

## MA/OR 706 Nonlinear Programming - Spring, 2000

- Prerequisites: Elementary linear algebra (including orthogonal complementary subspaces in  $\mathbb{R}^n$ ) and elementary real analysis, at levels normally presented in single-semester introductions. Linear programming is not a prerequisite.
- Content: A description of various classes of nonlinear optimization problems and equilibrium problems that arise in engineering, the mathematical, physical and social sciences, and operations research. Unified formulations within the relatively abstract framework of generalized geometric programming and related complementarity, variational inequality and fixed point frameworks. The convexity, monotonicity, and conjugate-transformation theory needed to develop basic duality relations and fundamental existence, uniqueness and characterization theorems. Emphasis on problem formulation and basic theory, but also descriptions of relevant algorithmic and computational issues. Bifunction parametric formulations and post-optimality analyses if time permits.
- Text: Lecture notes made available by the instructor.
- Instructor: E. L. Peterson, Harrellson Hall 232  
Office hours: Immediately after class or by appointment  
Phone: 515-4859 (to leave your name, a number at which you can be reached, and a short message) and also 881-9194 (when a quicker response is needed or on a weekend).  
E-mail: elpeters@eos.ncsu.edu (only when a single-sentence response would suffice).
- Grades: Normally determined by two midterm exams and a final exam that counts twice. In terms of the average of the three highest of these four exam scores, the transition from A to B is an average of approximately 90%, from B to C 80%, and so forth. In borderline cases, improving exam scores during the semester coupled with a relatively high final exam score and relatively strong performances on the homework (if graded) results in the higher letter grade.

### Course Outline

- A. Some important problem classes.
1. Generalized polynomial optimization (in engineering design and queuing theory)
  2. Entropy optimization (in statistical inference, communication and information theory, statistical mechanics and thermodynamics)
  3. Quadratic and  $L_p$  optimization (in regression analysis, portfolio selection, and general numerical nonlinear optimization)
  4. Optimal location (of public and private facilities, as well as political opinions and decisions)
  5. Optimal allocation (of goods and services)

6. Traffic assignment (in transportation networks and communication networks)
  7. Optimal control with linear dynamics (or dynamic programming with linear transition equations)
  8. Monotone network equilibrium (for electric and hydraulic networks)
  9. Traffic equilibrium (of the Wardrop type for congested roadway networks)
  10. Economic equilibrium (for exchange models and spatial network models)
- B. Some important problem reformulations
1. Generalized geometric programming (for inducing separability and, sometimes, convexity)
  2. Generalized complementarity (for the fundamental connection between optimization and equilibration -- providing criticality conditions for optimization problems and, when possible, helpful optimization reformulations and variational principles for equilibrium problems)
  3. Variational inequality and fixed point reformulations (for analyzing equilibrium problems that have no helpful optimization reformulations and variational principles)
- C. Some relevant background theory
1. Convexity and monotonicity
  2. Conjugate transformation
- D. Fundamental optimization theory
1. Criticality conditions in generalized geometric programming
    - a. The special case for orthogonal complementary subspaces
    - b. Further specialization to, and comparison with, the traditional definition of critical solutions
  2. Duality relations in generalized geometric programming
    - a. Relevant generalized Lagrangians
    - b. Construction of the dual problem
    - c. Duality inequality
    - d. Duality gaps
    - e. Generalized complementarity and criticality
  3. Existence of feasible solutions and optimal solutions
  4. Uniqueness of feasible solutions and optimal solutions
  5. Optimal solution characterizations
- E. Fundamental equilibration theory
1. Integrability conditions for generalized complementarity problems
    - a. Optimization formulations and variational principles
    - b. Dual optimization formulations and dual variational principles
- F. Some algorithmic approaches
1. Relevant types of algorithmic convergence
  2. Decomposition principles (if time permits)
- G. Bifunction parametric formulations (if time permits)
1. Post-optimality analyses, equilibrium prices and generalized Lagrange multipliers
  2. The strongest duality and existence theory in the closed convex case
  3. The duality between parameter deletion and suboptimization
    - a. The basic relations between various duality constructions
    - b. The basic relations between some important decomposition principles

## Geometric Programming

Geometric programming originated in 1961 with Zener's discovery (1, 2, 3, 4, 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 formulas that show how the minimum total cost and associated optimal design depend on the design parameters [such as the unit material costs and power costs, which are determined externally and hence can not 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 analogue of the "dual variational principles" that characterize the "network duality" originating from the two "Kirchoff laws" and "Ohm's laws" (originally discovered in the context of electrical networks). 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, Peterson and Zener (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

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

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  $P(t) \rightarrow +\infty$  as either  $t \rightarrow 0^+$  or  $t \rightarrow +\infty$ .

### Generalized polynomials

The “objective function”  $P(t)$  defined by equation (1) is an example of a “generalized polynomial”  $P(t) = \sum_i^n T_i$  -- namely, a sum of terms  $T_i = c_i \prod_j^m t_j^{a_{ij}}$ , each of which is a given coefficient  $c_i$  [usually determined by design parameters] multiplied into a product  $\prod_j^m t_j^{a_{ij}}$  of the independent design variables  $t_j$  raised to appropriate powers  $a_{ij}$ , termed exponents. In the “single-variable” generalized polynomial (1), the independent design variable  $t$  is the single scalar variable  $t_1$  while the exponents  $a_{11} = 1$  and  $a_{21} = -1$ . In the “multi-variable” 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  $a_{11} = -1$ ,  $a_{12} = 2$ ,  $a_{21} = -1/2$  and  $a_{22} = -3$ . Since non-integer 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  $P(t) = \sum_{i=1}^n T_i$ . Such generalized polynomials  $P(t)$ , including our example (1), are termed “posynomials” and are relatively easy to minimize via geometric programming. Since posynomials  $P(t)$  are obviously “unbounded from above” (because any term  $T_i = c_i \prod_{j=1}^m t_j^{a_{ij}}$  will approach  $+\infty$  as some variable  $t_j$  approaches  $0^+$  or  $+\infty$  while all other terms remain positive), the maximization of posynomials is trivial and tends not to occur in the real world. 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 either minimize or maximize.

### 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 log of a posymonomial  $c_i \prod_{j=1}^m t_j^{a_{ij}}$  is a linear function  $\ln c_i + \sum_{j=1}^m a_{ij} \ln t_j$  of the logs,  $\ln t_j$ , of its independent design variables  $t_j$  -- which makes it relatively easy to use in analytically approximating empirically determined relations (via “linear regression

analysis”). 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 equation (1), is to solve the “optimality condition”  $dP/dt = 0$  for  $t$ ; that is, solve

$$c_1 - c_2 t^{-2} = 0. \quad (2)$$

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”]

$$P^* = P(t^*) = (c_1 c_2)^{1/2} + (c_1 c_2)^{1/2} = 2(c_1 c_2)^{1/2} = 2(CL DLRI^2)^{1/2} = 2LI(CDR)^{1/2}.$$

However, more complicated “posynomial minimization problems” [with more terms  $T_i$  and/or more independent variables  $t_j$ ] usually can not be solved analytically by solving the appropriate optimality condition -- namely,  $dP/dt = 0$  in the single-variable case, or its multi-variable 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  $c_i$  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 dependencies that designers and other decision makers are normally very interested in.

### The geometric programming approach

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

$$c_1 t^1 - c_2 t^{-1} = 0, \quad (3)$$

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 formulas for differentiating and multiplying posynomials]

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

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

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

$$T_1 - T_2 = 0. \quad (5)$$

Since this optimality condition (5) 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

$$T_1/P + T_2/P = 1. \quad (6)$$

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

$$y_1 = T_1/P \text{ and } y_2 = T_2/P, \quad (7)$$

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, equation (5) divided by  $P > 0$ , and equation (6) show that these optimal ratios  $y_1$  and  $y_2$  satisfy both the “orthogonality condition”

$$y_1 - y_2 = 0 \quad (8)$$

and the “normality condition”

$$y_1 + y_2 = 1. \quad (9)$$

[It is worth noting here that the use of geometric concepts such as the “vector-space orthogonality” (8) is partly the origin of the terminology “geometric programming”.]

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

$$y^*_1 = y^*_2 = 1/2, \quad (10)$$

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  $(1/2, 1/2)$  while solving for the optimal value  $P^*$  and optimal solution  $t^*$  via the equations

$$y^*_1 = c_1 t^{1/P} \quad \text{and} \quad y^*_2 = c_2 t^{-1/P}, \quad (11)$$

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

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

$$\begin{aligned} \ln P &= \ln (c_1/y^*_1) + \ln t \\ \ln P &= \ln (c_2/y^*_2) - \ln t. \end{aligned} \quad (12)$$

This system (12) is most easily solved by first solving for  $\ln P$  -- simply by multiplying both sides of its two equations by  $y^*_1$  and  $y^*_2$  respectively and then adding the results to get

$$(y^*_1 + y^*_2) \ln P = y^*_1 \ln (c_1/y^*_1) + y^*_2 \ln (c_2/y^*_2) + (y^*_1 - y^*_2) \ln t, \quad (13)$$

which reduces to

$$\ln P = y^*_1 \ln (c_1/y^*_1) + y^*_2 \ln (c_2/y^*_2) \quad (14)$$

by virtue of the normality condition (9) and the orthogonality condition (8).

Exponentiation of both sides of this equation (14) shows that

$$P^* = (c_1/y^*_1)^{y^*_1} (c_2/y^*_2)^{y^*_2} = 2(c_1 c_2)^{1/2} = 2LI(CDR)^{1/2}, \quad (15)$$

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

$$y^*_1 \ln (c_1/y^*_1) + y^*_2 \ln (c_2/y^*_2) = \ln (c_1/y^*_1) + \ln t \quad (16)$$

$$y^*_1 \ln (c_1/y^*_1) + y^*_2 \ln (c_2/y^*_2) = \ln (c_2/y^*_2) - \ln t,$$

which is “over-determined” with individual solutions

$$\ln t = y^*_1 \ln (c_1/y^*_1) + y^*_2 \ln (c_2/y^*_2) - \ln (c_1/y^*_1) \quad (17)$$

and

$$\ln t = \ln (c_2/y^*_2) - y^*_1 \ln (c_1/y^*_1) - y^*_2 \ln (c_2/y^*_2)$$

respectively. Exponentiation of both sides of these equations (17) gives

$$t^* = (c_1/y^*_1)^{y^*_1} (c_2/y^*_2)^{y^*_2} (y^*_1/c_1) = (c_2/c_1)^{1/2} = I(DR/C)^{1/2}$$

and

$$t^* = (c_2/y^*_2)(y^*_1/c_1)^{y^*_1} (y^*_2/c_2)^{y^*_2} = (c_2/c_1)^{1/2} = I(DR/C)^{1/2} \quad (18)$$

respectively, which shows that the over-determined system (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 a **system of nonlinear equations**  $P = 0$ ] to determine  $t^*$  -- because the

geometric programming transformations (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 (1) consists of maximizing  $(c_1/y_1)^{y_1} (c_2/y_2)^{y_2}$  subject to the linear “orthogonality condition” (8), the linear “normality condition” (9), and the linear “positivity conditions”  $y_1 > 0$  and  $y_2 > 0$ . Since this maximization problem has a unique “dual feasible solution”  $y = (1/2, 1/2)$ , 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 formula (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 (15) for  $P^*$  that

$$\ln P^* / \ln c_i = y^*_i, \quad i = 1, 2, \quad (19)$$

by virtue of the invariance of  $y^*$  with respect to changes in  $c$ . In essence,  $y^*$  provides a “post-optimal 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 formulas (19) are used to show that

$$P^*/c_i = (P^*/\ln P^*)(\ln P^*/\ln c_i)(\ln c_i/c_i) = (P^*)(y_i^*)(1/c_i), \quad i = 1, 2, \quad (20)$$

which in turn implies via the “multi-variable chain rule” that for any design parameter  $p$

$$P^*/p = P^*[(y_1^*/c_1)(c_1/p) + (y_2^*/c_2)(c_2/p)]. \quad (21)$$

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

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

Exercises: 1. Compute formulas for  $P^*/C$ ,  $P^*/R$ ,  $P^*/I$ , and  $P^*/D$ .

A nonelementary logistical example  
[The optimal shipment of gravel]

Given that  $v$  cubic feet of gravel is to be transported across a river via a ferryboat, design a transport box with an open top whose bottom dimensions are  $t_1$  and  $t_2$  and whose height is  $t_3$  (measured in feet). Given that: (1) the bottom and two sides cost  $\$p/\text{ft}^2$ , (2) the two ends cost  $\$q/\text{ft}^2$ , (3) a round-trip on the ferry costs  $\$r$ , and (4) the salvage value of the box is  $\$0$ , determine an optimal design  $t = (t_1, t_2, t_3)$  that minimizes the total capital costs plus total transport costs, namely

$$P(t) = pt_1t_2 + 2pt_1t_3 + 2qt_2t_3 + r(v/t_1t_2t_3)$$

or

$$P(t) = c_1t_1t_2 + c_2t_1t_3 + c_3t_2t_3 + c_4t_1^{-1}t_2^{-1}t_3^{-1} \quad \text{where } c_1 = p, c_2 = 2p, c_3 = 2q, c_4 = rv.$$

Exercises: 1. Can you spot a simplifying assumption used in the formula for  $P(t)$ ? We will eventually see that this simplification does not negate the existence of an optimal design  $t^*$  for this simplified model. In fact, this  $t^*$  will be used to formulate a subsequent model whose optimal solution  $t^*$  will be a more accurate approximation to the true optimal solution.

2. Give the coefficient vector  $c$  and exponent matrix  $A$  for  $P(t)$ . Why is  $P$  a posynomial?

The elementary calculus approach to minimizing this posynomial  $P(t)$  is to first find a “critical solution”  $t$ , at which its gradient  $\nabla P(t) = 0$ ; in particular, find a solution  $(t_1, t_2, t_3)$  for which

$$c_1t_2 + c_2t_3 - c_4t_1^{-2}t_2^{-1}t_3^{-1} = 0$$

$$c_1t_1 + c_3t_3 - c_4t_1^{-1}t_2^{-2}t_3^{-1} = 0$$

$$c_2t_1 + c_3t_2 - c_4t_1^{-1}t_2^{-1}t_3^{-1} = 0$$

This is a nonlinear system that is not easily solved for the optimal design

$t^* = (t_1^*, t_2^*, t_3^*)$  even when numerical values are given for  $c$  (by specifying numerical values for  $p, q, r$ , and  $v$ ).

The geometric programming approach is to multiply the preceding equations by the unknowns  $t_1, t_2$  and  $t_3$  respectively and then divide each side by the unknown  $P(t) > 0$  to get the linear system

$$y_1 + y_2 - y_4 = 0$$

$$y_1 + y_3 - y_4 = 0 \quad \text{the "orthogonality conditions"}$$

$$y_2 + y_3 - y_4 = 0$$

$$y_1 + y_2 + y_3 + y_4 = 1 \quad \text{the "normality condition"}$$

where the new unknowns

$$y_1 = c_1 t_1 t_2 / P(t) \quad y_2 = c_2 t_1 t_3 / P(t) \quad y_3 = c_3 t_2 t_3 / P(t) \quad y_4 = c_4 t_1^{-1} t_2^{-1} t_3^{-1} / P(t)$$

can be interpreted as the fractional parts of the unknown objective value  $P(t)$  due to its unknown terms at optimality. The normality condition comes from this interpretation, and the orthogonality conditions constrain  $y$  to the "orthogonal complement" of the "column space" of the exponent matrix  $A$ .

