilon\, 2^{-k}.$$

Next, we notice that for a given k the closed sets $E_{k,j}$, the union of which yields F_k, are pairwise disjoint. Therefore, if we choose a point $y_{k,j} \in I_{k,j}$, the function

$$g_k(x) := y_{k,j} \qquad \text{if } x \in E_{k,j},\ j = 1, \dots, N$$

defines a function $g_k : F_k \to \mathbb{R}$ that is piecewise constant, and more precisely constant in closed and disjoint sets, and, consequently, continuous. Moreover, $|g_k(x) - f(x)| \le 1/k$ on F_k. If we set

$$F := \bigcap_k F_k,$$

then F is closed, $|E \setminus F| \le \sum_{k=1}^{\infty} |E \setminus F_k| \le \epsilon \sum_{k=1}^{\infty} 2^{-k} = \epsilon$ and $|g_k(x) - f(x)| < 1/k$ in F. This allows us to conclude that $g_k \to f$ uniformly in F, hence $f_{|F}$ is continuous.

When $|E| = +\infty$, we square $\mathbb{R}^n$ with closed cubes $\{R_j\}$ with disjoint interiors and decompose E as $E = \cup_j E_j$ $E_j := E \cap R_j$. For each j we have $|E_j| < +\infty$ and, by the above, we find a closed set $F_j \subset E_j$ such that $f_{|F_j}$ is continuous and $|E_j \setminus F_j| \le \epsilon 2^{-j}$. Since the family $\{F_j\}$ is locally finite, $F := \cup_j F_j$ is closed, $f_{|F}$ is continuous and $|E \setminus F| \le \epsilon$. □

Taking into account Tietze's extension theorem, we then deduce the following.

5.50 Proposition. *Let $E \subset \mathbb{R}^n$ be measurable and $f : E \subset \mathbb{R}^n \to \mathbb{R}$ be measurable in E. For any $\epsilon > 0$ there exists a continuous function $g_\epsilon : E \to \mathbb{R}$ such that*

$$\mathcal{L}^n\Big(\Big\{x \in E \,\Big|\, f(x) \neq g_\epsilon(x)\Big\}\Big) < \epsilon.$$

5.51 ¶. Show that $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{L}^n$-measurable if and only if it agrees $\mathcal{L}^n$-a.e. with a Borel function.

5.2.2 Lebesgue integral

Given a measure $(\mathcal{E}, \mu)$ on X, there are several equivalent ways of defining the integral of an $\mathcal{E}$-measurable function. Here we choose a specific path starting from the integral of simple and $\mathcal{E}$-measurable functions.

If one starts with the Lebesgue measure $(\mathcal{M}, \mathcal{L}^n)$, the resulting integral is the *Lebesgue integral*, as already presented in [GM4]. In this case, the integral of a nonnegative measurable function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ will agree with the $\mathbb{R}^{n+1}$ Lebesgue measure of the subgraph of f:

$$\int f(x)\, dx = \mathcal{L}^{n+1}\Big(\{(x,t) \in \mathbb{R}^{n+1} \mid f(x) > t\}\Big).$$

a. Definition of Lebesgue integral

Let X be a set and $(\mathcal{E}, \mu)$ be a measure in X. Denote also by $\mathcal{M}$ the class of the $\mathcal{E}$-measurable functions $f : X \to \overline{\mathbb{R}}$ and by $\mathcal{S}$ the class of simple functions that are $\mathcal{E}$-measurable. Recall that a simple function writes as

$$\varphi(x) = \sum_{k=1}^{N} a_k \chi_{E_k},$$

where $a_i \neq a_j$ for $i \neq j$ and $E_k := \Big\{ x \,\Big|\, \varphi(x) = a_k \Big\}$. Thus, φ is $\mathcal{E}$-measurable if and only if $E_k \in \mathcal{E}$ for all k.

The *integral* of a nonnegative $\mathcal{E}$-measurable simple function $\varphi \in \mathcal{S}$ is then defined as

$$I_\mu(\varphi) := \sum_{k=1}^{N} a_k \mu(E_k),$$

where we agree that $a_k \mu(E_k) = 0$ if $a_k = 0$ and $\mu(E_k) = \infty$.

5.52 ¶. Show that the integral is a linear functional on $\mathcal{S}$,

$$I_\mu(\alpha\varphi + \beta\psi) = \alpha I_\mu(\varphi) + \beta I_\mu(\psi) \qquad \forall \alpha, \beta \in \mathbb{R},\ \forall \varphi, \psi \in \mathcal{S}.$$

Deduce that if $\varphi = \sum_{i=1}^{N} a_i \chi_{E_i}$, $E_i \in \mathcal{E}$, then $I(\varphi) = \sum_{i=1}^{N} a_i \mu(E_i)$ even if the E_i's are not pairwise disjoint.

We then define the integral of a measurable and nonnegative function $f : X \to \overline{\mathbb{R}}$. The idea is the following. As we have seen, f is the pointwise limit of an increasing sequence $\{\varphi_k\}$ of $\mathcal{E}$-measurable simple functions. Therefore, the sequence of the corresponding integrals $\{I(\varphi_k)\}$ is increasing, hence the limit $\lim_{k\to\infty} I_\mu(\varphi_k)$ exists. One can show that this limit does not depend on the particular sequence $\{\varphi_k\}$ that approximates f, see Beppo Levi's theorem below. Therefore, it is reasonable to define the integral of f as the above limit, $\lim_{k\to\infty} I_\mu(\varphi_k)$.

More formally, we proceed as follows. Let $f : X \to \overline{\mathbb{R}}_+$ be $\mathcal{E}$-measurable and nonnegative. We define the integral of f with respect to the measure $(\mathcal{E}, \mu)$ as

$$\int_X f(x)\, d\mu(x) := \sup\Big\{ I_\mu(\varphi) \,\Big|\, \varphi \in \mathcal{S}, \varphi(x) \leq f(x)\ \forall x \in X \Big\}.$$

For a generic $\mathcal{E}$-measurable function $f : X \to \overline{\mathbb{R}}$, its positive and negative parts, defined by

$$f_+(x) = \max(f(x), 0), \qquad f_-(x) = \max(-f(x), 0),$$

are $\mathcal{E}$-measurable and nonnegative and $f(x) = f_+(x) - f_-(x)\ \forall x$. We then set the following.

5.53 Definition. *Let $f : X \to \overline{\mathbb{R}}$ be an $\mathcal{E}$-measurable function. We say that f is* μ-integrable *if at least one of the integrals $\int f_+(x)\,d\mu(x)$ or $\int f_-(x)\,d\mu(x)$ is finite; in this case the* Lebesgue integral *of f with respect to the measure μ is*

$$\int_X f(x)\,d\mu(x) := \int_X f_+(x)\,d\mu(x) - \int_X f_-(x)\,d\mu(x).$$

We say that $f : X \to \overline{\mathbb{R}}$ is μ-summable *if both the integrals of f_+ and f_- are finite.*

We often write

$$\int_X f(x)\,d\mu(x) \qquad \text{as} \qquad \int_X f\,d\mu.$$

If $E \in \mathcal{E}$ and $f : E \to \mathbb{R}$ is $\mathcal{E}$-measurable in E, then the extension $\widetilde{f} : X \to \overline{\mathbb{R}}$ of f defined by setting $\widetilde{f}(x) = 0$ if $x \notin E$ is $\mathcal{E}$-measurable. We say that f is μ-integrable in E (respectively μ-summable in E) if $\widetilde{f}$ is μ-integrable (respectively μ-summable) and define the integral of f in E as

$$\int_E f(x)\,d\mu(x) := \int_X \widetilde{f}(x)\,d\mu(x).$$

The class of μ-summable function on X is denoted by $\mathcal{L}^1(X,\mu)$ or by $\mathcal{L}^1(X)$ or $\mathcal{L}^1$ when no confusion may arise.

The claims collected in the following proposition easily follow.

5.54 Proposition. *Let $(\mathcal{E},\mu)$ be a measure on X and $E \in \mathcal{E}$. We have the following:*

(i) *Lebesgue's integral extends the integral of $\mathcal{E}$-measurable simple functions, meaning $\int \varphi(x)\,d\mu(x) = I_\mu(\varphi)\ \forall \varphi \in \mathcal{S}$.*
(ii) *If f is integrable on E, then f is integrable on every measurable subset $F \subset E$ and $\int_F f(x)\,d\mu(x) = \int_E f(x)\chi_F(x)\,d\mu(x)$.*
(iii) *Let $E, F \in \mathcal{E}$ and $f : E \cup F \to \overline{\mathbb{R}}$ be integrable in $E \cup F$; then*

$$\int_F f\,d\mu + \int_G f\,d\mu = \int_{F\cup G} f\,d\mu + \int_{F\cap G} f\,d\mu.$$

(iv) *If $|E| = 0$, then every function is summable on E and $\int_E f(x)\,d\mu(x) = 0$.*
(v) *The class of summable functions in E, $\mathcal{L}^1(E,\mu)$, is a vector space.*
(vi) *f is μ-summable in E if and only if f is $\mathcal{E}-$measurable and*

$$\int |f(x)|\,d\mu(x) < +\infty.$$

(vii) *Lebesgue's integral is monotone, i.e., if f, g are μ-integrable in E and $f \le g$, then*

$$\int_E f(x)\,d\mu(x) \le \int_E g(x)\,d\mu(x),$$

(viii) *If f is μ-integrable in E, then*

$$\left| \int_E f(x)\, d\mu(x) \right| \le \int_E |f(x)|\, d\mu(x).$$

5.55 ¶. Let $(\mathcal{E}, \mu)$ be a measure on a set X and let $E \in \mathcal{E}$ and $f, g : E \to \overline{\mathbb{R}}$ be two functions that agree for μ-a.e. $x \in E$. Show that

(i) f is integrable if and only if g is integrable,
(ii) f is summable if and only if g is summable.

Moreover, $\int_E f\, d\mu = \int_E g\, d\mu$.

5.56 ¶. Let $(\mathcal{E}, \mu)$ be a measure on a set X and $E \in \mathcal{E}$. Show that if f is μ-summable in E and g is measurable with $|g(x)| \le M$ for μ-a.e. $x \in E$, then fg is μ-summable in E and $\int_E |f(x)g(x)|\, d\mu(x) \le M \int_E |f|\, d\mu$.

5.57 ¶. If $\int_{B \cap E} f(x)\, dx = 0$ for every ball B centered at points of E, then $f(x) = 0$ for all $x \in E$.

b. Beppo Levi's theorem

5.58 Theorem (Beppo Levi). *Let $(\mathcal{E}, \mu)$ be a measure on X and let $\{f_k\}$ be a nondecreasing sequence of nonnegative μ-measurable functions $f_k : E \subset X \to \overline{\mathbb{R}}$. Then $f(x) := \lim_{k\to+\infty} f_k(x)$ is μ-measurable in E and*

$$\int_E f(x)\, d\mu(x) = \lim_{k\to\infty} \int_E f_k(x)\, d\mu(x).$$

Proof. By extending f to the whole of X by setting $f(x) = 0$ if $x \notin E^c$, we reduce ourselves to proving the claim when $E = X$.

As pointwise limit of μ-measurable functions, f is μ-measurable and being $f_k(x) \le f(x)$ for every k and every $x \in X$,

$$\lim_{k\to\infty} \int f_k\, d\mu \le \int f\, d\mu.$$

The opposite inequality remains to be proved. Let

$$\alpha := \lim_{k\to\infty} \int_E f_k\, d\mu.$$

It is enough to assume $\alpha < +\infty$. Let ϕ be a simple function with $\phi \le f$ and let $0 < \beta < 1$. Consider for $k = 1, 2, \dots$ the sets

$$A_k := \Big\{ x \in X \,\Big|\, f_k(x) \ge \beta\phi(x) \Big\}.$$

Clearly, $\{A_k\}$ is a nondecreasing sequence of μ-measurable sets, and we have

$$\beta \int_{A_k} \phi\, d\mu = \int_{A_k} \beta\phi\, d\mu \le \int_{A_k} f_k\, d\mu \le \int f_k\, d\mu. \tag{5.7}$$

On the other hand, $\phi(x) := \sum_{i=1}^N a_i \chi_{B_i}(x)$ with $B_i := \{x \,|\, \phi(x) = a_i\} \in \mathcal{E}$, hence

$$\int_{A_k} \phi\, d\mu = \sum_{i=1}^N a_i \int_{B_i \cap A_k} d\mu \to \sum_{i=1}^N a_i \int_{B_i} d\mu = \int \phi\, d\mu$$

as $k \to \infty$, by the continuity property of measure μ. Therefore, passing to the limit in (5.7),

$$\beta \int \phi \, d\mu = \beta \lim_{k\to\infty} \int_{A_k} \phi \, d\mu \le \alpha.$$

When $\beta \to 1$, we then get $\int \phi \, d\mu \le \alpha$ for all ϕ that is simple and $\mathcal{E}$-measurable with $0 \le \phi \le f$. This shows that $\int f \, d\mu \le \alpha$. □

Fatou's lemma below is in fact another formulation of Beppo Levi's theorem. Lebesgue's dominated convergence theorem then follows. We only quote the statement and refer the reader to Section 2.2 of [GM4]: The proofs there presented work for the Lebesgue integral, but they can be extended verbatim to the integral with respect to a generic measure $(\mathcal{E}, \mu)$ on an arbitrary set X.

5.59 Lemma (Fatou). *Let $(\mathcal{E}, \mu)$ be a measure on a set X and $\{f_k\}$ a sequence of nonnegative and $\mathcal{E}$-measurable functions in $E \in \mathcal{E}$. Then*

$$\int_E \liminf_{k\to\infty} f_k \, d\mu \le \liminf_{k\to\infty} \int_E f_k \, d\mu.$$

5.60 Theorem (Lebesgue dominated convergence theorem). *Let $(\mathcal{E}, \mu)$ be a measure on a set X and let $\{f_k\}$ be a sequence of $\mathcal{E}$-measurable functions in $E \in \mathcal{E}$. If*

(i) *$f_k(x) \to f(x)$ for μ-a.e. $x \in E$,*
(ii) *there exists a μ-summable function $\phi : E \to \overline{\mathbb{R}}$ such that $|f_k(x)| \le \phi(x)$ for all k and for μ-a.e. $x \in E$,*

then

$$\int_E |f_k(x) - f(x)| \, d\mu(x) \to 0;$$

in particular, $\int_E f_k \, d\mu \to \int_E f \, d\mu$.

c. Linearity of integral

5.61 Proposition. *Let $(\mathcal{E}, \mu)$ be a measure in a set X. The integral is a linear operator, more precisely:*

(i) *Let $E \in \mathcal{E}$ and let $f, g \in \mathcal{L}^1(E)$ be either μ-summable functions in E or μ-integrable and nonnegative in E. Then*

$$\int_E (f + g) \, d\mu = \int_E f \, d\mu + \int_E g \, d\mu.$$

(ii) *If f is μ-integrable in E and $\lambda \in \mathbb{R}$, then $\int_E \lambda f \, d\mu = \lambda \int_E f \, d\mu$.*

Proof. (i) Assume f, g are $\mathcal{E}$-measurable and nonnegative. According to Lemma 5.45, let $\{\varphi_k\}$, $\{\psi_k\} \subset \mathcal{S}$ be such that $\varphi_k \uparrow f$, $\psi_k \uparrow g$. Then trivially $\varphi_k + \psi_k \uparrow f + g$. Moreover, since the integral is linear on simple and $\mathcal{E}$-measurable functions, $I_\mu(\varphi_k) + I_\mu(\psi_k) = I_\mu(\varphi_k + \psi_k)$. Beppo Levi's theorem then yields

$$\int (f+g)\,d\mu = \lim_{k\to\infty} I_\mu(\varphi_k+\psi_k) = \lim_{k\to\infty} I_\mu(\varphi_k) + \lim_{k\to\infty} I_\mu(\psi_k)) = \int_E f\,d\mu + \int_E g\,d\mu.$$

(ii) Let f, g be μ-summable. By decomposing f and g in their positive and negative parts, we see, again by Lemma 5.45, that there are $\{\varphi_k\}$, $\{\psi_k\} \subset \mathcal{S}$ such that $\varphi_k \to f$, $\psi_k \to g$ pointwise, $|\varphi_k| \le |f|$, $|\psi_k| \le |g|$. Therefore $\varphi_k + \psi_k \to f + g$ pointwise and $|\varphi_k + \psi_k| \le |f| + |g|$. Moreover, since the integral is linear on simple and $\mathcal{E}$-measurable functions, $I_\mu(\varphi_k) + I_\mu(\psi_k) = I_\mu(\varphi_k + \psi_k)$. Lebesgue's dominated convergence then yields

$$I_\mu(\varphi_k) \to \int f\,d\mu, \qquad I_\mu(\psi_k) \to \int g\,d\mu,$$

and

$$I_\mu(\varphi_k) + I_\mu(\psi_k) = I_\mu(\varphi_k+\psi_k) \to \int (f+g)\,d\mu,$$

hence the first claim.

The second claim of the theorem follows at once from the definition of integral. □

d. Cavalieri formula

5.62 Theorem. *Let $(\mathcal{E}, \mu)$ be a measure on a set X, and E a measurable set in X, $E \in \mathcal{E}$. For every nonnegative $\mathcal{E}$-measurable function $f : E \subset \mathbb{R}^n \to \overline{\mathbb{R}}$ we have*

$$\int_E f\,d\mu = \int_0^{+\infty} \mu(\{x \in E \mid f(x) > t\})\,d\mathcal{L}^1(t).$$

Notice that the function $t \mapsto \mu(\{x \in E \mid f(x) > t\})$ is nonincreasing, hence Riemann integrable on bounded sets.

Proof. Let φ be a simple $\mathcal{E}$-measurable nonnegative function, $\varphi(x) = \sum_{j=1}^N a_j \chi_{E_j}(x)$ where the a_j's are distinct and $E_j := \{x \mid \varphi(x) = a_j\} \in \mathcal{E}$. For $t > 0$ set $\alpha_j(t) := \chi_{[0,a_j]}(t)$. For all $t > 0$ we have

$$\mu\Big(\Big\{x \in X \,\Big|\, \varphi(x) > t\Big\}\Big) = \sum_{j=1}^N \alpha_j(t)\mu(E_j),$$

hence

$$\begin{aligned}\int_0^{+\infty} \mu\Big(\{x \mid \varphi(x) > t\}\Big)\,dt &= \sum_{j=1}^N \Big(\int_0^{+\infty} \alpha_j(t)\,dt\Big)\mu(E_j)\\ &= \sum_{j=1}^N a_j\,\mu(E_j) = \int \varphi\,d\mu,\end{aligned}$$

i.e., the Cavalieri formula holds for the simple $\mathcal{E}$-measurable functions. The general case follows by approximation, Lemma 5.45 taking into account the continuity of the measure for increasing sequences of sets and Beppo Levi's theorem. □

e. Chebycev's inequality

Let $(\mathcal{E},\mu)$ be a measure on a set X, $E\in\mathcal{E}$. For a nonnegative $\mathcal{E}$-measurable function $f : E\subset X\to\overline{\mathbb{R}}$ and for $t>0$, let $E_{f,t} := \{x\in E\,|\,f(x)>t\}$. Trivially, $t\le f(x)$ on $E_{f,t}$; the monotonicity of the integral then yields

$$\mu(E_{f,t}) \le \frac{1}{t}\int_{E_{f,t}} f\,d\mu \qquad \forall t>0. \tag{5.8}$$

The inequality in (5.8) is very useful in many instances; for this reason it is referred to with many names: *weak estimate*, *Markov's inequality* and *Chebycev's inequality*. The nonincreasing function $t\to\mu(E_{f,t})$ is sometimes called the *distribution function* of f, see Chapter 9 of [GM2].

f. Null sets and the integral

5.63 Theorem. *Let $(\mathcal{E},\mu)$ be a measure on a set X, $E\in\mathcal{E}$ and $f : E\to\overline{\mathbb{R}}$ a nonnegative and μ-summable function. Then $f(x)<+\infty$ for μ-a.e. $x\in E$.*

Proof. Let $C := \int_E f\,d\mu$. According to (5.8), for any $k\in\mathbb{R}$ we have

$$\mu\Big(\Big\{x\in E\,\Big|\,f(x)=+\infty\Big\}\Big) \le \mu\Big(\Big\{x\in E\,\Big|\,f(x)>k\Big\}\Big) \le \frac{C}{k}.$$

Hence $\mu(\{\in E\,|\,f(x)=+\infty\})=0$. □

5.64 Theorem. *Let $(\mathcal{E},\mu)$ be a measure on X, $E\in\mathcal{E}$ and $f : E\to\overline{\mathbb{R}}$ a nonnegative and $\mathcal{E}$-measurable function. Then $\int_E f\,d\mu=0$ if and only if $f(x)=0$ for μ-a.e. $x\in E$.*

Proof. Suppose $f(x)=0$ for μ-a.e. x. Then any simple minorant φ of f is nonzero only on sets of zero measure, hence $I_\mu(\varphi)=0$ and, by definition of integral $\int_E f(x)\,d\mu(x)=0$.

Conversely, from (5.8)

$$k\,\mu\Big(\Big\{x\in E\,\Big|\,f(x)>1/k\Big\}\Big) \le \int_E f(x)\,d\mu(x)=0,$$

hence $\mu(\{x\in E\,|\,f(x)>1/k\})=0$ for all integers k. Since

$$\Big\{x\in E\,\Big|\,f(x)>0\Big\} = \bigcup_k \Big\{x\in E\,\Big|\,f(x)>1/k\Big\}$$

we conclude, see Proposition 5.18, that $\mu(\{x\in E\,|\,f(x)>0\})=0$. □

5.65 ¶. Show that, if f and g are integrable on E and $f(x)\le g(x)$ for μ-a.e. $x\in E$ and $\int_E f(x)\,dx=\int_E g(x)\,dx$, then $f(x)=g(x)$ for μ-a.e. $x\in E$.

g. Convergence theorems

Actually, all of the convergence results in Section 2.2 of [GM4] extend verbatim to integrals with respect to a generic measure $(\mathcal{E}, \mu)$ on an arbitrary set X. We only quote the statements, as the relative proofs are trivial variations of the proofs provided in Section 2.2 of [GM4].

5.66 Proposition (Total convergence of series). *Let $(\mathcal{E}, \mu)$ be a measure on a set X and $\{f_k\}$ be a sequence of nonnegative $\mathcal{E}$-measurable functions $f_k : E \to \overline{\mathbb{R}}$. Then*

$$\int_E \sum_{k=1}^{\infty} f_k \, d\mu = \sum_{k=1}^{\infty} \int_E f_k \, d\mu.$$

5.67 Theorem (Lebesgue). *Let $(\mathcal{E}, \mu)$ be a measure on a set X and let $\{f_k\}$ be a sequence of $\mathcal{E}$-measurable functions in E such that*

$$\sum_{k=0}^{\infty} \int_E |f_k(x)| \, d\mu(x) < +\infty.$$

Then the series $\sum_{k=0}^{\infty} f_k(x)$ converges absolutely for μ-a.e. $x \in E$ to a μ-summable function f and

$$\int_E \Big| f(x) - \sum_{k=0}^{p} f_k(x) \Big| \, dx \to 0 \qquad \text{as } p \to \infty. \tag{5.9}$$

In particular,

$$\int_E f \, d\mu = \sum_{k=0}^{\infty} \int_E f_k \, d\mu.$$

5.68 Theorem (Absolute continuity of the integral). *Let $(\mathcal{E}, \mu)$ be a measure on a set X and let $f : E \to \overline{\mathbb{R}}$ be μ-summable. Then for every $\epsilon > 0$ there exists $\delta > 0$ such that for every measurable subset $F \subset E$ with $\mu(F) < \delta$ we have $\int_F |f| \, d\mu < \epsilon$. Equivalently,*

$$\int_E f \, d\mu \to 0 \qquad \text{as } \ \mu(E) \to 0.$$

5.69 Theorem (Continuity). *Let A be a metric space, X a set and $(\mathcal{E}, \mu)$ a measure on X. Let $f : A \times X \to \overline{\mathbb{R}}$ be such that*

(i) *for μ-a.e. $x \in X$ the function $t \to f(t, x)$ is continuous in A,*
(ii) *$\forall t \in A$ the function $x \to f(t, x)$ is μ-summable,*
(iii) *there exists a μ-summable function ϕ such that*

$$|f(t, x)| \le \phi(x) \qquad \text{for all } t \in A \text{ and } \mu\text{-a.e. } x \in X, \tag{5.10}$$

then the function

$$F(t) := \int f(t,x)\, d\mu(x), \qquad t \in A,$$

is continuous in A.

5.70 Theorem (Differentiation). *Let A be an open set in $\mathbb{R}^k$, X a set and $(\mathcal{E}, \mu)$ a measure on X. Denote by $t = (t_1, t_2, \dots, t_k)$ the coordinates in A, and let $f : A \times X \to \overline{\mathbb{R}}$ $f = f(t,x)$ be such that*

(i) *$x \to f(t,x)$ is μ-summable for all $t \in A$,*
(ii) *f has a partial derivative in the variable t_j at (t,x) for all t and for μ-a.e. x,*
(iii) *there exists a μ-summable function ϕ such that*

$$\left|\frac{\partial f}{\partial t_j}(t,x)\right| \le \phi(x) \qquad \text{for all } t \in A \text{ and for } \mu\text{-a.e. } x. \tag{5.11}$$

Then the function

$$F(t) := \int f(t,x)\, d\mu(x), \qquad t \in A,$$

has a partial derivative with respect to t_j at t for all $t \in A$ and

$$\frac{\partial F}{\partial t_j}(t) = \int \frac{\partial f}{\partial t_j}(t,x)\, d\mu(x) \qquad \forall t \in A.$$

h. Riemann integrable functions

We now restrict ourselves to the special case of the Lebesgue measure $(\mathcal{M}, \mathcal{L}^n)$ in $\mathbb{R}^n$ where $\mathcal{M}$ denotes the σ-algebra of the $\mathcal{L}^{n*}$ measurable sets and, in fact, to $n = 1$.

Let $f(x)$, $x \in [a,b]$, be a function that is nonnegative, bounded, with compact support (zero outside an interval $[a,b]$) and Riemann integrable. According to the Riemann definition of integral, for any $\epsilon > 0$ there exist two simple functions $\varphi(x) = \sum_{j=1}^N a_j \chi_{I_j}(x)$ and $\psi(x) = \sum_{j=1}^N b_j \chi_{I_j}(x)$ constant on each interval I_j such that $\varphi(x) \le f(x) \le \psi(x)$ $\forall x$,

$$I(\varphi) \le \text{Riemann} \int f(x)\, dx \le I(\psi)$$

and $I(\psi) - I(\varphi) < \epsilon$; here, the integral of a simple function is given by $I(\varphi) = \sum_{j=1}^N a_j |I_j|$.

Denote by $SG_{f,[a,b]}$ the subgraph of f,

$$SG_{f,[a,b]} := \Big\{ (x,t) \,\Big|\, x \in [a,b], 0 \le t \le f(x) \Big\},$$

then trivially $SG_{\varphi,[a,b]} \subset SG_{f,[a,b]} \subset SG_{\psi,[a,b]}$ and, since $SG_{\psi,[a,b]}$ and $SG_{\varphi,[a,b]}$ are the union of finitely many intervals, we infer that

$$\mathcal{L}^{2*}(SG_{\psi,[a,b]} \setminus SG_{f,[a,b]}) < \epsilon.$$

Since ϵ is arbitrary, we infer that $SG_{f,[a,b]}$ is $\mathcal{L}^{n+1}$-measurable, and, on account of Fubini's theorem, see Theorem 5.84 below, that f is $\mathcal{L}^n$-measurable. Moreover, since φ and ψ are simple and $\mathcal{L}^n$-measurable functions, we also have

$$I(\varphi) \leq \text{Lebesgue} \int f(x)\,dx \leq I(\psi),$$

thus we can state the following.

5.71 Proposition. *If f is Riemann integrable, then f is $\mathcal{L}^n$-measurable and*

$$\textit{Riemann} \int f(x)\,dx = \textit{Lebesgue} \int f(x)\,dx.$$

Fubini's theorem, Theorem 5.84, has been used in the previous argument to prove that f is $\mathcal{L}^n$-measurable. Alternatively, the measurability of f follows from the following characterization of Riemann integrable functions due to Giuseppe Vitali (1875–1932) in terms of Lebesgue's measure.

Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function with compact support ($f = 0$ outside $[a, b]$). For $k = 1, 2 \ldots$, set

$$M_k(x) := \sup_{B(x,1/k)} f(t)$$

and

$$f^*(x) := \lim_{k\to\infty} M_k(x) = \limsup_{t\to x} \max(f(t), f(x)).$$

It is not difficult to prove that

(i) f^* is upper semicontinuous and $f^*(x) \geq f(x)$ for all $x \in \mathbb{R}$,
(ii) f^* is the smallest upper semicontinuous function that is not smaller than f.

Sometimes f^* is called the *upper semicontinuous regularization* of f. Similarly, one defines the *lower semicontinuous regularization* of f,

$$f_*(x) := \lim_{k\to\infty} m_k(x) = \liminf_{t\to x} \min(f(t), f(x)),$$

where $m_k(x) := \inf_{B(x,1/k)} f(t)$. Of course,

(i) *f is continuous at x_0 if and only if $f_*(x_0) = f^*(x_0)$,*
(ii) $f_*(x) \leq f(x) \leq f^*(x)$ $\forall x$,
(iii) *f_* and f^* are $\mathcal{L}^n$-measurable.*

Denote by $\mathcal{S}^{el}$ the class of elementary simple functions, i.e., piecewise constant functions on intervals and vanishing outside $[a,b]$. If $\varphi \in \mathcal{S}^{el}$, $\varphi \geq f$ in $[a,b]$ and φ is continuous at x_0, then $\varphi(x_0) \geq \sup_{B(x_0,r)} f(t)$ for some $r > 0$. This yields $f^*(x_0) \leq \varphi(x_0)$, hence $f^*(x) \leq \varphi(x)$ at all points of continuity of φ and in conclusion at all points of $[a,b]$ except finitely many. It follows

$$\int f^*(x)\,dx \leq \int \varphi(x)\,dx \qquad \forall \varphi \in \mathcal{S}^{el},\ \varphi \geq f.$$

Indeed we have

$$\int f^*(x)\,dx = \inf\Big\{ \int \varphi(x)\,dx \,\Big|\, \varphi \in \mathcal{S}^{el},\ \varphi \geq f \Big\}$$

and, similarly,

$$\int f_*(x)\,dx = \sup\Big\{ \int \varphi(x)\,dx \,\Big|\, \varphi \in \mathcal{S}^{el},\ \varphi \leq f \Big\}.$$

In other words, f is Riemann integrable if and only if

$$\int f_*(x)\,dx = \int f^*(x)\,dx$$

and, since $f_* \leq f^*$, if and only if $f_*(x) = f(x) = f^*(x)$ for a.e. x. In particular, f is $\mathcal{L}^n$-measurable and we can readily state the following.

5.72 Theorem (Vitali). *A bounded function with compact support is Riemann integrable if and only if it is continuous at $\mathcal{L}^1$-a.e. $x \in \mathbb{R}$.*

5.73 ¶. Provide all details to prove the previous claims.

5.74 ¶. Let f be a measurable function on E and let

$$\int_E^* f(x)\,dx := \inf\Big\{ \int_E \varphi(x)\,dx \,\Big|\, \varphi \text{ simple, } \varphi \geq f \text{ a.e.} \Big\},$$
$$\int_{*E} f(x)\,dx := \sup\Big\{ \int_E \varphi(x)\,dx \,\Big|\, \varphi \text{ simple, } \varphi \leq f \text{ a.e.} \Big\}.$$

Show that f is integrable if and only if

$$\int_E^* f(x)\,dx = \int_{*E} f(x)\,dx$$

and, in this case, $\int_E f(x)\,dx = \int_E^* f(x)\,dx = \int_{*E} f(x)\,dx$.

5.75 ¶. Prove the following claims.

(i) Every lower (upper) semicontinuous function with compact support is an increasing (decreasing) limit of continuous functions f_k with compact support, hence f is measurable and

$$\int f(x)\,dx = \lim_{k\to\infty} \int f_k(x)\,dx.$$

(ii) f is Lebesgue integrable if and only if for all $\epsilon > 0$ there are lower and upper semicontinuous functions, respectively g and h, such that

$$g \leq f \leq h \quad \text{and} \qquad \int (h(x) - g(x))\,dx < \epsilon.$$

5.3 Product Spaces and Measures

In this section we discuss the construction of product measures and how to compute the corresponding integrals by means of iterated integrals. We first consider the special case of the Lebesgue measure in $\mathbb{R}^n$ spaces.

a. $\mathbb{R}^n$ spaces and Lebesgue measures

We begin with the following lemma.

5.76 Lemma (Products). *Let E be a measurable set in $\mathbb{R}^n$ and $L > 0$. Then the sets $E\times]0,L[$, $E \times [0,L]$, $E\times]0,L]$ and $E \times [0,L[$ are $\mathcal{L}^{n+1}$-measurable and*

$$\mathcal{L}^{n+1}(E) = L \cdot \mathcal{L}^n(E).$$

Proof. Let us prove that $E\times]0,L[$ is $\mathcal{L}^{n+1}$-measurable if E is $\mathcal{L}^n$-measurable. The other claims are proved similarly.

(i) *The claim is true if $E = I$ is an interval, an open interval or a closed interval of $\mathbb{R}^n$.* If I is an interval, $I\times]0,L]$ is an interval in $\mathbb{R}^{n+1}$ and $|I\times]0,L]| = L \cdot |I|$. On the other hand, $I\times]0,L[$ differs from $I\times]0,L]$ possibly for some face of its boundary, hence for a zero set. Thus, $I\times]0,L[$ is measurable and $|I\times]0,L[| = L\,|I|$. A similar argument works for I open or closed.

(ii) *If the claim holds for disjoint sets E and F, then it holds for $E \cup F$*, and *if it holds for E and F and $E \subset F$, then it holds for $F \setminus E$.*

(iii) *Let $E = \cup_j E_j$ be the increasing union of measurable sets E_j and the claim holds for each E_j, then the claim holds for E.* In fact, $E\times]0,L[= \cup_j (E_j\times]0,L[)$ and the increasing union of measurable sets is measurable. Passing to the limit we also have

$$|E\times]0,L[| = \lim_{j\to\infty} |E_j\times]0,L[| = L \lim_{j\to\infty} |E_j| = L\,|E|.$$

From (i), (ii) and (iii) the claim holds when E is open, compact or a denumerable union of closed sets.

(vi) *The claim holds if E is a zero set.* In fact, for all $\epsilon > 0$ there is an open set $A \supset E$ with $|A| < \epsilon$. Hence $E\times]0,L[\subset A\times]0,L$ and, on account of (i) and (iii),

$$\mathcal{L}^{n+1}(E\times]0,L[) \le |A\times]0,L[| = L\,|A| \le L\,\epsilon.$$

Since ϵ is arbitrary, we conclude that $|E\times]0,L[| = 0$.

(vii) Finally, since every measurable set agrees with a denumerable union of closed sets apart from a zero set, again (ii) yields that the claim holds for all measurable sets E. □

5.77 ¶. Show that $E := \mathbb{R}^{n-1} \times \{0\} \subset \mathbb{R}^n$ has zero n-dimensional Lebesgue measure.

5.78 ¶. Show that for all $A \subset \mathbb{R}^n$ we have $\mathcal{L}^{(n+1)*}(A \times [0,L]) = L\,\mathcal{L}^{n*}(A)$.

5.79 ¶. Prove the following theorem.

Theorem. *Suppose that $E \subset \mathbb{R}^n$ is $\mathcal{L}^n$-measurable and $F \subset \mathbb{R}^k$ is $\mathcal{L}^k$-measurable. Then $E \times F \subset \mathbb{R}^n \times \mathbb{R}^k$ is $\mathcal{L}^{n+k}$-measurable and $\mathcal{L}^{n+k}(E \times F) = \mathcal{L}^n(E)\mathcal{L}^k(F)$. (Here $0 \cdot \infty = 0$.)*

Finally, show that $\mathcal{L}^{(n+k)*}(E \times F) = \mathcal{L}^{n*}(E)\mathcal{L}^{k*}(F)$ if at least one of E or F is measurable.

5.80 Theorem. *Let $f : E \subset \mathbb{R}^n \to \overline{\mathbb{R}}$ be a nonnegative measurable function defined in the measurable set E. Then the subgraph of f*

$$SG_{f,E} = \Big\{(x,t) \,\Big|\, x \in E,\ 0 < t < f(x)\Big\}$$

is $\mathcal{L}^{n+1}$-measurable and

$$\int_E f(x)\,dx = \mathcal{L}^{n+1}(SG_{f,E}). \tag{5.12}$$

Proof of Theorem 5.80. We divide the proof into two steps.

(i) The claim is true if f is simple and $E = \mathbb{R}^n$. For a measurable subset F in $\mathbb{R}^n$

$$\mathcal{L}^{n+1}(F[\times[0,a]) = a \cdot \mathcal{L}^n(F) = \int_{\mathbb{R}^n} a\chi_F(x)\,dx,$$

according to Lemma 5.76. Consequently, from the additivity of the $(n+1)$-dimensional measure we readily infer that whenever $\varphi \in \mathcal{S}$, $\varphi(x) = \sum_{i=1}^N a_i E_i$, then $SG_{\varphi,\mathbb{R}^n}$ is $\mathcal{L}^{n+1}$-measurable and

$$\mathcal{L}^{n+1}(SC_{\varphi,\mathbb{R}^n}) = \sum_{i=1}^n \mathcal{L}^{n+1}(E_i \times [0,a_i]) = \sum_{i=1}^n a_i \mathcal{L}^n(E_i) = \int_{\mathbb{R}^n} \varphi(x)dx$$

by the additivity of the $\mathcal{L}^{n+1}$-measure and the definition of integral.

