GEOMETRIC PROGRAMMING

Geometric programming originated in 1961 with Zener's discovery (1–5) of an ingenious method for designing equipment at minimum total cost—a method that is applicable when the component capital costs and operating costs can be expressed in terms of the design variables via a certain type of generalized polynomial (one whose exponents need not be positive integers). Unlike competing analytical methods, which require the solution of a system of *nonlinear* equations derived from the differential calculus, this method requires the solution of a system of *linear* equations derived from both the differential calculus and certain ingenious transformations. Unlike competing numerical methods, which minimize the total cost by either direct search or steepest descent or the Newton–Raphson method (or one of their numerous descendants), this method provides formulae that show how the minimum total cost and associated optimal design depend on the design parameters (such as unit material costs and power costs, which are determined externally and hence cannot be set by the designer).

In 1962, Duffin (6,7) significantly enlarged the class of generalized polynomials that can be minimized with this method, by introducing an ingenious analog of the *dual variational principles* that characterize the *network duality* originating from the two Kirchhoff laws and Ohm's laws. In 1964, Duffin and Peterson (8,9) extended this *geometric programming duality* and associated methodology to the minimization of generalized polynomials subject to inequality constraints on other generalized polynomials. In essence, that development provided a nonlinear generalization of *linear programming duality*—one that is frequently applicable to the optimal design of sophisticated equipment and complicated systems (such as motors, transformers, generators, heat exchangers, power plants, and their associated systems).

In 1967, Duffin et al. (10) published the first book on geometric programming, which included additional generalizations of the mathematical methodology as well as illustrative applications to a variety of realistic optimization problems in engineering design. In 1971, Zener (11) published a short introductory book to make geometric programming more accessible to design engineers. More recent developments and publications are discussed in later sections.

AN ELEMENTARY EXAMPLE: THE OPTIMAL DESIGN OF A POWER LINE

Suppose the capital cost is simply proportional to the volume of the line, namely the product of its desired length L (a design parameter) and its cross-sectional area t (an independent design variable or decision variable). In particular, then, the capital cost is CLt, where C (a design parameter) is the cost per unit volume of the material making up the line. Also, suppose the operating cost is simply proportional to the power loss, which is known to be proportional to both L and the line resistivity R (a design parameter) as well as to the square of the carried current I (a design parameter) while being inversely proportional to t. In particular, then, the operating cost is $DLRI^2/t$, where the proportionality constant D (a design parameter) is determined from the predicted lifetime of the line as well as the present and future unit power costs (via standard accounting procedures for expressing the sum of all such costs as a present value determined by interest rates). In summary, the problem is to find the cross-sectional area $t > 0$ that minimizes the total cost

$$P(t) = c_1 t^1 + c_2 t^{-1} \text{ for given coefficients } c_1 = CL \text{ and } c_2 = DLRI^2 \tag{1}$$

Such an optimal cross-sectional area t^* exists, because the positivity of the coefficients c_1 and c_2 clearly implies that, for $t > 0$, the continuous function $P(t) > 0$ and $P(t) \to +\infty$ as either $t \to 0^+$ or $t \to +\infty$.

GENERALIZED POLYNOMIALS

The *objective function* $P(t)$ defined by Eq. (1) is an example of a *generalized polynomial* $P(t) = \sum_{i=1}^{n} T_i$—namely, a sum of terms $T_i = c_i \prod_{j=1}^{m} t_j^{a_{ij}}$, each of which is a given coefficient c_i (usually determined by design parameters) multiplied into a product $\prod_{j=1}^{m} t_j^{a_{ij}}$ of the independent design variables t_j raised to appropriate powers a_{ij}, termed *exponents*. In the *single-variable* generalized polynomial Eq. (1), the independent design variable t is the scalar variable t_1, while the exponents $a_{11} = 1$ and $a_{21} = -1$. In the *multivariable* generalized polynomial $P(t) = c_1 t_1^{-1} t_2^2 + c_2 t_1^{-1/2} t_2^{-3}$, the independent design variable t is the vector variable (t_1, t_2), while the exponents $a_{11} = -1$, $a_{12} = 2$, $a_{21} = -\frac{1}{2}$, and $a_{22} = -3$. Since noninteger exponents a_{ij} are mathematically permissible and are, in fact, needed in many applications, the natural domain of a generalized polynomial $P(t)$ is normally $t > 0$ (meaning that each component t_j of t is positive)—so that $t_1^{-1/2}$, for example, is defined and real-valued.

Posynomials and Signomials

If each coefficient c_i is positive, each term $T_i = c_i \prod_{j=1}^{m} t_j^{a_{ij}}$ in $P(t)$ is clearly positive, and hence so is each value $P(t) = T_i$. Such generalized polynomials $P(t)$, including Eq. (1), are termed *posynomials* and are reasonably easy to minimize via geometric programming. Generalized polynomials $P(t)$ that can be expressed as the difference of two posynomials, such as our second example $P(t) = c_1 t_1^{-1} t_2^2 + c_2 t_1^{1/2} t_2^{-3}$ when $c_1 > 0$ but $c_2 < 0$, are termed *signomials* and are usually more difficult to minimize. Moreover, the maximization of either posynomials or signomials is usually more difficult than the minimization of posynomials.

Posymonomials and Modeling

It is clear that most, if not all, equipment-component volumes are posynomial or signomial functions of their various geometric dimensions—namely, some of the independent design variables t_j. Moreover, many physical and economic relations have been expressed in terms of single-term posynomials called *posymonomials*. Such posymonomials arise either because of the relevant geometric, physical, or economic laws or because the logarithm of a posymonomial $c_i \prod_{j=1}^{m} t_j^{a_{ij}}$ is a *linear* function $\log c_i + \sum_{j=1}^{m} a_{ij} \log t_j$ of the logarithms, $\log t_j$, of its independent design variables t_j (and hence is relatively easy to use in analytically approximating empirically determined relations). Consequently, it is not surprising that many realistic optimization problems can be modeled accurately with generalized polynomials of one type or another.

TRADITIONAL CALCULUS AND NUMERICAL APPROACHES

The differential-calculus approach to minimizing our power-line example $P(t)$, given by Eq. (1), is to solve the *optimality condition* $dP/dt = 0$ for t; that is, solve

$$c_1 - c_2 t^{-2} = 0 \tag{2}$$

The solution to this nonlinear equation, easily accomplished analytically in this simple case, gives the optimal design (or "optimal solution")

$$t^* = (c_2/c_1)^{1/2} = (DLRI^2/CL)^{1/2} = I(DR/C)^{1/2}$$

which in turn provides the minimum total cost (or "minimum value" or "optimal value")

$$\begin{aligned} P^* &= P(t^*) = (c_1c_2)^{1/2} + (c_1c_2)^{1/2} = 2(c_1c_2)^{1/2} \\ &= 2(CLDLRI^2)^{1/2} = 2LI(CDR)^{1/2} \end{aligned}$$

However, more complicated posynomial minimization problems (with more terms T_i and/or more independent variables t_j) usually cannot be solved analytically by solving the appropriate optimality condition—namely, $dP/dt = 0$ in the single-variable case, or its multivariable version $\nabla P = 0$.

Prior to the creation of geometric programming, such minimization problems had to be solved numerically, either via a type of Newton–Raphson method applied to $dP/dt = 0$ (or $\nabla P = 0$), or via a direct-search or descent method applied directly to $P(t)$. Since all such numerical methods require specific values for the posynomial coefficients c_i, they provide only a specific optimal solution t^* and optimal value P^*, which are optimal only for the specific coefficient values and hence a very limited range of design-parameter values. Consequently, resorting to such numerical approaches does not provide the complete functional dependence of the optimal solution t^* and the optimal value P^* on the design parameters—functional dependences that designers and other decisionmakers are normally very much interested in.

THE GEOMETRIC PROGRAMMING APPROACH

We replace the nonlinear optimality condition $dP/dt = 0$ (or $\nabla P = 0$) by an equivalent nonlinear optimality condition that can be transformed into an equivalent linear optimality condition (or system of linear optimality conditions in the multivariable case) whose solutions are easily obtainable via elementary linear algebra. To do so for our power-line example, multiply the nonlinear optimality condition Eq. (2) by the unknown $t > 0$ to get the equivalent nonlinear optimality condition

$$c_1 t^1 - c_2 t^{-1} = 0 \tag{3}$$

each of whose terms is the corresponding term of $P(t)$ multiplied by the exponent of t in that term [a result that holds for *all* generalized polynomials $P(t)$, by virtue of the formulae for differentiating and multiplying posymonomials].