Gaussian row reduction shows that the preceding linear system has a unique solution

$$y^*_1 = y^*_2 = y^*_3 = 1/5, y^*_4 = 2/5.$$

This solution shows that an optimally designed transport box always produces capital costs for which the bottom cost equals both the two side costs and the two end costs and for which the total capital costs and transport costs are in the ratio of 3 to 2 ("necessary conditions" for optimality that might not be "sufficient").

To find the solutions  $P^*$  and  $t^*$  for  $P$  and  $t$  respectively, take the logarithm of both sides of the defining equations for  $y$  to obtain

$$\ln P = \ln(c_1/y^*_1) + \ln t_1 + \ln t_2$$

$$\ln P = \ln(c_2/y^*_2) + \ln t_1 + \ln t_3$$

$$\ln P = \ln(c_3/y^*_3) + \ln t_2 + \ln t_3$$

$$\ln P = \ln(c_4/y^*_4) - \ln t_1 - \ln t_2 - \ln t_3$$

Multiplying these equations by  $y^*_1$ ,  $y^*_2$ ,  $y^*_3$ , and  $y^*_4$  respectively and then adding the

results gives

$$(y^*_1 + y^*_2 + y^*_3 + y^*_4)\ln P = y^*_1\ln(c_1/y^*_1) + y^*_2\ln(c_2/y^*_2) + y^*_3\ln(c_3/y^*_3) +$$

$$y^*_4\ln(c_4/y^*_4) +$$

$$(y^*_1 + y^*_2 - y^*_4)\ln t_1 +$$

$$(y^*_1 + y^*_3 - y^*_4)\ln t_2 +$$

$$(y^*_2 + y^*_3 - y^*_4)\ln t_3,$$

which becomes

$$\ln P = y^*_1\ln(c_1/y^*_1) + y^*_2\ln(c_2/y^*_2) + y^*_3\ln(c_3/y^*_3) + y^*_4\ln(c_4/y^*_4)$$

by virtue of the normality and orthogonality conditions. Exponentiation now shows that

$$P^* = (c_1/y^*_1)^{y^*_1}(c_2/y^*_2)^{y^*_2}(c_3/y^*_3)^{y^*_3}(c_4/y^*_4)^{y^*_4}.$$

$$= (5/2^{2/5})c_1^{1/5}c_2^{1/5}c_3^{1/5}c_4^{2/5} = 5p^{2/5}q^{1/5}r^{2/5}v^{2/5}.$$

To find  $t^*$ , substitute this  $P^*$  into the preceding system, which you will note is linear in

$\ln t$ , namely  $(\ln t_1, \ln t_2, \ln t_3)$ . Gaussian row reduction of this linear system produces the

equivalent linear system

$$\begin{aligned}
\ln t_1 &= [1/2]\ln(P^*[y^*_1/c_1][y^*_2/c_2][c_3/y^*_3]) \\
\ln t_2 &= [1/2]\ln(P^*[y^*_1/c_1][c_2/y^*_2][y^*_3/c_3]) \\
\ln t_3 &= [1/2]\ln(P^*[c_1/y^*_1][y^*_2/c_2][y^*_3/c_3]) \\
0 &= 0.
\end{aligned}$$

Exponentiation now shows that

$$\begin{aligned}
t^*_1 &= (P^*[y^*_1/c_1][y^*_2/c_2][c_3/y^*_3])^{1/2} = p^{-4/5}q^{3/5}r^{1/5}v^{1/5} \\
t^*_2 &= (P^*[y^*_1/c_1][c_2/y^*_2][y^*_3/c_3])^{1/2} = p^{1/5}q^{-2/5}r^{1/5}v^{1/5} \\
t^*_3 &= (P^*[c_1/y^*_1][y^*_2/c_2][y^*_3/c_3])^{1/2} = (1/2)p^{1/5}q^{-2/5}r^{1/5}v^{1/5}.
\end{aligned}$$

Since  $\ln P^* = \sum_{i=1}^4 y^*_i \ln c_i - \sum_{i=1}^4 y^*_i \ln y_i$ , and since  $y^*$  is determined only by  $A$  and

hence is a fixed constant relative to  $c$ , differentiation shows that  $(\ln P^*)/(\ln c_i) = y^*_i$  for

$i = 1, 2, 3, 4$ ; so the chain rule

$(P^*)/(c_i) = [(P^*)/(\ln P^*)][(\ln P^*)/(\ln c_i)][(\ln c_i)/(c_i)]$  implies that

$$(P^*)/(c_i) = [P^*][y_i][1/c_i] \text{ for } i = 1, 2, 3, 4.$$

Consequently, the multivariable chain rule implies that

$$(P^*)/(p) = P^* \sum_{i=1}^4 [y_i/c_i][(\ln c_i)/(p)] = 2p^{-3/5}q^{1/5}r^{2/5}v^{2/5}.$$

Exercises: 1. Compute formulas for  $(P^*)/(q)$ ,  $(P^*)/(r)$ , and  $(P^*)/(v)$ .

Unconstrained posynomial minimization via geometric programming  
[A summary of the zero degree-of-difficulty case]

If the posynomial  $P(t) = \sum_{i=1}^n c_i \prod_{j=1}^m t_j^{a_{ij}}$  is minimized by some  $t^* > 0$ , then

$P(t^*) = 0$ ; so

and  $\sum_{j=1}^m a_{ij} y_i = 0$  for  $j = 1, 2, \dots, m$  the “orthogonality conditions”  
 and  $\sum_{i=1}^n y_i = 1$  the “normality condition”

when the new “dual variables”

$$y_i = c_i \sum_{j=1}^m t_j^{a_{ij}} / P(t^*) \text{ for } i = 1, 2, \dots, n.$$

In the zero degree-of-difficulty case, the preceding orthogonality and normality conditions have a unique solution  $y^* > 0$ , which can be obtained via elementary linear algebra [normally “elementary row operations”] applied to the corresponding “augmented matrix”. The optimal value  $P^*$  and optimal solution  $t^*$  can then be obtained from  $y^*$  by solving the defining system

$$y_i^* = c_i \sum_{j=1}^m t_j^{a_{ij}} / P(t) \text{ for } i = 1, 2, \dots, n$$

for  $P$  and  $t$ . The solution to this nonlinear system can also be obtained via elementary linear algebra -- applied to the corresponding equivalent system

$$\ln P = \ln(c_i/y_i^*) + \sum_{j=1}^m a_{ij} \ln t_j \text{ for } i = 1, 2, \dots, n,$$

which is clearly linear in the unknowns  $\ln P$  and  $\ln t_j$ ,  $j = 1, 2, \dots, m$ . This solution is most easily accomplished by first multiplying each side of the preceding equations by the respective values  $y_i^*$  and then adding the resulting equations to get

$$\left( \sum_{i=1}^n y_i^* \right) \ln P = \sum_{i=1}^n y_i^* \ln(c_i/y_i^*) + \sum_{j=1}^m \left( \sum_{i=1}^n a_{ij} y_i^* \right) \ln t_j,$$

which reduces to

$$\ln P = \sum_{i=1}^n y_i^* \ln(c_i/y_i^*)$$

by virtue of the normality and orthogonality conditions. Exponentiation of each side of this equation now shows that

$$P^* = \sum_{i=1}^n (c_i/y_i^*)^{y_i^*}.$$

To solve for  $t^*$ , substitute this solution  $P^*$  for  $P$  into the original system to obtain the “reduced system”

$$\sum_{j=1}^m a_{ij} \ln t_j = \ln P^* + \ln(y_i^*/c_i) \quad \text{for } i = 1, 2, \dots, n,$$

which is clearly linear in the remaining unknowns  $\ln t_j$ ,  $j = 1, 2, \dots, m$  and can be solved via elementary linear algebra applied to the corresponding augmented matrix. This reduced system is normally “overdetermined” because  $n$  is normally greater than  $m$ ; but it must, of course, have a solution  $t^*$  if  $P(t)$  is minimized for some  $t^* > 0$ . Furthermore, the elementary theory of linear algebra asserts that such a solution  $t^*$  is unique if, and only if, the system coefficient matrix [namely the exponent matrix  $A$ ] has “full column rank” [that is, the columns of  $A$  are linearly independent]. Otherwise, infinitely many optimal  $t^*$ 's are obtained [a fact that is not easily seen without the geometric programming approach].

Since  $\ln P^* = \sum_{i=1}^n y_i^* \ln(c_i/y_i^*) = \sum_{i=1}^n y_i^* \ln c_i - \sum_{i=1}^n y_i^* \ln y_i^*$  and since  $y^*$  does not depend on the coefficient vector  $c$ ,  $(\ln P^*)/(c_i) = y_i^*$  for  $i = 1, 2, \dots, n$ ; so the chain rule implies that  $(P^*)/(c_i) = [P^*][y_i^*][1/c_i]$  for  $i = 1, 2, \dots, n$ , and hence the multivariable chain rule implies that

$$(P^*)/(u) = P^* \sum_{i=1}^n [y_i^*/c_i][c_i/(u)]$$

for any “design parameter”  $u$  that helps to determine at least one coefficient  $c_i$ . We shall eventually see that this formula for  $(P^*)/(u)$  is also valid when the degree of difficulty is positive [in which case the orthogonality and normality conditions have infinitely many solutions  $y > 0$  and the appropriate  $y^*$  does depend on  $c$ ].

The optimal shipment of gravel  
[A refined and more accurate model]

The number of ferry round-trips prescribed by the previous analysis is of course

$$v/t_1^*t_2^*t_3^* = 2p^{2/5}q^{1/5}r^{-3/5}v^{2/5},$$

which is generally not an integer and hence would have to be replaced by either the largest smaller integer or the smallest larger integer, say  $s$  for either case. This means, of course, that the “optimal solution”  $t^*$  previously given is only approximate and should be replaced by the optimal solution  $t^*$  to the problem of minimizing just the capital costs given that  $s$  trips are made with a box whose volume is “constrained” to be  $v/s$ . In particular:

Minimize just the capital cost  $\underline{P}(t) = c_1t_1t_2 + c_2t_1t_3 + c_3t_2t_3$

subject to the “constraint”  $\underline{c}_4t_1^{-1}t_2^{-1}t_3^{-1} = 1$

where  $c_1 = p$ ,  $c_2 = 2p$ ,  $c_3 = 2q$ ,  $\underline{c}_4 = v/s$ .

The constraint renders this posynomial minimization problem different from the examples previously treated, unless the constraint is solved for one of the design variables, say  $t_3$ , in terms of the other two design variables, and  $\underline{P}$  is reformulated in terms of those two design variables.

Exercises: 1. Minimize the reformulated posynomial  $\underline{P}$  via the geometric programming methodology previously given. Also, find the “sensitivity” of the resulting  $\underline{P}^*$  to  $p$ ,  $q$ ,  $r$ , and  $v$  via differentiation.

2. Give a heuristic method for determining the true optimal solution from a knowledge of the optimal solutions for the two different values of  $s$  [which differ by 1].

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

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

over its natural domain

$$T = \{t \in \mathbb{R}^m \mid t > 0\}, \quad (22b)$$

the “feasible solution set” for unconstrained posynomial minimization. Since there need not be an optimal solution  $t^*$ , this minimization actually consists of finding the “problem infimum”

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

which is used to define the “optimal solution set”

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

Although  $P^* = 2(c_1 c_2)^{1/2}$  and  $T^*$  contains a single point  $t^* = (c_2/c_1)^{1/2}$  for our power-line

example  $P(t) = c_1 t^1 + c_2 t^{-1}$ ,  $P^*$  is clearly 0 and  $T^*$  is clearly empty when either

$P(t) = c_1 t^1$  or  $P(t) = c_2 t^{-1}$  [because of the restrictions  $0 < t < \infty$ , which are enforced in

order to keep  $t$  within the domain of  $\ln t$ , so that the geometric programming

transformations previously described are applicable]. The detection and treatment of

posynomial minimization problems (22, 23) for which  $T^*$  is empty [because some

optimal  $t^*$ ; is either 0 or  $+\infty$ ] is usually not needed, because well-posed realistic models

normally do not imply “extreme” optimal designs [namely those involving 0 or  $+\infty$ ].

However, the detection and treatment of such “degenerate problems” is described in

references (9, 10).

Transformations

The key roles played by  $\ln t$  and  $\ln P$  in the geometric programming solution of our power-line and gravel-shipment examples suggests making the transformation defined by the following change of variables

$$z_j = \ln t_j, \quad j = 1, 2, \dots, m \quad (24a)$$

and

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

Since the log function is “monotone increasing” with range  $\mathbb{R}$ , 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

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

and

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

In particular,

$$P^* = \exp(p^*) \quad (26a)$$

and

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

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

$$x_i = \sum_{j=1}^m a_{ij} z_j, \quad i = 1, 2, \dots, n, \quad (27a)$$

and

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

Since  $x$  ranges over the “vector space”

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

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

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

and

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

even when the linear transformation  $z \mapsto x = Az$  is not “one-to-one” [i.e., when the exponent matrix  $A$  does not have full column rank]. In particular,

$$p^* = g^* \quad (30a)$$

and

$$Z^* = \{z \in \mathbb{R}^m \mid Az = x \text{ for some } x \in X^*\}. \quad (30b)$$

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

$$g^* = \inf_{x \in X} \ln \left[ \prod_{i=1}^n c_i \exp(x_i) \right] \quad (31a)$$

and corresponding optimal solution set

$$X^* = \{x \in X \mid \ln \left[ \prod_{i=1}^n c_i \exp(x_i) \right] = g^*\} \quad (31b)$$

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

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

and corresponding optimal solution set

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

Exercises: Give formulas for  $p(z)$ ,  $g(x)$ ,  $X$ ,  $X^*$ ,  $Z^*$  and  $T^*$  for:

1. the optimal-design-of-electrical-conductors problem,
2. the initial model of the optimal-shipment-of-gravel problem,
3. the refined model of the optimal-shipment-of-gravel problem.

### Existence and uniqueness of optimal solutions

The preceding relation (32b) between the optimal solution sets  $T^*$  and  $X^*$  clearly implies that  $T^*$  is non-empty if, and only if,  $X^*$  is non-empty. Moreover, the “strict convexity” of the functions  $c_i \exp(x_i)$  in equations (31) implies that  $X^*$  contains at most a single  $x^*$ . Consequently, relation (32b) shows that  $T^*$  contains at most a single  $t^*$ ; unless  $z^* = Ax^*$  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

$$\sum_{j=1}^m a_{ij} \ln t_j = x_i^*, \quad i = 1, 2, \dots, n. \quad (33)$$

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 and gravel-shipment examples and would normally be the case for a properly modeled problem from the real world], relation (26b) implies that  $Z^*$  contains at least one  $z^*$ . Moreover, since the defining formula (24b) for the objective function  $p(z)$  in the associated minimization problem (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,

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

In view of relations (27a) and (30b), these “optimality conditions” (34) for the problem formulation (24b, 25) imply that

$$\{ \prod_{i=1}^n [c_i \exp(x^*_i)] \}^{-1} \{ \prod_{i=1}^n [c_i \exp(x^*_i)] a_{ik} \} = 0, \quad k = 1, 2, \dots, m, \quad (35)$$

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

$$y^*_i = [c_i \exp(x^*_i)] / \{ \prod_{i=1}^n [c_i \exp(x^*_i)] \}, \quad i = 1, 2, \dots, n, \quad (36)$$

satisfies the conditions

$$\prod_{i=1}^n a_{ik} y_i = 0, \quad k = 1, 2, \dots, m, \quad \text{orthogonality conditions} \quad (37a)$$

$$\prod_{i=1}^n y_i = 1, \quad \text{normality condition} \quad (37b)$$

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

with the positivity conditions satisfied because each posynomial coefficient  $c_i > 0$  and each  $\exp(x^*_i) > 0$ . Conversely, references (9, 10) show that when 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 equations (36) a  $y^*$  that is a solution, but not necessarily the only solution, to linear system (37)]. Moreover, references (9, 10) also show that every “nontrivial” posynomial minimization problem (22, 23) can be “reduced” to an “equivalent” posynomial minimization problem whose “dual constraints” (37) can be satisfied. Consequently, **posynomial minimization problems whose dual constraints (37) can**

be satisfied are termed “canonical problems”; and canonical problems, and only canonical problems, have non-empty optimal solution sets  $T^*$ ,  $Z^*$  and  $X^*$ .

