DUALITY THEORY FOR COMPOSITE GEOMETRIC PROGRAMMING

by

Ya-Ping Wang

B.S. Mathematics, National Tsing-Hua University, 1973

M.A. Mathematical Statistics, Wayne State University, 1979

M.S. Applied Mathematics, Carnegie Mellon University, 1982

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Ya-Ping Wang

It was defended on

December 14, 2012

and approved by

Oleg A. Prokopyev, PhD, Associate Professor, Department of Industrial Engineering

Denis R. Saure, PhD, Assistant Professor, Department of Industrial Engineering

Zhi-Hong Mao, PhD, Associate Professor, Department of Electrical and Computer Engineering

Dissertation Director: Jayant Rajgopal, PhD, Associate Professor, Department of Industrial Engineering

DUALITY THEORY FOR COMPOSITE GEOMETRIC PROGRAMMING

Ya-Ping Wang, PhD

University of Pittsburgh, 2013

This research develops an alternative approach to the duality theory of the well-known subject of Geometric Programming (GP), and uses this to then develop a duality theory for Quadratic Geometric Programming (QGP), which is an extension of GP; it then develops a duality theory for (convex) Composite Geometric Programming (CGP), which in turn, is a generalization of QGP. The building block of GP is a special class of functions called *posynomials*, which are summations of terms, where the logarithm of each term is a linear function of the logarithms of its design variables. Such log-linear relationships often appear as an empirical fit in numerous engineering applications, most notably in engineering design. At other times, these relationships may simply follow the dictates of the laws of nature and/or economics (Varian 1978). Many functions that describe engineering systems are posynomials and hence GP is especially suitable for handling optimization problems involving such functions. For instance, the so-called Machining Economics Problem (MEP) for conventional metals can be handled successfully by GP (Tsai, 1986). The defining relationship of the tool life as a function of machining variables such as cutting speed, depth of cut, feed rate, tool change downtime, etc., is typically log-linear.

However, when the logarithm of the tool life for a given alloy is a *quadratic* (instead of a linear) function of the logarithms of the machining variables, GP is not able to handle this kind of MEP (Hough, 1978; Hough and Goforth, 1981a, b, c). Jefferson and Scott (1985) introduced

QGP to successfully handle the quadratic version of the MEP. They also gave primal-dual formulations of the QGP problem, an optimality condition, and three illustrative numerical examples of MEP (Jefferson and Scott 1985, p.144). A strong duality theorem for QGP was later proved by Fang and Rajasekera (1987) using a dual perturbation approach and two simple geometric inequalities.

However, more detailed duality theory for QGP and for CGP comparable to the development for GP by Duffin, Peterson, and Zener in chapters IV and VI of their seminal text (1967) are yet to be developed. This theory is rooted in the duality principle for óconjugateô pairs of convex functions, which translates the analysis of one optimization problem into an equivalent, yet very different optimization problem. It allows us to view the problem from two different angles, often with new insights and with other remarkable consequences such as suggesting an easier solution approach. In this thesis, we extend QGP problems to the more general CGP problems to account for the even more general log-convex (as opposed to log-linear) relationships. The aim of this dissertation is to develop a comprehensive duality theory for both QGP and CGP.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

In memory of my parents: Yue-Tao Wang and Yu-Guei Tian.

The fear of the LORD is the beginning of wisdom: and the knowledge of the holy is understanding. (**Proverbs 9:10**)

This dissertation was done in two stages with a very painful prolonged interruption of more than a decade in between. In the proposal stage, it was under the guidance of Dr. T.R. Jefferson with the title: *Composite Geometric Programming-Theory and Application*; and later in the actual writing stage it is under the guidance of Dr. J. Rajgopal with the current title: *Duality Theory for Composite Geometric Programming*.

The author hereby expresses his sincere gratitude towards the above two advisors for their patient advice and guidance over all these years. It was Dr. T.R. Jefferson who introduced me to the field of Petersonôs Generalized Geometric Programming, and this dissertation was actually inspired from a paper he coauthored with Dr. C.H. Scott on Quadratic Geometric Programming.

Then, after a number of years, the author communicated with Dr. J. Rajgopal, who took over as my new advisor and agreed to let me resume writing this dissertation under the current title. The author gratefully appreciates his kindness and the departmentøs approval for giving him this second chance.

The author appreciates the Department of Industrial Engineering of this school for the 4 year teaching assistantships he received during his first stage of Ph.D study and for occasional tuition waievers during his second stage. He also appreciates the emergency loans during his financial crisis at the later stage from the following friends: Dr. Jen-Luen Chu, Dr. Zengbiao Qi, Dr. Li Zou, and Dr. Ning Zhang, and relatives: Dr. Yue Chen and Dr. Wei Tian, without which this dissertation could not possibly be completed.

Lastly, I would like to express my thankfulness to the late Professor R.J. Duffin for getting me interested in the prototype posynomial programming from taking an introductory course he offered on this subject at Carnegie-Mellon University. Most of the ideas I used while developing this duality theory were obtained from the writings of Dr. J.M. Borwein, Dr. E.L. Peterson, and Dr. R.T. Rockafellar. The author expresses his sincere thankfulness for their pioneering work.

1.0 INTRODUCTION

This dissertation develops a comprehensive duality theory for (convex) Composite Geometric Programming (*CGP*) as well as Quadratic Geometric Programming (*QGP*). QGP is an extension of the well-known subject of Geometric Programming (*GP*), while CGP is in turn, a generalization of QGP. QGP was first introduced [Jefferson and Scott, 1985] as a means to solve a Machining Economics Problem (*MEP*) where the logarithm of the tool life for a given alloy is modeled as a *quadratic* function of the logarithms of machining variables such as cutting speed, depth of cut, feed rate, tool change downtime, etc. The MEP for conventional metals can be handled successfully by GP (Tsai, 1986), since the defining relationship is log-linear. But with alloys, the relationship becomes log-quadratic, and GP is not able to handle this case [Hough, 1978; Hough and Goforth, 1981a,b,c]. One has to then resort to QGP [Hough and Chang, 1998]. Jefferson and Scott [1985, p.144] gave the primal-dual formulations of the QGP problem, an optimality condition, and three illustrative numerical examples of MEP. The main duality theorem for QGP was later proved by Fang and Rajasekera [1987] using a dual perturbation approach and two geometric inequalities. Jefferson et al. [1990] extend QGP problems to the more general CGP problems to account for the more general log-*convex* relationships, where primal-dual formulations of the CGP problems are provided, along with an optimality condition (Theorems 4.1 and 5.1, pp.109-113), and two prior examples [Lidor and Wilde, 1978; Beightler and Phillips, 1976] to illustrate the idea of the dual solution approach. However, more detailed duality theory for QGP and for CGP comparable to the development for traditional GP by Duffin, Peterson, and Zener in chapters IV and VI of their seminal text [1967] has never been developed. This theory is rooted in the duality principle for óconjugateô pairs of convex functions, which translates the analysis of one optimization problem into an equivalent, yet very different optimization problem. It allows us to view the problem from two different angles, often with new insights and with other remarkable consequences such as suggesting an easier solution approach. The aim of this dissertation is to develop a comprehensive duality theory for CGP, and for several special cases of CGP including Exponential GP (*EGP*), QGP and l_pGP.

Additional application of CGP can be found in Scott et al. [1996]. In a straightforward manner, CGP formulations can also be further generalized to Composite Convex Programming (CCP) problems [Scott and Jefferson, 1991].

1.1 OPTIMIZATION PROBLEMS

Optimization problems are concerned with finding the maximum or minimum of a function subject to some set of constraints. In this dissertation, we shall adopt minimization as the main vehicle of exposition, leaving the parallel statements for maximization as unstated results, since maximizing a function is simply equivalent to minimizing its negative. We shall work in the real space R^n of column n-vectors equipped with the standard inner product. By an Optimization Problem (*OP*) in R^n we shall mean mathematically:

$$
\text{(1.1)} \quad \text{(OP)} \qquad \begin{cases} \underset{x}{inf}\ f_0(\boldsymbol{x}) \\ s.t.\ f_i(\boldsymbol{x}) \le 0,\ i \in I_1 \\ \quad f_i(\boldsymbol{x}) = 0,\ i \in I_2 \\ \qquad \boldsymbol{x} \in C \end{cases}
$$

where $\boldsymbol{x}=(x_1,$é $,x_n)^t \in R^n$ represents a *design* or *decision vector*, $C \subset R^n$ is a nonempty set of ñmeaningfulò values for the design vector, which often takes the form of a product set $C = C_1 \times \cdots \times C_n$, where each C_j is a closed interval in R with nonempty interior. I_1 and I_2 are two disjoint finite index sets with $I_1 \cup I_2 = I$. Each problem-defining function $f_i : C \to R,\ i \in \{0\} \cup I$ is real-valued, and the inequalities $f_i(\boldsymbol{x}) \le 0,\ i \in I_1$ and the equations $f_i(\boldsymbol{x}) = 0, i \in I_2$ representing design specifications or restrictions (physical, technical, financial, etc.) are called *constraints*. The classification of constraints into inequality ones and equality ones is only a matter of formality. Mathematically speaking, there is no loss of generality in assuming that all of the constraints (if any) are of inequality type, i.e. $I_2=\phi$, as $f_i(\boldsymbol{x}) = 0 \Leftrightarrow f_i(\boldsymbol{x}) \le 0$, and $-f_i(\boldsymbol{x}) \le 0$. There is also clearly no loss of generality in assuming that the objective function f_0, is linear; this can be done by mnimizing a new varaible x_0 and adding the new constraint $f_0(\boldsymbol{x})$ Òx_0 .

Problem (OP) (1.1) is concerned with seeking the *infimum* of the *objective function* f_0 over its *feasible region* $S := \left\{ \boldsymbol{x} \in C \middle| f_i(\boldsymbol{x}) \le 0, i \in I_1; f_i(\boldsymbol{x}) = 0, i \in I_2 \right\}$, $\inf_S f_0 := \inf\{ f_0(\boldsymbol{x}) \mid x \in S \}$, and its set of *minimizers*, $\text{argmin}_S f_0 := \left\{ \boldsymbol{x} \in S \middle| f_0(\boldsymbol{x}) = \inf_S f_0 \right\}$, when $\inf_S f_0 < \infty$. It is said to be *feasible*, if S is non-empty and *infeasible*, if otherwise. When (OP) is feasible (SÍ ϕ) and $\inf_S f_0$ = -Ð, we say that (OP) has an *unbounded* infimum. By convention, we define $\inf_S f_0$ = +Ð, if (OP) is infeasible (S=ϕ). The major issues being considered in this problem (OP) are usually arranged into three phases: *feasibility, optimality,* and *sensitivity* [Eiselt and Sandblom, 2007, p.57]. Note that when I_1=ϕ, C=R^n, and all of the problem-defining functions f_i, $i \in \{0\} \cup I$ are differentiable, (1.1) is simply a classical optimization problem, which is typically solved by the method of Lagrange multipliers [Luenberger, 1984, p.300]. It is termed a *Linear Program* (LP) or a Linear Optimization Problem, if all of the problem-defining functions f_i are affine (i.e. linear plus constant) and each C_i=[0,Ð), or =(-Ð,Ð) :=R; and an *Ordinary Convex Program* (OCP), if the set C and all of the functions f_i, $i \in \{0\} \cup I_1$, are convex, and f_i, $i \in I_2$, are affine. In the literature, a *Convex Optimization Problem* (COP) is referred to as a problem of minimizing a convex objective function over a convex feasible region, say, S. Since the feasible region of an (OCP) is clearly a convex set, an (OCP) is certainly a (COP); and the converse is also true: since the condition that $x \in S$ is equivalent to $i_S(x) \le 0$, where i_S is the *indicator function* defined by $i_S(x) = 0$, if $x \in S$; and =+∞, if $x \notin S$. However, this kind of formulation using an indicator function is not computationally very useful; a better formulation is to use a more tractable convex function. For instance, suppose S is the closed convex set in R^2 given by $S := \{(x, y) \in R^2 \mid xy \ge 1, x \ge 0, y \ge 0\}$. By an algebraic trick, we can derive that:

$$
\begin{aligned}
& xy \ge 1, x \ge 0, y \ge 0 \\
\Leftrightarrow\ & xy \ge 1, x + y \ge 0 \\
\Leftrightarrow\ & (x-y)^2 + 4 \le (x+y)^2, x + y \ge 0 \\
\Leftrightarrow\ & \sqrt{(x-y)^2 + 4} \le x + y
\end{aligned}
$$

Hence $S = \{(x, y) \in R^2 \mid g(x, y) \le 0\}$, where $g(x, y) := \sqrt{(x-y)^2 + 4} - x - y$ is a *differentiable* convex function; whereas the indicator function i_S is not differentiable.

In the LP case, the feasible region S is a *polyhedral* convex set, i.e., a set of the form $\left\{ \boldsymbol{x} \in R^n \middle| \boldsymbol{Ax} \le \boldsymbol{b} \right\}$, for some matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$, its set of minimizers, $\arg\min_{x \in S} f_0(\boldsymbol{x})$, is an *exposed face* of S with the form $S \cap H$, where $H =: \left\{ \boldsymbol{x} \in R^n \middle| f_0(\boldsymbol{x}) = \alpha \right\}$, for $\alpha = \inf_S f_0(\boldsymbol{x})$ (Rockafellar, 1970, p.162); and in the OCP case, both the feasible region S and its set of minimizers, $S \cap \left\{ \boldsymbol{x} \in R^n \middle| f_0(\boldsymbol{x}) \le \alpha \right\}$, are convex subsets of R^n. In the next section, we shall show that prototype GP (the Posynomial programs) can be easily identified as ordinary convex programs.

In summary, optimization is a branch of mathematics dealing with techniques for maximizing or minimizing an objective function subject to linear, nonlinear, and/or integer constraints on the variables [Luenberger, 1984, p.1]. It is a rich and thriving mathematical discipline, which has a very broad area of successful applications, most notably in engineering, statistics, economics, and mathematics itself. To address these applications, we need to formulate real-world problems as *mathematical models* such as (1.1), develop techniques (*algorithms*) for solving the models, and write *software* that execute the algorithms on computers based on the mathematical theory [Luenberger, 1984, p. *xxxiii*].

Since its inception in the early 1960's, GP has been well accepted by the engineering community as a viable means for optimally solving a number of engineering design problems, a task that is an important and challenging one for an engineer: the goal is to design a device or a system that performs a given function in an optimal way, e.g. at a minimum cost, a maximum production rate, a minimum weight, etc. Examples of engineering design cut across virtually every major engineering discipline: electrical, mechanical, civil, chemical, and industrial. Some good sources of early examples of GP application are: [Duffin et. al. 1967, chapter V], [Zener, 1971], [Beightler and Phillips, 1976, chapter 11], [Dinkel et.al. 1977], and [Wilde, 1978]. More recently, GP has found applications in entropy optimization [Fang et.al. 1997, Scott and Fang, 2001], and in probability and statistics, finance and economics, control theory, circuit design, and communication systems [Chiang, 2005]. Many applications of *OCP* to engineering problems are also listed in Boyd and Vandenberghe, [2004].

The Simplex method for solving *LP* problems was developed by Dantzig in 1947. In 1972 Klee and Minty showed that its worst case performance was exponential [Dantzig and Thapa, 1997]. However, this rarely happens in practice and the method has remained popular because of its practical efficiency. In 1979, Khachian developed a polynomial time Ellipsoidal algorithm, but unfortunately, it was not practically efficient. Five years later, Karmarkar [1984] developed a practical polynomial time projective method ï this in turn gave rise to *interior point methods* and today, both simplex and interior point methods coexist for solving *LP*s. Before the close of the last century, Nesterov and Nemirovski [1994] and Ye [1997] found that the family of interior point methods can also be used to solve the much broader class of *COP* problems. This also benefits the area of combinatorial optimization and global optimization, where *COP* is used to find bounds on the optimal value, and to find approximate solutions. Nowadays, some *COP* problems such as semi-definite programs and conic quadratic programs can be solved by these new methods almost as easily as *LP*s [Dantzig and Thapa, 1997, p. *xi*]. Convex Programs are generally much easier to solve than the non-convex ones, and a local solution is automatically a global solution. Indeed, the computational effort required to solve them is vastly different, as was pointed out by Ben-Tal and A. Nemirovski [2001, p. *xii*, p.336]:

"…Under minimal additional computability assumptions (which are satisfied in basically all applications), a convex optimization problem is computationally tractable—the

computational effort required to solve the problem to a given accuracy grows moderately with the dimensions of the problem and the required number of accuracy digits…. In contrast to this, general-type non-convex problems are too difficult for numerical solution; the computational effort required to solve such a problem, by the best numerical methods known, grows prohibitively fast with the dimensions of the problem and the number of accuracy digits…"

GP problems in convex form are also solvable in polynomial time by the interior-point methods ([Nesterov and Nemirovski, 1994, pp.229-232], [Kortanek, 1996]), and user-friendly software for GP is available online (www.mosek.com).

1.2 GP AS A SUBCLASS OF CONVEX PROGRAMS

Notational convention: In this thesis, the notation $\boldsymbol{x} = [x_j]_{j=1}^m = [x_1, \cdots x_m]^t$ shall denote a column vector $\boldsymbol{x}$ whose jth component is x_j, where the superscript t represents transpose. The relationship $\boldsymbol{x} > \boldsymbol{0}$ means $x_j > 0, \forall j$.

Below we shall define *GP* and look at it from the perspective of an Ordinary Convex Program (*OCP*).

Posynomial Program or **Prototype Geometric Program (*GP*) or GP in design space:**

(1.2) (GP)
$$\begin{cases} \inf_{\boldsymbol{t} \in R^m} G_0(\boldsymbol{t}) \\ s.t.\ G_k(\boldsymbol{t}) \le 1, \quad k = 1, \ldots, p \\ \quad U_i(\boldsymbol{t}) = 1, \quad i = n+1, \ldots, \tilde{n} \\ \quad\quad t_j > 0, \quad j = 1, \ldots, m \end{cases}$$

where $\boldsymbol{t} =: [t_j]_{j=1}^{m} \in R^m$ represents a *design vector* whose values are sought and which must be positive in order to be meaningful; each m-variate power function

$$(1.3) \qquad U_i(\boldsymbol{t}) =: C_i t_1^{a_{i1}} \ldots t_m^{a_{im}} = C_i \prod_{j=1}^{m} t_j^{a_{ij}}, \qquad \text{for } i = 1, \ldots, \tilde{n}$$

with positive *cost* coefficient C_i, and arbitrary real *technological* coefficient a_{ij}, is called a *posynomial term*; each m-variate problem-defining function

$$(1.4) \qquad G_k(\boldsymbol{t}) =: \sum_{i \in [k]} U_i(\boldsymbol{t}) = \sum_{i \in [k]} C_i \prod_{j=1}^{m} t_j^{a_{ij}}, \qquad \text{for } k = 0, 1, \cdots p$$

is called a *posynomial.* The LHS of each of the constraints defined by $U_i(\boldsymbol{t})=1$ has a single posynomial term; this special case of a posynomial is referred to as a *monomial*. Thus in (GP), one seeks to minimize a posynomial, subject to finitely many unit upper bound posynomial inequality constraints, plus perhaps some additional monomial equality constraints. We define

I:={1,é ,n}: the index set of terms in the objective posynomial and the p (posynomial) constraints;

$\tilde{I} := \{1, \ldots, \tilde{n}\}$ with nÒ$\tilde{n}$: the index set of *all* terms including any in monomial constraints;

K=:{1,é ,p}: the index set of the *posynomial* constraints;

J:={1,é ,m}: the index set of design variables;

[0]={1,é ,n_0}, [1]={n_0+1,é ,n_1}, é , [k]= {n_{k-1}+1,é ,n_k},é ,[p]={n_{p-1}+1,é ,n}, where 1Òn_0<n_1<é <n_p=n: the block index subsets of the objective and each constraint posynomial; note that for each $k \in K$, block [k] has size $|[k]| = n_k - n_{k-1} =: \Delta n_k$, and [0] has size $n_0 := \Delta n_0$, so that $\sum_{k=0}^{p} \Delta n_k = n_0 + n_p - n_0 = n_p = n$ and I is thus partitioned into $[0] \cup [1] \cup \cdots \cup [p]$.

The vector $\tilde{\boldsymbol{C}} =: (C_1, \cdots, C_{\tilde{n}})^T$ is called the *cost vector*, $\tilde{\boldsymbol{A}}$ with i^{th} row $\boldsymbol{a}^i =: (a_{i1}, \cdots, a_{im})$, for $i = 1, \ldots, \tilde{n}$, is called the *exponent matrix*. The data for the program (GP) is the matrix $(\tilde{\boldsymbol{A}} \quad \tilde{\boldsymbol{C}}) : \tilde{n} \times (m+1)$, and the partition structure of I into the block index subsets $[k]$. We may partition $(\tilde{\boldsymbol{A}} \quad \tilde{\boldsymbol{C}})$ into: $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \breve{\boldsymbol{A}} & \breve{\boldsymbol{C}} \end{pmatrix}$ with $(\boldsymbol{A} \quad \boldsymbol{C})$ being its first n rows, and $(\breve{\boldsymbol{A}} \quad \breve{\boldsymbol{C}})$ the remaining rows. We further partition $(\boldsymbol{A} \quad \boldsymbol{C})$ into sub-matrices: $(\boldsymbol{A}^{[k]} \; \boldsymbol{C}^{[k]})$ according to the partition structure of I into $[k]$. The above model of GP is slightly more general than what is commonly seen in the GP literature in that we also take into account *monomial* equality constraints to make our model slightly more versatile. Without these constraints (i.e., with $n = \tilde{n},\ I = \tilde{I},\ (\boldsymbol{A} \quad \boldsymbol{C}) = (\tilde{\boldsymbol{A}} \quad \tilde{\boldsymbol{C}})$), our model is the same as that in Duffin et.al. [1967].

The basic reason for calling the treatment of such problems GP [Duffin and Peterson, 1966] is because (i) they employed a so-called *geometric inequality* (a simple generalization of the well-known Arithmetic-Geometric Mean inequality) as the main tool in the proof of the duality theory for GP, and (ii) because of its intimate connections with geometric means and with some geometrical concepts such as orthogonality of vectors.

Clearly, program GP in its design space is a differentiable optimization problem, since both $U_i(\boldsymbol{t})$ and $G_k(\boldsymbol{t})$ are differentiable with partial derivatives and gradients (column vectors) given by:

$$(1.5) \qquad \begin{cases} \dfrac{\partial U_i(\boldsymbol{t})}{\partial t_j} = \dfrac{a_{ij} U_i(\boldsymbol{t})}{t_j},\ \forall i \& j; \qquad \dfrac{\partial G_k(\boldsymbol{t})}{\partial t_j} = \displaystyle\sum_{i \in [k]} \dfrac{\partial U_i(\boldsymbol{t})}{\partial t_j} = \dfrac{1}{t_j} \sum_{i \in [k]} a_{ij} U_i(\boldsymbol{t}),\ \forall k \& j \\ \nabla U_i(\boldsymbol{t}) = U_i(\boldsymbol{t}) \left[\dfrac{a_{ij}}{t_j} \right]_{j=1}^{m}, \forall i; \quad \nabla G_k(\boldsymbol{t}) = \left[\dfrac{1}{t_j} \displaystyle\sum_{i \in [k]} a_{ij} U_i(\boldsymbol{t}) \right]_{j=1}^{m} = \displaystyle\sum_{i \in [k]} U_i(\boldsymbol{t}) \left[\dfrac{a_{ij}}{t_j} \right]_{j=1}^{m}, \forall k \end{cases}$$

Using the following change of variables and parameters,

(1.6) $u_i(\boldsymbol{z}) =: \ln U_i(\boldsymbol{t}), g_k(\boldsymbol{z}) =: \ln G_k(\boldsymbol{t}),$ where $\boldsymbol{z} = (z_1, \cdots, z_m)^t$, $z_j := \ln t_j$, $c_i := \ln C_i, \forall i, j, k$,

We can turn equations (1.3) and (1.4) into:

(1.7) $$u_i(\boldsymbol{z}) = \boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z} + c_i,\ i = 1, \dots, \tilde{n}; \quad \text{and} \quad g_k(\boldsymbol{z}) = \ln\left\{\sum\nolimits_{i\in[k]} \exp u_i(\boldsymbol{z})\right\}, k = 0, 1, \cdots, p$$

Clearly $g_k(\boldsymbol{z})$ is also differentiable with partial derivatives and gradient:

(1.8) $$\begin{cases} \dfrac{\partial g_k(\boldsymbol{z})}{\partial z_j} = \dfrac{1}{G_k(\boldsymbol{t})} \displaystyle\sum_{i\in[k]} a_{ij} U_i(\boldsymbol{t}) = \dfrac{\partial G_k(\boldsymbol{t}) / \partial t_j}{G_k(\boldsymbol{t}) / t_j},\ \forall k \& j, \\ \nabla g_k(\boldsymbol{z}) = \dfrac{1}{G_k(\boldsymbol{t})} \left[\displaystyle\sum_{i\in[k]} a_{ij} U_i(\boldsymbol{t})\right]_{j=1}^{m} = \dfrac{1}{G_k(\boldsymbol{t})} \left[\displaystyle\sum_{i\in[k]} U_i(\boldsymbol{t}) (\boldsymbol{a}^{i})^t\right],\ \forall k. \end{cases}$$

Surprisingly, the function $g_k(\boldsymbol{z}) = \ln\left\{\sum\nolimits_{i\in[k]} \exp\left(u_i(\boldsymbol{z})\right)\right\}$ is also convex in $\boldsymbol{z} = \left(z_{1,\dots,} z_m\right)^t$ for all k [Duffin et. al. 1967, p.58, exercise 11(a)]. By taking logarithms of the objective function and of both sides of each constraints and replacing the design vector $\boldsymbol{t}$ by $\boldsymbol{z}$, (*GP*) (1.2) is readily turned into the following differentiable Ordinary Convex Program (GP_z) with the set C in (1.1) being the space R^m .

GP in log-design space [Duffin et. al. 1967, p.125, exercise 5]:

(1.9) $$(\mathrm{GP}_z) \qquad \begin{cases} \inf\limits_{z\in R^m} g_0(\boldsymbol{z}) \\ s.t.\ \ g_k(\boldsymbol{z}) \le 0, \quad k \in K \\ \qquad \boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z} + c_i = 0,\ i = n+1, \dots, \tilde{n} \quad (\Leftrightarrow \breve{\boldsymbol{A}}\boldsymbol{z} + \breve{\boldsymbol{c}} = \boldsymbol{0}) \end{cases}$$

where $\breve{\boldsymbol{A}}$ is defined by the last $\tilde{n} - n$ rows of the exponent matrix $\breve{\boldsymbol{A}}$, and $\breve{\boldsymbol{c}} =: (c_{n+1}, \cdots, c_{\tilde{n}})^T$. The two programs ($\mathrm{GP}_z$) and (GP) are clearly equivalent. Their respective optimum solutions $\boldsymbol{z}^*$, and $\boldsymbol{t}^*$, and respective optimum values are in one to one correspondence through the relationship $z_j^* = \ln t_j^*, \forall j,$ and $\inf\left(\mathrm{GP}_z\right) = \ln\left[\inf\left(\mathrm{GP}\right)\right]$.

Note that the original program (GP) is generally non-convex, but its hidden convexity can be easily induced by the above approach.

Other convex reformulations of (*GP*)

Transformed Primal Program A_z (Duffin et. al. 1967, p.82)

(1.10) $$(A_z)\quad \begin{cases} \inf\limits_{z\in R^m} \sum_{i\in[0]} \exp(\boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z}+c_i) \\ \text{s.t. } \sum_{i\in[k]} \exp(\boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z}+c_i) \le 1, \quad k\in K \\ \qquad \boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z}+c_i = 0, i=n+1,\ldots,\tilde n \;\; (\Leftrightarrow \breve{\boldsymbol{A}}\boldsymbol{z}+\breve{\boldsymbol{c}}=\boldsymbol{0}) \end{cases}$$

where $\sum_{i\in[k]} \exp(\boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z}+c_i) = \exp g_k(\boldsymbol{z}),\ k=0,1,\ldots,p$ are also convex. The above two reformulations of (*GP*) both bring out its hidden convexity. A further changes of variables $x_i =: \boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z},$ for $i=1,\ldots,\tilde n$ $(\Leftrightarrow \tilde{\boldsymbol{x}} =: \tilde{\boldsymbol{A}}\boldsymbol{z})$; recall that $\tilde{\boldsymbol{A}}: \tilde n \times m$ is the exponent matrix) also bring out their hidden *separability* and change the programs (GP_z) and A_z into two other equivalent convex programs (GP_x) and $A_{\boldsymbol{x}}$, respectively.

Transformed Primal Program $A_{\boldsymbol{x}}$ (Duffin et. al. 1967, p.167)

(1.11) $$(A_{\boldsymbol{x}})\quad \begin{cases} \inf\limits_{\tilde{\boldsymbol{x}}\in R^{\tilde n}} \sum_{i\in[0]} \exp(x_i+c_i) \\ s.t. \ \sum_{i\in[k]} \exp(x_i+c_i) - 1 \le 0, \ k\in K \\ \qquad \tilde{\boldsymbol{x}} \in \boldsymbol{\mathscr{P}}, \ \breve{\boldsymbol{x}} = -\breve{\boldsymbol{c}} \end{cases}$$

where $\tilde{\boldsymbol{x}} =: (\boldsymbol{x}, \breve{\boldsymbol{x}}) = (\boldsymbol{Az}, \breve{\boldsymbol{A}}\boldsymbol{z}) = \tilde{\boldsymbol{A}}\boldsymbol{z}$, $\boldsymbol{\mathscr{P}} =: \left\{\tilde{\boldsymbol{A}}\boldsymbol{z} \mid \boldsymbol{z}\in R^m\right\}$ is the column space of $\tilde{\boldsymbol{A}}$ (also called the *primal space* of the program), and $\boldsymbol{x}$ consists of the first n components of $\tilde{\boldsymbol{x}}$.

GP in GGP form [Peterson, 1976]

$$(1.12) \qquad (GP_{\boldsymbol{x}}) \qquad \begin{cases} \inf_{\tilde{\boldsymbol{x}} \in R^{\tilde{n}}} \ \ln\left[\sum_{i\in[0]} \exp(x_i + c_i)\right] + i_{\breve{c}}(\breve{\boldsymbol{x}}) \\ s.t. \ \ \ln\left[\sum_{i\in[k]} \exp(x_i + c_i)\right] \le 0, \ k = 1,\dots, p \\ \qquad \tilde{\boldsymbol{x}} \in \mathscr{P} \end{cases}$$

where $i_{\breve{c}}(\breve{\boldsymbol{x}}) =: 0$, if $\breve{\boldsymbol{x}} = -\breve{\boldsymbol{c}}$; and $=:\infty$, if $\breve{\boldsymbol{x}} \neq -\breve{\boldsymbol{c}}$, is an indicator function. This program ($GP_{\boldsymbol{x}}$) is an instance of Peterson's *GGP* model to which his *GGP* duality theory is readily applicable.

The above two programs ($GP_{\mathbf{z}}$) and ($GP_{\mathbf{x}}$) are respectively obtained from the transformed programs $A_{\mathbf{z}}$ and $A_{\mathbf{x}}$ by applying logarithmic transformations. The original duality theory for GP as developed by Duffin and Peterson was based on the transformed primal programs $A_{\mathbf{z}}$ and $A_{\mathbf{x}}$ (without the monomial equality constraints) and their corresponding dual programs. In Chapter 2 we shall review this duality theory based on ($GP_{\mathbf{z}}$) and ($GP_{\mathbf{x}}$) instead. Program ($GP_{\mathbf{x}}$) is equivalent to ($GP_{\mathbf{z}}$) (and hence to (GP) as well) in the following sense: If ($\boldsymbol{x}^*$, $\breve{\boldsymbol{x}}^* = -\breve{\boldsymbol{c}}$) and $\boldsymbol{z}^*$ are their respective optimum solutions, then $\boldsymbol{x}^*=\boldsymbol{A}\boldsymbol{z}^*$, which is not a one to one correspondence unless the matrix A has full column rank (i.e., $\boldsymbol{A}$ has nullity zero). Their optimum values are equal: inf ($GP_{\boldsymbol{x}}$) = inf (GP_{z}).

As a matter of convenience, in this thesis we shall simply call the important function $geo_n(\boldsymbol{x}) := \ln[\sum_{i=1}^{n} \exp x_i] : R^n \to R$ a *geometric function* (also called *logexp*($\boldsymbol{x}$) in [Rockafellar and Wets, 2004]). We will usually omit the subscript *n*, unless there is a danger of ambiguity. Trivially, one has $geo_1(\boldsymbol{x}) = \boldsymbol{x}$. This function is obviously *strictly isotone*: any points $\boldsymbol{x}^1 \le \boldsymbol{x}^2$ in R^n satisfy $geo(\boldsymbol{x}^1) \le geo(\boldsymbol{x}^2)$, and the latter equality holds only when $\boldsymbol{x}^1 = \boldsymbol{x}^2$. It has a unique *linearity vector* (up to a scalar multiple) [Rockafellar, 1970a, Theorem 8.8], namely, the sum vector $\boldsymbol{1}$=(1,é ,1)$^{\mathrm{T}}$: $geo(\boldsymbol{x}+t\boldsymbol{1}) = geo(\boldsymbol{x})+\boldsymbol{t},\ \forall \boldsymbol{x} \in R^n, \forall t \in R$. So it is strictly convex on any line that is not parallel to the vector ***1.*** It plays a dominant role in our development of GP duality theory based on the models (GP$_z$) and (GP$_x$). The composite of $geo_n(\boldsymbol{x})$ with any affine function $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z}+\boldsymbol{c}$, where $\boldsymbol{A}: n\times m,\ \boldsymbol{c}: n\times 1$ is another convex function that is given by $g(\boldsymbol{z}) =: geo_n(\boldsymbol{A}\boldsymbol{z}+\boldsymbol{c}): R^m \to R^1$.

Thus the problem functions, $g_k(\boldsymbol{z})$ in the above program (GP$_z$) maybe written as $geo(\boldsymbol{A}^{[k]}\boldsymbol{z}+\boldsymbol{c}^{[k]})$ and the program itself maybe rephrased as:

(1.13) $\quad$ $(\mathrm{GP}_z) \quad \inf_{z\in R^m}\ geo(\boldsymbol{A}^{[0]}\boldsymbol{z}+\boldsymbol{c}^{[0]})\ s.t.\ geo(\boldsymbol{A}^{[k]}\boldsymbol{z}+\boldsymbol{c}^{[k]}) \le 0,\ \forall k \in K,\ \breve{\boldsymbol{A}}\boldsymbol{z}+\breve{\boldsymbol{c}} = \breve{\boldsymbol{0}}$

where $\boldsymbol{A}^{[k]}$ is the k^{th} component sub-matrix of A, i.e., obtained by discarding all rows $\boldsymbol{a}^i$ of $\boldsymbol{A}$ for which i is not in [k], and the sub-vector $\boldsymbol{c}^{[k]}$ of $\boldsymbol{c}$ is similarly defined. Likewise, (GP$_x$) maybe rephrased as:

(1.14) $\quad$ $(\mathrm{GP}_{\tilde{x}}) \quad \inf_{\tilde{x}\in R^{\tilde{n}}}\ geo(\boldsymbol{x}^{[0]}+\boldsymbol{c}^{[0]}) + I_{-\breve{\boldsymbol{c}}}(\breve{\boldsymbol{x}})\ s.t.\ geo(\boldsymbol{x}^{[k]}+\boldsymbol{c}^{[k]}) \le 0,\ \forall k \in K,\ \tilde{\boldsymbol{x}} \in \mathscr{P}$

where $\forall k,\ \boldsymbol{x}^{[k]} =: \boldsymbol{A}^{[k]}\boldsymbol{z}$ is the sub-vector of $\boldsymbol{x}$ obtained by striking out all of its components x_i for which i is not in [k], and $\mathscr{P} =: \left\{\tilde{\boldsymbol{A}}\boldsymbol{z} \mid \boldsymbol{z} \in R^m\right\}$.

Meaning of the technological coefficient a_{ij} and of partial derivative $\partial g_k(z)/\partial z_j$:

Since $a_{ij} = \dfrac{\partial u_i(\boldsymbol{z})}{\partial z_j} = \dfrac{\partial \ln U_i(\boldsymbol{t})}{\partial \ln t_j} = \dfrac{\partial \ln U_i(\boldsymbol{t})/\partial t_j}{\partial \ln t_j/\partial t_j} = \dfrac{t_j}{U_i(\boldsymbol{t})}\cdot\dfrac{\partial U_i(\boldsymbol{t})}{\partial t_j} \approx \dfrac{\Delta U_i(\boldsymbol{t})/U_i(\boldsymbol{t})}{\Delta t_j/t_j}\cdot\dfrac{100\%}{100\%}$, a one percent increase in the value of the j^{th} design variable t_j will cause approximately a_{ij} percent increase in the value of the i^{th} term $U_i(\boldsymbol{t})$.

Since $\dfrac{\partial g_k(\boldsymbol{z})}{\partial z_j} = \dfrac{\partial \ln G_k(\boldsymbol{t})}{\partial \ln t_j} = \dfrac{\partial \ln G_k(\boldsymbol{t})/\partial t_j}{\partial \ln t_j/\partial t_j} = \dfrac{t_j}{G_k(\boldsymbol{t})}\cdot\dfrac{\partial G_k(\boldsymbol{t})}{\partial t_j} \approx \dfrac{\Delta G_k(\boldsymbol{t})/G_k(\boldsymbol{t})}{\Delta t_j/t_j}\cdot\dfrac{100\%}{100\%}$

and $\left(\frac{\partial g_k(\boldsymbol{z})}{\partial z_j}\right) G_k(\boldsymbol{t}) = \sum_{i\in[k]} a_{ij} U_i(\boldsymbol{t})$, by(1.8), a one percent increase in the value of the jth design variable t_j will cause approximately $\sum_{i\in[k]} a_{ij} U_i(\boldsymbol{t}) / G_k(\boldsymbol{t}) = (\partial g_k(\boldsymbol{z}) / \partial z_j)$ percent increase in the value of $G_k(t)$ and approximately $\sum_{i\in[k]} a_{ij} U_i(\boldsymbol{t})$ absolute increase in the value of $G_k(\boldsymbol{t})$.