The linear way in which the terms of P reappear in the transformed optimality condition Eq. (3) suggests that our focus on finding the optimal t should shift to finding the optimal terms

$$T_1 = c_1 t^1 \quad \text{and} \quad T_2 = c_2 t^{-1} \tag{4}$$

which, according to the *nonlinear* optimality condition Eq. (3), must satisfy the *linear* optimality condition

$$T_1 - T_2 = 0 \tag{5}$$

Since this condition is necessary but obviously not sufficient in itself to determine the optimal terms, another optimality condition is needed. The key to finding an appropriate linear one is to use the defining equation $P = T_1 + T_2$ and the fact that the minimum $P > 0$ to infer that

$$\frac{T_1}{P} + \frac{T_2}{P} = 1 \tag{6}$$

Then the linear way in which the ratios T_1/P and T_2/P appear in this optimality condition (6) suggests that our focus on finding the optimal terms T_1 and T_2 should further shift to finding the optimal ratios

$$y_1 = T_1/P \quad \text{and} \quad y_2 = T_2/P \tag{7}$$

which are simply the fractional parts of the minimum objective value P due to its optimal terms T_1 and T_2 respectively. Needless to say, Eq. (5) divided by $P > 0$ and Eq. (6) show that these optimal ratios y_1 and y_2 satisfy both the *orthogonality condition*

$$y_1 - y_2 = 0 \tag{8}$$

and the *normality condition*

$$y_1 + y_2 = 1 \tag{9}$$

[It is worth noting here that the use of geometric concepts such as the vector-space orthogonality Eq. (8) is part of the origin of the term "geometric programming."]

Now, the linear system consisting of the orthogonality and normality conditions in Eqs. (8), (9) clearly has a unique solution

$$y_1^* = y_2^* = 1/2 \tag{10}$$

which shows that an optimally designed power line always produces capital and operating costs that are the same—invariant with respect to the coefficient vector $c = (c_1, c_2)$ (and hence the design parameters L, C, R, I, D). Other important interpretations of the optimal-ratio vector $y^* = (y_1^*, y_2^*)$ will become transparent if we do not use its specific value $(\frac{1}{2}, \frac{1}{2})$ while solving for the optimal value P^* and optimal solution t^* via the equations

$$y_1^* = c_1 t^1 / P \quad \text{and} \quad y_2^* = c_2 t^{-1} P \tag{11}$$

which result from combining Eqs. (4) and (7).

The nonlinear system Eq. (11) with the unknowns P and t is actually a disguised version of an equivalent linear system in the corresponding unknowns logarithm P and logarithm t—one that can be obtained by taking the logarithm of both sides of the Eq. (11), which produces the *log-linear* system

$$\begin{aligned} \log P &= \log(c_1/y_1^*) + \log t \\ \log P &= \log(c_2/y_2^*) - \log t \end{aligned} \tag{12}$$

This system is most easily solved by first solving for log P—simply by multiplying both sides of its two equations by y_1^* and y_2^* respectively and then adding the results to get

$$(y_1^* + y_2^*) \log P = y_1^* \log(c_1/y_1^*) + y_2^* \log(c_2/y_2^*) + (y_1^* - y_2^*) \log t \tag{13}$$

which reduces to

$$\log P = y_1^* \log(c_1/y_1^*) + y_2^* \log(c_2/y_2^*) \tag{14}$$

by virtue of the normality condition Eq. (9) and the orthogonality condition Eq. (8). Exponentiation of both sides of this equation shows that

$$P^* = (c_1/y_1^*)^{y_1^*} (c_2/y_2^*)^{y_2^*} = 2(c_1 c_2)^{1/2} = 2LI(CDR)^{1/2} \tag{15}$$

which gives the minimum value P^* prior to having an optimal solution t^*. However, an optimal solution t^* can now be obtained by substituting the formula Eq. (14) for log P back into the log-linear system Eq. (12) to get the log-linear *reduced system*

$$\begin{aligned} y_1^* \log(c_1/y_1^*) + y_2^* \log(c_2/y_2^*) &= \log(c_1/y_1^*) + \log t \\ y_1^* \log(c_1/y_1^*) + y_2^* \log(c_2/y_2^*) &= \log(c_2/y_2^*) - \log t \end{aligned} \tag{16}$$

which is overdetermined, with individual solutions

$$\log t = y_1^* \log(c_1/y_1^*) + y_2^* \log(c_2/y_2^*) - \log(c_1/y_1^*) \tag{17a}$$

and

$$\log t = \log(c_2/y_2^*) - y_1^* \log(c_1/y_1^*) - y_2^* \log(c_2/y_2^*) \tag{17b}$$

respectively. Exponentiation of both sides of these equations gives

$$t^* = (c_1/y_1^*)^{y_1^*} (c_2/y_2^*)^{y_2^*} (y_1^*/c_1) = (c_2/c_1)^{1/2} = I(DR/C)^{1/2} \tag{18a}$$

and

$$t^* = (c_2/y_2^*)(y_1^*/c_1)^{y_1^*} (y_2^*/c_2)^{y_2^*} = (c_2/c_1)^{1/2} = I(DR/C)^{1/2} \tag{18b}$$

respectively, which shows that the overdetermined system Eq. (16) does indeed have a solution—the same solution found via the traditional differential-calculus approach.

Distinguishing Features

In the geometric-programming approach, y^* is determined first, then P^*, and finally t^*—all by elementary linear algebra. In contrast, this order is reversed in the traditional differential-calculus approach, in which t^* is determined first and then $P^* = P(t^*)$. This reversal of order generally requires the solution of a nonlinear equation $dP/dt = 0$ (or, in the multivariable case, a system of nonlinear equations $\nabla P = 0$) to determine t^*—because the geometric programming transformations Eqs. (4), (7) leading to y are not used.

Analogous to duality in linear programming, y^* is the optimal solution to a *dual* of the primal problem being solved. That dual for our power-line example Eq. (1) consists of maximizing $(c_1/y_1)^{y_1}(c_2/y_2)^{y_2}$ subject to the linear orthogonality condition Eq. (8), the linear normality condition Eq. (9), and the linear positivity conditions $y_1 > 0$ and $y_2 > 0$. Since this maximation problem has a unique *dual feasible solution* $y = (\frac{1}{2}, \frac{1}{2})$ and since a unique dual feasible solution y must, *a fortiori,* be a dual optimal solution y^*, the solution of this geometric dual problem is relatively easy (involving only the linear algebra already done in finding y^*). Moreover, the lack of a geometric programming "duality gap" between the primal minimum value P^* and the dual maximum value $(c_1/y_1^*)^{y_1^*}(c_2/y_2^*)^{y_2^*}$ is an immediate consequence of Eq. (15).

Geometric dual problems with a unique dual feasible solution y are said to have zero *degree of difficulty.* In general, for geometric dual problems with at least one dual feasible solution y, this degree of difficulty is simply the dimension of the smallest linear manifold containing the dual feasible solution set, namely, the dimension of the set of solutions to the orthogonality and normality conditions. It can remain zero as the problem size, determined primarily by both the number of posynomial terms and the number of independent variables, increases (as shown in the next section).