Degree of difficulty

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

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

is not negative. In fact, in the canonical case, if  $d = 0$ , linear algebra implies that the dual constraints (37) have a unique solution -- namely, the vector  $y^*$  defined by equations (36). Moreover, in the canonical case, if  $d > 0$ , linear algebra and elementary topology imply that the dual constraints (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 equations (36) can be obtained via only elementary linear algebra** -- as in our power-line example (1) and gravel-shipment examples. On the other hand, **if  $d > 0$  in the canonical case, the vector  $y^*$  defined by equations (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 (22, 23) or one of its equivalent reformulations (24b, 25) or (27b, 28, 29) -- or via a numerical solution of their “dual problem”** [which has been described for the power-line example (1) and gravel-shipment example, but is not generally defined until a later subsection]. Actually, posynomial minimization problem (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 references (8, 9, 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 (22, 23) and its equivalent reformulations (24b, 25) and (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 relations (22a) and (33) imply that equations (36) can be rewritten as

$$y^*_i = [c_i \exp(x^*_i)]/P^*, \quad i = 1, 2, \dots, n, \quad (39)$$

which show 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 equations (7) for our power-line example (1). Now, take the logarithm of both sides of equations (39) to get

$$\ln P^* = \ln(c_i/y^*_i) + x^*_i, \quad i = 1, 2, \dots, n; \quad (40)$$

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

$$\left( \sum_{i=1}^n y_i^* \right) \ln P^* = \sum_{i=1}^n y_i^* \ln (c_i/y_i^*) + \sum_{i=1}^n x_i^* y_i^*, \quad (41)$$

which reduces to

$$\ln P^* = \sum_{i=1}^n y_i^* \ln (c_i/y_i^*) \quad (42)$$

because  $y^*$  satisfies the normality condition (37b) and because  $x^*$  and  $y^*$  are orthogonal by virtue of the transformation equations (27a) and the orthogonality conditions (37a).

Needless to say, exponentiation of both sides of this equation (42) gives the desired optimal value

$$P^* = \sum_{i=1}^n (c_i/y_i^*)^{y_i^*}; \quad (43)$$

and substituting formula (42) for  $\ln P^*$  back into equations (40) gives the optimal

$$x_i^* = \left[ \sum_{i=1}^n y_i^* \ln (c_i/y_i^*) \right] - \ln (c_i/y_i^*), \quad i = 1, 2, \dots, n, \quad (44)$$

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

Some real-world problems to which the preceding theory can be applied originate with a need to solve a problem modeled directly by the dual constraints (37) [rather than by the primal posynomial minimization problem (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] 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 it’s “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

$$\sum_{j=1}^n p_{ij} = 1 \quad (45a)$$

$$p_{ij} \geq 0, \quad j = 1, 2, \dots, n. \quad (45b)$$

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

$$yP = y \quad (46a)$$

$$\sum_{i=1}^n y_i = 1 \quad (46b)$$

has a unique solution  $y^*$  that also satisfies

$$y > 0. \quad (46c)$$

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 system

(46a, b). In such cases, an “iterative approach” based on the preceding geometric programming theory might be successful.

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

$$a_{ij} = \begin{cases} p_{ij} - 1 & \text{if } i = j \\ p_{ij} & \text{if } i \neq j. \end{cases} \quad (4)$$

Since these dual constraints have a unique solution, namely  $y^* > 0$ , the corresponding transformed posynomial minimization problem (24b, 25) has zero degree-of-difficulty and is canonical [as defined and discussed following the dual constraints (37)]; so problem (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 equations (45a) and (47)], each  $z^*$  provides the desired equilibrium distribution  $y^*$  via the formula

$$y^*_i = c_i \exp \left( - \sum_{j=1}^m a_{ij} z^*_j \right) / \sum_{k=1}^m c_k \exp \left( - \sum_{j=1}^n a_{kj} z^*_j \right), \quad i = 1, 2, \dots, n, \quad (48)$$

which comes from combining formulas (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  $\ln \left[ \sum_{i=1}^n c_i \exp \left( - \sum_{j=1}^n a_{ij} z_j \right) \right]$  for minimization problem (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 these notes.

Unconstrained posynomial minimization via geometric programming  
[The positive degree-of-difficulty case]

In the gravel-shipment optimization problem, suppose that the transport-box height  $t_3$  has to be 1 [because it has to be transported below deck]. Then the posynomial to be minimized is clearly

$$P(t) = pt_1t_2 + 2pt_1 + 2qt_2 + r(v/t_1t_2)$$

or

$$P(t) = c_1t_1t_2 + c_2t_1 + c_3t_2 + c_4t_1^{-1}t_2^{-1} \quad \text{where } c_1 = p, c_2 = 2p, c_3 = 2q, c_4 = rv.$$

Note that the coefficient vector  $c$  has not changed, but the exponent matrix  $A$  has lost its third column; so the rank of  $A$  has decreased from 3 to 2, and hence the degree-of-difficulty has increased from 0 to 1. However, the vector  $y^* = (1/5, 1/5, 1/5, 2/5)$  clearly satisfies, a fortiori, the smaller set of orthogonality conditions, but there are now infinitely many such “dual feasible solutions”  $y$  [that also satisfy the normality and positivity conditions]. Since this modified problem is canonical [because of the existence of its dual feasible solution  $y^* > 0$ ], our theory asserts the existence of an optimal solution  $t^*$  for it. Note, however, that the previous optimal solution  $t^*$  [obtained from  $y^*$  and for which  $t^*_3 = (1/2)p^{1/5}q^{-2/5}r^{1/5}v^{1/5}$ ] can not be optimal for the modified problem when  $p = q = r = v = 1$  [because  $t^*_3 = (1/2)$  is not even feasible for the modified problem]. Consequently, the appropriate dual optimal solution  $y^*$  for the modified problem must be one of the infinitely many other dual feasible solutions  $y$ , but which one?

- Exercises:
1. Use your knowledge of linear algebra to obtain formulas for all dual feasible solutions  $y$  corresponding to the modified gravel-shipment optimization problem.
  2. In the power-line example, suppose that power consumption is to be encouraged via a tax that is proportional to the square of the power consumed. For the resulting problem of minimizing  $P(t) = c_1 t^1 + c_2 t^{-1} + c_3 t^{-2}$  for the given coefficients  $c_1 = CL$ ,  $c_2 = DLRI^2$  and  $c_3 = ED^2L^2R^2I^4$  [where  $E$  is the appropriate proportionality constant]: (a) give the exponent matrix, (b) give the degree of difficulty, (c) obtain formulas for all dual feasible solutions  $y$ .

We shall now see that the function  $\prod_1^n (c_i/y_i)^{y_i}$  appearing in the key equation (43) for determining  $P^*$  in the zero degree-of-difficulty case must be maximized over all dual feasible solutions  $y$  [infinite in number in the positive degree-of-difficulty case] to obtain the appropriate dual feasible solution  $y^*$  that makes  $\prod_1^n (c_i/y_i^*)^{y_i^*} = P^*$ . We shall also see that such a dual feasible solution  $y^*$ , termed a “dual optimal solution”, can be used to find the appropriate “primal optimal solutions”  $x^*$  and all optimal  $t^*$  (as well as all optimal  $z^*$ ) via the same relations [essentially (44) and (33)] already developed for the zero degree-of-difficulty case.

### 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 (22, 23) [including its equivalent formulations

(24b, 25) and (27b, 28, 29)], the dual feasible solution set consists of all solutions to the dual constraints

$$\sum_{j=1}^m a_{ij} y_j = 0, \quad j = 1, 2, \dots, m, \quad \text{the orthogonality conditions} \quad (49a)$$

$$\sum_{i=1}^n y_i = 1, \quad \text{the normality condition} \quad (49b)$$

$$y_i \geq 0, \quad i = 1, 2, \dots, n, \quad \text{the positivity conditions,} \quad (49c)$$

which differ from the originally motivated dual constraints (37) only in that positivity condition (49c) is a slightly relaxed version of positivity condition (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 relation (43) and is to be maximized, has formula

$$U(y) = \sum_{i=1}^n (c_i/y_i)^{y_i}, \quad (50)$$

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 (22, 23)] and if  $y$  is “dual feasible” [in that  $y$  satisfies the “dual constraints” (49) for the corresponding dual maximization problem (49, 50)], then

$$U(y) = P(t), \quad (51a)$$

with equality holding if, and only if,

$$y_i = [c_i \sum_{j=1}^m t_j^{a_{ij}}] / [\sum_{k=1}^n c_k \sum_{j=1}^m t_j^{a_{kj}}], \quad i = 1, 2, \dots, n; \quad (51b)$$

in which case  $t$  and  $y$  are primal and dual optimal, respectively [and the primal problem (22,23) and its dual problem (49,50) are canonical].

“Duality inequality” (51a) and the corresponding “primal-dual optimality condition” (51b) can be established with the aid of the well-known “Cauchy inequality”

$$\prod_{i=1}^n u_i^{y_i} \leq \prod_{i=1}^n y_i u_i \quad (52)$$

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

$$u_i = T_i/y_i = c_i \prod_{j=1}^m t_j^{a_{ij}}/y_i \quad i = 1, 2, \dots, n, \quad (53)$$

and then employ both the primal constraints  $t_j > 0$  and the dual constraints (49). A

byproduct is that the dual problem (49, 50) has a **unique optimal solution**  $y^*$

[determined via equations (32b) and (36)] when primal problem (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 (52) is partly the origin of the terminology “geometric programming”.]

Also, for canonical problems [of any degree of difficulty], the “implicit function theorem” [from multi-variable advanced calculus] can be used to show that

$$\ln P^*/ \ln c_i = y_i^*, \quad i = 1, 2, \dots, n, \quad (54)$$

which is the basis for “post-optimal sensitivity analyses” in posynomial geometric programming -- as previously illustrated in the zero degree-of-difficulty power-line example (1) and gravel-shipment examples.

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

### Dual reformulations

The dual constraints (49) are linear; so the dual feasible solutions  $y$  can be characterized in various ways via both linear algebra and linear programming. Both types of characterization supply important qualitative and quantitative information about a given problem.

#### Linear-algebraic dual reformulations

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

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

for values of the “basic variables”  $r_j$  for which

$$b^{(0)}_i + \sum_{j=1}^d r_j b^{(j)}_i = 0, \quad i = 1, 2, \dots, n. \quad (55b)$$

The vector  $b^{(0)}$ , which satisfies both the orthogonality and normality conditions (49a, b), is termed a “normality vector”. The vectors  $b^{(j)}$  for  $j = 1, \dots, d$ , which are “linearly independent” solutions to the “homogeneous counterpart” of the orthogonality and normality conditions (49a, b), are called “nullity vectors”. If the degree-of-difficulty

$d = 0$ , 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 normality and nullity 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 (43, 44, 33)], written in terms of the basic variables  $r_j$ , is

$$\begin{aligned} V(r) &= \left\{ \prod_{i=1}^n c_i^{[b^{(0)}_i + \sum_{j=1}^d r_j b^{(j)}_i]} \right\} \left( \prod_{i=1}^n y_i(r)^{-y_i(r)} \right) \\ &= K_0 \left( \prod_{j=1}^d K_j^{r_j} \right) \left( \prod_{i=1}^n y_i(r)^{-y_i(r)} \right) \end{aligned} \quad (56a)$$

where the “basic constants”

$$K_j = \prod_{i=1}^n c_i^{b^{(j)}_i}, \quad j = 0, 1, \dots, d, \quad (56b)$$

and where

$$y_i(r) = b^{(0)}_i + \sum_{j=1}^d r_j b^{(j)}_i, \quad i = 1, 2, \dots, n. \quad (56c)$$

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

$$b^{(0)}_i + \sum_{j=1}^d b^{(j)}_i r_j \geq 0, \quad i = 1, 2, \dots, n. \quad (57)$$

Prior to maximizing  $V(r)$ , useful qualitative information about its optimal value  $V^* \{ = U^* = P^* \}$  can be obtained from the defining formulas (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 dual problem (49, 50) [and hence primal problem (22, 23) and its equivalents] -- in that formula (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  $K_j$ ,  $j = 1, 2, \dots, d$ , are dimension-less. Moreover, for a fixed  $A = (a_{ij})$  [typically fixed by the non-changing laws of geometry, physical science and/or economics], the normality and nullity vectors  $b^{(j)}$  can be fixed independent of the coefficients  $c_i$  [typically not fixed but determined by changing design parameters, such as unit material costs, power costs and design specifications]. The basic constants  $K_j$  are then functions only of the coefficients  $c_i$ ; in fact, each  $\ln K_j$  is a linear function of the  $\ln c_i$ 's, as indicated by taking the log of each side of the defining equation (56b) for  $K_j$ . The resulting equations

$$\sum_{i=1}^n b^{(j)}_i \ln c_i = \ln K_j, \quad j = 0, 1, \dots, d, \quad (58)$$

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 equation (38)] is clearly always greater than the number  $d + 1$  of basic constants  $K_j$ . Each solution  $c$  to linear system (58) determines a different primal problem (22, 23), but the corresponding reformulated dual problems (56, 57) are all the same; so the minimum value  $P^*$  for each of these primal problems is the same even though the primal optimal solutions  $t^*$  are generally different. In summary, the solution of a specific problem (56, 57) [by the maximization of  $V(r)$  for a particular  $K$ ] solves infinitely many posynomial minimization problems (22, 23) [determined by all solutions  $c$  to linear system (58) for the particular  $K$ ].

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

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

Since the 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  $\ln V(r)$  at such an  $r^*$  implies that  $(\ln V)/r_j(r^*) = 0$  for  $j = 1, 2, \dots, d$ , which means that  $\ln K_j - \sum_i^n [\ln y_i^* + 1] b_i^{(j)} = 0$  for  $j = 1, 2, \dots, d$ , and hence that  $\ln K_j = \sum_i^n b_i^{(j)} \ln y_i^*$  for  $j = 1, 2, \dots, d$  [because  $\sum_i^n b_i^{(j)} = 0$  for  $j = 1, 2, \dots, d$ ]. Since  $\ln U$  is a “concave function” of  $y$ , the preceding derivation actually shows that a dual feasible solution  $y > 0$  is, in fact, dual optimal if, and only if,

$$\ln K_j = \sum_i^n b_i^{(j)} \ln y_i, \quad j = 1, 2, \dots, d, \quad (59a)$$

in which case  $\ln U = \ln K_0 - \sum_i^n b_i^{(0)} \ln y_i$  and hence

$$P^* = K_0 \sum_i^n y_i^{-b_i^{(0)}}. \quad (59b)$$

Note that the preceding “maximizing equations” (59a) map each dual feasible solution  $y > 0$  into basic constants  $K_j$  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 (22, 23) with a coefficient vector  $c$  that satisfies the resulting linear system (58).