Example 1.2.1 A three-term, two-variable (*n=3, m=2*) GP example

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \; U_i(\boldsymbol{t}) = C_i t_1^{a_{i1}} t_2^{a_{i2}}, \; i\text{=1,2,3.}$$

There are 4 possible GP formulations for this matrix A:

1) $\inf_{t>0} U_1(\boldsymbol{t}) + U_2(\boldsymbol{t} + U_3(\boldsymbol{t}),$

2) $\inf_{t>0} U_1(\boldsymbol{t}) + U_2(\boldsymbol{t}) \; s.t. \; U_3(\boldsymbol{t}) \le 1,$

3) $\inf_{t>0} U_1(\boldsymbol{t}) \; s.t. \; U_2(\boldsymbol{t}) + U_3(\boldsymbol{t}) \le 1,$

4) $\inf_{t>0} U_1(\boldsymbol{t}) \; s.t. \; U_2(\boldsymbol{t}) \le 1, \; U_3(\boldsymbol{t}) \le 1.$

Each of them corresponds to a different partition structure of *I*:

1) $I = [0] = \{1,2,3\},$

2) $I = \{1,2\} \cup \{3\},$

3) $I = \{1\} \cup \{2,3\},$

4) $I = \{1\} \cup \{2\} \cup \{3\},$

Convex reformulations GP_z of the above problems after the following variable transformations:

$$u_i(\boldsymbol{z}) = \boldsymbol{a}^i \boldsymbol{z} + c_i = a_{i1} z_1 + a_{i2} z_2 + c_i, \; i\text{=1,2,3.}$$

1) $\inf_z e^{\boldsymbol{a}^1 z + c_1} + e^{\boldsymbol{a}^2 z + c_2} + e^{\boldsymbol{a}^3 z + c_3},$

2) $\inf_z e^{\boldsymbol{a}^1 z + c_1} + e^{\boldsymbol{a}^2 z + c_2}, \; s.t. \; \boldsymbol{a}^3 \boldsymbol{z} + c_3 \le 0,$

3) $\inf_z \boldsymbol{a}^1 \boldsymbol{z} + c_1 \; s.t. \; e^{\boldsymbol{a}^2 z + c_2} + e^{\boldsymbol{a}^3 z + c_3} \le 1,$

4) $\inf_z \boldsymbol{a}^1 \boldsymbol{z} + c_1 \; s.t. \; \boldsymbol{a}^2 \boldsymbol{z} + c_2 \le 0, \; \boldsymbol{a}^3 \boldsymbol{z} + c_3 \le 0.$ ■

Clearly, when each block set [*k*] is a singleton {*k*}, the inequality constraints $g_k(\boldsymbol{z}) \le 0$ in (GP_z) become $\boldsymbol{a}^k \boldsymbol{z} + c_k \le 0,$ for $k = 1,..., p$, (GP_z) then becomes a linear program (LP):

$$\begin{cases} \inf_{z\in R^m} \boldsymbol{a^0 z} + c_0 \\ s.t.\ \boldsymbol{a^k z} + c_k \le 0,\ k=1,\dots,p \\ \quad \boldsymbol{a^k z} + c_k = 0, k = p+1,\dots,\tilde{p} \end{cases}$$

Thus (GP_z) may be viewed as being somewhere in between LP and OCP.

Observe that equation (1.3) is equivalent to the following log-linear equation:

$$\ln U_i(\boldsymbol{t}) = \ln C_i + \sum_{j=1}^{m} a_{ij} \ln t_j, \qquad i = 1,\dots,\tilde{n}$$

Such log-linear relationships occur in numerous engineering applications, and often show up on an engineering designerôs log-log plots as a result of his regression analysis among various engineering design variables. At other times, instead of an empirical fit, these relationships may simply follow the dictates of the laws of nature and/or economics [Varian 1978]. This explains why many of the functions describing an engineering system are posynomials and why GP is especially suitable for handling optimization problems involving such functions.

1.3 EXTENSIONS TO *EGP*, *QGP*, AND THEN TO *CGP*

Exponential Posynomial Programs

Duffin et al. [1967, p.100-101] have mentioned that if a primal program involves a function of the form $G(\boldsymbol{t}) + Ce^{V(\boldsymbol{t})}$ (*additive exponential of a posynomial term*), where $G(\boldsymbol{t})$ is a posynomial, $C > 0$, and $V(\boldsymbol{t})$ is a posynomial term, it can be handled by limiting techniques; in the same book, Duffin et al. [1967, p.210, 238] employed an abstract geometric inequality to derive a duality theory for a class of extended GP problems of the following form:

Example 1.3.1 (*additive logarithm of a posynomial term*)

$$\begin{cases} \min_{t>0} H_0(\boldsymbol{t}) + \ln U_0(\boldsymbol{t}) \\ s.t. \quad H_1(\boldsymbol{t}) + \ln U_1(\boldsymbol{t}) \le 1 \end{cases} \quad \text{where } \begin{cases} H_0, H_1 \text{ are posynomials,} \\ U_0, U_1 \text{ are posynomial terms} \end{cases}.$$

Later, Lidor and Wilde [1978] treated a class of optimization problems, called *Transcendental Geometric Program* (TGP), which involves posynomial-like functions whose variables may appear also as exponents or in logarithms

$$(TGP) \ \inf_{\mathbf{0}<\boldsymbol{t}\in R^m} \tilde{G}_0(\boldsymbol{t}) \ s.t. \ \tilde{G}_k(\boldsymbol{t}) \le 1, \ k \in K,$$

where

(1.15) $$\tilde{G}_k(\boldsymbol{t}) = \sum_{i\in[k]} U_i(\boldsymbol{t}) \cdot \exp\left[\sum_l D_l t_{l'}\right], \ l' \in J, \text{ and } D_l \text{ are reals},$$

viz., some exponential factor of a linear signomial may be multiplied to certain posynomial terms. [Lidor, 1975] reported that the dual of the above program models chemical equilibrium problems for *non-ideal* systems. *The dual model they derived has the primal variables also appear in the orthogonality conditions and thus these constraints are no longer linear and its objective value also does not serve as a lower bound for the primal objective value, thus loses most of the attractiveness of a dual program.*

Replacing the linear signomial factors in the above model with posynomials, we get a partial extension of the above model, which we call *Exponential Posynomial Programs* (EPP), whose problem functions are of the form:

(1.16) $$\tilde{G}_k(\boldsymbol{t}) := \sum_{i\in[k]} U_i(\boldsymbol{t}) \cdot \exp\left[\sum_l V_l(\boldsymbol{t})\right],$$

where $V_l(\boldsymbol{t}) = D_l \prod_{j\in J} t_j^{b_{lj}}, D_l > 0$. Following Lidorȏs terminology, we shall also call such a function *posynential*. With the positivity assumption on the coefficients D_l, we can easily transform this program into a convex one:

$$(EPP)_z \quad \inf_{z \in R^m} \tilde{g}_0(\boldsymbol{z}) \ s.t. \ \tilde{g}_k(\boldsymbol{z}) \leq 0, \ \forall k \in K \ ,$$

with

$$(1.17) \qquad \tilde{g}_k(\boldsymbol{z}) =: \ln \tilde{G}_k(\boldsymbol{t}) = \ln\left\{\sum\nolimits_{i \in [k]} \exp\left[\boldsymbol{a}^i \boldsymbol{z} + c_i + \sum\nolimits_l \exp\left(\boldsymbol{b}^l \boldsymbol{z} + d_l\right)\right]\right\}, \ k \in \breve{K},$$

where $d_l =: \ln D_l$ and $\boldsymbol{b}^l$ is the l^{th} row of the matrix $B = [b_{lj}]$. We are able to derive in this thesis a duality theory for EPP that is as powerful as the one for posynomial programs. Recall that the problem functions $g_k(\boldsymbol{z})$ in a posynomial program $(GP)_z$ are composites of the geometric function $geo(\boldsymbol{x}) = \ln(\sum e^{x_i})$ with affine functions $\boldsymbol{a}^i \boldsymbol{z} + c_i$ that are the log of the posynomial terms $U_i(\boldsymbol{t})$. Now, if we add to those affine functions a sum of exponential functions $\sum_l \exp\left(\boldsymbol{b}^l \boldsymbol{z} + d_l\right)$ of the log of some other posynomial terms $\ln V_l(\boldsymbol{t})$, we get the problem functions for $(EPP)_z$. This new class of programs opens a door to new area of applications such as maximum likelihood point estimation for normal probability distribution and for some other distributions.

Note that, upon exponentiating each of its problem functions, the above Example 1.3.1 is equivalent to an (EPP) problem.

$$\min_{\boldsymbol{t}>0} U_0(\boldsymbol{t}) \cdot e^{H_0(\boldsymbol{t})} \ s.t. \ \mathrm{e}^{-1} U_1(\boldsymbol{t}) \cdot e^{H_1(\boldsymbol{t})} \leq 1$$

Quadratic Posynomial Programs

The extension from *GP* to *QGP* was also prompted by a need arising from real world applications. As noted above, the building block of *GP* is the posynomial term $U_i(t)$, which is equivalent to a log-linear relationship. In the study of Machining Economics Problem (*MEP*), there is a well-known empirical tool life equation in posynomial form:

(1.18) $$T = C\prod_{j=1}^{m} t_j^{a_j}$$

originally formulated by F.W. Taylor [1907], based on the assumption that the logarithm of the machine tool life T is linear in the logarithms of the machining variables t_j such as cutting speed, depth of cut, feed, etc., viz.,

(1.19) $$\ln T = \ln C + \sum_{j=1}^{m} a_j \ln t_j.$$

Using this (extended) Taylor formula, the MEP can be formulated and solved as a GP. Hough and Goforth [1981b] hold that GP is one of the most straightforward techniques for solving the constrained MEP, and it gives the most insight into the problem; however, this kind of formulation of MEP had been met with limited acceptance and use by industry, mainly due to the inaccuracy of Taylorǿs tool-life equation (1.18) for some combinations of machine tools and materials. This inaccuracy is partially attributable to the *inadequacy* of the <u>linear logarithmic assumption</u> (1.19) for some types of MEP, (see, for example, Colding [1959, 1969]; [Colding and Konig, 1971]; [Wu, 1964]). To correct this, they suggested adopting instead a <u>quadratic logarithmic assumption</u>,

(1.20) $$\ln T = \ln C + \sum_{j=1}^{m} a_j \ln t_j + \tfrac{1}{2}\sum_{j=1}^{m}\sum_{l=1}^{m} q_{jl} \ln t_j \ln t_l$$

which, after exponentiation, leads to a more accurate Colding and Wu type tool-life equation:

(1.21) $$T = C\prod_{j=1}^{m} t_j^{a_j} \cdot \prod_{j=1}^{m} t_j^{\frac{1}{2}\sum_{l=1}^{m} q_{jl}\ln t_l} = C\prod_{j=1}^{m} t_j^{a_j + \frac{1}{2}\sum_{l=1}^{m} q_{jl}\ln t_l}$$

They reported that this change increased the usage of MEP by industry. In his Ph.D. dissertation, Hough [1978] named any finite sum of terms such as (1.21) a *Quadratic Posylognomials* (*QPL*) and studied the following (*QPL*) optimization problem:

$$(QPL)\ \ \inf_{\mathbf{0}<\boldsymbol{t}\in R^m} \tilde{G}_0(\boldsymbol{t})\ s.t.\ \tilde{G}_k(\boldsymbol{t}) \le 1,\ k \in K,\ \text{with}\ \boldsymbol{t}=[t_j]_{j=1}^m,$$

with

$$\text{(1.22)} \qquad \tilde{G}_k(\boldsymbol{t}) = \sum_{i\in[k]} \tilde{U}_i(\boldsymbol{t}) =: \sum_{i\in[k]} C_i \prod_{j=1}^m t_j^{a_{ij}+\frac{1}{2}\sum_{l=1}^m q_{jl}^i \ln t_l}, \forall k \in K.$$

The problem functions of a (*QPL*) are posynomial-like, except that the exponents of certain design variables in some QPL terms $\tilde{U}_i(\boldsymbol{t})$ has an added linear function of the $\ln t_l\text{'}s$ to the usual constants a_{ij}. Obviously, then the log of a *QPL* term is a quadratic function of the logarithms of the design variables:

$$\ln \tilde{U}_i(\boldsymbol{t}) = c_i + \sum_{j=1}^m (a_{ij} + \tfrac{1}{2}\sum_{l=1}^m q_{jl}^i \ln t_l) \ln t_j = c_i + \boldsymbol{a}^i \boldsymbol{z} + \tfrac{1}{2} < \boldsymbol{z}, Q^i \boldsymbol{z} >,$$

where $Q^i = [q_{jl}^i]_{m\times m}$, and $z_j = \ln t_j$. Therefore a (*QPL*) problem can be equivalently formulated in the $\boldsymbol{z}$ variables as:

$$(QPL)_z\ \ \inf_{z\in R^m} \tilde{g}_0(\boldsymbol{z})\ s.t.\ \tilde{g}_k(\boldsymbol{z}) \le 0,\ \forall k \in K,$$

with

$$\text{(1.23)} \qquad \tilde{g}_k(\boldsymbol{z}) =: \ln \tilde{G}_k(\boldsymbol{t}) = \ln\left\{\sum_{i\in[k]} \exp[\tfrac{1}{2} < \boldsymbol{z}, \boldsymbol{Q}^{\boldsymbol{i}} \boldsymbol{z} > + \boldsymbol{a}^i \boldsymbol{z} + c_i]\right\},\ k \in K,$$

which are composites of the geometric function with quadratic functions in the $\boldsymbol{z}$ variables. Of course, if each of the above matrices $\boldsymbol{Q}^{\boldsymbol{i}}$ is symmetric and positive semi-definite (p.s.d.), $(QPL)_z$ is a differentiable convex program, so this formulation $(QPL)_z$ is better than its predecessor (*QPL*), since the hidden convexity is brought out in the p.s.d. case.

In practice, many nonlinear empirical formulas used in engineering optimization are developed assuming a linear logarithmic relationship. Sometimes, a quadratic logarithmic relationship may be more appropriate. Therefore, QPL can have applications to areas other than *MEP*. When each set $[k]$ in (1.23) is a singleton $\{k\}$, $\tilde{g}_k(\boldsymbol{z}) = \frac{1}{2} < \boldsymbol{z}, \boldsymbol{Q}^k \boldsymbol{z} > + \boldsymbol{a}^k \boldsymbol{z} + c_k$, $\forall k \in \mathcal{K}$, the corresponding model becomes a Quadratically Constrained Quadratic Program (*QCQP*).

Hough found that this (*QPL*) problem cannot be transformed into a GP problem, and hence a second order logarithmic MEP, although more accurate, can no longer take advantage of the powerful techniques and niceties of a GP approach. Attempting to remedy this, Hough and Goforth [1981a,b,c] extended the theory of GP somewhat, <u>but almost all of the niceties of a GP approach were lost in their theory, as the authors themselves pointed out [Hough and Goforth, 1981b]</u>:

"The QPL theory is not as powerful or as clean as the posynomial case due to the nonlinearities and the fact that the primal variables appear in the dual (program)."

Jefferson and Scott [1985] proceeded to factor each $\boldsymbol{Q}^i$ as $\boldsymbol{Q}^i = (\boldsymbol{B}^i)^t \boldsymbol{B}^i$, where $\boldsymbol{B}^i$ is a $r_i \times m$ matrix with full row rank, so $\frac{1}{2} \boldsymbol{z}^t \boldsymbol{Q}^i \boldsymbol{z} = \frac{1}{2} < \boldsymbol{z}, (\boldsymbol{B}^i)^t \boldsymbol{B}^i \boldsymbol{z} > = \frac{1}{2} \| \boldsymbol{B}^i \boldsymbol{z} \|^2$, where $\| \cdot \|$ is the Euclidean norm in R^{r_i}. Thus the problem functions in $(QPL)_z$ become:

(1.24) $$\tilde{g}_k(\boldsymbol{z}) = \ln\left\{ \sum_{i \in [k]} \exp\left[\tfrac{1}{2} \| \boldsymbol{B}^i \boldsymbol{z} \|^2 + \boldsymbol{a}^i \boldsymbol{z} + c_i \right] \right\}, \; k \in \mathcal{K},$$

which are better than those in (1.23), because more linearity is brought out. By applying Peterson's GGP principle to the variables-separated version of this model, they gave a corresponding dual model and a set of optimality conditions.

Let us illustrate the above formulation by a very simple example.

Example 1.3.2 [Fang and Rajasekera, 1987]

$$(QPL)\begin{cases}\inf_{\mathbf{0}<\boldsymbol{t}\in R^2} \tilde{G}_0(\boldsymbol{t}) = t_1 t_2 \cdot t_1^{50\ln t_1} + t_1^{-1} t_2^{-1} \\ s.t.\ G_1(\boldsymbol{t}) = 0.01 t_1^{1/2} + t_2 \le 1, \end{cases},$$

This is a (QPL) problem, since the log of its first term has a quadratic term $50z_1^2 = \frac{1}{2} < \boldsymbol{z}, \boldsymbol{Q}^1\boldsymbol{z} >$

with $Q^1 = \begin{bmatrix}100 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}10\\ 0\end{bmatrix}\begin{bmatrix}10 & 0\end{bmatrix} = (B^1)^T B^1$, so $B^1 = \begin{bmatrix}10 & 0\end{bmatrix}$, $\frac{1}{2} < \boldsymbol{z}, \boldsymbol{Q}^1\boldsymbol{z} >= \frac{1}{2} \| \boldsymbol{B}^1\boldsymbol{z} \|^2 = \frac{1}{2}(10z_1)^2$. ■

Composite Posynomial Programs (A unified model for both EPP and QPL)

The problem functions for $(QPL)_z$ (1.23) and for $(EPP)_z$ (1.17) are respectively:

$$\ln\left\{\sum_{i\in[k]} \exp\left[\boldsymbol{a}^i\boldsymbol{z} + c_i + \tfrac{1}{2} < \boldsymbol{z}, \boldsymbol{Q}^i\boldsymbol{z} >\right]\right\} \text{ and } \ln\left\{\sum_{i\in[k]} \exp\left[\boldsymbol{a}^i\boldsymbol{z} + c_i + \sum_l \exp\left(\boldsymbol{b}^l\boldsymbol{z} + d_l\right)\right]\right\}$$

An obvious extension to both EPP and QPL models is to use a general convex function $\boldsymbol{h}_i : R^m \to R$ as the last term in the above expression to obtain more general problem functions

$$\text{(1.25)} \qquad \tilde{g}_k(\boldsymbol{z}) =: \ln\left\{\sum_{i\in[k]} \exp[\boldsymbol{a}^i\boldsymbol{z} + c_i + \boldsymbol{h}_i(\boldsymbol{z})]\right\}$$

for the resultant *Composite Posynomial Program* $(\text{CPP})_z$

$$(CPP)_z \ \inf_{z\in R^m} \tilde{g}_0(\boldsymbol{z})\ s.t.\ \tilde{g}_k(\boldsymbol{z}) \le 0,\ \forall k \in K,$$

1.4 THESIS PLAN

The intended research here is to develop for EPP, QPP, and CPP a unified basic duality theory which is in parallel with those basic duality theorems of GP contained in Chapter 4 of [Duffin et al., 1967].

These have 3 major elements:

a) The main lemma of GP, which is the basis for proving all other duality theorems,

b) The first duality theorem of GP, which asserts the existence of a dual optimal solution under mild conditions on the primal problem, and

c) The second duality theorem of GP, which asserts the existence of a primal optimal solution under mild conditions on the dual problem.

2.0 LITERATURE REVIEW ON DUALITY THEORY

Duality theory is essential for establishing optimality conditions, for performing sensitivity (post-optimality) analysis. It also provides insight into the optimization problem and a meaningful economic interpretation of the model. It infers about the relationships between the primal and the dual optimal objective values. The importance of LP duality was first emphasized by [Dantzig, 1965, Chapter 6], and that of Conic duality for Convex Programs (reformulated in conic form so that the potential reduction interior-point methods would apply) by [Nesterov and Nemirovski, 1994, Section 4.2] and by [Ben-Tal and Nemirovski, 2001, Lecture 2].

The celebrated duality theory of LP developed by [Gale-Kuhn-Tucker, 1951] has served as a model for all subsequent developments in more general settings. The first such extension by Fenchel to convex programs without explicit (nonlinear) constraints can be found in [Rockafellar, 1970, Section 31]. The duality theory of GP is somewhere in between the duality theories of LP and of OCP

2.1 DUAL PROGRAM GD AND DUALITY THEORY OF GP

In this section we shall review the duality theory for the GP problems studied by Duffin and Peterson [1966]. These problems are stated below and they belong to a special case of our earlier GP model (1.2) in Section 1.2 under the assumption that $n = \tilde{n}, I = \tilde{I}$, and $[\boldsymbol{A}\ \boldsymbol{C}] = [\tilde{\boldsymbol{A}}\ \tilde{\boldsymbol{C}}]$,

i.e., no monomial equality constraints are considered.

Primal posynomial program (GP)

(GP) $\inf_{\mathbf{0} < \boldsymbol{t} \in R^m} G_0(\boldsymbol{t})\ s.t.\ G_k(\boldsymbol{t}) \le 1,\ k \in K := \{1, \cdots, p\}$, where $G_k(\boldsymbol{t})$ are as given in (1.4).

As a special case of (1.9) we have

An equivalent convex reformulation of (GP)

(GP)$_z$ $\inf_{z \in R^m} g_0(\boldsymbol{z})\ s.t.\ g_k(\boldsymbol{z}) \le 0,\ \forall k \in K$, where $g_k(\boldsymbol{z})$ are as given in

Error! Reference source not found., and

$$\boldsymbol{a}^i \boldsymbol{z} + c_i = \ln U_i(\boldsymbol{t}) = \ln C_i + \sum_{j \in J} a_{ij} \ln t_j, \boldsymbol{z} = [z_j]_{j=1}^m, z_j = \ln t_j, c_i := \ln C_i, \text{ and } \boldsymbol{a}^i = i^{\text{th}} \text{ row of } A.$$

Likewise, as a special case of (1.14) we have

A GGP formulation of GP

(2.1) (GP)$_{\boldsymbol{x}}$ $\inf_{x \in R^n} geo(\boldsymbol{x}^0 + \boldsymbol{c}^0)\ \ s.t.\ \ geo(\boldsymbol{x}^k + \boldsymbol{c}^k) \le 0,\ \forall k \in K, \boldsymbol{x} \in \mathscr{P}$,

where $\mathscr{P} = \{\boldsymbol{Az} \mid \boldsymbol{z} \in R^m\} \subset R^n$ is the column space of A, $\boldsymbol{x}^k = [x_i]_{i \in [k]}, x_i =: \boldsymbol{a}^i \boldsymbol{z},\ i \in I, \boldsymbol{c}^k = [c_i]_{i \in [k]}$,

and $geo(\boldsymbol{x}^k) := \ln[\sum_{i \in [k]} \exp x_i] : R^{n_k} \to R$ is the geometric function (also called *logexp*(x)) defined in Section 1.2.

So one has

$$geo(\boldsymbol{x}^k + \boldsymbol{c}^k) = \ln\left\{\sum_{i \in [k]} \exp(x_i + c_i)\right\} = g_k(\boldsymbol{z}) = \ln G_k(\boldsymbol{t})$$

The variables in $geo(\boldsymbol{x}^k + \boldsymbol{c}^k)$ are separated for different $k \in K$.

[Duffin and Peterson, 1966] defined the dual program GD for the primal program GP as follows (note that each primal term U_i is assigned a dual variable δ_i):

Dual posynomial program (*GD*) under the convention that $0^0=1$.

$$
\text{(2.2)} \qquad \text{(GD)} \quad \begin{cases} \sup\limits_{\delta \in R^n} V(\delta) := \prod_{k \in \breve{K}} \prod_{i \in [k]} \left(C_i \lambda_k / \delta_i \right)^{\delta_i} & \text{(Dual function)} \\ s.t. \quad \delta_i \ge 0, \forall i \in I, \quad \lambda_0 = 1 & \text{(normality condition)} \\ \qquad \sum_{i=1}^{n} a_{ij} \delta_i = 0, \ \forall j \in J & \text{(orthogonality conditions)} \\ \text{where } \lambda_k =: \sum_{i \in [k]} \delta_i \ , \ \forall k \in \breve{K}. & \end{cases}
$$

Observe that $V(\boldsymbol{\delta}) = \prod_{i=1}^{n} \left(C_i / \delta_i \right)^{\delta_i} \prod_{k=1}^{p} \lambda_k^{\lambda_k} > 0$ for all $\delta \ge \boldsymbol{0}$. We set $\sup(\text{GD}) = 0$, if (GD) is infeasible.

The meaning of the orthogonality conditions in (GD)

In vector form, they are expressed as:

$\boldsymbol{a}_j \cdot \boldsymbol{\delta} = 0, \ \forall j \in J$, where $\boldsymbol{a}_j$ is the j^{th} column of the matrix A,

and in matrix form:

$$A^T \boldsymbol{\delta} = \boldsymbol{0} \ (\Leftrightarrow \boldsymbol{\delta} \perp \mathscr{P} \Leftrightarrow \sum_{i \in I} (\boldsymbol{a}^i)^T \delta_i = \boldsymbol{0}).$$

So these orthogonality conditions may be summarized as

$$\boldsymbol{\delta} \in \mathscr{D} \, (= \mathscr{P}^{\perp}).$$

Thus the orthogonality conditions in (GD) can be obtained simply by taking the orthogonal complement of the primal space $\mathscr{P}$ in $(\text{GP})_{\mathbf{x}}$. Moreover,

$$\text{(2.3)} \qquad \boldsymbol{x} \cdot \boldsymbol{\delta} = 0, \ \forall \boldsymbol{x} \in \mathscr{P}, \forall \boldsymbol{\delta} \in \mathscr{D}$$

The space $\mathscr{D}$ is also called the *dual space* of the program (GD).

The dimension of the dual feasible set (= $R_+^n \cap \mathcal{D} \cap$ normality hyperplane, called *dual flat*) is termed the *degree of difficulty* of the dual program (GD) in the geometric programming literature. This number is dd=n-m-1, when $\boldsymbol{A}$ is of full column rank m.

The concavity of the log-dual function in (GD) [Duffin et al. 1967, p. 121]

The log-dual function

$$v(\delta) =: \ln V(\delta) = \sum_{k=0}^{p} \sum_{i \in [k]} \left[\delta_i \ln \left(C_i \lambda_k / \delta_i \right) \right] = \sum_{i=1}^{n} \delta_i (c_i - \ln \delta_i) + \sum_{k=1}^{p} \lambda_k \ln \lambda_k \tag{2.4}$$

is concave on its domain of definition.

Thus the dual program (GD) is equivalent to an ordinary convex program. Note that

$$\min - v(\delta) \Leftrightarrow \max v(\delta) \Leftrightarrow \max V(\delta) \Leftrightarrow \min 1/V(\delta),$$

and that $1/V(\delta) = e^{-v(\delta)}$ is convex. Unlike the primal problem, the log-dual function $v(\delta)$ is *not* differentiable. In fact, it is not even continuous, since its domain has an empty interior due to the normality condition. Its partial derivatives exist only when $\delta_i > 0$:

$$\frac{\partial v(\delta)}{\partial \delta_i} = \begin{cases} c_i - \ln \delta_i - 1, & \forall i \in [0] \\ c_i - \ln \delta_i + \ln \lambda_k, & \forall i \in [k],\ k \in \mathrm{K} \end{cases} \tag{2.5}$$

In general, the λ_k's in (GD) are treated as dependent variables, and the dual objective $V(\boldsymbol{\delta})$ as a function of $\boldsymbol{\delta}$ alone. Computationally, [Dembo, 1978a, p.232] feels this is a very bad practice, as it obstructs the design of efficient and numerically stable software for (GD). He holds that for computational purposes, both the λ_k's and δ_i's should be regarded as independent variables, and thus the log-dual $v(\delta, \lambda)$ becomes a separable function.

Observe that the term coefficients C_i and the exponents a_{ij} in (GD) are separated in that the former appear only in the objective and the latter only in the (linear) constraints; whereas in

(GP) they are scattered all over the terms $U_i(\boldsymbol{t})$ of the functions $G_k(\boldsymbol{t})$. In summary, through the use of (GD), one can recover 3 exploitable structures: *linearity, separability, and convexity* which are originally hidden in the (GP) model. The founders of GP first pointed out these unique advantageous features.

Basic duality theorems of GP

These have three major components:

0) The main lemma of GP, which is the basis for proving all other duality theorems,
1) The first duality theorem of GP, which asserts the existence of a dual optimal solution under mild assumption on the primal problem, and
2) The second duality theorem of GP, which asserts the existence of a primal optimal solution under mild assumption on the dual problem.

Duffin and Peterson [1966] first developed a duality theory for posynomial GP based on a so-called *geometric inequality*, which is the only machinery needed to derive the dual program GD and to establish the main lemma of GP. This key lemma provided weak and incipient duality relationships between the primal and the dual GP programs.

The Geometric Inequality

It is a slight generalization of the classical arithmetic mean–geometric mean (*AGM*) inequality. This AGM inequality is stated and proved in any book on inequalities, e.g. [Beckenbach and Bellman, 1965], and [Hardy, Littlewood, and Polya, 1952, pp.17-18]. Its strict version is equivalent to the strict convexity of the exponential function *exp x*, and also equivalent to the strict concavity of the logarithmic function *ln x*. For easy reference, we list below some forms of the AGM inequality and the geometric inequality, and provide proofs in Appendix A.2.

Strict AGM Inequality $\forall T_i > 0, \forall \delta_i > 0,\ i$=1,...,$n$, with $\sum_{i=1}^{n} \delta_i = \lambda$, where $2 \le n \in N$,

(2.6) $$\left(\sum_{i=1}^{n} \delta_i T_i\right)^{\lambda} \ge \prod_{i=1}^{n} T_i^{\delta_i} \cdot \lambda^{\lambda},$$

where equality holds exactly when all the T_i's are equal. When $\lambda = 1$, this reduces to the more familiar equivalent form [Duffin and Peterson 1966, p.1316, Lemma 0]:

$$\sum_{i=1}^{n} \delta_i T_i \ge \prod_{i=1}^{n} T_i^{\delta_i},$$

where equality holds exactly when all the T_i's are equal.

For the remainder of this thesis, we shall adopt the convention:

$0^0 = 1$, or equivalently, $0 \ln 0 = 0$, since $\lim_{a \downarrow 0} a^a = 1$.

Thus we have $(u/0)^0 = u^0 = 1$, for $u \ge 0$.

Relaxed AGM Inequality $\forall T_i \ge 0, \forall \delta_i \ge 0,\ i$=1,...,$n$, with $\sum_{i=1}^{n} \delta_i = \lambda$, where $2 \le n \in N$

(2.7) $$\left(\sum_{i=1}^{n} \delta_i T_i\right)^{\lambda} \ge \prod_{i=1}^{n} T_i^{\delta_i} \cdot \lambda^{\lambda},$$

where equality holds exactly when all the T_i's for which $\delta_i > 0$ are equal.

We can slightly generalize the above weighted AGM Inequality into the following

Geometric Inequality *Let* $G = \sum_{i=1}^{n} U_i$, $\lambda = \sum_{i=1}^{n} \delta_i$, *with* $U_i \ge 0$, $\delta_i \ge 0, \forall i$, *then one has*

(2.8) $$\left(\sum_{i=1}^{n} U_i\right)^{\lambda} \ge \prod_{i=1}^{n} (U_i \lambda / \delta_i)^{\delta_i} = \prod_{i=1}^{n} (U_i / \delta_i)^{\delta_i} \lambda^{\lambda},$$

where equality holds iff $U_i \lambda = \delta_i G, \forall i$, (*i.e.*, U_i's *and* δ_i's *are in proportion*).

Incidentally, the inequality

(2.9) $$\sum U \ge \prod (U/\delta)^{\delta}.$$

that appears on the front cover of the first textbook on GP [Duffin et. al. 1967] is the special case (omitting indexes) of the above formula(2.8) for $\lambda = 1$.

The original version of this inequality [Duffin and Peterson 1966, p.1316, Lemma 1] requires that $U_i > 0, \forall i$. Under this assumption, by setting $U_i = e^{x_i}, \forall i$, taking logarithm of its two sides and rearranging terms, one obtains an equivalent form of (2.8) [Duffin et. al. 1967]:

(2.10) $$\lambda geo(\boldsymbol{x}) + \sum \delta_i \ln(\delta_i / \lambda) = \lambda geo(\boldsymbol{x}) + \sum \delta_i \ln \delta_i - \lambda \ln \lambda \geq \boldsymbol{x} \cdot \delta,$$

where $geo(\boldsymbol{x}) := \ln(\sum e^{x_i})$, as was already defined in section 1.2, and by our convention, $\sum \delta_i \ln(\delta_i / \lambda) = i(\boldsymbol{\delta} \mid \boldsymbol{0}),$ when λ=0. This inequality becomes an equality iff $\delta_i = (\lambda / G) e^{x_i}, \forall i$.

Lemma 2.1.0 (Main Lemma of GP**)** *If* $\boldsymbol{t}$ *is feasible for primal program* (GP) *and* δ *is feasible for dual program* (GD)*, then*

$$G_0(\boldsymbol{t}) \geq V(\delta).$$

Moreover, under the same conditions $G_0(\boldsymbol{t}) = V(\delta)$, *if, and only if, the following set of extremality conditions holds:*

(2.11) $$\delta_i = \begin{cases} U_i(\boldsymbol{t}) / G_0(\boldsymbol{t}), & i \in [0] \quad (1) \\ \lambda_k U_i(\boldsymbol{t}), & i \in [k], \forall k \in K \quad (2) \end{cases}$$

in which case $\boldsymbol{t}$ *is optimal for primal program* (GP) *and* δ *is optimal for dual program* (GD).

A primal geometric program (GP) is said to be *superconsistent* if $\exists \boldsymbol{t} > \boldsymbol{0}$ $s.t.$ $G_k(\boldsymbol{t}) < 1, \ \forall k \in K$. The *original Lagrangian* for (GP) is defined to be the function

$$L_o(\boldsymbol{t}, \boldsymbol{\mu}) =: G_0(\boldsymbol{t}) + \sum_{k \in K} \mu_k [G_k(\boldsymbol{t}) - 1], \text{ defined for } \boldsymbol{t} > \boldsymbol{0} \text{ in } R^m \text{ and } \boldsymbol{\mu} \in R_+^p,$$

A saddle point of the original Lagrangian $L_o(\boldsymbol{t}, \boldsymbol{\mu})$ is a point $(\boldsymbol{t}', \boldsymbol{\mu}')$ that satisfies:

$$\max_{\boldsymbol{\mu} \geq 0} L_o(\boldsymbol{t}', \boldsymbol{\mu}) = L_o(\boldsymbol{t}', \boldsymbol{\mu}') = \min_{\boldsymbol{t} > \boldsymbol{0}} L_o(\boldsymbol{t}, \boldsymbol{\mu}')$$

Theorem 2.1.1 (First Duality Theorem of GP**):** *Suppose that the primal program* (GP) *is superconsistent and has a minimum solution* ***t'***, *then there exists a Lagrange multiplier vector* $\boldsymbol{\mu}' \in R_+^p$ *for* $\boldsymbol{t}'$ *such that* $(\boldsymbol{t}',\boldsymbol{\mu}')$ *forms a saddle point for* $L_o(\boldsymbol{t},\boldsymbol{\mu})$, *and the dual program* (GD) *also has a maximum solution* δ' *with* $\boldsymbol{\lambda}' =: \boldsymbol{\mu}'/G_0(\boldsymbol{t}')$ *such that*

$$\min(\text{GP}) = G_0(\boldsymbol{t}') = V(\delta') = \max(\text{GD})$$

Moreover, each pair of primal and dual optimal solution $(\boldsymbol{t}',\boldsymbol{\delta}')$ *satisfies*

$$U_i(\boldsymbol{t}') = \begin{cases} \delta_i' V(\delta'), & i \in [0] \\ \delta_i'/\lambda_k', & i \in [k],\ k \in K, \text{ and } \lambda_k' > 0 \end{cases} \tag{2.12}$$

A dual program (GD) is said to be *canonical* if there exists a positive vector $\delta > 0$ in the dual space $\mathcal{D}$. One can also assume without loss of generality that this dual vector δ is feasible.

Theorem 2.1.2 (Second Duality Theorem of GP**):** *Suppose that primal program* (GP) *is consistent and dual program* (GD) *is canonical. Then primal program* (GP) *has a minimum solution* ***t'***.

Let us use an example to show why sometimes it is easier to solve a primal GP through its dual GD *without the need of calculating derivatives*.

Example 2.1.1 A two-term, one-variable (*n=2, m=1*) GP example

$\inf_{t>0} C_1 t^{-a_1} + C_2 t^{a_2}$, where all parameters: C_1, C_2, a_1, a_2 are positive.

Solution The dual program is

$$\sup_{\delta \in R_+^2} \left(\frac{C_1}{\delta_1}\right)^{\delta_1} \left(\frac{C_2}{\delta_2}\right)^{\delta_2}$$

s.t. $\delta_1 + \delta_2 = 1$; $-a_1\delta_1 + a_2\delta_2 = 0$

The exponent matrix and a unique dual solution are given (respectively) by

$$\begin{bmatrix} -a_1 \\ a_2 \end{bmatrix} \text{ and } \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} a_2/(a_1+a_2) \\ a_1/(a_1+a_2) \end{bmatrix}$$

From the extremality condition(2.11) (1), we see that

$$\frac{\delta_1^*}{\delta_2^*} = \frac{U_1(\boldsymbol{t}^*)}{U_2(\boldsymbol{t}^*)}, \text{ thus } \quad \frac{a_2}{a_1} = \frac{C_1}{C_2} t^{-(a_1+a_2)}, t^{a_1+a_2} = \frac{C_1 a_1}{C_2 a_2}, \text{ so } \quad t^* = \left(\frac{C_1 a_1}{C_2 a_2} \right)^{\frac{1}{a_1+a_2}}.$$

This example has an application to a simple inventory control problem [Nahmias, 1989, p.147-148]. Consider $\inf_{Q>0} G(Q) = K\lambda Q^{-1} + hQ/2 + \lambda c$, where $C_1=K\lambda$, $C_2=h/2$, $a_1=a_2=1$, λc=constant, so the economic order quantity (EOQ) is

$$Q^* = \sqrt{2K\lambda / h}. \qquad \blacksquare$$

According to Peterson [2001a], the birth of GP occurred in Pittsburgh in 1961 at Westinghouse R&D, when Zener [1961] discovered a simple formula for the minimum value of a posynomial cost function whose number of terms exceeds the number of design variables by just one. Zenerôs approach was initially based on Fermatôs principle in calculus, namely, set the first partial derivatives to zero and solve; however, instead of solving directly for these design variables, he associated with each term a new variable representing its relative contribution in the total cost function, and solved the problem completely in this *zero degree of difficulty* caseð we give a detailed description of his approach in Appendix B.

2.2 ASSESSING A DUAL APPROACH TO GP

In this section, we list the advantages and disadvantages of a GP dual approach.

Dual to Primal conversion of optimum solutions

When there is perfect duality, taking logarithms of both sides of equation (2.12) gives

$$\sum_{j=1}^{m} a_{ij} z_j + c_i = \begin{cases} \ln(\delta_i V(\delta)), & \mathrm{i} \in [0] \\ \ln\left(\delta_i / \lambda_k\right), & \mathrm{i} \in [k], \ \ k \in K \text{ for which } \lambda_k > 0 \end{cases} \tag{2.13}$$

where c_i :=lnC_i, i=1,é ,n and z_j:=lnt_j, j=1,é ,m.