The dual optimal solution y^* provides other important information that can be obtained by observing from the solution Eq. (15) for P^* that

$$\frac{\partial \log P^*}{\partial \log c_i} = y_i^*, \qquad i = 1, 2 \tag{19}$$

by virtue of the invariance of y^* with respect to changes in c. In essence, y^* provides a "postoptimal sensitivity analysis" analogous to that provided by the dual optimal solutions in linear programming. This sensitivity analysis becomes directly meaningful when the chain rule and the formulas (19) are used to show that

$$\begin{aligned} \frac{\partial P^*}{\partial c_i} &= \left(\frac{\partial P^*}{\partial \log P^*}\right)\left(\frac{\partial \log P^*}{\partial \log c_i}\right)\left(\frac{\partial \log c_i}{\partial c_i}\right) \\ &= (P^*)(y_i^*)\left(\frac{1}{c_i}\right), \qquad i = 1, 2 \end{aligned} \tag{20}$$

which in turn implies via the multivariable chain rule that for any design parameter p,

$$\frac{\partial P^*}{\partial p} = P^* \left[\left(\frac{y_1^*}{c_1}\right)\left(\frac{\partial c_1}{\partial p}\right) + \left(\frac{y_2^*}{c_2}\right)\left(\frac{\partial c_2}{\partial p}\right)\right] \tag{21}$$

For example, identifying p in Eq. (21) with the various design parameters L gives, via the formulas Eq. (15) for P, the formula Eq. (10) for y^*, and the formula Eq. (1) for c, the partial derivative

$$\begin{aligned}\frac{\partial P^*}{\partial L} &= 2(c_1c_2)^{1/2}\left[\left(\frac{1}{2c_1}\right)(C) + \left(\frac{1}{2c_2}\right)(DRI^2)\right] \\ &= 2(CL^2DRI^2)^{1/2}\left[\left(\frac{1}{2CL}\right)(C) + \left(\frac{1}{2DLRI^2}\right)(DRI^2)\right] \\ &= 2(CDRI^2)^{1/2} = 2C^{1/2}R^{1/2}I\end{aligned}$$

UNCONSTRAINED POSYNOMIAL MINIMIZATION VIA GEOMETRIC PROGRAMMING: THE GENERAL CASE

Given an $n \times 1$ coefficient vector $c > 0$ and an $n \times m$ exponent matrix $A = (a_{ij})$, consider the problem of minimizing the corresponding posynomial

$$P(t) = \sum_{i=1}^{n} c_i \prod_{j=1}^{m} t_j^{a_{ij}} \tag{22a}$$

over its natural domain

$$T = \{t \in R^m \mid t > 0\} \tag{22b}$$

which is the feasible solution set for unconstrained posynomial minimization. Since there need not be an optimal solution t^*, this minimization actually consists in finding the *problem infimum*

$$P^* = \inf_{t \in T} P(t) \tag{23a}$$

which is used to define the *optimal solution set*

$$T^* = \{t \in T \mid P(t) = P^*\} \tag{23b}$$

Although T^* contains a single point $t^* = (c_2/c_1)^{1/2}$ for our power-line example $P(t) = c_1t^1 + c_2t^{-1}$, it is clearly empty when either $P(t) = c_1t^1$ or $P(t) = c_2t^{-1}$ (because of the restrictions $0 < t < \infty$, which are enforced in order to keep t within the domain of $\log t$, so that the geometric programming transformations previously described are applicable). The detection and treatment of *degenerate* posynomial minimization problems in Eqs. (22), (23) for which T^* is empty (because some optimal t_j^* is 0 or ∞) is usually not needed (because well-posed realistic models normally do not imply extreme optimal designs, namely those involving 0 or ∞), but is described in Refs. 9 and 10.

Transformations

The key roles played by $\log t$ and $\log P$ in the geometric programming solution of our power-line example suggests making the transformation defined by the following change of variables:

$$z_j = \log t_j, \qquad j = 1, 2, \ldots, m \tag{24a}$$

and

$$p(z) = \log\left[\sum_{i=1}^{n} c_i \exp\left(\sum_{j=1}^{m} a_{ij}z_j\right)\right] = \log P(t) \tag{24b}$$

Since the log function is monotone increasing with range R, the other elementary properties of it and its inverse exp imply that the desired computation of P^* and T^* can be achieved via the computation of both

$$p^* = \inf_{z \in R^m} p(z) \tag{25a}$$

and

$$Z^* = \{z \in R^m \mid p(z) = p^*\} \tag{25b}$$

In particular,

$$P^* = \exp(p^*) \tag{26a}$$

and

$$T^* = \{t \in R^m \mid t_j = \exp(z_j),\ j = 1, 2, \ldots, m, \text{ for some } z \in Z^*\} \tag{26b}$$

Now, the defining formula Eq. (24b) for $p(z)$ suggests making the additional transformation defined by the following change of variables:

$$x_i = \sum_{j=1}^{m} a_{ij}z_j, \qquad i = 1, 2, \ldots, n \tag{27a}$$

and

$$g(x) = \log\left[\sum_{i=1}^{n} c_i \exp(x_i)\right] = p(z) \tag{27b}$$

Since x ranges over the vector space

$$X = [\text{column space of } A = (a_{ij})] \tag{28}$$

as z ranges over the vector space R^m, it is not hard to show that the computation of p^* and Z^* can be achieved via the computation of both

$$g^* = \inf_{x \in X} g(x) \tag{29a}$$

and

$$X^* = \{x \in X \mid g(x) = g^*\} \tag{29b}$$

even when the linear transformation $z \rightarrow x = Az$ is not one-to-one (i.e., when the exponent matrix A does not have full column rank). In particular,

$$p^* = g^* \tag{30a}$$

and

$$Z^* = \{z \in R^m \mid Az = x \text{ for some } x \in X^*\} \tag{30b}$$

In summary, Eqs. (22) through (30) show that, when X is the column space of the exponent matrix A for the posynomial $P(t)$ defined by Eqs. (22), the infimum

$$g^* = \inf_{x \in X} \log\left[\sum_{1}^{n} c_i \exp(x_i)\right] \tag{31a}$$

and corresponding optimal solution set

$$X^* = \left\{ x \in X \;\middle|\; \log\left[\sum_1^n c_i \exp(x_i)\right] = g^* \right\} \qquad (31b)$$

produce, for the posynomial minimization problem (22), (23), the desired infimum

$$P^* = \exp(g^*) \qquad (32a)$$

and corresponding optimal solution set

$$T^* = \{t \in R^m \mid t_j = \exp(z_j),\ j = 1, 2, \ldots, m \text{ for some } z \text{ such that } Az = x \text{ for some } x \in X^*\} \qquad (32b)$$

Existence and Uniqueness of Optimal Solutions

The preceding Eq. (32b) between the optimal solution sets T^* and X^* clearly implies that T^* is nonempty if and only if X^* is nonempty. Moreover, the *strict convexity* of the functions $c_i \exp(x_i)$ in Eqs. (31) implies that X^* contains at most a single x^*. Consequently, the relation (32b) shows that T^* contains at most a single t^*, unless $z \to x = Az$ is not one-to-one (because A does not have full column rank), in which case T^* has infinitely many t^* when it has at least one t^*. In any case, if T^* contains at least one t^*, then X^* contains a unique x^* from which all t^* in T^* can be computed as all those $t > 0$ that satisfy the log-linear system

$$\sum_{j=1}^m a_{ij} \log t_j = x_i^*, \qquad i = 1, 2, \ldots, n \qquad (33)$$

In particular then, all t^* in T^* can be computed from the unique x^* in X^* via *elementary linear algebra.*

When T^* is not empty [which is the case for our power-line example (1) and would normally be the case for a properly modeled problem from the real world], Eq. (26b) implies that Z^* contains at least one z^*. Moreover, since the defining formula Eq. (24b) for the objective function $p(z)$ in the associated minimization problem Eqs. (24b), (25) shows that $p(z)$ is differentiable on its feasible solution set R^m, we infer from the differential calculus that z^* satisfies the optimality condition $\nabla p(z) = 0$; that is,

$$\left\{\sum_{i=1}^n \left[c_i \exp\left(\sum_{j=1}^m a_{ij} z_j^*\right)\right]\right\}^{-1} \left\{\sum_{i=1}^n \left[c_i \exp\left(\sum_{j=1}^m a_{ij} z_j^*\right)\right] a_{ik}\right\} = 0 \qquad k = 1, 2, \ldots, m \qquad (34)$$

In view of Eqs. (27a) and (30b), these optimality conditions Eq. (34) for the problem formulation Eq. (24b), (25) imply that

$$\left(\sum_{i=1}^n c_i \exp(x_i^*)\right)^{-1} \left(\sum_{i=1}^n c_i \exp(x_i^*) a_{ik}\right) = 0, \qquad k = 1, 2, \ldots, m \qquad (35)$$

which are the optimality conditions for the problem formulation (27b), (28), (29)—the formulation with a unique x^* in X^*. Consequently, when T^* is not empty (and hence X^* contains a unique x^*), the vector y^* with components

$$y_i^* = \frac{c_i \exp(x_i^*)}{\sum_{i=1}^n [c_i \exp(x_i^*)]}, \qquad i = 1, 2, \ldots, n \qquad (36)$$

satisfies the conditions

$$\sum_{i=1}^n a_{ik} y_i = 0, \quad k = 1, 2, \ldots, m \qquad (\textit{orthogonality conditions}) \qquad (37a)$$

$$\sum_{i=1}^n y_i = 1, \qquad (\textit{normality condition}) \qquad (37b)$$

$$y_i > 0, \quad i = 1, 2, \ldots, n \qquad (\textit{positivity conditions}) \qquad (37c)$$

with the positivity conditions satisfied because each posynomial coefficient $c_i > 0$ and each $\exp(x_i^*) > 0$. Conversely, Refs. 9 and 10 show that when the conditions (37) can be satisfied (a situation that can, in principle, be detected by elementary linear algebra or linear programming), T^* is not empty and hence X^* contains a unique x^*, which produces via the Eq. (36) a y^* that is a solution, but not necessarily the only solution, to the linear system [Eq. (37)]. Moreover, Refs. 9 and 10 also show that every nontrivial posynomial minimization problem Eqs. (22), (23) can be reduced to an equivalent posynomial minimization problem whose dual constraints Eq. (37) can be satisfied. Consequently, posynomial minimization problems whose dual constraints Eq. (37) can be satisfied are termed *canonical problems;* and canonical problems, and only canonical problems, have nonempty optimal solution sets T^*, Z^*, and X^*.