(ii) When $f : E \subset \mathbb{R}^n \to \overline{\mathbb{R}}$ is a generic nonnegative measurable function, we think of f as defined on the whole $\mathbb{R}^n$ setting $f(x) = 0$ is $x \notin E$. According to Lemma 5.45, we find a nondecreasing sequence of simple functions $\{\varphi_k\}$ such that $\varphi_k \uparrow f$ pointwise. Thus, $\{SG_{\varphi_k,E}\}$ is a nondecreasing sequence of sets in $\mathbb{R}^{n+1}$ and

$$SG_{f,E} = \cup_k SG_{\varphi_k,E}.$$

Thus, $SG_{f,E}$ is $\mathcal{L}^{n+1}$-measurable and, from (i), the continuity property of measures and Beppo Levi's theorem, we conclude

$$\mathcal{L}^{n+1}(SG_{f,E}) = \lim_{k\to\infty} \mathcal{L}^{n+1}(SG_{\varphi_k,E}) = \lim_{k\to\infty} \int_E \varphi_k\,dx = \int_E f(x)\,dx.$$

□

5.81 ¶. Let $f : E \to \overline{\mathbb{R}}$ be a measurable function on the measurable set $E \subset \mathbb{R}^n$. Show that the set

$$\Big\{(x,y) \in \mathbb{R}^n \times \mathbb{R} \,\Big|\, x \in E,\ y < f(x)\Big\}$$

is $\mathcal{L}^{n+1}$-measurable. [*Hint.* Decompose f as $f = f_+ - f_-$ and apply Theorem 5.80 to f_+ e f_-.]

5.82 ¶. Prove the following statement.

Proposition. *Let $f : E \subset \mathbb{R}^n \to \overline{\mathbb{R}}$ be a nonnegative and measurable function on E. Then*

$$\int_E f(x)\,dx = \sup \sum_j \Big(\inf_{x\in E_j} f(x)\Big)|E_j|,$$

where the supremum is taken among all partitions $\{E_j\}$ of E into finitely many disjoint measurable sets.

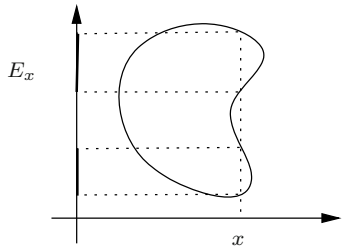


Figure 5.5. A slice E_x of E over x.

b. Fubini's theorem

In this paragraph we show how an n-dimensional integral can be computed by means of n 1-dimensional integrations.

Given a set E in $\mathbb{R}^{n+k}$ and a point $x \in \mathbb{R}^n$, we denote by

$$E_x := \left\{ y \in \mathbb{R}^k \,\middle|\, (x, y) \in E \right\}$$

the slice of E over x translated into the coordinate plane $\mathbb{R}^k$, see Figure 5.5.

5.83 Theorem (Fubini). *Let E be a $\mathcal{L}^{n+k}$-measurable set in $\mathbb{R}^{n+k}$. Then the following hold:*

(i) *For a.e. x the slice $E_x \subset \mathbb{R}^k$ is $\mathcal{L}^k$- measurable.*
(ii) *The function $x \to \mathcal{L}^k(E_x)$, $x \in \mathbb{R}^n$, is $\mathcal{L}^n$-measurable on $\mathbb{R}^n$.*
(iii) *We have*

$$\mathcal{L}^{n+k}(E) = \int_{\mathbb{R}^n} \mathcal{L}^k(E_x)\, dx.$$

The theorem, which on smooth sets E appears to be natural and even trivial, states that, under the assumption "E measurable", i.e., approximable in measure $\mathcal{L}^{n+k}$ by unions of intervals of dimension $n+k$, almost all of its slices E_x are approximable by unions of intervals of dimension k.

Proof. The proof consists in proving the claim for sets of increasing generality by means of the continuity property of measure and integral.

Step 1. The claim holds if E is an interval. Trivially, every interval $I \subset \mathbb{R}^{n+k}$ is a product $I = R \times S$, where $R \subset \mathbb{R}^n$ and $S \subset \mathbb{R}^k$ are intervals and $\mathcal{L}^{n+k}(R \times S) = \mathcal{L}^n(R)\mathcal{L}^k(S)$. For $x \in \mathbb{R}^n$ the slice $I_x \subset \mathbb{R}^k$ is then

$$I_x = \begin{cases} S & \text{if } x \in R, \\ \emptyset & \text{otherwise,} \end{cases}$$

hence I_x is $\mathcal{L}^k$-measurable. Moreover,

$$\mathcal{L}^k(I_x) = \begin{cases} \mathcal{L}^k(S) & \text{if } x \in R, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $x \to \mathcal{L}^k(I_x)$, $x \in \mathbb{R}^n$, is a simple function, thus $\mathcal{L}^n$-measurable and

XXXVII

(53). SUGLI INTEGRALI MULTIPLI

Rend. Lincei (5), 16, 1907, pp. 608-614.

1. Mi occuperò qui degli integrali superficiali di una funzione di due variabili x, y. E, come è oramai necessario in questo ordine di studi, mi riferirò agli integrali del LEBESGUE (1). Il teorema, che dimostreremo, è il seguente:

Se $f(x, y)$ è una funzione di due variabili x, y, limitata o illimitata, integrabile in un'area I' del piano (x, y), allora si ha sempre:

$$\int_{I'} f(x, y)\, d\sigma = \int dy \int f(x, y)\, dx = \int dx \int f(x, y)\, dy ,$$

quando con $d\sigma$ si intenda l'elemento d'area di I' (2).

Sull'area I' faremo dapprima l'ipotesi (del resto non essenziale) che la sua intersezione (3) con una qualsiasi retta $x = \text{cost}$, oppure $y = \text{cost}$ sia (linearmente) misurabile. (*Cfr.* il n° 2).

La Nota finisce con un breve cenno dell'estensione del precedente teorema a coordinate polari, o altre coordinate.

2. Introdurremo due definizioni: Diremo che un aggregato di punti nel piano (x, y) è linearmente misurabile, se ogni sua interse-

(1) LEBESGUE, *Intégrale, longueur, aire.* Annali di Matematica 1902.

(2) Quando la presente Nota era già in corso di stampa, mi fu fatto notare che in una osservazione a piè della pag. 30 della Memoria: *Sul principio di Dirichlet* (Rend. del Circ. Mat. di Palermo, tomo 22), il prof. B. LEVI accenna a questo teorema, partendo da alcuni lavori del sig. PRINGSHEIM sugli integrali superficiali di RIEMANN.

(3) Intersezione di un campo I', o di un aggregato E con una retta $x=\text{cost}$, oppure $y=\text{cost}$, è l'aggregato dei punti comuni al campo I', o l'aggregato E con la retta $x=\text{cost}$, (oppure $y=\text{cost}$) considerata. Al solito le x, y indicano coordinate cartesiane ortogonali.

LI

(71). SU UN TEOREMA RELATIVO AGLI INTEGRALI DOPPII (1)

Rend. Lincei (5), 22$_2$, 1913, p. 47.

Nel 1907 io avevo enunciato in questi Rendiconti un teorema che riduce il calcolo di un integrale superficiale a quello di un integrale doppio (iterato).

In una Nota recente ho enunciato il teorema reciproco, senza avvertire che nel 1909 (questi Rendiconti, 2° sem.) il prof. TONELLI aveva, a proposito di alcune sue ricerche *Sull'integrazione per parti*, dato un teorema, che è affatto equivalente a quello da me enunciato soltanto ora.

Credo quindi mio dovere riprodurre qui l'enunciato del prof. TONELLI:

Una funzione $f(x, y)$ misurabile superficialmente, per cui esista

$$\int_a^b dx \int_c^d |f(x, y)|\, dy ,$$

è integrabile superficialmente; e per essa vale la

$$\int_a^b dx \int_c^d f(x, y)\, dy = \int_a^b \int_c^d f(x, y)\, dx\, dy = \int_c^d dy \int_a^b f(x, y)\, dx .$$

(1) Pervenuto all'Accademia il 13 luglio 1913.

Figure 5.6. Two pages from the *Opere Scelte* by Guido Fubini (1879–1943) dealing with iterated integrals.

$$\int_{\mathbb{R}^n} \mathcal{L}^k(I_x)\, dx = \mathcal{L}^k(S)\mathcal{L}^n(R) = \mathcal{L}^{n+k}(R \times S).$$

Step 2. If the claim holds for disjoint measurable sets E and F, then it holds for $E \cup F$. In fact, $(E \cup F)_x = E_x \cup F_x$ is then $\mathcal{L}^k$-measurable for a.e. $x \in \mathbb{R}^n$, the function $x \to \mathcal{L}^k((E \cup F)_x) = \mathcal{L}^k(E_x) + \mathcal{L}^k(F_x)$ is $\mathcal{L}^n$-measurable and the equality (iii) for $E \cup F$ follows adding (iii) for E and F.

Step 3. If the claim holds for measurable sets E and F with $E \subset F$, then it holds for $F \setminus E$. One proceeds similarly to Step 2.

Step 4. If $E = \cup_j E_j$ is the union of a nondecreasing sequence $\{E_j\}$ of measurable sets and the claim holds for each E_j, then the claim holds for E. Since for every j the sets $E_{j,x}$ are $\mathcal{L}^k$-measurable for $\mathcal{L}^n$-a.e. x, and the j's are denumerable, we deduce that for $\mathcal{L}^n$-a.e. x, all of the sets $E_{j,x}$ are $\mathcal{L}^k$-measurable. Then E_x, that is equal to $\cup_j E_{j,x}$, is $\mathcal{L}^k$-measurable for $\mathcal{L}^n$-a.e. x. Moreover, $\mathcal{L}^k(E_x) = \lim_{j\to\infty} \mathcal{L}^k(E_{j,x})$ for $\mathcal{L}^n$-a.e. x, hence $x \to \mathcal{L}^k(E_x)$ is $\mathcal{L}^n$-measurable. Finally, because of Beppo Levi's theorem

$$\mathcal{L}^{n+k}(E) = \lim_{j\to\infty} \mathcal{L}^{n+k}(E_j) = \lim_{j\to\infty} \int_{\mathbb{R}^n} \mathcal{L}^k(E_{j,x})\, dx = \int_{\mathbb{R}^n} \mathcal{L}^k(E_x)\, dx.$$

A consequence of Steps 1, 2, 3 and 4 is that the theorem holds if E is open or a denumerable union of closed sets.

Step 5. The claim holds if $\mathcal{L}^{n+k}(E) = 0$. In this case, E is contained in the intersection of a denumerable and decreasing family of open sets A_j, $E \subset A = \cap_j A_j$, with $\mathcal{L}^{n+k}(A) = 0$. We may also assume that $\mathcal{L}^{n+k}(A_1) < +\infty$; as the theorem holds for A_1 that is open,

$$\int_{\mathbb{R}^n} \mathcal{L}^k(A_{1,x})\, dx = \mathcal{L}^{n+k}(A_1) < +\infty.$$

Therefore $\mathcal{L}^k(A_{1,x}) < +\infty$ for $\mathcal{L}^n$-a.e. x, consequently, since $A_x = \cap_{j=1}^{\infty} A_{j,x}$ for a.e. x,

$$\mathcal{L}^k(A_x) = \lim_{j\to\infty} \mathcal{L}^k(A_{j,x}) \qquad \text{for } \mathcal{L}^n\text{-a.e. } x$$

and, by Beppo Levi's theorem,

$$0 = \mathcal{L}^{n+k}(A) = \lim_{j\to\infty} \mathcal{L}^{n+k}(A_j) = \lim_{j\to\infty} \int_{\mathbb{R}^n} \mathcal{L}^k(A_{j,x})\,dx = \int_{\mathbb{R}^n} \mathcal{L}^k(A_x)\,dx,$$

that is $\mathcal{L}^k(A_x) = 0$ for $\mathcal{L}^n$-a.e. x.

Since $E \subset A$, also $\mathcal{L}^k(E_x) = 0$ for a.e. x and the function $x \to \mathcal{L}^k(E_x)$ vanishes a.e.. Trivially,

$$\int_{\mathbb{R}^n} \mathcal{L}^k(E_x)\,dx = \int_{\mathbb{R}^n} 0\,dx = 0 = \mathcal{L}^{n+k}(E).$$

Step 6. Since a measurable set E is a disjoint union of denumerable family of closed sets and of a set of zero measure, we see from the above that the theorem holds for E. □

5.84 Theorem (Fubini). *Let $E \subset \mathbb{R}^n$ be an $\mathcal{L}^n$-measurable set and $f : E \to \overline{\mathbb{R}}$ a nonnegative function. Then f is measurable in E if and only if its subgraph in $E \times \mathbb{R}$ is $\mathcal{L}^{n+1}$-measurable; in this case*

$$\mathcal{L}^{n+1}(SG_{f,E}) = \int_E f(x)\,dx.$$

Proof. In Theorem 5.80 we saw that the subgraph $SG_{f,E}$ of f is $\mathcal{L}^{n+1}$-measurable and

$$\mathcal{L}^{n+1}(SG_{f,E}) = \int_E f(x)\,dx$$

if f is measurable in E. It remains to prove that f is measurable if $SG_{f,E}$ is $\mathcal{L}^{n+1}$-measurable. For all $t > 0$

$$E_{f,t} = \Big\{x \in E \,\Big|\, f(x) > t\Big\} = \Big\{x \in \mathbb{R}^n, \,\Big|\, (x,t) \in SG_{f,E}\Big\} = (SG_{f,E})_t; \tag{5.13}$$

Theorem 5.83 then yields that $E_{f,t}$ is $\mathcal{L}^n$-measurable for $\mathcal{L}^1$-a.e. t. Consequently, $E_{f,t}$ is $\mathcal{L}^n$-measurable for all t, see Proposition 5.40. □

5.85 Theorem (Repeated integration). *Let $f : E \subset \mathbb{R}^{n+k} \to \overline{\mathbb{R}}$ be an integrable function in a measurable set E, and for $x \in \mathbb{R}^n$ let $E_x := \{y \in \mathbb{R}^k \,|\, (x,y) \in E\}$. Then the following hold:*

(i) *For a.e. x, the slice E_x is $\mathcal{L}^k$-measurable and the function $y \mapsto \varphi_x(y) := f(x,y)$ is $\mathcal{L}^k$-measurable on E_x.*

(ii) *The function*

$$x \to \int_{E_x} f(x,y)\,dy, \qquad x \in \mathbb{R}^n,$$

is $\mathcal{L}^n$-measurable.

(iii) *We have*

$$\int_E f(x,y)\,dx\,dy = \int_{\mathbb{R}^n} \left(\int_{E_x} f(x,y)\,dy \right) dx.$$

Proof. Assume in addition that f be nonnegative. Consider the subgraph of f in $\mathbb{R}^{n+k+1}$,

$$SG_{f,E} = \Big\{(x,y,t) \in \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R} \,\Big|\, (x,y) \in E, 0 < t < f(x,y)\Big\},$$

and, for $x \in \mathbb{R}^n$ the subgraph of $\varphi_x(y) := f(x,y)$, $y \in E_x$,

$$SG_{\varphi_x,E_x} := \Big\{(y,t) \in \mathbb{R}^k \times \mathbb{R} \,\Big|\, y \in E_x,\ 0 < t < f(x,y)\Big\} = (SG_{f,E})_x.$$

Fubini's theorem, Theorem 5.83, yields the following:

(i) E_x is $\mathcal{L}^k$-measurable and SG_{φ_x,E_x} is $\mathcal{L}^{k+1}$-measurable for a.e. x.
(ii) $x \to \mathcal{L}^{k+1}(SG_{\varphi_x,E_x})$, $x \in \mathbb{R}^n$, is $\mathcal{L}^n$-measurable.
(iii) We have

$$\mathcal{L}^{n+k+1}(SG_{f,E}) = \int_{\mathbb{R}^n} \mathcal{L}^{k+1}(SG_{\varphi_x.E_x})\,dx.$$

Theorem 5.84 then allows one to read (i), (ii) and (iii) above as the statement.

For generic f, it suffices to decompose f as $f = f_+ - f_-$ and apply the above. □

Fubini's theorem allows us to compute a multiple integral as iterated simple integrals, the order of integration being irrelevant. For instance,

$$\iint_E f(x,y)\,dx\,dy = \int \left(\int_{E_x} f(x,y)\,dy\right) dx,$$
$$\iint_E f(x,y)\,dx\,dy = \int \left(\int_{E_y} f(x,y)\,dx\right) dy,$$

where $E_x = \{y \,|\, (x,y) \in E\}$ and $E_y = \{x \,|\, (x,y) \in E\}$ both hold. More generally, we have the following.

5.86 Theorem (Tonelli). *Let $f : E \subset \mathbb{R}^{n+k} \to \overline{\mathbb{R}}$ be an integrable function, and let (x,y), $x \in \mathbb{R}^n$, $y \in \mathbb{R}^k$ be the coordinates in $\mathbb{R}^{n+k}$. The three integrals*

$$\int_{\mathbb{R}^n} \left(\int_{E_x} f(x,y)\,dy\right) dx, \qquad \int_{\mathbb{R}^k} \left(\int_{E_y} f(x,y)\,dx\right) dy$$

and

$$\iint_E f(x,y)\,dx\,dy$$

are well-defined and equal.

Notice that it is not assumed that the three integrals are finite but that f as function of $n+k$ variables is integrable. The following example shows that the *integrability assumption of f cannot be omitted.*

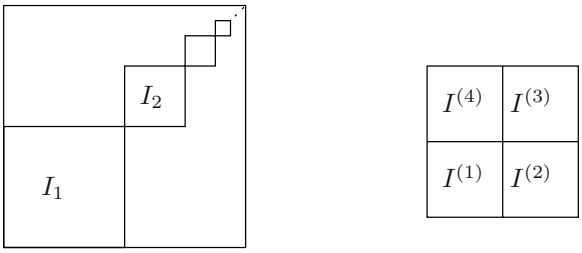


Figure 5.7. Illustrations (a) and (b) of Example 5.87.

5.87 Example. Let $a_k := 1 - 2^{-k}$ and $I_k :=]a_{k-1}, a_k] \times]a_{k-1}, a_k]$ be the sequence of squares in $I := [0,1] \times [0,1]$ shown in (a) of Figure 5.7. We now divide each I_k in four squares $I_k^{(1)}$ $I_k^{(2)}$, $I_k^{(3)}$ and $I_k^{(4)}$ as in (b) of Figure 5.7 and define $f : I \to \mathbb{R}$ by setting

$$f(x,y) := \begin{cases} \frac{1}{|I_k|} & \text{if } (x,y) \in I_k^{(1)} \cup I_k^{(3)}, \\ -\frac{1}{|I_k|} & \text{if } (x,y) \in I_k^{(2)} \cup I_k^{(4)}, \end{cases}$$

if $(x,y) \in \cup_k I_k$ and $f(x,y) = 0$ in $I \setminus \cup_k I_k$. It is readily seen that

$$\int_0^1 f(x,y)\,dy = 0 \ \forall x, \qquad \int_0^1 f(x,y)\,dx = 0 \ \forall y,$$

hence

$$\int_0^1 dx \int_0^1 f(x,y)\,dy = \int_0^1 dy \int_0^1 f(x,y)\,dx = 0.$$

However,

$$\iint_I f_+(x,y)\,dxdy = \sum_{k=1}^{\infty} \int_{I_k} f_+(x,y)\,dxdy = \sum_{k=1}^{\infty} \frac{1}{2} = +\infty$$

and, similarly, $\iint_I f_-(x,y)\,dxdy = +\infty$.

c. Product measures and repeated integration

We shall now deal with Fubini's theorem in the general abstract setting.

Let $(\mathcal{E}, \mu)$ and $(\mathcal{F}, \nu)$ be two measures on X and Y. Consider the set function $(\mathcal{I}, \lambda)$ in the Cartesian product $X \times Y$, where $\mathcal{I}$ is the family of "rectangles"

$$\mathcal{I} := \left\{ A = E \times F \,\middle|\, E \in \mathcal{E},\ F \in \mathcal{F} \right\}$$

and $\lambda : \mathcal{I} \to \overline{\mathbb{R}}_+$ is defined by $\lambda(A \times B) := \mu(A)\nu(B)$. It is readily seen that $\mathcal{I}$ *is a semiring and that* λ *is* σ*-additive on* $\mathcal{I}$. In fact, if $E \times F = \cup_k (E_k \times F_k)$, $E, E_k \in \mathcal{E}$, $F, F_k \in \mathcal{F}$, then

$$\chi_E(x)\,\chi_F(y) = \sum_{k=1}^{\infty} \chi_{E_k}(x)\chi_{F_k}(y) \qquad \forall x \in X,\ y \in Y;$$

by Beppo Levi's theorem, integrating on Y, we get

$$\chi_E(x)\,\nu(F) = \sum_{k=1}^{\infty} \nu(F_k)\chi_{E_k}(x) \qquad \forall x \in X,$$

and, integrating on X,

$$\lambda(E \times F) := \mu(E)\nu(F) = \sum_{k=1}^{\infty} \mu(E_k)\nu(F_k).$$

The *product measure* of $(\mathcal{E}, \mu)$ and $(\mathcal{F}, \nu)$ is, by definition, the measure $\mu \times \nu$ constructed from $(\mathcal{I}, \lambda)$ by Method I.

Since the set function λ is σ-additive, we infer, see Theorem 5.29 and Corollary 5.30, the following:

(i) For any μ-measurable set $A \subset X$ and any ν-measurable set $B \subset Y$ we have
$$(\mu \times \nu)(A \times B) = \lambda(A \times B) = \mu(A)\nu(B).$$
(ii) The rectangles in $\mathcal{I}$ are $\mu \times \nu$ measurable.
(iii) A set $E \subset X \times Y$ is $(\mu \times \nu)$ is σ-finite if and only if
$$E = \cap_{i=1}^{\infty} F_i \setminus N, \tag{5.14}$$
where $(\mu \times \nu)(N) = 0$ and $\{F_i\}$ is a nondecreasing sequence of $(\mu \times \nu)$-measurable sets with $(\mu \times \nu)(F_i) < +\infty$, each of which is the disjoint union of products $A \times B$ with A μ-measurable in X and B ν-measurable in Y.

5.88 ¶. Show that $\mathcal{L}^n \times \mathcal{L}^k = \mathcal{L}^{n+k}$ in $\mathbb{R}^n \times \mathbb{R}^k$.

5.89 ¶. Let μ be an outer measure on X and let $f : X \to \overline{\mathbb{R}}_+$ be a μ-summable function. Show that the subgraph of f,

$$SG_f := \Big\{(x,t) \in X \times \mathbb{R} \,\Big|\, 0 < t < f(x)\Big\},$$

is $\mu \times \mathcal{L}^1$-measurable and

$$\mu \times \mathcal{L}^1(SG_f) = \int_X f(x)\, d\mu.$$

[*Hint.* First assume that f is simple.]

Fubini's theorem then holds for the product measure $\mu \times \nu$. Let $A \subset X \times Y$. For $x \in A$ the slice A_x of A over x is the subset of Y

$$A_x := \Big\{y \in Y \,\Big|\, (x,y) \in A\Big\}.$$

5.90 Theorem (Fubini). *Let μ and ν be two outer measures in X and Y, respectively, and let $\mu \times \nu$ be the product measure of μ and ν on $X \times Y$. Suppose that $A \subset X \times Y$ is $(\mu \times \nu)$-measurable and $(\mu \times \nu)$ σ-finite. Then the following hold:*

(i) *A_x is ν-measurable for μ-a.e. $x \in X$.*

(ii) *$x \to \nu(A_x)$ is a μ-measurable function.*
(iii) *We have*

$$(\mu \times \nu)(A) = \int_X \nu(A_x)\, d\mu(x).$$

5.91 Example. The assumption that A is $\mu \times \nu$ σ-finite cannot be removed in general. For instance, consider the case in which $X = Y = \mathbb{R}$, $\mu = \mathcal{L}^1$ and ν is the counting measure. Let $S := \{(x, x) \,|\, x \in [0, 1]\}$ and let $f(x, y) = \chi_S(x, y)$ be its characteristic function. Since S is closed, S belongs to the smallest σ-algebra generated by intervals that in our case are sets $A \times B$ with $A \subset \mathbb{R}$ $\mathcal{L}^1$-measurable and B any set. Now $\mu \times \nu(S) = \infty$ but

$$\int d\mu \int f\, d\nu = \int_0^1 1\, dx = 1, \qquad \int d\nu \int f\, d\mu = \int_0^1 0\, d\nu = 0,$$

thus S is not $\mu \times \nu$ σ-finite.

Proof. The proof is a rewriting of the proof of Fubini's theorem for Lebesgue's measure. We include it for the reader's convenience.

Step 1. The claim holds if $I = E \times F$ is a rectangle with $E \in \mathcal{E}$ and $F \in \mathcal{F}$. For $x \in X$ the slice $I_x \subset Y$ is

$$I_x = \begin{cases} F & \text{if } x \in E, \\ \emptyset & \text{otherwise,} \end{cases}$$

hence I_x is ν-measurable for $x \in X$. Moreover,

$$\nu(I_x) = \begin{cases} \nu(F) & \text{if } x \in E, \\ 0 & \text{otherwise.} \end{cases}$$

hence $x \to \nu(I_x)$, $x \in X$, is a simple function, and, finally,

$$\int_X \nu(I_x)\, d\mu(x) = \nu(F)\mu(E) = \lambda(E \times F) = (\mu \times \nu)(I).$$

Step 2. If the claim holds for A and B disjoint, then it holds for $A \cup B$. In fact, $(A \cup B)_x = A_x \cup B_x$ is ν-measurable for all x, $\nu((A \cup B)_x) = \nu(A_x) + \nu(B_x)$ is μ-measurable, and we get (iii) for $A \cup B$.

Step 3. Inductively we find the following: *If the claim holds for a sequence of disjoint sets A_k, it holds for $\cup_{i=1}^N A_k$ for all N.*

Step 4. If $A = \cup_j A_j$ is the union of a nondecreasing sequence of sets $A_j \subset X \times Y$ and the claim holds for every A_j, then it holds for A, too. Since for each j the sets $A_{j,x}$ are ν-measurable for μ-a.e. x and the set of j's is denumerable, we deduce that for μ-a.e. x all the sets $A_{j,x}$ are ν-measurable. Therefore, $A_x = \cup_j A_{j,x}$ is ν-measurable for μ-a.e. x. By continuity, $\nu(A_x) = \lim_{j\to\infty} \nu(A_{j,x})$ for μ-a.e. x, hence $x \to \nu(A_x)$ is μ-measurable. Finally, by Beppo Levi's theorem

$$(\mu \times \nu)(A) = \lim_{j\to\infty} (\mu \times \nu)(A_j) = \lim_{j\to\infty} \int_X \nu(A_{j,x})\, d\mu(x) = \int_X \nu(A_x)\, d\mu(x).$$

Step 5. If the theorem holds for a nonincreasing sequence $\{A_j\}$ of sets with $\mu \times \nu(A_j) < +\infty$, then it holds for $A := \cap_{j=1}^\infty A_j$ too. Since $A_x = \cap_j A_{j,x}$ and the $A_{j,x}$ are ν-measurable for μ-a.e. $x \in X$, A_x is ν-measurable for μ-a.e. $x \in X$. From the assumption, in particular,

$$\int \nu(A_{1,x})\, d\mu(x) = (\mu \times \nu)(A_1) < +\infty,$$

hence $\nu(A_{1,x}) < +\infty$ for μ-a.e. x. Therefore, by continuity, we infer

$$\nu(A_x) = \nu(\cap_j A_{j,x}) = \lim_{j\to\infty} \nu(A_{j,x})$$

for μ-a.e. $x \in X$, thus concluding that $x \to \nu(A_x)$ is μ-measurable. Beppo Levi's theorem finally yields

$$\begin{aligned}(\mu\times\nu)(A) &= \lim_{j\to\infty}(\mu\times\nu)(A_j) = \lim_{j\to\infty}\int_X \nu(A_{j,x})\,d\mu(x)\\ &= \int_X \lim_{j\to\infty}\nu(A_{j,x})\,d\mu(x) = \int_X \nu(A_x)\,d\mu(x).\end{aligned}$$

Step 6. The claim holds if $\mu\times\nu(A) = 0$. In this case, $A \subset \cap B_i$, $B_i = \cup_j (E_{ij}\times F_{ij})$, $E_{ij} \in \mathcal{E}$, $F_{ij} \in \mathcal{F}$, $B_i \supset B_{i+1}$ $\forall i$ and $(\mu\times\nu)(B_i) < 1/i$. From the above, the claim holds for each B_i and for $B := \cap_i B_i$, hence

$$\int \nu(B_x)\,d\mu(x) = (\mu\times\nu)(B) = 0.$$

Therefore $\nu(B_x) = 0$ μ-a.e. $x \in X$. Since $A_x \subset B_x$, also A_x has zero ν-measure for μ-a.e. $x \in X$, hence E_x is ν-measurable for μ-a.e. $x \in X$, $x \to \nu(E_x)$ is μ-measurable and

$$\int_X \nu(A_x)\,d\mu(x) = \int_X 0\,d\mu(x) = 0 = (\mu\times\nu)(A).$$

Summing up, as in the case of Lebesgue's measure, we conclude the proof. □

5.92 ¶. Going through the proof of Fubini's theorem, show that if μ, ν are Borel measures on X and Y and if A is a Borel set, then A_x is Borel *for all* $x \in X$.

5.93 Definition. *Let μ be an outer measure on X. A function $f : X \to \overline{\mathbb{R}}$ is said to be μ σ-finite if the set $\{x \in X \mid f(x) \neq 0\}$ is μ σ-finite.*

5.94 Theorem (Fubini). *Let $(\mathcal{E}, \mu)$ and $(\mathcal{F}, \nu)$ be two measures on X and Y, respectively. If $f : X\times Y \to \overline{\mathbb{R}}$ is $(\mu\times\nu)$-integrable and if*

$$\left\{(x,y) \,\middle|\, f(x,y) \neq 0\right\}$$

is $(\mu\times\nu)$ σ-finite, then $y \to f(x,y)$ is ν-integrable for μ-a.e. $x \in X$, $x \to \int_Y f(x,y)\,d\nu(y)$ is μ-integrable and we have

$$\int f(x,y)\,d(\mu\times\nu)(x,y) = \int d\mu(x)\int f(x,y)\,d\nu(y).$$

Proof. If f is the characteristic function of a measurable and denumerable $\mu\times\nu$ finite set, we are back to Fubini's theorem, Theorem 5.90. Consequently, the theorem holds for $\mu\times\nu$-simple functions with $\mu\times\nu$ σ-finite level sets.

Suppose now that f is nonnegative and let $\{\varphi_k\}$ be the sequence of simple functions in Theorem 5.29. Then the φ_k's are simple, $\mu\times\nu$-measurable with $\mu\times\nu$ σ-finite level sets. Since the theorem holds for each φ_k, we see that it holds for f by applying the monotone convergence theorem of Beppo Levi. As usual, the general case then follows by splitting f into its positive and negative parts. □

5.95 Corollary (Tonelli). *Let $(\mathcal{E},\mu)$ and $(\mathcal{F},\nu)$ be two measures on X and Y, respectively. If $f : X \times Y \to \overline{\mathbb{R}}$ is $(\mu\times\nu)$-summable, then the following hold:*

(i) *$y \to f(x,y)$ is ν-summable for μ-a.e. $x \in X$.*
(ii) *$x \to \int_Y f(x,y)\,d\nu(y)$ is μ-summable.*
(iii) *We have*

$$\int f(x,y)\,d(\mu\times\nu)(x,y) = \int d\mu(x) \int f(x,y)\,d\nu(y).$$

Proof. In fact, a $\mu\times\nu$-summable function is $\mu\times\nu$-integrable and $\mu\times\nu$ σ-finite. □

We shall return to Fubini's theorem for Radon measures in Section 6.2.5.

5.4 Change of Variable in Lebesgue's Integral

This topic may be dealt with in several degrees of difficulties. Here we only deal with injective maps $\varphi : A \subset \mathbb{R}^n \to \mathbb{R}^n$ of class $C^1(A)$, A open.

a. Invariance under orthogonal transformations

5.96 Theorem. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. T maps measurable sets into measurable sets and*

$$\mathcal{L}^{n*}(T(E)) = |\det T|\,\mathcal{L}^{n*}(E) \qquad \forall E \subset \mathbb{R}^n. \tag{5.15}$$

In particular, Lebesgue's measure is invariant under orthogonal transformations.

Proof. Clear by intuition, (5.15) is not a one-line proof.

(i) First we prove that there is a constant $c(T)$ depending on T such that

$$\mathcal{L}^{n*}(T(E)) = c(T)\,\mathcal{L}^{n*}(E) \qquad \forall E \subset \mathbb{R}^n. \tag{5.16}$$

For that, we observe that in computing $\mathcal{L}^{n*}(E)$ we may just consider denumerable coverings of E made of cubes with sides parallel to the axes, that is,

$$\mathcal{L}^{n*}(E) = \inf\Big\{\sum_{k=1}^{\infty} |Q_k| \;\Big|\; \cup_k Q_k \supset E,\ Q_k \text{ cubes}\Big\}.$$

We also notice that each cube Q of the covering is congruent to $Q_1 := [0,1]^n$, that has volume 1 and that the outer measure is invariant under translations and positively homogeneous of degree n, i.e., if $Q = x_0 + \lambda Q_1$, then $|Q| = \mathcal{L}^n(Q) = \lambda^n\mathcal{L}^n(Q_1)$ and

$$\mathcal{L}^{n*}(T(Q)) = \mathcal{L}^{n*}(T(x_0) + \lambda T(Q_1)) = \mathcal{L}^{n*}(\lambda T(Q_1)) = \lambda^n\,\mathcal{L}^{n*}(T(Q_1)),$$

hence

$$\mathcal{L}^{n*}(T(Q)) = c(T)\,\mathcal{L}^n(Q), \qquad \text{where} \qquad c(T) := \frac{\mathcal{L}^{n*}(T(Q_1))}{\mathcal{L}^n(Q_1)}$$

for each cube Q with sides parallel to the axes. We then readily infer that

$$\mathcal{L}^{n*}(T(A)) = c(T)\,\mathcal{L}^{n*}(A) \qquad \text{for any open set } A \tag{5.17}$$

and

$$\mathcal{L}^{n*}(T(E)) \le c(T)\,\mathcal{L}^{n*}(E) \qquad \text{for any set } E \subset \mathbb{R}^n. \tag{5.18}$$

We now prove that $c(T)\mathcal{L}^{n*}(E) \le \mathcal{L}^{n*}(T(E))$. Of course, we may and do assume that $\mathcal{L}^{n*}(T(E)) < \infty$. For $\epsilon > 0$ we choose an open set B such that $B \supset T(E)$ and $\mathcal{L}^{n*}(B) \le \mathcal{L}^{n*}(T(E)) + \epsilon$. Then $T^{-1}(B)$ is open, $T^{-1}(B) \supset E$ and, because of (5.17),

$$c(T)\,\mathcal{L}^{n*}(E) \le c(T)\,\mathcal{L}^{n*}(T^{-1}(B)) = \mathcal{L}^{n*}(T(T^{-1}(B))) = \mathcal{L}^{n*}(B) \le \mathcal{L}^{n*}(T(E)) + \epsilon.$$

This proves the inequality and therefore (5.16).

(ii) Let $U : \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal transformation. U maps the unit ball into itself, hence, putting $E = B(0,1)$ in (5.16), we find $c(U) = 1$, i.e., $\mathcal{L}^{n*}(U(E)) = \mathcal{L}^{n*}(E)$: *$\mathcal{L}^{n*}$ is invariant under orthogonal transformations.*

(iii) Finally, let us prove that $c(T) = |\det T|$ for any linear map T. Using the singular value decomposition of T, we write T as $T = USV$, where U and V are orthogonal and S is diagonal with entries the singular values $\mu_1, \mu_2, \dots, \mu_n$ of T, the square roots of the eigenvalues of T^*T. S acts as a dilatation with factors $\mu_1, \mu_2, \dots, \mu_n$ in the axes directions. In particular, S transforms intervals into intervals and trivially

$$\mathcal{L}^n(S(I)) = \mu_1\mu_2\cdots\mu_n\mathcal{L}^n(I) = |\det(T^*T)|^{1/2}\,\mathcal{L}^n(I) = |\det T|\,\mathcal{L}^n(I),$$

hence, by (5.16), $\mathcal{L}^{n*}(S(E)) = |\det T|\,\mathcal{L}^{n*}(E)\ \forall E \subset \mathbb{R}^n$. Finally, from (ii)

$$\begin{aligned}\mathcal{L}^{n*}(T(E)) &= \mathcal{L}^{n*}(U(S(V(E))) = \mathcal{L}^{n*}(S(V(E)))\\ &= |\det T|\,\mathcal{L}^{n*}(V(E)) = |\det T|\,\mathcal{L}^{n*}(E).\end{aligned}$$

(iv) It remains to prove that $T(E)$ is measurable if E is measurable. This is a consequence of Proposition 5.98. □

5.97 ¶. Let S be the convex hull of $n+1$ points $x_0, \dots, x_n \in \mathbb{R}^n$ and let A be the $n \times n$ matrix with i-tuple column the vector $x_i - x_0$. Show that $\mathcal{L}^n(S) = |\det A|$.

b. Measurable maps and Lipschitz maps

5.98 Proposition. *Every locally Lipschitz map maps null sets into null sets and measurable sets into measurable sets.*

Proof. A continuous function $f : \mathbb{R}^n \to \mathbb{R}^n$ maps compact sets into compact sets, each closed set is union of a denumerable family of compact sets and $f(\cup_k E_k) = \cup_k f(E_k)$. Therefore, we readily conclude that a continuous function maps denumerable unions of closed sets into denumerable unions of closed sets. Consequently, $f(E)$ is measurable if E is closed. The theorem is completely proved because of the structure theorem for measurable sets if we show that locally Lipschitz maps map null sets into null sets.