The niceties of a GP dual approach

1. The primal problem generally has highly nonlinear constraints even in its convex form (GP$_z$), whereas the dual problem has only linear constraints. This is a clear advantage, since linear constraints are much easier to handle than are nonlinear constraints in developing numerical methods for solving optimization problems.
2. The above advantage becomes more apparent when one is to solve large-scale GP problems, because the often present sparsity structure of the exponent matrix ***A*** can be handled more directly in (GD) than in (GP).
3. The dual problem, just like the primal one, is also a convex program.
4. The dual to primal conversion only involves solving a system of linear equations(2.13), which is computationally easy. This is possible as long as there are enough indices k for which $\lambda_k > 0$ in(2.13).

The drawbacks of a GP dual approach

1. The log-dual objective function $v(\delta) = \sum_{i=1}^{n} \delta_i (c_i - \ln \delta_i) + \sum_{k=1}^{p} \lambda_k \ln\lambda_k$ is only differentiable at points where $\delta > 0$, since $\partial v(\delta) / \partial \delta_i \to +\infty$, as $\delta_i \to 0^+$. So if the optimum dual solution has some zero components, any gradient-based numerical code may fail to find this optimum solution.
2. There may not be enough indices k such that $\lambda_k > 0$ in order to solve for z in(2.13) (this happens when there are too many primal forced constraints in GP being slack at optimum), and solving the subsidiary problems may be cumbersome.

Some remedies for a GP dual approach

1. Fang et.al. [1988] implemented a well-controlled dual perturbation method for (GP) which guarantees to overcome the non-differentiability difficulty and the dual-to-primal conversion is via solving a simple LP.

2. Rajgopal and Bricker [2002] also proposed a generalized LP algorithm for (GP) which avoids all of the computational difficulties mentioned above.

Sensitivity (or Marginal) Analysis of the Optimum Objective Value in GP

The optimum solutions to (GP) and to its dual (GD) each provide sensitivity information of the optimum value to the parameter changes. To see this, we perturb each problem function $G_k(t)$ by dividing it by a positive amount B_k (with B_0=1), the corresponding log-dual objective becomes:

(2.14) $$v(\delta) = \sum_{k=0}^{p} \sum_{i\in[k]} \left[\delta_i \ln\left(C_i \lambda_k / \delta_i B_k \right) \right] = \sum_{i=1}^{n} \delta_i (c_i - \ln \delta_i) + \sum_{k=1}^{p} \lambda_k (\ln\lambda_k - b_k),$$

(where $b_k =: \ln B_k$) while subject to the same dual constraints as before [Dembo, 1982, p.3]. We define the log-dual Lagrangian function [Dembo, 1978b, p.334]:

$$l(\delta, z_0, \boldsymbol{z}) =: v(\delta) + (\lambda_0 - 1)z_0 + \sum\nolimits_{j=1}^{m} (\sum\nolimits_{i=1}^{n} a_{ij}\delta_i)z_j$$

which is simply equal to $v(\delta)$ at any dual feasible point δ. Now suppose that z^* and (δ^*, λ^*) are respectively optimum solutions to the perturbed pair of GP problems and there is no duality gap.

Then in the neighborhood of optimum solution satisfying some conditions specified in [Dembo, 1982, pp.7-8, pp.15-16, Table 3], we have

(2.15)
$$\begin{aligned} & g_0(\boldsymbol{z}^*) = v(\delta^*) = l(\delta^*, z_0^*, \boldsymbol{z}^*) \\ & = \sum_{i=1}^{n} \delta_i^*(c_i - \ln \delta_i^*) + \sum_{k=1}^{p} \lambda_k^*(\ln \lambda_k^* - b_k) + (\lambda_0^* - 1)z_0^* + \sum\nolimits_{j=1}^{m} (\sum\nolimits_{i=1}^{n} a_{ij}\delta_i^*)z_j^* . \end{aligned}$$

(2.16)
$$\frac{\partial g_0(\boldsymbol{z}^*)}{\partial c_i} = \delta_i^*, \quad \frac{\partial g_0(\boldsymbol{z}^*)}{\partial b_k} = -\lambda_k^*, \quad \frac{\partial g_0(\boldsymbol{z}^*)}{\partial a_{ij}} = \delta_i^* z_j^*$$

This information provides a quick estimate of the changes in optimum value of (*GP*), should there be a slight change in the parameter values:

- If the right hand side of the constraint $G_k(\boldsymbol{t}) \le B_k$ increases by 1%, the minimum value of (*GP*) will decrease by about λ_k^*%; if the term coefficient C_i increases by 1%, the minimum value of (*GP*) will increase by about δ_i^*%; and if the exponent a_{ij} increases by 1 (and $t_j^* > 1$), the minimum value of (*GP*) will increase by about $\delta_i^* z_j^*$%. In practice, the exponents are usually fixed by the laws of nature and/or economics.
- If, for a fixed k, all term coefficients C_i, $i \in [k]$, increase by 1%, the value of $G_k(\boldsymbol{t}^*)$ also obviously increases by 1%. By the first formula in(2.16), the minimum value of (*GP*)

should increase by about $\sum_{i\in[k]}\delta_i^* = \lambda_k^*$%. This is trivially confirmed, if k=0, since $\lambda_0^* = 1$.

For k>0, we argue as follows:

$$(1.01)G_k(\boldsymbol{t}^*) \le B_k \Leftrightarrow G_k(\boldsymbol{t}^*) \le B_k / (1.01) \approx (0.99)B_k$$

(by a first order Taylor approximation: $(1+x)^{-1} \approx 1-x,\text{for } x \approx 0$). Then by the second formula in(2.16), decreasing B_k by 1% will cause the minimum value of (GP) increase by about λ_k^*%. This shows coherence between the first two formulae in(2.16). A similar result also holds between the first and the third formulae: Increasing the exponent a_{ij} by 1 amounts to multiplying the i^{th} term coefficient C_i by t_j^*, or equivalently, to adding c_i by z_j^* since $\ln(C_i t_j^*) = c_i + z_j^*$, which, according to the first formula in(2.16), will in turn gives about $\delta_i^* z_j^*$ % increase in the minimum value of (GP).

2.3 FENCHEL'S DUALITY THEORY AND GENERALIZED GEOMETRIC INEQUALITY

In the previous two sections we have briefly reviewed the basic duality theory for the prototype GP. Our aim in this thesis is to establish similar basic duality theories for the extended GP cases. We have seen that the dual objective of GD and the main lemma can be derived solely by applying the geometric inequality(2.8). However, in section 1.3, we have also noted that this same approach was not successful for the EGP and QGP cases. In order to develop useful dual programs and the main lemma for these extended GP cases, we need to, first of all, assign additional dual variables to each second tier posynomial term in the primal program, and then utilize *generalized geometric inequalities* instead of just geometric inequality. In fact, even for the prototype GP, it is still beneficial to look at its duality theory from the perspective of Fenchelôs conjugate transform. *Specifically, the dual objective of (GD) is the exponential of the negative of the conjugate transform of the Lagrangian of the primal program (GP)$_x$, and the main lemma can be derived from the conjugate inequality for this Lagrangian and its conjugate transform.* We shall see this after we prove Lemma 3.0. In this section we will first introduce Fenchelôs conjugate transform that transforms an *extended-real-valued function* $f: R^n \to \overline{R}$, where $\overline{R} := R \cup \{\pm\infty\} = [-\infty, +\infty]$, into another function of the same type. We often find it more convenient to work with functions of this sort when dealing with optimization and conjugate duality. Below we use two examples to explain why.

Following (1.1), we consider an inequality-constrained program (P):

(2.17) (P) $\inf_{x \in R^n} \{f_0(x) \mid f_k(x) \le 0,\ k \in K,\ x \in C\} =: \inf(P),$

where $K = \{1, \cdots, p\},\ \phi \ne C \subset R^n,\ f_k : C \to R,\ \forall k \in \breve{K} := \{0\} \cup K,$ and the feasible set S is defined via $S := \left\{ \boldsymbol{x} \in C \middle| f_k(\boldsymbol{x}) \le 0,\ \forall k \in K \right\}$.

Note that if we define i_S to be the indicator of S, then this program (P) can be identified with its *objective function* $f := f_0 + i_S$ in the sense that inf(P) = $\inf_S f_0$ = inf f, and argmin(P) = $\operatorname{argmin}_S f_0$ = argmn f. This function f is from R^n to $(-\infty, +\infty]$. Another example is the *optimal value* (or *perturbation*) *function* of (P):

(2.18) $v(\boldsymbol{b}) := \inf_{x \in R^n} \{f_0(x) \mid f_k(x) \le b_k,\ k \in K,\ x \in C\},$ for $\boldsymbol{b} := [b_k]_p^{k=1}$.

This function v is from R^p to $[-\infty, +\infty]$.

Extended-real line $\overline{R}$ and extended-real-valued functions on R^n: some terminology

Just like the real line R, the extended real line $\bar{R}$, is also a *linearly* (or *totally*) *ordered set*: any two elements x and y in $\bar{R}$ are *comparable*, i.e. either $x \le y$ or $y \le x$. It is also equipped with 2 additional conventions:

1. $\infty - \infty = \infty = -\infty + \infty$ (Inf-addition rule). **2.** $0 \cdot (\pm\infty) = 0 = (\pm\infty) \cdot 0$.

The first rule is not symmetric, because we orient toward minimization. The implications of these rules are listed in Appendix C. Unlike in R, *every subset* C in $\bar{R}$ has in $\bar{R}$ a *supremum* $\sup C$ and an *infimum* $\inf C$. (Caution: $\inf \phi = \infty$ and $\sup \phi = -\infty$, so that $\inf \phi > \sup \phi$!).

In general, we use the capital letter F to denote the *effective domain* $\text{dom} f$ of a function $f : R^n \to \bar{R}$, that is, $F = \text{dom} f := \{x \in R^n \mid f(x) < \infty\}$, and we shall call f *proper* if it is never $-\infty$ and $F \ne \phi$, and *improper* if otherwise. So a proper function f is finite-valued on $F \ne \phi$ but $=\infty$ elsewhere, and an improper one is somewhere $-\infty$ or everywhere $+\infty$. For instance, the *support function* of C defined by $S_C(y) = S(y \mid C) =: \sup\{y \cdot x \mid x \in C\}$ for any set C in R^n, and the indicator function of C, i_C are both proper, when C is nonempty. When C is empty, however, they are both improper, $i_\phi \equiv \infty$, $S_\phi \equiv -\infty$. The objective function f of the above program (P) is improper exactly when program (P) is inconsistent, i.e. $f \equiv \infty \Leftrightarrow S = \phi$, and its perturbation function v is improper: $v(\boldsymbol{b}) = -\infty$ when (P) has an unbounded infimum for some perturbation vector $\boldsymbol{b}$. Proper functions are our central concern, but improper ones such as the above examples do arise indirectly and hence cannot be ruled out from our consideration. The sum of two proper functions f and g is proper exactly when $F \cap G \ne \phi$, or else $f + g \equiv \infty$. For a function $f : R^n \to \bar{R}$, we also define the *epigraph* of *f*, $\operatorname{epi} f := \left\{(x,\alpha) \in R^n \times R \mid f(x) \le \alpha\right\}$,

the *strict epigraph* of *f*, $\operatorname{epi}^s f := \left\{(x,\alpha) \in R^n \times R \mid f(x) < \alpha\right\}$,

the *hypograph* of *f*, $\operatorname{hpo} f := \left\{(x,\alpha) \in R^n \times R \mid f(x) \ge \alpha\right\}$,

the *strict hypograph* of *f*, $\operatorname{hpo}^s f := \left\{(x,\alpha) \in R^n \times R \mid f(x) > \alpha\right\}$,

the (*lower*) *level set* of *f*, $f^{\le\alpha} := \left\{x \in R^n \mid f(x) \le \alpha\right\}$, which corresponds to a ôhorizontal cross sectionô of epi *f*, and other similar level sets such as $f^{<\alpha}, f^{=\alpha}$, etc. For example, the level set of the objective function *f* of the above program (P) (2.17) is

(2.19) $$f^{\le\alpha} = C \cap f_0^{\le\alpha} \cap f_1^{\le 0} \cap \cdots \cap f_p^{\le 0}$$

The Minkowski sum of any two sets *C*, *D* in R^n is the set $C + D := \{x + y \mid x \in C, y \in D\}$, and the multiple of *C* by a scalar *t* in *R* is the set $tC := \{tx \mid x \in C\}$, $-C = (-1)C$. The Cartesian product of two sets $C \subset R^n, D \subset R^m$ is denoted (*C*, *D*). We will simply write (C, y) for $(C,\{y\})$ and likewise, if *n*=*m*, $C + y$ for $C + \{y\}$. Clearly, $t(C + D) = tC + tD$, but for any scalars *r* and *s*, one generally has $(r+s)C \subset rC + sC$ only, the reverse inclusion may fail as is evidenced by the set *C*={-1,1}and scalars *r*=*s*=1. Note that $0 \in C - D \Leftrightarrow C \cap D \ne \phi$, and $0 \in C - x \Leftrightarrow 0 \in x - C \Leftrightarrow x \in C$.

A set *C* is *convex* if $(1-t)C + tC \subset C, \forall t \in (0,1)$ (the reverse inclusion trivially holds). The sum, scalar multiple, and arbitrary intersection of convex sets are again convex sets. For a convex set *C* and non-negative scalars $r \ge 0, s \ge 0$, $(r+s)C = rC + sC$, thus *C*+ *C* =2*C*, *C*+ 2*C* =3*C*, etc.; if, in addition $0 \in C$, then $r \ge s \ge 0 \Rightarrow rC \supset sC$. A set *C* is a *cone* if $0 \in C$, and $tC \subset C$, $\forall t > 0$. A set *K* is a *convex cone* if it is both a cone and a convex set. For a nonempty set K, this condition is equivalent to $R_+K + R_+K \subset K$, and also to $R_+K \subset K,\ K + K \subset K$, where R_+ denotes the set of all non-negative reals.

A function $f : R^n \to \overline{R}$ is *convex* if for all points x^0, x^1 in F and scalar t in (0, 1), f and $x^t := (1-t)x^0 + tx^1$ satisfies *Jensen's inequality*:

(2.20) $$f(x^t) \le (1-t)f(x^0) + tf(x^1).$$

f is concave if its negative ï*f* is convex. The (lower) level sets $f^{\le\alpha}$ and $f^{<\alpha}$, epigraphs epi *f* and strict-epigraphs $\text{epi}^s f$ of a convex function *f* are convex. A function $h : R^n \to \overline{R}$ is *positively homogeneous* (*ph*) if *h* satisfies

$$0 \in H,\ h(tx) = th(x),\ \forall x \in H, \forall t > 0. \qquad \text{(Thus } h(0) = 0 \text{ or } -\infty\text{)}$$

It is *sublinear* if in addition

$$h(x^0 + x^1) \le h(x^0) + h(x^1),\ \forall x^0, x^1 \in H \qquad \text{(Subadditivity)}$$

Equivalently, a sublinear function is one which is both *ph* and convex. As an example, the support function of any subset *C* is sublinear.

It is easy to show that a function $f : R^n \to \overline{R}$ is convex (respectively: *ph*, sublinear) iff its epigraph epi *f* is a convex set (respectively: a cone, a convex cone) in $R^n \times R$.

We denote *ri* (*S*) the relative interior of any set *S* in R^n.

Proposition 2.3.0 [Rockafellar, 1970, Theorem 7.2] *If a convex function* $f : R^n \to \overline{R}$ *is somewhere* $-\infty$, *it is everywhere* $-\infty$ *on the relative interior of its domain ri (F); hence a convex function that is somewhere finite on ri (F) must be proper.*

Proposition 2.3.1 (convexity in composition) [Rockafellar and Wets, 2004, Exercise 2.20 (c)] *Suppose* (*F*,*G*): $R^n \to R^m \times R^l$ *is a mapping* (*where mÓ0, lÓ0*) *such that each of the m components of*

F is a convex function and each of the l components of G is an affine function, and $h: R^m \times R^l \to \bar{R}$ *is a convex function that is nondecreasing in its first m components. Then the composite function* $f(x) := h(F(x), G(x)) : R^n \to \bar{R}$ *is convex. (In particular, if* $l = 0, f(x) = h(F(x))$*, and if* $m = 0, f(x) = h(G(x))$.

Lower semicontinuity of a function

A function $f : R^n \to \bar{R}$ is *continuous at x* if $|y - x| \to 0 \Rightarrow f(y) \to f(x)$. The *lower* and *upper limits* of *f* at *x* are respectively defined by

(2.21) $$\underline{f}(x) = \liminf_{y \to x} f(y) := \sup_{t>0} \inf_{|y-x| \le t} f(y), \quad \overline{f}(x) = \limsup_{y \to x} f(y) := \inf_{t>0} \sup_{|y-x| \le t} f(y).$$

Clearly, one has $\underline{f}(x) \le f(x) \le \overline{f}(x)$, at any x, $\overline{f} = -\underline{g}$, and $\underline{f} = -\overline{g}$, for $g = -f$. We say f is *lower semicontinous* (*lsc*) at *x* if $\underline{f}(x) = f(x)$, *upper semicontinous* (*usc*) at *x* if $f(x) = \overline{f}(x)$. It can be shown that *f* is continuous at *x* iff *f* is both *lsc* and *usc* at *x* iff both *f* and ï *f* are *lsc* (or *usc*) at *x*. If we let $L(f;x) := \left\{ \alpha \in \bar{R} \mid \exists x^i \to x \text{ with } f(x^i) \to \alpha \right\}$, then $\underline{f}(x)$ and $\overline{f}(x)$ are respectively the smallest and largest element in the set $L(f;x)$ and

(2.22) $$\left\{ \underline{f}(x), f(x), \overline{f}(x) \right\} \subset L(f;x) \subset [\underline{f}(x), \overline{f}(x)] =: \left\{ \alpha \in \bar{R} \mid \underline{f}(x) \le \alpha \le \overline{f}(x) \right\}.$$

(Semi-) continuity on R^n means the property holds everywhere in R^n. Lower semicontinuity on R^n can be characterized as follows:

Proposition 2.3.2 *The following properties of a function* $f : R^n \to \bar{R}$ *are equivalent*:

(a) *f is lsc on* R^n;

(b) epi *f is closed in* $R^n \times R$;

(c) $f^{\le \alpha}$ *is closed is* R^n, $\forall \alpha \in R$.

For instance, the support function of any subset C is *lsc*; and in the program (P)(2.17), if the set C is closed, and all the problem functions $f_k, k \in R$Ĕ are *lsc*, then (2.19) shows the objective function f for program (P) is also *lsc*.

The closure, interior, and complement of a set S of R^n are respectively denoted by cl S, int S, and S^C. For any function $f : R^n \to \overline{R}$, one has cl (epi f) = [int (hpo f)] ^{C} = epi ($\underline{f}$) [Rockafellar and Wets, 2004, Exercise 1.13]. This condition implies that the lower limit function $\underline{f}$ is the largest *lsc* function g not exceeding f, also called the *lower semi-continuous hull* or the (*lower*) *closure* of f, and is denoted by cl f; thus

$$\mathrm{cl}\left(\mathrm{epi}\, f\right) = \mathrm{epi}(\mathrm{cl}\, f), \text{ and } \mathrm{cl}\, f(x) = \underline{f}(x) = \liminf_{y \to x} f(y)$$

Proposition 2.3.3 *Let $f : R^n \to \overline{R}$ be any function. If f is improper, so is* cl f; *if f is proper convex, so is* cl f. [Rockafellar, 1970, Theorem 7.4]

The (*closed*) *convex hull*, (cl) con f of a function f is defined to be the largest (*lsc*) convex function g not exceeding f. Clearly:

$$\mathrm{cl}\ \mathrm{con}\, f \le \{\mathrm{cl}\, f, \mathrm{con}\, f\} \le f .$$

For any *lsc* convex (respectively, *lsc*, convex) function g, $g \le f$ is equivalent to $g \le \mathrm{cl}\ \mathrm{con}\, f$ (respectively, cl f, con f).

A key duality result in convex and variational analysis is the following

Theorem 2.3.4 (envelope representation of convex sets and convex functions) [Rockafellar and Wets, 2004, theorem 6.20 and 8.13]

1). *A nonempty closed convex set in R^n is the intersection of its supporting half-spaces*, *hence is the intersection of a family of closed half-spaces.*

2). *An lsc proper convex function* $f : R^n \to \overline{R}$ *is the pointwise supremum of its affine supports, hence is the pointwise supremum of a family of affine functions.*

Remark An immediate consequence of this, in view of Proposition 2.3.3, is that any proper convex function must have at least one affine support, and hence one affine minorant.

Fenchel's Conjugate Inequality for a Function and its Conjugate Transform

The (*Fenchel*) *conjugate* of any function $f : R^n \to \overline{R}$ is another function $f^* : R^n \to \overline{R}$

(2.23) $$f^*(y) := \sup_x [y \cdot x - f(x)] = \sup_{x,\alpha} \{ l_{x,\alpha}(y) \mid f(x) \le \alpha \}, \text{ where } l_{x,\alpha}(y) := x \cdot y - \alpha$$

Hence $$\operatorname{epi} f^* = \bigcap_{(x,\alpha)\in \operatorname{epi} f} \operatorname{epi} l_{x,\alpha}, \text{ and}$$

$$\operatorname{epi} l_{x,\alpha} = \left\{ (y,\beta) \in R^n \times R \mid x \cdot y - \alpha \le \beta \right\} = \left\{ (y,\beta) \mid (x,-1)\cdot(y,\beta) \le \alpha \right\}$$

The latter sets are closed half-spaces. So $\operatorname{epi} f^*$ is a closed convex set and f^* is an *lsc* convex function. It may be improper due to 2 reasons:

(2.24) $$\begin{aligned} &\text{a) } f^* \text{ is somewhere } (\text{hence everywhere}) -\infty, \text{ and this happens exactly when } f \equiv \infty. \\ &\text{b) } f^* \text{ is everywhere } \infty, \text{ and this happens exactly when } f \text{ has no affine minorant, as} \end{aligned}$$

(2.25) $$\beta \ge f^*(y) \Leftrightarrow \beta \ge y \cdot x - f(x), \forall x \Leftrightarrow f(x) \ge y \cdot x - \beta =: l_{y,\beta}(x), \forall x$$

e.g. the proper function $f(x) = -|x|, \ \forall x \in R$ has no affine minorant and $f^* \equiv \infty$.

Proposition 2.3.5 *For any function* $f : R^n \to \overline{R}$, *one has*

(2.26) $$f^* = (\operatorname{cl} f)^* = (\operatorname{con} f)^* = (\operatorname{cl}\operatorname{con} f)^*$$

Proof Since an affine function is *lsc* convex, it minorizes f iff it minorizes any one of these 3 functions: cl f, con f, cl con f. The above equation then follows from (2.25) . ■

The conjugate of the indicator function i_C of any set C in R^n is the support function S_C of C. When C is a cone, then $S_C = i_{C^-}$, and the *barrier cone* of C, $B_C = \mathrm{dom} S_C$ equals its polar cone, C^-. The conjugate of the geometric function, $geo(\boldsymbol{x}) = \ln(\sum e^{x_i})$, is the *entropy function* $ent(\boldsymbol{y}) := \sum y_i \ln y_i$, if $y_i \geq 0$, $\sum y_i = 1$, and := ∞ if otherwise (Where $0 \ln 0 = 0$). Trivially, one has $-f^*(0) = \inf f$. So inf $geo = -ent(0) = -\infty$ and this can be approached by letting $x_i \to -\infty$ in $geo(\boldsymbol{x})$ for each i. The mapping $f \to f^*$ is called the (*Legendre-*) *Fenchel transform*. It is obviously order-reversing $f \leq g \Rightarrow f^* \geq g^*$.

The following result is an analogue of proposition 2.3.3.

Proposition 2.3.6 *Let $f : R^n \to \overline{R}$ be any function. If f is improper, so is f^*; if f is proper convex, so is f^*.* [Rockafellar, 1970, Theorem 12.2]

Proof The first part is easy. For the second part, if f is proper convex, the remark following Theorem 2.3.4 shows that f has at least one affine minorant, hence by (2.24) (b), f^* is not everywhere ∞; on the other hand, since f is not everywhere ∞, by (2.24) (a), f^* is never $-\infty$. Thus f^* is also proper. ■

The *subdifferential* (or *set of subgradients*) of a proper function f at a point x is the set

(2.27) $$\partial f(x) := \left\{ y \in R^n \mid \langle y, z - x \rangle \leq f(z) - f(x),\ \forall z \in F \right\}$$

This is clearly a closed convex set, which is empty when $x \notin F$; when nonempty, it consists of the gradients y of all affine supports $z \to \langle y, z - x \rangle + f(x) = y \cdot z - f^*(y)$ for f at x (see Proposition 2.3.8 below). The *domain* of the *multifunction* $x \to \partial f(x)$ is the set $\mathrm{dom}\, \partial f := \left\{ x \in R^n \mid \partial f(x) \neq \varnothing \right\}$, over which the function f is said to be *sub-differentiable*, which means geometrically that the set epi f has a non-vertical supporting hyperplane at the point $\left(x, f\left(x\right)\right)$. A proper convex function f with $x \in F$ is differentiable at x iff $\partial f(x)$ is a singleton, namely, $\{\nabla f(x)\}$ [Rockafellar, 1970, Theorem 25.1], in which case f has a unique affine support: $z \rightarrow \langle \nabla f(x), z - x \rangle + f(x)$ for f at x, whose graph is the non-vertical tangent hyperplane to the set epi f at the point $\left(x, f\left(x\right)\right)$.

For example, the proper convex function $f\left(x\right) := -\sqrt{x}$, for $x \geq 0$; and $:= +\infty$, for $x < 0$ has subdifferential $\partial f(x) = \varnothing$, for $x \leq 0$; and $= \{-\frac{1}{2\sqrt{x}}\}$, for $x > 0$; The proper non-convex function $g\left(x\right) := |x|$, for $x \neq 0$; and $:= -1$, for $x = 0$ has subdifferential $\partial g(x) = \{\operatorname{sgn} x\}$, for $x \neq 0$; and $=$ $[-1,1]$, for $x = 0$; The indicator function i_C has subdifferential at any point x in C, $\partial i_C(x) = \left\{ y \in R^n \mid \langle y, z - x \rangle \leq 0, \ \forall z \in C \right\} = (C - x)^- =: N_C(x)$, the *normal cone* to C at x.

It is often critical to tell if a subgradient set is non-empty.

Proposition 2.3.7 [ibid, Theorem 23.4] *For any proper convex function* $f : R^n \rightarrow \overline{R}$, *the subgradient set* $\partial f(x)$ *is non-empty if* $x \in ri(F)$, *non-empty and bounded iff* $x \in \operatorname{int} F$. *Hence one has* $\operatorname{ri} F \subset \operatorname{dom} \partial f \subset F$.

Trivially, for any proper function f finite at x, one has $0 \in \partial f(x) \Leftrightarrow x \in \arg\min f$. More generally, one has $y \in \partial f(x) \Leftrightarrow x \in \arg\min_z [f(z) - y \cdot z]$. This is because

Proposition 2.3.8 *For any function* $f : R^n \rightarrow \overline{R}$, *there holds the inequality*

(2.28) $$f(x) + f^*(y) \geq x \cdot y, \ \forall x \in F, y \in F^*$$ (Fenchelôs conjugate inequality)

When f is proper, finite at x, equality holds iff $y \in \partial f(x)$ iff $x \in \arg\max_z [y \cdot z - f(z)]$.

Proof The inequality follows easily from the definition of conjugate f^* in(2.23). When f is proper, finite at x, we see that

$$y \in \partial f(x) \Leftrightarrow y \cdot z - f(z) \le y \cdot x - f(x),\ \forall z \in F$$
$$\Leftrightarrow x \in \arg\max_z [y \cdot z - f(z)] \Leftrightarrow y \cdot x - f(x) = f^*(y)$$ ■

Dually, if f^* is proper, finite at y, one has $f^*(y) + f^{**}(x) \ge y \cdot x,\ \forall y \in F^*, x \in F^{**} =: \mathrm{dom} f^{**}$ with equality holding iff $x \in \partial f^*(y)$. The function $f^{**} = (f^*)^*$ is called the *bi-conjugate* of the function *f*. The conjugate inequality(2.28) also implies that

$$f(x) \ge \sup_y \{x \cdot y - f^*(y)\} = (f^*)^*(x),\ \forall x.$$

Hence $f \ge f^{**}$ holds in general.

Theorem 2.3.9 [Rockafellar and Wets, 2004, Proposition 11.1] *For any function* $f : R^n \to \overline{R}$ *with* con *f proper, both* f^* *and* f^{**} *are lsc proper convex, and* f^{**}= cl con *f*. *When f is proper convex, one has* f^{**}= cl *f*. *When f is itself lsc proper convex, one has* f^{**}=*f*.

Proof The first assertion follows from proposition 2.3.5 and 2.3.6. For the second one, since by proposition 2.3.3, cl con *f* is *lsc* proper convex, Theorem 2.3.4 (2) says that it must be the pointwise supremum of its affine minorants, which means for all *x*

$$\mathrm{cl\ con} f(x) = \sup_{y,\beta}\{l_{y,\beta}(x) \mid l_{y,\beta} \le f\},\ \text{where } l_{y,\beta}(x) = y \cdot x - \beta$$
$$= \sup_{y,\beta}\{y \cdot x - \beta \mid \beta \ge f^*(y)\},\ \text{ by (2.27)}$$
$$= \sup_y [y \cdot x - f^*(y)] = f^{**}(x)$$

The remaining assertions then follow easily from this. ■

The last equation in the theorem: $f^{**} = f$ also holds for improper functions $f \equiv \pm\infty$: since then $f^* \equiv \mp\infty$ and $f^{**} \equiv \pm\infty$, but it can be shown that these are the only ‘improper’ cases. Thus

$$f^{**} = f \Leftrightarrow f \text{ is } lsc \text{ proper convex or } f \equiv \pm\infty$$

A function for which this equation holds is also called *closed convex*. For proper convex functions closedness is equivalent to lower semi-continuity.

Theorem 2.3.10 [Rockafellar and Wets, 2004, Proposition 11.3] *For any lsc proper convex function f, one has*

(2.29)
$$
\begin{aligned}
&a)\ \partial f^* = (\partial f)^{-1} \text{ and } \partial f = (\partial f^*)^{-1} \\
&\text{b) } y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y) \\
&c)\ \partial f^*(y) = \text{argmax}_z \{y \cdot z - f(z)\}, \\
&\text{d) } \partial f(x) = \text{argmax}_w \{x \cdot w - f^*(w)\}
\end{aligned}
$$

For example, the function $geo(\boldsymbol{x})$, being finite convex, is clearly *lsc* proper convex, has its conjugate equal to the *lsc* proper convex function $ent(\delta)$, and the conjugate of which is then $geo(\boldsymbol{x})$. The conjugate inequality relating this pair of functions is:

$$\ln(\sum e^{x_i}) + \sum \delta_i \ln \delta_i \geq \sum x_i \delta_i \ (\Leftrightarrow \ln(\sum e^{x_i}) \geq \sum \delta_i \ln(e^{x_i} / \delta_i)),$$

which is equivalent to the original geometric inequality (2.9), $\sum U_i \geq \prod (U_i/\delta_i)^{\delta_i}$, (where $U_i = e^{x_i}, \forall i$). The condition for equality is $\delta = \nabla f(\boldsymbol{x})$, *i.e.* $\delta_i = e^{x_i} / G, \forall i$, with $G = \sum e^{x_i}$.

The *dual cone* of any set S in R^n is the closed convex cone

$S^+ =: \left\{ y \in R^n \mid y \cdot x \geq 0, \forall x \in S \right\}$. When K itself is a closed convex cone, one has $K^{++} = K$.

Fenchel's dual pair of conic programs

For any proper function f on R^n and any cone K in R^n, *Fenchel's conic program* (P) and its dual program (D) are respectively given by

(2.30) (P) $p = \inf\{f(x) \mid x \in F \cap K\}$ (D) $d = -\inf\{f^*(y) \mid y \in F^* \cap K^+\}$, where $p,\ d \in [-\infty, +\infty]$, F and F^* are the respective effective domains of f and f^* (actually, F and F^* can both be omitted from the above formulation). If in addition, f and K are both closed convex, the dual program of (D) is then (P); hence the formulation becomes completely symmetric. The following result is immediate.

Lemma 2.3.11 (Main Lemma of Fenchel's conic dual programs) *Under the above setting* (2.30), *one has* $f(x) \geq -f^*(y), \quad \forall x \in F \bigcap K,\ \forall y \in F^* \bigcap K^+,$

with equality holding if and only if $y \in \partial f(x)$, *and* $x \cdot y = 0$, *in which case x is optimal for primal problem* (P) *and y is optimal for dual problem* (D). *Thus* $p \geq d$ *is always satisfied.*

Theorem 2.3.12 (Fenchel's Conic Duality Theorem) [Rockafellar, 1970, Theorem 31.4] *Assume that in the above setting*(2.30), *f and K are both convex.*

a) *If* $p > -\infty$, *and* $\operatorname{ri}(F) \bigcap \operatorname{ri}(K) \neq \emptyset$, *then* $p = d$, and *d is attained.*

b) *If, furthermore, f and K are both closed,* $d < +\infty$, *and* $\operatorname{ri}(F^*) \bigcap \operatorname{ri}(K^+) \neq \emptyset$, *then* $p = d$, *and p is attained.*

In general, x and y satisfy

$$f(x) = \inf_K f = \sup_{K^+} - f^* = -f^*(y)$$

if, and only if

$$y \in \partial f(x), \quad x \in K, \quad y \in K^+, \ \ x \cdot y = 0 .$$

The hypotheses $\operatorname{ri}(F) \bigcap \operatorname{ri}(K) \neq \emptyset$ in (a) and $\operatorname{ri}(F^*) \bigcap \operatorname{ri}(K^+) \neq \emptyset$ in (b) are commonly called *Fenchel's hypothesis* for the primal problem (P) and the dual problem (D), respectively. If either of these conditions hold, and f and K are both closed, then $p = d$.

As an easy consequence of this, one has the following corollary:

Corollary 2.3.13 (Fenchel-Rockafellar's Duality Theorem) [Rockafellar, 1970, Corollary 31.2.1] *For any functions* $f_1 : R^n \to \overline{R},\ f_2 : R^m \to \overline{R}$*, and a linear map* $A : R^n \to R^m$*, let* $p, d \in \overline{R}$ *be optimal values defined by the pair of programs below*:

$$p = \inf_{x \in R^n} [f_1(x) + f_2(Ax)]$$

$$d = -\inf_{y \in R^m} [f_1^*(-A^T y) + f_2^*(y)].$$

These values satisfy $p \geq d$ *. If, furthermore,* f_1 *and* f_2 *are proper convex and satisfy the condition*

(2.31) $\exists x \in \text{ri}\ (F_1)\ s.t.\ Ax \in \text{ri}\ (F_2)$, (Fenchelôs hypothesis)

and $p > -\infty$*, then* $p = d$ *, and d is attained.*

This hypothesis is equivalent to the condition $\text{ri}\ (F_1) \bigcap A^{-1}\text{ri}\ (F_2) \neq \emptyset$, which implies that

$$\text{ri}\ (F) \neq \emptyset, \text{ where } F = F_1 \cap A^{-1} F_2 = \text{dom} f,\ f := f_1 + f_2 \circ A.$$

A special case of this result is obtained, when $n = m$ and A is the identity map.

Corollary 2.3.14 (Fenchel's Duality Theorem) [Rockafellar, 1970, Theorem 31.1] *For any functions* $f_1, f_2 : R^n \to \overline{R}$*, let* $p, d \in \overline{R}$ *be optimal values defined by the pair of programs below*:

$$p = \inf_{x \in R^n} [f_1(x) + f_2(x)]$$

$$d = -\inf_{y \in R^n} [f_1^*(-y) + f_2^*(y)].$$

These values satisfy $p \geq d$ *. If, furthermore,* f_1 *and* f_2 *are proper convex and satisfy the condition*

(2.32) $\exists x \in \text{ri}\ (F_1) \cap \text{ri}\ (F_2)$,

and $p > -\infty$, then $p = d$ *and d is attained.*

Generalized geometric inequalities

For any proper convex function $f: R^n \to \overline{R},\ \lambda \in R_+$, and $c \in R^n$, the Fenchelôs conjugate inequality for the function $\lambda f(x+c)$ is termed a *generalized geometric inequality*:

(2.33) $$\lambda f(x+c) + \lambda f^*(y/\lambda) \geq (x+c)\cdot y,\ \forall x, y \in R^n$$

with equality holding iff $y \in \lambda\partial f(x+c)$. When $\lambda = 0$, this inequality degenerates to

$$0 + i(y\,|\,0) \geq (x+c)\cdot y,\ \forall x, y \in R^n$$

with equality holding iff $\boldsymbol{y} = \boldsymbol{0}$ (Under the convention: $0 f^*(\boldsymbol{y}/0) = i(\boldsymbol{y}\,|\,0)$). Obviously, when $\lambda = 1, c = 0$, the generalized geometric inequality (2.33) reduces to the conjugate inequality (2.28) for the function *f*. As an example, if $f(\boldsymbol{x}) = geo(\boldsymbol{x}), \lambda \in R_+, \boldsymbol{c} = \boldsymbol{0}$ in the above generalized geometric inequality(2.33), then one has (with $\boldsymbol{\delta}$ replacing $\boldsymbol{y}$):

$$\lambda geo(\boldsymbol{x}) + \lambda ent(\delta/\lambda) = \lambda geo(\boldsymbol{x}) + \sum \delta_i \ln(\delta_i/\lambda) \geq \boldsymbol{x}\cdot\delta\ ,$$

which is the original geometric inequality (2.10), and the above condition for equality, $y \in \lambda\partial f(x+c)$ becomes $\delta = \lambda\nabla geo(\boldsymbol{x})$, which means $\delta_i = (\lambda/G)e^{x_i}, \forall i,$ with $G = \sum e^{x_i}$.

In particular, for any differentiable convex function $h: R \to R$ and parameters $y \in R_+$, and $d \in R$, the Fenchelôs conjugate inequality for the function $yh(\xi + d)$ yields a generalized geometric inequality:

(2.34) $$yh(\xi+d) + yh^*(\eta/y) \geq (\xi+d)\eta,\ \forall \xi, \eta \in R$$

with equality holding iff $\eta = yh'(\xi+d)$, and the degenerate case for $y = 0$ is

$$0+i(\eta\,|\,0)\geq(\xi+d)\eta,\ \forall\xi,\eta\in R$$

with equality holding iff $\eta=0$ (Under the convention: $0h^*(\eta/0)=i(\eta\,|\,0)$).