Degree of Difficulty

According to linear algebra, dual constraints Eq. (37)—in fact, just the orthogonality conditions (37a) and the normality condition Eq. (37b)—can be satisfied only when the integer

$$d = n - (\text{rank}\, A + 1) \qquad (38)$$

is nonnegative. In fact, in the canonical case, if $d = 0$, linear algebra implies that the dual constraints Eq. (37) have a unique solution—namely, the vector y^* defined by Eq. (36). Moreover, in the canonical case, if $d > 0$, linear algebra and elementary topology imply that the dual constraints Eq. (37) have a solution set whose dimension is d and hence have infinitely many solutions. Consequently, if $d = 0$ in the canonical case, the vector y^* defined by Eq. (36) can be obtained only via elementary linear algebra—as in our power-line example (1). On the other hand, if $d > 0$ in the canonical case, the vector y^* defined by Eq. (36) can *not* be obtained via only elementary linear algebra but can be obtained via a numerical solution of either the primal posynomial minimization problem Eq. (22, 23) or one of its equivalent reformulations Eq. (24b, 25) or Eq. (27b, 28, 29)−or via a numerical solution of their dual problem (which has been described for the power-line example (1) but is not generally defined until a later subsection). Actually, posynomial minimization problem Eq. (22, 23) is normally *not* solved numerically when $d > 0$, because it usually does not have the desirable property of being convex. However, its equivalent reformulations and their dual are convex, but choosing which of those three to solve numerically when $d > 0$ requires more information about the exponent matrix A. Since Refs. 8–10 show that the dual problem is only linearly constrained (with appropriate orthogonality, normality and positivity conditions) even when nonlinear posynomial constraints are present in the primal problem, the dual problem should normally be solved numerically when optimally designing equipment subject to constraints. Since we have already noted that d is the dimension of the dual feasible solution set, d has been termed the degree of difficulty of the dual problem, as well as the degree of difficulty of the corresponding primal posynomial problem Eq. (22, 23) and its equivalent reformulations Eq. (24b, 25) and Eq. (27b, 28, 29).

The Determination of the Optimal Value and All Optimal Solutions

Once y^* is obtained (usually, but not always, via the dual problem), the desired optimal value P^* and all optimal solutions t^* can easily be obtained from y^*, by first noting that Eqs. (22a) and (33) imply that the Eq. (36) can be rewritten as

$$y_i^* = \frac{c_i \exp(x_i^*)}{P^*}, \qquad i = 1, 2, \ldots, n \tag{39}$$

which shows that these components y_i^* of y^* are simply the fractional parts of the minimum objective value P^* due to its optimal terms $c_i \exp(x_i^*)]$ respectively—the same interpretation provided by the equations Eq. (7) for our power-line example Eq. (1). Now, take the logarithm of both sides of the equations Eq. (39) to get

$$\log P^* = \log(c_i/y_i^*) + x_i^*, \qquad i = 1, 2, \ldots, n \tag{40}$$

and then multiply both sides of Eq. (40) by y_i^*, $i = 1, 2, \ldots, n$, respectively. Now, add the resulting equations to get

$$\left(\sum_{i=1}^{n} y_i^*\right) \log P^* = \sum_{i=1}^{n} y_i^* \log \frac{c_i}{y_i^*} + \sum_{i=1}^{n} x_i^* y_i^* \tag{41}$$

which reduces to

$$\log P^* = \sum_{i=1}^{n} y_i^* \log \frac{c_i}{y_i^*} \tag{42}$$

because y^* satisfies the normality condition Eq. (37b) and because x^* and y^* are orthogonal by virtue of the transformation equations Eq. (27a) and the orthogonality conditions Eq. (37a). Needless to say, exponentiation of both sides of Eq. (42) gives the desired optimal value

$$P^* = \prod_{i=1}^{n} \left(\frac{c_i}{y_i^*}\right)^{y_i^*} \tag{43}$$

and substituting the formula Eq. (42) for $\log P^*$ back into the Eq. (40) gives the optimal

$$x_i^* = \left(\sum_{i=1}^{n} y_i^* \log \frac{c_i}{y_i^*}\right) - \log \frac{c_i}{y_i^*}, \qquad i = 1, 2, \ldots, n \tag{44}$$

from which all t^* in T^* can be computed as all those $t > 0$ that satisfy the log-linear system Eq. (33).

Some real-world problems to which the preceding theory can be applied originate with a need to solve a problem modeled by the dual constraints Eq. (37) rather than by the primal posynomial minimization problem Eqs. (22), (23).

An Important Example: The Numerical Solution of Regular Markov Chains

A physical system whose state can change randomly during each transition, but with a known probability distribution, can be accurately modeled as a *Markov process.* For example, the analysis and design of a complicated engineering system (such as a large telephone network or computer network) frequently requires the numerical solution of a *Markov chain,* for which the known probability distribution depends only on the system's current state (rather than on its history of previous states). A Markov chain with only a finite number n of discrete states i can be completely characterized by a single $n \times n$ matrix P—the *transition matrix* whose element p_{ij} is the known probability of going from a current state i to state j in one transition. In particular then, row i of P is a known probability distribution for which

$$\left.\begin{aligned} &\sum_{j=1}^{n} p_{ij} = 1, && (45a)\\ &p_{ij} \geq 0, \quad j = 1, 2, \ldots, n && (45b)\end{aligned}\right\} i = 1, 2, \ldots, n$$

Given a Markov chain that is *regular,* or *ergodic,* in that $P^q > 0$ for some positive integer q (the case for many engineering systems), it is well known that the corresponding linear system

$$yP = y \tag{46a}$$

$$\sum_{i=1}^{n} y_i = 1 \tag{46b}$$

has a *unique* solution y^* and that

$$y^* > 0 \tag{46c}$$

Since it is also known that y_i^* gives the probability of the system being in state i after a large number of transitions (regardless of the system's initial state i_0), the computation of this equilibrium distribution y^* is very important.

When the number n of system states i is extremely large, current computer limitations prevent the computation of y^* via the standard pivot operations of linear algebra applied to the system Eq. (46a, b). In such cases, an iterative approach based on the preceding geometric programming theory can be successful.

This approach comes from observing that the linear system Eq. (46) is equivalent to those dual constraints Eq. (37) whose exponent matrix A has elements

$$a_{ij} = \begin{cases} p_{ij} - 1 & \text{if} \quad i = j \\ p_{ij} & \text{if} \quad i \neq j \end{cases} \tag{47}$$

Since these dual constraints have a unique solution, namely $y^* > 0$, the corresponding transformed posynomial minimization problem Eqs. (24b), (25) has zero degree of difficulty and is canonical [as defined and discussed following the dual constraints Eq. (37)]; so the problem Eqs. (24b), (25) has an optimal solution z^* as long as each coefficient $c_i > 0$. Although z^* is not unique [because the column vectors of A sum to 0 by virtue of Eqs. (45a) and (47)], each z^* provides the desired equilibrium distribution y^* via the formula

$$y_i^* = \frac{c_i \exp\left(\sum_{j=1}^{n} a_{ij} z_j^*\right)}{\sum_{k=1}^{n} c_k \exp\left(\sum_{j=1}^{n} a_{kj} z_j^*\right)}, \qquad i = 1, 2, \ldots, n \tag{48}$$

which comes from combining Eqs. (30b) and (36).