Fix a cube K in $\mathbb{R}^n$ and denote by L_K the Lipschitz constant of f in K so that

$$|f(x) - f(y)| \le L_K|x - y| \qquad \forall x, y \in K.$$

A cube I of side r in K has diameter of length $r\sqrt{n}$. Since f is Lipschitz, $f(I)$ has diameter at most $r\sqrt{n}\,L_K$, hence it is contained in a cube of side $r\sqrt{n}\,L_K$, and in conclusion, $\mathcal{L}^{n*}(f(I)) \le n^{n/2}L_K^n r^n = n^{n/2}L_K^n\,\mathcal{L}^{n*}(I)$, i.e.,

$$\mathcal{L}^{n*}(f(E\cap K)) \le n^{n/2} L_K^n \, \mathcal{L}^{n*}(E\cap K) \qquad \forall E \subset \mathbb{R}^n. \tag{5.19}$$

In particular, we have proved that for fixed K, $f(E\cap K)$ has zero measure if E is a null set. By choosing an increasing sequence of cubes $\{K_j\}$ so that $\mathbb{R}^n = \cup_j K_j$, we conclude that $f(E) = \cup_j f(E\cup K_j)$ is a null set if E is a null set. □

In particular, functions of class C^1 map null sets into null sets. We notice that this is no longer true for continuous, or even α-Hölder-continuous functions with $0 < \alpha < 1$.

5.99 Example. Cantor–Vitali $f : [0,1] \to [0,1]$, see Section 5.1.3, maps the ternary Cantor set C, for which $\mathcal{L}^1(C) = 0$, into $f(C) = [0,1] \setminus \{y = k/2^n \,|\, k = 0, \dots, 2^n,\ n = 0, 1, \dots\}$ for which $\mathcal{L}^1(f(C)) = 1$. Recall that f is α-Hölder-continuous with exponent $\alpha = \frac{\log 2}{\log 3}$.

c. The area formula

5.100 Theorem. *Let A be an open set in $\mathbb{R}^n$ and let $\varphi : A \to \mathbb{R}^n$ be a map of class $C^1(A)$ and injective. For any measurable set $E \subset A$, $\varphi(E)$ is measurable and*

$$\mathcal{L}^n(\varphi(E)) = \int_E |\det \mathbf{D}\varphi(x)|\, dx. \tag{5.20}$$

Of course, the claim holds for $\cup_j E_j$ if it holds for each $E_j \subset A$. Moreover, we also know that it holds if $|E| = 0$, see Proposition 5.98. Therefore it suffices for the claim to be proved to assume that E is a closed cube contained in A.

The proof consists in approximating φ by piecewise linear maps and controlling the errors. Let $I \subset A$ be compact, and for $x_0 \in I$ denote by

$$Q(x_0, r) := \left\{ x \in \mathbb{R}^n \;\middle|\; |x^i - x_0^i| \le r,\ i = 1, \dots, n \right\}$$

the cube of sides r with sides parallel to the axes centered at x_0. Finally, denote by $L : \mathbb{R}^n \to \mathbb{R}^n$ the affine linear map

$$L(x) := \varphi(x_0) + \mathbf{D}\varphi(x_0)(x - x_0).$$

5.101 Lemma. *Suppose $\varphi : A \to \mathbb{R}^n$ is a diffeomorphism onto its image and let $I \subset A$ be compact. For each τ with $0 < \tau < 1$ there exists $\delta > 0$ such that for all $x_0 \in I$ and all ρ, $0 < \rho < \delta$, we have*

$$L(Q(x_0, (1-\tau)\rho)) \subset \varphi(Q(x_0, \rho)) \subset L(Q(x_0, (1+\tau)\rho)).$$

Proof. Introduce the norm $|x|_\infty := \max\{x^i \,|\, x = (x^1, x^2, \dots, x^n)\}$, so that $x \in Q(x_0, \rho)$ if and only if $|x - x_0|_\infty \le \rho$. Since $\varphi \in C^1(A)$, the mean value theorem yields

$$\varphi^i(x) - L^i(x) = (\mathbf{D}\varphi^i(\xi) - \mathbf{D}\varphi^i(x_0))(x - x_0)$$

for some $\xi = (\xi^1, \xi^2, \dots, \xi^n)$ in the segment joining x_0 and x (assuming that the segment joining x to x_0 is contained in A). Therefore, since $I \subset\subset A$, for all $\epsilon > 0$ there is a $\delta > 0$ such that

$$|\varphi(x) - L(x)|_\infty \le \epsilon\, |x - x_0|_\infty \tag{5.21}$$

whenever $x_0 \in I$ and $x \in Q(x_0, \delta)$.

On the other hand, since $\varphi : A \to \mathbb{R}^n$ is a diffeomorphism and I is strictly contained in A, we find a constant $\nu > 0$ such that

$$\nu\,|x-y|_\infty \le |\varphi(x)-\varphi(y)|_\infty, \qquad \nu\,|x-y|_\infty \le |L(x)-L(y)|_\infty \tag{5.22}$$

for all x, y in a neighborhood of I. For $x \in A$, define $x_1 := x_1(x) := L^{-1}(\varphi(x))$. Fix now $0 < \tau < 1$ and choose $\epsilon = \nu\tau$. Then, from (5.21) and (5.22), we infer the existence of $\delta > 0$ such that if $0 < \rho < \delta$ and $x \in Q(x_0, \rho)$, then

$$\nu|x_1 - x|_\infty \le |L(x_1) - L(x)|_\infty = |\varphi(x) - L(x)|_\infty \le \tau\,|x - x_0|_\infty \le \nu\tau\rho,$$

hence

$$|x_1 - x_0|_\infty \le |x_1 - x|_\infty + |x - x_0|_\infty \le \Big(\tau + 1\Big)\rho.$$

In other words, $\varphi(Q(x_0, \rho)) \subset L(Q(x_0, (1+\tau)\rho))$.

We proceed similarly to prove the other inclusion. In fact, let $x_1 := x_1(x) = \varphi^{-1}(L(x))$. Fix $0 < \tau < 1$ and choose $\epsilon = \nu\frac{\tau}{1-\tau}$. Then there exists $\delta > 0$ such that for $0 < \rho < \delta$ and $x \in Q(x_0, (1-\tau)\rho)$, we have

$$\nu|x_1 - x|_\infty \le |\varphi(x_1) - \varphi(x)|_\infty = |L(x) - \varphi(x)|_\infty \le \epsilon|x - x_0|_\infty \le \nu\frac{\tau}{1-\tau}(1-\tau)\rho,$$

hence

$$|x_1 - x_0|_\infty \le |x_1 - x|_\infty + |x - x_0|_\infty \le (1 - \tau + \tau)\rho = \rho.$$

This proves that $L(Q(x_0, (1-\tau)\rho)) \subset \varphi(Q(x_0, \rho))$. □

Proof of Theorem 5.100. (i) Suppose that $\varphi : A \to \mathbb{R}^n$ is a diffeomorphism. Given a cube $Q = Q(x_0, \rho)$, we denote by $Q_{-\tau}$ and Q_τ the cubes centered at x_0 with sides $(1-\tau)\rho$ and $(1+\tau)\rho$, respectively. For τ with $0 < \tau < 1$ and for any cube Q small enough with center in I, Lemma 5.101 yields

$$\mathcal{L}^n(L(Q_{-\tau})) \le \mathcal{L}^n(\varphi(Q)) \le \mathcal{L}^n(L(Q_\tau))$$

and, since L is affine, we find

$$(1-\tau)^n\,|\det \mathbf{D}\varphi(x_0)|\,\mathcal{L}^n(Q) \le \mathcal{L}^n(\varphi(Q)) \le (1+\tau)^n\,|\det \mathbf{D}\varphi(x_0)|\,\mathcal{L}^n(Q)$$

if we take into account (5.15). On the other hand, since $\mathbf{D}\varphi(x)$ is continuous, for sufficiently small cubes we have

$$\frac{1}{\mathcal{L}^n(Q)}\int_Q |\det \mathbf{D}\varphi(x)|\,dx - \tau \le |\det \mathbf{D}\varphi(x_0)| \le \frac{1}{\mathcal{L}^n(Q)}\int_Q |\det \mathbf{D}\varphi(x)|\,dx + \tau,$$

hence

$$(1-\tau)^n\bigg(\int_Q |\det \mathbf{D}\varphi(x)|\,dx - \tau\mathcal{L}^n(Q)\bigg) \le \mathcal{L}^n(\varphi(Q)) \le (1+\tau)^n\bigg(\int_Q |\det \mathbf{D}\varphi(x)|\,dx + \tau\mathcal{L}^n(Q)\bigg).$$

Covering I with sufficiently small cubes with disjoint interiors and summing the previous inequalities, we then conclude

$$(1-\tau)^n\bigg(\int_I |\det \mathbf{D}\varphi(x)|\,dx - \tau|I|\bigg) \le \mathcal{L}^n(\varphi(I)) \le (1+\tau)^n\bigg(\int_I |\det \mathbf{D}\varphi(x)|\,dx + \tau|I|\bigg).$$

Letting τ to zero yields the theorem under the extra assumption that $\varphi : A \to \mathbb{R}^n$ is a diffeomorphism.

(ii) In the general case, set $N := \{x \in A \mid \det \mathbf{D}\varphi(x) = 0\}$. Of course $A \setminus N$ is open and the implicit function theorem tells us that $\varphi : A \setminus N \to \mathbb{R}^n$ is a local diffeomorphism, hence a diffeomorphism since φ is injective. From the first part of the proof we then infer

$$\mathcal{L}^n(\varphi(E \setminus N)) = \int_{E\setminus N} |\det \mathbf{D}\varphi(x)|\, dx = \int_E |\det \mathbf{D}\varphi(x)|\, dx.$$

On the other hand, $\varphi(E) \setminus \varphi(E \setminus N) \subset \varphi(N)$ and $|\varphi(N)| = 0$ because Theorem 3.12 of [GM4], thus $\mathcal{L}^n(\varphi(E)) = \mathcal{L}^n(\varphi(E \setminus N))$. This concludes the proof. □

d. Change of variables in multiple integrals

As a consequence of the area formula we get the following.

5.102 Theorem (Change of variables formula). *Let $\varphi : A \to \mathbb{R}^n$ be a map of class C^1 and injective in an open set A. If $E \subset A$ is measurable, then $\varphi(E)$ is measurable. Moreover, $f : E \to \overline{\mathbb{R}}$ is integrable in $\varphi(E)$ if and only if $f \circ \varphi(x)|\det \mathbf{D}\varphi(x)|$ is integrable in E, and in this case*

$$\int_{\varphi(E)} f(y)\, dy = \int_E f(\varphi(x))|\det \mathbf{D}\varphi(x)|\, dx. \tag{5.23}$$

Proof. The theorem holds if f is the characteristic function of a measurable set, since it reduces to the area formula. By linearity, it holds for simple functions. As a consequence of Beppo Levi's theorem, it holds for nonnegative measurable function by passing to the limit on increasing sequences of simple functions, and, finally, it holds for general f's by decomposing them into their positive and negative parts. □

5.103 Remark. In specific situations (polar coordinates, etc.), a slight extension of Theorem 5.102 is convenient. Let $\varphi : A \subset \mathbb{R}^n \to \mathbb{R}^n$ be of class $C^1(A)$, A open, and let $E \subset A$ be a measurable set. Suppose φ is injective on an open set $B \subset E$ with $|E \setminus B| = 0$. Then the change of variable formula (5.23), and, consequently, the area formula (5.20) hold on E.

5.5 Exercises

5.104 ¶. Let $\mathcal{E} \subset \mathcal{P}(X)$ be a σ-algebra of subsets of X and let $\{E_i\} \subset \mathcal{E}$. Define

$$\liminf_{i\to\infty} E_i := \bigcup_{i=1}^{k} \bigcap_{i=k}^{\infty} E_i, \qquad \limsup_{i\to\infty} E_i := \bigcap_{i=1}^{k} \bigcup_{i=k}^{\infty} E_i$$

and show that $\liminf_{i\to\infty} E_i, \limsup_{i\to\infty} E_i \in \mathcal{E}$. Moreover, if $(\mathcal{E}, \mu)$ is a measure in X, show that

$$\mu(\liminf_{i\to\infty} E_i) \le \liminf_{i\to\infty} \mu(E_i), \qquad \limsup_{i\to\infty} \mu(E_i) \le \mu(\limsup_{i\to\infty} E_i).$$

[*Hint.* Notice that $\liminf_{i\to\infty} E_i$ is the set of points that are in all the E_i's except for a finite number and that $\limsup_{i\to\infty} E_i$ is the set of points of X that belong to infinitely many E_i's.]

5.105 ¶ Regularity property of measures. Show that the Lebesgue outer measure has the following regularity property.

Proposition (Regularity). *Let $E \subset \mathbb{R}^n$ be a set. For each $\epsilon > 0$ there is an open set $A \supset E$ such that $\mathcal{L}^{n*}(A) \leq \mathcal{L}^{n*}(E) + \epsilon$, hence*

$$\mathcal{L}^{n*}(E) = \inf\Big\{ \mathcal{L}^{n*}(A) \,\Big|\, A \text{ open},\ A \supset E \Big\}.$$

Prove then the following.

Corollary. *For each set $E \subset \mathbb{R}^n$ there exists a set F that is a denumerable intersection of open sets such that $F \supset E$ and $\mathcal{L}^{n*}(F) = \mathcal{L}^{n*}(E)$.*

[*Hint.* It is not restrictive to assume that E has finite outer measure. Cover E with a denumerable family of intervals $P = \cup_k I_k$ with $\sum_{k=1}^{\infty} |I_k| < \mathcal{L}^{n*}(E) + \epsilon$; and for all k consider an open interval J_k centered as I_k and homothetic to I_k such that $|J_k| \leq |I_k| + \epsilon\, 2^{-k}$.]

5.106 ¶. Let E and F be two measurable sets. Show that $|E \cup F| + |E \cap F| = |E| + |F|$.

5.107 ¶. Define the *inner measure* of E by

$$\mathcal{L}^n_*(E) := \sup\Big\{ |F| \,\Big|\, F \text{ closed},\ F \subset E \Big\}.$$

Under the assumption $\mathcal{L}^n_*(E) < +\infty$, show that E is measurable if and only if $\mathcal{L}^n_*(E) = \mathcal{L}^{n*}(E)$.

5.108 ¶. In Proposition 5.18 the hypothesis that at least one of the E_k's has finite measure cannot be omitted. Consider the cases $E_k =]k, +\infty[\subset \mathbb{R}$ and $E_k := \mathbb{R}^n \setminus B(0, k)$, k integer.

5.109 ¶. Show that for any increasing sequence $\{A_k\}$ of sets not necessarily measurable we have $\lim_{k\to\infty} \mathcal{L}^{n*}(A_k) = \mathcal{L}^{n*}(\cup_k A_k)$. [*Hint.* For each k consider a measurable set E_k with $|E_k| = \mathcal{L}^{n*}(A_k)$.]

5.110 ¶. Show that there are decreasing families of $\{E_k\}$ of subsets of $\mathbb{R}^n$ with $\mathcal{L}^{n*}(E_k) < +\infty$ such that

$$\mathcal{L}^{n*}\Big(\bigcup_{k=1}^{\infty} E_k \Big) < \sum_{k=1}^{\infty} \mathcal{L}^{n*}(E_k) \qquad \text{and} \qquad \lim_{k\to\infty} \mathcal{L}^{n*}(E_k) > \mathcal{L}^{n*}\Big(\bigcap_{k=1}^{\infty} E_k \Big).$$

5.111 ¶. Show a measurable set that is mapped into a nonmeasurable set by a continuous function. [*Hint.* Choose as the map the Cantor–Vitali function.]

5.112 ¶. (vii) of Proposition 5.40 can be used to define the class of measurable functions with values in $\mathbb{R}^k$: $f : E \subset \mathbb{R}^n \to \mathbb{R}^k$ is measurable if E is measurable and $f^{-1}(A)$ is measurable for every Borel set A in $\mathbb{R}^k$. Show that $f : E \subset \mathbb{R}^n \to \mathbb{R}^k$ is measurable if and only if all its components $f^1, \ldots, f^k : E \to \mathbb{R}$ are measurable.

5.113 ¶. Construct a measurable set $E \subset [0, 1]$ with the following property: For any subinterval $I \subset [0, 1]$, both $I \cap E$ and $I \setminus E$ have positive measure. [*Hint.* Consider a Cantor set C_σ, $\sigma < 1/3$, and on each interval of its complement construct another Cantor set and continue this way.]

5.114 ¶. Show measurable functions f and g for which $g \circ f$ is not measurable. [*Hint.* Choose as f the inverse function of $x + F(x)$ where $F(x)$ is the Cantor-Vitali function and as g the characteristic function of a suitable set E.]

5.115 ¶. Let $A := \{x \in \mathbb{R} \,|\, x = \sum_{k=i}^{\infty} \frac{a_k}{5^k},\ a_k \in \{0, 4\}\}$. Show that $|A| = 0$.

5.116 ¶. Let $(\mathcal{E}, \mu)$ be a measure on a set X and $p : X \to \overline{\mathbb{R}_+}$ be a nonnegative and μ-summable function. Show that $(\mathcal{E}, \nu)$, where ν is defined by

$$\nu(E) := \int_E p(x)\, d\mu(x), \qquad E \in \mathcal{E},$$

is a measure on X. Moreover, show that

$$\int_E f(x)\, d\nu(x) = \int_E f(x) p(x)\, d\mu(x)$$

for every $\mathcal{E}$-measurable and nonnegative function.

5.117 ¶ Functions with discrete range. Let $(\mathcal{E}, \mu)$ be a measure on a set Ω and let $f : \Omega \to \mathbb{R}$ be a nonnegative measurable function with a discrete range, i.e., that takes at most denumerable many values as, for instance, if Ω is denumerable. Show that

$$\int_\Omega f(x)\, d\mu(x) = \sum_{i=1}^{\infty} a_i \mu\Big(\Big\{x \,\Big|\, f(x) = a_i\Big\}\Big) + (+\infty)\mu\Big(\Big\{x \,\Big|\, f(x) = +\infty\Big\}\Big)$$

where $(+\infty) \cdot \mu(\{x \,|\, f(x) = +\infty\}) = 0$ when $\mu(\{x \,|\, f = +\infty\}) = 0$.

5.118 ¶ Probability on a denumerable set. Let $(\mathcal{E}, \mu)$ be a measure on a denumerable set Ω with density $p(x) := \mu(\{x\})$ and let $f : \Omega \to \mathbb{R}$ be a nonnegative $\mathcal{E}$-measurable function that may take the value $+\infty$. Show that

$$\int_\Omega f(x)\ d\mu(x) = \sum_{x \in \Omega} f(x) p(x).$$

5.119 ¶ The Dirac's delta. Let Ω be a set and $x_0 \in \Omega$. The function $\delta_{x_0} : \mathcal{P}(\Omega) \to \mathbb{R}$,

$$\delta_{x_0}(A) = \begin{cases} 1 & \text{if } x_0 \in A, \\ 0 & \text{otherwise,} \end{cases}$$

called the *Dirac's delta*[2] at x_0, is a probability measure on Ω. Show that

$$\int_\Omega f(x)\, d\delta_{x_0}(x) = f(x_0).$$

5.120 ¶ Sum of measures. Let $(\mathcal{E}, \alpha)$ and $(\mathcal{E}, \beta)$ be two measures in Ω and let $\lambda \in \mathbb{R}$. Show that $\alpha + \beta : \mathcal{E} \to \overline{\mathbb{R}}_+$ and $\lambda\alpha : \mathcal{E} \to \overline{\mathbb{R}}_+$ given by $(\alpha + \beta)(E) = \alpha(E) + \beta(E)$ and $(\lambda\alpha)(E) = \lambda\alpha(E)\ \forall E \in \mathcal{E}$, define two measures in Ω and

$$\int_\Omega f(x)\, (\alpha + \beta)(dx) = \int_\Omega f(x)\, \alpha(dx) + \int_\Omega g(x)\, \beta(dx),$$
$$\int_\Omega f(x)\, (\lambda\alpha)(dx) = \lambda \int_\Omega f(x)\, \alpha(dx).$$

for all $\mathcal{E}$-measurable and nonnegative f.

[2] Paul Dirac (1902–1984).

5.121 Example (Counting measure). Let Ω be a set and $\mathcal{H}^0(A) := \#\{x \,|\, x \in A\}$ the counting measure in Ω. Show that $(\mathcal{P}(\Omega), \mathcal{H}^0)$ is a measure, and

$$\int_\Omega f(x)\, \mathcal{H}^0(dx) = \sum_{x \in \Omega} f(x)$$

for all f that is nonzero only on finitely many points.

6. Hausdorff and Radon Measures

In this chapter we present the fundamental theorems of measure theory, such as the Lebesgue–Besicovitch differentiation theorem, the Stieltjes–Lebesgue theory of integral, the fundamental properties of Hausdorff measures and the general area and coarea formulas.

6.1 Abstract Measures

6.1.1 Positive Borel Measures

Let X be a metric space. A relevant class of measures in X are the *Borel-regular* measures, see Chapter 5. The Lebesgue measure in $\mathbb{R}^n$ is an example of Borel-regular measure. As we have seen, Method II of construction of measures produces Borel-regular measures. In the following, it is understood that Borel measures are defined on the σ-algebra of Borel sets $\mathcal{B}(X)$; we then write simply μ instead of $(\mathcal{B}(X), \mu)$ to denote a Borel measure on X.

Borel sets are quite complicated if compared to open sets that are simply denumerable unions of closed cubes with disjoint interiors. However, the following holds.

6.1 Theorem. *Let μ be a finite Borel measure on a metric space X. For each Borel set $E \subset \mathcal{B}(X)$ we have*

$$\mu(E) = \inf\Big\{\mu(A) \,\Big|\, A \supset E,\ A \text{ open}\Big\}, \tag{6.1}$$

$$\mu(E) = \sup\Big\{\mu(F) \,\Big|\, F \subset E,\ F \text{ closed}\Big\}. \tag{6.2}$$

Proof. Consider the family

$$\mathcal{A} := \Big\{E \text{ Borel} \,\Big|\, (6.1) \text{ holds true for } E\Big\}.$$

Of course, $\mathcal{A}$ contains the family of open sets. We prove that $\mathcal{A}$ *is closed under denumerable unions and intersections.* Let $\{E_j\} \subset \mathcal{A}$ and, for $\epsilon > 0$ and $j = 1, 2, \ldots,$

M. Giaquinta and G. Modica, *Mathematical Analysis, Foundations and Advanced Techniques for Functions of Several Variables*, DOI 10.1007/978-0-8176-8310-8_6,

let A_j be open sets with $A_j \supset E_j$ and $\mu(A_j) \le \mu(E_j) + \epsilon\, 2^{-j}$, that we rewrite as $\mu(A_j \setminus E_j) < \epsilon\, 2^{-j}$ since E_j and A_j are measurable with finite measure. Since

$$\Big(\bigcup_j A_j\Big) \setminus \Big(\bigcup_j E_j\Big) \subset \bigcup_j (A_j \setminus E_j), \qquad \Big(\bigcap_j A_j\Big) \setminus \Big(\bigcap_j E_j\Big) \subset \bigcup_j (A_j \setminus E_j),$$

we infer

$$\mu(A \setminus \cup_j E_j) \le \epsilon, \qquad \mu(B \setminus \cap_j E_j) \le \epsilon, \tag{6.3}$$

where $A := \cup_j A_j$ and $B := \cap_j A_j$. Since A is open and $A \supset \cup_j E_j$, the first of (6.3) yields $\cup_j E_j \in \mathcal{A}$. On the other hand, $C_N := \cap_{j=1}^N A_j$ is open, contains $\cap_j E_j$ and by the second of (6.3), $\mu(C_N \setminus \cap_j E_j) \le 2\,\epsilon$ for sufficiently large N. Therefore $\cap_j E_j \in \mathcal{A}$.

Moreover, since every closed set is the intersection of a denumerable family of open sets, $\mathcal{A}$ also contains all closed sets. In particular, the family

$$\widetilde{\mathcal{A}} := \Big\{A \in \mathcal{A},\ A^c \in \mathcal{A}\Big\}$$

is a σ-algebra that contains the family of open sets. Consequently, $\mathcal{A} \supset \widetilde{\mathcal{A}} \supset \mathcal{B}(X)$ and (6.1) holds for all Borel sets of X.

Since (6.2) for E is (6.1) for E^c, the proof is concluded. □

The following theorem slightly extends Theorem 6.1.

6.2 Proposition. *Let μ be a Borel-regular measure on a metric space X.*

(i) *If E is a subset of X with $E \subset \cup_j V_j$ where $\{V_j\}$ is an increasing sequence of* open *sets with $\mu(V_j) < +\infty$, then*

$$\mu(E) = \inf\Big\{\mu(A) \,\Big|\, A \supset E,\ A \text{ open}\Big\}. \tag{6.4}$$

(ii) *If E is μ-measurable and μ–σ-finite, then*

$$\mu(E) = \sup\Big\{\mu(F) \,\Big|\, F \subset E,\ F \text{ closed}\Big\}. \tag{6.5}$$

When μ satisfies the conclusion of (i) of Proposition 6.2, we say that μ is *outer-regular*. Additionally, if μ satisfies the conclusion (ii) of Proposition 6.2, one says that μ is *inner-regular*. Before proving the proposition, we introduce a notation: Given a measure μ in X and $A \subset X$, the *restriction* of μ to A is the measure

$$\mu \llcorner A(E) := \mu(A \cap E) \qquad \forall E \subset X. \tag{6.6}$$

It is easily seen that the μ-measurable sets are $\mu \llcorner A$-measurable (A need not be μ-measurable), $\mu \llcorner A$ is a Borel measure if μ is Borel-regular and either A is μ-measurable with $\mu(A) < \infty$ or A is a Borel set.

Proof of Proposition 6.2. Since μ is Borel-regular, without loss of generality we may assume that E is a Borel set.

(i) We may assume $\mu(E) < +\infty$ otherwise (6.4) is trivial. The measures $\mu \llcorner V_j$ are Borel and $\mu \llcorner V_j(X) = \mu(V_j) < +\infty$. Theorem 6.1 then yields that for any $\epsilon > 0$ there are open sets A_j with $A_j \supset E$ and $\mu \llcorner V_j(A_j \setminus E) < \epsilon\, 2^{-j}$. The set $A := \cup_j (A_j \cap V_j)$ is open, $A \supset E$ and, by the subadditivity of μ,

$$\mu(A \setminus E) = \mu(\cup_j ((A_j \cap V_j) \setminus E)) \le \sum_{j=1}^{\infty} \mu((A_j \cap V_j) \setminus E) \le \sum_{j=1}^{\infty} \mu \llcorner V_j (A_j \setminus E) \le \epsilon.$$

(ii) The claim easily follows applying (ii) of Theorem 6.1 to the measure $\mu \llcorner E$ if $\mu(E) < +\infty$. If $\mu(E) = +\infty$ and $E = \cup_j E_j$ with E_j measurable and $\mu(E_j) < +\infty$, then for every $\epsilon > 0$ and every j there exists a closed set F_j with $F_j \subset E_j$ and $\mu(E_j \setminus F_j) < \epsilon\, 2^{-j}$. The set $F := \cup_j F_j$ is contained in E and

$$\mu(E \setminus F) \le \mu(\cup_j (E_j \setminus F_j)) \le \sum_{j=i}^{\infty} \mu(E_j \setminus F_j) \le \epsilon,$$

hence for N sufficiently large, $G_N := \cup_{i=1}^{N} F_i$ is closed and $\mu(E \setminus G_N) < 2\epsilon$. □

a. Lusin theorem

The approximability in measure of μ-measurable sets by open and closed sets translates into the following approximability property for μ-measurable functions.

6.3 Theorem (Lusin). *Let μ be a Borel-regular and finite measure in a metric space X and $f : X \to \mathbb{R}$ a μ-measurable function. For any μ-measurable set $A \subset X$ and any $\epsilon > 0$ there is a closed set $F \subset A$ such that $f_{|F}$ is continuous and $\mu(A \setminus F) < \epsilon$.*

Proof. For any integer k, we divide the real line in a denumerable union of intervals $I_{k,j}$ of length $1/k$. The inverse images $A_{k,j} := f^{-1}(I_{k,j}) \cap A \in \mathcal{E}$ define then a measurable partition of A, and, according to Proposition 6.2, for any $\epsilon > 0$ and any j there is a closed set $E_{k,j}$ with

$$E_{k,j} \subset A_{k,j}, \qquad \mu(A_{k,j} \setminus E_{k,j}) < \epsilon 2^{-k} 2^{-j},$$

so that $\mu(A \setminus \cup_{j=1}^{\infty} E_{k,j}) \le \epsilon\, 2^{-k}$. Consequently, there is an integer N such that $F_k := \cup_{i=1}^{N} E_{k,j}$ is closed and

$$\mu(A \setminus F_k) < \epsilon\, 2^{-k}.$$

Notice that for each k the closed sets $E_{k,j}$, the union of which is F_k, are pairwise disjoint. We now choose a point $y_{k,j} \in I_{k,j}$ for each j and set

$$g_k(x) := y_{k,j} \qquad \text{if } x \in E_{k,j},\ j = 1, \dots, N,$$

defining in this way a function $g_k : F_k \to \mathbb{R}$ that is constant on disjoint closed sets hence continuous. Furthermore, we have $|g_k(x) - f(x)| \le 1/k$ on F_k. If we now define

$$F := \bigcap_k F_k$$

we have the following: F is closed, $\mu(A \setminus F) \le \sum_{k=1}^{\infty} \mu(A \setminus F_k) \le \epsilon \sum_{k=1}^{\infty} 2^{-k} = \epsilon$ and $|g_k(x) - f(x)| < 1/k$ in F. Consequently, as $k \to \infty$, $g_k \to f$ uniformly on F and $f_{|F}$ is continuous. □

Notice that the maps $\{g_k\}$ constructed in the proof of Theorem 6.3 are such that $\inf_X f \le g_k(x) \le \sup_X f\ \forall x \in X$.

6.4 ¶ Dirac measure. The set function in $\mathbb{R}^n$

$$\delta_{x_0}(A) := \begin{cases} 1 & \text{if } x_0 \in A, \\ 0 & \text{if } x_0 \notin A, \end{cases} \qquad A \subset \mathbb{R}^n,$$

defines a finite Borel-regular measure in $\mathbb{R}^n$, called the *Dirac measure* concentrated at x_0. Interpret Lusin theorem for the Dirac measure at zero.

Recalling Tietze's theorem, see [GM3], we infer at once the following.

6.5 Corollary. *Let X be a metric space, μ a Borel-regular measure in X, $f : X \to \mathbb{R}$ a μ-measurable function and $A \subset X$ a μ-measurable set with $\mu(A) < +\infty$. For any $\epsilon > 0$ there is a continuous function $g : X \to \mathbb{R}$ such that*

$$||g||_{\infty,X} \le ||f||_{\infty,X} \qquad \text{and} \qquad \mu\Big(\Big\{x \in A \,\Big|\, f(x) \neq g(x)\Big\}\Big) < \epsilon.$$

Here $||f||_{\infty,X}$ is the *essential-sup-norm* of f, i.e.,

$$\operatorname{ess\,sup}_X f = ||f||_{\infty,X} := \inf\Big\{t \,\Big|\, \mu(\{x \mid\ |f(x)| > t\}) = 0\Big\}.$$

6.1.2 Radon measures in $\mathbb{R}^n$

6.6 Definition. *A* Radon measure *in a metric space X is a Borel-regular measure that is* inner-regular*,*

$$\mu(E) = \sup\Big\{\mu(K) \,\Big|\, K \subset E,\ K \text{ compact}\Big\} \qquad \forall E \in \mathcal{B}(X), \tag{6.7}$$

and locally finite.

Since $\mathbb{R}^n$ is a denumerable union of open sets with compact closure, trivially, any Borel-regular measure in $\mathbb{R}^n$ that is finite on compacts sets is a Radon measure, and $\mathbb{R}^n$ is a denumerable union of open sets with finite μ-measure. Therefore, (i) of Proposition 6.2 applies and we infer the following.

6.7 Proposition. *If μ is a Borel-regular measure finite on compact sets in $\mathbb{R}^n$, then μ is locally finite and*

$$\begin{aligned} \mu(E) &= \sup\Big\{\mu(K) \,\Big|\, K \subset E,\ K \text{ compact}\Big\} \qquad \forall E \in \mathcal{B}(X), \\ \mu(E) &= \inf\Big\{\mu(A) \,\Big|\, A \supset E,\ A \text{ open}\Big\} \qquad \forall E \subset \mathbb{R}^n. \end{aligned} \tag{6.8}$$

6.8 ¶. Prove that a measure μ in $\mathbb{R}^n$ is a Radon measure if and only if μ is Borel-regular and σ-finite.

a. Support

We say that a measure $(\mathcal{E}, \mu)$ is concentrated on E if $E \in \mathcal{E}$ and $\mu(E^c) = 0$. We notice that not necessarily is there a *minimal* set in which μ is concentrated: Think of Lebesgue's measure in $\mathbb{R}$.

The *support* of a Borel measure in $\mathbb{R}^n$ is defined as the set

$$F := \left\{ x \in \mathbb{R}^n \,\middle|\, \mu(B(x,r)) > 0 \ \forall r > 0 \right\}$$

and denoted by $\operatorname{spt}\mu$. Trivially, $x \notin \operatorname{spt}\mu$ if and only if there is $r_x > 0$ such that $\mu(B(x, r_x)) = 0$; consequently, $(\operatorname{spt}\mu)^c$ is open with $\mu((\operatorname{spt}\mu)^c) = 0$ and $\operatorname{spt}\mu$ is the smallest *closed* set F for which $\mu(F^c) = 0$. Notice that $\mu((\operatorname{spt}\mu)^c) = 0$ if μ is a Radon measure. In fact, any compact set K contained in the open set $(\operatorname{spt}\mu)^c$ can be covered by finitely many balls of zero measure, and μ is inner-regular.

6.9 ¶. Show that $\operatorname{spt}\mathcal{L}^n = \mathbb{R}^n$.

6.10 ¶ Dirac's measure. Show that $\operatorname{spt}\delta_{x_0} = \{x_0\}$.

b. Lusin theorem for Radon measures

The following two theorems are variants of Lusin's theorem for Radon measures.

6.11 Corollary (Lusin). *Let μ be a Borel-regular measure in $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ a μ-measurable function and $A \subset \mathbb{R}^n$ a μ-measurable set with $\mu(A) < \infty$. For any $\epsilon > 0$ there is a compact set $F \subset A$ such that $f_{|K}$ is continuous and $\mu(A \setminus K) < \epsilon$.*

Proof. Apply Lusin's Theorem 6.3 to find a closed set F such that $f_{|}F$ is continuous and $\mu(A \setminus F) < \epsilon$, and observe that in $\mathbb{R}^n$ every closed set is a denumerable union of compact sets. Since μ is finite, we find then a compact set $K \subset F$ with $\mu(F \setminus K) < \epsilon$, concluding that $f_{|K}$ is continuous and $\mu(A \setminus K) < 2\epsilon$. □

6.12 Corollary. *Let μ be a Radon measure in $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a μ-measurable function that vanishes outside a set of finite measure. For any $\epsilon > 0$ there is a continuous function $g : \mathbb{R}^n \to \mathbb{R}$ with* compact support *such that*

$$||g||_{\infty,\mathbb{R}^n} \le ||f||_{\infty,\mathbb{R}^n} \qquad \text{and} \qquad \mu\left\{ x \,\middle|\, f(x) \neq g(x) \right\} < \epsilon.$$

Proof. We know that there is an open set A with finite measure that contains $\operatorname{spt} f$. Corollary 6.11 yields a compact set $K \subset A$ such that $f_{|K}$ is continuous and $\mu(A \setminus K) < \epsilon$. In order to conclude the proof, it suffices to extend the continuous function

$$\overline{f}(x) = \begin{cases} f(x) & \text{if } x \in K, \\ 0 & \text{if } x \in A^c \end{cases}$$

defined on $K \cup A^c$ to a continuous function $g : \mathbb{R}^n \to \mathbb{R}$ via Tietze's theorem. □

6.13 ¶. Let μ be a Radon measure in $\mathbb{R}^n$. Prove that $C_c(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n, \mu)$, $1 \le p < \infty$. [*Hint.* Approximate $f \in \mathcal{L}^p(\mathbb{R}^n, \mu)$ with a bounded function with compact support.]

c. Riesz's theorem

Measures are deeply related to linear functionals on linear spaces.

A family L of functions $f : X \to \mathbb{R}$ on a set X is called a *lattice* of functions.

(i) If $f, g \in L$ and $c \geq 0$, then $f + g$, cf, $\inf(f, g)$ and $\inf(f, c)$ are functions in L.
(ii) If $f, g \in L$ and $f \leq g$, then $g - f \in L$.

The family of $\mathcal{E}$-measurable functions ($\mathcal{E}$ being a σ-algebra on X) and the space of continuous functions in a topological space X are examples of lattices of functions. Moreover, if L is a lattice of functions on X, so is the subset $L^+ := \{f \in L \,|\, f \geq 0\}$.

Let $(\mathcal{E}, \mu)$ be a measure on X and let $L \subset \mathcal{L}^1(X, \mu)$ be a family of summable functions on X. As we have seen, $\varphi \to \int_X \varphi \, d\mu$, $\varphi \in L$, defines a linear operator that is continuous for the μ-a.e. increasing convergence. Indeed such a property characterizes the integral completely. In fact, one can prove the following.

6.14 Theorem (Riesz). *Let L be a lattice of functions on X and let $\lambda : X \to \mathbb{R}$ be a linear functional such that the following hold:*

(i) *$\forall f, g \in L$ and $c \geq 0$ we have $\lambda(f + g) = \lambda(f) + \lambda(g)$ and $\lambda(cf) = c\,\lambda(f)$.*
(ii) *If $f, g \in L$ and $f \leq g$, then $\lambda(f) \leq \lambda(g)$.*
(iii) *If $\{f_k\} \subset L$, $f \in L$ and $\{f_k\}$ converges increasingly to f, then $\lambda(f_k) \to \lambda(f)$.*

Then there is a measure $(\mathcal{E}, \mu)$ on X such that every function in L is $\mathcal{E}$-measurable and $\lambda(f) = \int_X f \, d\mu$ for all $f \in L$.

We shall not prove this theorem[1]. We confine ourselves to discussing the case of linear, monotone and continuous functionals on continuous functions in $\mathbb{R}^n$. First, let us recall some notation and notice a few facts.