2.4 KKT THEOREM FOR CONVEX PROGRAMS

In order to derive the first duality theorems for GP and for its extensions, we need to apply the Karush-Kuhn-Tucker (KKT) theorem to the convex formulations of the various extensions of GP. Let us consider in this section an (inequality-constrained) ordinary *convex* program (CP)

(2.35) $\quad$ (CP) $\inf_{\boldsymbol{x}\in F}\{f(\boldsymbol{x})\,|\,\boldsymbol{g}(\boldsymbol{x})\leq 0\}$

where $\boldsymbol{g}(\boldsymbol{x})=[g_i(\boldsymbol{x})]_{i=1}^m$, the functions $f,g_1,g_2,...,g_m:R^n\to\overline{R}$ are proper convex and satisfy $F=\operatorname{dom} f\subset\operatorname{dom} g_i=G_i,\forall i$, $S=\{\boldsymbol{x}\in F\,|\,\boldsymbol{g}(\boldsymbol{x})\leq 0\}$ is its feasible set. The *perturbation function* $v:R^m\to\overline{R}$ of (CP) is convex, where

(2.36) $\quad$ $v(\boldsymbol{b}):=\inf_{\boldsymbol{x}\in F}\{f(\boldsymbol{x})\,|\,\boldsymbol{g}(\boldsymbol{x})\leq\boldsymbol{b}\},\ \forall\boldsymbol{b}\in R^m.$

This function is an *antitone* $\boldsymbol{b}\leq\boldsymbol{u}\Rightarrow v(\boldsymbol{b})\geq v(\boldsymbol{u})$. Clearly, $v(\boldsymbol{0})=\inf(\text{CP})<\infty$ iff $S\neq\phi$.

The function $L(x,\lambda):=f(x)+\lambda\cdot g(x):R^n\times R_+^m\to(-\infty,+\infty]$ is the *Lagrangian function* of (CP). For each fixed $\lambda\in R_+^m$, it is a proper convex function in x with domain F. A vector $\overline{\lambda}\in R^m$ is a *Lagrange multiplier vector* for a point $\overline{x}\in F$ if

(1) $\overline{\lambda}\geq 0,\ g(\overline{x})\leq 0,\ \overline{\lambda}\cdot g(\overline{x})=0$ and

(2) $\overline{x}\in\arg\min_x L(x,\overline{\lambda})$, i.e. $\overline{x}$ minimizes the function $L(\cdot,\overline{\lambda})$ over R^n.

If, in addition the convex functions $f, g_1, g_2, ..., g_m$ are actually differentiable at $\bar{x}$, then condition (2) is equivalent to the gradient equation:

$$(2)\text{ô}\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x}) = 0 .$$

A vector pair $(\bar{x}, \bar{\lambda})$ with $\bar{x} \in F, \bar{\lambda} \geq 0$ is a *saddle-point* of the Lagrangian L and the corresponding value $L(\bar{x}, \bar{\lambda})$ a *saddle-value* of L if $L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda})$, $\forall x \in F$, $\forall \lambda \in R_+^m$.

Clearly, $(\bar{x}, \bar{\lambda})$ with $\bar{x} \in F, \bar{\lambda} \geq 0$ is a saddle-point of L iff $\bar{\lambda}$ is a Lagrange multiplier vector for $\bar{x}$.

Theorem 2.4.1 (Lagrangian sufficient condition) *If, in a convex program* (CP) (2.35), *a point* $\bar{x} \in F$ *has a Lagrange multiplier vector, say,* $\bar{\lambda}$ *then* $\bar{x}$ *is an optimal solution to* (CP).

Proof This vector $\bar{\boldsymbol{x}}$ is clearly feasible, i.e. $\bar{\boldsymbol{x}} \in S$. Moreover,

$$\begin{aligned} \boldsymbol{x} \in S \Rightarrow x \in F, \boldsymbol{g}(x) \leq 0 \Rightarrow \bar{\lambda} \cdot g(x) \leq 0 \\ \Rightarrow f(\boldsymbol{x}) \geq f(\boldsymbol{x}) + \bar{\lambda} \cdot g(x) = L(x, \bar{\lambda}) \geq L(\bar{x}, \bar{\lambda}) = f(\bar{\boldsymbol{x}}) \end{aligned}$$

Hence $\bar{x}$ is optimal for (CP). ■

Note this result actually does not depend on convexity.

We can summarize the above relationships to establish the following characterizations.

Theorem 2.4.2 *For a differentiable convex program* (CP)(2.35) *where all the problem functions are differentiable, its Lagrangian L, a vector* $\bar{\lambda} \in R^m$, *and a point* $\bar{x} \in F$, *the following three conditions are equivalent*:

1) *The vector* $\bar{\lambda}$ *is a Lagrange multiplier vector for* $\bar{x}$
2) *The vector pair* $(\bar{x}, \bar{\lambda})$ *forms a saddle-point* of L

3) *The vector pair* $(\bar{x}, \bar{\lambda})$ *satisfies the following Karush-Kuhn-Tucker* (*KKT*) *optimality conditions for* (CP):

(a) $\bar{\lambda} \geq 0,\ g(\bar{x}) \leq 0,\ \bar{\lambda} \cdot g(\bar{x}) = 0$

(b) $\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x}) = 0$.

Proof Show the implications $1) \Rightarrow 2) \Rightarrow 3) \Rightarrow 1)$ are true. ■

We say the convex program (CP) (2.35) satisfies a regularity condition known as the *Slater constraint qualification* (*Slater CQ*) if $\exists \hat{x} \in F$ s.t. $g_i(\hat{x}) < 0,\ \forall i = 1, 2, \ldots, m$. This condition means that $\boldsymbol{0} \in \operatorname{int} \operatorname{dom} v$. The following theorem, a central result in the study of convex programs, tells us when a Lagrange multiplier vector exists for a solution to a convex program.

Theorem 2.4.3 (*A Lagrange multiplier theorem*: Lagrangian necessary conditions). [Borwein and Lewis, 2006, Theorem 3.2.8] *Suppose that the convex program* (CP) (2.35) *satisfies Slater CQ and has an optimal solution* $\bar{x}$. *Then the set of all Lagrange multiplier vectors for* $\bar{x}$ *is a non-empty compact convex subset of* R_+^m.

Proof Since $v(\boldsymbol{0}) = f(\bar{\boldsymbol{x}})$ is finite and $\boldsymbol{0} \in \operatorname{int} \operatorname{dom} v$, the convex value function v is proper (Proposition 2.3.3), and the set $\partial v(\boldsymbol{0})$ is a non-empty compact convex subset of R^m (Proposition 2.3.4). Let $\bar{\boldsymbol{\lambda}}$ be any vector in $-\partial v(\boldsymbol{0})$. Then it must satisfy

(2.37) $$f(\bar{\boldsymbol{x}}) = v(\boldsymbol{0}) \leq v(\boldsymbol{b}) + \bar{\boldsymbol{\lambda}} \cdot \boldsymbol{b},\ \forall \boldsymbol{b} \in R^m .$$

Since $v(\boldsymbol{b}) \leq v(\boldsymbol{0}), \forall \boldsymbol{b} \in R_+^m$, this, together with (2.37) implies that

$$v(\boldsymbol{0}) \leq v(\boldsymbol{b}) + \bar{\boldsymbol{\lambda}} \cdot \boldsymbol{b} \leq v(\boldsymbol{0}) + \bar{\boldsymbol{\lambda}} \cdot \boldsymbol{b},\ \forall \boldsymbol{b} \in R_+^m \Rightarrow 0 \leq \bar{\boldsymbol{\lambda}} \cdot \boldsymbol{b},\ \forall \boldsymbol{b} \in R_+^m \Rightarrow \bar{\boldsymbol{\lambda}} \in R_+^m .$$

Note that $v(g(\boldsymbol{x})) \leq f(\boldsymbol{x}),\ \forall \boldsymbol{x} \in F$. This and (2.37) with $\boldsymbol{b} = g(\boldsymbol{x}),\ \boldsymbol{x} \in F$ implies that

(2.38) $$f(\bar{\boldsymbol{x}}) \le v(g(\boldsymbol{x})) + \bar{\boldsymbol{\lambda}} \cdot g(\boldsymbol{x}) \le f(\boldsymbol{x}) + \bar{\boldsymbol{\lambda}} \cdot g(\boldsymbol{x}) = L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}),\ \forall \boldsymbol{x} \in F$$

In particular, when $\boldsymbol{x} = \bar{\boldsymbol{x}}$, this implies that $0 \le \bar{\boldsymbol{\lambda}} \cdot g(\bar{\boldsymbol{x}}) \Rightarrow \bar{\boldsymbol{\lambda}} \cdot g(\bar{\boldsymbol{x}}) = 0$, as $\bar{\boldsymbol{\lambda}} \ge 0, g(\bar{\boldsymbol{x}}) \le 0$.

It then follows from (2.38) that $L(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}) = f(\bar{\boldsymbol{x}}) \le L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}),\ \forall \boldsymbol{x} \in R^n$, and we see that $\bar{\boldsymbol{\lambda}}$ a Lagrange multiplier vector for $\bar{\boldsymbol{x}}$. ■

Theorem 2.4.4 (***Karush-Kuhn-Tucker theorem***) [Rockafellar, 1970, Corollary 28.3.1] *For a differentiable convex program* (CP) (2.35) *satisfying Slater CQ, its Lagrangian L, and a vector* $\bar{x} \in R^n$, *the following conditions are equivalent*:

1) $\bar{x}$ *is an optimal solution to* (CP).

2) *There exists a vector* $\bar{\lambda} \in R^m$ *such that* $(\bar{x}, \bar{\lambda})$ *forms a saddle-point* of L.

3) *There exists a vector* $\bar{\lambda} \in R^m$ *such that* $(\bar{x}, \bar{\lambda})$ *satisfies the KKT conditions for* (CP):

(a) $\bar{\lambda} \ge 0,\ g(\bar{x}) \le 0,\ \bar{\lambda} \cdot g(\bar{x}) = 0$ (b) $\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x}) = 0$.

2.5 THE EXISTENCE OF A PRIMAL MINIMAL SOLUTION

A function $f : R^n \to \bar{R}$ is (*lower*) *level-bounded* if its level sets $f^{\le \alpha}$ are bounded $\forall \alpha \in \mathbb{R}$, i.e. $|x| \to \infty \Rightarrow f(x) \to \infty$. Note that

(2.39) $$\forall \alpha \in \bar{R},\ f^{\le \alpha} = \bigcap_{\beta > \alpha} f^{< \beta} = \bigcap_{\beta > \alpha} f^{\le \beta}.$$

In particular, if f is a proper function, $\alpha = \inf f < \infty$ then the minimum set, argmin $f = f^{=\alpha} = f^{\le \alpha}$ is the intersection of a nonempty family of nested (i.e. $\beta \le \beta' \Rightarrow f^{\le \beta} \subset f^{\le \beta'}$) non-empty level sets $f^{\le\beta}$ which are compact if *f* is also *lsc* and level-bounded. This minimum set must be non-empty, according to Cantorôs intersection theorem in real analysis.

Theorem 2.5.1 (**Attainment of a minimum**) [Rockafellar and Wets, 2004, 1.9] *Suppose* $f : R^n \to \overline{R}$ *is lsc, proper, and level-bounded. Then* inf f *is finite and the set* argmin f *is nonempty and compact.*

The functions we are going to deal with in CGP problems are finite-valued convex, hence they are *lsc* proper convex, with closed convex level sets whose bounded-ness property are equivalent to the lack of direction of recession. In R^n space, a non-zero vector *d* is a *direction of recession* of a non-empty set *C* if the *ray set* $x + R_+ d := \{x + td \mid t \in R_+\}$ is contained in *C* for all *x* in *C*. We define the (*global*) *recession cone* of *C*, R_C to be the set $\{d \in R^n \mid C + R_+ d \subset C\}$ and the *horizon cone* of *C*, C^∞ to be the unique set *S* with $\mathrm{cl}K \cap (0, R^n) = (0, S),$ where $K := R_+(1, C)$. Below we let $Z_+ = \{0,1,2,...\}$ and $R_{++} = R_+ \setminus \{0\} = (0, +\infty)$.

Theorem 2.5.2 *For any non-empty set C in* R^n, *the recession cone* $R_C = \bigcap\{\frac{C-x}{t} \mid x \in C, t > 0\}$ *is a convex cone, the horizon cone* $C^\infty = \{d \in R^n \mid \exists x^i \in C,\ t_i \to 0^+ \text{ with } t_i x^i \to d\}$ *is a closed cone, and they satisfy*:

$$R_C \subset \tfrac{1}{t} K_C \subset C^\infty, \forall t > 0, \quad R_C \subset R_C'(x) \subset C^\infty, \forall x \in C, \quad (0, C^\infty) = \mathrm{cl}K \setminus R_{++}(1, \mathrm{cl}C)$$

where

$$K_C := \bigcap_{x \in C}(C - x) = \{d \mid C + d \subset C\} = \{d \mid C + Z_+ d \subset C\} \subset \{d \mid x + Z_+ d \subset C\}, \forall x \in C$$
$$R_C'(x) := \bigcap_{t > 0} \tfrac{C-x}{t} = \{d \mid x + R_+ d \subset C\} \subset \{d \mid x + Z_+ d \subset C\} \text{ and } R_C'(x) \text{ is a cone } \forall x \in C,$$

When $C \neq \phi$ *is a closed convex set, these sets coincide into a closed convex cone*

(2.40) $$R_C = K_C = C^\infty = R_C^{'}(x),\ \forall x \in C.$$

If, furthermore, $\boldsymbol{0} \in C$ *then* $R_C = K_C = C^\infty = R_C^{'}(\boldsymbol{0}) = \bigcap_{\lambda>0} \lambda C$.

Note that $R_C = \bigcap_{x\in C} R_C^{'}(x) = \bigcap_{t>0} \frac{1}{t} K_C$ and that $\{d \mid x + Z_+ d \subset C\} \subset C^\infty, \forall x \in C$

Corollary 2.5.3 *For any family of closed convex sets* $\{C_i \mid i \in I\}$ *with non-empty intersection,* $C := \bigcap_{i\in I} C_i \neq \varnothing$ *one has* $C^\infty = \bigcap_{i\in I} C_i^\infty$.

Theorem 2.5.4 *In* R^n *space, a set C is bounded iff* $C^\infty = \{\boldsymbol{0}\} =: \mathscr{O}$, *a closed convex set* $C \neq \phi$ *is bounded iff* $C^\infty = \{d \mid x + R_+ d \subset C\} = \mathscr{O}$, $\forall x \in C$, i.e. $\exists$ no $d \neq \boldsymbol{0}$ and $x \in C$ with $x + R_+ d \subset C$.

Definition 2.5.5 *For any lsc proper convex function f its recession (or horizon) function* f^∞ *is given by*

(2.41) $$f^\infty(d) := \sup_{t>0} \tfrac{1}{t}\left[f(x+td) - f(x)\right] = \lim_{t\to\infty} \tfrac{1}{t}\left[f(x+td) - f(x)\right],\ \text{ if } x \in F$$

In particular,

(2.42) $$f^\infty(d) = \lim_{t\to\infty} \tfrac{1}{t} f(td) = \lim_{\lambda \downarrow 0} \lambda f(d/\lambda),\ \text{ if } 0 \in F$$

Moreover, if *f* is also ph, $f^\infty = f$, e.g. for any nonempty convex set *C*, the function S_C is *lsc* proper sublinear, hence $S_C^\infty = S_C$. If *f* has bounded domain *F*, then for any nonzero *d*, the ray set $x + td$ for large *t* will eventually leave *F*, hence $f^\infty = i_0$, e.g. since the entropy function $ent\left(= geo^*\right)$ has bounded domain $\Delta_n := \left\{\boldsymbol{y} \in R_+^n \mid \boldsymbol{1}^T \boldsymbol{y} = 1\right\}$ (the unit simplex), $ent^\infty = i_0$

Proposition 2.5.6 (A simple rule) *Let* $f : R^n \to \overline{R}$ *be an lsc proper convex function,* $A : n \times m,\ b \in R^n$, $g : R^m \to \overline{R}$ *be defined by* $g(z) := f(Az + b)$ *with domain* $G := A^{-1}(F - b) \neq \phi$.

Then g is also lsc proper convex with its recession function given by $g^\infty(d) = f^\infty(Ad) = f^\infty[(Ad+b)^\infty]$, where $(\boldsymbol{Ad}+\boldsymbol{b})^\infty = [(\boldsymbol{a}^i\boldsymbol{d}+\boldsymbol{b}^i)^\infty]_{i=1}^n = [\boldsymbol{a}^i\boldsymbol{d}]_{i=1}^n = \boldsymbol{Ad}$.

Proof g is clearly proper convex, it is also *lsc* [Rockafellar and Wets, 2004, 1.40 (a)]. If $z \in G$, then $Az + b \in F$ and

$$g^\infty(d) = \lim_{t\to\infty} \tfrac{1}{t}\left[g(z+td) - g(z)\right] = \lim_{t\to\infty} \tfrac{1}{t}\left[f(Az+b+tAd) - f(Az+b)\right] = f^\infty(Ad) \qquad \blacksquare$$

Our knowledge about horizon function can help us calculate the horizon cone of level sets and epigraph of an *lsc* proper convex function.

Theorem 2.5.7 *Let* $f : R^n \to \overline{R}$ *be any lsc proper convex function, then*

(2.43) $\quad (a)\ (f^{\le\alpha})^\infty = (f^\infty)^{\le 0} \qquad (b)\ (\text{epi}\, f)^\infty = \text{epi}\, f^\infty$.

Hence the minimum set argmin f *is non-empty and bounded iff* $\exists$ *no* $d \ne 0$ *such that* $f^\infty(d) \le 0$, *and the recession function* f^∞ *is lsc proper sublinear.*

Proof (a) $\forall \alpha \in (\overline{\alpha}, +\infty)$, *where* $\overline{\alpha} = \inf f < \infty$, the level sets $f^{\le\alpha}$ are non-empty closed convex, hence by (2.40) one has

$$\begin{aligned} d \in (f^{\le\alpha})^\infty &\Leftrightarrow x + td \in f^{\le\alpha},\ \forall t > 0,\ \text{if } x \in f^{\le\alpha} \\ &\Leftrightarrow f(x+td) \le \alpha,\ \forall t > 0,\ \text{if } f(x) \le \alpha \\ &\Leftrightarrow f(x+td) \le f(x),\ \forall t > 0,\ \text{if } x \in F \\ &\Leftrightarrow f^\infty(d) = \sup_{t>0} \tfrac{1}{t}[f(x+td) - f(x)] \le 0,\ \text{if } x \in F \end{aligned}$$

So $\qquad (f^{\le\alpha})^\infty = (f^\infty)^{\le 0},\ \forall \alpha \in (\overline{\alpha}, +\infty)$

(b) Since f is *lsc* proper convex, epi f is non-empty closed convex, hence by (2.40) one has

$$\begin{aligned} (d,\mu) \in (\text{epi}\, f)^\infty &\Leftrightarrow (x+td, \alpha+t\mu) \in \text{epi}\, f,\ \forall t > 0,\ \text{if } (x,\alpha) \in \text{epi}\, f \\ &\Leftrightarrow f(x+td) - t\mu \le \alpha,\ \forall t > 0,\ \text{if } f(x) \le \alpha \\ &\Leftrightarrow f(x+td) - t\mu \le f(x),\ \forall t > 0,\ \text{if } x \in F \\ &\Leftrightarrow f^\infty(d) = \sup_{t>0} \tfrac{1}{t}[f(x+td) - f(x)] \le \mu,\ \text{if } x \in F \end{aligned}$$

So $(\mathrm{epi}\, f)^{\infty} = \mathrm{epi}\, f^{\infty}$ is a closed convex cone, f^{∞} is *lsc* proper sublinear, and the rest are obvious. ■

In the above theorem, the closed convex cone $(f^{\infty})^{\leq 0}$ is called the *recession cone* of *f*, whose nonzero vectors are called *directions of recession* of *f*. The following is an important duality result. For example, for the function $f(x) = \sqrt{x^2+1}$, since $0 \in F$,

$f^{\infty}(d) = \lim_{t\to\infty} \frac{1}{t}\sqrt{t^2 d^2 + 1} = |d|$, the recession cone of *f* is $|d|^{\leq 0} = \mathscr{O}$, so *f* is level-bounded and its minimum set is non-empty and bounded, namely $\{0\}$. Indeed, we have the level sets

$f^{\leq \alpha} = \left[-\sqrt{\alpha^2-1},\ +\sqrt{\alpha^2-1}\right]$ all nonempty and bounded, $\forall \alpha \geq 1$, and $\arg\min f = f^{\leq 1} = \{0\}$.

Theorem 2.5.8 (horizon functions as support functions) [Rockafellar, 1970, Theorem 13.3] *Let $f : R^n \to \overline{R}$ be a proper convex function, then*

1) $f^{*\infty} = S_F$, and $(f^{*\infty})^{\leq 0} = (F)^{-}$ (recession cone of f^* = polar cone of *F*)

2) *if f is also lsc, then* $f^{\infty} = S_{F^*}$, and $(f^{\infty})^{\leq 0} = (F^*)^{-}$ (recession cone of *f* = polar cone of F^*)

Proof (1) By (2.23) we have $\mathrm{epi}\, f^* = \bigcap_{(x,\alpha)\in \mathrm{epi}\, f} \mathrm{epi}\ l_{x,\alpha} \neq \varnothing$, since f^* is proper. Then by Corollary 2.5.3 we have: $(\mathrm{epi}\, f^*)^{\infty} = \bigcap_{(x,\alpha)\in \mathrm{epi}\, f} (\mathrm{epi}\ l_{x,\alpha})^{\infty}$

and then by (2.43) (b), $\mathrm{epi}\, f^{*\infty} = \bigcap_{(x,\alpha)\in \mathrm{epi}\, f} \mathrm{epi}\ l_{x,\alpha}^{\infty} = \bigcap_{(x,\alpha)\in \mathrm{epi}\, f} \mathrm{epi}\ l_{x,0}$

which means $\forall d,\ f^{*\infty}(d) = \sup\{x\cdot d \mid (x,\alpha)\in \mathrm{epi}\, f\} = \sup\{x\cdot d \mid x\in F\} = S_F(d)$,

Hence we have $f^{*\infty} = S_F$, and $(f^{*\infty})^{\leq 0} = (S_F)^{\leq 0} = \left\{y \in R^n \mid \sup_{x\in F} y\cdot x \leq 0\right\} = F^{-}$

(2) If *f* is also *lsc*, then apply (1) to f^* and the result follows. ■

For instance, if *f* is the affine function on R^n, $l_{a,\beta}(x) = a \cdot x - \beta$, then

$$f^*(y) = i_a(y) + \beta,\ f^\infty(d) = a \cdot d = S_{F^*}(d), F^* = \{a\},\ f^{*\infty}(d) = i_0(d) = S_F(d), F = R^n .$$

Since the geometric function *geo* has domain $F = R^n$, $ent^\infty = geo^{*\infty} = S_F = i_0$.

Corollary 2.5.9 *Let C be a nonempty convex set, then*

1) $S_C^\infty = S_C,\ S_C^{\le 0} = C^-$, (recession cone of S_C = polar cone of *C*)

2) *if C is also closed, then* $C^\infty = (B_C)^-$, (recession cone of *C* = polar cone of barrier cone of *C*)

Proof Apply this theorem to the function $f = i_C$. ■

Theorem 2.5.10 (dual properties of minimization) [Rockafellar, 1970, Theorem 27.1] *Let* $f : R^n \to \overline{R}$ *be any lsc proper convex function. Then*

a) $\inf f = -f^*(0)$. *Thus f is bounded below iff* $0 \in F^*$

b) argmin $f = \partial f^*(0)$. *Thus* inf *f is attained iff* $0 \in \text{dom}\ \partial f^*$. *This is implied by* $0 \in \text{ri}\ F^*$.

c) argmin *f is non-empty and bounded iff* $0 \in \text{int}\ F^*$ *iff f has no directions of recession*.

d) argmin $f = \{x\}$ *iff* f^* *is differentiable at 0 and* $\nabla f^*(0) = x$.

The problem setting

We now consider an inequality-constrained convex program (CP) where the index set $K = \{1, \cdots, p\}$, $\breve{K} := \{0\} \cup K$, each function f_k, $\forall k \in \breve{K}$ is an *lsc* proper convex function,

(2.44) (CP) $\inf_{x \in R^n} \{f_0(x) \mid x \in F_0, f_k(x) \le 0,\ \forall k \in K\} =: \inf f$

where $f = f_0 + i_S$ is the objective function of this program (CP), $F_0 = \text{dom} f_0$, and the feasible set is $S = \{\boldsymbol{x} \in F_0 \,|\, f_k(\boldsymbol{x}) \le 0,\ \forall k \in K\}$. We assume that $S \ne \emptyset$ so that for α sufficiently large, the level set of *f*, $f^{\le\alpha} = f_0^{\le\alpha} \cap f_1^{\le 0} \cap \cdots \cap f_p^{\le 0}$ is non-empty and closed convex, and the function *f* is also *lsc* proper convex. The direction of recession of the objective function *f* shall be called the *direction of recession* of the program (CP), which as the following theorem shows, is the direction of recession common to all the problem functions in (CP).

Theorem 2.5.11 *In the above consistent convex program* (CP)(2.44), *the recession cone of its objective function f is*

(2.45) $$(f^{\infty})^{\le 0} = \left\{ d \in R^n \mid f_k^{\infty}(d) \le 0,\ \forall k \in \breve{K} \right\}.$$

Hence the direction of recession of the objective function f is the direction of recession common to all the problem functions f_k *in* (CP), *and the minimum set of the program* (CP) *is non-empty and bounded iff the system* $f_k^{\infty}(d) \le 0,\ k \in \breve{K}$ *has no non-trivial solution in d*, *which means that program* (CP) *has no directions of recession.*

Proof Since for α sufficiently large, the recession cone of *f* is

$$\begin{aligned}
(f^{\infty})^{\le 0} = (f^{\le\alpha})^{\infty} &= (f_0^{\le\alpha} \cap f_1^{\le 0} \cap \cdots \cap f_p^{\le 0})^{\infty}, && \text{by (2.45) (a)} \\
&= (f_0^{\le\alpha})^{\infty} \cap (f_1^{\le 0})^{\infty} \cap \cdots \cap (f_p^{\le 0})^{\infty}, && \text{by Corollary 2.5.3} \\
&= (f_0^{\infty})^{\le 0} \cap (f_1^{\infty})^{\le 0} \cap \cdots \cap (f_p^{\infty})^{\le 0}, && \text{by (2.45) (a)} \\
&= \left\{ d \in R^n \mid f_k^{\infty}(d) \le 0,\ \forall k \in \breve{K} \right\}
\end{aligned}$$

Hence the direction of recession of the objective function *f* is the direction of recession common to all the problem functions f_k in (CP). The minimum set of program (CP) is non-empty and bounded ⇔ f is level-bounded ⇔ the recession cone of *f* is a zero cone

⇔ the system $f_k^{\infty}(d) \le 0,\ k \in \breve{K}$ has no non-trivial solution in *d*

⇔ there are no direction of recession common to all problem functions of (CP).

⇔ Program (CP) has no direction of recession. ■

3.0 ALTERNATIVE PROOFS AND SOME REFINEMENTS OF THE DUALITY THEOREMS OF GP

Since the method we are going to use to derive the basic duality theories for the exponential, quadratic, l_p, and composite (posynomial) GP are significantly different from existing methods in the literature used for posynomial GP, it seems appropriate to test our method on posynomial GP first to confirm its validity, and also to familiarize the readers with this new approach for the subsequent rather complicated extended GP models in the later chapters.

CONVEXITY PROPERTY OF $geo(\boldsymbol{x}) := \ln[\sum_{i=1}^{n} \exp x_i] : R^n \to R$

First note that $geo(x)$ is a finite-valued function with inf $geo(x) = -\infty$ and sup $geo(x) = +\infty$.

Fact 3.1 (A variant of Holderôs inequality) [Hardy, et.al. 1952, pp.21-22] *Given positive weights* $\alpha > 0, \beta > 0, \alpha + \beta = 1,$ *and two sets of non-negative numbers* $\mathrm{U}_i \geq 0, \mathrm{V}_i \geq 0,$ *one has*

(3.1) $$\sum_{i=1}^{n} \mathrm{U}_i^{\alpha} V_i^{\beta} \leq (\sum_{i=1}^{n} \mathrm{U}_i)^{\alpha} (\sum_{i=1}^{n} V_i)^{\beta}$$

where '=' holds exactly when $U_i(\sum_{i=1}^{n} V_i) = V_i(\sum_{i=1}^{n} \mathrm{U}_i), \forall i,$ *namely, the sum of the geometric means is not greater than the geometric mean of the sums, with'=' holding if and only if the two sets of numbers are in proportion.*

Proof When one of the sums $U =: \sum_{i=1}^{n} \mathrm{U}_i$ or $V =: \sum_{i=1}^{n} V_i$ equals zero, the inequality trivially holds as an identity: 0=0. Otherwise, we can divide both sides by $U^{\alpha} V^{\beta}$ and show that

$$\sum_{i=1}^{n}\left(\frac{U_i}{U}\right)^{\alpha}\left(\frac{V_i}{V}\right)^{\beta} \le \sum_{i=1}^{n}\left[\alpha\left(\frac{U_i}{U}\right)+\beta\left(\frac{V_i}{V}\right)\right] = \alpha+\beta = 1,$$

by AGM inequality, then (3.1) follows. Clearly, equality holds iff $U_i/U=V_i/V$, for all i, which means, $U_iV=V_i U$, for all i. ■

The convexity of geometric functions *geo(x)* is an easy consequence of the above inequality.

Fact 3.2 *The geometric function* $geo_n(x) := \ln[\sum_{i=1}^{n}\exp x_i]$ *is convex on* R^n *, specifically*

(3.2) $\quad geo_n(\alpha x+\beta y) \le \alpha\ geo_n(x)+\beta\ geo_n(y), \quad \forall x, y \in R^n, \forall \alpha>0, \beta>0, \alpha+\beta=1,$

with equality holding iff $\exists t \in R, y = x + t1,$ where $1=(1,\cdots,1)^t \in R^n$. (*So it is not strictly convex*)

Proof Setting $U_i = e^{x_i}, V_i = e^{y_i}, \forall i$ in (3.1) yields

$$\sum_{i=1}^{n}(e^{x_i})^{\alpha}(e^{y_i})^{\beta} = \sum_{i=1}^{n}(e^{\alpha x_i+\beta y_i}) \le \left(\sum_{i=1}^{n}e^{x_i}\right)^{\alpha}\left(\sum_{i=1}^{n}e^{y_i}\right)^{\beta}.$$

Take logarithms of both sides of the above, (3.2) follows. Moreover, equality holds iff the two sets of numbers e^{x_i} and e^{y_i} are in proportion: $\exists T>0,\ e^{y_i} = Te^{x_i}, \forall i$, or equivalently, $\exists t(=\ln T), y = x+t1,$ where $1=(1,\cdots,1)^t \in R^n$. ■

Fact 3.3 (Faithful convexity of the function *geo*) [Rockafellar, 1970b] *If geo*(x) *is affine on any line segment* $\mathrm{L}[x,y] =: \{x+r(y-x) \mid r\in[0,1]\}$ *joining two points x and y, it is affine on the line extending that segment*: $L(x,y) =: \{x+r(y-x) \mid r \in R\}$.

Proof If *geo*(x) is affine on [x, y], then by this proposition, (3.2) must hold as equality and y $=x+t\mathbf{1}$, for some real t. We readily see that along the line $L(x,y) =: \{x+r(t1) \mid r\in R\}$, *geo*($x$) is affine with slope t:

(3.3) $\quad geo(x+r(t1)) = geo(x)+rt, \forall r,t \in R$ ■

This proposition says: The function $geo(x)$ is affine along any line in the direction of the sum vector $\mathbf{1}= (1,é\ , 1)^T$, yet it is strictly convex along lines in any other direction. Thus the function $geo(x)$ has linearity 1 and rank n-1 [Rockafellar 1970, p.70]. It is also *strictly isotone*: any points $x^1 \le x^2$ in R^n satisfy $geo(x^1) \le geo(x^2)$ with equality holding iff $x^1 = x^2$.

3.1 MAIN LEMMA OF GP

The GGP formulation of (GP) in the variables $x_i =: \boldsymbol{a}^i \boldsymbol{z},\ i \in I$ is

(3.4) $$(GP)_{\boldsymbol{x}}\ \ \inf_{\boldsymbol{x} \in R^n} geo(\boldsymbol{x}^0 + \boldsymbol{c}^0)\ s.t.\ geo(\boldsymbol{x}^k + \boldsymbol{c}^k) \le 0,\ \forall k \in K,\ \boldsymbol{x} \in \mathscr{P},$$

where $\boldsymbol{a}^i$ is the i^{th} row of the exponent matrix A, $\mathscr{P} = \left\{ A\boldsymbol{z} \mid \boldsymbol{z} \in R^m \right\}$ is the column space of A, $\boldsymbol{x}^k$ is the sub-vector of $\boldsymbol{x}$ obtained by deleting all of its components x_i for which i is not in $[k]$, and $geo(\boldsymbol{x}^k + \boldsymbol{c}^k) = \ln\left\{ \sum_{i \in [k]} \exp(x_i + c_i) \right\}$ is convex by Fact 3.2. Of course, we have $geo(\boldsymbol{x}^k + \boldsymbol{c}^k) = g_k(\boldsymbol{z}) = \ln G_k(\boldsymbol{t}), \forall k \in K$. The associated dual program (GD) is given below with each dual variable δ_i corresponding to a primal term U_i.

Dual posynomial program (*GD*) under the convention that 0^0=1.

$$
\text{(3.5)} \qquad \text{(GD)} \begin{cases} \sup\limits_{\delta \in R_+^n} V(\delta) := \prod_{k\in \breve{K}} \prod_{i\in[k]} \left(C_i\lambda_k/\delta_i\right)^{\delta_i} \\ \qquad\qquad = \prod_{i=1}^{n} \left(C_i/\delta_i\right)^{\delta_i} \cdot \prod_{k=1}^{p} \lambda_k^{\lambda_k} \\ s.t. \qquad \lambda_0 = 1, \\ \qquad \sum_{i=1}^{n} a_{ij}\delta_i = 0,\ j \in J \\ \text{where } \lambda_k =: \sum_{i\in[k]} \delta_i\ ,\ k \in \breve{K}. \end{cases}
$$

We now provide an alternative proof for the main lemma of (GP) (*cf* proof of Lemma 2.0).

Lemma 3.1.0 (Main Lemma of GP) *If **t** is feasible for primal program* (GP) *and* δ *is feasible for dual program* (GD)*, then*

$$G_0(\boldsymbol{t}) \geq V(\delta).$$

Moreover, under the same conditions $G_0(\boldsymbol{t}) = V(\delta)$, *if, and only if, one of the following two sets of equivalent conditions holds:*

$$
\text{(3.6)} \qquad \begin{cases} G_k(\boldsymbol{t})^{\lambda_k} = 1,\ \forall k \in K, & \text{(a)} \\ \delta_i G_k(\boldsymbol{t}) = \lambda_k U_i(\boldsymbol{t}),\ \forall i \in [k] & \text{(b)},\ \forall k \in \{0\} \cup K \end{cases} \qquad \text{(Extremality condition 1)}
$$

$$
\text{(3.7)} \qquad \delta_i = \begin{cases} U_i(\boldsymbol{t})/G_0(\boldsymbol{t}), & i \in [0] \quad \text{(c)} \\ \lambda_k U_i(\boldsymbol{t}), & i \in [k], \quad \text{(d)},\ \forall k \in K \end{cases} \qquad \text{(Extremality condition 2)}
$$

*in which case **t** is optimal for primal program* (*GP*) *and* δ *is optimal for dual program* (*GD*).

(Recall from Lemma 2.1.0 that the above two sets of extremality conditions are equivalent.)

Proof: Assume that ***t*** is feasible for primal program (*GP*) and δ for dual program (GD), then the transformed variables $\boldsymbol{x} = A\boldsymbol{z}$, with $x_i = \boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z}$, $z_j = \ln t_j$, $j \in J$, $i \in I$ are feasible for (G*P*)$_{\boldsymbol{x}}$, thus we have $geo(\boldsymbol{x}^k + \boldsymbol{c}^k) \leq 0,\ \forall k \in K,\ \boldsymbol{x} \in \mathscr{P},$ and $\lambda_k \geq 0,\ \forall k \in K$, so that by (2.3), $\boldsymbol{x} \cdot \boldsymbol{\delta} = 0$ and

$$
\text{(3.8)} \qquad geo(\boldsymbol{x}^0 + \boldsymbol{c}^0) \geq geo(\boldsymbol{x}^0 + \boldsymbol{c}^0) + \sum_{k=1}^{p} \lambda_k geo(\boldsymbol{x}^k + \boldsymbol{c}^k).
$$

Applying the generalized geometric inequality (2.33) to each function $\lambda_k geo(\boldsymbol{x}^k + \boldsymbol{c}^k)$, $k \in \breve{K}$, with λ_0=1 yields

$$(3.9)\qquad \begin{aligned} \lambda_k geo(\boldsymbol{x}^k + \boldsymbol{c}^k) &\geq -\sum\nolimits_{i\in[k]} \delta_i \ln\left(\delta_i / \lambda_k\right) + \sum\nolimits_{i\in[k]} (x_i + c_i)\delta_i \\ &= \sum\nolimits_{i\in[k]} \delta_i \ln\left(C_i \lambda_k / \delta_i\right) + \sum\nolimits_{i\in[k]} x_i \delta_i, \quad \forall k \in \breve{K} \end{aligned}$$

where equality holds iff $\delta^k = \lambda_k \nabla geo(\boldsymbol{x}^k + \boldsymbol{c}^k)$, i.e., $\delta_i = (\lambda_k U_i(\boldsymbol{t}) / G_k(\boldsymbol{t})),\ \forall i \in [k]$.