If the degree-of-difficulty  $d$  is only one, there is only one maximizing equation (59a), one  $K_j = K_1$  and one  $r_j = r_1$ . In that case, treating  $r_1$  as the independent variable and  $\ln K_1$  as the dependent variable and simply graphing the resulting maximizing equation

$$\ln K_1 = \sum_i^n b_i^{(1)} \ln (b_i^{(0)} + b_i^{(1)} r_1)$$

essentially solves all posynomial minimization problems (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 equations (58) for  $\ln K_j$  give a particular

$$\ln K_1 = \sum_{i=1}^n b^{(1)}_i \ln c_i,$$

which determines, via the graph, the corresponding optimal  $r^*_1$  and hence dual optimal solution  $y^* = b^{(0)} + b^{(1)} r^*_1$ ; from which the desired  $P^*$ ,  $x^*$ ,  $z^*$  and  $t^*$  can be determined via the duality equations (43, 44, 33). In retrospect, it is worth noting that the  $r_1$ -versus- $\ln K_1$  graph: (1) always has “range”  $R$  because the range of  $\ln K_1$  in the preceding displayed formula involving the  $\ln c_i$  is clearly always  $R$ , (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 more sophisticated numerical solution technique [such as the Newton-Raphson method] to determine  $r^*$  from a knowledge of the  $\ln K_j$ .

Exercises: Construct the vectors  $b^{(0)}$  and  $b^{(1)}$  and the resulting graph of  $r_1$ -versus- $\ln K_1$  for:

1. the modified power-line example [with power-consumption tax]
2. the modified gravel-shipment example [with  $t_3 = 1$ ].

[Notes: Exercises 1. and 2. do not have unique solutions because the vectors  $b^{(0)}$  and  $b^{(1)}$  are not uniquely determined by their exponent matrix  $A = [a_{ij}]$ .

Moreover, you can expedite your work and increase its accuracy if you use Maple, Mathematica, Matlab or some other mathematical computer software to graph  $r_1$  -versus-  $\ln K_1$  for your choices of  $b^{(0)}$  and  $b^{(1)}$  for each exercise.]

### Linear-programming dual reformulations

These reformulations characterize the dual feasible solutions  $y$  as “convex combinations” of the “basic dual feasible solutions”  $y^k$ ,  $k = 1, 2, \dots, p$ , to the dual feasibility conditions (49). Unlike the normality and nullity vectors  $b^{(j)}$ ,  $j = 0, 1, \dots, d$ , the basic dual feasible solutions  $y^k$  are unique and can be determined from the linear system (49) via “phase I” of the “simplex method” for “linear programming”. In particular, if the dual is consistent [which we assume to be the case], the “terminal tableau” for phase I produces  $y^1$  -- simply by setting the terminal “nonbasic” [or independent] variables equal to zero. Then, after deleting the phase I objective row from the terminal tableau, “pivoting” in each “nonbasic column” with the “pivot row” determined by the usual “minimal ratio” produces other  $y^k$ ,  $k = 2, 3, \dots$ . Moreover, pivoting in each nonbasic column of each new tableau can produce still more  $y^k$ ; but after a finite number of pivots, no new tableaus and hence no new  $y^k$  will be produced. In summary, a finite number  $p$  of basic dual feasible solutions  $y^k$ ,  $k = 1, 2, \dots, p$ , can be computed via the “elementary row operations” used in “pivoting” -- to form the columns of the “basic dual feasible solution matrix”

$$Y = [y^1, y^2, \dots, y^p].$$

Moreover, according to the “resolution theorem” [sometimes called the “decomposition theorem” or “Weyl’s theorem” or “Goldman’s theorem”] for “polytopes” [namely, “bounded polyhedral sets”], each dual feasible solution  $y$  is a “convex combination” of the basic dual feasible solutions  $y^k$ ; that is,  $y$  is dual feasible if, and only if,

$$y = \sum_{k=1}^p \lambda_k y^k \text{ [or } y = Y \lambda \text{] for “weights” } \lambda_k \text{ [for which } \lambda_k \geq 0 \text{ and } \sum_{k=1}^p \lambda_k = 1 \text{]}. \quad (60)$$

To reveal important properties of the columns  $y^k$  in  $Y$ , and to estimate the total number  $p$  of those columns  $y^k$ , first note that the total number of basic (dependent) variables in each phase-I tableau is clearly  $(\text{rank } A + 1)$ ; so the total number of nonbasic (independent) variables is  $n - (\text{rank } A + 1)$  -- namely, the degree-of-difficulty  $d$ .

Consequently, each basic feasible solution  $y^k$  has at least  $d$  zero components [in fact, exactly  $d$  zero components if, and only if,  $y^k$  is “non-degenerate in the linear programming sense”]. It follows then that each basic feasible solution  $y^k$  also has at most  $(n - d)$  positive components [in fact, exactly  $(n - d)$  positive components if, and only if,  $y^k$  is non-degenerate in the linear programming sense]. Consequently, the total number of positive entries in the whole matrix  $Y$  satisfies the inequality

$$p(n - d), \text{ with equality holding if, and only if, each } y^k \text{ is non-degenerate.} \quad (60a)$$

On the other hand, the characterization  $y = Y^{-1}b$  of the dual feasible  $y$  [given by relation (60)] clearly implies that the given problem is canonical if, and only if, each row of  $Y$  has at least one positive entry. Consequently,

$$n \geq p \text{ when the given problem is canonical.} \quad (60b)$$

It follows then from relations (60a) and (60b) that

$$p = n/(n - d) = n/(\text{rank } A + 1) \text{ when the given problem is canonical.} \quad (60c)$$

On the other hand, we know that

$$p = n!/d!(n - d)! = n!/[(n - (\text{rank } A + 1))!(\text{rank } A + 1)!] \quad (60d)$$

because each basic feasible solution  $y^k$  comes from a tableau, and the number of tableaus (including those whose basic solutions are infeasible) can not exceed  $n!/d!(n - d)!$ .

Since  $y^k$  is orthogonal to each column of the exponent matrix  $A$  [by virtue of both the dual feasibility of  $y^k$  and the orthogonality condition (49a)], it is clear that the vector  $\underline{y}^k$  that results from deleting the zero components of  $y^k$  is orthogonal to each column of the matrix  $\underline{A}$  that results from deleting the corresponding rows of  $A$ . Moreover, since  $\underline{y}^k$  obviously inherits normality and positivity from  $y^k$ , it is a dual feasible solution for minimizing the posynomial  $\underline{P}_k$  that results from deleting the corresponding terms of  $P$ . In fact, minimizing  $\underline{P}_k$  is a canonical problem because  $\underline{y}^k > 0$ ; and it has zero degree-of-difficulty because the components of  $\underline{y}^k$  are uniquely determined by the zero values for the “non-basic variables” relative to the “simplex tableau” that determines  $y^k$  as a basic feasible solution to linear system (49). In essence, minimizing  $\underline{P}_k$  is a meaningful [though not necessarily accurate] 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 at least one additional term from a posynomial  $\underline{P}_k$  would produce a posynomial whose infimum is zero -- indicating that its minimization could not possibly be a meaningful approximation to minimizing the original posynomial  $P$ . In summary, for  $k = 1, 2, \dots, p$ , the non-zero 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  $\underline{P}_k$  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  $\underline{P}_k$  along with the duality inequality (50, 51) imply that

$$\underline{P}_k^* = \min \underline{P}_k = \prod_{i=1}^n (c_i/y_i^k)^{y_i^k} < \min P = P^*, \quad k = 1, 2, \dots, p,$$

where the strict inequality results from the fact that  $y^k$  has at least one zero component and hence can not 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  $\max_k \{ \underline{P}_k^* \mid k = 1, 2, \dots, p \}$ , use relation (60) to reformulate the dual objective function  $U(y)$  [defined by equation (50)] in terms of  $\lambda$  as

$$\begin{aligned} W(\lambda) &= \left\{ \prod_{i=1}^n c_i^{y_i(\lambda)} \right\} \left\{ \prod_{i=1}^n y_i(\lambda)^{-y_i(\lambda)} \right\} \\ &= \left\{ \prod_{k=1}^p L_k^{\lambda_k} \right\} \left\{ \prod_{i=1}^n y_i(\lambda)^{-y_i(\lambda)} \right\} \end{aligned}$$

where

$$y_i(\lambda) = \prod_{k=1}^p y_i^k \lambda_k, \quad i = 1, 2, \dots, n,$$

and where the “basic constants”

$$L_k = \prod_{i=1}^n c_i^{y_i^k} = \underline{P}_k^* \prod_{i=1}^n (y_i^k)^{y_i^k}, \quad k = 1, 2, \dots, p.$$

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

$$0 \text{ and } \prod_{k=1}^p \lambda_k = 1.$$

Exercises: Construct the vectors  $\underline{y}^k$ , the resulting posynomials  $\underline{P}_k$ , their optimal values  $\underline{P}_k^*$ , and their optimal solutions  $t^{*k}$  for:

1. the modified power-line example [with power-consumption tax]
2. the modified gravel-shipment example [with  $t_3$  required to be 1].

Also, for each of these two posynomial-minimization problems, determine the smallest and largest possible fractional part of the total minimum cost  $P^*$  due to each posynomial term  $i$ . [Hint: In view of the interpretation of  $y_i^*$  relative to the  $i$ 'th term of  $P^*$ , you need only find the smallest and largest possible values for each  $y_i$  for dual feasible  $y$  -- linear optimization problems that can be readily solved via the "simplex method" simply by inspecting the entries in the corresponding basic dual feasible solution matrix  $Y$ .] Note that this linear analysis provides a relatively quick check that can show a proposed design  $t$  is not optimal.

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

Another important example

[Entropy optimization in information theory, thermodynamics and statistical mechanics]

Given a finite “sample space”  $\{s_1, s_2, \dots, s_n\}$  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”

$$y_i \geq 0 \quad (61a)$$

$$\sum_{i=1}^n y_i = 1 \quad (61b)$$

from both given “moment conditions”

$$\sum_{i=1}^n s_{ij} y_i = \mu_j, \quad j = 1, \dots, m, \quad (62)$$

and a given “a-priori distribution”

$$q_i \geq 0 \quad (63a)$$

$$\sum_{i=1}^n q_i = 1. \quad (63b)$$

The moment conditions (62) typically result from statistically obtained “expected values”  $\mu_j$  of known “random variables”  $s_{ij}$ ; 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 references (12, 13) from certain reasonable axioms in probability theory] is that the “best inference” for the unknown probability distribution  $y$  from the given moment conditions (62) and a-priori distribution (63) is the optimal solution  $y^*$  to the maximization problem:

$$\text{Maximize the “cross entropy” } H(y) = \sum_{i=1}^n y_i \ln (nq_i/y_i) \quad (64)$$

subject to constraints (61) and (62).

Since  $H(y) = \ln U(y)$  when  $c_i = nq_i$  [by virtue of equation (50)] and since condition (61b) makes the moment conditions (62) equivalent to the orthogonality conditions  $\sum_{i=1}^n (\mu_j - v_{ij}) y_i = 0$ ,  $j = 1, \dots, m$ , maximization problem (64) is essentially dual problem (49, 50) when

$$c_i = n q_i \text{ and } a_{ij} = \mu_j - v_{ij}. \quad (65)$$

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

$$\begin{aligned} \text{Minimize } G(z) &= \ln \left[ \prod_{i=1}^n n q_i \exp\left(- \prod_{j=1}^m \{\mu_j - v_{ij}\} z_j\right) \right] \\ &= \left\{ - \prod_{j=1}^m \mu_j z_j + \ln \left[ \prod_{i=1}^n n q_i \exp\left(- \prod_{j=1}^m v_{ij} z_j\right) \right] \right\} \end{aligned} \quad (66)$$

subject to the constraint  $z \in \mathbb{R}^m$ .

Since  $n \gg \gg m$  and hence the degree-of-difficulty  $d = n - (\text{rank } A + 1) \gg \gg m$ , problem (66) is probably much easier to solve numerically than problem (64). Moreover, the previously described canonicity theory for posynomial programming implies that problem (66) has an optimal solution  $z^*$  if, and only if, constraints (61) and (62) have a feasible solution  $y > 0$ . Since the sample space  $\{s_1, s_2, \dots, s_n\}$  can obviously be made smaller if there is no such feasible distribution  $y > 0$ , we can assume, without loss of generality, that problems (64) and (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 reference (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 (62) are linearly independent.
- (3) The solution pairs  $(y^*, z^*)$  constitute the solution set for the system consisting of the probability-distribution conditions (61), the moment conditions (62) and the “duality equation”

$$H(y) = G(z). \quad (67)$$

- (4) The solution pairs  $(y^*, z^*)$  also constitute the solution set for the system consisting of the moment conditions (62) and the “primal-dual optimality conditions”

$$y_i = [q_i \exp(-\sum_{j=1}^m \mu_j z_j)] / \{ \sum_{k=1}^n [q_k \exp(-\sum_{j=1}^m \mu_j z_j)] \}, \quad i = 1, 2, \dots, n, \quad (68)$$

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

- (5) If each  $\mu_j = 0$  and each  $\mu_j = 0$ , then the primal-dual optimality conditions (68) show that  $y^* = q$  [by virtue of the a-priori probability-distribution condition (63b)]. This means that setting  $y = q$  maximizes the cross-entropy  $H(y)$  when the only constraints on  $y$  are the probability-distribution conditions (61). It follows then that:

- (a) the inferred distribution  $y^*$  is simply the a-priori distribution  $q$  when  $q$  satisfies the moment conditions (62),
- (b) when  $q$  satisfies the moment conditions (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  $\epsilon_i = \epsilon_{i1}$ ] and given that the sample space  $\{s_1, s_2, \dots, s_n\}$  consists of the possible “states”  $i$  of a “physical system” that has “energy”  $\epsilon_i$  in state  $i$  [with  $\mu$  being the system’s average energy or “internal energy”], then the primal-dual optimality conditions (68) further simplify to

$$y_i = [\exp(-\epsilon_i z)] / \{ \sum_{k=1}^n \exp(-\epsilon_k z) \}, \quad i = 1, 2, \dots, n; \quad (69)$$

in which case

- (a) the denominator in the primal-dual optimality conditions (69) is the system’s “partition function”,
  - (b) the system’s “absolute temperature”  $T = 1/z^*$  where  $k$  is “Boltzmann’s constant”,
2. the primal-dual optimality conditions (69) and the internal-energy condition

$$\sum_{i=1}^n \epsilon_i y_i = \mu \quad (70)$$

along with the interpretation  $z^* = 1/T$  constitute the “fundamental law” [described in reference (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 (22, 23), (24b, 25), (27b, 28, 29) and (49, 50).