We denote by $C_c(\mathbb{R}^n)$ the class of continuous functions with compact support in $\mathbb{R}^n$ and with $C_0(\mathbb{R}^n)$ its completion with respect to the uniform norm $||\ ||_{\infty,\mathbb{R}^n}$. Of course, $C_0(\mathbb{R}^n)$ is a Banach space with the uniform norm. Furthermore, since the uniform limit of continuous functions produces a continuous function, it is not difficult to show that f belongs to $C_0(\mathbb{R}^n)$ if and only if f is continuous and for any $\epsilon > 0$ there exists a compact set K such that $||f||_{\infty,K^c} < \epsilon$.

Recall that a linear map $L : C_0(\mathbb{R}^n) \to \mathbb{R}$ is *continuous on* $C_0(\mathbb{R}^n)$ if and only if there is a constant $K > 0$ such that $|L(f)| \leq K\,||f||_{\infty,\mathbb{R}^n}$. The *norm* of L is then defined as

$$||L|| := \inf\Big\{K \;\Big|\; |L(f)| \leq K||f||_{\infty,\mathbb{R}^n} \ \forall f \in C_0(\mathbb{R}^n)\Big\}$$

[1] The interested reader may refer to, for example, H. Federer, *Geometric Measure Theory*, Springer-Verlag, 1969, Section 2.5.2.

or, equivalently, as

$$||L|| = \sup\Big\{ L(f) \,\Big|\, f \in C_0(\mathbb{R}^n), ||f||_{\infty,\mathbb{R}^n} \le 1 \Big\};$$

therefore, $L : C_0(\mathbb{R}^n) \to \mathbb{R}$ is continuous on $C_0(\mathbb{R}^n)$ if and only if $||L|| < \infty$.

Let μ be a *finite* Radon measure. The functional

$$L(f) := \int_{\mathbb{R}^n} f(x)\, d\mu$$

is linear on $C_0(\mathbb{R}^n)$, positive, meaning that $L(f) \ge 0$ if $f \ge 0$, and continuous on $C_0(\mathbb{R}^n)$, since

$$|L(f)| \le \int_{\mathbb{R}^n} |f|\, d\mu \le ||f||_{\infty,\mathbb{R}^n} \mu(\mathbb{R}^n).$$

In particular, $||L|| \le \mu(\mathbb{R}^n)$.

6.15 Theorem (Riesz). *Let $L : C_0(\mathbb{R}^n) \to \mathbb{R}$ be a linear, positive and continuous functional on $C_0(\mathbb{R}^n)$. There exists a unique Borel-regular and finite measure μ such that*

$$L(f) = \int_{\mathbb{R}^n} f\, d\mu \qquad \forall f \in C_0(\mathbb{R}^n).$$

Furthermore, $\mu(\mathbb{R}^n) = ||L||$.

Proof. Denote by $\mathcal{A}$ the family of open sets in $\mathbb{R}^n$. For $A \in \mathcal{A}$, set

$$\zeta(A) := \sup\Big\{ L(f) \,\Big|\, f \in C_0(\mathbb{R}^n),\ 0 \le f(x) \le \chi_A(x) \Big\}$$

and let μ be the outer measure obtained from $(\mathcal{A}, \zeta)$ by Method I of construction of measures,

$$\mu(E) := \inf\Big\{ \sum_{i=1}^{\infty} \zeta(A_i) \,\Big|\, \cup_i A_i \supset E,\ A_i \text{ open} \Big\} = \inf\Big\{ \zeta(A) \,\Big|\, A \supset E,\ A \text{ open} \Big\}.$$

We now prove that $(\mathcal{B}(\mathbb{R}^n), \mu)$ is the measure μ in the claim.

(i) Trivially, $\mu = \zeta$ on $\mathcal{A}$. Therefore,

$$\mu(\mathbb{R}^n) = \zeta(\mathbb{R}^n) = \sup\Big\{ L(f) \,\Big|\, f \in C_0(\mathbb{R}^n), 0 \le f(x) \le 1 \Big\} = ||L||, \tag{6.9}$$

in particular, μ *is finite.*

(ii) μ *is a Borel measure.* It is easily seen that ζ is finitely additive on $\mathcal{A}$. Consider two generic sets E and F in $\mathbb{R}^n$ with $d(E,F) > 0$. For any given $\epsilon > 0$, we find open sets A and B with $A \supset E$, $B \supset F$, $A \cap B = \emptyset$ and $\zeta(A \cup B) \le \mu(E \cup F) + \epsilon$. Hence, $\zeta(A) + \zeta(B) = \zeta(A \cup B)$ and

$$\mu(E \cup F) \ge \zeta(A \cup B) - \epsilon = \zeta(A) + \zeta(B) - \epsilon \ge \mu(E) + \mu(F) - \epsilon,$$

concluding, ϵ being arbitrary,

$$\mu(E \cup F) = \mu(E) + \mu(F)$$

if E and F have positive distance. The Carathéodory test, Theorem 5.36, implies then that μ is a Borel measure.

(iii) *μ is a Borel-regular measure.* In fact, if $E \subset \mathbb{R}^n$ is a generic set and $\{A_k\}$ a family of open sets with $A_k \supset E$ and $\mu(A_k) \le \mu(E) + 1/k$, we have $\mu(\cap_k A_k) = \mu(E)$.

(iv) *$L(f) \le \int_{\mathbb{R}^n} f\,d\mu$ for all nonnegative $f \in C_0(\mathbb{R}^n)$.* Of course, functions in $C_0(\mathbb{R}^n)$ are bounded. Set $b := ||f||_{\infty,\mathbb{R}^n}$. Given ϵ with $0 < \epsilon < 1$, choose an integer $k > 1/\epsilon$ and divide the interval $]-b/k, b+b/k]$ into $k+2$ closed on the right intervals

$$E_i = \Big\{ x \,\Big|\, y_i < f(x) \le y_{i+1} \Big\}$$

of length $1/k$, where $y_i = \frac{b}{k} i$, $i = -1, 0, \dots, k+1$. Clearly, $\{E_i\}$ is a partition of $\mathbb{R}^n$. Moreover, for ϵ with $0 < \epsilon < 1$, choose an open set V_i with $E_i \subset V_i \subset E_i \cup E_{i+1}$ and $\mu(V_i) \le \mu(E_i) + \frac{\epsilon}{k(k+1)}$. The family $\{V_i\}$ is a pointwise finite covering of $\mathbb{R}^n$, and there is a *partition of unity* associated to it, see [GM4], i.e., functions $h_i \in C_c(V_i)$ with $0 \le h_i(x) \le \chi_{V_i}(x)$ and $\sum_i h_i(x) = 1$. Consequently,

$$\begin{aligned} L(f) &= L\Big(\sum_{i=-1}^{k-1} h_i f \Big) = \sum_{i=-1}^{k-1} L(h_i f) \le \sum_{i=-1}^{k-1} y_{i+2} L(h_i) \le \sum_{i=-1}^{k-1} y_{i+2} \mu(V_i) \\ &\le \sum_{i=-1}^{k-1} \Big(y_i + 2\frac{b}{k} \Big) \Big(\mu(E_i) + \frac{\epsilon}{k(k+1)} \Big) \\ &\le \sum_{i=-1}^{k-1} y_i \mu(E_i) + 2b\epsilon \sum_{i=-1}^{k-1} \mu(E_i) + 2\epsilon \frac{b}{k^2(k+1)} + \frac{\epsilon}{k(k+1)} \sum_{i=-1}^{k-1} y_i \\ &\le \int_{\mathbb{R}^n} f\,d\mu + \epsilon \Big(2b\mu(\mathbb{R}^n) + 2b + 1 \Big), \end{aligned}$$

hence $L(f) \le \int_{\mathbb{R}}^n f\,d\mu$ if f is nonnegative.

(v) *$L(f) = \int_{\mathbb{R}^n} f\,d\mu$ for all $f \in C_0(\mathbb{R}^n)$.* It suffices to prove

$$L(f) \le \int_{\mathbb{R}^n} f\,d\mu \qquad \forall f \in C_0(\mathbb{R}^n), \tag{6.10}$$

since the required equality follows by applying (6.10) to f and $-f$.

Since by the Lusin theorem $C_0(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n, \mu)$, it follows from (iv) that L extends of a linear functional $\widetilde{L} : L^1(X,\mu) \to \mathbb{R}$ continuous on $L^1(X,\mu)$. Choosing $\{\varphi_n\} \subset C_0(\mathbb{R}^n)$ such that $\varphi_n \uparrow 1$, Beppo Levi's theorem yields $L(\varphi_n) \to \widetilde{L}(1)$ and, since trivially $L(\varphi_n) \le ||L|| \le \widetilde{L}(1)$, we get

$$\widetilde{L}(1) = ||L|| = \mu(\mathbb{R}^n).$$

Let $f \in C_0(\mathbb{R}^n)$; then for some constant c, $f + c$ is nonnegative, and, consequently, by (iv) and the above,

$$\begin{aligned} L(f) = \widetilde{L}(f) &= \widetilde{L}(f+c) - \widetilde{L}(c) = L(f+c) - c\widetilde{L}(1) \\ &\le \int (f+c)\,d\mu - \int c\,d\mu = \int f\,d\mu. \end{aligned}$$

(vi) Finally, let us prove uniqueness of the measure μ. Let μ_1 and μ_2 fulfill the thesis of the theorem and let $E \subset \mathbb{R}^n$. Since μ_1 and μ_2 are Borel-regular and finite, hence Radon measures, there exists a compact set C and an open set A such that $C \subset E \subset A$ and $\mu(A \setminus C) < \epsilon$. Furthermore, it is not restrictive to assume that A is bounded. The function

$$f(x) := \frac{d(x, A^c)}{d(x, A^c) + d(x, C)}$$

is, of course, continuous and with compact support and we have

$$\mu_1(E) - \epsilon \le \mu_1(K) \le \int f\, d\mu_1 = L(f) = \int f\, d\mu_2 \le \mu_2(A) \le \mu_2(E) + \epsilon.$$

Exchanging μ_1 and μ_2, we then conclude that $|\mu_1(E) - \mu_2(E)| < 2\,\epsilon$ for all $\epsilon > 0$. □

6.2 Differentiation of Measures

In this section we discuss the following. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a nonnegative and measurable function. The function

$$\lambda(A) := \int_A f(x)\, dx \tag{6.11}$$

defines trivially a new measure in $\mathbb{R}^n$ for which Lebesgue measurable sets are λ-measurable. It is easily seen that λ uniquely determines f for a.e. x; we would like to have an explicit formula for f in terms of λ.

We also characterize measures λ in $\mathbb{R}^n$ that can be written as in (6.11).

Finally, we discuss how to compare two measures λ and μ in $\mathbb{R}^n$.

6.2.1 Differentiation of Lebesgue integral

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a nonnegative and measurable function and let λ be the Borel-regular measure

$$\lambda(A) := \int_A f(x)\, dx. \tag{6.12}$$

In this subsection we characterize f in terms of λ.

We know that if f is continuous at x, the integral mean value theorem yields

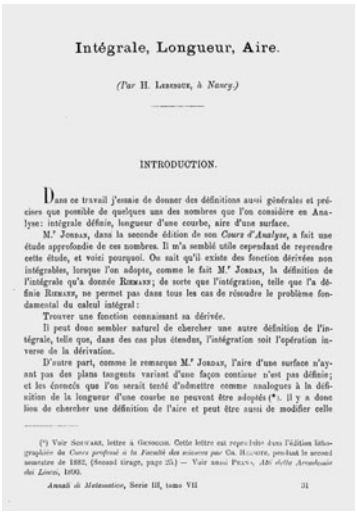
$$\lim_{r\to 0} \frac{\lambda(B(x,r))}{|B(x,r)|} = \lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\, dy = f(x).$$

If f is merely locally summable, the number

$$\frac{d\lambda}{d\mathcal{L}^n}(x) := \lim_{r\to 0} \frac{\lambda(B(x,r))}{|B(x,r)|},$$

if it exists, is called the *Radon–Nikodym derivative*, or simply the *derivative* at x of the measure λ in (6.12) with respect to the measure $\mathcal{L}^n$.

In this subsection we prove the following.



Intégrale, Longueur, Aire.

(Par H. Lebesgue, à Nancy.)

INTRODUCTION.

Dans ce travail j'essaie de donner des définitions aussi générales et précises que possible de quelques uns des nombres que l'on considère en Analyse: intégrale définie, longueur d'une courbe, aire d'une surface.

M.r Jordan, dans la seconde édition de son *Cours d'Analyse*, a fait une étude approfondie de ces nombres. Il m'a semblé utile cependant de reprendre cette étude, et voici pourquoi. On sait qu'il existe des fonction dérivées non intégrables, lorsque l'on adopte, comme le fait M.r Jordan, la définition de l'intégrale qu'a donnée Riemann; de sorte que l'intégration, telle que l'a définie Riemann, ne permet pas dans tous les cas de résoudre le problème fondamental du calcul intégral:

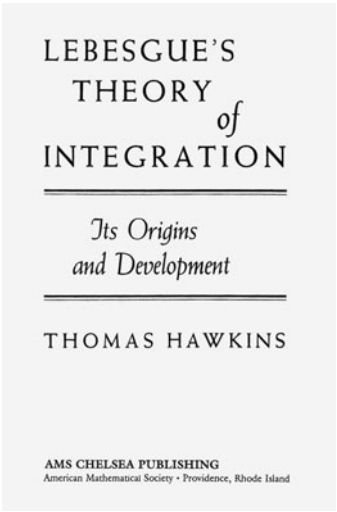
Trouver une fonction connaissant sa dérivée.

Il peut donc sembler naturel de chercher une autre définition de l'intégrale, telle que, dans des cas plus étendus, l'intégration soit l'opération inverse de la dérivation.

D'autre part, comme le remarque M.r Jordan, l'aire d'une surface n'ayant pas des plans tangents variant d'une façon continue n'est pas définie; et les énoncés que l'on serait tenté d'admettre comme analogues à la définition de la longueur d'une courbe ne peuvent être adoptés (*). Il y a donc lieu de chercher une définition de l'aire et peut être aussi de modifier celle

(*) Voir Serret, lettre à Genocchi. Cette lettre est reproduite dans l'édition lithographiée du *Cours professé à la Faculté des sciences par* Ch. Hermite, pendant le second semestre de 1882. (Second tirage, page 25.) — Voir aussi Peano, *Atti della Accademia dei Lincei*, 1890.

Annali di Matematica, Serie III, tomo VII 31



LEBESGUE'S
THEORY
of
INTEGRATION

Its Origins and Development

THOMAS HAWKINS

AMS CHELSEA PUBLISHING
American Mathematical Society • Providence, Rhode Island

Figure 6.1. The first page of the doctorate thesis of Henri Lebesgue (1875–1941) that appeared in 1902, and an essay on the hystorical developments of Lebesgue's integral.

6.16 Theorem (Vitali–Lebesgue). *Let f be a nonnegative and locally summable function and let λ be defined by* (6.11). *Then $\frac{d\lambda}{d\mathcal{L}^n}(x)$ exists, is finite and $\frac{d\lambda}{d\mathcal{L}^n}(x) = f(x)$ for $\mathcal{L}^n$-a.e. x.*

The proof is based on the notion of the *Hardy–Littlewood maximal function* and is quite robust with respect to the family of averaging sets. Later, see Section 6.2.4, we shall present the classical proof that uses the celebrated *Vitali covering theorem.*

a. Maximal function

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function in $L^1(\mathbb{R}^n)$. The *Hardy–Littlewood maximal function* of f is the function

$$Mf(x) := \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|\, dy, \qquad x \in \mathbb{R}^n.$$

Clearly, $Mf(x)$ is nonnegative, depends only on the a.e. equivalence class of f and is upper semicontinuous since the integral means are continuous functions of the radius. Moreover, if $f \in L^\infty(\mathbb{R}^n)$, then $Mf \in L^\infty(\mathbb{R}^n)$ and $||Mf||_\infty \le ||f||_\infty$.

6.17 ¶. Show that if $f \in \mathcal{L}^1(\mathbb{R}^n)$ and is nonzero on a set of positive measure, then $Mf(x) \ge c/|x|^n$ for $|x| \ge 1$ where c is a constant independent of x. In particular, $Mf(x) \notin L^1(\mathbb{R}^n)$.

Results on the differentiation of integrals for noncontinuous functions are granted on suitable selections of coverings. Here, the following lemma will suffice.

6.18 Lemma. *Let X be a metric space with distance d and let $\mathcal{B} = \{B(x, r(x))\}$ be a covering with open balls of a compact set $K \subset X$. We may select from $\mathcal{B}$ a finite and disjoint family of balls $\mathcal{B}' := \{B(x_i, r_i)\}_{i=1,N}$ such that $K \subset \cup_{i=1}^{N} B(x_i, 3r_i)$.*

Proof. Since K is compact, we may and do assume that $\mathcal{B}$ is a finite covering of K. We order the balls in $\mathcal{B}$ according to decreasing radii and iteratively select disjoint balls by choosing first a ball B_1 of maximal radius, then a ball B_2 of maximal radius among the balls that do not intersect B_1, a ball B_3 of maximal radius among the balls of the covering that do not intersect $B_1 \cup B_2$ and so on. In other words, at each step we choose a ball of maximum radius among the avalaible balls and then remove all balls that intersect it.

The family $\mathcal{B}' := \{B_1, \dots, B_N\}$ of the selected balls have the properties stated in the lemma. In fact, $\mathcal{B}'$ is a disjoint family of balls and, if $B(x, r) \in \mathcal{B}$ is a nonselected ball, then $B(x, r)$ intersects at least a ball $B_j \in \mathcal{B}'$ with $r_j \geq r$, hence $d(x, x_j) < r + r_j \leq 2r_j$. If follows that $B(x, r) \subset B(x_j, 3r_j)$ and, in conclusion, $K \subset \cup_{B \in \mathcal{B}} B \subset \cup_{j=1}^{N} B(x_j, 3r_j)$. □

The key information on maximal functions that follows from the argument in Lemma 6.18 is contained in the following estimate known as the *weak* $(1-1)$ *estimate* or *Hardy–Littlewood weak estimate.*

6.19 Proposition (Hardy–Littlewood). *Let $f \in L^1(\mathbb{R}^n)$. For all $t > 0$ we have*

$$\left| \left\{ x \,\middle|\, Mf(x) > t \right\} \right| \leq \frac{3^n}{t} \int_{\mathbb{R}^n} |f(x)|\, dx. \tag{6.13}$$

Proof. Let K be a compact set contained in $\{Mf(x) > t\}$. For all $x \in K$ there exists a ball $B(x, r(x))$ such that

$$\frac{1}{|B(x, r(x))|} \int_{B(x, r(x))} |f(y)|\, dy > t.$$

The family $\mathcal{B} = \{B(x, r(x))\}_{x \in K}$ is a covering of K with open balls for which Lemma 6.18 yields a finite subfamily $\{B(x_j, r_j)\}$ of disjoint balls such that $K \subset \cup_{j=1}^{N} B(x_j, 3r_j)$. Consequently, $|K| \leq 3^n \sum_{j=1}^{N} |B(x_j, r_j)|$ and

$$|K| \leq 3^n \sum_{j=1}^{N} |B(x_j, r_j)| \leq \frac{3^n}{t} \sum_{j=1}^{N} \int_{B(x_j, r_j)} |f(y)|\, dy \leq \frac{3^n}{t} \int_{\mathbb{R}^n} |f(y)|\, dy$$

since the balls are disjoint. Since $K \subset \{x \,|\, Mf(x) > t\}$ is arbitrary, the claim is proved. □

b. Differentiation of Lebesgue's integral

6.20 Theorem (Lebesgue). *Let $f \in L^p(E)$, $1 \leq p < +\infty$, where E is a measurable set in $\mathbb{R}^n$. For $\mathcal{L}^n$-a.e. $x \in E$ we have*

$$\frac{1}{|B(x, r)|} \int_{E \cap B(x, r)} |f(y) - f(x)|^p\, dy \to 0 \qquad \text{as } r \to 0^+.$$

In particular, for a.e. $x \in \mathbb{R}^n$

$$\frac{1}{|B(x, r)|} \int_{E \cap B(x, r)} f(y)\, dy \to f(x) \qquad \text{as } r \to 0^+.$$

Then Theorem 6.16 follows.

Proof. First we notice that it suffices to prove the theorem for functions $f \in L^p(\mathbb{R}^n)$. In fact, if $f \in L^p(E)$, then $f\chi_E \in L^p(\mathbb{R}^n)$, hence from

$$\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)\chi_E(y) - f(x)\chi_E(x)|^p\, dy \to 0$$

for a.e. $x \in \mathbb{R}^n$, we have, in particular,

$$\frac{1}{|B(x,r)|}\int_{E\cap B(x,r)} |f(y)\chi_E(y) - f(x)\chi_E(x)|^p\, dy \to 0,$$

that is,

$$\frac{1}{|B(x,r)|}\int_{E\cap B(x,r)} |f(y)\chi_E(y) - f(x)|^p\, dy \to 0$$

for a.e. $x \in E$.

Assume $f \in L^p(\mathbb{R}^n)$ and set

$$V(f,x) := \limsup_{r\to 0}\left(\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y) - f(x)|^p\, dy\right)^{1/p}.$$

We shall prove that $V(f,x) = 0$ for a.e. $x \in \mathbb{R}^n$. This is true if $f : \mathbb{R}^n \to \mathbb{R}$ is also continuous. In fact, the integral mean theorem yields

$$\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y) - f(x)|^p\, dy \le \Big(\operatorname{osc}_{B(x,r)} f\Big)^p,$$

hence

$$V(f,x) \le \limsup_{r\to 0^+} \operatorname{osc}_{B(x,r)} f = 0.$$

It f is not continuous, by the density theorem, Theorem 1.25, there is a sequence of continuous and p-summable functions $\varphi_k : \mathbb{R}^n \to \mathbb{R}$ such that $||f - \varphi_k||_p \to 0$. Now

$$|f(y) - f(x)| \le |f(y) - \varphi_k(y)| + |\varphi_k(y) - \varphi_k(x)| + |\varphi_k(x) - f(x)|,$$

hence

$$\begin{aligned} V(f,x) &\le |f(x) - \varphi_k(x)| + V(\varphi_k, x) \\ &\quad + \limsup_{r\to 0}\left(\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y) - \varphi_k(y)|^p\, dy\right)^{1/p} \\ &\le |f(x) - \varphi_k(x)| + \Big(M(|f - \varphi_k|^p)(x)\Big)^{1/p} \end{aligned} \tag{6.14}$$

Notice that $V(\varphi_k, x) = 0$ since φ_k is continuous. For a given $\epsilon > 0$, (6.14) yields

$$\Big\{x \,\Big|\, V(f,x) > \epsilon\Big\} \subset \Big\{x \,\Big|\, (M(|f - \varphi_k|^p)(x))^{1/p} > \frac{\epsilon}{2}\Big\} \bigcup \Big\{x \,\Big|\, |\varphi_k(x) - f(x)| > \frac{\epsilon}{2}\Big\}.$$

From (6.13) we infer

$$\Big|\Big\{x \,\Big|\, (M(|f - \varphi_k|^p)(x))^{1/p} > \frac{\epsilon}{2}\Big\}\Big| \le \frac{2^p 3^n}{\epsilon^p}\int_{\mathbb{R}^n} |f - \varphi_k|^p\, dy$$

and from Chebichev's inequality

$$\Big|\Big\{x \,\Big|\, |\varphi_k(x) - f(x)| > \frac{\epsilon}{2}\Big\}\Big| = \Big|\Big\{x \,\Big|\, |\varphi_k(x) - f(x)|^p > (\epsilon/2)^p\Big\}\Big| \le \frac{2^p}{\epsilon^p}\int_{\mathbb{R}^n} |f - \varphi_k|^p\, dy.$$

We then conclude

$$\left|\left\{x \,\middle|\, V(f,x) > \epsilon\right\}\right| \le \frac{2^p(3^n+1)}{\epsilon^p} \int |f-\varphi_k|^p \, dy \qquad \forall k.$$

Letting $k \to \infty$, we conclude $|\{x \,|\, V(f,x) > \epsilon\}| = 0$, and, since ϵ is arbitrary,

$$\left|\left\{x \,\middle|\, V(f,x) > 0\right\}\right| = \lim_{j\to\infty} \left|\left\{x \,\middle|\, V(f,x) > \frac{1}{j}\right\}\right| = 0.$$

□

6.21 Example. If $f \in L^1(]-1,1[)$, for a.e. $x \in]-1,1[$, we have

$$\lim_{r\to 0^+} \frac{1}{2r} \int_{x-r}^{x+r} f(y)\, dy = f(x).$$

c. Some variants of Lebesgue differentiation

The previous argument is sufficiently robust to allow us to infer variants of Lebesgue's theorem. Not only averages over balls converge almost everywhere: We can replace balls by cubes or even other families. For instance, assume that A is a bounded set of positive measure, say

$$A \subset B(0,100) \subset \mathbb{R}^n, \qquad |A| = c\,|B_1|.$$

For all $x \in \mathbb{R}^n$, $r > 0$ and $A_{x,r} := x + r\,A$ we clearly have $A_{x,r} \subset B(x, 100\,r)$ and $|A_{x,r}| = r^n\,|A| = c\,r^n\,B_1 = c\,|B(x,r)|$. The following analogous of Lebesgue's theorem holds.

6.22 Theorem. *Let $f \in L^p(E)$, $E \subset \mathbb{R}^n$. For a.e. $x \in E$ we have*

$$\frac{1}{|A_{x,r}|} \int_{E\cap A_{x,r}} |f(y) - f(x)|^p \, dy \to 0 \qquad \textit{per } r \to 0^+.$$

Proof. As previously, we may assume $f \in L^p(\mathbb{R}^n)$ and for $x \in \mathbb{R}^n$ set

$$V_A(f,x) := \limsup_{r\to 0} \left(\frac{1}{|A_{x,r}|} \int_{A_{x,r}} |f(y) - f(x)|^p \, dy \right)^{1/p}.$$

We get

$$V_A(f,x) \le \left(M_A(|f-\varphi_k|^p)(x)\right)^{1/p} + |f(x) - \varphi_k(x)|,$$

where for $g \in L^1(\mathbb{R}^n)$ we have set

$$M_A(g,x) := \sup_{r>0} \frac{1}{|A_{x,r}|} \int_{A_{x,r}} |g(y)|\, dy.$$

The proof then follows the same path of the proof of Theorem 6.20, provided we prove a *Hardy–Littlewood inequality* for the modified maximal function M_A, i.e.,

$$\left|\left\{x \,\middle|\, M_A f(x) > t\right\}\right| \le \frac{C}{t} \int_{\mathbb{R}^n} |f(y)|\, dy, \qquad \forall t > 0 \tag{6.15}$$

for some constant C independent of f and t. For that, we notice that

$$\begin{aligned} \frac{1}{|A_{x,r}|} \int_{A_{x,r}} |f(y)|\, dy &\le \frac{1}{c}\, \frac{1}{|B(x,r)|} \int_{B(x,100\,r)} |f(y)|\, dy \\ &\le \frac{1}{C}\, \frac{1}{|B(x,100\,r)|} \int_{B(x,100\,r)} |f(y)|\, dy \end{aligned}$$

with $C = c(100)^{-n}$. It follows that $M_A f(x) \le \frac{1}{C} M f(x)$, hence $\{x \,|\, M_A f(x) > t\} \subset \{x \,|\, M f(x) > C\,t\}$ and therefore, by (6.15),

$$\left|\left\{x \,\middle|\, M_A f(x) > t\right\}\right| \le \left|\left\{x \,\middle|\, M f(x) > C\,t\right\}\right| \le \frac{3^n}{C\,t} \int_{\mathbb{R}^n} |f(y)|\,dy \qquad \forall t > 0.$$

□

6.23 ¶. Let $f \in L^1(\mathbb{R})$. Show that for a.e. $x \in \mathbb{R}$

$$\lim_{r\to 0^+} \frac{1}{r} \int_0^r f(x+t)\,dt = \lim_{r\to 0^+} \frac{1}{r} \int_{-r}^0 f(x+t)\,dt = f(x)$$

and

$$\lim_{r\to 0^+} \frac{1}{r} \int_r^{2r} f(x+t)\,dt = f(x).$$

6.24 ¶. For $f \in L^1(\mathbb{R}^n)$ set $\mathcal{M} f(x) = \sup_{Q \ni x} \frac{1}{|Q|} \int_Q |f(y)|\,dy$, where the supremum is taken among all cubes containing x with sides parallel to the axes. Show that there exists a constant $c = c(n)$ such that for all $f \in L^1(\mathbb{R}^n)$ and $t > 0$

$$\left|\left\{x \,\middle|\, \mathcal{M} f(x) > t\right\}\right| \le \frac{c}{t} \int_{\mathbb{R}^n} |f(y)|\,dy.$$

Deduce that, if $f \in L^p_{\mathrm{loc}}(E)$, $E \subset \mathbb{R}^n$, $p \ge 1$, then for a.e. $x \in E$ we have

$$\lim_{\substack{|Q|\to 0 \\ Q \ni x}} \frac{1}{|Q|} \int_{E\cap Q} |f(y) - f(x)|^p\,dy = 0.$$

6.25 ¶. Let $E \subset \mathbb{R}^n$ be a measurable set and $x \in \mathbb{R}^n$. The *upper density*, *lower density* and *density* of E at x is respectively defined as

$$\theta^*(E,x) := \limsup_{r\to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|}, \qquad \theta_*(E,x) := \liminf_{r\to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|}$$

and

$$\theta(E,x) := \lim_{r\to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|}.$$

Show that

(i) $\theta^*(E,x)$, $\theta_*(E,x)$ are measurable functions, actually Borel functions,
(ii) $\theta(E,x) = 1$ for a.e. $x \in E$ and $\theta(E,x) = 0$ for a.e. $x \in E^c$.

d. Lebesgue's points

The following theorem allows us to identify a representative in each equivalence class of functions in L^1.

6.26 Definition. *Let $f : E \subset \mathbb{R}^n \to \mathbb{R}$ be a summable function in E. The Lebesgue points of f are the points of the set*

$$\mathcal{L}_f := \Big\{ x \in E \in \mathbb{R}^n \,\Big|\, \exists \lambda \in \mathbb{R} \text{ such that } \frac{1}{|B(x,r)|} \int_{E\cap B(x,r)} |f(y) - \lambda|\,dy \to 0 \Big\}.$$

The set $\mathcal{L}_f$ and the function $\lambda(x) : \mathcal{L}_f \to \mathbb{R}$ depend on the a.e. equivalence class of f and not on f directly. The function $\lambda(x) : \mathcal{L}_f \to \mathbb{R}$ is called the *Lebesgue representative* of f.

6.27 Theorem. *If $\lambda : \mathcal{L}_f \to \mathbb{R}$ is the Lebesgue representative of f, then $|\mathbb{R}^n \setminus \mathcal{L}_f| = 0$, i.e., $f(x) = \lambda(x)$ for a.e. x.*

6.28 ¶. Prove Theorem 6.27.

6.2.2 Radon–Nikodym theorem

In this subsection we deal with a comparison argument between two measures. As a byproduct, we characterize the measures λ for which a differentiation formula such as (6.11) holds.

6.29 Definition. *Two measures $(\mu, \mathcal{E})$ and $(\nu, \mathcal{E})$ on X are said to be* mutually singular *if there is $E \in \mathcal{E}$ such that $\mu(E) = 0$ and $\nu(E^c) = 0$. We say that $(\nu, \mathcal{E})$ is* absolutely continuous *with respect to $(\mu, \mathcal{E})$, and we write $\nu \ll \mu$ if $\nu(E) = 0$ whenever $\mu(E) = 0$.*

6.30 Example. The measure $\mu(A) = \int_A f \, d\mathcal{L}^n$, $f \in L^1(\mathbb{R}^n)$, $f \geq 0$, is absolutely continuous with respect to Lebesgue's measure $\mathcal{L}^n$. The Dirac measure at 0,

$$\delta_0(A) := \begin{cases} 1 & \text{if } 0 \in A, \\ 0 & \text{if } 0 \notin A \end{cases}$$

is singular with respect to $\mathcal{L}^n$. In fact, for $Z := \mathbb{R}^n \setminus \{0\}$, we have $\delta_0(Z) = 0$ and $\mathcal{L}^n(Z^c) = 0$.

6.31 Proposition. *Let $\lambda_1, \lambda_2, \mu : X \to \overline{\mathbb{R}}$ be three measures in X. We have the following:*

(i) *If f is nonnegative, μ-measurable and $\nu(E) := \int_E f \, d\mu$, then $\nu << \mu$.*
(ii) *If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 + \lambda_2 \perp \mu$.*
(iii) *If $\lambda_1 \ll \mu$ and $\lambda_2 \ll \mu$, then $\lambda_1 + \lambda_2 \ll \mu$.*
(iv) *If $\lambda_1 \ll \lambda_2$ and $\lambda_2 \perp \mu$, then $\lambda_1 \perp \mu$.*
(v) *If $\lambda_1 \ll \mu$ and $\lambda_1 \perp \mu$, then $\lambda_1 = 0$.*

6.32 ¶. Prove Proposition 6.31.

6.33 Proposition. *Let $(\lambda, \mathcal{E})$ and $(\mu, \mathcal{E})$ be two measures and suppose that $(\lambda, \mathcal{E})$ is finite. Then $(\lambda, \mathcal{E})$ is absolutely continuous with respect to $(\mu, \mathcal{E})$ if and only if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $E \in \mathcal{E}$ with $\mu(E) < \delta$ we have $\lambda(E) < \epsilon$.*


Figure 6.2. Otto Nikodým (1887–1974) and Thomas Jan Stieltjes (1856–1894).

Proof. Suppose $\lambda \ll \mu$. Assume by contradiction that for $\epsilon > 0$ and for a sequence $\{E_k\} \subset \mathcal{E}$ we have $\mu(E_k) < 2^{-k}$ and $\lambda(E_k) \geq \epsilon$. Then for $E := \cap_{k=1}^{\infty} \cup_{j=k}^{\infty} E_j \in \mathcal{E}$ we have

$$\mu(E) = \lim_{k\to\infty} \mu\Big(\bigcup_{k=j}^{\infty} E_j\Big) \leq \lim_{k\to\infty} \sum_{j=k}^{\infty} \mu(E_j) = 0$$

and, since λ is finite,

$$\lambda(E) = \lim_{k\to\infty} \lambda\Big(\bigcup_{k=j}^{\infty} E_j\Big) \geq \liminf_{j\to\infty} \lambda(B_j) \geq \epsilon,$$

a contradiction. The converse is trivial. □

6.34 ¶. In Proposition 6.33 it is essential that $(\lambda, \mathcal{E})$ is finite. Show that Proposition 6.33 does not hold for $\lambda(A) := \int_A f(t)\, d\mathcal{L}^1(t)$ and $\mu(A) := \mathcal{L}^1(A)$ for a suitable choice of $f \in L^1_{loc}(\mathbb{R})$.

6.35 Theorem. *Let $(\lambda, \mathcal{E})$ and $(\mu, \mathcal{E})$ be two σ-finite measures in X. Then there is a unique decomposition of λ as a sum of two measures $\lambda = \lambda^a + \lambda^s$ with $\lambda^a \ll \mu$ and $\lambda^s \perp \mu$.*

More precisely, there exists an $\mathcal{E}$-measurable and nonnegative function θ with $\theta(x) < +\infty$ for μ-a.e. x such that setting $Z := \{x \,|\, \theta(x) = +\infty\}$, we have $\mu(Z) = 0$ and

$$\lambda^a(E) = \int_E \theta\, d\mu, \qquad \lambda^s(E) = \lambda(E \cap Z) \qquad \forall E \in \mathcal{E}. \tag{6.16}$$

Proof. The ideas in the following proof go back to von Neumann.

Step 1. As usual, it easy to see, using the standard exhaustion argument, that is suffices to prove the theorem under the extra assumption that μ and λ are finite. In fact, if μ and λ are only σ-finite, we may then assume that there exists a disjoint sequence $\{E_k\} \subset \mathcal{E}$ with $\mu(E_k), \lambda(E_k) < \infty$ such that $X = \cup_k E_k$. By applying the theorem to $\lambda \llcorner E_k$ and $\mu \llcorner E_k$ for each k, we find a nonnegative $\mathcal{E}$-measurable function $\theta_k : E_k \to \overline{\mathbb{R}}$ such that if $Z_k := \{x \in E_k \,|\, \theta_k(x) = +\infty\}$, then $\mu(Z_k) = 0$ and $\lambda \llcorner E_k = \lambda^a_k + \lambda^s_k$, where

$$\lambda_k^a(E) = \int_{E\cap E_k} \theta_k \, d\mu, \qquad \lambda_k^s(E) = \lambda(E \cap Z_k).$$

Summing in k one gets the Lebesgue decomposition formula for λ with respect to μ.

Step 2. From now on we shall assume that λ and μ are finite. The linear functional $L(\varphi) := \int \varphi \, d\lambda$ is continuous on $L^2(X, \mu + \lambda)$ since

$$|L(\varphi)| \le \int_X |\varphi| \, d\lambda \le \lambda(X)^{1/2} \Big(\int_X |\varphi|^2 \, d\lambda \Big)^{1/2} \le \lambda(X)^{1/2} \Big(\int_X |\varphi|^2 \, d(\lambda + \mu) \Big)^{1/2}.$$

Therefore, Riesz's theorem yields $g \in L^2(X, \mu + \lambda)$ such that

$$\int \varphi \, d\lambda = L(\varphi) = (\varphi | g)_{2,\mu+\lambda} = \int \varphi g \, d(\lambda + \mu)$$

i.e.,

$$\int \varphi(1 - g) \, d\lambda = \int \varphi g \, d\mu \tag{6.17}$$

for every $\varphi \in L^2(X, \mu + \lambda)$, in particular, for all bounded and $\mathcal{E}$-measurable φ since λ and μ are finite.

Equation (6.17) for φ equal to the characteristic function of the sets $\{x \,|\, g(x) < 0\}$, $\{x \,|\, g(x) > 1\}$ and $\{x \,|\, g(x) = 1\}$ yields respectively $0 \le g(x)$ and $g(x) \le 1$ for $(\lambda+\mu)$-a.e. x, and $g(x) < 1$ for μ-a.e. x.