Summing over all these inequalities yields

$$(3.10)\qquad \begin{aligned} \sum\nolimits_{k=0}^{p} \lambda_k geo(\boldsymbol{x}^k + \boldsymbol{c}^k) &\geq \sum\nolimits_{k=0}^{p} \sum\nolimits_{i\in[k]} \delta_i \ln\left(C_i \lambda_k / \delta_i\right) + \sum\nolimits_{k=0}^{p} \sum\nolimits_{i\in[k]} x_i \delta_i \\ &= \sum\nolimits_{k=0}^{p} \sum\nolimits_{i\in[k]} \delta_i \ln\left(C_i \lambda_k / \delta_i\right) = \ln V(\delta),\ \because \boldsymbol{x} \cdot \boldsymbol{\delta} = 0 \end{aligned}.$$

Combining inequalities (3.8) and (3.10) shows

$$(3.11)\qquad geo(\boldsymbol{x}^0 + \boldsymbol{c}^0) \geq \sum\nolimits_{k=0}^{p} \lambda_k geo(\boldsymbol{x}^k + \boldsymbol{c}^k) \geq \sum\nolimits_{k=0}^{p} \sum\nolimits_{i\in[k]} \delta_i \ln\left(C_i \lambda_k / \delta_i\right) = \ln V(\delta)$$

Hence $geo(\boldsymbol{x}^0 + \boldsymbol{c}^0) \geq \ln V(\delta)$ with equality holding iff the two inequalities (3.8) and (3.9) both hold as equalities, which means complementary slackness conditions $\lambda_k geo(\boldsymbol{x}^k + \boldsymbol{c}^k) = 0, \forall k \in K$ as well as the gradient (or proportionality) conditions $\delta_i G_k(\boldsymbol{t}) = \lambda_k U_i(\boldsymbol{t}),\ \forall i \in [k],\ \forall k \in \breve{K}$ must hold true. Exponentiating the far two sides of inequality (3.11) and the above complementary slackness conditions, one obtains $G_0(\boldsymbol{t}) \geq V(\delta)$ with equality holding iff the Extremality condition 1 hold true. ■

In (3.10), the function on the left is the Lagrangian of the primal program (GP)$_{\boldsymbol{x}}$ given by $geo(\boldsymbol{x}^0 + \boldsymbol{c}^0) + \sum\nolimits_{k=1}^{p} \lambda_k geo(\boldsymbol{x}^k + \boldsymbol{c}^k)$, and the negative of its conjugate is the last expression, namely $\sum\nolimits_{k=0}^{p} \sum\nolimits_{i\in[k]} \delta_i \ln\left(C_i \lambda_k / \delta_i\right) = \ln V(\delta)$. Thus the dual objective $V(\delta)$ of (GD) is the exponential of the negative of the conjugate of the Lagrangian of the primal program

(GP)$_{\boldsymbol{x}}$. We have actually proved the main lemma of (GP) by the conjugate inequality for the Lagrangian of (GP)$_{\boldsymbol{x}}$ and its conjugate. *We shall see that this same principle also works for* (CGP).

It follows from this lemma that we always have $\infty \geq \inf(\text{GP}) \geq \sup(\text{GD}) \geq 0$. If (GP) and (GD) are both feasible, and inf (GP) = sup (GD), we say there is no *duality gap* between these two programs. One way to show this is to prove that $\forall \varepsilon > 0, \exists \boldsymbol{t} \in T, \delta \in \Delta$, s.t. $G_0(\boldsymbol{t}) - V(\delta) \leq \varepsilon$, where T and æ are the feasible sets for the programs (GP) and (GD), respectively. If, in addition, the two optimum values are both attained, i.e. min (GP) = max (GD), we say *perfect duality* holds. In this case, if an optimum solution for one program can be easily converted to an optimum solution for the other program, then the two programs (GP) and (GD) may be deemed equivalent in this sense. The duality theory for GP attempts to find the conditions under which the above desirable outcomes would result, so that it is possible to solve the primal through its dual. This is referred to as a GP dual approach.

EXAMPLE 3.1.1 (Maximum Likelihood Estimator of a Bernoulli Parameter) [Ross, 2004, p.231] *Suppose that n independent trials, each of which is a success with probability p, are performed. What is the maximum likelihood estimator of p*?

Solution The likelihood function to be maximized is

$$f(x_{1,} \cdots, x_n \mid p) = p^k (1-p)^{n-k}, \text{ with } k = \sum\nolimits_{i=1}^{n} x_i,\ x_i = 0 \text{ or } 1, \forall i$$

This can be formulated as a three-term, two-variable (*n=3, m=2*) GP example:

$$\inf_{(p,q)>0} f^{-1} = p^{-k} q^{k-n} \quad \text{s.t. } p + q \leq 1$$

The exponent matrix and a unique dual solution is

$$\begin{array}{c} \\ \delta_1 \\ \delta_2 \\ \delta_3 \end{array} \begin{array}{c} \begin{array}{cc} p & q \end{array} \\ \begin{bmatrix} -k & k-n \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \delta^* \\ 1 \\ k \\ n-k \end{array} \begin{array}{c} \lambda_1^* = n \\ \\ \\ \\ \end{array} ,$$

From extremality condition 2, i.e., (3.7) (d), we see that

$$\frac{k}{n} = \frac{\delta_2^*}{\lambda_1^*} = p^*, \text{ so } \breve{p} = \frac{\sum_1^n X_i}{n}. \qquad \blacksquare$$

The following corollary is immediate from the main lemma of GP.

Weak Duality Theorem of GP

i) *Always,* $\infty \geq \inf(\text{GP}) \geq \sup(\text{GD}) \geq 0.$

ii) *When* (GP) *is feasible,* $\infty > \inf(\text{GP}) \geq \sup(\text{GD}).$

i.e. if sup(GD)=∞ , (GP) *is infeasible*.

iii) *When* (GD) *is feasible,* $\inf(\text{GP}) \geq \sup(\text{GD}) > 0.$

i.e. if inf (GP)=0 , (GD) *is infeasible*.

iv) *When* (GP) *and* (GD) *are both feasible,* $\infty > \inf(\text{GP}) \geq \sup(\text{GD}) > 0.$

3.2 FIRST DUALITY THEOREM OF GP

In this section we shall work with the $(\text{GP})_z$ formulation in the variables $z_j := \ln t_j,\ j \in J$

$$(\text{GP})_z \quad \inf_{z \in R^m} g_0(z) \ s.t.\ g_k(z) \leq 0,\ \forall k \in K$$

where $\boldsymbol{z} = (z_1, \cdots, z_m)$ and $g_k(\boldsymbol{z}) =: \ln G_k(\boldsymbol{t}) = \ln\left\{\sum_{i\in[k]} \exp[\boldsymbol{a}^i \boldsymbol{z} + c_i]\right\} = geo(\boldsymbol{A}^k \boldsymbol{z} + \boldsymbol{c}^k)$ is convex by Proposition 2.3.1 and Fact 3.2. The Lagrangian for this differentiable convex program (GP)$_z$ is

$$l(\boldsymbol{z}, \lambda) := g_0(\boldsymbol{z}) + \sum_{k\in K} \lambda_k g_k(\boldsymbol{z}),\ \forall \lambda \in R_+^p, \boldsymbol{z} \in R^m.$$

Theorem 3.2.1 (cf Theorem 2.1.1) **(First Duality Theorem of GP)** *Suppose that primal program* (GP) *is superconsistent. Then the following 3 conditions are equivalent*:

1) ***t'*** *is a minimal solution to* (GP).
2) *There exists a vector* $\lambda' \in R_+^p$ *for* ***z'*** (*where* ***z'*** *= ln* ***t'***) *such that* $(\boldsymbol{z}', \lambda')$ *forms a saddle point of* $l(\boldsymbol{z}, \lambda)$.
3) *There exists a vector* $\lambda' \in R_+^p$ for ***z'*** (*where* ***z'*** *= ln* ***t'***) *such that* $(\boldsymbol{z}', \lambda')$ *satisfies the KKT conditions for* (GP)$_z$:

$$(a)\ \lambda_k' \geq 0,\ g_k(\boldsymbol{z}') \leq 0,\ \lambda_k' g_k(\boldsymbol{z}') = 0, \forall k \in K$$
$$(b)\ \nabla_z l(\boldsymbol{z}', \lambda') = \sum_{k\in \bar{K}} \lambda_k' \nabla g_k(\boldsymbol{z}') = \boldsymbol{0},\ where\ \lambda_0' = 1$$

in which case the set of all such vectors $\boldsymbol{\lambda}'$ *is a non-empty compact convex subset of* R_+^p, *and the dual program* (GD) *also has a maximum solution* δ' *such that*

$$\min(\text{GP}) = G_0(\boldsymbol{t}') = V(\delta') = \max(\text{GD}) \qquad \text{(Perfect duality)}$$

Moreover, each pair of primal and dual optimal solution $(\boldsymbol{t}', \boldsymbol{\delta}')$ *satisfies*

$$U_i(\boldsymbol{t}') = \begin{cases} \delta_i' V(\delta'), & i \in [0] \\ \delta_i' / \lambda_k', & i \in [k],\ k \in K, \text{ and } \lambda_k' > 0 \end{cases} \tag{3.12}$$

Proof: By assumption, the differentiable convex program (GP)$_z$ satisfies the Slater CQ. Hence by KKT Theorem (Theorem 2.4.4), the above conditions 1) through 3) are equivalent and therefore, by Theorem 2.4.3 the set of all such vectors $\boldsymbol{\lambda}'$ is a non-empty compact convex subset of R_{+}^{p}, and conditions (a) and (b) in 3) hold. We note that

Condition (b) $\Rightarrow \sum_{k\in K} \lambda_k^{'} \frac{\partial g_k(\boldsymbol{z}')}{\partial z_j} = 0,\ \forall j \in J \Rightarrow \sum_{k\in K} \frac{\lambda_k^{'}}{G_k(\boldsymbol{t}')} \sum_{i\in[k]} a_{ij} U_i(\boldsymbol{t}') = 0,\ \forall j \in J$, by (1.8)

If we define $\delta_i^{'} = \lambda_k^{'} U_i(\boldsymbol{t}') / G_k(\boldsymbol{t}') \geq 0, \forall i \in [k], \forall k \in K$ in the above, we have the orthogonality condition in (GD): $\sum_{k\in K} \sum_{i\in[k]} a_{ij}\delta_i^{'} = \sum_{i\in I} a_{ij}\delta_i^{'} = 0,\ \forall j \in J.$

Condition (a) $\Rightarrow G_k(\boldsymbol{t}')^{\lambda_k^{'}} = 1,\ \forall k \in K$, this is condition (a) in the extremality condition 1 (3.6) for main lemma of (GP), and the condition (b) in (3.6) is clearly true by our definition of δ'. Lastly, we see that $\sum_{i\in[k]} \delta_i^{'} = \sum_{i\in[k]} [\lambda_k^{'} U_i(\boldsymbol{t}') / G_k(\boldsymbol{t}')] = \lambda_k^{'},\ \forall k \in K$. We have thus defined a dual feasible solution δ' which, together with the minimal solution t', satisfies the extremality condition 1 in (3.6). Hence by main lemma of (GP), this δ' is a maximal solution satisfying

$$\min(\text{GP}) = G_0(\boldsymbol{t}') = V(\delta') = \max(\text{GD})$$

Since (3.6) and (3.7) are equivalent, and the latter condition implies(3.12). ■

Obviously, our Theorem 3.1 has a stronger conclusion than that of Theorem 2.1.

3.3 SECOND DUALITY THEOREM OF GP

To prove the second duality theorem of (GP) we will apply **Theorem 2.5.10** to $(\text{GP})_z$ which fits into the problem setting (2.44), because all of the problem functions $g_k(\boldsymbol{z}) = \ln\left\{\sum_{i\in[k]} \exp[\boldsymbol{a}^i \boldsymbol{z} + c_i]\right\} = geo(\boldsymbol{A}^k \boldsymbol{z} + \boldsymbol{c}^k)$ are finite convex on R^m. Hence they are *lsc* proper convex. Thus we need to compute the recession functions g_k^∞ for all of the problem functions g_k, and that leads us to the computation of geo^∞, as $g_k^\infty(\boldsymbol{d}) = geo^\infty(\boldsymbol{A}^k\boldsymbol{d})$ by Proposition 2.5.6 and by the fact that $g_k(\boldsymbol{z})=geo(\boldsymbol{A}^k\boldsymbol{z}+\boldsymbol{c}^k)$.

Recession function of *geo* (*x*)

Since $geo(x)$ is finite-valued convex with $\boldsymbol{0}$ in the domain, it is *lsc* proper convex and we can apply (2.42) to compute its recession function:

(3.13) $$geo^\infty(\boldsymbol{d}) = \lim_{\lambda\downarrow 0} \lambda geo(\boldsymbol{d}/\lambda) = \lim_{t\uparrow\infty} \frac{1}{t} geo(t\boldsymbol{d})$$

We define the max function $\max : R^n \to R$ by $max_n(x) =: \max_{i=1}^n x_i$ for $x = (x_1,\cdots,x_n) \in R^n$ (The subscript n can usually be omitted). This function is clearly finite sublinear, hence it is *lsc* and proper. It is also an *isotone*: any points $x^1 \le x^2$ in R^n satisfy $max(x^1) \le max(x^2)$, and is not strictly convex (since $max(tx) = t\, max(x)$, for all $t \ge 0$).

Fact 3.4 (recession function of geometric function *geo*) *The non-smooth function* $max_n(x)$ *can be approximated uniformly, to any accuracy, by a smooth function* $(\lambda\star geo_n)(x) =: \lambda geo_n(x/\lambda)$:

(3.14) $$0 \le \sup_x[\lambda \mathrm{geo}_n(x/\lambda) - \max_n(x)] \le \lambda \ln n, \ \forall\lambda > 0.$$

(3.15) $$\mathrm{geo}^\infty(x) = \max(x), \forall x$$

Proof Let $max_n(x) = x_k$. Then $e^{x_k} \le \sum_1^n e^{x_i} \le ne^{x_k}$ and thus $x_k \le geo_n(x) \le x_k + \ln n$, which implies

(3.16) $$0 \le geo_n(x) - \max_n(x) \le \ln n, \ \forall x$$

Homogenize this and we get(3.14). Let $\lambda\downarrow 0$ in(3.14), then $\lambda geo_n(x/\lambda) \to \max_n(x)$, uniformly, so (3.15) follows by(3.13). ■

An alternative proof of (3.15)

Since $geo^* = ent$ with domain $F^* = \Delta_n$ (the unit simplex), one has by Theorem 2.5.8

$$
\begin{aligned}
geo^\infty(\boldsymbol{d}) = S_{\Delta_n}(\boldsymbol{d}) &= \sup\left\{\boldsymbol{d}\cdot\boldsymbol{y} \mid \boldsymbol{1}^T\boldsymbol{y} = 1,\ \boldsymbol{y}\in R_+^n\right\}, \text{ a knapsack problem} \\
&= \inf\left\{\lambda \mid \lambda \ge d_i, \forall i = 1,\cdots,n\right\}, \text{ by LP duality theory} \\
&= \max(\boldsymbol{d})
\end{aligned}
$$
■

Since $g_k(\boldsymbol{z}) = geo(\boldsymbol{A}^k\boldsymbol{z} + \boldsymbol{c}^k)$, by Proposition 2.5.6 we have $g_k^\infty(\boldsymbol{d}) = geo^\infty(\boldsymbol{A}^k\boldsymbol{d}) = \max(\boldsymbol{A}^k\boldsymbol{d})$.

A theorem of the alternatives from linear programming duality theory

Of the following two linear systems exactly one has a solution (*where A is an n by m matrix*):

(I) *Find* $\boldsymbol{z}$ *with* $0 \ne A\boldsymbol{z} \le 0$ (II) *Find* $\boldsymbol{y}$ *with* $\boldsymbol{y} > 0,\ A^T\boldsymbol{y} = 0$

Recall that program (GD) is canonical if system (II) has a solution with A being its exponent matrix.

Theorem 3.3.1 (cf Theorem 2.1.2) **(Second Duality Theorem of GP)** *Suppose that primal program* (GP) *is consistent. Then the minimum set of program* (GP)$_z$ *is non-empty and bounded iff dual program* (GD) *is canonical, in which case program* (GP) *has a minimum solution* ***t'***.

Obviously, this theorem is stronger than the original second duality theorem of GP (Theorem 2.2), and it also includes its converse.

Proof By assumption, program (GP)$_z$ is also consistent. Now

$$
\begin{aligned}
g_k^\infty(\boldsymbol{d}) \le 0,\ k \in \check{K} &\Leftrightarrow \max(\boldsymbol{A}^k\boldsymbol{d}) \le 0,\ k \in \check{K} \\
&\Leftrightarrow \boldsymbol{A}^k\boldsymbol{d} \le \boldsymbol{0},\ k \in \check{K} \\
&\Leftrightarrow \boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{d} \le 0,\ i \in [k], k \in \check{K} \\
&\Leftrightarrow \boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{d} \le 0,\ i \in I \Leftrightarrow \boldsymbol{Ad} \le \boldsymbol{0}
\end{aligned}
$$

Since A is of full column rank, it is one-to-one, hence $\boldsymbol{d} \ne 0 \Rightarrow \boldsymbol{Ad} \ne 0$. Therefore by Theorem 2.5.10, one has

The minimum set of program $(GP)_z$ is non-empty and bounded

⇔ The system $g_k^{\infty}(d) \leq 0,\ k \in \breve{K}$ has no non-trivial solution in d

⇔ ∃ no d with $0 \neq Ad \leq 0$, i.e. system (I) above has no solution

⇔ System (II) above has a solution

⇔ Program (GD) is canonical

In this case, program (GP) must have a minimum solution ***t'***. ■

The original second duality theorem (Theorem 2.2) has been refined by the original authors:

Theorem 3.3.2 (Strong Duality Theorem of GP) [Duffin et.al 1967, p.173, Theorem 1(ii)]

If dual program (GD) *is canonical and has a finite positive supremum, then primal program* (GP) *has a minimum solution* $\boldsymbol{t}'$ *closing the duality gap*:

$$G_0(\boldsymbol{t}') = \min(GP) = \sup(GD)$$

Compared to Theorem 2.2, the assumption of this theorem is weaker, since by the weak duality theorem of GP (iv) the condition that primal program (GP) is consistent and dual program (GD) is canonical implies the condition that dual program (GD) is canonical and has a finite positive supremum. The conclusion of this theorem is also much stronger, because it also says that there is no duality gap between (GP) and (GD).

On the other hand, since we have shown in Fact 3.3 that the geometric function is faithfully convex, Theorem 3 of [Rockafellar, 1970b] implies the following refinement of the first and second duality theorems (Theorem 2.1 and 2.2).

Theorem 3.3.3 *If primal program* (GP) *is consistent, then*

$$\infty > \inf(GP) = \sup(GD) \geq 0.$$

Moreover, if (GP) *is superconsistent and if the common value in the above equation is positive, then* (GD) *has a maximal solution*.

On the other hand, if dual program (GD) *is canonical, then*

$$\infty \geq \inf(GP) = \sup(GD) > 0 .$$

Moreover, if the common value in the above equation is finite, then (GP) *has a minimal solution*.

(The second part of this theorem is a rephrase of Theorem 3.4).

Corollary 3.3.4 *If primal program* (GP) *is consistent and dual program* (GD) *is canonical, then* (GP) *has a minimal solution and*

$$\infty > \min(GP) = \sup(GD) > 0 .$$

Moreover, if (GP) *is superconsistent, then* (GD) *also has a maximal solution, hence*

$\min(GP) = \max(GD)$ *, and both are attained.* (Perfect duality)

We can also use Theorem 2.5.7 to prove some side facts.

Fact 3.5 [Rockafellar, 1970a, p.68]. *The vector* $y \neq 0$ *is a direction of recession for the function* $geo(x)$, *iff one of the following equivalent conditions holds*

0) $y \leq 0$.

1) *The half-line* $H =: \{x + ty \mid t \geq 0\}$ *is contained in a level set* $geo^{\leq \alpha} =: \{x \mid geo(x) \leq \alpha\}$.

2) *The function* $geo(x)$ *is bounded above on the half-line* $H =: \{x + ty \mid t \geq 0\}$.

3) *The function* $geo(x)$ *is non-increasing on the line* $L =: \{x + ty \mid t \in R\}$.

Fact 3.6 [Rockafellar, 1970a, p.69]. *Along any direction* $\boldsymbol{y} \neq \boldsymbol{0}$ *the function* $geo(\boldsymbol{x})$ *is not constant on the line* $L =: \{x + ty \mid t \in R\}$. *Equivalently put*:

1) *The line* $L =: \{x + ty \mid t \in R\}$ *is not contained in any level set* $geo^{\leq \alpha} =: \{x \mid geo(x) \leq \alpha\}$.

2) *The function* $geo(x)$ *is unbounded above on the line* $L =: \{x + ty \mid t \in R\}$.

Fact 3.7 [Rockafellar, 1970a, p.70]. *The vector* $\boldsymbol{y \neq 0}$ *is an affine direction for the function geo*(x) *with slope* β *iff* y=[illegible]1.

4.0 DUALITY THEORY OF EXPONENTIAL GP

Exponential Geometric Program (EGP) has many applications as we shall point out later in this chapter. A partial special case of this, called *Transcendental Geometric Program* [Lidor and Wilde, 1978] was reported to have its dual modeling chemical equilibrium problems for *non-ideal* systems [Lidor, 1975]. An EGP problem arises, when, in a posynomial program, some posynomial term, say, $U_i(\boldsymbol{t})$ is multiplied by an exponential factor of another posynomial, $\exp E_i(\boldsymbol{t})$, where $E_i(\boldsymbol{t}) := \sum_{l \in <i>} V_l(\boldsymbol{t})$ and $V_l(\boldsymbol{t})$ are posynomial terms.

4.1 PROBLEM FORMULATION OF EGP

The problem formulation of an EGP is:

Primal EGP problem

$$(EGP)\ \inf_{\mathbf{0} < \boldsymbol{t} \in R^m} \tilde{G}_0(\boldsymbol{t})\ s.t.\ \tilde{G}_k(\boldsymbol{t}) \le 1,\ k \in K,$$

where

(4.1)
$$\tilde{G}_k(\boldsymbol{t}) := \sum_{i \in [k]} \left[U_i(\boldsymbol{t}) \cdot \exp \sum_{l \in <i>} V_l(\boldsymbol{t}) \right],\ k \in \breve{K} := \{0\} \cup K$$
$$U_i(\boldsymbol{t}) = C_i \prod_{j \in J} t_j^{a_{ij}},\ \text{and}\ V_l(\boldsymbol{t}) = D_l \prod_{j \in J} t_j^{b_{lj}},\ \text{are posynomial terms},$$

Alternatively put:

(4.2) $$\tilde{G}_k(\boldsymbol{t}) := \sum\nolimits_{i\in[k]} \tilde{U}_i(\boldsymbol{t}),\ \tilde{U}_i(\boldsymbol{t}) := U_i(\boldsymbol{t})\cdot \exp E_i(\boldsymbol{t}),\ E_i(\boldsymbol{t}) := \sum\nolimits_{l\in<i>} V_l(\boldsymbol{t})$$

For easy reference, we call $E_i(\boldsymbol{t})$ a *second tier posynomial* whose terms $V_l(\boldsymbol{t})$ are said to be associated with the first tier term $U_i(\boldsymbol{t})$, $\tilde{U}_i(\boldsymbol{t})$ a *posynential term*, and $\tilde{G}_k(\boldsymbol{t})$ a *posynential function*. The index sets $<i>$ will be explained below.

In this program, the index sets *I*, *J*, *K* and the partition of *I* into subsets [*k*] are the same as those for GP in section 2.1. We assume that $\forall i \in I$ the posynomial $E_i(\boldsymbol{t})$ consists of r_i second tier posynomial terms, where $r_i = |<i>|$ =size of $<i>$, $r_i \geq 0$, with $r_i = 0$ indicating that $<i> = \phi$ and $E_i(\boldsymbol{t})$ is non-existent. Let $r = r_1 + \ldots + r_n$.

If $r > 0$, we define a new index set $L = \{1, \cdots, r\}$ corresponding to an ordered-list of second tier posynomial terms $V_1, \cdots, V_r$, whose exponents also form a *second tier exponent matrix* $B = [b_{lj}] : r \times m$, together with *A*, they constitute the (*composite*) *exponent matrix* [illegible] = [illegible] for this program. The set *L* also has a sequential partition

$$L = <1> \cup <2> \cup \cdots \cup <n>,$$

where $<1>$ consists of the first r_1 integers from *L*, $<2>$ the next r_2 integers, and so forth, $<n>$ the last r_n integers. The data for this program is the $(n+r)\times(m+1)$ matrix [illegible] and the partition structures of *I* and *L*, where $c_i = \ln C_i,\ i \in I,\ d_l = \ln D_l,\ l \in L$.

The elements of $<i>$ can also be explicitly listed. Indeed, if $S_k =: r_1 + \ldots r_k$, $\forall k \in I$, and $S_0 = 0$, then $S_n = r$, $r_i = S_i - S_{i-1}$, and

(4.3) $$<1> = \{1, \cdots, S_1\}, \cdots, <i> = \{S_{i-1}+1, \cdots, S_i\}, \cdots, <n> = \{S_{n-1}+1, \cdots, S_n\}$$

Let $I^+ =: \{i \in I \,|< i >\neq \phi\}$, then $L = \cup_{i\in I^+} < i >$, and $r = \sum_{i\in I^+} r_i$.

$$\begin{aligned}&\text{For } i \in I \setminus I^+,\ \ < i >= \phi,\ E_i(\boldsymbol{t}) = 0,\ \tilde{U}_i(\boldsymbol{t}) = U_i(\boldsymbol{t})\\ &\text{For } [k] \subset I \setminus I^+ \text{ (i.e. } [k] \cap I^+ = \phi),\ \tilde{G}_k(\boldsymbol{t}) = \sum_{i\in[k]} U_i(\boldsymbol{t}) := G_k(\boldsymbol{t})\end{aligned}$$

If $r = 0,$ then $L = \phi, < i >= \phi,\ \forall i \in I$, program (EGP) reduces to a posynomial program (GP), which we shall call *the underlying* GP of this (EGP). This is the same (GP) as we would obtain, if we set the second tier exponent matrix B to zero, i.e. set $r = 0$. Since $E_i(\boldsymbol{t}) \geq 0$, one has

$$\tilde{U}_i(\boldsymbol{t}) \geq U_i(\boldsymbol{t}),\ \tilde{G}_k(\boldsymbol{t}) \geq G_k(\boldsymbol{t}),\ \text{and } \inf(EGP) \geq \inf(GP)$$

Let us illustrate this model with some examples.

Example 4.1.1 An EGP (with n=4, r=3)

$$\begin{cases}\min_{t>0} \tilde{G}_0(\boldsymbol{t}) = U_1 + U_2 e^{V_1+V_2} \\ \ s.t. \quad \tilde{G}_1(\boldsymbol{t}) = U_3 e^{V_3} + U_4 \leq 1\end{cases}, \text{ where } \begin{cases}U_1, \cdots, U_4 \text{ are the first tier posynomial terms}\\ V_1, \cdots, V_3 \text{ are the second tier posynomial terms}\end{cases}$$

$<1>= \phi, <2>= \{1,2\}, <3>= \{3\}$, and $<4>= \phi$. The composite exponent matrix is of order 7 by m, where m = the number of design variables, and the degree of difficulty is 6-m. ■

Example 4.1.2 When the objective function to be minimized is a posynomial multiplied by an exponential factor of another posynomial, e.g.

$$\min_{\boldsymbol{t}>\boldsymbol{0}} \left(\sum_{i\in[0]} U_i(\boldsymbol{t})\right) \cdot \exp\left(\sum_{l\in<0>} V_l(\boldsymbol{t})\right)$$, with m = the dimension of $\boldsymbol{t}$.

This can be modeled as an EPP:

$$\begin{cases}\min\limits_{(t_0,\boldsymbol{t})>\boldsymbol{0}} \ t_0 \cdot \exp\left(\sum_{l\in<0>} V_l(\boldsymbol{t})\right) \\ \ \text{s.t.} \quad t_0^{-1}\left(\sum_{i\in[0]} U_i(\boldsymbol{t})\right) \leq 1\end{cases} \text{ where } \begin{cases}n = 1 + |[0]|,\ r = |<0>|,\\ \text{degree of difficulty } d_1 = |[0]| + |<0>| - m - 1\end{cases} \quad ■$$

Caution: This problem also has an equivalent EPP formulation:

$$\min_{t>0} \sum_{i\in[0]} \left[U_i(\boldsymbol{t}) \cdot \exp\left(\sum_{l\in<0>} V_l(\boldsymbol{t}) \right) \right], \text{ where } n = |[0]|,\ r = |[0]| \cdot |<0>|,$$

but the degree of difficulty $d_2 = |[0]| + |[0]| \cdot |<0>| - m - 1$ could be much higher than d_1.

Since in the variables $z_j =: \ln t_j, j \in J$, $U_i = e^{\boldsymbol{a}^i z + c_i}$, $V_l = e^{\boldsymbol{b}^l z + d_l}$, where $d_l =: \ln D_l$,

and $\boldsymbol{b}^l$ is the l^{th} row of the matrix B, the problem functions of program (EGP) are:

$$\tilde{G}_k(\boldsymbol{t}) = \sum_{i\in[k]} \left[U_i(\boldsymbol{t}) \cdot \exp\ \sum_{l\in<i>} V_l(\boldsymbol{t}) \right] = \sum_{i\in[k]} \exp\left[\boldsymbol{a}^i \boldsymbol{z} + c_i + \sum_{l\in<i>} \exp\left(\boldsymbol{b}^l \boldsymbol{z} + d_l \right) \right].$$

We therefore have a

Convex reformulation of (EGP) in the variables $\boldsymbol{z}$

$$(EGP)_z \quad \inf_{z\in R^m} \tilde{g}_0(\boldsymbol{z}) \ s.t.\ \tilde{g}_k(\boldsymbol{z}) \le 0,\ \forall k \in K,$$

where each problem function

(4.4) $$\tilde{g}_k(\boldsymbol{z}) =: \ln \tilde{G}_k(\boldsymbol{t}) = \ln \left\{ \sum_{i\in[k]} \exp\left[\boldsymbol{a}^i \boldsymbol{z} + c_i + \sum_{l\in<i>} \exp\left(\boldsymbol{b}^l \boldsymbol{z} + d_l \right) \right] \right\},\ k \in \breve{K},$$

is a composite of the convex geometric function *geo* with a mapping $: R^m \to R^{n_k}$, $n_k = |[k]|$ whose components

$$\boldsymbol{a}^i \boldsymbol{z} + c_i + \sum_{l\in<i>} \exp\left(\boldsymbol{b}^l \boldsymbol{z} + d_l \right),\ \forall i \in [k],$$

are clearly convex in $\boldsymbol{z}$,. Hence by proposition 2.3.1 $\tilde{g}_k(\boldsymbol{z})$ is convex.

In terms of the new variables

$$(\boldsymbol{x}, \boldsymbol{\xi}) \in R^n \times R^r, \text{ with } \boldsymbol{x} = \boldsymbol{Az},\ \boldsymbol{\xi} = \boldsymbol{Bz},\ x_i =: \boldsymbol{a}^i \boldsymbol{z},\ i \in I,\ \xi_l =: \boldsymbol{b}^l \boldsymbol{z},\ l \in L,$$

the problem functions $\tilde{g}_k(\boldsymbol{z})$ can be written in form where the variables are separated over different $k \in \breve{K}$:

(4.5) $$\tilde{g}_k(\boldsymbol{z}) = \ln \left\{ \sum_{i\in[k]} \exp\left[x_i + c_i + \sum_{l\in<i>} \exp\left(\xi_l + d_l \right) \right] \right\} =: \tilde{g}_k(\boldsymbol{x}, \boldsymbol{\xi}).$$

GGP formulation of (EGP) in the variables $(\boldsymbol{x},\boldsymbol{\xi})$

$$(EGP)_{x,□} \inf_{(\boldsymbol{x},\xi)\in R^{n+r}} \tilde{g}_0(\boldsymbol{x},\boldsymbol{\xi}) \; s.t. \; \tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi}) \le 0, \; \forall k \in K, \; (\boldsymbol{x},\boldsymbol{\xi}) \in \mathscr{P},$$

where $\mathscr{P}$ is called the *primal space* of this program, and is the column space of the composite exponent matrix □ = □ □ □ □, and the variables in the convex problem functions $\tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi})$ are separated for different $k \in K$.

The notation $\tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi}) = \tilde{g}_k(\boldsymbol{z})$ that we used in the last equation was inadequate, since $(\boldsymbol{x},\xi) \neq \boldsymbol{z}$. This is because we do not yet have a proper name for the function $\ln\left\{\sum_{i=1}^{n} \exp\left[x_i + \sum_{l\in<i>} \exp \xi_l\right]\right\}$ in $(\boldsymbol{x},\boldsymbol{\xi})$ variables, but in next section we will remedy this by defining this function as composite geometric function, and show that it plays a role in CGP similar to that of the geometric function *geo* in GP. Specifically, recall that we can describe the problem functions in $(\text{GP})_{\boldsymbol{x}}$, $\ln\left\{\sum_{i\in[k]} \exp(x_i + c_i)\right\}$ as $geo(\boldsymbol{x}^k + \boldsymbol{c}^k)$, express the conjugate of the Lagrangian of the primal program $(\text{GP})_{\boldsymbol{x}}$ in terms of the conjugate of the function *geo*, and express the recession functions of the problem functions $g_k(\boldsymbol{z})$ in $(\text{GP})_z$ in terms of that of the function *geo*. In the next section we shall do a similar thing for the composite GP.

4.2 COMPOSITE GEOMETRIC FUNCTION

Let us call the function $\ln\left\{\sum_{i=1}^{n}\exp\left[x_i+\sum_{l\in<i>}\exp\xi_l\right]\right\}$ an *exponential geometric function*, and more generally, the function $\ln\left\{\sum_{i=1}^{n}\exp\left[x_i+\sum_{l\in<i>}h_l(\xi_l)\right]\right\}$ a *composite geometric function*, and denote them $\tilde{g}eo(\boldsymbol{x},\boldsymbol{\xi})$, with $(\boldsymbol{x},\boldsymbol{\xi})$ in (R^n,R^r), where for each $l\in L=\{1,\cdots,r\}$, h_l is a given differentiable and strictly convex function$: R\rightarrow R$. This function $\tilde{g}eo$ is clearly differentiable and convex. Each index set $<i>$ has $r_i=|<i>|$ elements and they form a sequential partition of L: $L=<1>\cup<2>\cup\cdots\cup<n>$ such that $<1>$ consists of the first r_1 integers from L, $<2>$ the next r_2 integers, and so forth, $<n>$ the last r_n integers. We also let $I=\{1,\cdots,n\}, I^+:=\left\{i\in I\,|<i>\neq\phi\right\}$. Of course, we have $L=\cup_{i\in I^+}<i>$, $r=\sum_{i\in I^+}r_i$. To simplify our notation, we shall also denote the function $\sum_{l\in<i>}h_l(\xi_l)$, $\forall i\in I^+$ by $H_i(\boldsymbol{\xi}^i)$, where $\boldsymbol{\xi}^i=[\xi_l]_{l\in<i>}\in R^{r_i}$. For $i\in I\setminus I^+$, H_i is regarded as zero. Then we define a mapping $H:R^r\rightarrow R^n$ by $H(\boldsymbol{\xi}):=\left[H_i(\boldsymbol{\xi}^i)\right]_{i\in I}$. We can now give this function a more formal definition.

Definition 4.2.1 *The composite geometric function* $\tilde{g}eo:(R^n,R^r)\rightarrow R$ with each h_l being a given differentiable strictly convex function$: R\rightarrow R$ is defined by

(4.6) $$\tilde{g}eo(\boldsymbol{x},\boldsymbol{\xi}):=geo[\boldsymbol{x}+H(\boldsymbol{\xi})]=\ln\left\{\sum_{i=1}^{n}\exp\left[x_i+\sum_{l\in<i>}h_l(\xi_l)\right]\right\}$$

An *exponential geometric function* is a special case of this function when $h_l(\xi_l)=\exp\xi_l, \forall l\in L$.

The problem functions $\tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi})$ in $(\text{EGP})_{x,\xi}$ (4.5) can now be described as

(4.7) $$\tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi})=\tilde{g}eo(\boldsymbol{x}^k+\boldsymbol{c}^k,\boldsymbol{\xi}^{[k]}+\boldsymbol{d}^{[k]})\text{ with } h_l=\exp, \forall l\in L,$$

where $\boldsymbol{\xi}^{[k]} = [\boldsymbol{\xi}^i]_{i\in[k]} \in R^{r(k)}, r(k) := \sum_{i\in[k]} r_i$ and $\boldsymbol{d}^{[k]}$ is similarly defined.

Proposition 4.2.2 (conjugate of a composite geometric function): *The conjugate* $\tilde{g}eo^* : (R^n, R^r) \to R$ *of the function* $\tilde{g}eo : (R^n, R^r) \to R$ *is given by*

(4.8)
$$\tilde{g}eo^*(\boldsymbol{y},\eta) = \begin{cases} \sum_{i\in I} y_i \ln y_i + \sum_{i\in I^+} \sum_{l\in<i>} y_i h_l^*(\eta_l / y_i), \text{ if } \boldsymbol{y} \in \Delta_n \\ +\infty, \text{ otherwise} \end{cases}$$

under the convention that $0\ln 0 = 0,$ *and* $0h_l^*(\eta_l / 0) = i(\eta_l \mid 0)$, *where* Δ_n *= the unit simplex in* R^n.

Moreover, under the same convention, for $\lambda \geq 0, (\boldsymbol{c},\boldsymbol{d}) \in (R^n, R^r)$ *the conjugate of the function* $\lambda\tilde{g}eo(\boldsymbol{x}+\boldsymbol{c}, \boldsymbol{\xi}+\boldsymbol{d})$ *is given by*

(4.9)
$$\begin{cases} \sum_{i\in I} y_i \ln(y_i / C_i\lambda) + \sum_{i\in I^+} \sum_{l\in<i>} \left[y_i h_l^*(\eta_l / y_i) - d_l\eta_l \right], \text{ if } \boldsymbol{y} \in \lambda\Delta_n \\ +\infty, \text{ otherwise} \end{cases}$$

Note that in both equations the second summation term is a sublinear function in $(\boldsymbol{y},\eta)$, and the first equation is a special case of the second when $\lambda = 1,$ and $(\boldsymbol{c},\boldsymbol{d}) = (\boldsymbol{0},\boldsymbol{0})$.