Physicists who write textbooks on statistical mechanics and thermodynamics have traditionally based their presentations [such as in reference (15)] on analyzing conditions (69) and (70) with  $z$  replaced by  $1/T$ , namely, the conditions

$$y_i = [\exp(-\epsilon_i/T)] / \sum_{k=1}^n [\exp(-\epsilon_k/T)], \quad i = 1, 2, \dots, n, \quad \text{and} \quad \sum_{i=1}^n \epsilon_i y_i = \mu.$$

These two conditions or “laws” [comparable in importance to “Newton’s laws of motion for mechanics” and seemingly unrelated to optimization] originated from fundamental physical observations. In any given context, they relate the probability distribution vector  $y$  and state energy vector  $\epsilon$  [statistical-mechanics or microscopic variables] to the absolute temperature  $T$  and internal energy  $\mu$  [thermodynamic or macroscopic variables]. The “entropy” usually employed with these two laws is actually a function of only the thermodynamic or macroscopic variables [such as the absolute temperature  $T$ , the volume, the pressure, the electric and magnetic field intensities, etceteras] and is not directly related to the “cross entropy” function  $H(y) = \sum_{i=1}^n y_i \ln(1/y_i) = -\sum_{i=1}^n y_i \ln(y_i)$  employed here. This function  $H(y)$ , actually its negative  $-\sum_{i=1}^n y_i \ln(y_i)$ , is also called “entropy”, but  $H(y)$  originated much later [in the 1940’s] with the development of “communication theory” and “information theory” by engineers and applied mathematicians [as described in references (12) and (13)], who were influenced somewhat by the statistical mechanics and thermodynamics of the physicists but based their development on the [dual] optimization problem (64). Subsequently, the “variational principle” that connects the cross-entropy maximization problem (64) with the fundamental law (69, 70) for statistical mechanics and thermodynamics was noted and

pedagogically exploited by other engineers [in references (16, 17)], but the alternative variational principles provided by minimization problems (22, 23), (24b, 25) and (27b, 28, 29) [appropriately specialized as in equations (66)] for information theory and statistical mechanics and thermodynamics seem to have been noted for the first time in reference (14). It is presently unknown whether the theory or pedagogy of either information theory or statistical mechanics and thermodynamics would benefit from an exploitation of the properties of these [primal] minimization problems (22, 23), (24b, 25) and (27b, 28, 29). Other connections between geometric programming, statistical mechanics and thermodynamics had previously been given in references (18, 19).

Finally, the significance of the cross-entropy maximization problem (64) in statistical theory and its applications is thoroughly described in reference (20), but the significance of the corresponding [primal] geometric programming minimization problems (22, 23), (24b, 25) and (27b, 28, 29) in statistical theory and its applications is yet to be determined. Moreover, all of this mathematical methodology seems to be applicable to the inference of “mass distributions”  $y^*$  in “nondestructive testing” [where the moment conditions (62) come from X-ray data, sonar data, or any other data resulting from non-destructive tests performed on the “body” whose internal mass distribution  $y^*$  is to be determined].

#### Constrained algebraic optimization via geometric programming

References (8, 9, 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  $Q(t) \leq q$  on other posynomials  $Q(t)$ . Although such minimization problems are generally “non-convex”, the reformulations that result from extending the geometric programming transformations described herein are “convex” when all constraints are of the “prototype form”  $Q(t) \leq q$ . These generalizations greatly enlarge the applicability of posynomial minimization to engineering design and other areas, as can be seen in many references [such as (10, 11, 21, 22, 23, 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 reference (10)]. Moreover, reference (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  $<$ ,  $>$ , and  $=$ ] as equivalent posynomial minimization problems with posynomial constraints of both the desired prototype  $Q(t) \leq q$  and the “reversed type”  $R(t) \geq r$ . Moreover, reference (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, references (27, 28, 29) show how the amazingly general posynomial minimization problems with reversed constraints  $R(t) \geq r$  can be “conservatively approximated” by those with only constraints of the desired prototype  $Q(t) \leq q$ .

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 reference (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 references (31, 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 non-destructive testing. The general theory of geometric programming includes: (i) very strong existence, uniqueness, and characterization theorems, (ii) useful parametric and post-optimality analyses, (iii) illuminating decomposition principles, and (iv) powerful numerical solution techniques.

A comprehensive survey of the whole field as it existed in 1980 can be found in reference (33), and a much more recent and comprehensive treatment of entropy optimization can be found in reference (34). Finally, reference (35) will provide a current state-of-the-art survey in 2000 [or shortly thereafter].

### Bibliography

1. C. Zener, A mathematical aid in optimizing engineering designs, *Proc. Nat. Acad. Sci.* **47**, 537 (1961).
2. C. Zener, A further mathematical aid in optimizing engineering designs, *Proc. Nat. Acad. Sci.* **48**, 518 (1962).
3. C. Zener, Minimization of system costs in terms of subsystem costs, *Proc. Nat. Acad. Sci.* **51**, 164 (1964).
4. C. Zener, An example of design for minimum total costs, counter-flow heat exchangers, *IEEE Trans. on Mil. Elec.* **MIL-8**, 63 (1964).
5. C. Zener, Redesign over-compensation, *Proc. Nat. Acad. Sci.* **53**, 242 (1965).
6. R. J. Duffin, Dual programs and minimum cost, *SIAM Jour.* **10**, 119 (1962).

7. R. J. Duffin, Cost minimization problems treated by geometric means, *Operations Research* **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 Jour. Appl. Math.*, **14**, 6, 1307 (1966).
10. R. J. Duffin, E. L. Peterson and C. Zener, *Geometric Programming-Theory and Application*, New York: John Wiley & Sons (1967).
11. C. Zener, *Engineering Design by Geometric Programming*, New York: John Wiley & Sons - Interscience (1971).
12. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press (1949).
13. A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover Publications (1957).
14. R. J. Duffin and E. L. Peterson, Optimization and insight by geometric programming, *Jour. Appl. Phys.*, **60**, 1860 (1986).
15. R. P. Feynman, *Statistical Mechanics - A Set of Lectures*, W. A. Benjamin (1972).
- ?? A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover Publications (1949).
16. E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.*, **108** (1957).

17. A. Katz, *Principles of Statistical Mechanics -- The Information Theory Approach*, 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, *Jour. Phys. Chem.*, **74**, 2419 (1970).
20. S. Kullback, *Information and Statistics*, Wiley, and Chapman and Hall (1959).
21. M. Avriel, M. J. Rijckaert and D. J. Wilde (editors), *Optimization and Design*, Prentice Hall (1973).
22. R. E. D. Woolsey and H. S. Swanson, *Operations Research for Immediate Application -- A Quick and Dirty Manual*, Harper and Row (1975).
23. C. S. Beightler and D. T. Phillips, *Applied Geometric Programming*, John Wiley and Sons (1976).
24. D. J. Wilde, *Globally Optimal Design*, Wiley Interscience (1978).
25. R. J. Duffin and E. L. Peterson, Geometric programming with signomials, *Jour. Opt. Th. & Appls.*, **11**, 3 (1973).
26. R. J. Duffin and E. L. Peterson, The proximity of algebraic geometric programming to linear programming, *Jour. 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 Jour. Appl. Math.*, **19**, 125 (1970).

29. R. J. Duffin and E. L. Peterson, Reversed geometric programs treated by harmonic means, *Indiana Univ. Math. Jour.*, **22**, 531 (1972).
30. R. J. Duffin, Nonlinear networks, Ila, *Bull. Amer. Math. Soc.*, 53, 963 (1947).
31. E. L. Peterson, Symmetric duality for generalized unconstrained geometric programming, *SIAM Jour. Appl. Math.*, **19**, 487 (1970).
32. E. L. Peterson, Geometric programming, *SIAM Rev.*, **19**, 1 (1976).
33. M. Avriel (editor), *Advances in Geometric Programming*, Plenum Press (1980).
34. S. C. Fang, J. R. Rajasekera, and H-S Jacob Tsao, *Entropy Optimization and Mathematical Programming*, International Series in Operations Research and Management Science, Kluwer Academic Publishers (1997).
35. S. C. Fang and C. Scott (editors), *Geometric Programming and Its Generalizations, with Special Emphasis on Entropy Optimization -- A Special Issue of the Annals of Operations Research* (to appear in 2000).
- ?? Authors(?), Use of geometric programming in the design of an Algerian water conveyance system, *Interfaces*, **8**, No. 3, May (1978).
- ?? E. L. Peterson and G. Plenert, Geometric programming, *Wiley Encyclopedia of Electrical and Electronic Engineering* (John G. Webster, editor) Volume **8**, John Wiley and Sons, 325 (1999).
- ?? Cover and Thomas, *Information Theory* (currently a standard textbook in the field of information theory).

?? Proakis, *Digital Communication*, 3rd edition (currently a standard textbook in the field of communication theory -- one that applies information theory to communication theory in its Chapter 7).

### Cauchy's Arithmetic-Geometric Mean Inequality

The following derivation of this inequality differs considerably from Cauchy's original derivation, but the following derivation also indicates how to derive other inequalities needed in the development of useful duality theories for other important classes of optimization problems.

First, note that the single-variable function  $f(z) = z \ln z$  with natural domain  $z > 0$ , for which  $f'(z) = \ln z + 1$  and for which  $f''(z) = 1/z$ , has the graph

$$f(z) = z \ln z$$

Since  $f(z)$  is strictly convex [because  $f''(z) > 0$

for  $z > 0$ ], the "tangent line" at any point

$(\underline{z}, \underline{z} \ln \underline{z})$  for which  $\underline{z} > 0$  lies under the graph

$1/e$   $\underline{z}$   $1$   $z$  of  $f(z)$  except at the point  $(\underline{z}, \underline{z} \ln \underline{z})$  where

they are tangent.

slope  $x$  any number in  $\mathbb{R}$ .

Consequently,  $z \ln z \geq \underline{z} \ln \underline{z} + (\ln \underline{z} + 1)(z - \underline{z})$  for each  $\underline{z} > 0$  and for each  $z > 0$ , with equality holding if, and only if,  $z = \underline{z}$ . Reformulated in terms of the slope  $x = \ln \underline{z} + 1$  [for which  $\underline{z} = e^{x-1}$ ], this inequality and its equality characterization become, after algebraic simplification and manipulation,

$$x z \geq e^{x-1} + z \ln z \text{ for each } x \in \mathbb{R} \text{ and for each } z > 0,$$

with equality holding if, and only if,  $z = e^{x-1}$ .

Now, the preceding inequality and its equality characterization assert that

$$x_i z_i \geq e^{x_i - 1} + z_i \ln z_i \text{ for each } x_i \in \mathbb{R} \text{ and for each } z_i > 0, \quad i = 1, 2, \dots, n,$$

with equality holding in each inequality if, and only if,  $z_i = e^{x_i - 1}$ ,  $i = 1, 2, \dots, n$ ,

where the variables  $x_i$  and  $z_i$  are the individual components of vector variables  $x \in \mathbb{R}^n$  and

$z \in \mathbb{R}^n$ , respectively, for which  $z > 0$ . Adding all  $n$  of these inequalities implies that

$$\sum_{i=1}^n x_i z_i \geq \sum_{i=1}^n e^{x_i - 1} + \sum_{i=1}^n z_i \ln z_i \text{ for each } x \in \mathbb{R}^n \text{ and for each } z \in \mathbb{R}^n \text{ for which } z > 0,$$

with equality holding if, and only if,  $z_i = e^{x_i - 1}$ ,  $i = 1, 2, \dots, n$ .

Now, note that: (i) for each vector  $z > 0$  there is a vector  $y > 0$  for which  $\sum_{i=1}^n y_i = 1$

and for which  $z = sy$  for some scalar  $s > 0$  [where  $s = \sum_{i=1}^n z_i$  and  $y = z/s$ ]; and

(ii) conversely, a vector  $z$  satisfies the condition  $z > 0$  when there is a vector  $y > 0$  for

which  $\sum_{i=1}^n y_i = 1$  and for which  $z = sy$  for some scalar  $s > 0$ . Consequently, substituting

$sy$  for  $z$  into the preceding displayed inequality and its equality characterization shows,

after algebraic simplification, that

$$\sum_{i=1}^n x_i y_i \geq (1/s) \sum_{i=1}^n e^{x_i - 1} + (\ln s) + \sum_{i=1}^n y_i \ln y_i$$

for each vector  $x \in \mathbb{R}^n$  and for each vector  $y \in \mathbb{R}^n$  for which  $y > 0$  and  $\sum_{i=1}^n y_i = 1$

and for each scalar  $s > 0$ ,

with equality holding if, and only if,  $y_i = e^{x_i - 1}/s$ ,  $i = 1, 2, \dots, n$ .

Now, for each vector  $x \in \mathbb{R}^n$  and for each vector  $y \in \mathbb{R}^n$  for which  $y > 0$  and

$\sum_{i=1}^n y_i = 1$ , choose  $s > 0$  to minimize the expression  $(1/s) \sum_{i=1}^n e^{x_i - 1} + (\ln s)$  that appears in

the preceding displayed inequality. Elementary differential calculus shows that the

optimal  $s^* = \prod_{i=1}^n e^{x_i - 1}$ . Substituting this optimal value  $s^*$  for  $s$  into the preceding displayed inequality and its equality characterization shows, after algebraic simplification, that

$$\prod_{i=1}^n x_i y_i \leq \ln \left( \prod_{i=1}^n e^{x_i} \right) + \prod_{i=1}^n y_i \ln y_i$$

for each vector  $x \in \mathbb{R}^n$  and for each vector  $y \in \mathbb{R}^n$  for which  $y_i > 0$  and  $\prod_{i=1}^n y_i = 1$ ,

with equality holding if, and only if,  $y_i = e^{x_i} / \left( \prod_{i=1}^n e^{x_i} \right)$ ,  $i = 1, 2, \dots, n$ .

Since  $y_i \ln y_i$  approaches 0 as  $y_i$  approaches  $0^+$  [by virtue of Le Hospital's rule], not every  $y_i$  must be  $> 0$  to maintain the validity of the preceding displayed inequality. In particular, the natural domain  $y_i > 0$  of  $y_i \ln y_i$  can be extended to include  $y_i = 0$  while maintaining the continuity of the function  $y_i \ln y_i$ , simply by defining  $0 \ln 0$  to be 0. Then, since both  $x_i y_i$  and  $y_i \ln y_i$  in the preceding displayed inequality are 0 when  $y_i = 0$  [independent of the value of  $x_i$ ], the preceding displayed inequality is actually a strict inequality when some  $y_i = 0$  [because the corresponding term  $e^{x_i}$  in  $\ln \left( \prod_{i=1}^n e^{x_i} \right)$  can not be 0 for any value of  $x_i$  and because the log function is monotone strictly increasing on its domain]. Consequently, the validity of the preceding displayed inequality and its equality characterization are maintained when the "positivity condition"  $y_i > 0$  is replaced by the weaker "positivity condition"  $y_i \geq 0$  [but note that not all  $y_i$  can be simultaneously 0 because of the "normality condition"  $\prod_{i=1}^n y_i = 1$ ].

Now, given a [coefficient] vector  $c > 0$ , replacing each  $x_i$  by  $x_i + \ln c_i$  in the preceding displayed inequality and its equality characterization establishes the following usable version of Cauchy's arithmetic-geometric mean inequality

$$\sum_{i=1}^n x_i y_i \ln \left( \sum_{i=1}^n c_i e^{x_i} \right) + \sum_{i=1}^n y_i \ln (y_i/c_i) \quad (70)$$

for each vector  $x \in \mathbb{R}^n$  and for each vector  $y \in \mathbb{R}^n$  for which  $y \geq 0$  and  $\sum_{i=1}^n y_i = 1$   
[with the understanding that  $0 \ln 0 = 0$ ]

with equality holding if, and only if,  $y_i = c_i e^{x_i} / \left( \sum_{k=1}^n c_k e^{x_k} \right)$ ,  $i = 1, 2, \dots, n$ .

This version will be used in later sections, but is not normally the version attributed to Cauchy. His version can be derived from this version after replacing the positivity condition  $y \geq 0$  with the stronger positivity condition  $y > 0$ , so that the further change of variables  $u_i = c_i e^{x_i} / y_i$ ,  $i = 1, \dots, n$ , can be made.