Step 3. Let

$$Z := \{x \,|\, g(x) = 1\}, \qquad \lambda^a(E) := \lambda(A \cap Z^c) \quad \text{and} \quad \lambda^s(E) := \lambda(E \cap Z).$$

Trivially, $\lambda^s \perp \mu$ since $\mu(Z) = 0$ and $\lambda^s(Z^c) = 0$. Moreover, if $\mu(N) = 0$, again by (6.17),

$$\int_{N\cap Z^c} (1 - g) d\lambda = \int \chi_{N\cap Z^c}(1 - g) \, d\lambda = \int_{N\cap Z^c} g \, d\mu = 0.$$

Since $g < 1$ on Z^c, we then conclude that $\lambda^a(E) = \lambda(N \cap Z^c) = 0$.

Step 4. Let φ be $\mathcal{E}$-measurable and nonnegative. For any integer n, we get from (6.17)

$$\int \varphi \Big(\sum_{k=0}^{n} g^k \Big)(1 - g) \, d\lambda = \int \varphi g \Big(\sum_{k=0}^{n} g^k \Big) \, d\mu,$$

hence Beppo Levi's theorem yields, as $n \to \infty$,

$$\int_{Z^c} \varphi \, d\lambda = \int \varphi \frac{g}{1 - g} \, d\mu.$$

In conclusion, setting $\theta := \frac{g}{1-g}$, we have

$$Z = \Big\{ x \,\Big|\, \theta(x) = \infty \Big\} \qquad \text{and} \qquad \lambda^a(E) = \lambda(E \cap Z^c) = \int_E \theta(x) \, d\mu.$$

Step 5. Let us prove the uniqueness of the Lebesgue decomposition of λ with respect to μ.

Suppose $\lambda = \lambda^a + \lambda^s = \lambda^1 + \lambda^2$ with $\lambda^a, \lambda^1 \ll \mu$ and $\lambda^2 \perp \mu$, $\lambda^s \perp \mu$. Let B and B_1 be such that $\mu(B) = \mu(B_1) = 0$ and $\lambda^s(B^c) = \lambda^2(B_1^c) = 0$. For $A \in \mathcal{E}$ and $A \subset B \cup B_1$, since $\mu(A) = 0$ and $\lambda^a, \lambda^1 \ll \mu$ necessarily $\lambda^a(A) = \lambda^1(A) = 0$, hence $\lambda^s(A) = \lambda(A) = \lambda^2(A)$. If $A \subset B^c \cap B_1^c$, then necessarily $\lambda^s(A) = \lambda^2(A) = 0$. Consequently, for all $A \in \mathcal{E}$

$$\begin{aligned} \lambda^2(A) &= \lambda^2(A \cap (B \cup B_1)) + \lambda^2(A \cap (B^c \cap B_1^c)) \\ &= \lambda^s(A \cap (B \cup B_1)) + \lambda^s(A \cap (B^c \cap B_1^c)) = \lambda^s(A), \end{aligned}$$

i.e., $\lambda^s = \lambda^2$. It follows that $\lambda^a(A) = \lambda^1(A)$ for all A with $\lambda(A) < +\infty$, and since λ is σ-finite, we conclude that $\lambda^a = \lambda^1$. □

6.36 Corollary. *Let $(\lambda, \mathcal{E})$ and $(\mu, \mathcal{E})$ be two σ-finite measures in X. Then $\lambda \ll \mu$ if and only if there exists an $\mathcal{E}$-measurable and nonnegative function θ with $\theta(x) < +\infty$ μ-a.e. such that*

$$\lambda(E) = \int_E \theta \, d\mu, \qquad \forall E \in \mathcal{E}. \tag{6.18}$$

Proof. Consider $E \in \mathcal{E}$ such that $\mu(E) = 0$. If (6.18) holds, then $\lambda(E) = 0$ by the definition of integral. This proves that $\lambda \ll \mu$. Conversely, assume that $\lambda \ll \mu$. The uniqueness of the Lebesgue decomposition of λ yields $\lambda^a = \lambda$ and $\lambda^s = 0$. Therefore (6.18) follows from Theorem 6.35. □

If $(\lambda, \mathcal{E})$ and $(\mu, \mathcal{E})$ are two σ-finite measures, we write $\lambda = \theta \, \mu$ or $d\lambda = \theta \, d\mu$ instead of (6.18). Of course, (6.18) implies that

$$\int \varphi \, d\lambda = \int \varphi\theta \, d\mu$$

for all nonnegative and $\mathcal{E}$-measurable functions φ, by the usual procedure, using simple functions and Beppo Levi's theorem. A trivial consequence is the *chain rule* for the Radon–Nikodym derivatives: If $d\nu = \rho \, d\lambda$ and $d\lambda = \theta \, d\mu$, then $d\mu = \rho\theta \, d\mu$ since for $E \in \mathcal{E}$

$$\nu(E) = \int \chi_E \rho \, d\lambda = \int \chi_E \rho\theta \, d\mu.$$

6.2.3 Doubling measures in metric spaces

a. Differentiation of the integral

6.37 Definition. *Let X be a metric space with distance d and let μ be a measure in X. Denote by $B(x, r) := \{y \in X \mid d(x, y) < r\}$. We say that μ has the* doubling property *if there is a constant $C > 0$ such that $\mu(B(x, 2r)) \le C\mu(B(x, r))$ $\forall x \in X$, $\forall r > 0$.*

Trivially, μ is doubling if and only if there exists a constant C' such that $\mu(B(x, 5r)) \le C'\mu(B(x, r))$ $\forall x \in X$, $\forall r > 0$.

Going through the proof of the weak-type estimate for the Hardy–Littlewood function, Proposition 6.19, it clearly appears that the only properties of $\mathcal{L}^n$ that enter the proof are (i) $\mathcal{L}^n$ is a Radon measure and (ii) $\mathcal{L}^n$ has the doubling property. Therefore, assuming that μ is a Radon measure, the same proof yields the following.

6.38 Proposition. *Let μ be a measure on a metric space X satisfying the doubling property and $f \in L^1_{loc}(X)$. Define the maximal function of f with respect to μ at $x \in \operatorname{spt} \mu$ by*

$$M_\mu f(x) := \sup_{r>0} \frac{1}{\mu(B(x, r))} \int_{B(x,r)} |f(y)| \, d\mu(y). \tag{6.19}$$

Then, for any $t > 0$,

$$\mu\Big(\Big\{x \in \operatorname{spt}\mu \;\Big|\; M_\mu f(x) > t\Big\}\Big) \le \frac{C}{t}\int_X |f(y)|\, d\mu(y). \tag{6.20}$$

The proof follows the same line of the proof of Proposition 6.19 if μ is a Radon measure with the doubling property. In the general case, a slight improvement in the covering argument is needed.

Let us describe the needed covering argument. If B is a ball in X, we denote by $r(B)$ the radius of B and by $\widetilde{B}$ the ball with the same center as B and radius five times $r(B)$,

$$\widetilde{B} := B(x, 5\,r) \qquad \text{if} \qquad B = B(x, r).$$

If $\mathcal{B}$ is a family of balls of X, we denote by $\widetilde{\mathcal{B}}$ the family $\{B(x, 5r) \,|\, B(x, r) \in \mathcal{B}\}$.

6.39 Lemma. *Let X be a metric space and let $\mathcal{B}$ be a family of balls in X with bounded diameters,*

$$\sup\Big\{r(B) \;\Big|\; B \in \mathcal{B}\Big\} < +\infty$$

(it is irrelevant whether the balls are open or closed). Then there exists a subfamily $\mathcal{B}'$ of $\mathcal{B}$ of disjoint balls such that for any $B \in \mathcal{B}$ there is $B' \in \mathcal{B}'$ such that $B \cap B' \neq \emptyset$ and $B \subset \widetilde{B'}$. Consequently,

$$\bigcup_{B\in\mathcal{B}} B \subset \bigcup_{B\in\mathcal{B}'} \widetilde{B}.$$

Proof. When $\mathcal{B}$ is denumerable, we can order the balls of $\mathcal{B}$ in a sequence of decreasing radii and inductively select the subfamily as in Lemma 6.18. As, in general, X need not be separable, one proceeds similarly using the axiom of choice. Let

$$\mathcal{B}_h = \Big\{B \in \mathcal{B} \;\Big|\; R\,2^{-h-1} < r(B) \le R\,2^{-h}\Big\}$$

where $R := \sup\{r \;\big|\; B(x, r) \in \mathcal{B}\}$. By Zorn's lemma there is a maximal set of disjoint balls $\mathcal{B}'_0 \subset \mathcal{B}_0$. We define inductively $\mathcal{B}'_h$ as the maximal set of disjoint balls in

$$\Big\{B \in \mathcal{B}_h \;\Big|\; B \cap C = \emptyset \;\forall C \in \cup_{j=1}^{h-1}\mathcal{B}'_j\Big\}.$$

The set $\mathcal{B}' := \cup_{h=1}^{\infty}\mathcal{B}'_h$ has the requested properties. In fact, the balls in $\mathcal{B}'$ are disjoint as for each h the balls in $\mathcal{B}'_h$ are disjoint and also disjoint from the balls previously chosen in $\mathcal{B}'_j$ for $j = 0, \dots, h-1$. On the other hand, if $B \in \mathcal{B}_h \setminus \mathcal{B}'$, then B meets a ball $B' \in \mathcal{B}'$ of radius at least $R\,2^{-h-1}$, hence $B \subset \widetilde{B'}$. It follows that $\cup_{B\in\mathcal{B}} B \subset \cup_{B\in\mathcal{B}'} \widetilde{B}$. □

Proof of Proposition 6.38. We can assume that $\int_X |f(y)|\,d\mu(y) < \infty$, otherwise there is nothing to prove. For $x \in \operatorname{spt}\mu$ and $R > 0$, set

$$M_{\mu,R}f(x) := \sup_{0<r<R} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y)|\,d\mu(y)$$

and consider the set

$$M_R := \Big\{ x \in \operatorname{spt}\mu \,\Big|\, M_{\mu,R}f(x) > t \Big\}.$$

For any $x \in A_R$ there is a ball $B(x, r(x))$ with $r(x) < R$ such that $\mu(B(x, r(x))) < \frac{1}{t}\int_{B(x,r(x))} |f(y)|\,d\mu(y)$. Let $\mathcal{B}$ be the family of such balls. Lemma 6.39 yields a disjoint subfamily $\mathcal{B}' \subset \mathcal{B}$ such that $\widetilde{\mathcal{B}'}$ is a covering of M_R. Using also the doubling property of μ, we then get

$$\mu(M_R) \le \sum_{B\in\mathcal{B}'} \mu(\widetilde{B}) \le C \sum_{B\in\mathcal{B}'} \mu(B) \le \frac{C}{t} \sum_{B\in\mathcal{B}'} \int_B |f(y)|\,d\mu(y) \le \frac{C}{t}\int_X |f(y)|\,dy,$$

where C depends on the doubling constant of μ. Finally, since R is arbitrary, we find

$$\mu\Big(\Big\{ x \in \operatorname{spt}\mu \,\Big|\, M_\mu f(x) > t \Big\}\Big) = \lim_{R\to\infty} \mu(M_R) \le \frac{C}{t}\int_X |f(y)|\,dy.$$

□

Following the same path of the proof of Theorem 6.20, taking also into account Proposition 6.38, we get the following.

6.40 Theorem. *Let X be a metric space with distance d, let μ be a measure on X with the doubling property and let $f \in L^p(X,\mu)$, $1 \le p < +\infty$. For μ-a.e. $x \in \operatorname{spt}\mu$ we have*

$$\frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|^p\,d\mu(y) \to 0 \qquad \text{as } r \to 0^+.$$

b. Differentiation of measures

Let λ and μ be σ-finite measures on a metric space X and let $\lambda = \lambda^a + \lambda^s$ be the Lebesgue decomposition of λ with respect to μ. From Theorem 6.35 we infer that $\lambda(E) = \int \theta\,d\mu$ for some non-negative, μ-a.e. finite and $\mathcal{E}$-measurable function θ and $\lambda^s(E) = \lambda(E \cap J)$, where $J = \{x \,|\, \theta(x) + = \infty\}$.

However, the density function θ is obtained with a global argument; a pointwise description of θ is likely to be more useful in analytic and geometric applications. We deal precisely with this aspect of the Lebesgue decomposition in the following sections. Here we restrict ourselves to the case in which λ and μ are Radon measures and μ *has the doubling property.*

For $x \in \operatorname{spt}\mu$ we define

$$D^+_\mu\lambda(x) = \limsup_{\rho\to 0^+} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))}, \qquad D^-_\mu\lambda(x) = \liminf_{\rho\to 0^+} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))},$$

and if $D^+_\mu\lambda(x) = D^-_\mu\lambda(x)$ at $x \in \operatorname{spt}\mu$, the common value is called the *Radon–Nikodym derivative* at x of λ with respect to μ and denoted by $\frac{d\lambda}{d\mu}(x)$,

$$\frac{d\lambda}{d\mu}(x) = \lim_{\rho\to 0^+} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))}.$$

Since closed balls can be approximated by open balls and conversely, the upper and lower derivatives of λ with respect to μ do not change if we replace closed balls with open balls. Since

$$D_\mu^+\lambda(x) = \lim_{k\to\infty} \sup_{0<\rho<1/k} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))},$$
$$D_\mu^-\lambda(x) = \lim_{k\to\infty} \inf_{0<\rho<1/k} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))},$$

we have the following.

6.41 Proposition. *The functions $D_\mu^+\lambda$ and $D_\mu^-\lambda$ are Borel functions.*

6.42 Theorem. *Let λ and μ be two Radon measures on a metric space X and let $\lambda = \lambda^a + \lambda^s$ be the Lebesgue decomposition of λ with respect to μ. Assume that μ has the doubling property. Then for μ-a.e. x the Radon–Nikodym derivative of λ^a with respect to μ exists, is finite and*

$$\lambda^a(E) = \int_E \frac{d\lambda^a}{d\mu}(x)\, d\mu(x) \qquad \forall E \in \mathcal{B}(X).$$

If moreover, $X = \cup_j X_j$ where X_j are open with $\lambda(X_j) < +\infty$, then for μ-a.e. x, $\frac{d\lambda^s}{d\mu}(x)$ exists, is finite and $\frac{d\lambda^s}{\delta\mu}(x) = 0$. Consequently, $\frac{d\lambda}{d\mu}(x)$ exists and is finite for μ-a.e. $x \in X$. Moreover, we have

$$\lambda^s(E) = \lambda(E \cap J) \qquad \forall E \in \mathcal{B}(X)$$

where $J := (\operatorname{spt}\mu)^c \cup \{x \in \operatorname{spt}\mu \,|\, D_\mu^+\lambda(x) = +\infty\}$.

6.43 Lemma. *Let λ and μ be two Borel measures in a metric space X. Assume that μ has the doubling property. If*

$$E \subset \Big\{x \in \operatorname{spt}\mu \,\Big|\, \limsup_{r\to 0} \frac{\lambda(B(x,r))}{\mu(B(x,r))} > t\Big\},$$

then there is a constant C such that

$$\mu(E) \le \frac{C}{t} \inf\Big\{\lambda(A) \,\Big|\, A \supset E, A \text{ open}\Big\}. \tag{6.21}$$

In particular, $\mu(E) \le \frac{C}{t}\lambda(E)$ for any Borel set $E \subset X$ if λ is outer-regular.

Proof. Without loss of generality we may assume that A is an open set in X such that $A \supset E$ and $\lambda(A) < +\infty$. For any $x \in E$, there is a ball $B(x, r_x)$ with $r_x < 1$ such that $B(x, r_x) \subset A$ and $t\mu(B(x, r_x)) < \lambda(B(x, r_x))$. Let $\mathcal{B}$ be the collection of these balls. Lemma 6.39 yields a disjoint family $\mathcal{B}' \subset \mathcal{B}$ such that $\widetilde{\mathcal{B}'}$ covers E. Therefore,

$$\mu(E) \le \sum_{B\in\mathcal{B}'} \mu(\widetilde{\mathcal{B}'}) \le C \sum_{B\in\mathcal{B}'} \mu(B) \le \frac{C}{t} \sum_{B\in\mathcal{B}'} \lambda(B) \le \frac{C}{t}\lambda(A).$$

□

Actually, one can get $\mu(E) \le \frac{1}{t}\inf\{\lambda(A) \mid A \supset E,\ A \text{ open}\}$.

Proof of Theorem 6.42. (i) From Theorem 6.35, λ decomposes uniquely as $\lambda = \lambda^a + \lambda^s$ and there exists a nonnegative μ-measurable function θ with $\theta < +\infty$ μ-a.e. such that

$$\lambda^a(E) = \int_E \theta \, d\mu, \qquad \lambda^s(E) = \lambda(E \cap Z).$$

Since λ is locally finite and $\theta \in L^1_{loc}(X, \mu)$, Lebesgue's differentiation theorem, Theorem 6.16, yields, in turn, that

$$\theta(x) = \lim_{r \to 0^+} \frac{\lambda^a(B(x,r))}{\mu(B(x,r))}$$

for μ-a.e. x, i.e., $\frac{d\lambda^a}{d\mu}(x)$ exists and is finite for μ-a.e. $x \in X$.

(ii) We now prove that $\frac{d\lambda^s}{d\mu}(x) = 0$ for μ a.e. $x \in X$. To this purpose, for $t > 0$, set

$$E_t := \Big\{ x \in \operatorname{spt}\mu \,\Big|\, D^+_\mu \lambda^s(x) > t \Big\}.$$

Of course, $\mu(E_t) = \mu(E_t \cap J) + \mu(E_t \cap J^c)$, $\mu(E_t \cap J) = 0$ and $\lambda^s(E_t \cap J^c) = 0$. Since, by assumption, λ^s is an outer-regular Borel measure, see Proposition 6.2, we can apply Lemma 6.43 with $\lambda := \lambda^s$, and $E := E_t \cap J^c$ to get $t\mu(E_t \cap J^c) \le \lambda^s(E_t \cap J^c) = 0$. Therefore $\mu(E_t) = 0$. Since t is arbitrary, $\frac{d\lambda^s}{d\mu}(x) = 0$ for μ-a.e. $x \in X$.

(iii) From (i) and (ii)

$$\theta(x) = \frac{d\lambda^a}{d\mu}(x) = \frac{d\lambda^a}{d\mu}(x) + \frac{d\lambda^s}{d\mu}(x) = \frac{d\lambda}{d\mu}(x)$$

for μ-a.e. $x \in X$.

(iv) Finally, let us prove that $\lambda \llcorner J \perp \mu$. Let $K \subset J \cap \operatorname{spt}\mu$ be compact. Then for any $t > 0$ we have

$$K \subset F_t := \Big\{ x \in \operatorname{spt}\mu \,\Big|\, D^+_\mu \lambda(x) > t \Big\}.$$

Since by assumption λ is an outer-regular Borel measure, we can apply Lemma 6.43 to find $t\mu(K) \le C\lambda(K)$, hence $\mu(K) = 0$ if we let $t \to \infty$. Taking the supremum in K, we conclude that $\mu(J) = 0$. On the other hand, $\lambda \llcorner J(J^c) = \lambda(J \cap J^c) = 0$. □

c. Monotone functions

d. Stieltjes–Lebesgue's integral

Let $h : \mathbb{R} \to \mathbb{R}$ be a bounded and monotone function that, to be definite, we assume nondecreasing. The function h has left and right limits $h(x^-)$ and $h(x^+)$ respectively, at each point x, and $h(x^-) \le h(x) \le h(x^+)$; moreover, h is continuous at x if and only if h has no jump at x, i.e., $h(x^+) = h(x^-) = h(x)$. Since for every integer k there are only finitely many points where h has a jump larger than $1/k$, the *discontinuity points of h are at most denumerable.*

Let $h : \mathbb{R} \to \mathbb{R}$ be a nondecreasing, nonnegative and bounded function that, moreover, is *left-continuous.* Starting from the semiring of left-closed and right-open intervals,

$$\mathcal{I} = \Big\{ [a, b[\,\Big|\, a < b \Big\},$$

and the nonnegative set function $\alpha : \mathcal{I} \to \overline{\mathbb{R}}_+$ defined by $\alpha([a,b[) := h(b) - h(a)$, we construct a measure $(\mathcal{M}, \mu_h)$ by means of Method I, i.e., we set for $E \subset \mathbb{R}$

$$\mu_h(E) := \inf\Big\{\sum_{k=1}^{\infty}(h(b_k) - h(a_k)) \;\Big|\; E \subset \cup_k [a_k, b_k[\Big\}.$$

It is easily seen that α is finitely additive and subadditive. Moreover, the following holds.

6.44 Proposition. *Let $h : \mathbb{R} \to \mathbb{R}$ be nondecreasing, nonnegative, bounded and left-continuous. Then the set function $\alpha([a,b[) := h(b) - h(a)$ defined in the class of left-closed and right-open intervals is σ-additive.*

Proof. Let $\{I_i\} \subset \mathcal{I}$, $I_i = [x_i, y_i[$, be a disjoint covering of $I = [a,b[$. From the finite additivity we infer

$$\sum_{i=1}^{\infty} \alpha(I_i) \le \alpha(I).$$

Let us prove the opposite inequality. For $\epsilon > 0$, let $\{\delta_i\}$ be such that $h(x_i - \delta_i) \ge h(x_i) - \epsilon\, 2^{-i}$. The open intervals $]x_i - \delta_i, y_i[$ form an open covering of $[a, b-\epsilon]$, hence finitely many among them cover again $[a, b-\epsilon[$. Therefore, by the finite additivity,

$$h(b-\epsilon) - h(a) \le \sum_{k=1}^{N}(h(y_{i_k}) - h(x_{i_k} - \delta_i) = \sum_{k=1}^{N}(\alpha(I_{i_k}) + \epsilon\, 2^{-i_k}) \le \sum_{i=1}^{\infty}\alpha(I_i) + \epsilon.$$

When ϵ tends to 0, we conclude

$$\alpha(I) = h(b) - h(a) \le \sum_{i=1}^{\infty}\alpha(I_i).$$

□

6.45 Example. If h is not left-continuous, the set function $\alpha : \mathcal{I} \to \mathbb{R}$, $\alpha([a,b[) := h(b) - h(a)$ is not, in general, subadditive. For instance, for $0 \le a \le 1$, set

$$h(t) = \begin{cases} 0 & \text{if } -1 \le t < 0, \\ a & \text{if } t = 0, \\ 1 & \text{if } 0 < t \le 1. \end{cases}$$

If $\mathcal{I} = \{[-\frac{1}{j}, -\frac{1}{j+1}[\}_j \cup [0,1[$, clearly $\cup_{I\in\mathcal{I}} I = [-1,1[$, but

$$1 = h(1) - h(-1) = \alpha\Big(\cup_{I\in\mathcal{I}} I\Big) > \sum_{I\in\mathcal{I}}\alpha(I) = h(1) - h(0) = 1 - a$$

as soon as $a > 0$.

Because of Proposition 6.44, we get from Theorem 5.29 that μ_h agrees with α on $\mathcal{I}$, $\mu_h([a,b[) = h(b) - h(a)$, and that the Borel sets are μ_h-measurable, so that μ_h is a finite Borel measure in $\mathbb{R}$, called the *Stieltjes–Lebesgue measure* associated to h. The corresponding integral denoted by

$$\int \varphi \, dh := \int \varphi \, d\mu_h$$

is called the *Stieltjes–Lebesgue integral* with respect to h.

The same considerations apply also to nondecreasing, nonnegative and bounded functions that, moreoever, are *right-continuous*. In this case one may start with the class $\mathcal{J} := \{]a,b] \,|\, a < b\}$ of left-open and right-closed intervals and with $\alpha(]a,b]) := h(b) - h(a)$. If now $h : \mathbb{R} \to \mathbb{R}$ is an arbitrary nondecreasing, bounded and nonnegative function, by changing h in at most a denumerable set of points, one gets two nondecreasing, nonnegative and bounded functions labeled h' and h'' that are respectively left- and right-continuous. It is easy to see that the corresponding measures $\mu_{h'}$ and $\mu_{h''}$ constructed by Method I agree. We refer to them as the *Lebesgue–Stieltjes measure* associated to the nondecreasing function h.

6.46 Theorem (Vitali). *Let $h : \mathbb{R} \to \mathbb{R}$ be a nondecreasing, nonnegative and bounded function. Then h is differentiable at a.e. $x \in \mathbb{R}$, h' is Borel measurable and*

$$\mu_h(E) = \int_E h'(x)\, dx + \mu_h^s(E), \qquad \mu_h^s \perp \mathcal{L}^1. \tag{6.22}$$

In particular, $h' \in L^1(\mathbb{R})$, h' is nonnegative and

$$0 \le \int_x^y h'(x)\, dx \le h(y) - h(x) \qquad \forall x < y.$$

Proof. Let λ be the Stieltjes–Lebesgue measure associated to h, let $\lambda = \lambda^a + \lambda^s$ be its Lebesgue decomposition with respect to $\mathcal{L}^1$, $g \in L^1(\mathbb{R})$ be such that $\lambda^a(E) = \int g(x)\, dx$ and $Z := \{x \,|\, g = +\infty\}$ so that $|Z| = 0$ and $\lambda^s(E) = \lambda(E \cap Z)$.

(i) From the differentiation theorem for integrals, if $A \subset B(0,1)$, $|A| \ge c|B(0,1)|$ and $A_{x,r} := x + rA$, we have

$$g(x) = \lim_{r \to 0} \frac{\mu_h^a(A_{x,r})}{|A_{x,r}|} \qquad \text{for } \mathcal{L}^1\text{-a.e. } x \in \mathbb{R}.$$

(ii) For $t \ge 0$, set

$$E_t := \Big\{ x \,\Big|\, \limsup_{r \to 0} \frac{\lambda^s(B(x,r))}{2r} > t \Big\}.$$

Of course, $E_t = (E_t \cap Z) \cup (E_t \cap Z^c)$ and therefore $|E_t \cap Z| = 0$ and $\lambda^s(E_t \cap Z^c) = 0$. Applying Lemma 6.43 with $\lambda := \lambda^s$, $\mu = \mathcal{L}^1$ and $E := E_t \cap Z^c$, it follows that $|E_t \cap Z^c| = 0$, from which we deduce that $|E_t| = 0$. Since t is arbitrary,

$$\limsup_{r \to 0} \frac{\lambda^s(B(x,r)}{2r} = 0$$

for $\mathcal{L}^1$-a.e. $x \in X$ and, since $A_{x,r} \subset B(x,r)$ and $|A_{x,r}| \ge Cr$, we also have

$$\limsup_{r \to 0} \frac{\lambda^s(A_{x,r})}{|A_{x,r}|} = 0$$

for $\mathcal{L}^1$-a.e. $x \in \mathbb{R}$.

(iii) From (i) and (ii) we infer that

$$g(x) = \lim_{r\to 0} \frac{\mu_h(A_{x,r})}{|A_{x,r}|}$$

exists finite for $\mathcal{L}^1$-a.e. $x \in \mathbb{R}$. If we now choose $A_{x,r} := [x, x+r[$ and $A_{x,r} = [x-r, x[$, we find

$$g(x) = \lim_{r\to 0^+} \frac{\mu_h([x, x+r[)}{r} = \lim_{r\to 0^+} \frac{h(x+r)-h(x)}{r},$$
$$g(x) = \lim_{r\to 0^-} \frac{\mu_h([x-r, x[)}{r} = \lim_{r\to 0^-} \frac{h(x-r)-h(x)}{r}$$

for a.e. x, i.e., h is differentiable and $g(x) = h'(x)$ for $\mathcal{L}^1$-a.e. $x \in \mathbb{R}$. □

6.47 ¶. Let h be nondecreasing, nonnegative, bounded and left-continuous. Prove that $\int_a^b d\mu_h = h(b) - h(a)$ and $\mu_h(\{a\}) = h(a^+) - h(a-)$.

6.48 ¶. For $x \in \mathbb{R}$, let $h_x(y) = \chi_{]x,+\infty[}(y)$. Show that μ_h is the Dirac measure δ_x at x. Consequently, for every function $f : \mathbb{R} \to \mathbb{R}$ we have

$$\int f(y)\, d\mu_h(y) = f(x).$$

6.49 Theorem (Integration by parts). *Let $f : \mathbb{R} \to \mathbb{R}$ be of class $C^1(\mathbb{R})$ and let h be nondecreasing, nonnegative and left-continuous. Then*

$$\int_a^b f'(y)h(y)\, dy + \int_{[a,b[} f(y) d\mu_h(y) = f(b)h(b) - f(a)h(a).$$

Proof. Consider the procuct measure $\mu_h(x) \times \mathcal{L}^1(y)$ on $\mathbb{R}^2$ and let $E := \{(x,y) \in [a,b[\times[a,b[\,\big|\, a \le y \le x\}$. Fubini's theorem yields the equalities

$$\begin{aligned}
\iint_E f'(y)\, d\mu_h \times \mathcal{L}^1 &= \int_{[a,b[} \left(\int_0^x f'(y)\, dy \right) d\mu_h(x) = \int_{[a,b[} (f(x) - f(a)) d\mu_h(x) \\
&= \int_{[a,b[} f(x)\, d\mu_h(x) - f(a)(h(b) - h(a)), \\
\iint_E f'(y)\, d\mu_h \times \mathcal{L}^1 &= \int_a^b f'(y)\, d\mu_h([y,b[) = \int_a^b f'(y)(h(b) - h(y))\, dy \\
&= -\int_a^b f'(y)h(y)\, dy + h(b)(f(b) - f(a)),
\end{aligned}$$

hence the conclusion. □

6.50 ¶. If f is measurable, then $\phi(t) := \mathcal{L}^n(\{x \,|\, |f| > t\})$ is nonincreasing and right-continuous. Show that $\int |f|^p\, d\mathcal{L}^n(x) = -\int_0^\infty t^p d\phi(t)$. [*Hint.* Use Cavalieri's formula and integrate by parts.]

A function $f : \mathbb{R} \to \mathbb{R}$ has *bounded variation* in $[a,b]$ if its *total variation* in $[a,b]$ defined by

$$V_a^b(f) := \sup\Big\{ \sum_k |f(x_{k+1}) - f(x_k)| \,\Big|\, a = x_0 < x_1, \dots, x_N = b \Big\}$$

is finite. A function with bounded variation f can be written as $f = f_+ - f_-$ where f_+ and f_- are nondecreasing and with bounded variations, simply by taking $f_+ := V_a^x(f)$. In particular, from Vitali's theorem, Theorem 6.46, we infer that *every function with bounded total variation is a.e. differentiable.* However, notice that, generally, the fundamental theorem of calculus does not hold for even continuous functions with bounded variation.

6.51 Example. Let $C \subset [0,1]$ be the ternary Cantor's set and $f : [0,1] \to [0,1]$ the Cantor–Vitali function, see Chapter 5. As we have seen f is nondecreasing, $f([0,1]) = [0,1]$ and f is constant in each of the open intervals $A_{k,j}$ in which $[0,1]\backslash C$ is decomposed. It follows that $\mu_f(A_{k,j}) = 0$ and $\mu_f([0,1] \setminus C) = 0$, i.e., $\operatorname{spt} \mu_f \subset C$. Since $|C| = 0$, μ_f *and* $\mathcal{L}^1$ *are mutually singular.* Additionally, μ_f *is singular with respect to the counting measure*, in fact, $\mu_f(\{x\}) = 0\ \forall x \in [0,1]$ since f is continuous.

Finally, since $f'(x) = 0$ for a.e. $x \in [0,1]$, the fundamental theorem of calculus does not hold:

$$1 = f(1) - f(0) \neq \int_0^1 f'(t)\,dt = 0.$$

e. Absolutely continuous functions

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *absolutely continuous* if for all $\epsilon > 0$ there is a $\delta > 0$ such that whenever $\{x_k\}$ and $\{y_k\}$ are such that $\sum_{k=1}^{\infty} |x_k - y_k| < \delta$, then $\sum_{k=1}^{\infty} |f(x_k) - f(y_k)| < \epsilon$. Of course, absolutely continuous functions are uniformly continuous and map zero sets into zero sets; of course, Lipschitz functions are absolutely continuous.

Notice that Hölder-continuous functions with exponent $\alpha < 1$ are not absolutely continuous, in general. For instance, the Cantor–Vitali function, which is Hölder-continuous, see Section 5.1.3, maps the Cantor ternary set, which is of zero measure, onto a set of positive measure.

As a consequence of Lebesgue's absolute continuity theorem, for every $g \in L^1(\mathbb{R})$ the function

$$f(x) := \int_a^x g(s)\,ds$$

is absolutely continuous. We now prove that a function is absolutely continuous if and only if the fundamental theorem of calculus holds for it.

6.52 Theorem (Vitali). *A function $f : [a,b] \to \mathbb{R}$ is absolutely continuous in $[a,b]$ if and only if f is differentiable for a.e. $x \in [a,b]$, $f'(x) \in L^1([a,b])$ and*

$$f(y) - f(x) = \int_x^y f'(s)\,ds \qquad \forall x, y \in \mathbb{R},\ x < y.$$

It is readily seen that an absolutely continuous function in $[a,b]$ has finite total variation in $[a,b]$. Moreover, the following holds.

6.53 Lemma. *If $f : [a,b] \to \mathbb{R}$ is absolutely continuous, then $x \to V_a^x(f)$ is absolutely continuous in $[a,b]$, too.*

Proof. For $\epsilon > 0$ let $\delta > 0$ and let $\{x_k\},\ \{y_k\} \subset [a,b]$ be such that $\sum_k |x_k - y_k| < \delta$ and $\sum_k |f(x_k) - f(y_k)| < \epsilon$. Since the bounded variation of f is finite, for $k = 1, 2 \dots, n$ we can find a subdivision $x_k = c_1^{(k)} < c_2^{(k)} < \cdots < c_{p_k}^{(k)} = y_k$ of $[x_k, y_k]$ such that

$$V_{x_k}^{y_k}(f) \le \sum_{j=1}^{p_k - 1} |f(c_{j+1}^{(k)}) - f(c_j^{(k)})| + \epsilon.$$

Since $\sum_{j=1}^{n} |c_{j+1}^{(k)} - c_j^{(k)}| < \delta$, we have

$$\sum_{k=1}^{n} V_{x_k}^{y_k}(f) \le \epsilon + \epsilon,$$

hence $\sum_{k=1}^{n} |V_a^{x_k}(f) - V_a^{y_k}(f)| \le 2\,\epsilon$. □

Proof of Theorem 6.52. Let μ_+, μ_- be the Stieltjes–Lebesgue measures associated to f_+ and f_- given by

$$f_+(x) := V_a^x(f), \qquad f_-(x) := V_a^x(f) - f(x),$$

respectively. From the absolute continuity of f_+ and f_- we infer that $\mu_+(E) = \mu_-(E) = 0$ if $|E| = 0$, i.e., μ_+ and μ_- are absolutely continuous measures with respect to $\mathcal{L}^1$. From Theorem 6.46 we then infer that f_+ and f_- are differentiable for a.e. $x \in \mathbb{R}$, that f_+, f_- are locally summable and that

$$f_+(y) - f_+(x) = \mu_+([x,y[) = \int_x^y f_+'(x)\,dx,$$
$$f_-(y) - f_-(x) = \mu_-([x,y[) = \int_x^y f_-'(x)\,dx.$$

The proof is completed, as $f = f_+ - f_-$. □

Since Lipschitz-continuous functions are absolutely continuous, we can state the following.

6.54 Corollary. *Let $f : \mathbb{R} \to \mathbb{R}$ be Lipschitz-continuous with Lipschitz constant L. Then f is differentiable at a.e. $x \in \mathbb{R}$, $f'(x)$ is measurable, $Lip(f) = ||f'||_{\infty,\mathbb{R}}$ and, for $x < y$, we have*

$$f(y) - f(x) = \int_x^y f'(s)\,ds.$$

Here $||f||_{\infty,\mathbb{R}} := \operatorname{ess\,sup}_{t\in\mathbb{R}} |f'(t)|$.

Another consequence of Vitali's theorem, Theorem 6.52, is the following.

6.55 Proposition (Integration by parts). *Let f and g be two absolutely continuous functions in $[a,b]$. Then*

$$\int_a^b f'(x)\,g(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f(t)g'(t)\,dt.$$

f. Rectifiable curves

A curve $\gamma : [a,b] \to \mathbb{R}^n$ is said to be *rectifiable* if $\gamma(t)$ is continuous and with bounded variation. In this case, by Theorem 6.52, $\gamma'(t)$ exists for $\mathcal{L}^1$-a.e. t and $|\gamma'(t)| \in L^1([a,b])$. In turn, the arclength $s(t) := \int_0^t |\gamma'(s)|\, ds$ is absolutely continuous, too, and $s'(t) = |\gamma'(t)|$ for $\mathcal{L}^1$-a.e. $t \in [a,b]$. Because of Vitali's theorems, Theorems 6.46 and 6.52, we have the following.

6.56 Theorem (Tonelli). *Let $\gamma : [a,b] \to \mathbb{R}^n$ be a rectifiable curve. Then γ is differentiable for $\mathcal{L}^1$-a.e. $t \in [a,b]$ and the following inequality holds for the length L of $\gamma(t)$:*

$$L \geq \int_a^b |\gamma'(t)|\, dt$$

with equality if and only if the components of $\gamma(t)$ are absolutely continuous functions.

g. Lipschitz functions in $\mathbb{R}^n$

We recall that a map $f : X \to Y$ between two metric spaces is said to be *Lipschitz-continuous* if there is a constant $L > 0$ such that

$$d_Y(f(x), f(y)) \leq L\, d_X(x,y) \qquad \forall x, y \in X. \tag{6.23}$$

The best constant for which (6.23) holds is called the *Lipschitz constant* of f and is denoted by $Lip(f)$, or $Lip(f, X)$ if we want to emphasize the domain.

In many respects, Lipschitz-continuous functions are easier to handle than C^1 functions; for instance, we need only the metric structure to define them, and if f and g are Lipschitz-continuous with real values, then $\max(f,g)$, $\min(f,g)$ and $|f|$ are Lipschitz-continuous, too. Moreover, an extension theorem holds.