Proof: The conjugate of $\tilde{g}eo(\boldsymbol{x},\boldsymbol{\xi})$ is:

$$\begin{aligned} \tilde{g}eo^*(\boldsymbol{y},\eta) &= \sup_{(\boldsymbol{x},\xi)} \left\{ (\boldsymbol{x},\boldsymbol{\xi})\cdot(\boldsymbol{y},\eta) - geo[\boldsymbol{x} + H(\boldsymbol{\xi})] \right\} \\ &= \sup_{\xi} \sup_{\boldsymbol{x}} \left\{ [\boldsymbol{x} + H(\boldsymbol{\xi})]\cdot\boldsymbol{y} - geo[\boldsymbol{x}+H(\boldsymbol{\xi})] + \boldsymbol{\xi}\cdot\eta - \boldsymbol{y}\cdot H(\boldsymbol{\xi}) \right\} \\ &= geo^*(\boldsymbol{y}) + \sup_{\xi} \sum_{i\in I^+} \left[\boldsymbol{\xi}^i \cdot \eta^i - y_i H_i(\boldsymbol{\xi}^i) \right] \\ &= \sum_{i\in I} y_i \ln y_i + i(\boldsymbol{y} \mid \Delta_n) + \sum_{i\in I^+} y_i H_i^*(\eta^i / y_i) \\ &= \begin{cases} \sum_{i\in I} y_i \ln y_i + \sum_{i\in I^+} \sum_{l\in<i>} y_i h_l^*(\eta_l / y_i), \text{ if } \boldsymbol{y} \in \Delta_n \\ +\infty, \text{ otherwise} \end{cases} \end{aligned}$$

The conjugate of $\lambda\tilde{g}eo(\boldsymbol{x}+\boldsymbol{c},\boldsymbol{\xi}+\boldsymbol{d})$ is:

$$\lambda \tilde{g}eo^*(\boldsymbol{y}/\lambda, \eta/\lambda) - \boldsymbol{c}\cdot\boldsymbol{y} - \boldsymbol{d}\cdot\eta$$
$$= \begin{cases} \sum_{i\in I} y_i[\ln(y_i/\lambda) - c_i] + \sum_{i\in I^+}\sum_{l\in<i>}\left[y_i h_l^*(\eta_l/y_i) - d_l\eta_l\right], \text{ if } \boldsymbol{y}\in\lambda\Delta_n \\ +\infty, \text{ otherwise} \end{cases}$$
$$= \begin{cases} \sum_{i\in I} y_i \ln(y_i/C_i\lambda) + \sum_{i\in I^+}\sum_{l\in<i>}\left[y_i h_l^*(\eta_l/y_i) - d_l\eta_l\right], \text{ if } \boldsymbol{y}\in\lambda\Delta_n \\ +\infty, \text{ otherwise} \end{cases}$$

■

Corollary 4.2.3 *The conjugate* $\tilde{g}eo^*(\boldsymbol{y},\eta)$ *of the exponential geometric function* $\tilde{g}eo(\boldsymbol{x},\xi) := \ln\left\{\sum_{i=1}^n \exp\left[x_i + \sum_{l\in<i>}\exp\xi_l\right]\right\}$ *is*

$$\begin{cases} \sum_{i\in I} y_i \ln y_i + \sum_{i\in I^+}\sum_{l\in<i>}\eta_l\ln(\eta_l/ey_i), \text{ if } \boldsymbol{y}\in\Delta_n, \eta_l\in y_iR_+, \forall l\in<i>, \forall i\in I^+ \\ +\infty, \text{ otherwise} \end{cases} \tag{4.10}$$

where $0\ln 0 = 0$. *Moreover, for* $\lambda\ge 0, (\boldsymbol{c},\boldsymbol{d})\in(R^n,R^r)$ *the conjugate of the function* $\lambda\tilde{g}eo(\boldsymbol{x}+\boldsymbol{c},\boldsymbol{\xi}+\boldsymbol{d})$ *is*

$$\begin{cases} \sum_{i\in I} y_i \ln(y_i/C_i\lambda) + \sum_{i\in I^+}\sum_{l\in<i>}\eta_l\ln(\eta_l/eD_ly_i), \text{ if } \boldsymbol{y}\in\lambda\Delta_n, \eta_l\in y_iR_+, \forall l\in<i>, \forall i\in I^+ \\ +\infty, \text{ otherwise} \end{cases} \tag{4.11}$$

Proof For $h(\xi) = e^\xi$, $h^*(\eta) = \eta\ln(\eta/e) + i(\eta\,|\,R_+)$, $yh^*(\eta/y) = \eta\ln(\eta/ey) + i(\eta\,|\,yR_+), \forall y\ge 0$, so by (4.8) the conjugate $\tilde{g}eo^*(\boldsymbol{y},\eta)$ of $\tilde{g}eo(\boldsymbol{x},\xi)$ is:

$$\begin{cases} \sum_{i\in I} y_i \ln y_i + \sum_{i\in I^+}\sum_{l\in<i>}\eta_l\ln(\eta_l/ey_i), \text{ if } \boldsymbol{y}\in\Delta_n, \eta_l\in y_iR_+, \forall l\in<i>, \forall i\in I^+ \\ +\infty, \text{ otherwise} \end{cases}$$

And by (4.9) the conjugate of $\lambda\tilde{g}eo(\boldsymbol{x}+\boldsymbol{c},\boldsymbol{\xi}+\boldsymbol{d})$ is:

$$\begin{cases} \sum_{i\in I} y_i \ln(y_i/C_i\lambda) + \sum_{i\in I^+}\sum_{l\in<i>}\eta_l\ln(\eta_l/eD_ly_i), \text{ if } \boldsymbol{y}\in\lambda\Delta_n, \eta_l\in y_iR_+, \forall l\in<i>, \forall i\in I^+ \\ +\infty, \text{ otherwise} \end{cases}$$

■

4.3 DUAL PROGRAM EGD

In the dual program (EGD) for (EGP), there is a dual variable y_i that corresponds to each first tier posynential term $\tilde{U}_i(\boldsymbol{t})$, just as in the posynomial GP case. Additionally, there is also a second tier dual variable η_l that corresponds to each second tier posynomial term $V_l(\boldsymbol{t})$. We will denote $\boldsymbol{y} = [y_i]_n^1 \in R^n$, $\boldsymbol{\eta} = [\eta_l]_r^1 \in R^r$. The dual program is given by:

$$\text{(4.12)} \qquad \text{(EGD)} \quad \begin{cases} \sup \tilde{V}(\boldsymbol{y},\eta) := \exp[-\tilde{l}_\lambda^*(\boldsymbol{y},\eta)] \\ s.t.\ (\boldsymbol{y},\eta) \in \mathcal{D} = \text{dual space of (EGD)} \end{cases}$$

where $\tilde{l}_\lambda^*(\boldsymbol{y},\boldsymbol{\eta})$ is the conjugate of the Lagrangian $\tilde{l}_\lambda(\boldsymbol{x},\xi) = \tilde{l}(\boldsymbol{x},\boldsymbol{\xi};\lambda)$ of $(\text{EGP})_{\boldsymbol{x},\xi}$ and $\mathcal{D} := \mathcal{P}^\perp =$ null space of M^T, that is, $(\boldsymbol{y},\eta)$ must satisfy $A^T\boldsymbol{y} + B^T\eta = \boldsymbol{0}$. In vector form, they can be expressed as $\boldsymbol{a}_j \cdot \boldsymbol{y} + \boldsymbol{b}_j \cdot \eta = 0,\ \forall j \in J$, where $\boldsymbol{a}_j, \boldsymbol{b}_j$ are the j^{th} columns of the matrix A and B, respectively. In scalar form, this means: $\sum_{i=1}^{n} a_{ij} y_i + \sum_{l=1}^{r} b_{lj}\eta_l = 0,\ \ \forall j \in J$ (orthogonality conditions)

Proposition 4.3.1 *The dual objective of* (EGD) *is*

$$\text{(4.13)} \qquad \begin{aligned} &\tilde{V}(\boldsymbol{y},\eta) = V(\boldsymbol{y}) \cdot \prod_{i \in I^+} \prod_{l \in <i>} \left(\frac{D_l y_i}{\eta_l} \right)^{\eta_l} \cdot e^{\beta}, \\ &\text{if } \boldsymbol{y} \geq \boldsymbol{0}, \sum_{i \in [k]} y_i = \lambda_k,\ \forall k \in \breve{K},\ \ \lambda_0 = 1,\ \eta \geq \boldsymbol{0}, \text{ and satisfy} \\ &(*) \qquad y_i = 0 \Rightarrow \eta_l = 0,\ \forall l \in <i>,\ \ \forall i \in I^+ \\ &\text{otherwise, its value is 0.} \end{aligned}$$

where $V(\boldsymbol{y}) = \prod_{k \in \breve{K}} \prod_{i \in [k]} \left(\frac{C_i \lambda_k}{y_i} \right)^{y_i}, \beta = \sum_{l \in L} \eta_l$.

Proof: By(4.7) the Lagrangian of $(\text{EGP})_{x,\xi}$ is:

$$\tilde{l}(\boldsymbol{x},\boldsymbol{\xi};\lambda) = \sum_{k\in\breve{K}} \lambda_k \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}), \text{ with } \lambda_0 = 1$$

By (4.11) its conjugate is

$$\begin{aligned}
\tilde{l}_\lambda^*(\boldsymbol{y},\boldsymbol{\eta}) &= \sum_{k\in\breve{K}} \Big[\sum_{i\in[k]} y_i \ln(y_i / C_i\lambda_k) + \sum_{i\in[k]^+} \sum_{l\in<i>} \eta_l \ln(\eta_l / eD_l y_i) \Big] \\
&= \sum_{k\in\breve{K}} \sum_{i\in[k]} y_i \ln(y_i / C_i\lambda_k) + \sum_{i\in I^+} \sum_{l\in<i>} \eta_l \ln(\eta_l / eD_l y_i),
\end{aligned}$$

if $\boldsymbol{y} \ge \boldsymbol{0}, \sum_{i\in[k]} y_i = \lambda_k, \ \forall k \in \breve{K}, \ \ \lambda_0 = 1, \ \eta \ge \boldsymbol{0}$

and satisfy (*) $\ y_i = 0 \Rightarrow \eta_l = 0, \ \forall l \in <i>, \ \ \forall i \in I^+ \ \ (\because \eta_l \in y_i R_+)$

otherwise, its value is ∞.

Therefore, the dual objective of (EGD) is

$$\begin{aligned}
\tilde{V}(\boldsymbol{y},\eta) = \exp[-\tilde{l}_\lambda^*(\boldsymbol{y},\eta)] &= \prod_{k\in\breve{K}} \prod_{i\in[k]} \left(\frac{C_i\lambda_k}{y_i} \right)^{y_i} \cdot \exp\left[-\sum_{i\in I^+} \sum_{l\in<i>} \eta_l \ln(\eta_l / eD_l y_i) \right] \\
&= V(\boldsymbol{y}) \cdot \prod_{i\in I^+} \prod_{l\in<i>} \left(\frac{D_l y_i}{\eta_l} \right)^{\eta_l} \cdot e^\beta, \quad \because \beta = \sum_{l\in L} \eta_l
\end{aligned}$$

if $\boldsymbol{y} \ge \boldsymbol{0}, \sum_{i\in[k]} y_i = \lambda_k, \ \forall k \in \breve{K}, \ \ \lambda_0 = 1, \ \eta \ge \boldsymbol{0},$ and satisfy

(*) $\qquad y_i = 0 \Rightarrow \eta_l = 0, \ \forall l \in <i>, \ \ \forall i \in I^+$

otherwise, its value is 0.

■

Thus the dual program (EGD) can be elaborated as

Dual EGP problem (Under the convention that $0^0 = 1$)

$$
(4.14) \qquad (EGD) \begin{cases} \sup\limits_{(\mathbf{y},\boldsymbol{\eta})\in(R^n,R^r)} \tilde{V}(\mathbf{y},\eta) =: \prod\limits_{k\in K}\prod\limits_{i\in[k]} \left(\frac{C_i\lambda_k}{y_i}\right)^{y_i} \cdot \prod\limits_{i\in I^+}\prod\limits_{l\in<i>} \left(\frac{D_l y_i}{\eta_l}\right)^{\eta_l} \cdot e^{\beta} \\ s.t. \qquad \lambda_0 = 1, \qquad\qquad \text{(Normality Condition)} \\ \sum\limits_{i=1}^{n} a_{ij} y_i + \sum\limits_{l=1}^{r} b_{lj}\eta_l = 0, \ j \in J \quad \text{(Orthogonality Conditions)} \\ \qquad y_i \ge 0, \ \forall i \in I, \ \eta_l \ge 0, \ \forall l \in L, \\ (*) \qquad y_i = 0 \Rightarrow \eta_l = 0, \ \forall l \in <i>, \ \ \forall i \in I^+ \\ \text{where} \quad \lambda_k := \sum\limits_{i\in[k]} y_i, \ \forall k \in K, \ \ \beta := \sum\limits_{l\in L} \eta_l \end{cases}
$$

Alternatively put:

$$
\tilde{V}(\mathbf{y},\eta) = \prod_{1}^{n} \left(\frac{C_i}{y_i}\right)^{y_i} \prod_{1}^{p} \lambda_k^{\lambda_k} \cdot \prod_{1}^{r} \left(\frac{D_l}{\eta_l}\right)^{\eta_l} \prod_{i\in I^+} y_i^{\beta_i} \cdot e^{\beta}, \text{ where } \beta_i := \sum_{l\in<i>} \eta_l, \ i \in I^+
$$

Of course, $\beta = \sum_{i\in I^+} \beta_i$ and at any feasible point $(\boldsymbol{y},\eta)$, $\beta_i = 0 \Leftrightarrow \eta_l = 0, \ \forall l \in <i>$. The (*) constraints are sublinear, hence potentially nonlinear, but they will be fulfilled if a pair of primal and dual solutions $(\boldsymbol{t};\boldsymbol{y},\eta)$ satisfy the extremality condition in(4.20) (b). Note that the second factor of the dual objective $\tilde{V}(y,\eta)$ in (4.14) is exactly similar to the first factor in that η_l is to y_i in the second tier just as y_i is to λ_k in the first tier, and that β in the last exponential factor is the total sum of all the second tier dual variables. For the first tier dual variables, we could similarly define $\alpha := \sum\limits_{i\in I} y_i = 1 + \sum\limits_{k\in K} \lambda_k$. Thus if each of the cost coefficients C_i and D_l is discounted by a factor of θ, $0<\theta<1$, then the dual function is discounted by a factor of $\theta^{\alpha+\beta}$, (cf. Duffin et.al [1967, p.183]). Note that

$$
(4.15) \qquad (\boldsymbol{x},\boldsymbol{\xi})\cdot(\boldsymbol{y},\eta) = \boldsymbol{x}\cdot\boldsymbol{y} + \boldsymbol{\xi}\cdot\eta = 0, \ \forall(\boldsymbol{x},\boldsymbol{\xi}) \in \mathcal{P}, \forall(\boldsymbol{y},\eta) \in \mathcal{D}
$$

Concavity of the log-dual function $\tilde{v}(\boldsymbol{y},\eta) := \ln \tilde{V}(\boldsymbol{y},\eta)$ in any (CGD) program

Since the Lagrangian $\tilde{l}_\lambda(\boldsymbol{x},\boldsymbol{\xi})$ is finite convex in $(\boldsymbol{x},\boldsymbol{\xi})$, $\forall \lambda \in R_+^p$, it is *lsc* proper convex and so is its dual $\tilde{l}_\lambda^*(\boldsymbol{y},\boldsymbol{\eta})$. So $\tilde{v}(\boldsymbol{y},\eta) = \ln \tilde{V}(\boldsymbol{y},\eta) = -\tilde{l}_\lambda^*(\boldsymbol{y},\eta)$ is *usc* proper concave in $(\boldsymbol{y},\eta)$.

In the (EGD) case, the log-dual takes the form:

(4.16)
$$\begin{aligned} \tilde{v}(\boldsymbol{y},\eta) &= \sum_{k\in K}\sum_{i\in[k]} y_i \ln\left(\frac{C_i\lambda_k}{y_i}\right) + \sum_{i\in I^+}\sum_{l\in<i>} \eta_l \ln\left(\frac{D_l y_i}{\eta_l}\right) + \beta \\ &= \sum_{i\in I} y_i(c_i - \ln y_i) + \sum_{k\in K}\lambda_k \ln\lambda_k + \sum_{l\in L}\eta_l(d_l - \ln\eta_l) + \sum_{i\in I^+}\beta_i \ln y_i + \beta \end{aligned}$$

So the dual program (EGD) is equivalent to a convex program with (m+1) linear equality constraints, a set of nonlinear constraints (*) and non-negativity restrictions on its (n+r) dual variables y_i and η_l. Its degree of difficulty d, defined as the dimension of its dual flat, equals (n+r−m−1). This is r greater than that of its underlying GD, due to the fact that its primal program contains r more posynomial terms in the second tier and is therefore more complicated than its underlying GP.

Observe also that in the dual program (EGD) the cost coefficients C_i and D_l only appear in its objective and the technological coefficients a_{ij} and b_{lj} only in the constraints; whereas in the primal program (EGP) they are scattered all over the terms $\tilde{U}_i(t)$ of the functions $\tilde{G}_k(t)$. The log-dual program (EGD) possesses 3 exploitable structures: *concavity*, *partial-separability*, and *linearity* that are originally hidden in the (EGP) model. Obviously, these features would make solving the primal through its dual an attractive approach, provided that the two programs have identical optimal values and conversion of optimum solutions from dual to primal is computationally economical. The duality theory of EGP attempts to confirm the above fact under mild conditions.

It is also clear that when $r = 0 \Leftrightarrow L = \phi$, the matrix B and the vector η are not present, and the above program readily reduces to the program dual to the underlying (GP):

$$
\text{(GD)} \begin{cases} \sup\limits_{\mathbf{y}\in R_+^n} V(\boldsymbol{y}) = \prod\limits_{k\in \breve{K}} \prod\limits_{i\in[k]} \left(\dfrac{C_i \lambda_k}{y_i} \right)^{y_i} = \prod\limits_{1}^{n} \left(\dfrac{C_i}{y_i} \right)^{y_i} \prod\limits_{1}^{p} \lambda_k^{\lambda_k} \\ s.t. \qquad \lambda_0 = 1, \text{ and } \sum\limits_{i=1}^{n} a_{ij} y_i = 0, \ \forall j \in J \\ \qquad \text{where} \quad \lambda_k := \sum\limits_{i\in[k]} y_i, \quad \forall k \in \breve{K} \end{cases}
$$

An observation on the relationships between inf GP, inf EGP, sup GD, and sup EGD.

Denote $\tilde{T}$ and T the feasible regions of programs (EGD) and (GD), respectively. Observe that

$$\boldsymbol{y} \in T \Rightarrow (\boldsymbol{y}, \boldsymbol{0}) \in \tilde{T}, \text{ and } V(\boldsymbol{y}) = \tilde{V}(\boldsymbol{y}, \boldsymbol{0}) \le \sup_{\tilde{T}} \tilde{V}(\boldsymbol{y}, \eta) = \sup EGD.$$

Thus one has $\sup GD \le \sup EGD$ in addition to the previous relationship $\inf GP \le \inf EGP$.

4.4 MAIN LEMMA OF EGP

Before proving this lemma, we need an additional geometric inequality which is related to the exponential function.

Lemma 4.4.1 *Under the convention that* $0^0 = 1$, *the inequality*

(4.17) $e^x \ge e^\eta (x/\eta)^\eta$ *holds* $\forall x \ge 0, \ \eta \ge 0$, *with equality holding if and only if* $x = \eta \ge 0$.

Proof: The inequality $e^x \ge ex, \forall x$ holds, with equality holding if, and only if, x=1. Hence

$e^{x/\eta} \ge ex/\eta > 0, \ \forall x > 0, \ \eta > 0$ holds, with equality holding if, and only if, $x = \eta > 0$.

Raising both sides of the above inequality to the η^{th} power yields

$$e^x \geq e^\eta (x/\eta)^\eta, \ \forall x > 0, \ \eta > 0, \text{ with equality holding if and only if } x = \eta > 0.$$

For the remaining cases:

$x > \eta = 0$ or $\eta > x = 0$, the inequality is strict,
$x = \eta = 0$, the inequality is an equality: 1=1. ■

The Fenchel's inequality for the function e^ξ and its conjugate $\eta \ln \eta - \eta$ is:

(4.18) $\quad e^\xi + \eta \ln \eta - \eta \geq \xi\eta$, valid for $\xi \in R$ and $\eta \geq 0$, with equality holding iff $\eta = e^\xi$.

After a change of variable: $x = e^\xi$, this becomes

$$x + \eta \ln \eta - \eta \geq \eta \ln x \Leftrightarrow x \geq \eta + \eta \ln(x/\eta) \Leftrightarrow e^x \geq e^\eta (x/\eta)^\eta, \text{ valid for } x > 0 \text{ and } \eta \geq 0,$$

where equality holds iff $(e^\xi =)x = \eta > 0$. Since in the proof below, the x variable stands for $y_i V_l$, which could be zero when the dual variable $y_i = 0$, we need to extend the inequality a little as in Lemma 4.4.1.

Lemma 4.4.2 (Main Lemma of EGP) *If $\boldsymbol{t}$ is feasible for primal program* (EGP) *and* $(\boldsymbol{y}, \eta)$ *is feasible for dual program* (EGD)*, then* $\tilde{G}_0(\boldsymbol{t}) \geq \tilde{V}(\boldsymbol{y}, \eta)$

Moreover, under the same conditions, $\tilde{G}_0(\boldsymbol{t}) = \tilde{V}(\boldsymbol{y}, \eta)$ *if, and only if, one of the following two sets of equivalent conditions holds:*

$$\text{(4.19)} \quad (a) \begin{cases} \tilde{G}_k(\boldsymbol{t})^{\lambda_k} = 1, \ k \in K, \\ y_i \tilde{G}_k(\boldsymbol{t}) = \lambda_k \tilde{U}_i(\boldsymbol{t}), \ i \in [k], \ k \in \breve{K} \end{cases} \qquad \text{(Extremality condition 1)}$$

$$(b) \quad \eta_l = y_i V_l(\boldsymbol{t}), \ l \in \langle i \rangle, \ i \in I^+$$

$$\text{(4.20)} \quad (c) \quad y_i = \begin{cases} \tilde{U}_i(\boldsymbol{t}) \big/ \tilde{G}_0(\boldsymbol{t}), & i \in [0] \\ \lambda_k \tilde{U}_i(\boldsymbol{t}), & i \in [k], \ k \in K \end{cases} \qquad \text{(Extremality condition 2)}$$

$$(d) \quad \eta_l = y_i V_l(\boldsymbol{t}), \ l \in \langle i \rangle, \ i \in I^+$$

in which case $\boldsymbol{t}$ *is optimal for primal program* (EGP) *and* $(\boldsymbol{y},\eta)$ *is optimal for dual program* (EGD). (Note the similarity in the above extremality conditions 2 between (d) and the second set of equations in (c). Recall from Lemma 2.1.0 that the above conditions (a) and (c) are equivalent.)

Proof: Let $(\boldsymbol{x},\boldsymbol{\xi})=(A\boldsymbol{z},B\boldsymbol{z})$, where $z_j=\ln t_j, \forall j\in J$. Then $(\boldsymbol{x},\boldsymbol{\xi})\in\mathscr{P}$ is feasible for program $(EPP)_{x,\xi}$ and hence $(\boldsymbol{x},\boldsymbol{\xi})\cdot(\boldsymbol{y},\eta)=0$. Applying geometric inequality (2.8) to each problem function $\tilde{G}_k(\boldsymbol{t})$ in (*EGP*) (4.1) yields

$$\begin{aligned}\tilde{G}_0 &\geq \tilde{G}_0\cdot\prod_{k\in K}\tilde{G}_k^{\lambda_k}\geq\prod_{k\in K}\prod_{i\in[k]}\left(\tilde{U}_i\lambda_k/y_i\right)^{y_i}=\prod_{k\in K}\prod_{i\in[k]}\left(U_i\lambda_k/y_i\right)^{y_i}\cdot\prod_{i\in I}e^{y_iE_i}\quad\left(\because\tilde{U}_i=U_ie^{E_i}\right)\\&=\prod_{k\in K}\prod_{i\in[k]}\left(C_i\lambda_k/y_i\right)^{y_i}\cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}}\cdot\prod_{i\in I^+}\prod_{l\in\langle i\rangle}e^{y_iV_l}\qquad\left(\because U_i=C_ie^{x_i},\ E_i=\sum_{l\in\langle i\rangle}V_l\right)\end{aligned}\tag{4.21}$$

So far, we have employed the geometric inequality as before to convert each first tier posynential term $\tilde{U}_i(\boldsymbol{t})$ to a corresponding dual variable y_i, but the last expression still involves $\boldsymbol{t}$ through the second tier posynomial terms $V_l(\boldsymbol{t})$. If one simply uses this as the dual objective function, he will get a mixed dual program, and this is the same reason why prior attempts on deriving useful duality theories for (TGP) by [Lidor and Wilde, 1978] and for (QPL) by [Hough and Goforth, 1981a, 1981b], respectively, are not successful. One need to use a second geometric inequality to convert the remaining second tier terms $V_l(\boldsymbol{t})$ to corresponding dual variables η_l to derive a pure dual program (EPD) and to establish the main lemma. To this end, we apply the inequality (4.17) (setting $x=y_iV_l$) to the last factor in the expression in (4.21) and continue with a further lower bound:

$$\geq \prod_{k\in K}\prod_{i\in[k]}(C_i\lambda_k/y_i)^{y_i}\cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}}\cdot\prod_{i\in I^+}\prod_{l\in<i>}\left[e^{\eta_l}(y_iV_l/\eta_l)^{\eta_l}\right]$$

$$= V(\boldsymbol{y})\cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}}\cdot\prod_{i\in I^+}\prod_{l\in<i>}(D_ly_i/\eta_l)^{\eta_l}\cdot e^{\xi\cdot\eta}\cdot e^{\beta}\qquad\left(\because V_l=D_le^{\xi_l}\right)$$

$$= V(\boldsymbol{y})\cdot\prod_{i\in I^+}\prod_{l\in<i>}(D_ly_i/\eta_l)^{\eta_l}\cdot e^{\beta}\qquad\left(\because \boldsymbol{x}\cdot\boldsymbol{y}+\boldsymbol{\xi}\cdot\eta=0\right)$$

$$= \tilde{V}(\boldsymbol{y},\eta)$$

Therefore, $\tilde{G}_0(\boldsymbol{t})\geq\tilde{V}(\boldsymbol{y},\eta)$ and equality holds iff the extremality conditions (4.19) or (4.20) are satisfied. ■

Instead of the above proof based on (4.17), an alternative approach based on conjugate function theory is possible.

Alternative proof of the Main Lemma

First, note that $y_iV_l=y_ie^{\xi_l+d_l}$. Then the generalized geometric inequality (2.34) with $y\geq 0,\ h=\exp=h',\ h^*(\eta)=\eta\ln\eta-\eta+i(\eta\,|\,R_+)$ is

(4.22) $$ye^{\xi+d}+\eta\ln(\eta/y)-\eta\geq(\xi+d)\eta,\ \text{valid for } \xi\in R \text{ and } y,\eta\geq 0,$$

with equality holding iff $\eta=ye^{\xi+d}$. Applying this inequality, instead of (4.17), to the last factor in the expression in (4.21) and continue with a further lower bound:

$$= V(\boldsymbol{y})\cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}}\cdot\prod_{i\in I^+}\prod_{l\in<i>}\exp\left[y_ie^{\xi_l+d_l}\right]$$

$$\geq V(\boldsymbol{y})\cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}}\cdot\prod_{i\in I^+}\prod_{l\in<i>}\exp\left[\xi_l\,\eta_l+(d_l+1)\eta_l+\eta_l\ln(y_i/\eta_l)\right]$$

$$= V(\boldsymbol{y})\cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}}\cdot e^{\xi\cdot\eta}\cdot\prod_{i\in I^+}\prod_{l\in<i>}\left[(D_le)^{\eta_l}(y_i/\eta_l)^{\eta_l}\right],\ \because e^{d_l}=D_l\ ,$$

$$= V(\boldsymbol{y})\cdot\prod_{i\in I^+}\prod_{l\in<i>}(D_ly_i/\eta_l)^{\eta_l}\cdot e^{\beta}\qquad\because \boldsymbol{x}\cdot\boldsymbol{y}+\boldsymbol{\xi}\cdot\eta=0$$

$$= \tilde{V}(\boldsymbol{y},\eta)$$

and conditions for equality are $\eta_l=y_ie^{\xi_l+d_l}=y_iV_l(\boldsymbol{t}),\forall l\in<i>,\forall i\in I^+$, which constitute condition (b) in (4.20). ■

This main lemma clearly extends the earlier main lemma of GP.

Corollary 4.4.3 (Weak Duality Theorem for (EGP))

1) *Always,* $\infty \geq \inf EGP \geq \sup EGD \geq 0$

2) *When both* (*EGP) and* (*EGD*) *are feasible,* $\infty > \inf EGP \geq \sup EGD > 0$.

Corollary 4.4.4 *Under mild conditions one has*

$$\inf EGP - \inf GP = \sup EGD - \sup GD \geq 0$$

Proof By the main lemmas for GP and for EGP, and by our previous observations, we have

$$\sup GD \leq \inf GP \leq \inf EGP, \text{ and } \sup GD \leq \sup EGD \leq \inf EGP.$$

The duality theorems show that the first inequality in the first expression and the second inequality in the second expression above are actually equalities under mild conditions. So the conclusion follows. ■

4.5 FIRST AND SECOND DUALITY THEOREMS OF EGP

Superconsistency for an exponential geometric program (EGP) as well as for any other composite geometric programs (CGP) shall be the same: $\exists \boldsymbol{t} > \boldsymbol{0} \ s.t. \ \tilde{G}_k(\boldsymbol{t}) < 1, \ \forall k \in K.$ In the convex formulation (EGP)$_z$: $\inf_{z \in R^m} \tilde{g}_0(\boldsymbol{z}) \ s.t. \ \tilde{g}_k(\boldsymbol{z}) \leq 0, \ \forall k \in K$, each problem function

$$\tilde{g}_k(\boldsymbol{z}) =: \ln \tilde{G}_k(\boldsymbol{t}) = \ln \left\{ \sum\nolimits_{i \in [k]} \exp \left[\boldsymbol{a}^i \boldsymbol{z} + c_i + \sum\nolimits_{l \in <i>} \exp\left(\boldsymbol{b}^l \boldsymbol{z} + d_l \right) \right] \right\}, \ k \in K,$$

has partial derivatives:

$$\frac{\partial \tilde{g}_k(\boldsymbol{z})}{\partial z_j} = \frac{1}{\tilde{G}_k(\boldsymbol{t})} \sum_{i \in [k]} \tilde{U}_i(\boldsymbol{t}) \left[a_{ij} + \sum_{l \in <i>} V_l(\boldsymbol{t}) b_{lj} \right], \ \forall j \in J \tag{4.23}$$

The Lagrangian for this differentiable convex program $(\text{EGP})_z$ is

$$\tilde{l}(\boldsymbol{z},\lambda) := \tilde{g}_0(\boldsymbol{z}) + \sum_{k\in K} \lambda_k \tilde{g}_k(\boldsymbol{z}),\ \forall \lambda \in R_+^p, \boldsymbol{z} \in R^m.$$

Theorem 4.5.1 (First Duality Theorem of EGP) *Suppose that primal program* (EGP) *is superconsistent. Then the following 3 conditions are equivalent*:

1) ***t'*** *is a minimal solution to* (EGP).

2) *There exists a vector* $\lambda' \in R_+^p$ *for* $\boldsymbol{z}$, $\boldsymbol{z}'$ (*where* $\boldsymbol{z}' = \ln \boldsymbol{t}'$) *such that* $(\boldsymbol{z}', \lambda')$ *forms a saddle point of* $\tilde{l}(\boldsymbol{z},\lambda)$.

3) *There exists a vector* $\lambda' \in R_+^p$ for $\boldsymbol{z}'$ (*where* $\boldsymbol{z}' = \boldsymbol{z}' = \ln \boldsymbol{t}'$) *such that* $(\boldsymbol{z}', \lambda')$ *satisfies the KKT conditions for* $(\text{EGP})_z$:

$$(a)\ \lambda_k' \geq 0,\ \tilde{g}_k(\boldsymbol{z}') \leq 0,\ \lambda_k' \tilde{g}_k(\boldsymbol{z}') = 0, \forall k \in K$$
$$(b)\ \nabla_z \tilde{l}(\boldsymbol{z}', \lambda') = \sum_{k\in \bar{K}} \lambda_k' \nabla \tilde{g}_k(\boldsymbol{z}') = \boldsymbol{0},\ where\ \lambda_0' = 1$$

in which case the set of all such vectors $\boldsymbol{\lambda}'$ *is a non-empty compact convex subset of* R_+^p, *and the dual program* (EGD) *also has a maximum solution* $(\boldsymbol{y}', \eta')$ *such that*

$$\min(\text{EGP}) = \tilde{G}_0(\boldsymbol{t}') = \tilde{V}(\boldsymbol{y}', \eta') = \max(\text{EGD}) \qquad \text{(Perfect duality)}$$

Proof By assumption, the differentiable convex program $(\text{EGP})_z$ satisfies the Slater CQ. Hence by KKT Theorem (Theorem 2.4.4), the above conditions 1) through 3) are equivalent and in which case, by Theorem 2.4.3 the set of all such vectors $\boldsymbol{\lambda}'$ is a non-empty compact convex subset of R_+^p, and conditions (a) and (b) in 3) hold true. We note that

$$\text{Condition (b)} \quad \Rightarrow \sum_{k\in \bar{K}} \lambda_k' \frac{\partial \tilde{g}_k(\boldsymbol{z}')}{\partial z_j} = 0,\ \forall j \in J$$

$$\Rightarrow \sum_{k \in \breve{K}} \frac{\lambda_k'}{\tilde{G}_k(\boldsymbol{t}')} \sum_{i \in [k]} \tilde{U}_i(\boldsymbol{t}) \left[a_{ij} + \sum_{l \in <i>} V_l(\boldsymbol{t}) b_{lj} \right] = 0,\ \forall j \in J \text{, by (4.23).}$$

Now we define $y_i' = \lambda_k' \tilde{U}_i(\boldsymbol{t}') / \tilde{G}_k(\boldsymbol{t}') \geq 0, \forall i \in [k], \forall k \in \breve{K}$; and $\eta_l' = y_i' V_l(\boldsymbol{t}') \geq 0,\ l \in <i>,\ i \in I^+$,

then we have condition (*) in (EGD) being satisfied and

$$\sum_{k \in \breve{K}} \sum_{i \in [k]} \left[a_{ij} y_i' + \sum_{l \in <i>} b_{lj} \eta_l' \right] = \sum_{i=1}^{n} a_{ij} y_i' + \sum_{l=1}^{r} b_{lj} \eta_l' = 0,\ \ j \in J$$

This is the orthogonality condition for (EGD).

Condition (a) $\Rightarrow G_k(\boldsymbol{t}')^{\lambda_k'} = 1,\ \forall k \in K$, this is the first set of conditions in part (a) of the extremality conditions 1 (4.19) for main lemma of (EGP) and the remaining conditions there are trivially satisfied by our definition of $(\boldsymbol{y}', \eta')$. Lastly, we see that $\sum_{i \in [k]} y_i' = \sum_{i \in [k]} [\lambda_k' \tilde{U}_i(\boldsymbol{t}') / \tilde{G}_k(\boldsymbol{t}')] =$

$\lambda_k',\ \forall k \in \breve{K}$. So we have defined a dual feasible solution $(\boldsymbol{y}', \eta')$ which together with t' satisfies the extremality conditions 1 (4.19). Hence by main lemma of (EGP), this $(\boldsymbol{y}', \eta')$ is a maximum solution satisfying

$$\min(\text{EGP}) = \tilde{G}_0(\boldsymbol{t}') = \tilde{V}(\mathbf{y}', \eta') = \max(\text{EGD}) \qquad \blacksquare$$

Next, we have a result that is very similar to Proposition 2.5.6.

Proposition 4.5.2 *The recession function of the composite geometric function* $\tilde{geo}(\boldsymbol{x}, \boldsymbol{\xi}) = geo[\boldsymbol{x} + H(\boldsymbol{\xi})]$ *is given by*

(4.24) $$\tilde{geo}^{\infty}(\boldsymbol{x}, \boldsymbol{\xi}) = \max_{i \in I} \left[x_i + \sum\nolimits_{l \in <i>} h_l^{\infty}(\xi_l) \right] = geo^{\infty}[\boldsymbol{x} + H^{\infty}(\boldsymbol{\xi})]$$

where $H^{\infty}(\boldsymbol{\xi}) = \left[H_i^{\infty}(\boldsymbol{\xi}^i) \right]_{i \in I}$, and $H_i^{\infty}(\boldsymbol{\xi}^i) = \sum\nolimits_{l \in <i>} h_l^{\infty}(\xi_l)$.

Proof Since $\tilde{geo}(\boldsymbol{x}, \boldsymbol{\xi}) = geo[\boldsymbol{x} + H(\boldsymbol{\xi})]$, we have by (3.16) the approximation:

(4.25) $$0 \leq geo[\boldsymbol{x} + H(\boldsymbol{\xi})] - \max[\boldsymbol{x} + H(\boldsymbol{\xi})] \leq \ln n,\ \forall (\boldsymbol{x}, \boldsymbol{\xi})$$

Homogenizing this, we obtain

$$0 \le \left(\frac{1}{s}\right) geo[s\boldsymbol{x} + H(s\boldsymbol{\xi})] - \max[\boldsymbol{x} + \left(\frac{1}{s}\right) H(s\boldsymbol{\xi})] \le \left(\frac{1}{s}\right) \ln n, \ \forall(\boldsymbol{x},\boldsymbol{\xi}), \forall s > 0$$

Let $s \to \infty$, we get a uniform convergence:

$$\begin{aligned} \tilde{g}eo^{\infty}(\boldsymbol{x},\boldsymbol{\xi}) &= \lim_{s\to\infty} \max[\boldsymbol{x} + \left(\frac{1}{s}\right) H(s\boldsymbol{\xi})] = \max[\boldsymbol{x} + \lim_{s\to\infty}\left(\frac{1}{s}\right) H(s\boldsymbol{\xi})] \\ &= \max_{i\in I}\left[x_i + \lim_{s\to\infty}\left(\frac{1}{s}\right)\sum\nolimits_{l\in<i>} h_l(s\xi_l)\right] = \max_{i\in I}\left[x_i + \sum\nolimits_{l\in<i>} h_l^{\infty}(\xi_l)\right] \\ &= \max_{i\in I}\left[x_i + H_i^{\infty}(\boldsymbol{\xi}^i)\right] = \max[\boldsymbol{x} + H^{\infty}(\boldsymbol{\xi})] = geo^{\infty}[\boldsymbol{x} + H^{\infty}(\boldsymbol{\xi})] \end{aligned}$$

■

Corollary 4.5.3 *The recession function of the exponential geometric function is*

(4.26) $$\tilde{g}eo^{\infty}(\boldsymbol{x},\boldsymbol{\xi}) = \max(\boldsymbol{x}) + i(\boldsymbol{\xi} \mid R_-^r)$$

Proof: Since the recession function of $h(\xi) = e^{\xi}$ is $h^{\infty}(\xi) = \lim_{s\to\infty} \frac{1}{s} e^{s\xi} = \begin{cases} 0, & \xi \le 0 \\ +\infty, & \xi > 0 \end{cases} = i(\xi \mid R_-)$

we have by (4.24) that

$$\tilde{g}eo^{\infty}(\boldsymbol{x},\boldsymbol{\xi}) = \max_{i\in I}\left[x_i + \sum\nolimits_{l\in<i>} i(\xi_l \mid R_-)\right] = \max(\boldsymbol{x}) + i(\boldsymbol{\xi} \mid R_-^r) \qquad ■$$

In (EGP)$_z$ the problem functions given by (4.4) and (4.7) are $\tilde{g}_k(\boldsymbol{z}) = \tilde{g}eo(\boldsymbol{A}^k\boldsymbol{z} + \boldsymbol{c}^k, \boldsymbol{B}^{[k]}\boldsymbol{z} + \boldsymbol{d}^{[k]})$, with each $h_l = \exp, \forall l \in L$, where $\boldsymbol{B}^{[k]}$ consists of those rows b^l of $\boldsymbol{B}$ for which $l \in <i>$, $i \in [k]$, and $\boldsymbol{d}^{[k]}$ is similarly defined. We then have by proposition 2.5.6 and corollary 4.5.3

$$\tilde{g}_k^{\infty}(\boldsymbol{z}) = \tilde{g}eo^{\infty}(\boldsymbol{A}^k\boldsymbol{z}, \boldsymbol{B}^{[k]}\boldsymbol{z}) = \max(\boldsymbol{A}^k\boldsymbol{z}) + i(\boldsymbol{B}^{[k]}\boldsymbol{z} \mid R_-^{r(k)}), \text{ where } r(k) := \sum\nolimits_{i\in[k]} r_i$$

A theorem of the alternatives from linear programming duality theory

Of the following two linear systems exactly one has a solution (*where* $M = \begin{bmatrix} A \\ B \end{bmatrix} : (n+r)\times m$):

(I) *Find* $\boldsymbol{z}$ *with* $0 \neq \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \boldsymbol{z} \leq 0$ (II) *Find* $\begin{pmatrix} \boldsymbol{y} \\ \eta \end{pmatrix} > 0$ with $A^T \boldsymbol{y} + B^T \eta = \boldsymbol{0}$

We say that program (EGD) is *canonical* if system (II) has a solution when M is the program's composite exponent matrix.