Exercise: Using the preceding version of Cauchy's inequality, prove the following Cauchy's version

$$\sum_{i=1}^n u_i^{y_i} \geq \sum_{i=1}^n y_i u_i$$

for each vector  $u \in \mathbb{R}^n$  for which  $u > 0$  and

for each vector  $y \in \mathbb{R}^n$  for which  $y \geq 0$  and  $\sum_{i=1}^n y_i = 1$

with equality holding if, and only if, there is a scalar  $u > 0$  such that  $u_i = u$ ,  $i = 1, \dots, n$ .

### Generalized geometric programming (without explicit constraints)

Posynomial minimization and cross-entropy maximization, treated via the reformulations already described, are frequently referred to as "prototype geometric programming" -- in fact, "prototype dual geometric programming". "Generalized geometric programming" is an abstraction of the prototype reformulations -- one that permits analogous treatments of other important classes of optimization problems.

By definition, a generalized geometric programming problem  $A$  consists of minimizing a given function  $g$  with given domain  $C \subseteq \mathbb{R}^n$  over the intersect of  $C$  with a given vector (sub)space  $X \subseteq \mathbb{R}^n$ . Stated more formally for future reference, we have

**Problem A:** Given a function  $g:C$  whose domain  $C \subseteq \mathbb{R}^n$  and given a vector space

$X \subseteq \mathbb{R}^n$ , use the resulting “feasible solution set”

$$S = X \cap C,$$

to calculate both the “problem infimum”

$$= \inf_{x \in S} g(x)$$

and the “optimal solution set”

$$S^* = \{x \in S \mid g(x) = \inf_{x \in S} g(x)\}.$$

Since any optimization problem can be formulated as problem  $A$  -- simply by letting  $g$  be the objective function to be minimized,  $C$  the feasible solution set over which the minimization is to occur, and  $X$  the whole vector space  $\mathbb{R}^n$  -- there is no loss of generality in studying the properties of problem  $A$ . Actually, it is the extra flexibility of being able to let  $X$  be a “proper subspace” of  $\mathbb{R}^n$  that makes problem  $A$  an important formulation for general optimization problems.

As a concrete example, we have already seen that, for the important posynomial-minimization example previously treated, the given function  $g:C$  has the concrete formulas

$$g(x) = \ln \left( \prod_{i=1}^n c_i e^{x_i} \right) \text{ and } C = \mathbb{R}^n$$

while the given vector space  $X$  has the concrete formula

$$X = \{x \in \mathbb{R}^n \mid x_i = \prod_{j=1}^m a_{ij} z_j, \quad i = 1, 2, \dots, n\}$$

[rather than  $g(x)$  itself being the posynomial  $\prod_{i=1}^n c_i \prod_{j=1}^m x_j^{a_{ij}}$  and  $C$  being the positive orthant

$\{x \in \mathbb{R}^n \mid x_i > 0\}$  while  $X$  is all of  $\mathbb{R}^n$ ].

As another concrete example, we have already seen that, for the important cross-entropy maximization example previously treated, the given function  $g:C$  has the concrete formulas

$$g(x) = -\sum_{i=1}^n x_i \ln(x_i/c_i) \text{ and } C = \{x \in \mathbb{R}^n \mid x_i \geq 0 \text{ and } \sum_{i=1}^n x_i = 1\}$$

while the given vector space  $X$  has the concrete formula

$$X = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i a_{ij} = 0, j = 1, 2, \dots, m\}.$$

[rather than  $C$  itself being  $\{x \in \mathbb{R}^n \mid x_i \geq 0 \text{ and } \sum_{i=1}^n x_i = 1 \text{ and } \sum_{i=1}^n x_i a_{ij} = 0, j = 1, 2, \dots, m\}$

while  $X$  is all of  $\mathbb{R}^n$ ]. Note that we are reformulating the original cross-entropy

maximization problem as an “equivalent” minimization problem -- an equivalence that

occurs because  $\ln(x_i/c_i) = -\ln(c_i/x_i)$ . For brevity, we shall also refer to the function being

minimized as cross-entropy [rather than negative cross-entropy]. We shall also find it

convenient to study other maximization problems by reformulating them as the equivalent

problem of minimizing minus what you wish to maximize. In doing so, we need only

keep in mind that the optimal values are negatives of each other, while the optimal

solution sets are identical to each other.

We shall eventually see that the “geometric programming duality” between the posynomial-minimization and cross-entropy minimization examples comes from two sources: (i) the duality [frequently called “conjugacy”] resulting between the two different concrete functions  $g:C$  in the inner-product version of Cauchy’s arithmetic-geometric mean inequality, and (ii) the duality [frequently called “orthogonal

complementarity”] resulting between the two different concrete vector spaces  $X$  characterized as the column space of a given matrix  $A$  and all vectors orthogonal to that column space, respectively. Source (ii) should be familiar from a previous study of linear algebra, and source (i) will be further explained after we reformulate other important classes of optimization problems as geometric programming problem  $A$ .

Example 0 (linear optimization problems)

The “standard formulation” is:

Minimize the “linear function”  $cz$  subject to the “linear constraints”

$$Mz = b \text{ and } z \geq 0,$$

where  $c$  is a given row vector with  $m$  components,  $z$  is a variable column vector with  $m$  components, and  $b$  is a given column vector whose number  $p$  of components is the same as the number  $p$  of rows in the given matrix  $M$  [whose number of columns is, of course,  $m$ ].

A reformulation of the standard formulation as a geometric programming problem  $A$  comes from choosing:

$$C = \{x = (x_1, x_M, x_I)^t \mid x_1 \in \mathbf{R}, x_M = b, x_I \geq 0\} \text{ and } g(x) = x_1 + 0x_M + 0x_I$$

while

$$X = \{x = (x_1, x_M, x_I)^t \mid x = \begin{matrix} c \\ M \\ I \end{matrix} z \text{ for some } z \in \mathbf{R}^m\},$$

where  $I$  is the  $m \times m$  identity matrix. Note that  $X$  is the column space of a “partitioned matrix” whose number  $n$  of rows is  $n = 1 + p + m$  -- the dimension of the appropriate

vector space  $\mathbb{R}^n$ . Also, note that  $X$  itself is not  $\mathbb{R}^n$  [because the presence of the  $m \times m$  identity matrix  $I$  in the definition of  $X$  implies that  $X$  has dimension  $m = n - (1 + p)$ ].

Exercises: 1. If the constraints  $Mz = b$  are replaced by the constraints  $Mz \leq b$ , how should the previous choice of  $g:C$  and  $X$  change?

2. If, in addition to the replacement of Exercise 1, the constraints  $z \geq 0$  are deleted [so that  $z$  is now “unrestricted”], how should the previous choice of  $g:C$  and  $X$  change?

#### Example 1 (quadratic optimization problems)

The “standard formulation” for the “convex case” [the case that occurs in the famous “Markowitz’s portfolio selection model”, as well as quadratic models frequently used to approximate general differentiable convex programming problems] is:

Minimize the “positive semidefinite quadratic function”  $(1/2)z^t H z + h z$  subject to the “linear constraints”

$$Mz \leq b,$$

where  $H$  is a given symmetric positive semidefinite matrix (i.e.,  $H = H^t$  and  $z^t H z \geq 0$  for each

$z \in \mathbb{R}^m$ ),  $h$  is a given row vector with  $m$  components,  $z$  is a variable column vector with  $m$  components, and  $b$  is a given column vector whose number  $p$  of components is the same as the number  $p$  of rows in the given matrix  $M$  [whose number of columns is, of course,  $m$ ].

A reformulation of the preceding standard formulation as a geometric programming problem  $A$  comes from first expressing the matrix  $H$  as  $H = D^t D$  in terms of its “square root” matrix  $D$  [whose computation is relatively elementary and is described in introductory books on quadratic forms], and then choosing:

$$C = \{x = (x_D, x_I, x_M)^t \mid x_D \in \mathbb{R}^d, x_I \in \mathbb{R}, x_M \in \mathbb{R}, b, \} \text{ and } g(x) = (1/2) \sum_1^d x_i^2 + x_I + 0x_M$$

$$X = \{x = (x_D, x_I, x_M)^t \mid x = \begin{matrix} D \\ h \\ M \end{matrix} z \text{ for some } z \in \mathbb{R}^m\},$$

where  $d$  is the number of rows in the square root matrix  $D$  [a number known to be  $m$ , with equality holding if, and only if,  $H$  is nonsingular].

Exercises: 1. If the constraints  $Mz \leq b$  are replaced by the constraints  $Mz = b$ , how should the previous choice of  $g;C$  and  $X$  change?

2. Modify the preceding choice of  $g;C$  and  $X$  so that the resulting problem  $A$  is the “standard linear regression problem” -- namely, “find a vector  $Dz$  in the column space of a given matrix  $D$  that is as close as possible, in the Euclidean-distance sense, to a given vector  $v$ ”.

Example 2 (posynomial minimization problems)

Previously described.

Example 3 (cross-entropy minimization problems)

Previously described [as the “geometric dual” to Example 2].

Some of you may know that Karmarker's "interior-point" algorithm for linear programming can be described to a large extent via the introduction of a nonlinear "barrier function" into the linear objective function for the standard formulation discussed in our Example 0. In particular, the "logarithmic barrier function"  $(-\ln z_j)$  and a positive parameter  $\mu$  can be used in the reformulation

Minimize the nonlinear function  $cz + \mu \sum (-\ln z_j)$  subject to the "linear constraints"

$$Mz = b \text{ and } z > 0,$$

which is a nonlinear minimization problem that gets "arbitrarily close" to the standard formulation in Example 0 as  $\mu \rightarrow 0^+$ . Note that the "interior point condition"  $z > 0$  replaces the condition  $z \geq 0$  because  $\ln 0$  is undefined; and note also that there is a "barrier" to  $z_j \rightarrow 0^+$  because  $(-\ln z_j) \rightarrow +\infty$  as  $z_j \rightarrow 0^+$  [the "wrong direction" for minimization]. Consequently, for each  $\mu > 0$ , the optimal solution  $z^*(\mu)$  [which is known to exist when feasible solutions exist] is "interior" to the "positive orthant" in that  $z^*(\mu) > 0$ .

More recently, Professor Fang of NCSU has found it advantageous to replace the logarithmic barrier function  $(-\ln z_j)$  in the preceding reformulation with the "entropic barrier function"  $(z_j \ln z_j)$ . The resulting barrier to  $z_j \rightarrow 0^+$  is more subtle because  $(z_j \ln z_j) \rightarrow 0$  as  $z_j \rightarrow 0^+$ ; in essence, the resulting barrier is because the derivative of  $(z_j \ln z_j)$

$\rightarrow -\infty$  as  $z_j \rightarrow 0^+$  [a property shared with  $(-\ln z_j)$  that clearly keeps  $z_j$  away from 0 when minimizing].

Exercises: For a given constant  $\mu > 0$ , reformulate in a "separable way":

1. the logarithmic-barrier reformulation as Problem A,
2. the entropic-barrier reformulation as Problem A.

[“Separable way” means that  $g:C$  is a “separable function” -- in that the formula for  $g(x)$  is a sum of terms, each of which depends on only one component of  $x$ , while  $C$  is the “Cartesian product” or “direct sum” of one-dimensional intervals.]

Example 4 (discrete optimal control with linear dynamics, or dynamic programming with linear transition equations)

The “standard formulation” is:

$$\begin{aligned} &\text{Minimize } \sum_{i=1}^p g_i(r^i, d^i) \text{ subject to the constraints} \\ &\text{“state” } r^i \in R_i \text{ and “decision” } d^i \in D_i, \quad i = 1, 2, \dots, p, \\ &\quad \text{and} \\ &r^1 = B_1 d^1, \text{ while } r^i = A_i r^{i-1} + B_i d^i \text{ for } i = 2, 3, \dots, p, \end{aligned}$$

where the given functions  $g_i(r^i, d^i)$  have given domains  $R_i \times D_i$ , and where the given matrices  $A_i$  and  $B_i$  have “sizes compatible with the sizes” of  $r^i$  and  $d^i$  as required in the preceding displayed “linear dynamics” or “linear transition equations”.

A reformulation of the preceding standard formulation as a geometric programming problem  $A$  comes from choosing:

$$\begin{aligned} C &= \{x = (r^1, d^1, \dots, r^p, d^p) \mid r^i \in R_i, d^i \in D_i, \dots, r^p \in R_p, d^p \in D_p\} \text{ and } g(x) = \sum_{i=1}^p g_i(r^i, d^i) \\ X &= \{x = (r^1, d^1, \dots, r^p, d^p) \mid r^1 = B_1 d^1, \text{ while } r^i = A_i r^{i-1} + B_i d^i \text{ for } i = 2, 3, \dots, p\} \end{aligned}$$

Exercises: 1. Given that  $p = 3$ , each  $r^i$  is a scalar variable  $r_i$ , each  $d^i$  is a scalar variable  $d_i$  [and hence each matrix  $A_i$  and  $B_i$  is  $1 \times 1$ ], and given that each  $g_i(r^i, d^i) = (1/2)(r_i^2 + d_i^2)$ , state explicit formulas for the corresponding problem  $A$ .

Example 5 (optimal location)

The “Weber problem” [or “generalized Fermat problem”] is:

$$\text{Minimize } \sum_{i=1}^p d_i(z, b^i) \text{ subject to the simple constraint } z \in \mathbb{R}^m$$

where  $d_i(z, b^i)$  is an appropriate “distance” from the location  $z$  of a “new facility” to the location  $b^i$  of a “previously constructed facility”  $i$ .

A reformulation of the preceding Weber problem as a geometric programming problem  $A$  comes from choosing:

$$C = \{x = (x^1, x^2, \dots, x^p) \mid x^1 \in \mathbb{R}^m, x^2 \in \mathbb{R}^m, \dots, x^p \in \mathbb{R}^m\} \text{ and } g(x) = \sum_{i=1}^p d_i(x^i, b^i)$$

$$X = \text{column space of the partitioned matrix } [I \ I \ \dots \ I]^t, \text{ with } p \text{ identity matrices } I \text{ that are } m \times m.$$

Exercises: 1. Given that  $p = 3$  and  $m = 1$  and that each distance  $d_i(z, b^i) = |z - b^i|$ , state explicit formulas for the corresponding problem  $A$ .

Additional examples are most easily described after problem  $A$  is made even more flexible, by only requiring  $X$  to be a “cone” -- namely, a set that is required to be “algebraically closed” only under all nonnegative linear combinations [rather than under all linear combinations, as required for a vector space]. However, prior to doing so, it is instructive to see how meaningful “optimality conditions” and meaningful “duals” can be constructed for all of the preceding examples [not just examples 2 and 3, as already done].

First-order optimality conditions

The standard first-order optimality conditions that come from setting the “gradient” of an objective function equal to zero are applicable to geometric programming problem  $A$  only when the underlying vector space  $X$  is all of  $\mathbb{R}^m$  [which we already know is normally not the case]. The appropriate first-order optimality conditions for problem  $A$  involve the orthogonal complement  $Y = X^\perp$  of  $X$ , as given in the following definition.

**DEFINITION.** A critical solution [stationary solution, or equilibrium solution] for problem  $A$  is any vector  $x^*$  that satisfies the following [geometric programming] first-order optimality conditions:

$$x^* \in X \cap C \text{ and } g(x^*) \in Y$$

Observation: If  $X$  is the vector space  $\mathbb{R}^n$ , then  $Y = \{0\}$  and hence the geometric programming first-order optimality conditions become the more familiar first-order optimality conditions  $x^* \in C$  and  $g(x^*) = 0$ .