6.57 Theorem (Kirszbraun). *Let X be a metric space, $A \subset X$ and let $f : A \to \mathbb{R}$ be a Lipschitz-continuous function and L:=Lip(f,A). Then the functions*

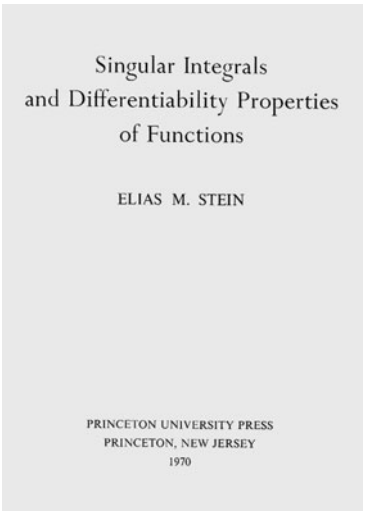
$$\widehat{f}(x) := \inf_{y \in A} (f(y) + Ld(x,y)), \qquad \widetilde{f}(x) := \sup_{y \in A} (f(y) - Ld(x,y)) \tag{6.24}$$

extend f to the whole of X with $Lip(\widehat{f}, X) = Lip(\widetilde{f}, X) = Lip(f, A)$.

Proof. We have

$$\begin{aligned} \widehat{f}(x_2) - \widehat{f}(x_1) &= \sup_{y_2 \in A} \inf_{y_1 \in A} \Big(f(y_1) + Ld(x_1, y_1) - f(y_2) - Ld(x_2, y_2) \Big) \\ &\leq \sup_{y_2 \in A} \Big(L\, d(x_1, y_2) - L\, d(x_2, y_2) \Big) \\ &\leq L\, d(x_1, x_2). \end{aligned}$$

The proof for $\widetilde{f}$ is similar. □



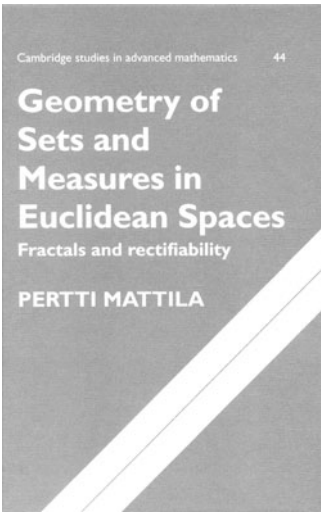


Figure 6.3. Frontispieces of two monographs that deal with the differentiation theory.

6.58 ¶. Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz-continuous function with Lipschitz constant $L := Lip(f, A)$. Show that if $g : \mathbb{R}^n \to \mathbb{R}$ is any Lipschitz extension of f to the whole of $\mathbb{R}^n$ with Lipschitz constant not greater than L, then

$$\widetilde{f}(x) \le g(x) \le \widehat{f}(x) \qquad \forall x \in \mathbb{R}^n$$

where $\widetilde{d}$ and $\widehat{f}$ are defined in (6.24).

The previous theorem allows one to extend a Lipschitz function $f : A \subset X \to \mathbb{R}^n$, $n > 1$, to the whole of X with $Lip(\widehat{f}, X) \le \sqrt{n}\, Lip(f, A)$. Actually, with a nonelementary construction, f can be extended to the whole of X with the same Lipschitz constant, $Lip(\widehat{f}, X) = Lip(f, A)$.

6.59 Theorem (Rademacher). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz function. Then f is differentiable at $\mathcal{L}^n$-a.e. point $x \in \mathbb{R}^n$, its Jacobian matrix is measurable and $Lip(f) = \operatorname{ess\,sup}_{x\in\mathbb{R}^n} |Df(x)|$.*

Proof. Let $\nu \in S^{n-1}$ be a direction and let A_ν be the set of points in $\mathbb{R}^n$ at which the derivative of f in the direction ν does not exist. Since $\frac{\partial f}{\partial \nu}(x)$ exists at points in which the Borel functions $\limsup_{t\to 0} \frac{f(x+t\nu)-f(x)}{t}$ and $\liminf_{t\to 0} \frac{f(x+t\nu)-f(x)}{t}$ agree, A_ν is a Borel set. For any $x \in \mathbb{R}^n$ and $\nu \in S^{n-1}$, $t \mapsto \varphi_{x,\nu}(t) := f(x + t\nu)$ is a Lipschitz function on $\mathbb{R}$, hence is differentiable for a.e. t. In other words, the intersection of A_ν with a line parallel to ν has zero measure. Moreover, $Lip(\varphi_{x,\nu}) = \operatorname{ess\,sup}_{t\in\mathbb{R}} |\varphi'_{x,\nu}(t)|$. Fubini's theorem then yields that A_ν has zero measure, i.e.,

$$\begin{aligned} \frac{\partial f}{\partial \nu}(x) = \lim_{t\to 0} \frac{f(x+t\nu) - f(x)}{\theta} \text{ exists for a.e. } x \in \mathbb{R}^n, \\ \Big|\frac{\partial f}{\partial \nu}(x)\Big| \le Lip(f) \end{aligned} \tag{6.25}$$

and

$$\begin{aligned} Lip(f) &= \sup_{x,\nu} Lip(\varphi_{x,\nu}) = \sup_{\nu} \operatorname{ess\,sup}_{x\in\mathbb{R}^n} \operatorname{ess\,sup}_{t\in\mathbb{R}} \left|\frac{\partial f}{\partial \nu}(x+t\nu)\right| \\ &= \sup_{\nu} \operatorname{ess\,sup}_{z\in\mathbb{R}^n} \left|\frac{\partial f}{\partial \nu}(z)\right| \end{aligned} \tag{6.26}$$

for a.e. $x \in \mathbb{R}^n$.

Now, for every $\varphi \in C_c^1(\mathbb{R}^n)$

$$\int_{\mathbb{R}^n} \frac{f(x+h\nu)-f(x)}{h}\,\varphi(x)\,dx = \int_{\mathbb{R}^n} \frac{\varphi(x)-\varphi(x-h\nu)}{h}\,f(x)\,dx;$$

notice that we have used Fubini's theorem and the absolute continuity of f to integrate by parts, see Proposition 6.55. Hence, letting $h \to 0$, because of Lebesgue's dominated convergence theorem and (6.25),

$$\int \frac{\partial f}{\partial \nu}\varphi\,dx = -\int f\frac{\partial \varphi}{\partial \nu}\,dx, \qquad \int D_j f\varphi\,dx = -\int f D_j\varphi\,dx,\ \forall j$$

and therefore,

$$\begin{aligned} \int \frac{\partial f}{\partial \nu}\varphi\,dx &= -\int f\frac{\partial \varphi}{\partial \nu}\,dx = -\int f(\,\nu\bullet\nabla\varphi\,)\,dx = -\sum_{j=1}^n \int f\nu^j D_j\varphi\,dx \\ &= \sum_{j=1}^n \int \nu^j\varphi D_j f\,dx = \int \varphi(\,\nu\bullet\nabla f\,)\,dx. \end{aligned}$$

Since φ is arbitrary, we conclude

$$\frac{\partial f}{\partial \nu}(x) = \nu\bullet\nabla f(x) \qquad \text{for a.e. } x\in\mathbb{R}^n \tag{6.27}$$

and, from (6.25) and (6.26),

$$Lip(f) = ||\nabla f||_{\infty,\mathbb{R}^n}. \tag{6.28}$$

The differentiability of f a.e.. remains to be proved. Let $\{\nu_1,\nu_2,\dots\}$ be a denumerable dense set of S^{n-1} and let

$$A_k := \left\{x \,\middle|\, \nabla f(x), \frac{\partial f}{\partial \nu_k}f(x) \text{ exists and } \frac{\partial f}{\partial \nu_k}f(x) = \nu_k\bullet\nabla f(x)\,,\ |\nabla f(x)|\le L\right\},$$

$L := Lip(f)$. From (6.25), (6.27) and (6.28), if $A := \cap_k A_k$, we have $|A^c| = 0$ $\frac{\partial f}{\partial \nu_k}f(x) = \nu_k\bullet\nabla f(x)$ $\forall x\in A$ and all $k = 1,2,\dots$, and $|\nabla f(x)|\le L$.

We now prove that f is differentiable at every $x\in A$. For $x\in A$, $\nu\in S^{n-1}$ and $h>0$, we set

$$Q(x,\nu,h) := \frac{f(x+h\nu)-f(x)}{h} - \nu\bullet\nabla f(x)\,.$$

Clearly, for $x\in A$ and $\nu,\nu'\in S^{n-1}$ and $h>0$, we have

$$|Q(x,\nu,h)-Q(x,\nu',h)| \le (n+1)\,L\,|\nu-\nu'|. \tag{6.29}$$

Given $\epsilon>0$, we now choose p large enough so that for every $\nu\in S^{n-1}$ we have

$$|\nu-\nu_k| < \frac{\epsilon}{2(n+1)\,L} \qquad \text{for some } k\in\{1,\dots,p\}. \tag{6.30}$$

As $\lim_{h\to 0} Q(x,\nu_l,h) = 0$ $\forall l$, we can find $\delta>0$ such that

$$|Q(x,\nu_k,h)| < \frac{\epsilon}{2} \qquad \text{for } 0<h<\delta \text{ and } k\in\{1,\dots,p\}. \tag{6.31}$$

On the other hand, $|Q(x,\nu,h)| \le |Q(x,\nu_k,h)| + |Q(x,\nu,h)-Q(x,\nu_k,h)|$, therefore, we conclude from (6.29), (6.30) and (6.31) that

$$|Q(x,\nu,h)| < \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon$$

whenever $0<h<\delta$ uniformly with respect to $\nu\in S^{n-1}$. This proves the theorem. □

6.2.4 Differentiation of measures in $\mathbb{R}^n$

The differentiation theorem for Radon measures with respect to a doubling measure, Theorem 6.42, was first proved by Vitali for Radon measures in $\mathbb{R}$ with respect to $\mathcal{L}^1$ with a direct argument which uses neither the Lebesgue–Nikodym theorem nor Lusin's theorem. His proof is grounded on a covering argument and on taking into account the homogeneity of the measure with respect to dilations, see Theorem 6.66.

In the middle of the twentieth century, Abram Besicovitch (1891–1970) proved a much stronger covering theorem. The Vitali covering property and, consequently, the differentiability of measures in $\mathbb{R}^n$ extend this way to arbitrary Radon measures.

a. The Besicovitch and Vitali covering theorems

6.60 Theorem (Besicovitch). *Let $\ell > 1$ and $n \in \mathbb{N}$, $n \geq 1$. There is a constant $c(n,\ell)$ with the following property: For any $E \subset \mathbb{R}^n$ and any bounded function $r : E \to \mathbb{R}$ there is a denumerable subset $X \subset E$ that decomposes as disjoint union $X = \cup_{j=0}^{\infty} X_j$ such that the following hold:*

(i) $E \subset \cup_{x\in X} \operatorname{int}(B(x, r(x)))$.
(ii) *The family* $\left\{B(x, \frac{r(x)}{2\ell})\right\}_{x\in X}$ *is disjoint.*
(iii) *Every ball $B(a, r(a))$, $a \in X_h$, meets at most $c(n,\ell)$ balls $B(x, r(x))$ with $x \in X_j$, $j \leq h$.*
(iv) *Every point of $\mathbb{R}^n$ lies in at most $c(n,\ell)$ balls $B(x, r(x))$, $x \in X$.*
(v) *The X_j's are finite sets if E is bounded.*

Here $B(x, r(x))$ denotes the round closed ball *centered at x of radius $r(x)$.*

Proof. If E is denumerable, the selection principle is as follows. First order the balls in a sequence of decreasing radii. Take the first one and, inductively, assuming that B_{n_1}, ..., B_{n_k} have already been selected, choose as $B_{n_{k+1}}$ the first ball B_j such that $j > n_k$ and does not meet the centers of B_{n_1}, ..., B_{n_k}.

Since, in general, E is not denumerable, one proceeds using the axiom of choice.[2]

Let $M := \sup_{x\in E} r(x)$. Define for $j = 0, 1, \ldots$

$$E_j := \left\{x \in E \,\middle|\, \ell^{-j-1}M < r(x) \leq \ell^{-j}M\right\},$$

and setting $X_{-1} = \emptyset$, inductively we define for $j = 0, 1, \ldots$ $X_j \subset E_j$ in such a way that $X := \cup_j X_j$ has the requested properties. Suppose $X_0, X_1, \ldots, X_{j-1}$ have been defined, and let $\mathcal{C}_j$ be the class of all denumerable subsets Y of $E_j \setminus \cup_{i<j} \cup_{x\in X_i} B(x, r(x))$, such that

$$|y - y'| \geq \ell^{-j-1}M.$$

Clearly, $\mathcal{C}_j$ is partially ordered by inclusion and every chain $Y_1 \subset Y_2 \subset \cdots$ in $\mathcal{C}_j$ has maximal element $\cup_i Y_i$. Zorn's lemma then yields the existence of a maximal element X_j in $\mathcal{C}_j$. Let us collect the properties of the points of $X = \cup_j X_j$. If $x, x' \in X_j$, $x \neq x'$, then

[2] Taken from E. Bombieri, *Notes on Geometric Measure Theory*, unpublished, Pisa, 1974.

$$\begin{cases} |x - x'| \geq \ell^{-j-1} M, \\ \ell^{-j-1} M < r(x), r(x') \leq \ell^{-j} M; \end{cases} \tag{6.32}$$

if $x \in X_j$, $x' \in X_h$ and $j < h$, then

$$\begin{cases} |x - x'| \geq \max(r(x), r(x')), \\ \ell^{-j-1} M < r(x) \leq \ell^{-j} M, \\ \ell^{-h-1} M < r(x') \leq \ell^{-h} M. \end{cases} \tag{6.33}$$

We divide the proof in five steps.

(i) We have $E_j \subset \cup_{x \in X_j} \operatorname{int} B(x, r(x))$. In fact, if $y \in E_j \setminus \cup_{x \in X_j} \operatorname{int} B(x, r(x))$, then $|y - x| \geq r(x) > \ell^{-j-1} M$ for all $x \in X_j$, consequently $X_j \cup \{y\} \in \mathcal{C}_j$, and this contradicts the maximality of X_j.

(ii) If $y \in B(x, r(x)/(2\ell)) \cap B(x', r(x')/(2\ell))$, then $|x' - x| \leq |x - y| + |x' - y| \leq \frac{r(x)}{2\ell} + \frac{r(x')}{2\ell} \leq \frac{1}{\ell} \max(r(x), r(x'))$. However, this contradicts (6.32) if x and x' are in the same X_j or (6.33) if x and x' belong to different X_j.

(iii) Fix h and for any $a \in X_h$ let

$$X(a) := \Big\{ x \in \cup_{j \leq h} X_j \;\Big|\; B(x, r(x)) \cap B(a, r(a)) \neq \emptyset \Big\}.$$

We prove that

$$\operatorname{card}(X(a) \cap X_j) \leq c_1(n, \ell) \qquad \forall j \leq h,$$

$$\operatorname{card}\Big\{ j \leq h - 1 \;\Big|\; X(a) \cap X_j \neq \emptyset \Big\} \leq c_2(n, \ell),$$

with constants c_1 and c_2 independent of h. It follows that $\operatorname{card} X(a) \leq c(n, \ell) := c_1(n, \ell)\,(c_2(n, \ell) + 1)$.

Let $x, x' \in X(a) \cap X_j$, $j \leq h$. Then

$$\begin{cases} |x - x'| > \ell^{-j-1} M, \\ |x - a| \leq r(x) + r(a) \leq 2\ell^{-j} M, \\ |x' - a| \leq r(x') + r(a) \leq 2\ell^{-j} M. \end{cases}$$

Thus, the number of points in $X(a) \cap X_j$ is at most the number of points in $B(0, 2)$ that have distance at least $1/\ell$, hence

$$\operatorname{card}(X(a) \cap X_j) \leq c_1(n, \ell).$$

Notice that $c_1(n, \ell)$ is increasing with ℓ and does not depend on h.

If now $\{x_i\} \in X(a) \cap X_i$, $i \leq h - 1$, since $a \in X_h$ and $x_i \in X_i$, we have

$$r(a) \leq \ell^{-h} M \leq \ell^{-i-1} M < \frac{1}{\ell} r(x_i) \qquad \forall i \leq h - 1,$$

hence

$$r(a) < \frac{1}{\ell} \min_{i \leq h-2} r(x_i)$$

and $\forall i, j \leq h - 1$

$$\begin{cases} |x_i - x_j| \geq \max(r(x_i), r(x_j)), \\ r(x_i) \geq |x_i - a| - r(a). \end{cases}$$

Therefore, if x and x' are two points chosen among the x_i's,

$$|x-x'| \geq \max(r(x), r(x')) \geq \max(|x-a|, |x'-a|) - r(a)$$
$$\geq \max(|x-a|, |x'-a|) - \frac{1}{\ell}\min(|x-a|, |x'-a|).$$

Assuming that $|x-a| > |x'-a|$, we then conclude

$$\frac{|x-x'| - |x-a|}{|x'-a|} > -\frac{1}{\ell},$$

hence

$$\Big|\frac{x-a}{|x-a|} - \frac{x'-a}{|x'-a|}\Big| = \Big|\frac{x-x'}{|x'-a|} - (x-a)\Big(\frac{1}{|x'-a|} - \frac{1}{|x-a|}\Big)\Big|$$
$$\geq \frac{|x-x'| - |x-a|}{|x'-a|} + 1 \geq 1 - \frac{1}{\ell} > 0.$$

It follows that the number of x_i's is not larger than the number of points of S^{n-1} with distance at least $1 - 1/\ell$, hence it is not larger than a constant $c_2(n, \ell)$ independent of h that this time is decreasing with ℓ. In conclusion,

$$\operatorname{card}\Big\{j \leq h-1 \;\Big|\; X(a) \cap X_j \neq \emptyset\Big\} \leq c_2(n, \ell),$$

and (iii) is completely proved.

(iv) This follows from (iii).

(v) If E is bounded, all X_j are finite by definition. □

6.61 Corollary (Besicovitch). *Let $E \subset \mathbb{R}^n$ be a bounded set and let $r : E \to \mathbb{R}$ be a bounded function. There is a denumerable subset $X \subset E$ and a constant $c(n)$ such that*

(i) $E \subset \cup_{x\in X} \operatorname{int}(B(x, r(x)))$,
(ii) *the family $\{B(x, r(x))\}_{x\in X}$ decomposes in at most $c(n)$ subfamilies of disjoint balls.*

Proof. By choosing $\ell = 2$ in Theorem 6.60, we find a denumerable subset $X \subset E$ that is a disjoint union of finite sets $X = \cup X_j$ with property (i) such that each ball $B(a, r(a))$ with $a \in X_h$ meets at most $c(n)$ balls $B(x, r(x))$, $x \in X_j$, $j \leq h$. We order the centers in a sequence $\{x_j\}$ first enumerating the points in X_0, then those in X_1 and so on. Suppose that inductively we have inserted the balls $B(x, r(x))$ with the first $j-1$ centers in p families $\mathcal{B}_1, \ldots \mathcal{B}_p$ of disjoint balls. We put the next ball $B_j := B(x_j, r(x_j))$ in the first family $\mathcal{B}_k$ for which $\{B_j\} \cup \mathcal{B}_k$ is again a disjoint family or we start a new family $\mathcal{B}_{p+1}$ with B_j. This second alternative holds if B_j meets at least p balls with preceding indices, but by construction $p \leq c(n)$. □

6.62 Remark. Going through the proof of Theorem 6.60, one can see that the theorem still works if the balls are open or, with a slightly different proof, if we replace balls with cubes. We notice instead that the boundedness of $r : E \to \mathbb{R}$ is essential: Conclusion (ii) does not hold if $E = [0, 1] \subset \mathbb{R}$ and $r(x) = 2|x|$; one cannot replace centered intervals $B(x, r)$ with half-intervals. If $A =]0, 1[$ and $B(x, r) := [x, x+1[$, conclusions (i) and (ii) cannot hold at the same time.

6.63 Definition. *Let $\mathcal{F}$ be a family of closed subsets of a metric space X. We say that $\mathcal{F}$* covers finely *$A \subset X$ if for any $x \in A$ and for any $\epsilon > 0$ there is an $F \in \mathcal{F}$ with $x \in F$ and diameter$(F) < \epsilon$.*

6.64 Theorem (Vitali). *Every Radon measure μ in $\mathbb{R}^n$ has the following property: If $A \subset \mathbb{R}^n$ is a bounded Borel set and $\mathcal{F}$ is a family of closed balls that finely covers A, then there is a disjoint denumerable subfamily $\mathcal{F}' \subset \mathcal{F}$ such that*

$$\mu(A \setminus \bigcup \mathcal{F}') = 0,$$

where we shorten $\bigcup_{B\in\mathcal{F}'} B$ in $\cup\mathcal{F}'$.

Proof. Let $c(n)$ be the constant in Besicovitch's theorem and let $\delta := 1 - (2c(n))^{-1}$. Set $\mathcal{F}_0 := \mathcal{F}$ and $A_0 := A$. For one of the denumerable families $\mathcal{B}_0$ of disjoint balls in the thesis of Besicovitch's theorem, Corollary 6.61, we have

$$\mu\Big(A \cap \bigcup \mathcal{B}_0\Big) \geq \frac{1}{2\,c(n)}\mu(A),$$

consequently, if we set $A_1 := A \setminus \bigcup \mathcal{B}_0$, we have $\mu(A_1) \leq \delta\,\mu(A)$. Since $\bigcup \mathcal{B}_0$ is compact, the family

$$\mathcal{F}_1 := \Big\{ B \in \mathcal{F} \;\Big|\; B \cap \bigcup \mathcal{B}_0 = \emptyset \Big\}$$

is again a fine covering of A_1, and we can repeat the argument for A_1 and $\mathcal{F}_1$ in place of A_0 and $\mathcal{F}_0$, respectively. By induction, one contructs for $k = 1, 2, \dots$ a set A_k and a subfamily $\mathcal{B}_k$ of disjoint balls of $\mathcal{F}$ that are pairwise disjoint and also disjoint of the balls in the families $\mathcal{B}_0, \mathcal{B}_1, \dots \mathcal{B}_{k-1}$, such that

$$A_{k+1} := A_k \setminus \bigcup \mathcal{B}_k, \qquad \mu(A_{k+1}) \leq \delta\mu(A_k).$$

Therefore, if $\mathcal{F}' := \cup_k \mathcal{B}_k$, we have

$$\mu\Big(A \setminus \bigcup \mathcal{F}'\Big) \leq \mu\Big(\bigcap_k A_k\Big) = 0.$$

□

Theorem 6.64 was first proved by Giuseppe Vitali (1875–1932) for Lebesgue's measure $\mathcal{L}^1$ in $\mathbb{R}$ with a much simpler proof grounded on the selection algorithm of Lemma 6.18 and taking into account the homogeneity of the measure with respect to dilations. The same proof by Vitali works for Borel-regular measures in metric spaces with the *doubling property*. For the reader's convenience, we present a proof of Theorem 6.64 that works for doubling measures in metric spaces and that avoids the use of Besicovitch's covering argument.

Recall the notations of Lemma 6.39.

6.65 Proposition. *Lat A be a bounded Borel set in a metric space X and $\mathcal{B}$ be a family of closed balls of bounded diameters that finely covers $A \subset X$. Then there exists a subfamily $\mathcal{B}'$ of $\mathcal{B}$ with the following property: For every finite choice of elements $B_1, \dots B_N \in \mathcal{B}'$, we have*

$$A \setminus \bigcup_{i=1}^{N} B_i \subset \bigcup_{B\in\mathcal{B}'\setminus\{B_1,\cdots,B_N\}} \widetilde{B}.$$

Proof. Let $\mathcal{B}'$ be the family chosen as in Lemma 6.39 and $x \in A \setminus \cup_{i=1}^{N} B_i$. Since $\mathcal{B}$ finely covers A and $X \setminus \cup_{i=1}^{N} B_i$ is open, there exist $B \in \mathcal{B}$ with $x \in B$ such that $B \cap (\cup_{i=1}^{N} B_j) = \emptyset$, and $S \in \mathcal{B}'$ that meets B and $\widetilde{S} \supset B$. S may be none of the $B_1, \dots, B_N$, hence $x \in \cup_{B \in \mathcal{B}' \setminus \{B_1, \cdots, B_N\}} \widetilde{B}$. □

Proof of Theorem 6.64 for doubling measures. Let Ω be an open set such that $\Omega \supset A$ and $\mu(\Omega) < +\infty$. The family of closed balls with bounded diameters

$$\mathcal{B} := \Big\{ B \in \mathcal{F} \,\Big|\, B \subset \Omega,\ \operatorname{diam}(B) \le 1 \Big\}$$

finely covers A. By Proposition 6.65, there is a subfamily of disjoint balls $\mathcal{B}' \subset \mathcal{B}$ such that for any finite choice of $B_1, \dots, B_p \in \mathcal{B}'$,

$$A \setminus \bigcup_{k=1}^{p} B_k \subset \bigcup_{\substack{B \in \mathcal{B}' \\ B \neq B_1, \cdots, B_p}} \widetilde{B}.$$

Since $\sum_{B \in \mathcal{B}'} \mu(B) \le \mu(\Omega) < +\infty$, we have $\mu(B) > 0$ for at most denumerable many of them. Let $\mathcal{F}' = \{B_n\}$ be the family of these balls. For any integer n we then have

$$A \setminus \bigcup_{k=1}^{n} B_k \subset \bigcup_{\substack{B \in \mathcal{B}' \\ B \neq B_1, \cdots B_n}} \widetilde{B},$$

hence

$$\mu\Big(A \setminus \bigcup_{k=1}^{n} B_k\Big) \subset \sum_{\substack{B \in \mathcal{B}' \\ B \neq B_1, \cdots B_n}} \mu(\widetilde{B}) \le C \sum_{\substack{B \in \mathcal{B}' \\ B \neq B_1, \cdots B_n}} \mu(B) = C \sum_{k=n+1}^{\infty} \mu(B_k),$$

thus concluding that $\mu\Big(A \setminus \bigcup \mathcal{F}'\Big) = 0$ since $\sum_{k=0}^{\infty} \mu(B_k) \le \mu(\Omega) < +\infty$. □

b. Radon–Nikodym's derivative

Let μ and λ be two Radon measures $\mathbb{R}^n$. For $x \in \operatorname{spt} \mu$ let

$$D_\mu^+ \lambda(x) = \limsup_{\rho \to 0^+} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))}, \qquad D_\mu^- \lambda(x) = \liminf_{\rho \to 0^+} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))},$$

and if $D_\mu^+ \lambda(x) = D_\mu^- \lambda(x) x \in \operatorname{spt} \mu$, let

$$\frac{d\lambda}{d\mu}(x) = \lim_{\rho \to 0^+} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))}$$

be the common value, that is called the *Radon–Nikodym derivative* of λ with respect to μ at x.

Recall that $D_\mu^+ \lambda$ and $D_\mu^- \lambda$ are Borel functions and that they do not change if we replace the open balls used to define them with closed balls.

6.66 Theorem (Lebesgue–Besicovitch). *Let λ and μ be two Radon measures in $\mathbb{R}^n$. Then for μ-a.e. $x \in \operatorname{spt} \mu$ the Radon–Nikodym derivative*

$$\frac{d\lambda}{d\mu}(x) = \lim_{\rho \to 0} \frac{\lambda(B(x,\rho))}{\mu(B(x,\rho))}$$

exists and is finite. Moreover, if

$$I := (\operatorname{spt}\mu)^c \cup \Big\{ x \in \operatorname{spt}\mu \;\Big|\; D_\mu^- \lambda(x) = +\infty \Big\},$$

$\lambda \llcorner I$ *is singular with respect to* μ *and*

$$\lambda(E) = \int_E \frac{d\lambda}{d\mu}(x)\, d\mu + \lambda \llcorner I(E)$$

for every Borel set $E \subset \mathbb{R}^n$.

In order to prove the theorem, we first state the following two lemmas.

6.67 Lemma. *Let* λ *and* μ *be two Radon measures in* $\mathbb{R}^n$, *let* $t \geq 0$ *and let* $E \subset \operatorname{spt}\mu$ *be a Borel set.*

(i) *If* $D_\mu^+ \lambda(x) \geq t \;\forall x \in E$, *then* $\lambda(E) \geq t\mu(E)$.
(ii) *If* $D_\mu^- \lambda(x) \leq t \;\forall x \in E$, *then* $\lambda(E) \leq t\mu(E)$.

Proof. (i) We may and do assume that E is bounded, For $\epsilon > 0$ let $A \supset E$ be an open set with compact closure such that $A \supset E$ and $\mu(A \setminus E) < \epsilon$. For each $x \in E$ we consider the family $\mathcal{B}$ of closed balls $B(x,r)$ with $x \in E$ and $B(x,r) \subset A$ such that $(t-\epsilon)\mu(B(x,r)) \leq \lambda(B(x,r))$. It is easily seen that $\mathcal{B}$ finely covers E. Vitali's theorem provides us with a disjoint denumerable subfamily $\mathcal{B}' \subset \mathcal{B}$ such that $\mu(E \setminus \bigcup \mathcal{B}') = 0$. Consequently, we have

$$(t-\epsilon)\mu(E) \leq (t-\epsilon) \sum_{B \in \mathcal{B}'} \mu(B) \leq \sum_{B \in \mathcal{B}'} \lambda(B) \leq \lambda(A) \leq \lambda(E) + \epsilon,$$

and letting $\epsilon \to 0$, we have the thesis.

(ii) This is proved as in (i) using Vitali's theorem for the measure λ. □

6.68 Lemma. *Let* λ *and* μ *be two Radon measures in* $\mathbb{R}^n$ *and let*

$$I := \Big\{ x \in \operatorname{spt}\mu \;\Big|\; D_\mu^- \lambda(x) = +\infty \Big\}, \qquad J := \Big\{ x \in \operatorname{spt}\mu \;\Big|\; D_\mu^+ \lambda(x) = +\infty \Big\}.$$

Then $\mu(J) = 0$ *and* $\lambda \llcorner J$ *is singular with respect to* μ. *Moreover,* $I \subset J$, $\lambda \llcorner I$ *is singular with respect to* μ, $\lambda \llcorner I^c$ *is absolutely continuous with respect to* μ *and*

$$\lambda\Big(\Big\{ x \in \operatorname{spt}\mu \;\Big|\; D_\mu^- \lambda(x) = 0 \Big\} \Big) = 0. \tag{6.34}$$

Proof. Of course, $I \subset J$ and for

$$J_t := \Big\{ x \in \operatorname{spt}\mu \;\Big|\; D_\mu^+ \lambda(x) > t \Big\},$$

$t > 0$, (i) of Lemma 6.67 yields $t\,\mu(J_t) \leq \lambda(\mathbb{R}^n)$. For $t \to +\infty$ we conclude $\mu(J) = 0$, hence $\lambda \llcorner J$ is singular with respect to μ.

We now prove that $\lambda \llcorner I^c$ is absolutely continuous with respect to μ. Given B with $\mu(B) = 0$, Lemma 6.67 yields $\lambda(B \cap I_t) \leq t\mu(B) = 0$ for all $t > 0$, where

$$I_t := \{ x \in \operatorname{spt}\mu \;\big|\; D_\mu^- \lambda(x) \leq t \},$$

hence $\lambda(B \cap I^c) = 0$. Finally, (6.34) follows from (ii) of Lemma 6.67. □

Proof of Theorem 6.66. First, assume that λ and μ are finite and set $\lambda^a := \lambda \llcorner I^c$ and $\lambda^s := \lambda \llcorner I$. We have $\lambda = \lambda^a + \lambda^s$ and, according to Lemma 6.68, λ^s is singular with respect to μ, and λ^a is absolutely continuous with respect to μ. It suffices then to prove that the Radon–Nikodym derivative of λ with respect to μ exists for μ-a.e. $x \in \operatorname{spt}\mu \setminus I$.

To prove this, we set

$$\lambda^+(E) := \int_{E\cap \operatorname{spt}\mu} D^+_\mu(x)\, d\mu, \qquad \lambda^-(E) := \int_{E\cap \operatorname{spt}\mu} D^-_\mu(x)\, d\mu$$

and show that $\lambda^+(E) \le \lambda^a(E) \le \lambda^-(E)$ for every Borel set E from which we clearly infer that $D^+_\mu\lambda(x) = D^-_\mu\lambda(x) \in \mathbb{R}$ for μ-a.e. $x \in \operatorname{spt}\mu$ and $\lambda^a(E) = \int_E \frac{d\lambda}{d\mu}\, d\mu$.

Given a Borel set E, $t > 1$ and $m \in \mathbb{Z}$, we set

$$E_m := \Big\{ x \in E \cap I^c \;\Big|\; D^+_\mu\lambda(x) \in]t^m, t^{m+1}] \Big\}$$

and $E_\infty := \Big\{ x \in E \cap I^c \;\Big|\; D^+_\mu\lambda(x) = +\infty \Big\}$. We have

$$\lambda^+(E_m) \le t^{m+1}\mu(E_m) \le t\lambda(E_m) = t\lambda^a(E_m),$$

whereas, according to Lemma 6.68, $\mu(E_\infty) = 0$, hence $\lambda^+(E_\infty) = 0$. Summing on $m \in \mathbb{Z}$ we obtain

$$\lambda^+(E) = \sum_{m\in\mathbb{Z}} \lambda^+(E_m) \le t \sum_{m\in\mathbb{Z}} \lambda^a(E_m) \le t\lambda^a(E)$$

and, for $t \to 1$, we infer $\lambda^+(E) \le \lambda^a(E)$. Similarly, for

$$E^m := \Big\{ x \in E \setminus I \;\Big|\; D^-_\mu\lambda(x) \in]t^m, t^{m+1}] \Big\},$$

we infer

$$t^{-1}\lambda^a(E^m) = t^{-1}\lambda(E^m) \le t^m\mu(E^m) \le \lambda^-(E).$$

Since by Lemma 6.68 we have $\lambda(\{x \in \operatorname{spt}\mu \mid D^-_\mu\lambda(x) = 0\}) = 0$, we conclude

$$t^{-1}\lambda^a(E) = t^{-1}\sum_{m\in\mathbb{Z}} \lambda^a(E^m) \le \sum_{m\in\mathbb{Z}} \lambda^-(E^m) \le \lambda^-(E)$$

that yields $\lambda^a(E) \le \lambda^-(E)$ when $t \to 1$. The theorem is then proved when λ and μ are finite.

To prove the theorem in the general case, it suffices to decompose $\mathbb{R}^n$ as $\mathbb{R}^n = \cup_h X_h$, where the X_h are open and bounded sets. □

6.2.5 Disintegration of measures

In this subsection we state and prove a generalization of Fubini's theorem to Radon measures.

6.69 Example. Let $\rho : \mathbb{R}^2 \to \mathbb{R}$ be a summable function in $\mathbb{R}^2$ with nonzero integral. Fubini's theorem yields that $\sigma(x) := \int_{\mathbb{R}} \rho(x,y)\, d\mathcal{L}^1(y)$ exists for a.e. $x \in \mathbb{R}$, is finite and nonzero and summable. Moreover, for any nonnegative Borel function φ on $\mathbb{R}^2$ we have

$$\int_{\mathbb{R}^2} \varphi(x,y)\,\rho(x,y)\, dxdy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{\rho(x,y)}{\sigma(x)}\, d\mathcal{L}^1(y) \right) \sigma(x)\, dx. \tag{6.35}$$

If ν denotes the Radon measure $d\nu(x,y) := \rho(x,y)\, d\mathcal{L}^2(x,y)$, then the measure $\mu := \pi_\#\nu$ on $\mathbb{R}$ defined by $\pi_\#\nu(A) := \nu(A \times \mathbb{R})$ is just $d\mu(x) = \sigma(x)\, dx$. If for a.e. $x \in \mathbb{R}$ we set

$$\nu_x(A) := \int_A \frac{\rho(x,y)}{\sigma(x)}\, d\mathcal{L}^1(y), \qquad A \subset \mathcal{B}(\mathbb{R}),$$

then ν_x is a finite Radon measure with $\nu_x(\mathbb{R}) = 1$ and (6.35) writes as

$$\int_{\mathbb{R}^2} \varphi(x,y)\, d\nu(x,y) = \int_{\mathbb{R}} \left(\int \varphi(x,y)\, d\nu_x(y) \right) d\mu(x). \tag{6.36}$$

This is the *disintegration formula* for ν with respect to the x-variable.

The above can be done for general measures.