If the coefficient vector c is chosen to be an *a priori* estimate of y^* (or the uniform distribution $c_i = 1/n$ when no such estimate is available), differentiation of the objective function

$$\log\left[\sum_{i=1}^{n} c_i \exp\left(\sum_{j=1}^{n} a_{ij} z_j\right)\right]$$

for the minimization problem Eqs. (24b), (25) shows that 0 should be the initial estimate of z^*. In particular, the gradient of this objective function evaluated at 0 can then serve as a *residual* in the usual numerical linear-algebraic sense to help determine an improved estimate of z^* and hence an improved estimate of y^*. A discussion of strategies for producing rapid convergence to z^* and hence rapid convergence to y^* lies at the interface of numerical linear algebra and numerical convex optimization—topics beyond the scope of this article.

The Dual Problem

Like any optimization problem, the dual problem has both a feasible solution set, the *dual feasible solution set,* and an objective function, the *dual objective function.* For the posynomial minimization problem Eqs. (22), (23) [including its equivalent formulations Eqs. (24b), (25) and Eqs. (27b), (28), (29)], the dual feasible solution set consists of all solutions to the dual constraints

$$\sum_{i=1}^{n} a_{ik} y_i = 0, \quad k = 1, 2, \ldots, m \qquad \text{the orthogonality conditions} \tag{49a}$$

$$\sum_{i=1}^{n} y_i = 1, \qquad \text{the normality condition} \tag{49b}$$

$$y_i \geq 0, \quad i = 1, 2, \ldots, n \qquad \text{the positivity conditions} \tag{49c}$$

which differ from the originally motivated dual constraints Eqs. (37) only in that the positivity condition Eq. (49c) is a slightly relaxed version of the positivity condition Eq. (37c)—a relaxation that is needed to obtain the most complete duality theory for posynomial programming. The dual objective function U, which is motivated by Eq. (43) and is to be maximized, has the formula

$$U(y) = \prod_{i=1}^{n} \left(\frac{c_i}{y_i}\right)^{y_i} \tag{50}$$

with the understanding that $0^0 = 1$—so that $U(y)$ is a continuous function for $y \geq 0$.

The Main Duality Theorem. If t is primal feasible [in that t satisfies the primal constraints $t > 0$ for the posynomial minimization problem Eqs. (22), (23)] and if y is dual feasible [in that y satisfies the constraints Eq. (49) for the corresponding dual problem Eqs. (49), (50)], then

$$U(y) \leq P(t) \tag{51a}$$

with equality holding if, and only if,

$$y_i = \frac{\left[c_i \prod_{j=1}^{m} t_j^{a_{ij}}\right]}{\left[\sum_{k=1}^{n} c_k \prod_{j=1}^{m} t_j^{a_{kj}}\right]} \qquad i = 1, 2, \ldots, n \tag{51b}$$

in which case t and y are primal and dual optimal, respectively and the primal problem Eqs. (22, 23) and its dual problem Eqs. (49, 50) are canonical. Duality inequality Eq. (51a) and the corresponding primal-dual optimality condition Eq. (51b) can be established with the aid of the well-known *Cauchy's inequality*

$$\prod_{i=1}^{n} u_i^{y_i} \leq \sum_{i=1}^{n} y_i u_i \tag{52}$$

between the geometric mean $\prod_{i=1}^{n} u_i^{y_i}$ and the arithmetic mean $\sum_{i=1}^{n} y_i u_i$ of n numbers $u_i \geq 0$, where n weights $y_i \geq 0$ and $\sum_{i=1}^{n} y_i = 1$. This geometric-mean arithmetic-mean inequality Eq. (52) becomes an equality if, and only if, there is some $u \geq 0$ such that $u_i = u$ for $i = 1, 2, \ldots, n$. To use these facts to establish the duality inequality (51a) and primal-dual condition (51b), let

$$u_i = \frac{T_i}{y_i} = c_i \prod_{j=1}^{m} \frac{t j^{a_{ij}}}{y_i} \tag{53}$$

and then employ both the primal constraints $t > 0$ and the dual constraints Eq. (49). A by-product is that the dual problem Eqs. (49), (50) has a unique optimal solution y^* [determined via Eqs. (32b) and Eq. (36)] when the primal problem Eqs. (22), (23) has at least one optimal solution t^*—the situation for canonical problems. (It is worth noting here that this use of the geometric mean $\prod_{i=1}^{n} u_i^{y_i}$ in Cauchy's inequality Eq. (52) is partly the origin of the term *geometric* programming.) Also, for canonical problems, the implicit function theorem from multivariable calculus can be used to show that

$$\frac{\partial \log P^*}{\partial \log c_i} = y_i^*, \qquad i = 1, 2, \ldots, n \tag{54}$$

which is the basis for postoptimal sensitivity analyses—as previously illustrated in the power-line example Eq. (1).

Since the reformulations Eqs. (24b), (25) and Eqs. (27b), (28), (29) of the primal posynomial minimization problem Eq. (22), (23) have provided key insights into posynomial minimization, it should not be surprising to learn that certain reformulations of its dual problem Eqs. (49), (50) also provide valuable insights into posynomial minimization.

Dual Reformulations

The dual constraints Eq. (49) are linear; so the dual feasible solutions y can be characterized in various ways via linear algebra and linear programming.

Linear-Algebraic Reformulations. These reformulations characterize the dual feasible solutions y in terms of the general solutions y to the orthogonality and normality conditions Eq. (49a, b). In particular, for a dual problem Eqs. (49), (50) with degree of difficulty d [defined by Eq. (38)], such a characterization results from constructing *basic vectors* $b^{(j)}$ for $j = 0, 1, \ldots, d$ so that each dual feasible solution

$$y = b^{(0)} + \sum_{j=1}^{d} r_j b^{(j)} \qquad (55)$$

for values of the *basic variables* r_j for which $b^{(0)} + \sum_{j=1}^{d} r_j b^{(j)} \geq 0$. The vector $b^{(0)}$, which satisfies both the orthogonality and *normality conditions* Eqs. (49a, b), is termed a normality vector. The vectors $b^{(j)}$ for $j = 1, \ldots, d$, which are linearly independent solutions to the homogeneous counterpart of the orthogonality and normality conditions Eqs. (49a, b), are called *nullity vectors.* If $d = 0$, then $b^{(0)}$ is unique (and equal to y^*) and the nullity vectors do not exist. If $d > 0$ (the case to be treated in this subsection), the basic vectors $b^{(j)}$ are not unique and can usually be chosen so that they have special meaning for the special problem being treated.

In any event, the dual objective function $U(y)$ [to be maximized to determine U^* and y^* so that the desired P^* and t^* can be determined via the duality equations Eqs. (43), (44), (33)], written in terms of the basic variables r_j, is

$$V(r) = \left(\prod_{i=1}^{n} c_i^{\,b_i^{(0)} + \sum_{j=1}^{d} r_j b_i^{(j)}} \right) \left(\prod_{i=1}^{n} y_i(r)^{-y_i(r)} \right)$$

$$= K_0 \left(\prod_{j=1}^{d} K_j^{r_j} \right) \left(\prod_{i=1}^{n} y_i(r)^{-y_i(r)} \right) \qquad (56a)$$

where the basic constants are

$$K_j = \prod_{i=1}^{n} c_i^{b_i^{(j)}}, \qquad j = 0, 1, \ldots, d \qquad (56b)$$

and where

$$y_i(r) = b_i^{(0)} + \sum_{j=1}^{d} r_j b_i^{(j)}, \qquad i = 1, 2, \ldots, n \qquad (56c)$$

In summary, the dual problem Eqs. (49), (50) (and hence the primal problem Eqs. (22), (23) and its equivalents Eqs. (24b), (25) and Eqs. (27b), (28), (29)) can be solved by maximizing the reformulated dual objective function $V(r)$ defined by Eqs. (56), subject to the reformulated positivity conditions

$$b_i^{(0)} + \sum_{j=1}^{d} b_i^{(j)} r_j \geq 0, \qquad i = 1, 2, \ldots, n \qquad (57)$$