Theorem 1. Under the hypothesis that  $g$  is “differentiable” at some point  $x^* \in (\text{int } C)$  [the “interior” of  $C$ ]:

- (1) if  $x^*$  is a “locally optimal solution” for problem  $A$  [eg, if  $x^*$  is an optimal solution for problem  $A$ ], then  $x^*$  is a critical solution for problem  $A$ ,
- (2) if  $g:C$  is “convex” and if  $x^*$  is a critical solution for problem  $A$ , then  $x^*$  is an optimal solution for problem  $A$  [and hence  $x^*$  is a “locally optimal solution” for problem  $A$ ].

Proof: To prove part (1), first recall that the local optimality of  $x^*$  implies that  $x^* \in X \cap C$ . Then, notice that the local optimality of  $x^*$ , the differentiability of  $g$  at  $x^* \in (\text{int } C)$ , and the vector-space properties of  $X$  imply that the “directional derivative”  $\nabla g(x^*) x = 0$  for each “feasible direction”  $x \in X$ ; so  $\nabla g(x^*) \in Y$ .

To prove part (2), first recall that the convexity of  $g$  and the differentiability of  $g$  at  $x^*$  imply that  $g(x) - g(x^*) \geq \nabla g(x^*) (x - x^*)$  for each  $x \in C$ . Consequently, the assumption that  $x^* \in X$  and the assumption that  $\nabla g(x^*) \in Y$ , along with the vector-space properties of  $X$ , imply that  $g(x) - g(x^*) \geq 0$  for each  $x \in X \cap C$ ; so  $x^*$  is an optimal solution for problem A.

Exercises: 1. To which of our six examples of Problem A does Theorem 1 always apply, and why?

2. To which of our six examples of Problem A does Theorem 1 sometimes apply, and what are the extra conditions on the given  $g:C$  that make it applicable?

3. To which of our six examples of Problem A does Theorem 1 never apply, and why?

[Note: A generalization of the gradient operation in the context of convex functions will permit a further generalization of the geometric programming criticality conditions -- a generalization that will apply to some of the examples to which the current Theorem 1 does not apply.]

The key to duality theory

[The general case]

Recall that the duality theorem relating posynomial minimization [example 2] and cross-entropy minimization [example 3] comes from a version (70) of Cauchy's arithmetic-geometric mean inequality -- in the general inner-product form of

$$xy \leq g(x) + h(y), \quad (71)$$

where  $xy$  is the inner product of  $n$ -dimensional vectors  $x$  and  $y$  while  $g(x)$  is the primal objective function and  $h(y)$  is the dual objective function. Since dual feasible solutions  $x$  and  $y$  are orthogonal, their inner product  $xy = 0$ ; so their corresponding dual feasible objective values  $g(x)$  and  $h(y)$  satisfy the inequality  $0 \leq g(x) + h(y)$ , which becomes an equality when  $x$  and  $y$  are actually optimal solutions. This is the essence of duality theory for mathematical programming in general, but the details must be supplied for each class of problems [as already done for the prototype posynomial and cross-entropic classes].

Clearly, the key question is whether there are other inequalities of the inner-product type (71); in particular, given an objective function  $g(x)$  with domain  $C$  [such as one of those specified in the preceding examples], is there a function  $h(y)$  with domain  $D$  such that  $xy \leq g(x) + h(y)$  for each  $x \in C$  and each  $y \in D$ ? Obviously, such function values  $h(y)$  must satisfy the inequality  $yx - g(x) \leq h(y)$  for each  $x \in C$ ; so  $h(y)$  must satisfy the inequality

$$\sup_{x \in C} [yx - g(x)] \leq h(y). \quad (72)$$

Consequently, the domain  $D$  of  $h$  can only contain points  $y$  for which  $\sup_{x \in C} [yx - g(x)]$  is finite. If there are no such points  $y$ , there is no function  $h:D$ , and hence no duality theory relevant to minimizing  $g(x)$ . On the other hand, if there is at least one such point  $y$ , there are infinitely many such functions  $h:D$ , all with the same nonempty domain  $D = \{y \in \mathbb{R}^n \mid \sup_{x \in C} [yx - g(x)] < +\infty\}$ .

To have the best chance of achieving equality in inequality (71) when  $x$  and  $y$  are optimal, it is reasonable to pick  $h(y)$  so that inequality (72) is actually an equality. This choice is called the “conjugate transformation”, and the resulting function  $h:D$  is called the “conjugate transform” of the given function  $g:C$ .

Fenchel's conjugate transformation  
[Conjugate functions]

Each function  $g:C$  whose domain  $C \subseteq \mathbb{R}^n$  has a conjugate transform  $h:D$  whose domain

$$D = \{y \in \mathbb{R}^n \mid \sup_{x \in C} [yx - g(x)] < +\infty\}$$

and whose function value

$$h(y) = \sup_{x \in C} [yx - g(x)],$$

provided, of course, that  $D$  is not empty.

Example: If  $C = \mathbb{R}^n$  and  $g(x) = \ln(\prod_{i=1}^n c_i e^{x_i})$ , then  $\sup_{x \in C} [yx - g(x)] =$

$\sup_{x \in \mathbb{R}^n} [\sum_{i=1}^n y_i x_i - \ln(\prod_{i=1}^n c_i e^{x_i})]$ , which is finite only if:

(i) each  $y_i \geq 0$  [because it is clearly  $+\infty$  from  $x_i \rightarrow -\infty$  when  $y_i < 0$ ]

(ii)  $\sum_{i=1}^n y_i = 1$  [because letting each  $x_i = s$  makes  $\sum_{i=1}^n y_i x_i - \ln(\prod_{i=1}^n c_i e^{x_i}) =$

$s(\sum_{i=1}^n y_i - 1) - \ln(\prod_{i=1}^n c_i)$ , which can clearly be sent to  $+\infty$  when  $\sum_{i=1}^n y_i > 1$ ].

On the other hand, if each  $y_i > 0$  and if  $\prod_{i=1}^n y_i = 1$ , then Cauchy's arithmetic-geometric mean inequality

$$\prod_{i=1}^n y_i x_i - \ln \left( \prod_{i=1}^n c_i e^{x_i} \right) = \prod_{i=1}^n y_i \ln(y_i/c_i) \text{ for each } x \text{ in } \mathbb{R}^n$$

implies that this  $\sup_{x \in \mathbb{R}^n} [\prod_{i=1}^n y_i x_i - \ln(\prod_{i=1}^n c_i e^{x_i})]$  is  $< +\infty$ . In fact, letting  $x_i = -$  for zero  $y_i$  and choosing  $x_i = \ln(y_i/c_i)$  for positive  $y_i$  shows that this

$\sup_{x \in \mathbb{R}^n} [\prod_{i=1}^n y_i x_i - \ln(\prod_{i=1}^n c_i e^{x_i})] = \prod_{i=1}^n y_i \ln(y_i/c_i)$ . Consequently,  $D = \{y > 0 \mid \prod_{i=1}^n y_i = 1\}$  and

$h(y) = \prod_{i=1}^n y_i \ln(y_i/c_i)$  for this example.

Exercises: 1. If  $C = \{x > 0 \mid \prod_{i=1}^n x_i = 1\}$  and  $g(x) = \prod_{i=1}^n x_i \ln(x_i/c_i)$ , find  $h:D$ . [Hint: Use Cauchy's arithmetic-geometric mean inequality with the symbols  $x$  and  $y$  interchanged.]

2. In view of the preceding example, what does exercise 1. say about the conjugate transform of the conjugate transform of  $g:C$  when  $C = \mathbb{R}^n$  and  $g(x) = \ln(\prod_{i=1}^n c_i e^{x_i})$ ?

The preceding example and exercises expose a very subtle and important duality [or symmetry] between the functions  $\ln(\prod_{i=1}^n c_i e^{x_i}):\mathbb{R}^n$  and  $[\prod_{i=1}^n y_i \ln(y_i/c_i)]:\{y > 0 \mid \prod_{i=1}^n y_i = 1\}$ . Each is the conjugate transform of the other. Such function pairs  $g:C$  and  $h:D$  are called "conjugate functions". In essence, the preceding conjugate functions, along with orthogonal complementary vector spaces  $X$  and  $Y$  [another important duality or symmetry] are the critical ingredients in posynomial and cross-entropy minimization duality theory [including all existence, uniqueness and characterization theorems, as well

as post-optimality analyses]. In fact, appropriate conjugate functions  $g:C$  and  $h:D$ , along with appropriate “dual cones”  $X$  and  $Y$  [an important generalization of orthogonal complementary vector spaces  $X$  and  $Y$ ] are the critical ingredients in all optimization duality theories known to the author [including those for the six example classes previously described]. This fact has unified seemingly unrelated duality theories previously constructed for various important classes of optimization problems, and [most importantly] has guided the construction of helpful duality theories for other important classes of optimization problems. In many cases, the problems must undergo an initial transformation [eg, from posynomial to logarithmic-exponential type] prior to computing an appropriate pair of dual cones [or orthogonal complementary vector spaces] and an appropriate pair of conjugate functions.

The preceding facts indicate the need for a thorough understanding of the conjugate transformation and the conditions under which it produces conjugate functions, as well as a thorough understanding of dual cones. Since the properties of certain types of conjugate functions imply many of the important properties of dual cones, we shall focus first on the conjugate transformation. The conjugate transformation is most easily understood when applied to a “separable function”  $g:C$ , namely a function for which  $g(x) = \prod_{i=1}^n g_i(x_i)$  and  $C = \times_{i=1}^n C_i$ . The reason is that

$$\sup_{x \in \times_{i=1}^n C_i} [\prod_{i=1}^n y_i x_i - \prod_{i=1}^n g_i(x_i)] = \prod_{i=1}^n \sup_{x_i \in C_i} [y_i x_i - g_i(x_i)],$$

which implies that the conjugate transform of  $[\prod_{i=1}^n g_i]: [\times_{i=1}^n C_i]$  is simply  $[\prod_{i=1}^n h_i]: [\times_{i=1}^n D_i]$

where  $h_i : D_i$  is the conjugate transform of  $g_i : C_i$ , for  $i = 1, 2, \dots, n$ . In essence, the conjugate transformation preserves separability; and the computation of the conjugate transform of a separable function can be reduced to the computation of the conjugate transforms of  $n$  functions  $g_i : C_i$ , each of which is a one-dimensional optimization problem that can frequently be solved analytically by techniques from calculus I.

Exercises: With the aid of analytical techniques from algebra and calculus I, find the conjugate transform of the following functions  $g: C$  of a single scalar variable  $x$  that occur as terms in the geometric programming formulations [problem A] of the previously given examples 0 and 1 [including the modified versions of example 0 mentioned after examples 2 and 3], as well as example 5.

1.  $C = \mathbb{R}$  and  $g(x) = x$ .
2.  $C = \{b\}$  and  $g(x) = 0$ .
3.  $C = \{x \geq 0\}$  and  $g(x) = 0$ .
4.  $C = \mathbb{R}$  and  $g(x) = (1/2)x^2$ .
5.  $C = \{x \geq b\}$  and  $g(x) = 0$ .
6.  $C = \mathbb{R}$  and  $g(x) = (1/2)(x - v)^2$ .
7.  $C = \{x > 0\}$  and  $g(x) = -\ln x$
8.  $C = \{x > 0\}$  and  $g(x) = x \ln x$

Now, find the conjugate transform of each of the conjugate transforms found in exercises

1

through 8. Do you see any relationship between the given function, its conjugate

transform, and the conjugate transform of its conjugate transform for exercises 1 through

8?

Given the function  $g:C$  where  $C = \mathbb{R}^n$  and  $g(x) = \sum_{i=1}^n c_i e^{x_i}$ :

9. This function is separable with  $C_i = \underline{\hspace{1cm}}$  and  $g_i(x_i) = \underline{\hspace{1cm}}$

Given that  $c_i > 0, i = 1, 2, \dots, n$ :

$$a. \sup_{x_i \in \mathbb{R}} [y_i x_i - c_i e^{x_i}] = \begin{matrix} \underline{\hspace{1cm}} & \text{when } y_i < 0 \\ \underline{\hspace{1cm}} & \text{when } y_i = 0 \\ \underline{\hspace{1cm}} & \text{when } y_i > 0 \end{matrix}$$

[Hint: When  $y_i < 0$  and when  $y_i = 0$ , let  $x_i = \underline{\hspace{1cm}}$ ; but when  $y_i > 0$ , use the differential calculus.]

b. Use the results of part a to give formulas for  $D$  and  $h(y)$  where  $h:D$  is the conjugate transform of the given function  $g:C$ .

10. Use the results of exercise 9 to formulate an alternative minimization

problem

[or maximization problem] that is dual to minimizing a posynomial

$$\sum_{i=1}^n c_i \prod_{j=1}^m t_j^{a_{ij}}; \text{ and hence show that a given optimization problem can have}$$

more

than one dual problem. {Note: It can be shown that a given convex

optimization

problem [the case here because each coefficient  $c_i > 0$ ] normally has infinitely many dual problems, but experience has indicated that one is usually more desirable computationally and theoretically than the others. Why do you think the original dual, based on the conjugate transform of the nonseparable function  $\ln \left( \prod_{i=1}^n c_i e^{x_i} \right)$ , is more desirable than the dual constructed in this exercise, even though the original is clearly more difficult to motivate and calculate?}

Given the function  $g:C$  where  $C = \mathbb{R}^n$  and  $g(x) = \prod_{i=1}^n c_i e^{x_i}$ , but at least one coefficient is negative, say  $c_1 < 0$ :

11. the  $\sup_{x_1 \in \mathbb{R}} [y_1 x_1 - c_1 e^{x_1}] = \underline{\hspace{2cm}}$  for each  $y_1 \in \mathbb{R}$

[Hint: Since  $c_1 < 0$ , the term  $-c_1 e^{x_1}$  + exponentially as  $x_1 \rightarrow +\infty$ , but  $y_1 x_1$  can approach  $-\infty$  in only a linear way.]

12. Use the result of exercise 11 to show that  $D$  is empty and hence  $g:C$  has no conjugate transform  $h:D$  when at least one coefficient  $c_i$  in a generalized polynomial is negative. {Note: The implication of this fact is that meaningful duals for the maximization of posynomials or the minimization or maximization

of signomials [namely, generalized polynomials whose coefficients have a mixture of algebraic signs] can be constructed only after such optimization problems have been transformed into posynomial-minimization form -- a task that is known to be possible only within the more general format of posynomial minimization subject to posynomial constraints [constraints that are nonlinear and hence can not be treated the way we have been treating linear constraints]. Since there are many important classes of nonlinear optimization problems with only linear constraints, it will be the end of the semester before we can treat nonlinear constraints, such as posynomial constraints.]

Since we shall eventually see that many other important optimization problems are either separable or can be reformulated in separable form [as we did with examples 0 through 5], the preceding facts indicate the importance of a thorough understanding of the conjugate transformation and the conditions under which it produces conjugate functions in the one-dimensional case. Since

$x = (x_1)$  and  $y = (y_1)$  in the one-dimensional case, we can simplify our notation by eliminating all subscripts; and we can easily think of the functions  $g:C$  and  $h:D$  in terms of their planar graphs.

One-dimensional conjugacy

Actually, it is best to think of functions in the one-dimensional case in terms of their “epigraphs”, namely, all points “on or above” their graphs [as illustrated below].

Closed  
and convex

Not closed  
but convex

Closed  
but not convex

Not closed  
and not convex

## SOME FUNCTION EPIGRAPHS

“Closed functions” are those functions whose epigraphs are “closed sets” [in the sense that every “boundary point for the set” is itself always in the set], and “convex functions” are those functions whose epigraphs are “convex sets” [in the sense that the “straight line joining two points” in the set is itself always in the set].