6.70 Theorem (Disintegration of measures). *Let ν be a finite Radon measure on $\mathbb{R}^n \times \mathbb{R}^m$, $n, m \geq 1$, let $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be the orthogonal projection of the first factor, $\pi(x,y) = x$ and let $\mu := \pi_\# \nu$ be the projection of ν into $\mathbb{R}^n$ defined as the finite Borel measure $\mu(A) = \pi_\#\nu(A) := \nu(A \times \mathbb{R}^m)$, $A \in \mathcal{B}(\mathbb{R}^n)$. Then for μ-a.e. x there exists a finite Borel measure ν_x on $\mathbb{R}^m$ such that the following hold:*

(i) *For any $B \in \mathcal{B}(Y)$ the function $x \to \nu_x(B)$ is μ-measurable.*
(ii) *For μ-a.e. x we have*

$$\nu_x(B) = \lim_{r \to 0} \frac{\nu(B(x,r) \times B)}{\nu(B(x,r) \times \mathbb{R}^m)},$$

in particular, ν_x is a probability measure, $\nu_x(\mathbb{R}^m) = 1$.
(iii) *For any $A \in \mathcal{B}(X)$ and any $B \in \mathcal{B}(Y)$ we have*

$$\nu(A \times B) = \int_A \nu_x(B)\, d\mu(x).$$

Consequently, if $f \in L^1(X \times Y, \nu)$, then the following hold:

(i) *For μ-a.e. $x \in X$ the function $y \to f(x,y)$ is ν_x-measurable.*
(ii) *$x \to \int f(x,y)\, d\nu_x(y)$ belongs to $L^1(X, \mu)$.*
(iii) *We have*

$$\int_{X \times Y} f(x,y)\, d\nu(x,y) = \int_X \left(\int_Y f(x,y)\, d\nu_x(y) \right) d\mu(x).$$

Proof. For any $B \in \mathcal{B}(Y)$ we consider the measure

$$\nu_B(A) := \nu(A \times B), \qquad A \in \mathcal{B}(X).$$

Trivially, $\nu_B << \mu$ and, according to Besicovitch's differentiation theorem, the following hold:

(i) There is a set $N_B \subset X$ with $\mu(N_B) = 0$ such that for any $x \notin N_B$ there exists

$$h(x,B) := \frac{d\nu_B}{d\mu}(x) \in \mathbb{R} \qquad \text{with} \qquad 0 \leq h(x,B) \leq 1.$$

(ii) The function $x \mapsto h(x,B)$ is μ-measurable.
(iii) For any $A \subset \mathcal{B}(X)$

$$\nu(A \times B) = \int_A h(x,B)\, d\mu(x).$$

Moreover, $h(x,B)=0$ for any $x\notin N_B$ if $\nu(\mathbb{R}^n\times B)=0$ and, if $\{B_i\}\subset\mathcal{B}(\mathbb{R}^m)$ is a disjoint family with $B=\cup_i B_i$, then

$$h_B(x)=\sum_{i=1}^{\infty}h_{B_i}(x)\qquad\forall x\notin N_B\cup\bigcup_i N_{B_i}. \tag{6.37}$$

Denote by $\mathcal{R}$ the family of half-intervals R with vertices of rational coordinates in $\mathbb{R}^n$. Of course, $\mathcal{R}$ is denumerable and the set $N:=\cup_{R\in\mathcal{R}}N_R$ has zero μ-measure. Additionally, for $x\notin N$, $h(x,R)$ is defined for all $R\in\mathcal{R}$. It is easy to see that the set function $\alpha_x:\mathcal{R}\to\mathbb{R}$, $\alpha_x(R):=h(x,R)$ is σ-additive on $\mathcal{R}$. Let ν_x be the measure constructed by Method I from $(\mathcal{R},\alpha_x)$. Again by (6.37), for any $B\in\mathcal{B}$

$$\nu_x(B)=h_(x,B)\qquad\text{for } x\notin N\cup N_B,$$

consequently μ-a.e.. It follows that for any $B\in\mathcal{B}(\mathbb{R}^m)$ the function $x\mapsto\nu_B(x)$ is μ-measurable and for $A\in\mathcal{B}(X)$ and $B\in\mathcal{B}(Y)$ we have

$$\nu(A\times B)=\int_A\nu_x(B)\,d\mu(x).$$

The other claims easily follow, by writing $f=f_+-f_-$ and approximating f_+ and f_- with simple functions. □

6.71 Remark. In Probability, the above leads to the definition of *conditional distribution*. Let X and Y be two random variables (i.e., two $\mathcal{E}$-measurable functions) on a probability space $(\Omega,\mathcal{E},P)$ and let P_X and P_Y be their distributions, respectively. Denote by $P_{X,Y}$ their joint distribution, which is a nonzero finite Radon measure in $X\times Y$. If $\nu:=P_{X,Y}$, then $\mu=\pi_{\#}\nu=P_X$. Then Theorem 6.70 yields

$$\int_{\mathbb{R}}\varphi(y)\,dP_X(y)=\int_{\mathbb{R}}\left(\int\varphi(y)\,d\nu_x(y)\right)dP_X(x).$$

For μ-a.e. x the finite Radon measure ν_x on $\mathbb{R}$ is called the *conditional distribution of Y with respect to X at x* and is usually denoted by $P_{(Y\,|\,X=x)}$ from which derives the formula

$$P(Y\in A)=\int P_{(Y\,|\,X=x)}(A)\,dP_X(x).$$

6.3 Hausdorff Measures

In this section we discuss Hausdorff measures in $\mathbb{R}^n$, that allow us to measure "k-dimensional sets in $\mathbb{R}^n$" similar to k-dimensional submanifolds in $\mathbb{R}^n$.

Given $s\in\mathbb{R}$, $s\geq 0$, we set

$$\omega_s:=\frac{\pi^{s/2}}{\Gamma(1+s/2)},$$


Sonderabdruck aus Mathematische Annalen. LXXV. Band. 3. Heft.
Druck und Verlag von B. G. Teubner in Leipzig.

Bemerkung über den Inhalt von Punktmengen.

Von

F. Hausdorff in Greifswald.

Bekanntlich haben E. Borel und H. Lebesgue das Problem der Inhaltsbestimmung von Punktmengen wenigstens eine Strecke weit *axiomatisch* zu behandeln versucht. Es soll jeder beschränkten Menge A des n-dimensionalen euklidischen Raumes E_n als Inhalt eine nichtnegative Zahl $f(A)$ unter folgenden Bedingungen zugeordnet werden:

(α) Der Einheitswürfel hat den Inhalt 1.

(β) Kongruente Mengen haben denselben Inhalt.

(γ) Es ist $f(A+B) = f(A) + f(B)$.

(δ) Es ist $f(A+B+C+\cdots) = f(A) + f(B) + f(C) + \cdots$ für eine beschränkte Summe von abzählbar vielen Summanden.*)

Die von Lebesgue gegebene *konstruktive* Inhaltsdefinition erfüllt zwar diese vier Forderungen, aber sie ordnet nicht allen (beschränkten) Mengen, sondern nur den „meßbaren" Inhalte zu. Das erste Beispiel einer im Lebesgueschen Sinne nicht meßbaren Menge stammt von G. Vitali**); übrigens hat das System der unmeßbaren Mengen dieselbe Mächtigkeit wie das der meßbaren und das aller Mengen. Die Frage bleibt also offen, ob das durch die Forderungen (α) bis (δ) gestellte Inhaltsproblem, *in der Ausdehnung auf alle beschränkten Mengen*, überhaupt lösbar ist oder nicht.

Man bemerkt sofort, daß eine Lösung des Inhaltsproblems für den E_{n+1} auch eine für den E_n liefern würde (indem man einer n-dimensionalen Menge A_n den $(n+1)$-dimensionalen zylindrischen Körper A_{n+1} mit der Basis A_n und der Höhe 1 zuordnet und $f_n(A_n) = f_{n+1}(A_{n+1})$ setzt);

*) Vgl. H. Lebesgue, Leçons sur l'intégration (Paris 1904), S. 103. Wir bezeichnen eine Summe von Mengen nur dann mit $A+B+\cdots$, wenn diese Mengen paarweise keinen Punkt gemein haben.

**) A. Schoenflies, Entwickelung der Mengenlehre (Leipzig und Berlin 1913), S. 374, wo außer weiteren Beispielen von E. B. van Vleck und Lebesgue auch das im Text folgende von mir mitgeteilt ist.

Figure 6.4. A poster for the celebration of Felix Hausdorff (1869–1942) in Bonn and the first page of one of his papers.

recalling that $\omega_s = \mathcal{L}^s(B(0,1))$ for integer s's, $B(0,1)$ being the unit s-dimensional ball in $\mathbb{R}^s$. We consider the set function

$$\alpha(E) := \frac{\omega_s}{2^s}(\operatorname{diam} E)^s, \qquad E \in \mathcal{P}(\mathbb{R}^n),$$

and construct starting from $(\mathcal{P}(\mathbb{R}^n), \alpha)$ a Borel measure $\mathcal{H}^s$ by means of Method II of construction, i.e., we define for any $\delta > 0$

$$\mathcal{H}^s_\delta(E) := \inf\Big\{ \frac{\omega_s}{2^s} \sum_{j=1}^{\infty} (\operatorname{diam} E_j)^s \;\Big|\; E \subset \cup_j E_j,\ \operatorname{diam}(E_j) \le \delta \Big\}$$

and set

$$\mathcal{H}^s(E) := \lim_{\delta \to 0^+} \mathcal{H}^s_\delta(E),$$

(notice that $\delta \to \mathcal{H}^s_\delta(E)$ is nondecreasing). It is not difficult to check the following:

(i) $\mathcal{H}^s_\delta(E) < +\infty$, $\delta > 0$, for any bounded set E.
(ii) $\mathcal{H}^s_\delta(E)$ is an outer measure.
(iii) In general, Borel sets are not measurable for $\mathcal{H}^s_\delta$, $\delta > 0$ (for instance, the line $y = 0$ in $\mathbb{R}^2$ is not $\mathcal{H}^1_\delta$-measurable for any $\delta > 0$).
(iv) $\mathcal{H}^0$ is the counting measure.
(v) $\mathcal{H}^s$ is not a Radon measure, since, in general, it is not finite on compact sets; for instance, if E has nonempty interior and $s < n$, then $\mathcal{H}^s(E) = +\infty$; for similar reasons, $\mathcal{H}^s$ in general is not σ-finite.

(vi) Since the definition of $\mathcal{H}^s_\delta(E)$ involves only the diameters of the sets of the covering of E, we may require that the covering is made by closed sets or convex and closed sets and even of open sets, since a closed set is the intersection of open sets with slightly larger diameters.

We may simplify even further and allow only coverings of E by balls

$$\mathcal{H}^s_{\delta,sph}(E) := \inf\Big\{\omega_s \sum_{j=1}^{\infty} r_j^s \;\Big|\; E \subset \cup_j B(x_j, r_j)\ r_j \le \delta\Big\}$$

and

$$\mathcal{H}^s_{sph}(E) := \lim_{\delta\to 0} \mathcal{H}^s_{\delta,sph}(E).$$

$\mathcal{H}^s_{sph}(E)$ is again a Borel-regular measure, called the *spherical Hausdorff measure*. However, if, for instance, E is an equilateral triangle in $\mathbb{R}^2$ with diameter less than δ, then $\mathcal{H}^s_{\delta,sph}(E) > \mathcal{H}^s_\delta(E)$, and, in general, one sees that $\mathcal{H}^s$ and $\mathcal{H}^s_{sph}$ are different, although they agree on "sufficiently regular sets". This way we have at least two different ways of measuring s-dimensional sets in $\mathbb{R}^n$, $s < n$. Finally, it is easily seen that these two measures that do not agree are in fact comparable: Since for every set $E \subset \mathbb{R}^n$ there exists a ball B that contains E and has diameter less than $2\,\mathrm{diam}(E)$, we see at once that $\mathcal{H}^s(E) \le \mathcal{H}^s_{sph}(E) \le 2^s\mathcal{H}^s(E)\ \forall E \in \mathcal{B}(\mathbb{R}^n)$.

The following proposition collects some of the elementary properties of Hausdorff measures $\mathcal{H}^s$.

6.72 Proposition. *The Hausdorff measure $\mathcal{H}^s$ is Borel-regular; furthermore, $\mathcal{H}^s \llcorner E$ is a Radon measure if $\mathcal{H}^s(E) < +\infty$. Moreover, the following hold:*

(i) *$\mathcal{H}^s$ is invariant under translation and rotation, and is positively homogeneous of degree s,*

$$\mathcal{H}^s(x + R(E)) = \mathcal{H}^s(E), \qquad \mathcal{H}^s(\lambda E) = \lambda^s\,\mathcal{H}^s(E)$$

if $R : \mathbb{R}^n \to \mathbb{R}^n$ is linear and $R^T R = \mathrm{Id}$.

(ii) *$\mathcal{H}^s = 0$ if $s > n$.*

(iii) *If $0 \le s < t \le n$, then $\mathcal{H}^t \le \mathcal{H}^s$. More precisely, $\mathcal{H}^t(E) > 0$ implies $\mathcal{H}^s(E) = +\infty$ and $\mathcal{H}^s(E) < \infty$ implies $\mathcal{H}^t(E) = 0$.*

(iv) *$\mathcal{H}^s(E) = 0$ if and only if $\mathcal{H}^s_\infty(E) = 0$.*

(v) *If $f : \mathbb{R}^n \to \mathbb{R}^k$ is a Lipschitz map, then $\mathcal{L}^k(f(E)) \le (Lip f)^k\,\mathcal{H}^k(E)$.*

Proof. By construction, see Proposition 6.2, $\mathcal{H}^s$ is a Borel-regular measure, while trivially $\mathcal{H}^s \llcorner E$ is a Radon measure if $\mathcal{H}^s(E) < \infty$.

(i) This simply follows from the definition of $\mathcal{H}^s$.

(ii) Let $s > n$ and let Q be a cube of side 1. Since Q can be covered by p^n cubes of side $1/p$, we infer for $\delta \ge \sqrt{n}/p$ that $\mathcal{H}^s_\delta(Q) \le \omega_s 2^{-s}(\sqrt{n}/p)^s p^n \le c(n,s)p^{n-s}$, hence for $p \to \infty$, we get $\mathcal{H}^s(Q) = 0$. It follows that $\mathcal{H}^s(\mathbb{R}^n) = 0$.

(iii) Since $(\operatorname{diam} E)^t \le (\operatorname{diam} E)^s \delta^{t-s}$ if $\operatorname{diam} E \le \delta$, we get $\mathcal{H}^t_\delta(E) \le \delta^{t-s}\mathcal{H}^s_\delta(E)$. The claim follows at once letting $\delta \to 0$.

(iv) Clearly $\mathcal{H}^s_\infty(E) \le \mathcal{H}^s(E)$, hence $\mathcal{H}^s_\infty(E) = 0$ if $\mathcal{H}^s(E) = 0$. Conversely, suppose that $\mathcal{H}^s_\infty(E) = 0$. For any $\epsilon \in]0,1[$ we can find a denumerable covering $\{E_j\}$ of E with $\operatorname{diam}(E_j) < 2r_j$ and $\omega_s \sum_{j=1}^\infty r_j^s < \epsilon$. Since the supremum of the r_j's can be estimated by $(\epsilon/\omega_s)^{1/s}$, we get

$$\mathcal{H}^s_{\delta(\epsilon)}(E) < \sigma \qquad \text{with } \delta(\epsilon) := 2(\epsilon/\omega_s)^{1/s}$$

and, letting $\epsilon \to 0$, we infer $\mathcal{H}^s(E) = 0$.

(v) From $\operatorname{diam}(f(E)) \le Lip(f)\operatorname{diam}(E)$ we get $\mathcal{H}^k(f(E)) \le Lip(f)\,\mathcal{H}^k(E)$. The claim then follows since $\mathcal{H}^k = \mathcal{L}^k$ in $\mathbb{R}^k$, see Theorem 6.75 below. □

6.73 Remark. In a less formal way, (iv) can be stated as follows: $\mathcal{H}^s(E) = 0$ *if and only if for any* $\epsilon > 0$ *there exists a sequence of open sets* $\{E_j\}$ *such that* $E \subset \cup_j E_j$ *and* $\sum_{j=1}^\infty (\operatorname{diam} E_j)^s < \epsilon$.

We notice that (v) is convenient in order to estimate from below the Hausdorff measure $\mathcal{H}^s$ of a set. To get an upper estimate, one usually estimates $\mathcal{H}^s_\delta(E)$ by suitably choosing a covering of E.

The conclusion (iii) of Proposition 6.72 implies that for $E \subset \mathbb{R}^n$, $\mathcal{H}^s(E)$ is finite and nonzero at most for a unique value of s, $0 \le s \le n$. This motivates the following.

6.74 Definition. *Let* $E \subset \mathbb{R}^n$. *The* Hausdorff dimension *of* E *is the number in* $[0,n]$ *given by*

$$\begin{aligned}\dim_{\mathcal{H}}(E) :&= \sup\Big\{s \,|\, \mathcal{H}^s(E) > 0\Big\} = \sup\Big\{s \,|\, \mathcal{H}^s(E) = +\infty\Big\}\\ &= \inf\Big\{s \,|\, \mathcal{H}^s(E) < \infty\Big\} = \inf\Big\{s \,|\, \mathcal{H}^s(E) = 0\Big\}.\end{aligned}$$

Of course, the four different ways of defining $\dim_{\mathcal{H}}(E)$ agree because of (iii) of Proposition 6.72 that also implies that if $0 < \mathcal{H}^s(E) < +\infty$, then $\dim_{\mathcal{H}}(E) = s$. Notice, however, that not necessarily $0 < \mathcal{H}^s(E) < +\infty$ if $\dim_{\mathcal{H}}(E) = s$.

6.75 Theorem. *In* $\mathbb{R}^n$ *we have* $\mathcal{H}^n = \mathcal{H}^n_\delta = \mathcal{L}^n \ \forall \delta > 0$.

Proof. We first prove that $\mathcal{L}^n(E) \le \mathcal{H}^n_\delta(E)$ for any $\delta > 0$. Of course, it is not restrictive to assume $\mathcal{H}^n_\delta(E) < +\infty$. Consider a generic covering $\{E_j\}$ of E with $\operatorname{diam}(E_j) \le \delta$. From the *isodiametric inequality*, see [GM4], we have

$$\mathcal{L}^n(E) \le \sum_{j=1}^\infty \mathcal{L}^n(E_j) \le \omega_n 2^{-n} \sum_{j=1}^\infty (\operatorname{diam} E_j)^n,$$

hence, by taking the infimum among all coverings, we get $\mathcal{L}^n(E) \le \mathcal{H}^n_\delta(E)\ \forall \delta > 0$.

Proving the opposite inequality is more complicated. First we notice that it suffices to prove that $\mathcal{H}^n(E) \le \mathcal{L}^n(E)$ for bounded sets E. In this case, as in (ii) of Proposition 6.72, we see that $\mathcal{H}^n$ is a Radon measure that has the doubling property by the n-homogeneity and the invariance under translations. Given $\delta > 0$ and $A \supset E$ with $\mathcal{L}^n(A) \le \mathcal{L}^n(E) + \epsilon$, Vitali's covering theorem, Theorem 6.64, yields a covering of E

made of closed and disjoint closed balls $B(x_i, r_i)$ with $x_i \in E$, $r_i \le \delta$, $B(x_i, r_i) \subset A$ and $\mathcal{H}^n(E \setminus \cup_i B(x_i, r_i)) = 0$. From the subadditivity of $\mathcal{H}^n_\delta$ we infer

$$\mathcal{H}^n_\delta(E) \le \sum_{i=1}^{\infty} \mathcal{H}^n_\delta(B(x_i, r_i)) \le \omega_n \sum_{i=1}^{\infty} r_i^n = \sum_{i=1}^{\infty} \mathcal{L}^n(B(x_i, r_i)) \le \mathcal{L}^n(A) \le \mathcal{L}^n(E) + \epsilon,$$

and, letting first $\delta \to 0$ and then $\epsilon \to 0$, we conclude the proof. □

We notice that in $\mathbb{R}^n$, we have, instead, $\mathcal{H}^s \ne \mathcal{H}^s_\delta$ if $s < n$ and $\delta > 0$.

6.3.1 Densities

a. Densities and Hausdorff measures

The Radon–Nikodym derivative of a Borel measure with respect to a Hausdorff measure is meaningless since $\mathcal{H}^s(B(x, r)) = +\infty \ \forall s < n \ \forall x$ and $\forall r > 0$. A suitable replacement is the so-called *s-dimensional density*.

Let λ be a Borel measure in $\mathbb{R}^n$ and $0 < s \le n$. The *upper s-dimensional density* and the *lower s-dimensional density* of λ at x are defined by

$$\theta^{s*}(\lambda, x) := \limsup_{r \to 0} \frac{\lambda(B(x, r))}{r^s}, \qquad \theta^s_*(\lambda, x) := \liminf_{r \to 0} \frac{\lambda(B(x, r))}{r^s},$$

respectively. If the two values agree, the common value

$$\theta^s(\lambda, x) := \lim_{r \to 0} \frac{\lambda(B(x, r))}{r^s}$$

is called the *s-density* of λ at x. In the previous definitions it is irrelevant whether the balls are open or closed. Again, arguing as in Proposition 6.41, we have the following.

6.76 Proposition. *Let $\lambda : \mathbb{R}^n \to \mathbb{R}$ be a Borel measure in $\mathbb{R}^n$. The functions $x \mapsto \theta^{*s}(\lambda, x)$ and $x \mapsto \theta^s_*(\lambda, x)$ are Borel functions.*

As for Radon measures, the following result is very useful in many instances that we shall not discuss here.

6.77 Theorem. *Let λ be a Borel-regular measure, $E \subset \mathbb{R}^n$ and $t \ge 0$.*

(i) *If $\theta^{*s}(\lambda, x) > t$ for all $x \in E$, then $t\,\mathcal{H}^s(E) \le \lambda(A)$ for any open set $A \supset E$. In particular, if λ is a Radon measure, then*

$$t\,\mathcal{H}^s(E) \le \lambda(E).$$

(ii) *If $\theta^{s*}(\lambda, x) \le t \ \forall x \in E$, then $\lambda(E) \le 2^s t\,\mathcal{H}^s(E)$.*

Proof. (i) We may and do assume $\lambda(A)$ finite and $t > 0$. Fix $\delta > 0$; the family

$$\mathcal{B} := \Big\{ B(x,r) \,\Big|\, B(x,r) \text{ closed},\ x \in E,\ 0 < r < \delta/2,\ B(x,r) \subset A,\ \lambda(B(x,r) > t\,\omega_s r^s \Big\}$$

finely covers E. Let $\mathcal{B}'$ be the subfamily selected according to Proposition 6.65. Since $\lambda(A \cap B(x,5r)) > 0$ for any $B = B(x,r) \in \mathcal{B}'$, $\mathcal{B}'$ is denumerable since each ball $B \in \mathcal{B}'$ has positive radius, and $\lambda(B) > 0$. Let $\mathcal{B}' = \{B_n\}$, $B_n := B(x_n, r_n)$, be the family of these balls. It follows that

$$\mathcal{H}^s_\delta(E) \le \omega_s \sum_{i=1}^{n} r_i^s + 5^s \omega_s \sum_{i=n+1}^{\infty} r_i^s. \tag{6.38}$$

On the other hand,

$$\omega_s \sum_{i=1}^{\infty} r_i^s \le \frac{1}{t} \sum_{i=1}^{\infty} \lambda(B(x_i, r_i)) \le \frac{1}{t} \lambda(\cup_i B(x_i, r_i)) \le \frac{1}{t} \lambda(A) < +\infty,$$

hence, letting $n \to \infty$ in (6.38),

$$\mathcal{H}^s(E) \le \omega_s \sum_{i=1}^{\infty} r_i^s \le \frac{1}{t} \lambda(A).$$

The last part of claim (i) follows since every Radon measure in $\mathbb{R}^n$ is outer-regular, see Proposition 6.2.

(ii) We decompose E as $E = \cup_k E_k$ with

$$E_k := \Big\{ x \in E \,\Big|\, \lambda(B(x,r)) \le t\omega_s r^s \ \forall r \in]0, 1/k[\Big\}.$$

Clearly, $E_k \subset E_{k+1}$, hence $\lambda(E_k) \to \lambda(E)$ (even if the E_k's are nonmeasurable). It suffices then to prove

$$\lambda(E_k) \le 2^s t \mathcal{H}^s(E_k) \qquad \forall k = 1, 2, \dots$$

For this, fix $\delta < 1/(2k)$ and consider a covering of E with sets $\{C_j\}$ with $\operatorname{diam}(C_j) < \delta$ and such that $C_j \cap E_k \neq \emptyset$. Let $B(x_j, r_j)$ be a ball with center in $E_k \cap C_j$ and radius $r_j := \operatorname{diam}(C_j)$ that contains C_j. Then $r_j < 1/k$ and

$$\begin{aligned} \lambda(E) \le \lambda(\cup_j C_j) &\le \sum_{j=1}^{\infty} \lambda(C_j) \le \sum_{j=1}^{\infty} \lambda(B(x_j, r_j)) \\ &\le t\,\omega_s \sum_{j=1}^{\infty} r_j^s = t\,\omega_s \sum_{j=1}^{\infty} (\operatorname{diam} C_j)^s. \end{aligned}$$

By taking the infimum on the coverings we finally get

$$\lambda(E) \le 2^s\, t\, \mathcal{H}^s_\delta(E) \le 2^s\, t\, \mathcal{H}^s(E).$$

□

6.78 Corollary. *Let λ be a Borel-regular measure in $\mathbb{R}^n$ and let $0 < s \le n$. Then the following hold:*

(i) *If λ is finite on $\mathbb{R}^n$, then $\theta^{*s}(\lambda, x) < +\infty$ for $\mathcal{H}^s$-a.e. $x \in \mathbb{R}^n$.*
(ii) *If λ is a Radon measure and $\lambda(E) = 0$, then $\theta^s(\lambda, x) = 0$ for $\mathcal{H}^s$-a.e. $x \in E$.*
(iii) *If E is λ-measurable and $\lambda(E) < \infty$, then $\theta^s(\lambda \llcorner E, x) = 0$ for $\mathcal{H}^s$-a.e. $x \in \mathbb{R}^n \setminus E$.*

(iv) *If E is λ-measurable and $\lambda(E) < \infty$, then $2^{-s} \le \theta^{*s}(E,x) \le 1$ for $\mathcal{H}^s$-a.e. $x \in E$.*

Proof. (i) If $A_t := \{x \,|\, \theta^{*s}(\lambda, x) > t\}$, we have $\mathcal{H}^s(A_t) \le \frac{1}{t}\lambda(\mathbb{R}^n)$ and the claim follows for $t \to +\infty$.

(ii) Since λ is Radon, then for each $t > 0$, for

$$E_t := \Big\{ x \in B \,|\, \theta^{*s}(\lambda, x) > t \Big\}$$

we have $\mathcal{H}^s(E_t) \le t\lambda(B_t) = 0$, hence $\{x \,|\, \theta^{*s}(\lambda, x) > 0\}$ has zero $\mathcal{H}^s$-measure.

(iii) Define $A_t := \{x \in E^c \,\big|\, \theta^{*s}(\lambda \llcorner E, x) > t\}$ for $t > 0$. It suffices to prove that $\mathcal{H}^s(A_t) = 0$ for all $t > 0$. Conclusion (i) of Theorem 6.77 yields

$$t\,\mathcal{H}^s(A_t) \le \lambda \llcorner E(A) = \lambda(E \cap A)$$

for all open sets A that contain E^c. Since E is λ-measurable and with finite measure then $\lambda \llcorner E$ is a finite Borel measure, hence outer-regular, see Theorem 6.1. Therefore,

$$t\,\mathcal{H}^s(A_t) \le \inf\Big\{ \lambda \llcorner E(A) \,|\, A \supset E^c,\ A \text{ open} \Big\} = \lambda \llcorner E(E^c) = 0.$$

(iv) Since E has finite measure, $\mathcal{H}^s \llcorner E$ is a Radon measure, hence for $C_t := \{x \in E \,|\, \theta^{*s}(E,x) > t\}$ we have $t\,\mathcal{H}^s(C_t) \le \mathcal{H}^s \llcorner E(C_t) = \mathcal{H}^s(C_t)$. It follows $\mathcal{H}^s(C_t) = 0$ if $t > 1$ since $\mathcal{H}^s(C_t) < \infty$.

Similarly, if $E_t := \{x \in E \,|\, \theta^{*s}(E,x) \le t\}$, then (ii) of Theorem 6.77 yields

$$\mathcal{H}^s(E_t) = \mathcal{H}^s \llcorner E(E_t) \le 2^s t \mathcal{H}^s(E_t),$$

i.e., $\mathcal{H}^s(E_t) = 0$ if $t < 2^{-s}$ since $\mathcal{H}^s(E_t) < +\infty$. □

6.79 ¶. Let $E \subset \mathbb{R}^n$. The *upper* and *lower s-densities* of E at a point x are defined by

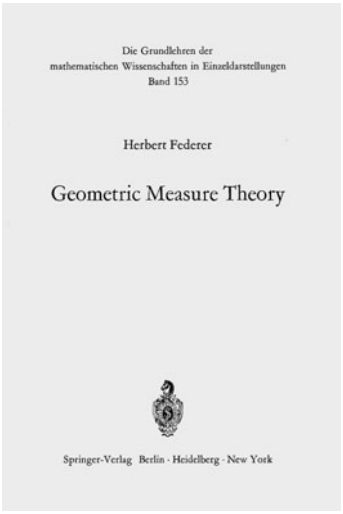
$$\theta^{*s}(E,x) := \theta^{*s}(\mathcal{H}^s \llcorner E, x), \qquad \theta^s_*(E,x) := \theta^s_*(\mathcal{H}^s \llcorner E, x),$$

respectively. Suppose that E is $\mathcal{H}^s$ measurable and $\mathcal{H}^s(E) < +\infty$. Show that

$$\theta^s(E,x) = 0 \qquad \text{for } \mathcal{H}^s\text{-a.e. } x \in \mathbb{R}^n \setminus E.$$

6.4 Area and Coarea Formulas

We conclude this chapter by proving two important formulas, the so-called *area* and *coarea* formulas. The formulas hold true for Lipschitz maps, but in the sequel we restrict ourselves to the case of C^1 maps.



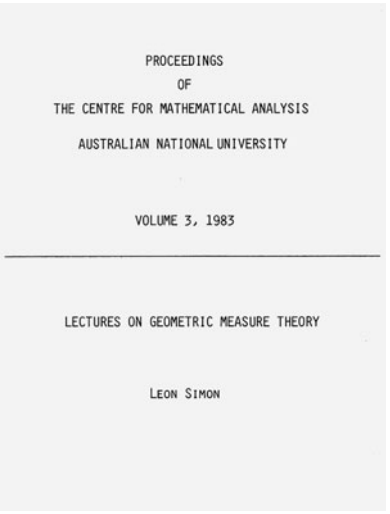


Figure 6.5. Frontispieces of two monographs dealing with geometric measure theory.

6.4.1 The area formula

6.80 Theorem (Area formula). *Let Ω be an open set in $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}^N$, $N \geq n \geq 1$, a map of class C^1 and $A \subset \Omega$ an $\mathcal{L}^n$-measurable set. Then the function*

$$y \in \mathbb{R}^N \to \mathcal{H}^0(A \cap f^{-1}(y))$$

that for every y counts the points that are mapped into y is $\mathcal{H}^n$-measurable and

$$\int_A J(\mathbf{D}f)(x)\,dx \;=\; \int_{\mathbb{R}^N} \mathcal{H}^0(A \cap f^{-1}(y))\,d\mathcal{H}^n(y), \tag{6.39}$$

where

$$J(\mathbf{D}f)(x) := \sqrt{\det(\mathbf{D}f(x)^T\,\mathbf{D}f(x))}. \tag{6.40}$$

In particular,

$$\mathcal{H}^n(f(A)) \;=\; \int_A J(\mathbf{D}f)(x)\,dx \tag{6.41}$$

if f is injective in A.

The function

$$y \to \mathcal{H}^0(A \cap f^{-1}(y)) =: N(f, A, y)$$

is called the *multiplicity function* or *Banach indicatrix*. Notice that

$$f(A) = \Big\{ y \in \mathbb{R}^N \;\Big|\; N(f, A, y) \neq 0 \Big\}.$$

Since $J(\mathbf{D}f)$ is continuous and nonnegative, the theorem states that both integrals in (6.39) exist, finite or infinite, and agree.

By approximating a nonnegative function u by simple functions and using Beppo Levi's theorem we readily infer the following.

6.81 Theorem (Formula of change of variables). *Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^N$, $n \le N$, Ω open, be a function of class C^1 and let $u : \mathbb{R}^n \to \mathbb{R}$ be either a $\mathcal{L}^n$-measurable and nonnegative function, or a function such that $|u|\, J(\mathbf{D}f)$ is summable. Then*

$$y \to \sum_{x \in f^{-1}(y)} u(x)$$

is $\mathcal{H}^n$-measurable and

$$\int_\Omega u(x)\, J(\mathbf{D}f)(x)\, dx = \int_{\mathbb{R}^N} \Big(\sum_{x \in f^{-1}(y)} u(x) \Big)\, d\mathcal{H}^n(y). \tag{6.42}$$

In particular, if $v : \mathbb{R}^N \to \mathbb{R}$ is $\mathcal{H}^n$-measurable and nonnegative, then

$$\int_A v(f(x))\, J(\mathbf{D}f)(x)\, dx = \int_{\mathbb{R}^N} v(y)\, N(f, A, y)\, d\mathcal{H}^n(y). \tag{6.43}$$

Proof. If $u = \chi_A$ is the characteristic function of a measurable set A, then (6.42) and (6.39) agree since

$$\sum_{x \in f^{-1}(y)} \chi_A(x) = \mathcal{H}^0(A \cap f^{-1}(y)).$$

By linearity, (6.42) holds for simple functions. If u is measurable and nonnegative, it is the limit of a nondecreasing sequence of simple functions; passing to the limit, we infer the result by Beppo Levi's theorem. If $|u|J(\mathbf{D}f)$ is summable, we decompose u as $u = u^+ - u_-$ and conclude by subtracting (6.42) for u^+ e u_-. □

Proof of Theorem 6.80. We shall prove the theorem when $A \subset\subset \Omega$. The general case follows easily by invading Ω with a sequence of compact sets $\Omega_k \subset\subset \Omega$, with $\Omega = \cup_k \Omega_k$, $\Omega_k \subset \Omega_{k+1}$, writing the formula for $A_k := A \cap \Omega_k$ and passing to the limit.

Let $A \subset\subset \Omega$. It is not restrictive to assume that f is Lipschitz in Ω, i.e., there is a constant L such that $|f(x) - f(y)| \le L\,|x - y|\ \forall x, y \in A$.

Step 1. $f(A)$ is $\mathcal{H}^n$-measurable. Let $\{K_h\}$ be a sequence of compact sets such that $K_h \subset A$, $K_h \subset K_{h+1}$ and $\mathcal{L}^n(A) = \mathcal{L}^n(\cup_h K_h)$. Since f is continuous, we have $f(\cup_h K_h) = \cup_h f(K_h)$, hence $f(\cup_h K_h)$ is a Borel set as denumerable union of compact sets. Now,

$$f(A) \;=\; f(\cup_h K_h) \cup f(A \setminus \cup_h K_h)$$

and

$$\mathcal{H}^n(f(A \setminus \bigcup_h K_h)) \le L^n\, \mathcal{H}^n(A \setminus \cup_h K_h) = L^n \mathcal{L}^n(A \setminus \cup_h K_h) = 0;$$

therefore, $f(A)$ is the union of a Borel set and of a set of zero $\mathcal{H}^n$ measure, hence $f(A)$ is $\mathcal{H}^n$-measurable.

Step 2. $y \to \mathcal{H}^0(A \cap f^{-1}(y))$ is $\mathcal{H}^n$-measurable and

$$\int_{\mathbb{R}^N} \mathcal{H}^0(A \cap f^{-1}(y))\, d\mathcal{H}^n(y) \;\le L^n\, \mathcal{L}^n(A). \tag{6.44}$$

This is proved by constructing a sequence $\{g_k\}$ of functions from $\mathbb{R}^N$ into $\mathbb{R}$ that are $\mathcal{H}^n$-measurable and that converge a.e. to the function $y \to \mathcal{H}^0(A \cap f^{-1}(y))$.

Decompose $\mathbb{R}^N$ into union of cubes $\{Q_i^k\}$ with disjoint interiors points, sides that are parallel to the coordinate axes, congruent and with sides of length 2^{-k}. Set

$$g_k(y) \;:=\; \sum_{i=1}^{\infty} \chi_{f(A \cap Q_i^k)}(y), \qquad y \in \mathbb{R}^N.$$

Since $f(A \cap Q_i^k)$ is $\mathcal{H}^n$-measurable for all i, see Step 1, g_k is $\mathcal{H}^n$-measurable for all k. Moreover, $g_k(y) \le g_{k+1}(y)\ \forall y$. We now remark the following:

1. If $f^{-1}(y) \cap A = \emptyset$, then $g_k(y) = 0 = \mathcal{H}^0(f^{-1}(y) \cap A)$ for all k.
2. If $\mathcal{H}^0(A \cap f^{-1}(y)) < \infty$, then for sufficiently large k, each point of $f^{-1}(y) \cap A$ belongs to exactly one of the cubes Q_i^k, hence
$$g_k(y) = \mathcal{H}^0(A \cap f^{-1}(y)).$$
3. If $\mathcal{H}^0(A \cap f^{-1}(y)) = +\infty$, trivially $\lim_{k\to\infty} g_k(y) = +\infty$.

In conclusion,
$$g_k(y) \ \uparrow \ \mathcal{H}^0(A \cap f^{-1}(y)) \qquad \forall y \in \mathbb{R}^N.$$
Finally, taking into account Beppo Levi's theorem,
$$\begin{aligned}\int_{\mathbb{R}^N} \mathcal{H}^0(A \cap f^{-1}(y))\, d\mathcal{H}^n(y) &= \lim_{k\to\infty} \int_{\mathbb{R}^N} g_k(y)\, d\mathcal{H}^n(y)\\ &= \lim_{k\to\infty} \sum_i \mathcal{H}^n(f(A \cap Q_i^k))\\ &\le \limsup_{k\to\infty} \sum_i L^n\, \mathcal{L}^n(A \cap Q_i^k)\\ &= L^n\, \mathcal{L}^n(A),\end{aligned}$$
i.e., (6.44).

Step 3. Let
$$B := \left\{ x \in \mathbb{R}^n \;\middle|\; J(\mathbf{D}f)(x) > 0 \right\}$$
and let $t > 1$. We now prove the following. *There exist a decomposition of B into disjoint Borel sets $\left\{B_j\right\}$ and injective linear maps $T_j : \mathbb{R}^n \to \mathbb{R}^N$ such that the following hold:*

(i) *$f_{|B_j}$ is injective.*
(ii) $\mathrm{Lip}(f_{|B_j} \circ T_j^{-1}) \le t$ and $\mathrm{Lip}(T_j \circ f_{B_j}^{-1}) \le t$.
(iii) *We have*
$$\frac{1}{t^n} |\det T_j| \ \le \ J(\mathbf{D}f)(x) \ \le \ t^n |\det T_j| \qquad \forall\, x \in E_j. \tag{6.45}$$

Since $\mathbf{D}f(x_0)$, $x_0 \in B$, has maximal rank, the implicit function theorem yields $r_0 > 0$ such that $f_{|B(x_0,r_0)}$ is injective; moreover, if $T_{x_0} : \mathbb{R}^n \to \mathbb{R}^N$ is the linear tangent map $x \mapsto \mathbf{D}f(x_0)(x)$, then T_{x_0} is invertible and $\mathbf{D}(f \circ T_{x_0}^{-1})(0) = \mathrm{Id}$, hence $\mathrm{Lip}(f_{|B(x_0,r_0)} \circ T_{x_0}^{-1}) \le t$, $\mathrm{Lip}(T_{x_0} \circ f^{-1}_{|B(x_0,r_0)}) \le t$, and (6.45) holds for all $x \in B(x_0, r_0)$ possibly for a smaller r_0.