Prior to maximizing $V(r)$, useful qualitative information about optimal value V^* ($= U^* = P^*$) can be obtained from the defining formulas Eq. (56) for $V(r)$ and K_j. In essence, constructing the K_j [using only linear algebra on the exponent matrix $A = (a_{ij})$] performs a dimensional analysis of the dual problem Eqs. (49), (50) [and hence the primal problem Eqs. (22), (23) and its equivalents]—in that the formula Eq. (56a) for $V(r)$ and the duality equation $V^* = P^*$ indicate that K_0 has the dimensions of the posynomial P (dollars, in cost minimization) while the other K_j, $j = 1, 2, \ldots, d$, are dimensionless. Moreover, for a fixed $A = (a_{ij})$ (typically fixed by the unchanging laws of geometry, physical science, and economics), the normality and nullity vectors $b^{(j)}$ can be fixed independently of the coefficients c_i (typically not fixed but determined by changing design parameters, such as material prices and design specifications). The basic constants K_j are then functions only of the coefficients c_i; in fact, each $\log K_j$ is a linear function of the $\log c_i$'s, as indicated by taking the logarithm of each side of the defining Eq. (56b) for K_j. The resulting equations

$$\sum_{i=1}^{n} b_i^{(j)} \log c_i = \log K_j, \qquad j = 0, 1, \ldots, d \qquad (58)$$

are satisfied by *infinitely many* coefficient vectors c for a given basic constant vector K (resulting from *one particular* choice of c), because the number $n = d + \text{rank}\, A + 1$ of coefficients c_i [obtained from Eq. (38)] is clearly always greater than the number $d + 1$ of basic constants K_j. Each solution c to the preceding linear system Eq. (58) determines a different primal problem Eqs. (22), (23), but the corresponding reformulated dual problems Eqs. (56), (57) are all the same; so the minimum value P^* for each of these primal problem is the same even though the primal optimal solutions t^* are generally different. In summary, the solution of a specific problem Eqs. (56), (57) (by the maximization of $V(r)$ for a particular K) solves infinitely many posynomial minimization problems Eqs. (22), (23) (determined by all solutions c to the linear system Eq. (58) for the particular K).

Maximizing $V(r)$ can, of course, be achieved by maximizing

$$\log V(r) = \log K_0 + \sum_{j=1}^{d} (\log K_j) r_j - \sum_{i=1}^{n} y_i(r) \log y_i(r)$$

Since previously described theory (for canonical problems) asserts the existence of an optimal $y^* > 0$, there is a corresponding optimal r^* such that $y(r^*) = y^*$. The differentiability of $\log V(r)$ at such an r^* implies that $\partial(\log V)/\partial r_j\,(r^*) = 0$, $j = 1, 2, \ldots, d$, which means that

$$\log K_j - \sum_{i=1}^{n} (\log y_i^* + 1) b_i^{(j)} = 0, \qquad j = 1, 2, \ldots, d$$

and hence that

$$\log K_j = \sum_{i=1}^{n} b_i^{(j)} \log y_i^*,\ j = 1, 2, \ldots, d$$

because

$$\sum_{i=1}^{n} b_i^{(j)} = 0,\ j = 1, 2, \ldots, d$$

Since log U is a concave function of y, the preceding computation actually shows that a dual feasible solution $y > 0$ is in fact dual optimal if and only if

$$\log K_j = \sum_{i=1}^{n} b_i^{(j)} \log y_i, \qquad j = 1, 2, \ldots, d \tag{59a}$$

in which case $\log U = \log K_0 - \sum_{i=1}^{n} b_i^{(0)} \log y_i$ and hence

$$P* = K_0 \prod_{i=1}^{n} y_i^{-b_i^{(0)}} \tag{59b}$$

Note that the *maximizing equations* Eq. (59a) map each dual feasible solution $y > 0$ into basic constants K in such a way that the dual feasible solution y is actually the dual optimal solution y^* for each of the infinitely many posynomial minimization problems Eqs. (22), (23) with a coefficient vector c that satisfies the resulting linear system Eq. (58).

If the degree of difficulty d is only one, there is only one maximizing Eq. (59a), one $K_j = K_1$, and one $r_j = r_1$. In that case, simply graphing the resulting maximizing equation

$$\log K_1 = \sum_{i=1}^{n} b_i^{(1)} \log(b_i^{(0)} + b_i^{(1)} r_1)$$

treating r_1 as the independent variable and $\log K_1$ as the dependent variable, essentially solves all posynomial minimization problems Eq. (22), (23) that have the exponent matrix A used in constructing the normality and nullity vectors $b^{(j)}$. The reason is that, given a particular coefficient vector c, the defining Eq. (58) for $\log K_j$ give a particular

$$\log K_1 = \sum_{i=1}^{n} b_i^{(1)} \log c_i$$

which determines, via the graph, the corresponding optimal r_1^* and hence the dual optimal solution $y^* = b^{(0)} + b^{(1)} r_1^*$, from which the desired P^* and t^* can be determined via the duality Eqs. (43), (44), (33). In retrospect, it is worth noting that the graph of $\log K_1$ versus r_1: (1) always has range R, because the range of $\log K_1$ in the preceding displayed formula is clearly always R, and (2) is always one-to-one, because the dual optimal solution y^* is unique, and hence so is r^* (by virtue of the linear independence of the $b^{(j)}$).

If the degree of difficulty d is larger than one, the graph of the maximizing equations is in at least a four-dimensional space; so the preceding solution procedure generally requires a numerical solution technique (such as the Newton–Raphson method) to determine r^* from a knowledge of $\log K_j$.

Linear Programming Reformulations. These reformulations characterize the dual feasible solutions y in terms of the basic dual feasible solutions y^k, $k = 1, 2, \ldots, p$, to the dual feasibility conditions Eq. (49). Unlike the normality and nullity vectors $b^{(j)}$, $j = 0, 1, \ldots, d$, the basic dual feasible solutions y^k are unique and can be determined from the linear system Eq. (49) via phase I of the simplex method for linear programming. Since the number of nonbasic variables relative to the simplex tableau that determines y^k is $n -$ (rank $A + 1$) [namely, the degree of difficulty d by virtue of Eq. (38)], each basic dual feasible solution y^k has at least d zero components (in fact, exactly d zero components if, and only if, y^k is *nondegenerate* in the linear programming sense). Moreover, according to the resolution theorem (sometimes called the "decomposition theorem" or "Weyl's theorem" or "Goldman's theorem") for polytopes, each dual feasible solution y is a convex combination of the basic dual feasible solutions y^k; that is,

$$y = \sum_{k=1}^{p} \delta_k y^k$$

$$\text{for appropriate weights } \delta \text{ for which } \delta \geq 0 \text{ and } \sum_{k=1}^{p} \delta k = 1 \tag{60}$$

Moreover, for nontrivial canonical problems [those for which the (rank A) ≥ 1], linear algebra and the simplex method can be used to show that $n/2 \leq p \leq n!/d!\,(n - d)!$.

Since y^k is orthogonal to each column of the exponent matrix A [by virtue of the dual feasibility of y^k and the orthogonality condition Eq. (49a)], it is clear that the vector y^k that results from deleting the zero components of y^k is orthogonal to each column of the matrix A that results from deleting the corresponding rows of A. Moreover, since y^k obviously inherits normality and positivity from y^k, it is a dual feasible solution for minimizing the posynomial P_k that results from deleting the corresponding terms of P. In fact, minimizing P_k is a canonical problem because $y^k > 0$; and it has zero degree of difficulty, because the components of y^k are uniquely determined by the zero values for the nonbasic variables relative to the simplex tableau that determines y^k as a basic feasible solution to the linear system Eq. (49). In essence, minimizing P_k is a meaningful approximation to minimizing the original posynomial P—an approximation that is easy to solve because of its zero degree of difficulty. Similar reasoning, combined with Tucker's positivity theorem concerning orthogonal complementary subspaces, shows that deleting even one additional term from the posynomial P_k would produce a posynomial whose infimum was zero—indicating that its minimization could not possibly be a meaningful approximation to minimizing the original posynomial P. In summary, for $k = 1, 2, \ldots, p$, the nonzero components of the basic dual feasible solution y^k constitute the dual optimal solution y^k to a meaningful (though not necessarily accurate) minimal-size, zero-degree-of-difficulty, canonical approximation to the problem of minimizing P—namely, the problem of minimizing the posynomial P_k that results from deleting the terms of P that correspond to the zero components of y^k.

Since $c_i^0 = 1$ and since we have defined $0^0 = 1$, the zero degree of difficulty in minimizing P_k along with the duality inequality Eq. (50), (51) implies that

$$P_k^* = \min P_k = \prod_{i=1}^{n} \left(\frac{c_i}{y_i^k} \right)^{y_i^k} < \min P = P^*, \qquad k = 1, 2, \ldots, p$$

where the strict inequality results from the fact that y^k has at least one zero component and hence cannot be dual optimal for minimizing P (which we know has a unique dual optimal $y^* > 0$). To improve on the resulting best extreme-point lower bound for P^*, namely $\max_k \{P_k^* | k = 1, 2, \ldots, p\}$, use Eq. (60) to reformulate the dual objective function $U(y)$ [defined by Eq.