Theorem 4.5.4 (Second Duality Theorem of EGP) *Suppose that primal program* (EGP) *is consistent. Then the minimum set of program* $(\text{EGP})_z$ *is non-empty and bounded iff dual program* (EGD) *is canonical*, *in which case program* (EGP) *has a minimum solution* $\boldsymbol{t}'$.

Proof: By assumption, program $(\text{EGP})_z$ is also consistent. The system

$$
\begin{aligned}
\tilde{g}_k^{\infty}(\boldsymbol{z}) \leq 0,\ k \in \breve{K} &\Leftrightarrow \max(\boldsymbol{A}^k \boldsymbol{z}) + i(\boldsymbol{B}^{[k]} \boldsymbol{z} \mid R_-^{r(k)}) \leq 0,\ k \in \breve{K} \\
&\Leftrightarrow \boldsymbol{A}^k \boldsymbol{z} \leq \boldsymbol{0},\ \boldsymbol{B}^{[k]} \boldsymbol{z} \leq \boldsymbol{0},\ k \in \breve{K} \\
&\Leftrightarrow \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \boldsymbol{z} \leq \boldsymbol{0}
\end{aligned},
$$

Since M is of full column rank, it is one-to-one, hence $\boldsymbol{z} \neq 0 \Rightarrow \boldsymbol{M}\boldsymbol{z} \neq 0$. Therefore by theorem 2.5.10, the minimum set of program $(\text{EGP})_z$ is non-empty and bounded

$\Leftrightarrow$ The system $\tilde{g}_k^{\infty}(\boldsymbol{z}) \leq 0,\ k \in \breve{K}$ has no non-trivial solution in $\boldsymbol{z}$

$\Leftrightarrow \exists$ no $\boldsymbol{z}$ with $0 \neq \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \boldsymbol{z} \leq 0$, i.e. system (I) above has no solution

$\Leftrightarrow$ System (II) above has a solution

$\Leftrightarrow$ Program (EGD) is canonical.

In this case, program (EGP) must have a minimum solution $\boldsymbol{t}'$. ■

5.0 DUALITY THEORY OF COMPOSITE GP

5.1 PROBLEM FORMULATIONS OF CGP AND CGD

Recall that in an (EGP) problem (4.1) the problem functions are of the form:

$$\tilde{G}_k(\boldsymbol{t}) = \sum_{i\in[k]} \left[U_i(\boldsymbol{t}) \cdot \exp \sum_{l\in<i>} V_l(\boldsymbol{t}) \right], \text{ for } k \in \breve{K} := \{0\} \cup K$$

Now, if in the above expression, each second tier term $V_l(\boldsymbol{t})$ is replaced by $h_l(\ln V_l(\boldsymbol{t}))$, where $h_l : R \to R$ is a differentiable and strictly convex function, then we call the resultant function

(5.1) $$\tilde{G}_k(\boldsymbol{t}) = \sum_{i\in[k]} \left[U_i(\boldsymbol{t}) \cdot \exp \sum_{l\in<i>} h_l(\ln V_l(\boldsymbol{t})) \right]$$

a *composite posynomial*, its terms $\tilde{U}_i(\boldsymbol{t}) := U_i(\boldsymbol{t}) \cdot \exp \sum_{l\in<i>} h_l(\ln V_l(\boldsymbol{t}))$ *composite posynomial terms*, and the resultant program a *composite* (*posynomial*) *geometric program*, which can be formally stated as

Primal CGP Program

$$(CGP) \ \inf_{\boldsymbol{0}<\boldsymbol{t}\in R^m} \tilde{G}_0(\boldsymbol{t}) \ s.t. \ \tilde{G}_k(\boldsymbol{t}) \le 1, \ k \in K,$$

where each problem function $\tilde{G}_k(\boldsymbol{t})$ is as given in (5.1). With $E_i(\boldsymbol{t}) := \sum_{l\in<i>} h_l(\ln V_l(\boldsymbol{t}))$, one has $\tilde{U}_i(\boldsymbol{t}) = U_i(\boldsymbol{t}) \cdot \exp E_i(\boldsymbol{t})$ and $\tilde{G}_k(\boldsymbol{t}) = \sum_{i\in[k]} \tilde{U}_i(\boldsymbol{t})$ as before. Everything else stays the same as in

(EGP): the index sets $I=\{1,...,n\}, J=\{1,...,m\}$, $K=\{1,\cdots,p\}$, and $L=\{1,\cdots,r\}$ together with their partition structures:

$$I=\cup_{k\in K}[k], \qquad L=\cup_{i\in I^+}<i>=\cup_{i\in I}<i>$$

where $I^+ := \{i\in I \mid <i> \neq \phi\}$, the sub-index sets $[k]$ and $<i>$ having $n_k>0$ and $r_i \geq 0$ number of elements, respectively, the exponent matrix $M=\begin{bmatrix} A \\ B \end{bmatrix}$, where $A=[a_{ij}]: n\times m,\ B=[b_{lj}]: r\times m$, and the data matrix $\begin{bmatrix} A & c \\ B & d \end{bmatrix}$, where $c_i=\ln C_i,\ i\in I,\ d_l=\ln D_l,\ l\in L$ all are the same as in Chapter 4. The only thing extra is a list of differentiable and strictly convex functions $h_l: R\to R, \forall l\in L$. So the problem is identifiable by an 8-tuple $(m,n,r,p,\boldsymbol{A},\boldsymbol{B},\boldsymbol{c},\boldsymbol{d})$ plus a list of functions: $h_1,\cdots,h_r$ as well as the partition structures of I and L. Clearly, the previous program (EGP) is a special case of this program if $h_l=\exp, \forall l\in L$. This generalization of (EGP) opens the door to model many more applications. As before,

when $i\notin I^+$, we regard $E_i(\boldsymbol{t})=0$, and $\tilde{U}_i(\boldsymbol{t})=U_i(\boldsymbol{t})$

when $[k]\cap I^+=\phi$, $\tilde{G}_k(\boldsymbol{t})=\sum_{i\in[k]}U_i(\boldsymbol{t}):=G_k(\boldsymbol{t})$

When $r=0,\ L=\phi, <i>=\phi,\ \forall i\in I,\ \tilde{G}_k(\boldsymbol{t})=G_k(\boldsymbol{t}), \forall k\in K$, this reduces to a posynomial program GP. The *underlying* (GP) of this (CGP) is meant to be the program when the entire list of second tier terms $V_l(\boldsymbol{t})$ are dropped off, i.e. when $L=\phi$, and it is identifiable by a 5-tuple $(m,n,p,\boldsymbol{A},\boldsymbol{c})$. We shall assume here that $r=r_1+\ldots+r_n>0$ so that (CGP) is not merely a GP problem.

If we take the logarithm of all problem functions in (CGP) and express them in the logarithm of the design variables t_j, we arrive at an equivalent program:

A convex form of (CGP) in the variables $z_j =: \ln t_j, j \in J$

$$(CGP)_z \ \inf_{z \in R^m} \tilde{g}_0(\boldsymbol{z}) \ s.t.\ \tilde{g}_k(\boldsymbol{z}) \le 0,\ \forall k \in K,$$

where the problem functions

(5.2) $$\tilde{g}_k(\boldsymbol{z}) := \ln \tilde{G}_k(\boldsymbol{t}) = \ln\left\{\sum_{i\in[k]} \exp\left[\boldsymbol{a}^i \boldsymbol{z} + c_i + \sum_{l\in<i>} h_l\left(\boldsymbol{b}^l \boldsymbol{z} + d_l\right)\right]\right\},\ \forall k \in K$$

are clearly differentiable (recall that $U_i = e^{a^i z + c_i}$, $V_l = e^{b^l z + d_l}$, $d_l =: \ln D_l$, and $\boldsymbol{b}^l$ is the l^{th} row of the matrix B).

It is known that the composition of some posynomials with positive exponents can be easily reformulated as a GP [Duffin et.al. pp. 96-97]. Although here the problem functions $\tilde{g}_k$ are also of composite type, these composite functions are defined on the z-space, not on the t-space. Specifically, each $\tilde{g}_k$ is a composite of the geometric function $geo: R^{n_k} \to R$ with a mapping $: R^m \to R^{n_k}$ whose components $\boldsymbol{a}^i \boldsymbol{z} + c_i + \sum_{l\in<i>} h_l(\boldsymbol{b}^l \boldsymbol{z} + d_l),\ \forall i \in [k]$ are clearly convex in $\boldsymbol{z}$. Hence by proposition 2.3.1 $\tilde{g}_k(\boldsymbol{z})$ is convex.

In terms of the variables $\boldsymbol{x} = \boldsymbol{Az},\ \boldsymbol{\xi} = \boldsymbol{Bz} \Leftrightarrow x_i = \boldsymbol{a}^i \boldsymbol{z},\ i \in I,\ \xi_l =: \boldsymbol{b}^l \boldsymbol{z},\ l \in L$, one has

(5.3) $$\tilde{g}_k(\boldsymbol{z}) = \ln\left\{\sum_{i\in[k]} \exp\left[x_i + c_i + \sum_{l\in<i>} h_l\left(\xi_l + d_l\right)\right]\right\} = \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}),$$

where $\tilde{g}eo$ is the composite geometric function we defined in (4.6). Thus we obtain an equivalent form of CGP problems:

GGP form of (CGP) in the variables $x_i =: \boldsymbol{a}^i \boldsymbol{z},\ i \in I$, and $\xi_l =: \boldsymbol{b}^l \boldsymbol{z},\ l \in L.$

$$(CGP)_{x,\xi} \inf_{(\boldsymbol{x},\xi)\in R^n \times R^r} \tilde{g}eo(\boldsymbol{x}^0 + \boldsymbol{c}^0, \boldsymbol{\xi}^{[0]} + \boldsymbol{d}^{[0]})\ s.t.\ \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}) \le 0, \forall k \in K, (\boldsymbol{x}, \boldsymbol{\xi}) \in \mathcal{P}$$

where the *primal space* $\mathscr{P}$ is the same as in (EGP): the column space of the exponent matrix M. The variables in the problem functions $\tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]})$ are also separated for different $k \in \breve{K}$.

Problem formulation of (CGD)

The format of dual program (CGD) is basically the same as that of the dual program (EGD), where the dual variables y_i and η_l correspond to the first and second tier terms $\tilde{U}_i(\boldsymbol{t})$ and $V_l(\boldsymbol{t})$, respectively. As before $\boldsymbol{y} = [y_i]_n^1 \in R^n$, $\boldsymbol{\eta} = [\eta_l]_r^1 \in R^r$, and the dual program is also defined as:

$$(5.4) \qquad (\text{CGD}) \quad \begin{cases} \sup \tilde{V}(\boldsymbol{y}, \eta) := \exp[-\tilde{l}_\lambda^*(\boldsymbol{y}, \eta)] \\ s.t. \ (\boldsymbol{y}, \eta) \in \mathscr{D} = \text{dual space of (CGD)} \end{cases}$$

where $\tilde{l}_\lambda^*(\boldsymbol{y}, \boldsymbol{\eta})$ is the conjugate of the Lagrangian $\tilde{l}_\lambda(\boldsymbol{x}, \xi) = \tilde{l}(\boldsymbol{x}, \boldsymbol{\xi}; \lambda)$ of $(\text{CGP})_{\boldsymbol{x},\xi}$ and $\mathscr{D} := \mathscr{P}^\perp =$ null space of M^T, that is, $(\boldsymbol{y}, \eta)$ must satisfy $A^T \boldsymbol{y} + B^T \eta = \boldsymbol{0}$, or equivalently:

$$\sum_{i=1}^n a_{ij} y_i + \sum_{l=1}^r b_{lj} \eta_l = 0, \quad \forall j \in J \text{ (Orthogonality conditions)}$$

Proposition 5.1.1 *The dual objective of* (CGD) *is*

$$(5.5) \qquad \begin{aligned} &\tilde{V}(\mathbf{y}, \eta) = \prod_{k \in \breve{K}} \prod_{i \in [k]} \left(\frac{C_i \lambda_k}{y_i} \right)^{y_i} \cdot \exp\left\{ \sum_{i \in I^+} \sum_{l \in <i>} \left[d_l \eta_l - y_i h_l^*(\eta_l / y_i) \right] \right\} \\ &\text{if } y_i \ge 0, \ \textstyle\sum_{i \in [k]} y_i = \lambda_k, \forall k \in \breve{K}, \ \lambda_0 = 1, \text{ and } \eta_l \in y_i J_l, \forall l \in <i>, i \in I^+ \\ &\text{otherwise, its value is } 0, \end{aligned}$$

where, J_l is the domain interval of the conjugate function h_l^*.

Proof: The Lagrangian of $(\text{CGP})_{\boldsymbol{x},\xi}$ is

$$\tilde{l}(\boldsymbol{x}, \boldsymbol{\xi}; \lambda) = \textstyle\sum_{k \in \breve{K}} \lambda_k \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}), \text{ with } \lambda_0 = 1$$

Its conjugate, by (4.9), is

$$\tilde{l}_{\lambda}^{*}(\boldsymbol{y},\boldsymbol{\eta}) = \sum_{k\in\breve{K}}\Big[\sum_{i\in[k]} y_i \ln(y_i / C_i\lambda_k) + \sum_{i\in[k]^{+}} \sum_{l\in<i>}\big[y_i h_l^{*}(\eta_l / y_i) - d_l\eta_l \big]\Big],$$

if $y_i \ge 0,\ \sum_{i\in[k]} y_i = \lambda_k, \forall k \in \breve{K},\ \lambda_0 = 1,$ and $\eta_l \in y_i J_l, \forall l \in <i>, i \in I^{+}$; otherwise, its value is ∞.

So the dual objective of (CGD) is

$$\tilde{V}(\boldsymbol{y},\eta) = \exp[-\tilde{l}_{\lambda}^{*}(\boldsymbol{y},\eta)] = \prod_{k\in\breve{K}}\prod_{i\in[k]}\left(\frac{C_i\lambda_k}{y_i}\right)^{y_i} \cdot \exp\left\{\sum_{i\in I^{+}}\sum_{l\in<i>}\left[d_l\eta_l - y_i h_l^{*}(\eta_l / y_i)\right]\right\},$$

with the above-mentioned restrictions. ■

Thus the dual program (CGD) can be elaborated as

Dual program (CGD) (Under the convention: $0^0 = 1,\ 0h_l^{*}(\eta_l / 0) = i(\eta_l \mid 0)$)

$$(5.6)\qquad (CGD)\begin{cases} \sup\limits_{(\mathbf{y},\boldsymbol{\eta})\in R^n\times R^r} \tilde{V}(\mathbf{y},\eta) := \prod\limits_{k\in\breve{K}}\prod\limits_{i\in[k]}\left(\frac{C_i\lambda_k}{y_i}\right)^{y_i} \cdot \exp\left\{\sum\limits_{i\in I^{+}}\sum\limits_{l\in<i>}\left[d_l\eta_l - y_i h_l^{*}(\eta_l / y_i)\right]\right\} \\ s.t. \qquad \lambda_0 = 1, \qquad \text{(Normality Condition)} \\ \qquad \sum\limits_{i=1}^{n} a_{ij} y_i + \sum\limits_{l=1}^{r} b_{lj}\eta_l = 0,\ \ j\in J \quad \text{(Orthogonality Conditions)} \\ \qquad y_i \ge 0, \forall i \in I, \sum\limits_{i\in[k]} y_i = \lambda_k,\ \ \forall k \in \breve{K} \\ \qquad \eta_l \in y_i J_l,\ \ \forall l \in <i>,\ \ i \in I^{+} \qquad (*) \end{cases}$$

The domain conditions in (*) constraints are sublinear, hence potentially nonlinear, but they will be fulfilled if a pair of primal and dual solutions $(\boldsymbol{t};\boldsymbol{y},\eta)$ satisfy the extremality condition to be described in equation (5.10) (b). In particular, they imply that

$$(5.7)\qquad y_i = 0 \Rightarrow \eta_l = 0,\ \ \forall l \in <i>, i \in I^{+}$$

The convention $0h_l^{*}(\eta_l / 0) = i(\eta_l \mid 0)$ implies that $\tilde{V}(\boldsymbol{y},\eta) = 0,$ if $\exists l \in <i>, s.t.\ y_i = 0, \eta_l \ne 0$

As with (EGD), *the underlying* (GD) of this (CGD) is meant to be the program where the second tier vector variable η , the constant vector $\boldsymbol{d}$, and the matrix B are all absent from (CGD), i.e. set $L = \phi$. This program is then the geometric dual of the underlying (GP).

The dual functions $\tilde{V}(\boldsymbol{y}, \eta)$ and ***V(y)*** satisfy

(5.8) $$\tilde{V}(\boldsymbol{y}, \eta) = V(\boldsymbol{y}) \cdot \exp\left\{ \sum_{i \in I^+} \sum_{l \in \langle i \rangle} \left[d_l \eta_l - y_i h_l^*(\eta_l / y_i) \right] \right\}$$

and their respective log-dual functions $\tilde{v}(\boldsymbol{y}, \eta) = \ln \tilde{V}(\boldsymbol{y}, \eta)$ and $v(\boldsymbol{y}) = \ln V(\boldsymbol{y})$ satisfy:

(5.9) $$\tilde{v}(\boldsymbol{y}, \eta) = v(\boldsymbol{y}) + \sum_{i \in I^+} \sum_{l \in \langle i \rangle} \left[d_l \eta_l - y_i h_l^*(\eta_l / y_i) \right].$$

As already pointed out in chapter 4, this function $\tilde{v}(\boldsymbol{y}, \eta)$ is *usc* proper and concave.

So the dual program (CGD) is also equivalent to a convex program with (m+1) linear equality constraints and non-negativity conditions on its n dual variables y_i, except for *except* for those constraints in (*). Its *degree of difficulty d* is also defined as the number $n + r - m - 1$, which is usually the dimension of its dual flat. This number is r greater than that of its underlying GD, because CGP is a more complex model than its underlying GP, containing r more posynomial terms in the second tier.

Just like the EGD case, the log-dual program of (CGD) also possesses 3 exploitable structures: *concavity, partial-separability,* and *linearity* that are originally hidden in the (CGP) model. Obviously, these features would make solving the primal through its dual an attractive approach, provided that the two programs have identical optimal values and conversion of optimum solutions from dual to primal is computationally economical. As in the precious chapter, the duality theory of CGP attempts to confirm the above fact under mild conditions.

5.2 MAIN LEMMA OF CGP

Lemma 5.2.1 (Main Lemma of CGP) *If* $\boldsymbol{t}$ *is feasible for primal program* (CGP) *and* $(\boldsymbol{y}, \eta)$ *is feasible for dual program* (CGD)*, then* $\tilde{G}_0(\boldsymbol{t}) \geq \tilde{V}(\boldsymbol{y}, \eta)$. *Moreover, under the same conditions,* $\tilde{G}_0(\boldsymbol{t}) = \tilde{V}(\boldsymbol{y}, \eta)$ *if and only if one of the following two sets of equivalent conditions holds:*

(5.10)
$$(a)\begin{cases} \tilde{G}_k(\boldsymbol{t})^{\lambda_k} = 1, \ k \in K, \\ y_i \tilde{G}_k(\boldsymbol{t}) = \lambda_k \tilde{U}_i(\boldsymbol{t}), \ i \in [k], k \in \check{K} \end{cases}$$
$$(b) \ \eta_l = y_i h_l^{'}(\ln V_l(\boldsymbol{t})), \ l \in <i>, \ i \in I^+$$
(Extremality condition 1)

(5.11)
$$(c) \ y_i = \begin{cases} \tilde{U}_i(\boldsymbol{t}) \big/ \tilde{G}_0(\boldsymbol{t}), & i \in [0] \\ \lambda_k \tilde{U}_i(\boldsymbol{t}), & i \in [k], \ k \in K \end{cases}$$
$$(d) \ \eta_l = y_i h_l^{'}(\ln V_l(\boldsymbol{t})), \ l \in <i>, \ i \in I^+$$
(Extremality condition 2)

in which case $\boldsymbol{t}$ *is optimal for primal program* (CGP) *and* $(\boldsymbol{y}, \eta)$ *is optimal for dual program* (CGD).

(Recall from Lemma 2.1.0 that the above conditions (*a*) and (*c*) are equivalent.)

Proof: Let $(\boldsymbol{x}, \xi) = (A\boldsymbol{z}, B\boldsymbol{z})$, where $z_j = \ln t_j, \forall j \in J$. Then $(\boldsymbol{x}, \xi) \in \mathscr{P}$ is feasible for program $(\text{CGP})_{\boldsymbol{x},\xi}$ Since $(\boldsymbol{y}, \eta) \in \mathscr{D}$, $(\boldsymbol{x}, \xi) \cdot (\boldsymbol{y}, \eta) = 0$ Applying geometric inequality (2.8) to each problem function $\tilde{G}_k(\boldsymbol{t})$ in (*CGP*), (5.1) yields

$$\tilde{G}_0 \geq \tilde{G}_0 \cdot \prod_{k \in K} \tilde{G}_k^{\lambda_k} \geq \prod_{k \in \check{K}} \prod_{i \in [k]} \left(\tilde{U}_i \lambda_k / y_i \right)^{y_i} = \prod_{k \in \check{K}} \prod_{i \in [k]} \left(U_i \lambda_k / y_i \right)^{y_i} \cdot \prod_{i \in I} e^{y_i E_i}, \quad \because \tilde{U}_i = U_i e^{E_i}$$

$$= \prod_{k \in \check{K}} \prod_{i \in [k]} \left(C_i \lambda_k / y_i \right)^{y_i} \cdot e^{\boldsymbol{x} \cdot \boldsymbol{y}} \cdot \exp\left[\sum_{i \in I^+} \sum_{l \in <i>} y_i h_l \left(\xi_l + d_l \right) \right], \ \because U_i = C_i e^{x_i}, \ E_i = \sum\nolimits_{l \in <i>} h_l \left(\xi_l + d_l \right)$$

$$\geq V(\boldsymbol{y}) \cdot e^{\boldsymbol{x} \cdot \boldsymbol{y}} \cdot \exp\left[\sum_{i \in I^+} \sum_{l \in <i>} \left[\left(\xi_l + d_l \right) \eta_l - y_i h_l^*(\eta_l / y_i) \right] \right],$$ by generalized geometric inequality (2.36)

$$= V(\boldsymbol{y}) \cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}} \cdot e^{\xi\cdot\eta} \cdot \exp\left[\sum_{i\in I^+} \sum_{l\in<i>} \left[d_l \eta_l - y_i h_l^*(\eta_l / y_i) \right] \right]$$

$$= V(\boldsymbol{y}) \cdot \exp\left[\sum_{i\in I^+} \sum_{l\in<i>} \left[d_l \eta_l - y_i h_l^*(\eta_l / y_i) \right] \right], \qquad \because \boldsymbol{x}\cdot\boldsymbol{y} + \boldsymbol{\xi}\cdot\eta = 0$$

$$= \tilde{V}(\boldsymbol{y}, \eta)$$

Therefore, $\tilde{G}_0(\boldsymbol{t}) \geq \tilde{V}(\boldsymbol{y}, \eta)$ and conditions for equality in the last inequality are $\eta_l = y_i h_l^{'}(\xi_l + d_l),\ l \in <i>,\ i \in I^+$, which are the same as $\eta_l = y_i h_l^{'}(\ln V_l(\boldsymbol{t})),\ l \in <i>,\ i \in I^+$, and this is condition (b) of (5.10). Therefore, equality holds iff the extremality conditions (5.10) or (5.11) are satisfied. ■

Alternate proof Assume that we have $(\boldsymbol{x}, \xi) \in \mathscr{P}$ and $(\boldsymbol{y}, \eta) \in \mathscr{D}$ feasible for program $(\text{CGP})_{\boldsymbol{x},\xi}$ and (CGD), respectively, thus $(\boldsymbol{x}, \xi) \perp (\boldsymbol{y}, \eta)$ and

(5.12) $$\tilde{g}eo(\boldsymbol{x}^0 + \boldsymbol{c}^0, \boldsymbol{\xi}^{[0]} + \boldsymbol{d}^{[0]}) \geq \sum_{k\in K} \lambda_k \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}) = \tilde{l}_\lambda(\boldsymbol{x}, \xi), \text{ where } \lambda_0 = 1$$

with equality holding iff $\lambda_k \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}) = 0, \forall k \in K$.

In $(\text{CGP})_{\boldsymbol{x},\xi}$, each problem function

$$\tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}) = \ln\left\{ \sum_{i\in[k]} \exp\left[x_i + c_i + \sum_{l\in<i>} h_l\left(\xi_l + d_l\right) \right] \right\}, \forall k \in K$$

has partial derivatives:

(5.13) $$\left[\frac{\partial}{\partial x_i} = \frac{\tilde{U}_i(\boldsymbol{t})}{\tilde{G}_k(\boldsymbol{t})}, \quad \frac{\partial}{\partial \xi_l} = \frac{\tilde{U}_i(\boldsymbol{t})}{\tilde{G}_k(\boldsymbol{t})}[h_l^{'}(\ln V_l(\boldsymbol{t}))], \forall l \in <i> \right], \ \forall i \in [k], \forall k \in K$$

Apply (2.33) to each function $\lambda_k \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \mathbf{d}^{[k]}),\ k \in K$ using (4.9) to get

(5.14) $$\begin{aligned} &\lambda_k \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}) + \sum_{i\in[k]} y_i \ln(y_i / C_i \lambda_k) + \sum_{i\in[k]^+} \sum_{l\in<i>} \left[y_i h_l^*(\eta_l / y_i) - d_l \eta_l \right] \\ &\geq \sum_{i\in[k]} \left[x_i y_i + \sum_{l\in<i>} \xi_l \eta_l \right], \ \forall k \in K \end{aligned}$$

where equality holds iff $(\boldsymbol{y}^{[k]}, \eta^{[k]}) = \lambda_k \nabla \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]})$, viz.,

$$(5.15)\left[y_i = \lambda_k \frac{\tilde{U}_i(\boldsymbol{t})}{\tilde{G}_k(\boldsymbol{t})},\ \eta_l = \lambda_k \frac{\tilde{U}_i(\boldsymbol{t})}{\tilde{G}_k(\boldsymbol{t})}[h_l^{'}(\ln V_l(\boldsymbol{t}))] = y_i h_l^{'}(\ln V_l(\boldsymbol{t})), \forall l \in <i> \right], \forall i \in [k] \text{ by}(5.13).$$

Summing over all of the inequalities in(5.14), we get

$$(5.16) \quad \begin{aligned} \tilde{l}_\lambda(\boldsymbol{x}, \boldsymbol{\xi}) &\geq -\left\{ \sum_{k \in \breve{K}} \left[\sum_{i \in [k]} y_i \ln(y_i / C_i \lambda_k) + \sum_{i \in [k]^+} \sum_{l \in <i>} \left[y_i h_l^*(\eta_l / y_i) - d_l \eta_l \right] \right] \right\} \\ &\quad + \sum_{k \in \breve{K}} \sum_{i \in [k]} \left[x_i y_i + \sum_{l \in <i>} \xi_l \eta_l \right] \\ &= -\tilde{l}_\lambda^*(\boldsymbol{y}, \boldsymbol{\eta}) + (\boldsymbol{x}, \xi) \cdot (\boldsymbol{y}, \eta) = -\tilde{l}_\lambda^*(\boldsymbol{y}, \boldsymbol{\eta}), \qquad \because (\boldsymbol{x}, \xi) \perp (\boldsymbol{y}, \eta) \end{aligned}$$

with equality holding iff(5.15) holds $\forall k \in \breve{K}$.

Combining inequalities (5.12) and (5.16) yields

$$(5.17) \qquad \tilde{g}eo(\boldsymbol{x}^0 + \boldsymbol{c}^0, \boldsymbol{\xi}^{[0]} + \boldsymbol{d}^{[0]}) \geq \tilde{l}_\lambda(\boldsymbol{x}, \xi) \geq -\tilde{l}_\lambda^*(\boldsymbol{y}, \boldsymbol{\eta})$$

Exponentiating the far two sides of this inequality, we get $\tilde{G}_0(\boldsymbol{t}) \geq \tilde{V}(\boldsymbol{y}, \eta)$. This holds as an equality iff $\lambda_k \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]}) = 0, \forall k \in K$ and (5.15) holds $\forall k \in \breve{K}$, which is clearly equivalent to Extremality condition 1 (5.10). ■

5.3 FIRST AND SECOND DUALITY THEOREMS OF CGP

In the convex formulation (CGP)z : $\inf_{z \in R^m} \tilde{g}_0(\boldsymbol{z})\ s.t.\ \tilde{g}_k(\boldsymbol{z}) \leq 0,\ \forall k \in K$, each problem function

$$\tilde{g}_k(\boldsymbol{z}) = \ln\left\{ \sum_{i \in [k]} \exp\left[\boldsymbol{a}^i \boldsymbol{z} + c_i + \sum_{l \in <i>} h_l(\boldsymbol{b}^l \boldsymbol{z} + d_l) \right] \right\},\ k \in \breve{K}$$

has partial derivatives:

$$(5.18) \qquad \frac{\partial \tilde{g}_k(\boldsymbol{z})}{\partial z_j} = \frac{1}{\tilde{G}_k(\boldsymbol{t})} \sum_{i \in [k]} \tilde{U}_i(\boldsymbol{t}) \left[a_{ij} + \sum_{l \in <i>} h_l^{'}(\ln V_l(\boldsymbol{t})) b_{lj} \right],\ \forall j \in J$$

The Lagrangian for this differentiable convex program (CGP)$_z$ is

$$\tilde{l}(\boldsymbol{z},\lambda) := \tilde{g}_0(\boldsymbol{z}) + \sum_{k\in K} \lambda_k \tilde{g}_k(\boldsymbol{z}),\ \forall \lambda \in R_+^p, \boldsymbol{z} \in R^m.$$

Theorem 5.3.1 (First Duality Theorem of CGP) *Suppose that primal program* (CGP) *is superconsistent. Then the following 3 conditions are equivalent*:

1) ***t'*** *is a minimal solution to* (CGP).

2) *There exists a vector* $\lambda' \in R_+^p$ *for* $\boldsymbol{z}'$ *(where* $\boldsymbol{z}' = \ln \boldsymbol{t}'$*) such that* $(\boldsymbol{z}',\lambda')$ *forms a saddle point of* $\tilde{l}(\boldsymbol{z},\lambda)$.

3) *There exists a vector* $\lambda' \in R_+^p$ for $\boldsymbol{z}'$ *(where* $\boldsymbol{z}' = \ln \boldsymbol{t}'$*) such that* $(\boldsymbol{z}',\lambda')$ *satisfies the KKT conditions for* (CGP)$_z$:

$$(a)\ \lambda_k' \ge 0,\ \tilde{g}_k(\boldsymbol{z}') \le 0,\ \lambda_k' \tilde{g}_k(\boldsymbol{z}') = 0, \forall k \in K$$
$$(b)\ \nabla_z \tilde{l}(\boldsymbol{z}',\lambda') = \sum_{k\in \breve{K}} \lambda_k' \nabla \tilde{g}_k(\boldsymbol{z}') = \boldsymbol{0},\ where\ \lambda_0' = 1$$

in which case the set of all such vectors $\boldsymbol{\lambda}'$ *is a non-empty compact convex subset of* R_+^p, *and the dual program* (CGD) *also has a maximum solution* $(\boldsymbol{y}',\eta')$ *such that*

$$\min(\text{CGP}) = \tilde{G}_0(\boldsymbol{t}') = \tilde{V}(\boldsymbol{y}',\eta') = \max(\text{CGD}) \qquad \text{(Perfect duality)}$$

Proof: By assumption, the differentiable convex program (CGP)$_z$ satisfies the Slater CQ. Hence by KKT Theorem (Theorem 2.4.4), the above conditions 1) through 3) are equivalent and in which case, by Theorem 2.4.3 the set of all such vectors $\boldsymbol{\lambda}'$ is a non-empty compact convex subset of R_+^p, and conditions (*a*) and (*b*) in 3) hold true. We note that

$$\text{Condition } (b)\ \Rightarrow \sum_{k\in \breve{K}} \lambda_k' \frac{\partial \tilde{g}_k(\boldsymbol{z}')}{\partial z_j} = 0,\ \forall j \in J$$

$$\Rightarrow \sum_{k\in \breve{K}} \frac{\lambda_k'}{\tilde{G}_k(\boldsymbol{t}')} \sum_{i\in[k]} \tilde{U}_i(\boldsymbol{t}) \left[a_{ij} + \sum_{l\in<i>} h_l'(\ln V_l(\boldsymbol{t})) b_{lj} \right] = 0,\ \forall j \in J \text{, by (5.18)}$$

Now define $y_i' = \lambda_k' \tilde{U}_i(\boldsymbol{t}') / \tilde{G}_k(\boldsymbol{t}') \geq 0, \forall i \in [k], \forall k \in \breve{K}$; and $\eta_i' = y_i' h_l'(\ln V_l(\boldsymbol{t}'))$, $l \in <i>$, $i \in I^+$.

Then condition (*) in (CGD): $\eta_l \in y_i J_l$, $\forall l \in <i>$, $i \in I^+$ is satisfied $(\because$ range $(h_l') \subset$ dom $h_l^*)$

and we have

$$\sum_{k \in \breve{K}} \sum_{i \in [k]} \left[a_{ij} y_i' + \sum_{l \in <i>} b_{lj} \eta_l' \right] = \sum_{i=1}^{n} a_{ij} y_i' + \sum_{l=1}^{r} b_{lj} \eta_l' = 0, \ \ j \in J$$

which are the orthogonality conditions for (CGD). It is then easy and routine to check that the above given $(\boldsymbol{y}', \eta')$ is a dual feasible solution, which together with $\boldsymbol{t'}$ satisfy the extremality condition (5.10). Therefore, by main lemma of (CGP), this $(\boldsymbol{y}', \eta')$ is a maximum solution satisfying

$$\min(\text{CGP}) = \tilde{G}_0(\boldsymbol{t}') = \tilde{V}(\boldsymbol{y}', \eta') = \max(\text{CGD}) \qquad \blacksquare$$

Since in $(\text{CGP})_z$ the problem functions are $\tilde{g}_k(\boldsymbol{z}) = \tilde{g}eo(\boldsymbol{A}^k \boldsymbol{z} + \boldsymbol{c}^k, \boldsymbol{B}^{[k]} \boldsymbol{z} + \boldsymbol{d}^{[k]})$,

$$\tilde{g}_k^\infty(\boldsymbol{z}) = \tilde{g}eo^\infty(\boldsymbol{A}^k \boldsymbol{z}, \boldsymbol{B}^{[k]} \boldsymbol{z}) = \max_{i \in [k]} \left[\boldsymbol{a}^i \boldsymbol{z} + \sum\nolimits_{l \in <i>} h_l^\infty(\boldsymbol{b}^l \boldsymbol{z}) \right], \text{ by (4.24)}$$

The directions of recession of $(\text{CGP})_z$ are the nontrivial solutions to the system:

$$\begin{aligned} \tilde{g}_k^\infty(\boldsymbol{z}) \leq 0, \forall k \in \breve{K} &\Leftrightarrow \boldsymbol{a}^i \boldsymbol{z} + \sum\nolimits_{l \in <i>} h_l^\infty(\boldsymbol{b}^l \boldsymbol{z}) \leq 0, \forall i \in [k], \forall k \in \breve{K} \\ &\Leftrightarrow \boldsymbol{a}^i \boldsymbol{z} + H_i^\infty(\boldsymbol{B}^i \boldsymbol{z}) \leq 0, \forall i \in I \\ &\Leftrightarrow \boldsymbol{A}\boldsymbol{z} + H^\infty(\boldsymbol{B}\boldsymbol{z}) \leq \boldsymbol{0} \end{aligned} \tag{5.19}$$

Since in general we do not know the exact form of the dual linear system of inequalities to the above system (this depends on the nature of all the functions h_l that are involved), we cannot provide a general duality theorem here. We can however, supply an existence theorem.

Theorem 5.3.2 (Existence Theorem of CGP) *Suppose that primal program* (CGP) *is consistent. Then the minimum set of program* $(\text{CGP})_z$ *is non-empty and bounded if and only if the system* $\boldsymbol{A}\boldsymbol{z} + H^\infty(\boldsymbol{B}\boldsymbol{z}) \leq \boldsymbol{0}$ *has no nontrivial solutions in* $\boldsymbol{z}$.

6.0 DUALITY THEORY OF QUADRATIC GP

The content of the duality theory of QGP to be presented in this chapter is independent of the material in chapters 4 and 5. However, in chapter 7, we shall present a more general and elegant approach to the same theory but this time as a special case of the duality theory of CGP developed in Chapter 5. In this way one can contrast the two different approaches to the same subject.

6.1 PROBLEM FORMULATIONS OF QGP

We define a quadratic geometric program (QGP) as follows: First, the index sets $I = \{1,...,n\}, J = \{1,...,m\}$, $K = \{1,\cdots,p\}$, and $L = \{1,\cdots,r\}$ together with the partition structures:

$$I = \cup_{k\in K}[k], \qquad L = \cup_{i\in I^+} < i > = \cup_{i\in I} < i > = \cup_{k\in K} \cup_{i\in[k]} < i >$$

where $I^+ := \{i \in I \mid r_i > 0\} = \{i \in I \mid < i > \neq \phi\}$, are the same as in (EGP) (Chapter 4). Each index subset <*i*> has $r_i\,(\geq 0)$ elements, $< i > = \{S_{i-1}+1,\cdots,S_i\}, \forall i \in I$, where $S_i := r_1 + \ldots + r_i, \forall i \in I$ and $S_0 = 0$.