Functions that are both closed and convex [called “closed convex functions”] are important because we shall see that the family of all such functions is both identical to the “range” of the conjugate transformation and “properly contained within” the “domain” of the conjugate transformation. In fact, the conjugate transformation “restricted” to the family of all closed convex functions maps that family “onto” itself in “one-to-one” fashion, in such a way that the conjugate transformation is its own “inverse”. Consequently, closed convex functions come in conjugate pairs, termed “conjugate functions”, each being the conjugate transform of the other. In particular, the first

function graphed above is half of a conjugate pair, but the other three are not, even though they have conjugate transforms.

To verify the preceding facts, first note that the linear function  $yx - t$  of  $x$  has a straight-line graph with slope  $y$  and vertical intercept  $-t$ , as well as a nonempty epigraph

$$\{(x, s) \in \mathbb{R}^2 \mid yx - t \leq s\} \quad (73)$$

that constitutes a “closed half-plane” in  $\mathbb{R}^2$  [“closed” in the topological sense, namely,  $\{(x, s) \in \mathbb{R}^2 \mid yx - t \leq s\}$  contains its “boundary points”].

Now, given a function  $g:C$  with a nonempty domain  $C$  and hence a nonempty epigraph

$$\{(x, s) \in \mathbb{R}^2 \mid x \in C \text{ and } g(x) \leq s\} \quad (\text{epi } g), \quad (74)$$

recall that its conjugate transform  $h:D$  exists if, and only if, it has a nonempty domain  $D$  and hence a nonempty epigraph

$$\{(y, t) \in \mathbb{R}^2 \mid y \in D \text{ and } h(y) \leq t\} \quad (\text{epi } h). \quad (75)$$

Elementary algebraic manipulations used in conjunction with the preceding defining equations for  $\{(x, s) \in \mathbb{R}^2 \mid yx - t \leq s\}$ ,  $(\text{epi } g)$ , and  $(\text{epi } h)$  [as well as the defining equations for  $h:D$  in terms of  $g:C$ ] shows that

$$(\text{epi } g) \neq \emptyset \text{ if, and only if, } (\text{epi } h) \neq \emptyset. \quad (76)$$

An immediate consequence of this fundamental relation is that  $g:C$  has a conjugate transform  $h:D$  [in that  $(\text{epi } h) \neq \emptyset$ ] if, and only if,  $(\text{epi } g) \neq \emptyset$  for some  $(y, t) \in \mathbb{R}^2$ .

Given that a function  $g:C$  has a conjugate transform  $h:D$  [with  $C$  and  $D$  not empty],

for each

$x \in C$  consider the closed half-plane

$$\{(y, t) \in \mathbb{R}^2 \mid xy - g(x) \leq t\} = \bigcap_{x \in C} H_x^{g(x)} \quad (77)$$

whose “boundary” is the straight line with slope  $x$  and vertical intercept  $-g(x)$ . Since the

definition of  $h:D$  implies that the “conjugate inequality”  $xy \leq g(x) + h(y)$  is valid for each

$x \in C$  and each

$y \in D$ , the  $(y, h(y)) \in H_x^{g(x)}$  for each  $x \in C$ ; so  $(y, h(y)) \in \bigcap_{x \in C} H_x^{g(x)}$ . On the other hand, if

$(y, t) \in \bigcap_{x \in C} H_x^{g(x)}$ , then  $(y, t) \in H_x^{g(x)}$  for each  $x \in C$ , and hence  $xy - g(x) \leq t$  for each  $x$

$x \in C$ ; so  $\sup_{x \in C} [xy - g(x)] \leq t$ , which means that  $y \in D$  and  $h(y) \leq t$ , so  $(y, t) \in (\text{epi } h)$ .

Consequently,

$$(\text{epi } h) = \bigcap_{x \in C} H_x^{g(x)} = H, \quad (78)$$

which is a closed convex set [because each half plane  $H_x^{g(x)}$  is clearly closed and convex

and because the intersect of closed convex sets is closed and convex]. We conclude that

the conjugate transform  $h:D$  is always a closed convex function [even when  $g:C$  is not a

closed convex function].

Given that a function  $g:C$  has a conjugate transform  $h:D$  [with  $C$  and  $D$  not empty],

$h:D$  itself has a conjugate transform, say  $g:C$ , where

$$C \subseteq C, \text{ and } g(x) = g(x) \text{ for each } x \in C, \quad (79)$$

because the conjugate inequality  $xy \leq g(x) + h(y)$  implies that  $xy - h(y) \leq g(x)$  for each  $y$

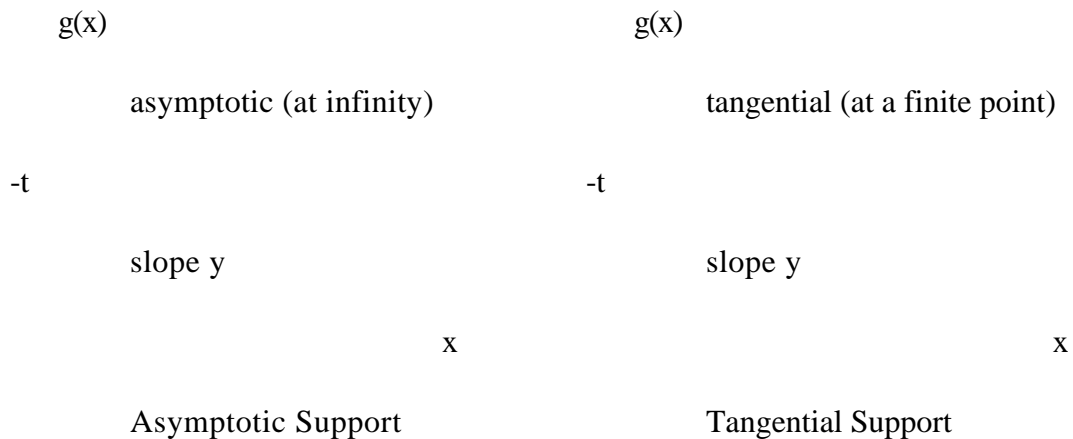
$y \in D$  and each  $x \in C$ ; so  $\sup_{y \in D} [xy - h(y)] \leq g(x)$  for each  $x \in C$ . Moreover,  $g:C$  is a

closed convex function [even when  $g:C$  is not a closed convex function] because we have just seen that a conjugate transform  $g:C$  is always a closed convex function. In fact, it can be shown via the “separation theory for convex sets” applied to function epigraphs that both and in relation (79) become equalities if, and only if,  $g:C$  is itself a closed convex function; in which case  $g:C$  and  $h:D$  constitute a “conjugate pair”. However, time limitations require us to simply illustrate this important fact with examples. Prior to doing so, it is worth noting that that preceding derivations clearly remain valid when  $R^2$  is replaced by  $R^{n+1}$  and  $xy$  is interpreted as the inner product of vectors  $x$  and  $y$  in  $R^n$ .

Since

$$\sup_{x \in C} [yx - g(x)] = - \inf_{x \in C} [g(x) - yx],$$

it is geometrically clear that  $y \in D$  if, and only if, there is a number  $t$  such that the graph of the linear equation  $yx - t$  “supports”  $g:C$ , in the sense that it is “under” and comes “arbitrarily close” to  $(\text{epi } g)$ , as shown in the following two examples.



### SOME FUNCTION EPIGRAPH SUPPORTS

It is also clear that such a  $y \in D$  gives rise to the conjugate-transform value  $h(y) = t$ . In essence,  $h:D$  comes from the “envelope” of  $g:C$ , namely, all “linear supports”  $(y, t)$  of

$g:C$ . If there are no such supports,  $D$  is empty and  $g:C$  has no conjugate transform. For example,  $g(x) = -e^x$  has no supports and hence no conjugate transform when  $C = \mathbb{R}$ , but has supports and hence a conjugate transform when  $C$  is bounded from above, say  $C = (-\infty, +1]$ .

If a support  $(y, t)$  is tangential at a point, say  $(\underline{x}, g(\underline{x}))$ , and if  $g:C$  is differentiable at  $\underline{x}$ , note that  $y = g'(\underline{x})$ . On the other hand, note that not every derivative  $g'(\underline{x})$  at some  $\underline{x} \in C$  produces a support  $(y, t)$  with  $y = g'(\underline{x})$ . Also, note that the following example shows that tangential supports can occur at a point  $(\underline{x}, g(\underline{x}))$  where  $g:C$  is not differentiable.

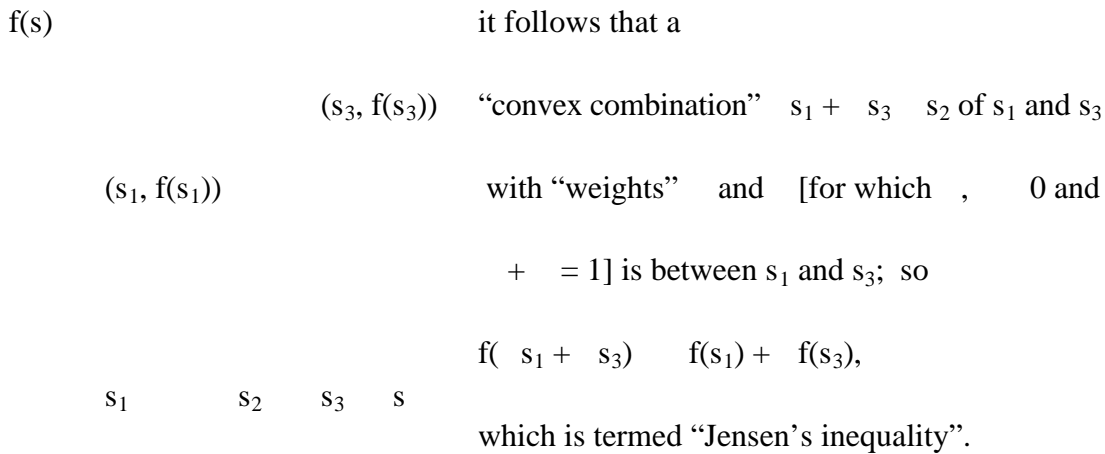
$g$  This function  $g:C$ , where  $C = [-1, +1]$  and  $g(x) = |x|$ , has  
 $-1$  infinitely many supports at the point  $(0, 0)$ , namely, all straight  
lines through  $(0, 0)$  with slopes  $y \in [-1, +1]$ . It also has  
 $-1$   $1 \ x$  infinitely many supports at the point  $(-1, +1)$ , namely, all  
 $g$  straight lines through  $(-1, +1)$  with slopes  $y \in (-\infty, 1]$ .  
 $-1$  Finally, it also has infinitely many supports at the point  
 $(+1, +1)$ , namely, all straight lines through  $(+1, +1)$  with  
 $-1$   $1 \ x$  slopes  $y \in [+1, +\infty)$ . None of these supports occur where  
 $g:C$  is differentiable, and no points at which  $g:C$  is  
differentiable produce a support. The tangential-support  
slopes  $y$  constitute the “range” of a multivalued function  $g$ ,  
 $h$  termed the “subderivative” of  $g$ , whose domain is a subset of  $C$

-1 [that is sometimes all of  $C$ ]. The range of  $g$ ,  $R$  for this example, is a subset of the domain  $D$  of the conjugate transform  $h:D$  [but  $D$  can also include other points, such as the slopes of asymptotic supports]. The graph and epigraph of  $h:D$  for this example are easily obtained analytically or geometrically [as shown]. Note that  $g:C$  is not a closed convex function [because it is not convex], but  $h:D$  is a closed convex function [as previously established in the general case]. The subderivative  $h$  of  $h:D$  is in fact just  $h(y)$ , except at points  $y$  where  $h:D$  is not differentiable; at which points it has a continuum of “multivalued” [a property that is shared by all convex functions]. The range of  $h$ , just  $[-1, +1]$  in this example, is of course the domain  $C$  of the conjugate transform  $g:C$  of  $h:D$ . The graph and epigraph of  $g:C$  are easily obtained either analytically or geometrically, as is the graph of its subderivative  $g$ . Note that  $C \subseteq C$  and  $g(x) \leq g(x)$  for each  $x \in C$  [as previously proved in the general case]. In fact, it can be shown that  $g:C$  is generally the “closed convex hull” of  $g:C$ , in that the  $(\text{epi } g)$  is the “closed convex hull” of the  $(\text{epi } g)$ , namely the intersect of all closed convex sets containing the  $(\text{epi } g)$ . Note also that the graph

$\{(x, y) \mid x \in C \text{ and } y = g(x)\}$  of  $g$  is a subset of the graph  $\Gamma = \{(x, y) \mid x \in C \text{ and } y = g(x)\}$  of  $g$ , and that both are “monotone” in the sense that  $(y^1 - y^2)(x^1 - x^2) \geq 0$  when points  $(x^1, y^1)$  and  $(x^2, y^2)$  belong to such a graph. In fact,  $\Gamma$  is the “completion” of  $\Gamma$ , in the sense that there is no “monotone curve” larger than  $\Gamma$  [i.e., properly containing  $\Gamma$  and hence  $\Gamma$ ]. These properties of  $\Gamma$  and  $\Gamma$  can be proved in the general case, as can the property that the graph of  $g$  is simply the “inverse” of the graph of  $h$  -- a fact that will enable us to calculate the conjugate transform  $h:D$  of  $g:C$  by “subdifferentiation” of  $g:C$  followed by “completion” of  $\Gamma$ , then “inversion” of the resulting  $\Gamma$ , and finally “integration” of  $\Gamma^{-1}$ . In “equilibrium problems” [rather than optimization problems], we shall see that “complete monotone curves” [rather than objective functions  $g:C$ ] arise from the modeling process.

Since we have just seen that convex functions and their subderivatives arise naturally from the conjugate transformation [even when the function being transformed is not convex], we need to take a careful look at convex functions and their subdifferentiability properties [including their differentiability properties].

It is not hard to show that a function  $f$  is convex only if its domain is a convex set [an “orthogonal projection” of the convex epigraph set]. In the single-variable case being considered, such a convex domain must either contain only a single point  $s$  or be an interval  $I$  [easily seen to be the only convex subsets of  $\mathbb{R}$ ]. If the domain contains only a single point  $s$ ,  $f$  is clearly closed [in addition to being convex] but  $f$  can not be differentiable at  $s$  [because  $s$  is not in the interior of the domain of  $f$ ]. However,  $f(s)$  is clearly  $\mathbb{R}$  and hence the graph of  $f$  is simply a vertical line -- one of the simplest types of complete monotone curves. On the other hand, if the domain is an interval  $I$ , given two points  $s_1$  and  $s_3$  in  $I$  with  $s_1 < s_3$ , the straight line segment connecting the corresponding (epi)graph points  $(s_1, f(s_1))$  and  $(s_3, f(s_3))$  must be in the epigraph of  $f:I$  [because  $f:I$  is assumed to be convex]. As illustrated below,



Since it is not hard to verify that the weight  $w_1 = \frac{s_3 - s_2}{s_3 - s_1}$  and the weight  $w_2 = \frac{s_2 - s_1}{s_3 - s_1}$ , we see that:

for a function  $f$  that is convex on an interval  $I$ , for each set of three points  $s_1 < s_2 < s_3$  in  $I$ ,

$$f(s_2) \leq \frac{s_3 - s_2}{s_3 - s_1} f(s_1) + \frac{s_2 - s_1}{s_3 - s_1} f(s_3) \quad \text{“Jensen’s inequality”} \tag{80}$$