(50)] in terms of δ as

$$W(\delta) = \left(\prod_{i=1}^{n} c_1^{y_i(\delta)}\right)\left(\prod_{i=1}^{n} y_i(\delta)^{-y_i(\delta)}\right) = \left(\prod_{k=1}^{p} L_k^{\delta_k}\right)\left(\prod_{i=1}^{n} y_i(\delta)^{-y_i(\delta)}\right)$$

where

$$y_i(\delta) = \sum_{k=1}^{p} y_i^k \delta_k, \qquad i = 1, 2, \ldots, n$$

and where the basic constants

$$L_k = \prod_{i=1}^{n} c_i^{y_i^k} = P_k^* \prod_{i=1}^{n} (y_i^k)^{y_i^k}, \qquad k = 1, 2, \ldots, p$$

Then, maximize $W(\delta)$ subject to the reformulated dual constraints

$$\delta \geq 0 \quad \text{and} \quad \sum_{k=1}^{p} \delta_k = 1$$

Additional problems to which the preceding theory applies originate with a need to solve a problem modeled by the dual maximization problem Eqs. (49), (50) [rather than its corresponding primal posynomial minimization problem Eqs. (22), (23)].

AN IMPORTANT EXAMPLE: ENTROPY OPTIMIZATION IN INFORMATION THEORY, THERMODYNAMICS, AND STATISTICAL MECHANICS

Given a finite sample space $\{s_1, s_2, \ldots, s_n\}$ with possible outcomes s_i (not necessarily numbers) a fundamental problem having to do with probability and statistics is to *infer* the associated probability distribution

$$y \geq 0 \tag{61a}$$

$$\sum_{i=1}^{n} y_i = 1 \tag{61b}$$

from given *moment conditions*

$$\sum_{i=1}^{n} \nu_{ij} y_i = \mu_j, \qquad j = 1, \ldots, m \tag{62}$$

and a given a priori distribution

$$q \geq 0 \tag{63a}$$

$$\sum_{i=1}^{n} q_i = 1 \tag{63b}$$

The moment conditions Eq. (62) typically result from statistically obtained expected values μ_j of known random variables ν_{ij}; and the a priori distribution q is *uniform* (i.e., $q_i = 1/n$) when no other information is available about y.

The *fundamental principle of information theory* (which is derived in Refs. 12 and 13 from certain reasonable axioms in probability theory) is that the best inference for unknown probability distribution y from the given moment conditions Eq. (62) and a priori distribution Eq. (63) is the optimal solution to the following optimization problem:

$$\left.\begin{array}{l}\text{Maximize the cross entropy } H(y) = \displaystyle\sum_{i=1}^{n} y_i \log \frac{nq_i}{y_i} \\ \text{subject to the constraints (61) and (62)}\end{array}\right\} \tag{64}$$

Since $H(y) = \log U(y)$ when $c_i = nq_i$ [by virtue of Eq. (50)] and since the condition Eq. (61b) makes the moment conditions Eq. (62) equivalent to the orthogonality conditions

$$\sum_{i=1}^{n} (\mu_j - \nu_{ij}) y_i = 0, \qquad j = 1, \ldots, m \tag{65}$$

the maximization problem Eq. (64) is essentially the dual problem Eqs. (49), (50) when

$$c_i = nq_i \quad \text{and} \quad a_{ij} = \mu_j - \nu_{ij}$$

Consequently, the corresponding primal problem [Eqs. (24b, 25)] which we shall see is more suitable and relevant than both its posynomial equivalent [Eqs. (22, 23)] and vector space equivalent [Eqs. (27b, 28, 29)] is

$$\left.\begin{array}{rl}\text{Minimize } G(z) = & \log\left[\displaystyle\sum_{i=1}^{n} nq_i \exp\left(\sum_{j=1}^{m} \{\mu_j - \nu_{ij}\} z_j\right)\right] \\ = & \left\{\displaystyle\sum_{j=1}^{m} \mu_j z_j + \log\left[\sum_{i=1}^{n} nq_i \exp\left(-\sum_{j=1}^{m} \nu_{ij} z_j\right)\right]\right\}\end{array}\right\} \tag{66}$$

Since $n \ggg m$ and hence the degree-of-difficulty $d = n - (\text{rank } A + 1) \ggg m$, problem Eq. (66) is probably much easier to solve numerically than problem Eq. (64). Moreover, the previously described canonicality theory for posynomial programming implies that problem Eq. (66) has an optimal solution z^* if, and only if, constraints Eq. (61) and Eq. (62) have a feasible solution $y > 0$. Since the sample space $\{s_1, s_2, \ldots, s_n\}$ can obviously be made smaller if there is no such feasible distribution $y > 0$, we can assume, without loss of generality, that problems Eq. (64) and Eq. (66) are canonical. Then, the previously described posynomial programming theory implies the following facts (many of which were first established via geometric programming and reported in Ref. 14):

(1) There is a unique optimal y^* [the inferred distribution], and $y^* > 0$.
(2) There is an optimal z^*; and z^* is unique if, and only if, the moment conditions Eq. (62) are linearly independent.
(3) The solution pairs (y^*, z^*) constitute the solution set for the system consisting of the moment conditions Eq. (62) and the "primal-dual optimality conditions"

$$y_i = \left[q_i \exp\left(-\sum_{j=1}^{m} \nu_{ij} z_j\right)\right] \Big/ \left\{\sum_{i=1}^{n}\left[q_i \exp\left(-\sum_{j=1}^{m} \nu_{ij} z_j\right)\right]\right\}, \quad i = 1, 2, \ldots, n \tag{67}$$

which come from conditions Eqs. (24a, 51b, 65) and algebraic simplification.

(4) The solution pairs (y^*, z^*) also constitute the solution set for the system consisting of the probability-distribution conditions Eq. (61), the moment conditions Eq. (62) and the "duality equation"

$$H(y) = G(z) \tag{68}$$

(5) If each $\nu_{ij} = 0$ and each $\mu_j = 0$, then the primal-dual optimality conditions Eq. (67) show that $y^* = q$ [by virtue of the *a priori* probability-distribution condition Eq. (63b)]. This means that setting $y = q$ maximizes the cross-entropy $H(y)$ when the only constraints on y are the probability-distribution conditions Eq. (61). It follows then that:
   (a) the inferred distribution y^* is simply the *a priori* distribution q when q satisfies the moment conditions Eq. (62),
   (b) when q satisfies the moment conditions Eq. (62) and $q_i = 1/n$, then $y^*_i = 1/n$ (so the principle of maximum cross entropy generalizes "LaPlace's principle of insufficient reason").

(6) Given that $q_i = 1/n$ and that $m = 1$ (with simplified notation $z = z_1$, $\mu = \mu_1$ and $\nu_i = \nu_{i1}$) and given that the sample space $\{s_1, s_2, \ldots, s_n\}$ consists of the possible "states" i of a "physical system" that has "energy" ν_i in state i (with μ being the system's average energy or "internal energy"), then the primal-dual optimality conditions Eq. (67) further simplify to

$$y_i = [\exp(-\nu_i z)] \Big/ \left\{ \sum_{i=1}^{n} [\exp(-\upsilon_i z)] \right\}, \quad i = 1, 2, \ldots, n \tag{69}$$

in which case
   (a) the denominator in the primal-dual optimality conditions Eq. (69) is the system's "partition function" Q,
   (b) the system's "absolute temperature" $T = 1/\kappa z^*$ where κ is "Boltzmann's constant,"
   (c) the primal-dual optimality conditions Eq. (69) and the internal-energy condition

$$\sum_{i=1}^{n} \nu_i y_i = \mu \tag{70}$$

along with the interpretation $z^* = 1/\kappa T$ constitute the "fundamental law" (described in Ref. 15 and elsewhere) relating statistical mechanics to thermodynamics—a law which, according to the geometric programming theory described herein, can also be expressed in terms of the "dual variational principles" provided by optimization problems Eqs. (22, 23), (24b, 25), (27b, 28, 29) and (49, 50).