In this way, we find a denumerable covering $\{B(x_i, r_i)\}$ of B for which (i), (ii) and (iii) hold with $B_j := B(x_j, r_j)$ and $T_j := T_{x_j}$. Then, by choosing inductively $B_1 := B(x_1, r_1)$, $T_1 := T_{x_1}$, $B_2 := B(x_2, r_2) \setminus B_1$, $T_2 := T_{x_2}$, $B_3 := B(x_3, r(x_3)) \setminus (B_1 \cup B_2)$ and so on, we find the requested decomposition.

Step 4. Let $A \subset B := \{x \in \mathbb{R}^n \mid J(\mathbf{D}f)(x) > 0\}$, and let $\{B_j\}$, T_j be as in Step 3. If we set $A_j := A \cap B_j$, we have
$$\begin{aligned}&\mathcal{H}^n(f(A_j)) = \mathcal{H}^n(f_{|B_j} \circ T_j^{-1} \circ T_j(A_j)) \le t^n \mathcal{H}^n(T_j(A_j)),\\ &\mathcal{H}^n(T_j(A_j)) = \mathcal{H}^n(T_j \circ f_{|B_j}^{-1} \circ f_{|B_j}(A_j)) \le t^n \mathcal{H}^n(f(A_j)),\\ &\frac{1}{t^n} J(T_j) \mathcal{L}^n(A_j) \le \int_{A_j} J(\mathbf{D}f)(x)\, dx \le t^n J(T_j) \mathcal{L}^n(A_j)\end{aligned}$$
and, because of the area formula for linear maps, Theorem 5.100, we have
$$\mathcal{H}^n(T_j(A_j)) = J(T_j)\, \mathcal{L}^n(A_j)$$
and, therefore,

$$\frac{1}{t^n} J(T_j)\mathcal{L}^n(A_j) \le \mathcal{H}^n(f(A_j)) \le t^n J(T_j)\mathcal{L}^n(A_j),$$
$$\frac{1}{t^n} J(T_j)\mathcal{L}^n(A_j) \le \int_{A_j} J(\mathbf{D}f)(x)\,dx \le t^n J(T_j)\mathcal{L}^n(A_j).$$

For $t \to 1$, we conclude

$$\int_{A_j} J(\mathbf{D}f)\,dx = \mathcal{H}^n(f(A_j))$$

and, summing in j, we infer the area formula, since

$$\mathcal{H}^0(A \cap f^{-1}(y)) = \sum_j \mathcal{H}^0(A_j \cap f^{-1}(y)) = \sum_j \chi_{f(A_j)}(y)$$

and

$$\int_{\mathbb{R}^N} \mathcal{H}^0(A \cap f^{-1}(y)) = \sum_j \mathcal{H}^n(f(A_j)).$$

Step 5. $A \subset \{x \,|\, J(\mathbf{D}f)(x) = 0\}$. In this case it suffices to prove that $\mathcal{H}^n(f(A)) = 0$. Given $\epsilon > 0$, let $g_\epsilon : \mathbb{R}^n \to \mathbb{R}^N \times \mathbb{R}^n$, $g_\epsilon(x) := (f(x), \epsilon x)$. Since $f(x)$ is the first factor of g_ϵ and the projection map $(x, y) \to x$ has Lipschitz constant 1,

$$\mathcal{H}^n(f(A)) \le \mathcal{H}^n(g_\epsilon(A)).$$

On the other hand, $J(\mathbf{D}g_\epsilon)^2 = \det(\mathbf{D}f^T\mathbf{D}f + \epsilon^2\,\mathrm{Id})$, hence $J(\mathbf{D}g_\epsilon)$ and, for $\epsilon \to 0$, $J(\mathbf{D}g_\epsilon) \to J(\mathbf{D}g_0) = J(\mathbf{D}f)$. Step 4 then yields

$$\mathcal{H}^n(g_\epsilon(A)) = \int_A J(\mathbf{D}g_\epsilon)(x)\,dx \to \int_A J(\mathbf{D}f)\,dx = 0.$$

Step 6. To conclude the proof, it suffices to decompose A as

$$A = \left(A \cap \left\{x \,\middle|\, J(\mathbf{D}f)(x) = 0\right\}\right) \bigcup \left(A \cap \left\{x \,\middle|\, J(\mathbf{D}f)(x) > 0\right\}\right)$$

and apply Step 4 and Step 5. □

6.4.2 The coarea formula

6.82 Theorem (Coarea formula). *Let $f : \Omega \to \mathbb{R}^N$, $\Omega \subset \mathbb{R}^n$ open, $N \le n$, be a function of class $C^1(\Omega)$, and let $A \subset \Omega$ be an $\mathcal{L}^n$-measurable function. Then for $\mathcal{L}^N$-a.e. $y \in \mathbb{R}^N$, the set $A \cap f^{-1}(y)$ is $\mathcal{H}^{n-N}$-measurable, the function $y \to \mathcal{H}^{n-N}(A \cap f^{-1}(y))$ is $\mathcal{L}^N$-measurable and*

$$\int_A J(\mathbf{D}f)(x)\,d\mathcal{L}^n(x) = \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap f^{-1}(y))\,d\mathcal{L}^N(y). \tag{6.46}$$

Here

$$J(\mathbf{D}f)(x) := \sqrt{\det(\mathbf{D}f(x)\mathbf{D}f(x)^T)}. \tag{6.47}$$

Again we notice that Theorem 6.82 states that if one of the two integrals in (6.46) exists, then the other exists too, and they are equal irrespective of their finiteness.

By an argument similar to that of Theorem 6.81, the following result can be proved starting from Theorem 6.82.

6.83 Theorem. *Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^N$, Ω open, $N \le n$, be a function of class C^1, and let $u : \Omega \to \mathbb{R}$ be a nonnegative and $\mathcal{L}^n$-measurable function or, alternatively, let $|u|\, J(\mathbf{D}f)$ be $\mathcal{L}^n$-summable. We have*

$$\int_{\mathbb{R}^n} u(x)\, J(\mathbf{D}f)(x)\, dx = \int_{\mathbb{R}^N} \left(\int_{f^{-1}(y)} u(x)\, d\mathcal{H}^{n-N}(x) \right) d\mathcal{L}^N(y). \quad (6.48)$$

Theorems 6.82 and 6.83 may be seen as a "curvilinear" extension of Fubini's theorem.

By applying (6.46) with $A = B := \{x \in \Omega \,|\, J(\mathbf{D}f)(x) = 0\}$, we infer

$$\mathcal{H}^{n-N}(B \cap f^{-1}(y)) = 0 \qquad \text{for } \mathcal{L}^N\text{-a.e. } y \in \mathbb{R}^N.$$

On the other hand, if $x \in B^c$, the implicit function theorem yields an open neighborhood U_x of x in $\mathbb{R}^n$ such that $U_x \cap f^{-1}(y)$ is an $\mathcal{H}^{n-N}$-submanifold of class C^1 of $\mathbb{R}^n$. We can therefore state the following.

6.84 Corollary (Sard-type theorem). *Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^N$, $n \ge N$, be a map of class $C^1(\Omega)$ and let $B := \{x \in \Omega \,|\, J(\mathbf{D}f)(x) = 0\} = \{x \,|\, \operatorname{Rank} \mathbf{D}f(x) < N\}$. For $\mathcal{L}^N$-a.e. $y \in \mathbb{R}^N$ we can decompose $f^{-1}(y)$ as*

$$f^{-1}(y) = (f^{-1}(y) \setminus B) \cup (f^{-1}(y) \cap B)$$

where $\mathcal{H}^{n-N}(B \cap f^{-1}(y)) = 0$ and $f^{-1}(y) \setminus B$ is a C^1 (n-N)-submanifold of $\mathbb{R}^n$.

Proof of Theorem 6.82. We divide the proof in eight steps.

Step 1. We prove the theorem for linear maps of maximal rank. Assume $f : \mathbb{R}^n \to \mathbb{R}^N$ is a linear map of rank N. From the polar decomposition formula, $f = (ff^T)^{1/2}U^*$, where $U^*U = \mathrm{Id}_{\mathbb{R}^N}$ and (ff^T) is an isomorphism. Therefore, if $\mathbb{R}^n = \mathbb{R}^N \times \mathbb{R}^{n-N}$ and $\pi : \mathbb{R}^n \to \mathbb{R}^N$ denotes the orthogonal projection onto $\mathbb{R}^N$, then $f = \sigma \circ \pi \circ R$, where $R : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal and $\sigma := (ff^T)^{1/2} : \mathbb{R}^N \to \mathbb{R}^N$ is an isomorphism. The invariance of the measure with respect to orthogonal transformations and Fubini's theorem then yield

$$\mathcal{H}^n(A) = \mathcal{H}^n(R(A)) = \int \mathcal{H}^{n-N}\Big(R(A) \cap \pi^{-1}(z)\Big)\, d\mathcal{L}^N(z).$$

Moreover, changing variables in the integral with $z = \sigma(y)$, we get $dy = |\det \sigma|\, dz$, consequently,

$$\int \mathcal{H}^{n-N}\Big(R(A) \cap \pi^{-1}(z)\Big)\, d\mathcal{L}^N(z) = \frac{1}{|\det \sigma|} \int \mathcal{H}^{n-N}\Big(R(A) \cap \pi^{-1}\sigma^{-1}(y)\Big)\, d\mathcal{L}^N(y)$$

and, since R is orthogonal and $R(A) \cap \pi^{-1}\sigma^{-}1(y)) = R(A \cap f^{-1}(y))$, for any $y \in \mathbb{R}^N$

$$\mathcal{H}^{n-N}\Big(R(A) \cap \pi^{-1}\sigma^{-1}(y)\Big) = \mathcal{H}^{n-N}\Big(A \cap f^{-1}(y)\Big),$$

that is,

$$|\det \sigma|\, \mathcal{H}^n(A) = \int \mathcal{H}^{n-N}\Big(A \cap f^{-1}(y)\Big)\, d\mathcal{L}^N(y).$$

The claim then follows, since from $RR^T = \mathrm{Id}$ and $\pi\pi^T = \mathrm{Id}_{\mathbb{R}^N}$, we have $J(\mathbf{D}f)^2 = \det(ff^T) = |\det \sigma|^2$.

We now prove the theorem for $f \in C^1(\Omega)$. As for the area formula, it suffices to prove the theorem for $A \subset\subset \Omega$, and we may assume f Lipschitz in A, i.e.,

$$|f(x) - f(y)| \le L|x - y| \qquad \forall x, y \in A$$

for some $L > 0$.

Step 2. $f^{-1}(y)$ *is closed in* Ω, hence a Borel set, consequently a $\mathcal{H}^{n-N}$-measurable set.

Step 3. $f(A)$ *is* $\mathcal{L}^N$*-measurable.* This may be proved as in Step 1 of the proof of Theorem 6.80.

Step 4. We now prove that

$$\int_{\mathbb{R}^N}^* \mathcal{H}^{n-N}(A \cap f^{-1}(y))\, d\mathcal{L}^N(y) \le \frac{\omega_{n-N}\,\omega_N}{\omega_n}\, L^N\, \mathcal{L}^n(A). \tag{6.49}$$

This proves the coarea formula when A is a null set. The symbol $\int^*$ deserves a definition. When $\varphi : \mathbb{R}^N \to \mathbb{R}$, we define

$$\int^* \varphi(x)\, d\mathcal{L}^N(x) := \inf\Big\{ \int h(x)\, dx \,\Big|\, h \text{ measurable},\ h \ge \varphi \Big\}.$$

Trivially, $\int^* \varphi\, dx \le \int^* g(x)\, dx$ if $\varphi \le g$ and $\int^* \varphi\, dx = \int \varphi\, dx$ if φ is $\mathcal{L}^N$-measurable.

For $j = 1, 2, \dots$ we choose a family of closed ball $\{B_i^j\}_i$ such that $A \subset \cup_{i=1}^\infty B_i^j$, $\operatorname{diam} B_i^j \le 1/j$, $B_i^j \subset \Omega$ and $\sum_{i=1}^\infty |B_i^j| \le |A| + 1/j$, and we set

$$g_i^j(y) := \omega_{n-N} \Big(\frac{\operatorname{diam} B_i^j}{2}\Big)^{n-N} \chi_{f(B_i^j)}(y).$$

According to Step 2, the functions g_i^j are $\mathcal{L}^N$-measurable. Moreover, for each j, by choosing as covering of $A \cap f^{-1}(y)$ the balls B_i^j that cover $f^{-1}(y)$, i.e., those for which $y \in f(B_i^j)$, from the definition of Hausdorff measure we get

$$\mathcal{H}^{n-N}_{1/j}(A \cap f^{-1}(y)) \le \omega_{n-N} \sum_{y \in f(B_i^j)} \Big(\frac{\operatorname{diam} B_i^j}{2}\Big)^{n-N} = \sum_{i=1}^\infty g_i^j(y).$$

Since the functions g_i^j are $\mathcal{L}^N$-measurable,

$$\begin{aligned} \int_{\mathbb{R}^N}^* \mathcal{H}^{n-N}(A \cap f^{-1}(y))\, dy &= \int_{\mathbb{R}^N}^* \liminf_{j\to\infty} \mathcal{H}^{n-N}_{1/j}(A \cap f^{-1}(y))\, dy \\ &\le \int_{\mathbb{R}^N}^* \liminf_{j\to\infty} \sum_{i=1}^\infty g_i^j\, dy = \int_{\mathbb{R}^N} \liminf_{j\to\infty} \sum_{i=1}^\infty g_i^j\, dy \end{aligned} \tag{6.50}$$

and, applying Fatou's lemma,

$$\begin{aligned} \int_{\mathbb{R}^N} \liminf_{j\to\infty} \sum_{i=1}^\infty g_i^j\, dy &\le \liminf_{j\to\infty} \sum_{i=1}^\infty \int_{\mathbb{R}^N} g_i^j\, dy \\ &= \liminf_{j\to\infty} \sum_{i=1}^\infty \omega_{n-N} \Big(\frac{\operatorname{diam} B_i^j}{2}\Big)^{n-N} \mathcal{H}^N(f(B_i^j)). \end{aligned} \tag{6.51}$$

On the other hand, by using the isodiametric inequality, see [GM4], we get

$$\mathcal{H}^N(f(B_j^i)) \le L^N \mathcal{L}^N(B_i^j) \le \omega_N \Big(\frac{\operatorname{diam}(B_i^j)}{2}\Big)^N,$$

hence, joining with (6.50) and (6.51), we conclude

$$\begin{aligned}\int_{\mathbb{R}^N}^* \mathcal{H}^{n-N}(A\cap f^{-1}(y))\,dy &\le L^N \liminf_{j\to\infty}\sum_{i=1}^{\infty}\omega_{n-N}\omega_N\Big(\frac{\operatorname{diam} B_i^j}{2}\Big)^n\\ &= L^N\frac{\omega_{n-N}\omega_N}{\omega_n}\liminf_{j\to\infty}\sum_{i=1}^{\infty}\mathcal{L}^n(B_i^j)\\ &\le L^N\frac{\omega_{n-N}\omega_N}{\omega_n}\mathcal{L}^n(A).\end{aligned}$$

Step 5. If A is $\mathcal{L}^n$-measurable, the map $y\to\mathcal{H}^{n-N}(A\cap f^{-1}(y))$ is $\mathcal{L}^n$-measurable.

(i) *If A is compact or open, then $y\to\mathcal{H}^{n-N}(A\cap f^{-1}(y))$ is a Borel, hence measurable, map.*
If A is compact and $y_h\to y$, then $A\cap f^{-1}(y)$ contains all limit points of $A\cap f^{-1}(y_h)$. Consequently, every open covering of $A\cap f^{-1}(y)$ necessarily covers $A\cap f^{-1}(y_h)$ for h sufficiently large, therefore, for any $\delta>0$

$$\limsup_{h\to\infty}\mathcal{H}^{n-N}_\delta(A\cap f^{-1}(y_h))\le\mathcal{H}^{n-N}_\delta(A\cap f^{-1}(y)).$$

It follows that for any $t\in\mathbb{R}$ the set

$$\Big\{y\in\mathbb{R}^n\;\Big|\;\mathcal{H}^{n-N}(A\cap f^{-1}(y))\ge t\Big\}$$

is closed, i.e., that the map $y\to\mathcal{H}^{n-N}_\delta(A\cap f^{-1}(y))$ is Borel, actually upper semicontinuous $\forall\delta>0$.
If A is open, then $A=\cup_h K_h$, $K_h\subset K_{h+1}$, K_h compact. The claim then follows from the compact case since

$$\mathcal{H}^{n-N}(A\cap f^{-1}(y))=\lim_{h\to\infty}\mathcal{H}^{n-N}(K_h\cap f^{-1}(y)).$$

(ii) *If A is an $\mathcal{L}^n$ null set, then $y\to\mathcal{H}^{n-N}(A\cap f^{-1}(y))=0$ for $\mathcal{L}^N$ a.e. $y\in\mathbb{R}^N$.*
Consider a sequence of open sets A_j such that $A\subset A_j$, $A_{j+1}\subset A_j$ $\forall j$ and $\mathcal{L}^n(A_j)<1/j$. From Step 3 and Step 4 we infer

$$\int^*\mathcal{H}^{n-N}(A\cap f^{-1}(y))\,dy\le\int^*\mathcal{H}^{n-N}(A_j\cap f^{-1}(y))\,dy\le C\,\mathcal{L}^n(A_j)\le\frac{C}{j}$$

and, for $j\to\infty$,

$$\int^*\mathcal{H}^{n-N}(A\cap f^{-1}(y))\,dy=0,$$

that is, $\mathcal{H}^{n-N}(A\cap f^{-1}(y))=0$ for $\mathcal{L}^N$-a.e. $y\in\mathbb{R}^N$.

(iii) *If A is $\mathcal{L}^n$-measurabile, the map $y\to\mathcal{H}^{n-N}(A\cap f^{-1}(y))$ is $\mathcal{L}^N$-measurable.* It suffices to decompose A as an at most denumerable union of compact sets and of $\mathcal{L}^n$-null sets and apply (i) and (ii) to each term of the decomposition.

Step 6. We prove the coarea formula when $A\subset B$ where

$$B:=\Big\{x\in\mathbb{R}^n\;\Big|\;J(\mathbf{D}f)(x)>0\Big\}.$$

Since A is foliated by the smooth submanifods $f^{-1}(y)$, we try to reduce to the coarea formula for linear maps.

Let $I(n-N,n)$ denote the set of $(n-N)$-tuples of increasing numbers between 1 and n. For each $\lambda=(\lambda_1,\lambda_2,\dots,\lambda_{n-N})\in I(n-N,n)$ let π_λ be the orthogonal projection onto the $n-N$-coordinate plane of coordinates $x^{\lambda_1},\dots,x^{\lambda_{n-N}}$,

$$\pi_\lambda(x^1,\dots,x^n)=(x^{\lambda^1},\dots,x^{\lambda^{n-N}}).$$

Since $Df(x_0)$ has maximal rank, for every $x_0\in B$ there exists $\lambda\in I(n-N,n)$ such that $T_{x_0}:=(\mathbf{D}f(x_0),\mathbf{D}\pi_\lambda)^T$ is invertible. According to the local invertibility

theorem, see [GM4], there is an open neighborhood U_{x_0} of x_0, such that the function $h_\lambda : U_{x_0} \to \mathbb{R}^n$ given by $h_\lambda(x) := (f(x), \pi_\lambda(x))$ is a diffeomorphism with open image, consequently, $g_\lambda := T_{x_0}^{-1} \circ h_\lambda : U_{x_0} \to \mathbb{R}^n$ is a diffeomorphism. Moreover, $\mathbf{D}g_\lambda(x_0) = \mathrm{Id}$ since $Dh_\lambda(x_0) = T_{x_0}$. Therefore, for $t > 1$,

$$\mathrm{Lip}(g_\lambda) \le t, \qquad \mathrm{Lip}(g_\lambda^{-1}) \le t$$

and

$$t^{-N} \le \frac{J(\mathbf{D}f)(x)}{J(\mathbf{D}f)(x_0)J(\mathbf{D}g_\lambda)(x)} \le t^N \qquad \forall x \in U_{x_0},$$

possibly taking a smaller neighborhood U_{x_0}.

Therefore, see Step 4 of the proof of Theorem 6.80, for any $t > 1$ there exist a disjoint family of Borel sets $\{B_k\}$, orthogonal projections π_k onto $n - N$ dimensional coordinate planes and linear isomorphisms $T_k : \mathbb{R}^n \to \mathbb{R}^n$ of the form $(L_k(x), \pi_k(x))$ with $L_k : \mathbb{R}^n \to \mathbb{R}^N$ of maximal rank such that the following hold:

(i) $B = \cup_k B_k$, where $\{B_k\}$ is a family of Borel sets,

(ii) for any k the map $x \to h_k(x) := (f(x)), \pi_k(x))$ is a diffeomorphism in a neighborhood of B_k.

(iii) For any k the map $g_k := T_k^{-1} \circ h_{\lambda_k}$ is a diffeomorphism in an open neighborhood of B_k with $\mathrm{Lip}(g_k) \le t$, $\mathrm{Lip}(g_k^{-1}) \le t$ and

$$t^{-N} J(L_k)J(\mathbf{D}g_k)(x) \le J(\mathbf{D}f)(x) \le t^N J(L_k)J(\mathbf{D}g_k)(x) \qquad \forall x \in B_k. \quad (6.52)$$

Set

$$A_k \ := \ A \cap B_k,$$

From the area formula and (6.52), trivially,

$$t^{-N} J(L_k)\, \mathcal{L}^n(g_k(A_k)) \le \int_{A_k} J(\mathbf{D}f)(x)\, dx \le t^N J(L_k)\, \mathcal{L}^n(g_k(A_k)), \quad (6.53)$$

while, since $f(x) = L_k \circ g_k(x)$, we infer by Step 1 that

$$\begin{aligned} J(L_k)\, \mathcal{L}^n(g_k(A_k)) &= \int_{\mathbb{R}^N} \mathcal{H}^{n-N}\Big(g_k(A_k) \cap L_k^{-1}(y)\Big)\, d\mathcal{L}^{n-N}(y) \\ &= \int_{\mathbb{R}^N} \mathcal{H}^{n-N}\Big(g_k(A_k \cap f^{-1}(y))\Big)\, d\mathcal{L}^{n-N}(y) \end{aligned} \quad (6.54)$$

since $g_k(A_k) \cap L_k^{-1}(y) = g_k(A_k \cap f^{-1}(y))$. On the other hand, for any $y \in \mathbb{R}^N$

$$\begin{aligned} t^{N-n}\mathcal{H}^{n-N}\Big(A_k \cap f^{-1}(y)\Big) &\le \mathcal{H}^{n-N}\Big(g(A_k \cap f^{-1}(y))\Big) \\ &\le t^{n-N}\mathcal{H}^{n-N}\Big(A_k \cap f^{-1}(y)\Big), \end{aligned} \quad (6.55)$$

since $\mathrm{Lip}(g_k),\ \mathrm{Lip}(g_k^{-1}) \le t$. By integrating (6.55) in y with respect to $\mathcal{L}^N$, we then infer

$$\begin{aligned} t^{N-n} \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A_k \cap f^{-1}(y))\, dy &\le J(L_k) \int_{A_k} J(\mathbf{D}g_k)(x)\, dx \\ &\le t^{n-N} \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A_k \cap f^{-1}(y))\, dy. \end{aligned} \quad (6.56)$$

The proof is then concluded from (6.53) and (6.56) if we sum in k and let $t \to 1$.

Step 7. We now prove the coarea formula for sets $A \subset B^c := \{x \,|\, J(\mathbf{D}f)(x) = 0\}$. For $\epsilon > 0$ set

$$\begin{aligned} &g : \mathbb{R}^n \times \mathbb{R}^N \to \mathbb{R}^N, \qquad g(x, w) := f(x) + \epsilon w, \quad x \in \mathbb{R}^n,\, w \in \mathbb{R}^N \\ &\widehat{\pi} : \mathbb{R}^n \times \mathbb{R}^N \to \mathbb{R}^N, \qquad \widehat{\pi}(x, w) := w, \quad x \in \mathbb{R}^n,\, w \in \mathbb{R}^N \end{aligned}$$

so that $\mathbf{D}g(x,w) = \mathbf{D}f + \epsilon\,\mathrm{Id}$ and $J(\mathbf{D}g) \le C\epsilon$. Since for any $w \in \mathbb{R}^N$

$$\int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap f^{-1}(y))\,dy = \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap f^{-1}(y - \epsilon w))\,dy,$$

we infer

$$\int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap f^{-1}(y))\,dy = \frac{1}{\omega_N} \int_{B(0,1)} dw \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap f^{-1}(y - \epsilon w))\,dy. \quad (6.57)$$

Next, we notice that if $C := A \times B(0,1)$, we have

$$C \cap g^{-1}(y) \cap \widehat{\pi}^{-1}(w) = \begin{cases} \emptyset & \text{if } w \notin B(0,1), \\ (A \cap f^{-1}(y - \epsilon w)) \times \{w\} & \text{if } w \in B(0,1) \end{cases} \quad (6.58)$$

for all y, $w \in \mathbb{R}^N$. In fact, $(x,z) \in C \cap g^{-1}(y) \cap \widehat{\pi}^{-1}(w)$ if and only if

$$x \in A, \qquad z \in B(0,1), \qquad f(x) + \epsilon z = y, \qquad z = w,$$

i.e., if and only if

$$x \in A, \qquad z = w \in B(0,1), \qquad f(x) = y - \epsilon z,$$

or if and only if

$$w \in B(0,1), \qquad (x,z) \in (A \cap f^{-1}(y - \epsilon w)) \times \{w\}.$$

Returning to (6.57), we now use Fubini's theorem and Step 5 to get

$$\begin{aligned} \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap f^{-1}(y))\,dy &= \frac{1}{\omega_N} \int_{\mathbb{R}^N} dw \Big(\int_{\mathbb{R}^N} \mathcal{H}^{n-N}(C \cap g^{-1}(y) \cap \widehat{\pi}^{-1}(w))\,dy \Big) \\ &= \int_{\mathbb{R}^N} \mathcal{H}^n(C \cap g^{-1}(y))\,dy = \int_C J(\mathbf{D}g)\,dx\,dz \\ &\le \sup_C J(\mathbf{D}g) \mathcal{H}^{n+N}(C) \le C\,\epsilon\,\mathcal{H}^n(A). \end{aligned}$$

For $\epsilon \to 0$ we conclude that

$$\int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap f^{-1}(y))\,dy = 0 = \int_A J(\mathbf{D}f)\,dx.$$

Step 8. In the general case, we write $A = (A \cap B) \cup (A \cap B^c)$, where $B := \{x \mid J(\mathbf{D}f)(x) > 0\}$. Applying Step 6 to $A \cap B$ and Step 7 to $A \cap B^c$, we conclude the proof. □

6.5 Exercises

6.85 ¶ Stereographic projection. Let $S^n = \{(x,z) \in \mathbb{R}^n \times \mathbb{R} \mid |x|^2 + z^2 = 1\}$ be the unit sphere in $\mathbb{R}^{n+1}$. The *stereographic projection* of the sphere onto $\mathbb{R}^n$ from the South Pole $P_S = ((0,\dots,0),1)$ is the map $\sigma : S^n \to \mathbb{R}^n$ given by $\sigma((x,z)) := x/(1+z)$, $(x,z) \in S^n$. Show that the inverse of the stereographic projection is the map $u : \mathbb{R}^n \to S^n \subset \mathbb{R}^{n+1}$ given by

$$u(x) := \Big(\frac{2}{1+|x|^2} x, \frac{1-|x|^2}{1+|x|^2} \Big), \qquad x \in \mathbb{R}^n.$$

If $A_1(x), \dots, A_n(x)$ denote the columns of $\mathbf{D}u(x)$, show that $A_i(x) \bullet A_j(x) = \lambda^2(x)\delta_{ij}$. Infer that

$$J_u^2(x) = \frac{1}{n^n}|\mathbf{D}u|^{2n} = \frac{1}{n^n}\frac{1}{(1+|x|^2)^n}$$

and conclude that

$$\mathcal{H}^n(S^n) = \frac{1}{n^{n/2}}\int_{\mathbb{R}^n}\frac{1}{(1+|x|^2)^n}\,dx = \frac{1}{n^{n/2}}\int_{\mathbb{R}^n}|\mathbf{D}u|^n\,dx.$$

6.86 ¶ Kantorovich's inequality. Let μ be a probability measure in $[0,1]$, let $f : [0,1] \to \mathbb{R}$ be a continuous function and let $m := \inf_{[0,1]} f(x)$ and $M := \sup_{[0,1]}$. Show that if

$$0 < m \le M < +\infty,$$

then

$$\Big(\int_0^1 f\,d\mu\Big)\Big(\int_0^1 \frac{1}{f}\,d\mu\Big) \le \frac{(m+M)^2}{4\,m\,M}$$

with equality if and only if μ is concentrated on the sets $\{x \,|\, f(x) = m\}$ and $\{x \,|\, f(x) = M\}$ with

$$\mu\Big(\{x \,|\, f(x) = m\}\Big) = \mu\Big(\{x \,|\, f(x) = M\}\Big) = \frac{1}{2}.$$

[*Hint.* Notice that

$$\sqrt{\Big(\int_0^1 f\,d\mu\Big)\Big(\int_0^1 \frac{1}{f}\,d\mu\Big)} \le \frac{1}{2}\int_0^1 \Big(\lambda f + \frac{1}{\lambda f}\Big)\,d\mu$$

for all $\lambda \in [0,1]$.]

A. Mathematicians and Other Scientists

André-Marie Ampère (1775–1836)
Cesare Arzelà (1847–1912)
Eugenio Beltrami (1835–1899)
Johann Bernoulli (1667–1748)
Sergi Bernstein (1880–1968)
Abram Besicovitch (1891–1970)
Enrico Betti (1823–1892)
Jacques Binet (1786–1856)
Jean-Baptiste Biot (1774–1862)
George Birkhoff (1884–1944)
T. Bonnesen
Emile Borel (1871–1956)
Luitzen E. J. Brouwer (1881–1966)
Georg Cantor (1845–1918)
Constantin Carathéodory (1873–1950)
Elie Cartan (1869–1951)
Bonaventura Cavalieri (1598–1647)
Pafnuty Chebycev (1821–1894)
Rudolf Clausius (1822–1888)
Charles Coulomb (1736–1806)
Antoine Cournot (1801–1877)
Georges de Rham (1903–1990)
Paul Dirac (1902–1984)
Peter Lejeune Dirichlet (1805–1859)
Paul du Bois–Reymond (1831–1889)
Dimitri Egorov (1869–1931)
Leonhard Euler (1707–1783)
Michael Faraday (1791–1867)
Gyula Farkas (1847–1930)
Werner Fenchel (1905–1986)
Joseph Fourier (1768–1830)
Ivar Fredholm (1866–1927)
Guido Fubini (1879–1943)
Carl Friedrich Gauss (1777–1855)
J. Willard Gibbs (1839–1903)
Jacques Hadamard (1865–1963)
William R. Hamilton (1805–1865)
Godfrey H. Hardy (1877–1947)
Felix Hausdorff (1869–1942)
Hermann von Helmholtz (1821–1894)
Heinrich Hertz (1857–1894)
David Hilbert (1862–1943)
William Hodge (1903–1975)
Otto Hölder (1859–1937)
Robert Hooke (1635–1703)
Adolf Hurwitz (1859–1919)
Christiaan Huygens (1629–1695)
Carl Jacobi (1804–1851)
Jean Pierre Kahane (1926–)
Erich Kähler (1906–2000)
Leonid Kantorovich (1912–1986)
Yitzhak Katznelson (1934–)
Harold Kuhn (1925–)
Joseph-Louis Lagrange (1736–1813)
Henri Lebesgue (1875–1941)
Adrien-Marie Legendre (1752–1833)
Gottfried W. von Leibniz (1646–1716)
Beppo Levi (1875–1962)
Tullio Levi–Civita (1873–1941)
Rudolf Lipschitz (1832–1903)
John E. Littlewood (1885–1977)
Hendrik Lorentz (1853–1928)
Nikolai Lusin (1883–1950)
Andrei Markov (1856–1922)
James Clerk Maxwell (1831–1879)
Adolph Mayer (1839–1903)
Hermann Minkowski (1864–1909)
Gaspard Monge (1746–1818)
Oskar Morgenstern (1902–1976)
Charles Morrey (1907–1984)
Harald Marston Morse (1892–1977)
John Nash (1928–)
Sir Isaac Newton (1643–1727)
Otto Nikodým (1887–1974)
Emmy Noether (1882–1935)
Marc-Antoine Parseval (1755–1836)
Gabrio Piola (1794–1850)
Siméon Poisson (1781–1840)
John Poynting (1852–1914)
Johann Radon (1887–1956)
Lord William Strutt Rayleigh (1842–1919)
Georg F. Bernhard Riemann (1826–1866)
Felix Savart (1791–1841)
Erwin Schrödinger (1887–1961)
Hermann Schwarz (1843–1921)
Sergei Sobolev (1908–1989)
Robert Solovay (1938–)

M. Giaquinta and G. Modica, *Mathematical Analysis, Foundations and Advanced Techniques for Functions of Several Variables*, DOI 10.1007/978-0-8176-8310-8,

Thomas Jan Stieltjes (1856–1894)
Brook Taylor (1685–1731)
Leonida Tonelli (1885–1946)
Albert Tucker (1905–1995)
Charles de la Vallée–Poussin (1866–1962)
Giuseppe Vitali (1875–1932)
John von Neumann (1903–1957)
Karl Weierstrass (1815–1897)
Hermann Weyl (1885–1955)
Hassler Whitney (1907–1989)
Ernst Zermelo (1871–1951)

There exist many web sites dedicated to the history of mathematics, we mention, e.g., `http://www-history.mcs.st-and.ac.uk/~history`.

B. Bibliographical Notes

We collect here a few suggestions for the readers interested in delving deeper into some of the topics treated in this volume. Of course, the list could be either longer or shorter. It reflects the taste and the knowledge of the authors.

General references:

- A. Friedman, *Foundations of Modern Analysis*, Dover, New York, 1982.
- E. Hewitt, K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, New York, 1965.
- L .Royden, *Real Analysis*, Macmillan, New York, 1988.
- W. Rudin, *Real and Complex Analysis*, McGraw–Hill, New York, 1986.

About *partial differential equations* and *calculus of variations*:

- V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, 1978.
- G. A. Bliss, *Lectures on the Calculus of Variations*, The University of Chicago Press, Chicago, 1946.
- G. Buttazzo, M. Giaquinta, S. Hildebrandt, *One–dimensional Variational Problems. An Introduction*, Clarendon Press, Oxford, 1998.
- R. Courant, *Dirichlet's Principle, Conformal Mapping, and Minimal Surfaces*, Springer–Verlag, New York, 1977.
- R. Courant, D. Hilbert, *Methods of Mathematical Physics*, 2 vols., Wiley, New York, 1953–1962.
- L. C. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, RI, 1998.
- M. Giaquinta, *Multiple Integrals in the Calculus of Variations and Nonlinear Elliptic Systems*, Princeton University Press, Princeton, NJ, 1983.
- M. Giaquinta, *Introduction to Regularity Theory for Nonlinear Elliptic Systems*, Birkäuser, Basel, 1993.
- M. Giaquinta, S. Hildebrandt, *Calculus of Variations*, 2 vols., Springer-Verlag, Berlin, 1996.
- M. Giaquinta, G. Modica, J. Souček, *Cartesian Currents in the Calculus of Variations*, 2 vols., Springer-Verlag, Berlin, 1998.
- H. Hofer, E. Zehnder, *Symplectic Invariants and Hamiltonian Dynamics*, Birkäuser, Basel, 1994.
- J. Jost, X. Li–Jost, *Calculus of Variations*, Cambridge University Press, Cambridge, 1999.
- E. H. Lieb, *Analysis*, Americam Mathematical Society, Providence, RI, 1997.
- H. F. Weinberger, *Partial Differential Equations*, Wiley, New York, 1965.

About *convexity* and its applications:

- E. F. Beckenbach, R. Bellman, *Inequalities*, Springer-Verlag, New York, 1965.
- T. Bonnesen, W. Frenchel, *Theorie der Konvexen Körper*, Springer-Verlag, New York, 1965.

- I. Ekeland, R. Temam, *Analyse Convexe et Problèmes Variationnels*, Dunod, Paris, 1974.
- J. Franklin, *Methods of Mathematical Economics*, Springer-Verlag, New York, 1980.
- L. Hörmander, *Notion of Convexity*, Birkäuser, Boston, 1994.
- S. Karlin, *Mathematical Methods and Theory in Games, Programming and Economics*, 2 vols., Addison–Wesley, Reading, MA, 1959.
- H. W. Kuhn, *Lectures on the Theory of Games*, Princeton University Press, Princeton, NJ, 2003.
- A. W. Roberts, D. E. Varberg, *Convex functions*, Academic Press, New York, 1973.
- R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

About *differential forms*:

- R. Bott, L. W .Hu, *Differential Forms in Algebraic Topology*, Springer-Verlag, New York, 1982.
- M. do Carmo, *Differential Forms and Applications*, Springer-Verlag, New York, 1994.
- H. F. Flanders, *Differential Forms with Applications to the Physical Sciences*, Dover, New York, 1989.
- M. Spivak, *Calculus on Manifolds*, Benjamin, Reading, MA, 1965.

About *measure* and *geometric measure theory*:

- L. C. Evans, R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, FL, 1992.
- H. Federer, *Geometric Measure Theory*, Springer-Verlag, New York, 1969.
- P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces*, Cambridge University Press, Cambridge, 1995.
- M. E. Munroe, *Measure and Integration*, Addison–Wesley, Reading, MA, 1971.
- S. Saks, *Théorie de l'Intégrale*, Monografje Matematyczne, Warszawa, 1933.
- L. Simon, *Lectures on Geometric Measure Theory*, The Centre for Mathematical Analysis, Canberra, 1983.
- R. L. Wheeden, A. Zygmund, *Measure and Integral*, Marcel Dekker, New York, 1977.

C. Index