The variational principle that connects the cross-entropy maximization problem Eq. (64) with the fundamental law Eqs. (69, 70) for statistical mechanics and thermodynamics had previously been noted and pedagogically exploited in Refs. 16 and 17, but the alternative variational principles provided by problems Eqs. (22, 23), (24b, 25) and (27b, 28, 29) seem to have originated in Ref. 14. Other connections between geometric programming, statistical mechanics and thermodynamics had previously been given in Refs. 18 and 19.

Finally, the significance of the cross-entropy maximization problem (64) in statistical theory and its applications is thoroughly described in Ref. 20, but the significance of the corresponding geometric programming problems (22, 23), (24b, 25) and (27b, 28, 29) in statistical theory and its applications is yet to be determined.

Constrained Algebraic Optimization Via Geometric Programming

References 8–10 show how essentially all of the theory and methodology described herein can be extended to the minimization of posynomials $P(t)$ subject to "inequality constraints" of the type $Q(t) \le q$ on other posynomials $Q(t)$. Although such minimization problems are generally "nonconvex," the reformulations that result from extending the geometric programming transformations described herein are "convex" when all constraints are of the "prototype form" $Q(t) \le q$. These generalizations greatly enlarge the applicability of posynomial minimization to engineering design and other areas, as can be seen in many references (such as Refs. 10, 11, and 21–24). They also include the "chemical equilibrium problem" as an important example of the resulting geometric dual problem, while including the extremely important "linear programming duality theory" as a special case of the resulting geometric programming duality theory (as can be seen in Ref. 10). Moreover, Ref. 25 shows how to reformulate all well-posed "algebraic optimization problems" (those with meaningful algebraic objective and constraint functions and any type of constraint involving the relations $\le$, $\ge$, and $=$) as equivalent posynomial minimization problems with posynomial constraints of both the desired prototype $Q(t) \le q$ and the "reversed type" $R(t) \ge r$. Moreover, Ref. 26 shows that this reformulation taken to its logical conclusion results in objective and constraint posynomials with at most two terms each—very close to the special linear programming case of exactly one term each. Finally, Refs. 27–29 show how the amazingly general posynomial minimization problems with reversed constraints $R(t) \ge r$ can be "conservatively approximated" by those with only constraints of the desired prototype $Q(t) \le q$.

Generalized Geometric Programming

Geometric programming is not just a special methodology for studying the extremely important class of algebraic optimization problems and their entropy-like dual problems. Its mathematical origin is actually the prior use of certain "orthogonal complementary subspaces" and the "Legendre transformation" in the study of electrical networks (in Ref. 30). Replacing the orthogonal complementary subspaces with the more general "dual convex cones" while replacing the Legendre transformation with the more general "conjugate transformation" has produced an extremely general mathematical theory and methodology for treating *all* linear and nonlinear optimization problems, as well as most (if not all) equilibrium problems. This generalized theory and methodology (developed primarily in Refs. 31 and 32) is especially useful for studying a large class of "separable problems." Its practical significance is due mainly to the fact that many important (seemingly inseparable) problems can actually be reformulated as separable generalized geometric programming problems—by fully exploiting their linear-algebraic structure (which is frequently hidden, as in the case of posynomial minimization). Some examples are *quadratic programming* (which should be treated separately from the general algebraic case), *discrete optimal control with linear dynamics* (or dynamic programming with linear transition equations), *economic equilibria* (either in the context of relatively simple exchange models or in the more sophisticated context of spatial and temporal models), *network analysis and operation* (particularly "monotone networks" of electric or hydraulic type, and certain types of transportation networks and transshipment networks, including both single-commodity and multi-commodity cases, as well as traffic assignment), *optimal location/allocation analysis, regression analysis, structural analysis,* and *design, tomography,* and *nondestructive testing.* The general theory of geometric programming includes (1) very strong existence, uniqueness, and characterization theorems, (2) useful parametric and post-optimality analyses, (3) illuminating decomposition principles, and (4) powerful numerical solution techniques.

A comprehensive survey of the whole field as it existed in 1980 can be found in Ref. 33. Finally, Ref. 34 will provide a current state-of-the-art survey in 1999 (or shortly thereafter).

BIBLIOGRAPHY

1. C. Zener, A mathematical aid in optimizing engineering designs, *Proc. Natl. Acad. Sci.,* **47**: 537, 1961.
2. C. Zener, A further mathematical aid in optimizing engineering designs, *Proc. Natl. Acad. Sci.,* **48**: 518, 1962.
3. C. Zener, Minimization of system costs in terms of subsystem costs, *Proc. Natl. Acad. Sci.,* **51**: 162, 1964.
4. C. Zener, An example of design for minimum total costs, counter-flow heat exchangers, *IEEE Trans. Mil. Electron.,* **MIL-8**: 63, 1964.
5. C. Zener, Redesign over-compensation, *Proc. Natl. Acad. Sci.,* **53**: 242, 1965.
6. R. J. Duffin, Dual programs and minimum cost, *SIAM J.,* **10**: 119, 1962.
7. R. J. Duffin, Cost minimization problems treated by geometric means, *Oper. Res.,* **10**: 668, 1962.
8. R. J. Duffin and E. L. Peterson, *Constrained minima treated by geometric means,* Westinghouse Scientific Paper 64-158-129-P3, 1964.
9. R J. Duffin and E. L. Peterson, Duality theory for geometric programming, *SIAM J. Appl. Math.,* 1966.
10. R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric Programming—Theory and Application,* New York: Wiley, 1967.
11. C. Zener, *Engineering Design by Geometric Programming,* New York: Wiley, 1971.
12. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication,* Urbana: Univ. Illinois Press, 1949.
13. A. I. Khinchin, *Mathematical Foundations of Information Theory,* New York: Dover, 1957.
14. R. J. Duffin and E. L. Peterson, Optimization and insight by geometric programming, *J. Appl. Phys.,* **60**: 1860, 1986.
15. R. P. Feynman, *Statistical Mechanics—A Set of Lectures,* New York: Benjamin, 1972.
16. E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.,* **108**, 1957.
17. A. Katz, *Principles of Statistical Mechanics—The Information Theory Approach,* San Francisco: Freeman, 1967.
18. R. J. Duffin and C. Zener, Geometric programming, chemical equilibrium, and the anti-entropy function, *Proc. Nat. Acad. Sciences,* **63**: 629, 1969.
19. R. J. Duffin and C. Zener, Geometric programming and the Darwin-Fowler method in statistical mechanics, *J. Phys. Chem.,* **74**: 2419, 1970.
20. S. Kullback, *Information and Statistics,* New York: Wiley, 1959.
21. M. Avriel, M. J. Rijckaert, and D. J. Wilde (eds.), *Optimization and Design,* Englewood Cliffs, NJ: Prentice-Hall, 1973.
22. R. E. D. Woolsey and H. S. Swanson, *Operations Research for Immediate Application—A Quick and Dirty Manual,* New York: Harper & Row, 1975.
23. C. S. Beightler and D. T. Phillips, *Applied Geometric Programming,* New York: Wiley, 1976.
24. D. J. Wilde, *Globally Optimal Design,* New York: Wiley Interscience, 1978.
25. R. J. Duffin and E. L. Peterson, Geometric programming with signomials, *J. Opt. Theory & Appl.,* **11**: 3, 1973.
26. R. J. Duffin and E. L. Peterson, The proximity of algebraic geometric programming to linear programming, *J. Math. Programming,* **3**: 250, 1972.
27. R. J. Duffin, Linearizing geometric programs, *SIAM Rev.,* **12**: 211, 1970.
28. M. Avriel and A. C. Williams, Complementary geometric programming, *SIAM J. Appl. Math.,* **19**: 125, 1970.
29. R. J. Duffin and E. L. Peterson, Reversed geometric programs treated by harmonic means, *Indiana Univ. Math. J.,* **22**: 531, 1972.
30. R. J. Duffin, Nonlinear networks, IIa, *Bull. Amer. Math. Soc.,* **53**: 963, 1947.
31. E. L. Peterson, Symmetric duality for generalized unconstrained geometric programming, *SIAM J. Appl. Math.,* **19**: 487, 1970.
32. E. L. Peterson, Geometric programming, *SIAM Rev.,* **19**: 1, 1976.
33. M. Avriel (ed.), *Advances in Geometric Programming,* New York: Plenum Press, 1980.
34. S. C. Fang and C. Scott (eds.), *Geometric Programming and Its Generalizations, with Special Emphasis on Entropy Optimization—A Special Issue of the Annals of Operations Research,* in press.

ELMOR L. PETERSON
North Carolina State University
GERHARD PLENERT
Brigham Young University