Primal program (QGP)

$$(QGP)\ \ \inf_{\boldsymbol{0}<\boldsymbol{t}\in R^m} \tilde{G}_0(\boldsymbol{t})\ s.t.\ \tilde{G}_k(\boldsymbol{t}) \le 1,\ k \in K,$$

with

$$(6.1)\qquad \begin{aligned} &\tilde{G}_k(\boldsymbol{t}) = \sum\nolimits_{i\in[k]} \tilde{U}_i(\boldsymbol{t}),\ \tilde{U}_i(\boldsymbol{t}) := U_i(\boldsymbol{t}) \cdot \exp H_i(\boldsymbol{t}),\ U_i(\boldsymbol{t}) = C_i \prod\nolimits_{j\in J} t_j^{a_{ij}},\ k \in \breve{K} \\ &H_i(\boldsymbol{t}) := \sum\nolimits_{l\in<i>} \tfrac{1}{2} \ln^2 V_l(\boldsymbol{t}),\ \text{and } V_l(\boldsymbol{t}) = D_l \prod\nolimits_{j\in J} t_j^{b_{lj}},\ \forall l \in <i>, \forall i \in I^+ \end{aligned}$$

For easy reference, we adopt the name given by Hough (1978) and call $\tilde{U}_i(\boldsymbol{t})$ a Quadratic Posylognomial (QPL) term, when $i \in I^+ (\Leftrightarrow <i> \ne \phi \Leftrightarrow r_i > 0)$, and call $\tilde{G}_k(\boldsymbol{t})$ a Quadratic Posylognomial, when $[k] \cap I^+ \ne \phi$. The exponent matrix for this program is $M = \begin{bmatrix} A \\ B \end{bmatrix}$, where $A = [a_{ij}] : n \times m,\ B = [b_{lj}] : r \times m$. The data for this program is the $(n+r)\times(m+1)$ matrix $\begin{bmatrix} A & c \\ B & d \end{bmatrix}$ and the partition structures of I and L, where $c_i = \ln C_i,\ i \in I,\ d_l = \ln D_l,\ l \in L$.

When $i \notin I^+$, we may regard $H_i(\boldsymbol{t}) = 0$, and $\tilde{U}_i(\boldsymbol{t}) = U_i(\boldsymbol{t})$.

When $[k] \cap I^+ = \phi,\ \tilde{G}_k(\boldsymbol{t}) = \sum\nolimits_{i\in[k]} U_i(\boldsymbol{t}) := G_k(\boldsymbol{t}).$

When $I^+ = \phi\ (\Leftrightarrow L = \phi \Leftrightarrow r = 0),\ \tilde{G}_k(\boldsymbol{t}) = G_k(\boldsymbol{t}), \forall k \in \breve{K}$,

then our program reduces to a posynomial program GP. By *the underlying* (GP) of this (QGP) we mean the program where all of the second tier terms $V_l(\boldsymbol{t})$ are absent $(\Leftrightarrow r = 0)$. Since $H_i(\boldsymbol{t}) \ge 0$, we have $\tilde{U}_i(\boldsymbol{t}) \ge U_i(\boldsymbol{t}), \forall i,\ \tilde{G}_k(\boldsymbol{t}) \ge G_k(\boldsymbol{t}), \forall k$, thus $\inf QGP \ge \inf GP$. We assume here that $r = r_1 + \ldots + r_n > 0$ so that (QGP) is not merely a GP problem.

A convex form of (QGP) in the variables $z_j =: \ln t_j,\ j \in J$

$$(QGP)_z \ \inf_{z \in R^m} \tilde{g}_0(\boldsymbol{z}) \ s.t.\ \tilde{g}_k(\boldsymbol{z}) \le 0,\ \forall k \in K,$$

where the problem functions

$$\tilde{g}_k(\boldsymbol{z}) =: \ln\left\{\sum\nolimits_{i\in[k]} \exp\left[c_i + \boldsymbol{a}^i \boldsymbol{z} + \sum\nolimits_{l\in<i>} \tfrac{1}{2} |\boldsymbol{b}^l \boldsymbol{z} + d_l|^2\right]\right\},\ k \in \breve{K} \tag{6.2}$$

are clearly convex (recall that $U_i = e^{\boldsymbol{a}^i z + c_i}$, $V_l = e^{\boldsymbol{b}^l z + d_l}$, $d_l =: \ln D_l$, and $\boldsymbol{b}^l$ is the l^{th} row of the matrix B). These functions become those in equation (1.24), if $D_l = 1,\ \forall l \in L$, since $\| \boldsymbol{B}^i \boldsymbol{z} \|^2 = \sum_{l\in<i>} |\boldsymbol{b}^l \boldsymbol{z}|^2$, where $\boldsymbol{b}^l$ is a row of the matrix $\boldsymbol{B}^i$. So our (*QGP*) model is slightly more flexible than the (*QPL*) model studied by previous researchers. It is worth noting that if the coefficient ½ that appears in the (QGP) model formula for $\tilde{G}_k(\boldsymbol{t})$ or for $\tilde{g}_k(\boldsymbol{z})$ is any other positive constant, say s_l, the program is still convertible to a QGP problem by simply multiplying the l^{th} row $[\boldsymbol{b}^l \ \ d_l]$ of the data matrix $\left[B \quad d\right]$ by $\sqrt{2s_l}$, since $s_l\left(\boldsymbol{b}^l \boldsymbol{z} + d_l\right)^2 = \frac{1}{2}\left[\sqrt{2s_l}\left(\boldsymbol{b}^l \boldsymbol{z} + d_l\right)\right]^2$.

A GGP reformulation of (QGP) in the variables $x_i =: \boldsymbol{a}^i \boldsymbol{z},\ i \in I,\ \xi_l =: \boldsymbol{b}^l \boldsymbol{z},\ l \in L.$

$$(QGP)_{x,\xi} \inf_{(\boldsymbol{x},\boldsymbol{\xi})\in R^{n+r}} \tilde{g}_0(\boldsymbol{x},\boldsymbol{\xi}) \ s.t.\ \tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi}) \le 0,\ \forall k \in K,\ (\boldsymbol{x},\boldsymbol{\xi}) \in \mathscr{P},$$

where $\mathscr{P}$ (the *primal space* of this program) is the column space of the exponent matrix *M*, and the variables in these convex problem functions

$$\tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi}) := \ln\left\{\sum\nolimits_{i\in[k]} \exp\left[x_i + c_i + \sum\nolimits_{l\in<i>} \tfrac{1}{2} |\xi_l + d_l|^2\right]\right\}, \tag{6.3}$$

are separated for different $k \in \breve{K}$.

In the dual program (QGD) for (QGP), each first tier term $\tilde{U}_i(\boldsymbol{t})$ has a corresponding dual variable y_i, and each second tier term $V_l(\boldsymbol{t})$ has a corresponding second tier dual variable η_l. Below, we let $\boldsymbol{y} = [y_i]_n^1 \in R^n$, $\boldsymbol{\eta} = [\eta_l]_r^1 \in R^r$.

Dual program (QGD) (Under the convention $\eta_l^2 / 2y_i = 0$ for $y_i = 0 = \eta_l$ and $=\infty$ for $y_i = 0 \neq \eta_l$)

$$
(6.4) \qquad (QGD) \quad \begin{cases} \sup\limits_{(\boldsymbol{y},\boldsymbol{\eta}) \in R_+^n \times R^r} \tilde{V}(\boldsymbol{y},\eta) := \prod\limits_{k \in \breve{E}} \prod\limits_{i \in [k]} \left(\dfrac{C_i \lambda_k}{y_i} \right)^{y_i} \cdot \prod\limits_{1}^{r} D_l^{\eta_l} \cdot \exp\left[-\sum\limits_{i \in I^+} \sum\limits_{l \in <i>} \dfrac{\eta_l^2}{2y_i} \right] \\ s.t. \qquad \lambda_0 = 1, \qquad \text{(Normality Condition)} \\ \qquad \sum\limits_{i=1}^{n} a_{ij} y_i + \sum\limits_{l=1}^{r} b_{lj} \eta_l = 0, \; j \in J \quad \text{(Orthogonality Conditions)} \\ \qquad y_i = 0 \Rightarrow \eta_l = 0, \; \forall l \in <i>, \; \forall i \in I^+ \qquad (*) \\ \text{where} \quad \lambda_k := \sum\limits_{i \in [k]} y_i, \; \forall k \in \breve{E} \end{cases}
$$

The convention implies $\tilde{V}(\boldsymbol{y},\eta) = 0$ if $\exists\, i,l$ s.t. $l \in <i>$, $y_i = 0$, but $\eta_l \neq 0$; hence the constraints in (*). After proving the main lemma of QGP we shall see that a solution which satisfies the extremality conditions automatically satisfies these constraints in (*). The dual variables η_l are *not* restricted in sign.

By *the underlying* geometric programming dual (GD) of this (QGD) we mean the program where all of the second tier terms $V_l(\boldsymbol{t})$, and therefore the matrix *B*, and the vectors $\boldsymbol{d},\eta$ are all absent $(\Leftrightarrow r = 0 \Leftrightarrow L = \phi \Leftrightarrow I^+ = \phi)$. This program is clearly the geometric dual of the underlying (GP) of (QGP). A similar observation can also be made on the relationships between inf GP, inf QGP, sup GD, and sup QGD. Denote $\tilde{T}$ and T the feasible regions of programs (QGD) and (GD), respectively and observe that

$$\boldsymbol{y} \in T \Rightarrow (\boldsymbol{y},\boldsymbol{0}) \in \tilde{T}, \text{ and } V(\boldsymbol{y}) = \tilde{V}(\boldsymbol{y},\boldsymbol{0}) \leq \sup_{\tilde{T}} \tilde{V}(\boldsymbol{y},\eta) = \sup QGD.$$

Thus one has: sup $GD = \sup_T V(\mathrm{y}) \le \sup QGD$, in addition to the previous relationship $\inf GP \le \inf QGP$.

Note that the dual function satisfies

(6.5) $$\tilde{V}(\boldsymbol{y},\eta) = V(\boldsymbol{y}) \cdot \exp\left[-\sum_{i\in I^+}\sum_{l\in<i>}\left(\frac{\eta_l^2}{2y_i} - d_l\eta_l\right)\right]$$

where $V(\boldsymbol{y}) = \prod_{k\in K}\prod_{i\in[k]}\left(\frac{C_i\lambda_k}{y_i}\right)^{y_i}$ is the dual function of the underlying GD. In terms of their respective log-dual functions: $\tilde{v}(\boldsymbol{y},\eta) = \ln \tilde{V}(\boldsymbol{y},\eta)$, $v(\boldsymbol{y}) = \ln V(\boldsymbol{y})$ the above equation becomes

(6.6) $$\tilde{v}(\boldsymbol{y},\eta) = v(\boldsymbol{y}) - \sum_{i\in I^+}\sum_{l\in<i>}\left(\frac{\eta_l^2}{2y_i} - d_l\eta_l\right).$$

Proposition 6.1.1 *The log-dual function* $\tilde{v}(\boldsymbol{y},\eta)$ *of* (QGD) is concave.

Proof Since we already known that the function $v(\boldsymbol{y})$ is concave from GP duality theory, and $\forall i \in I^+$ the function $\sum_{l\in<i>}\left(\frac{\eta_l^2}{2y_i} - d_l\eta_l\right)$ is the conjugate of the function $\sum_{l\in<i>}\frac{y_i(\xi_l + d_l)^2}{2}$ (which comes from the exponential factor y_iH_i attached to the term $\tilde{U}_i(\boldsymbol{t})$ in the primal program), it is *lsc* and convex, therefore $\tilde{v}(\boldsymbol{y},\eta)$ must also be *usc* concave. ■

So the dual program (QGD) is equivalent to a convex program with (m+1) linear equality constraints and non-negativity conditions on its (n+r) dual variables, except for those constraints in (*). Its degree of difficulty d, defined as the dimension of its dual flat, equals $n + r - m - 1$. This is greater than that of its underlying GD by r, due to the fact that its primal program contains r more posynomial terms in the second tier and is therefore more complicated than its underlying GP. Observe also that in the dual program (QGD) the cost coefficients C_i and D_l only appear in its objective function and the technological coefficients a_{ij} and b_{lj} only in the constraints; whereas in (*QGP*) they are scattered all over the terms $\tilde{U}_i(t)$ of the functions $\tilde{G}_k(t)$. The log-dual program (*QGD*) possesses 3 exploitable structures: *concavity*, *partial-separability*, and *linearity* that are originally hidden in the (*QGP*) model. Obviously, these features would make solving the primal through its dual an attractive approach, provided that the two programs have identical optimal values and conversion of optimum solutions from dual to primal is not computationally economical. The duality theory of QGP confirms the above fact under mild conditions.

6.2 MAIN LEMMA OF QGP

Similar to the EGP case, we also need an additional geometric inequality, which is related to the quadratic function, to prove this lemma. Fenchelôs inequality for the quadratic function $\frac{1}{2}\xi^2$ is:

$$\tfrac{1}{2}\xi^2 + \tfrac{1}{2}\eta^2 \geq \xi\eta,\ \forall \xi, \eta \in R \text{ with equality holding iff } \xi = \eta.$$

Fenchelôs inequality for the function $\frac{1}{2}y(\xi + d)^2$ with $y \geq 0$ (under the convention: $\eta^2/(2y) = 0$ for y=0=η; and =+∞ for y=0$\neq \eta$) is:

$$\text{(6.7)} \qquad \frac{y(\xi+d)^2}{2} + \frac{\eta^2}{2y} \geq (\xi + d)\eta,\ \forall \xi, \eta, d \in R, y \in R_+,$$

with equality holding iff $\eta = y(\xi + d)$ iff $\xi + d = y/\eta$, when $y > 0$.

Lemma 6.2.1 (Main Lemma of QGP) *If* $\boldsymbol{t}$ *is feasible for primal program* (QGP) *and* $(\boldsymbol{y}, \eta)$ *is feasible for dual program* (QGD)*, then* $\tilde{G}_0(\boldsymbol{t}) \geq \tilde{V}(\mathbf{y}, \eta)$

Moreover, under the same conditions, $\tilde{G}_0(\boldsymbol{t}) = \tilde{V}(\boldsymbol{y},\eta)$ *if and only if the following extremality conditions hold:*

(6.8) (a) $y_i = \begin{cases} \tilde{U}_i(\boldsymbol{t}) / \tilde{G}_0(\boldsymbol{t}), & i \in [0] \\ \lambda_k \tilde{U}_i(\boldsymbol{t}), & i \in [k], \quad k \in K \end{cases}$, and (b) $\eta_l = y_i \ln V_l(\boldsymbol{t}),\ l \in <i>,\ i \in I^+$

in which case $\boldsymbol{t}$ *is optimal for primal program* (QGP) *and* $(\boldsymbol{y},\eta)$ *is optimal for dual program* (QGD).

Proof: Let $(\boldsymbol{x},\boldsymbol{\xi}) = (A\boldsymbol{z}, B\boldsymbol{z})$, where $z_j = \ln t_j, \forall j \in J$. Then $(\boldsymbol{x},\boldsymbol{\xi}) \in \mathcal{P}$ is feasible for program (QGP)$_{x,\xi}$ and hence $\boldsymbol{x}\cdot\boldsymbol{y} + \boldsymbol{\xi}\cdot\eta = 0$. First, apply the geometric inequality to each problem function $\tilde{G}_k(\boldsymbol{t})$ in (QGP) to get:

$$\tilde{G}_0 \geq \tilde{G}_0 \cdot \prod_{k\in K} \tilde{G}_k^{\ \lambda_k} \geq \prod_{k\in K}\prod_{i\in[k]} (C_i\lambda_k / y_i)^{y_i} \cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}} \cdot \prod_{i\in I^+} e^{y_i H_i}, \quad \because \tilde{U}_i = C_i e^{x_i + H_i},\ H_i = \tfrac{1}{2}\sum_{l\in<i>} \ln^2 V_l$$

$$= \prod_{k\in K}\prod_{i\in[k]} (C_i\lambda_k / y_i)^{y_i} \cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}} \cdot \prod_{i\in I^+}\prod_{l\in<i>} \exp\left[\tfrac{y_i(\xi_l + d_l)^2}{2}\right], \quad \because y_i H_i = \sum_{l\in<i>} \tfrac{y_i(\xi_l+d_l)^2}{2}$$

Then apply a second geometric inequality (6.7) to the exp factor in the last expression to obtain a further lower bound:

$$\begin{aligned}
&\geq \prod_{k\in K}\prod_{i\in[k]} (C_i\lambda_k / y_i)^{y_i} \cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}} \cdot \prod_{i\in I^+}\prod_{l\in<i>} \exp\left[(\xi_l + d_l)\eta_l - (\eta_l^2 / 2y_i)\right] \\
&= \prod_{k\in K}\prod_{i\in[k]} (C_i\lambda_k / y_i)^{y_i} \cdot e^{\boldsymbol{x}\cdot\boldsymbol{y}} \cdot e^{\xi\cdot\eta} \cdot \prod_1^r D_l^{\eta_l} \cdot \exp\left[-\sum_{i\in I^+}\sum_{l\in<i>} \frac{\eta_l^2}{2y_i}\right],\ \because e^{d_l} = D_l, \\
&= \prod_{k\in K}\prod_{i\in[k]} (C_i\lambda_k / y_i)^{y_i} \cdot \prod_1^r D_l^{\eta_l} \cdot \exp\left[-\sum_{i\in I^+}\sum_{l\in<i>} \frac{\eta_l^2}{2y_i}\right], \quad \because \boldsymbol{x}\cdot\boldsymbol{y} + \boldsymbol{\xi}\cdot\eta = 0 \\
&= \tilde{V}(\boldsymbol{y},\eta)
\end{aligned}$$

and additional conditions for equality are $\eta_l = y_i(\xi_l + d_l) = y_i \ln V_l(\boldsymbol{t}),\ \forall l \in <i>,\ i \in I^+$, which are the same as conditions (b) in the lemma(6.8). Therefore, $\tilde{G}_0(\boldsymbol{t}) \geq \tilde{V}(\boldsymbol{y}, \eta)$ and equality holds iff the extremality conditions are satisfied. ■

This main lemma clearly extends the earlier main lemma of GP.

Corollary 6.2.2 (Weak Duality Theorem for (QGP))

1) *Always* $\infty \geq \inf QGP \geq \sup QGD \geq 0$

2) *When* (*QGP) and* (*QGD*) *are feasible,* $\infty > \inf QGP \geq \sup QGD > 0$.

Example 6.2.1 A (QGP) problem [Duffin et.al, p.214-5]

$$\min_{t>0} U_1 + U_2 \text{ s.t. } \ln^2 V_1 + \ln^2 V_2 \leq 1, \text{ where } U_1 = 4/\left(t_1 t_2 t_3\right),\ U_2 = 4t_2 t_3,\ V_1 = 2t_1 t_3,\ V_2 = t_1 t_2$$

Since the constraint is equivalent to: $e^{-1}\exp\left[\ln^2 V_1 + \ln^2 V_2\right] \leq 1$ with $U_3 = e^{-1}, <3> = \{1,2\}$, it is equivalent to a (QGP) problem with $n = 3, r = 2, m = 3$. ■

Example 6.2.2 [Fang and Rajasekera, 1987] revisited

$$(QPL) \quad \begin{cases} \inf\limits_{\mathbf{0}<t\in R^2} \tilde{G}_0(\boldsymbol{t}) = t_1 t_2 \cdot t_1^{50\ln t_1} + t_1^{-1} t_2^{-1} \\ s.t.\ G_1(\boldsymbol{t}) = 0.01 t_1^{1/2} + t_2 \leq 1, \end{cases},$$

In our QGP formulation, it is

$$(QGP) \quad \begin{cases} \inf\limits_{\mathbf{0}<t\in R^2} \tilde{G}_0(\boldsymbol{t}) = t_1 t_2 \cdot \exp\left(\tfrac{1}{2}\ln^2 V_1(\boldsymbol{t})\right) + t_1^{-1} t_2^{-1} \\ s.t.\ G_1(\boldsymbol{t}) = 0.01 t_1^{1/2} + t_2 \leq 1, \end{cases}$$

where $V_1(\boldsymbol{t}) = t_1^{10}, <1> = \{1\}$. So $n = 4,\ r = 1,\ m = 2$, the exponent matrix is $M = \begin{bmatrix} A \\ B \end{bmatrix}$, with $B = \begin{bmatrix} 10 & 0 \end{bmatrix}$, and A as is usual.

Its convex form is:

$$(QGP)_z \quad \begin{cases} \inf_{z \in R^2} \tilde{g}_0(\boldsymbol{z}) = \ln\left[\exp(z_1 + z_2 + \tfrac{1}{2}(10z_1)^2) + \exp(-z_1 - z_2)\right] \\ s.t.\ \ g_1(\boldsymbol{z}) = \ln\left[\exp(-2\ln 10 + \tfrac{1}{2}z_1) + \exp(z_2)\right] \le 0 \end{cases}$$

■

Dual to Primal conversion of optimum solutions

Suppose we have found an optimum solution $(\delta, \boldsymbol{v})$ for the dual program (QGD), and want to convert it to optimum solutions $\boldsymbol{t}$ for the primal program (QGP), believing that they have same optimum values $\tilde{G}_0(\boldsymbol{t}) = V(\delta, \boldsymbol{v}) := V$. We can derive from (Extremality Condition 2) that

(6.9)
$$\tilde{U}_i(\boldsymbol{t}) = \begin{cases} \delta_i V, \text{ for } i \in [0] \\ \delta_i / \lambda_k, \text{ for } i \in [k],\ k \in K, \text{ and } \lambda_k > 0 \end{cases}$$
$$B^i \boldsymbol{z} = v^i / \delta_i, \forall i \in I \text{ and } \delta_i > 0$$

Recall from that $\tilde{U}_i(\boldsymbol{t}) = \exp\{\tfrac{1}{2} | B^i \boldsymbol{z} |^2 + \boldsymbol{a}^i \boldsymbol{z} + c_i\}$. Taking logarithms of both sides of the first set of equations above and replacing $B^i \boldsymbol{z}$ using the last equation yields

$$\frac{|v^i|^2}{2\delta_i^2} + \boldsymbol{a}^i \boldsymbol{z} + c_i = \begin{cases} \ln(\delta_i V), & \text{for } i \in [0] \\ \ln\left(\delta_i / \lambda_k\right), & \text{for } i \in [k],\ k \in K,\ \delta_i > 0, \text{ and } \lambda_k > 0 \end{cases}$$

The last condition in (6.9) says $\forall l \in \langle i \rangle,\ i \in I,$ if $\delta_i > 0,\ \boldsymbol{b}^l \boldsymbol{z} = v_l / \delta_i$.

So all optimum solutions $\boldsymbol{z}$ for $(QGP)_z$ must satisfy the following system of linear equations:

(6.10)
$$\begin{cases} \sum_{j=1}^{m} a_{ij} z_j = \begin{cases} \ln(\delta_i V / C_i) - |v^i|^2 / 2\delta_i^2 & \text{for } i \in [0] \\ \ln\left(\delta_i / C_i \lambda_k\right) - |v^i|^2 / 2\delta_i^2, & \text{for } i \in [k],\ k \in K,\ \delta_i > 0, \text{ and } \lambda_k > 0 \end{cases} \\ \sum_{j=1}^{m} b_{lj} z_j = v_l / \delta_i, \text{ for } l \in \langle i \rangle,\ i \in I \text{ s.t. } \delta_i > 0 \end{cases}$$

Any solution $\boldsymbol{z}$ to the above system together with the given dual optimum solution $(\delta, \boldsymbol{v})$ will satisfy extremality condition 2 and $\tilde{g}_k(\boldsymbol{z}) = 0,\ \forall k \in K^+ =: \{k \in K \mid \lambda_k > 0\}$. If the coefficient matrix of the above system has rank m, e.g., if the optimum solution $(\delta, \boldsymbol{v})$ happens to have all $\delta_i>0$, and M has full column rank m, then the primal problem $(QGP)_z$ has a unique optimum solution determined by the above system. If not, the optimum solutions $\boldsymbol{z}$ can still be characterized by adjoining the equations: $\tilde{g}_k(\boldsymbol{z}) \le 0,\ \forall k \in K \setminus K^+$ into the above system.

Observe that for the (GP) case equation (6.10) easily reduces to that of [Duffin et.al. 1967, p. 185].

6.3 FIRST AND SECOND DUALITY THEOREMS OF QGP

Just as in the case of GP, we call a QGP problem *superconsistent* if all of its constraints can be satisfied strictly, i.e. there is a point $\boldsymbol{t}^0$ such that $\tilde{G}_k(\boldsymbol{t}^0) < 1$ for all k in K. We define the Lagrangian for $(QGP)_z$: $l(\boldsymbol{z}, \lambda) = \tilde{g}_0(\boldsymbol{z}) + \sum_{k \in K} \lambda_k \tilde{g}_k(\boldsymbol{z}),$ for $\lambda \in R_+^p$, and $\boldsymbol{z} \in R^m$.

This corresponds to the product function we used in the proof of the main lemma for QGP

$$L(\boldsymbol{t}, \lambda) = \exp l(\boldsymbol{z}, \lambda) = \tilde{G}_0(\boldsymbol{t}) \prod_{k=1}^{p} \tilde{G}_k(t)^{\lambda_k} .$$

Similarly, the following formulae include equation as a special case.

(6.11) $$\nabla \tilde{u}_i(\boldsymbol{z}) = (\boldsymbol{a}^i)^T + Q^i \boldsymbol{z}, \qquad \nabla \tilde{g}_k(\boldsymbol{z}) = \frac{1}{\tilde{G}_k(\boldsymbol{t})} \sum_{i \in [k]} \tilde{U}_i(\boldsymbol{t})[(\boldsymbol{a}^i)^T + Q^i \boldsymbol{z}]$$

Theorem 6.3.1 (First Duality theorem of QGP) *Suppose that primal program QGP is superconsistent. Then a point* $\boldsymbol{t}'$ *is a minimum solution for* (QGP) *if and only if there exists a* $\lambda' \in R_+^p$ *such that* $(\boldsymbol{z}', \lambda')$ *is a saddle point of the Lagrangian* $l(\boldsymbol{z}, \lambda)$ (*where* $\boldsymbol{z}' = \ln \boldsymbol{t}'$)

$$\max_{\lambda \ge 0} l(\boldsymbol{z}', \lambda) = l(\boldsymbol{z}', \lambda') = \min_{z} l(\boldsymbol{z}, \lambda'),$$

in which case the dual program (QGD) *also has a maximum solution* $(\delta', \boldsymbol{v}')$ *such that*

$$\min(QGP) = \tilde{G}_0(\boldsymbol{t}') = V(\delta', \boldsymbol{v}') = \max(QGD)$$

Moreover, the set of all such λ' and δ' is compact.

Proof Superconsistency of (QGP) means the existence of a Slater point $\boldsymbol{z}^0$ ($=ln\ \boldsymbol{t}^0$) for the convex program $(QGP)_z$. Thus $\boldsymbol{z}'$ is a minimum solution for $(QGP)_z$ if, and only if, there exist $\lambda' \in R_+^p$ such that $(\boldsymbol{z}', \lambda')$ form saddle point(s) for the Lagrangian (the set of all such λ' must be compact). This in turn implies that

$$\tilde{g}_0(\boldsymbol{z}') = l(\boldsymbol{z}', 0) \le l(\boldsymbol{z}', \lambda') \Rightarrow\ 0 \le \sum\nolimits_{k\in K} \lambda_k' \tilde{g}_k(\boldsymbol{z}') \le 0\ \Rightarrow \lambda_k' \tilde{g}_k(\boldsymbol{z}') = 0,\ \forall k \in K .$$

The last condition is equivalent to (a) being satisfied at $\boldsymbol{t}'$ and λ'. Setting $\lambda_0' = 1$, we define $(\delta', \boldsymbol{v}')$ using (b) and (c)

(6.12) $$\delta_i' = \lambda_k' \tilde{U}_i(\boldsymbol{t}') / \tilde{G}_k(\boldsymbol{t}'),\ \forall i \in [k], k \in \{0\} \cup K,\ \text{and}\ \ v^i{}' = \delta_i' B^i \boldsymbol{z}',\ \ \forall i \in I .$$

The solution $(\delta', \boldsymbol{v}')$ thus constructed satisfies the extremality condition 1. We continue to show that it is also dual feasible. From the previous equation it is clear that

$$\forall k \in \{0\} \cup K,\ \sum\nolimits_{i\in[k]} \delta_i' = \lambda_k',\ \ \delta_i' > 0, \forall i \in [k],\ \text{if}\ \lambda_k' > 0;\ \text{and}\ \delta_i' = 0, \forall i \in [k],\ \text{if}\ \lambda_k' = 0$$

So the set of all such δ' is also compact. Since $l(\boldsymbol{z}', \lambda') = \min_{z} l(\boldsymbol{z}, \lambda')$ and $l(\boldsymbol{z}, \lambda') = \tilde{g}_0(\boldsymbol{z}) + \sum\nolimits_{k\in K} \lambda_k' \tilde{g}_k(\boldsymbol{z})$ is a differentiable convex function in $\boldsymbol{z}$, we should have $\nabla_z l(\boldsymbol{z}', \lambda') = 0$, that is

$$0 = \sum_{k=0}^{p} \lambda_k' \nabla \tilde{g}_k(\boldsymbol{z}') = \sum_{k=0}^{p} \frac{\lambda_k'}{\tilde{G}_k(t')} \sum_{i\in[k]} \tilde{U}_i(\boldsymbol{t}')[(\boldsymbol{a}^i)^T + (B^i)^T B^i \boldsymbol{z}'],\ \text{by equation (6.11)}$$

$$= \sum_{k=0}^{p} \sum_{i \in [k]} [(\boldsymbol{a}^i)^T \delta_i' + (B^i)^T \boldsymbol{v}^i{}'] \text{, by equation (6.12)}$$

Hence $\qquad A^T \delta' + B^T \boldsymbol{v}' = 0$, by equation (i)

So by lemma 4.3, $(\delta', \boldsymbol{v}')$ is an optimum solution for the dual program (*QGD*) and it satisfies

$$\min(QGP) = \tilde{G}_0(\boldsymbol{t}') = V(\delta', \boldsymbol{v}') = \max(QGD) \qquad \blacksquare$$

So we see that the orthogonally condition in the constraints of (*QGD*) is equivalent to the partial gradient w.r.t. to $\boldsymbol{z}$ of the Lagrangian $l(\boldsymbol{z},\lambda)$ of the primal program $(QGP)_z$ being equal to zero. The above theorem is slightly more detailed than the earlier result in [Duffin et.al, 1967, p.80]. In addition, we have also provided a partial converse to this first duality theorem. The above proof also suggests the following.

Primal to Dual conversion of optimum solutions

Suppose we have found an optimum solution $\boldsymbol{t}'$ for the primal program (*QGP*), and want to convert it to optimum solutions $(\delta', \boldsymbol{v}')$ for the dual program (*QGD*), believing that they have the same optimum values $\tilde{G}_0(\boldsymbol{t}') = V(\delta', \boldsymbol{v}') := V$. First, we need to find the optimum Lagrange multipliers λ', and this can be done by solving the following linear system $\sum_{k \in K'} \lambda_k \nabla \tilde{g}_k(\boldsymbol{z}') = -\nabla \tilde{g}_0(\boldsymbol{z}')$ for $\lambda_k \geq 0,\ k \in K' =: \{k \in K \mid \tilde{g}_k(\boldsymbol{z}') = 0\}$ and set $\lambda_k = 0, \forall k \in K \setminus K'$.

Specifically, we are to solve for λ_k :

$$\text{(6.13)} \qquad \begin{aligned} &\sum_{k \in K'} \lambda_k \sum_{i \in [k]} \tilde{U}_i(t')[a_{ij} + \boldsymbol{b}_j^i \cdot u^i{}'] = -V^{-1} \sum_{i \in [0]} \tilde{U}_i(t')[a_{ij} + \boldsymbol{b}_j^i \cdot \boldsymbol{u}^i{}'],\ \forall j \in J \\ &\lambda_k \geq 0,\ \forall k \in K',\ \text{where } \boldsymbol{u}^i{}' = B^i \boldsymbol{z}',\ \text{and } \boldsymbol{b}_j^i = j^{th} \text{ column of } B^i. \end{aligned}$$

This system has *m* equations in at most *p* nonnegative variables, and can be easily solved by any method that finds a starting solution to an *LP* program. Once this is done, $(\delta', \boldsymbol{v}')$ is readily available from (6.12). Observe that for the (*GP*) case equation easily specializes to that of [Duffin et.al. 1967, p. 186].

We say that a dual program QGD is *canonical*, if it has a feasible solution $(\delta, \boldsymbol{v})$ with δ>0. The following fact is obvious.

Lemma 6.3.2 *If a uni-variate function* $f(x) =: ax^2 + bx + c$, *with* $a \geq 0$, *is bounded above* $\forall x \geq 0$, *then* $a = 0,\ b \leq 0$.

Theorem 6.3.3 (Second Duality theorem of QGP) *Suppose that primal program QGP is consistent and dual program QGD is canonical. Then the minimum set of* $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$ *is nonempty compact. The minimum set of* $(QGP)_{\boldsymbol{z}}$ *is also nonempty compact, provided that M is of full column rank.*

Proof Suppose that primal program $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$ has a recession direction, that is, a nonzero vector (x, u) in ▯ such that all of its problem functions $g_k(x^{[k]}, u^{[k]}) = \ln\left[\sum_{i\in[k]} \exp(c_i + x_i + \frac{1}{2}|u^i|^2)\right]$ are bounded above on the half-line $\{(0,0) + t(x,u) \mid t \geq 0\}$, say, by a real b. Then

$$g_k(tx^{[k]}, tu^{[k]}) = \ln\left[\sum_{i\in[k]} \exp(c_i + tx_i + \tfrac{1}{2}t^2 |u^i|^2)\right] \leq b,\ \forall t \geq 0,\ \forall k \in \{0\} \cup K$$

Hence by equation (3.16)

$$\phi_i(t) =: c_i + tx_i + \tfrac{1}{2}t^2 |u^i|^2 \leq b,\ \forall t \geq 0,\ \forall i \in [k],\ \forall k \in \{0\} \cup K$$

By Lemma 4.6, one must have $u^i = 0,\ x_i \leq 0,\ \forall i \in I \Rightarrow u = 0,\ x \leq 0$ and this x is nonzero. However, since the dual program (*QGD*) is canonical, it has a feasible solution $(\delta, \boldsymbol{v}) \in$ ▯ with δ>**0**. It follows that $0 = (\boldsymbol{x}, \boldsymbol{u}) \cdot (\boldsymbol{\delta}, \boldsymbol{v}) = \boldsymbol{x} \cdot \boldsymbol{\delta} < 0$, a contradiction. Hence primal program $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$ has no recession direction, and its optimum solution set is a nonempty compact subset of ▯. If

M is of full column rank, since $(\boldsymbol{x}, \boldsymbol{u}) = M\boldsymbol{z}$, the optimum solution set of $(QGP)_z$ is also a nonempty compact set in R^m . ■

So we see that the existence of a feasible solution $(\boldsymbol{\delta}, \boldsymbol{v})$ with $\delta > \mathbf{0}$ in (QGD) prevents the primal program $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$ from having a recession direction: a vector $(\boldsymbol{x}, \boldsymbol{u})$ in ⍰ with $0 \neq \boldsymbol{x} \leq \mathbf{0},\ \boldsymbol{u} = \mathbf{0}$. We shall show that the converse of this is also true. Note that this theorem is more detailed than that of [Duffin et.al, 1967, p.81].

We now state a classical result, due to [Tucker, 1956], which is more general than that of [Duffin et.al, 1967, p.20, p.168].

Proposition 6.3.4 (Existence of Tucker solutions) *Of the two dual systems of homogeneous linear relations*:

I. $\quad A\boldsymbol{z} \leq 0, \quad B\boldsymbol{z} = 0 \qquad\qquad$ II. $\quad A^T \boldsymbol{y} + B^T v = 0, \quad \boldsymbol{y} \geq 0$

there exist a pair of solutions $\boldsymbol{z}^*$ *and* $(\boldsymbol{y}^*, \boldsymbol{v}^*)$ such that $\boldsymbol{y}^* - A\boldsymbol{z}^* > 0$ (*component-wise*).

By this proposition, if (QGD) is not canonical, $\boldsymbol{y}^*$ must have some zero component, say, $y_i^* = 0$, then $\boldsymbol{x}^* = A\boldsymbol{z}^*$ has a negative component $x_i^* < 0$. This gives rise to a recession direction in $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$: a vector $(\boldsymbol{x}^*, 0)$ in ⍰ with $x^* \leq 0$, and $x_i^* < 0$. For this direction, consider the i^{th} term in $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$ over the ray (half-line) $(x,u) + s(x^*,0) = (x + sx^*, u),\ s \geq 0$ starting from any feasible point $(\boldsymbol{x}, \boldsymbol{u})$

(6.14) $\quad \tilde{U}_i(x_i + sx_i^*, u^i) = \exp(c_i + x_i + sx_i^* + \tfrac{1}{2}|u^i|^2) = \tilde{U}_i(x_i, u^i) \cdot (\exp x_i^*)^s \to 0, \text{ as } s \to \infty.$

And for all terms:

(6.15) $$\tilde{U}_i(x_i + sx_i^*, u^i) = \tilde{U}_i(x_i, u^i) \cdot (\exp x_i^*)^s \begin{cases} < \tilde{U}_i(x_i, u^i), \ \forall s > 0, \text{ if } x_i^* < 0 \\ = \tilde{U}_i(x_i, u^i), \ \forall s > 0, \text{ if } x_i^* = 0 \end{cases}$$

That is, over this ray the i^{th} term in $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$ can be driven to zero without causing any other terms to increase. Consequently, over this ray the problem function $\tilde{G}_k(t)$ containing this term will strictly decrease, while all other problem functions stay non-increasing and hence maintain feasibility. We call such a term *vanishing term*. If it appears in the objective function, the primal infimum clearly cannot be attained; if it only appears in a constraint function, the minimum set of the primal problem, if non-empty, must contain the above ray, in view of (6.15), and hence is unbounded. In any case, we have shown that if (*QGD*) is not canonical, the minimum set of $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$ is either empty or unbounded.

Theorem 6.3.5 (Converse of the Second Duality theorem of QGP) *Suppose that primal program QGP is consistent and dual program QGD is not canonical. Then the minimum set of* $(QGP)_{\boldsymbol{x},\boldsymbol{u}}$ *is not nonempty and bounded, i.e., it is either empty or unbounded.*

The above result already holds in the *GP* case [Abrams, 1975].

Classification of QGP Problems

By proposition 4.8, $\tilde{U}_i(\boldsymbol{t})$ is a vanishing term if and only if $\delta_i = 0$, whenever $(\boldsymbol{\delta},\boldsymbol{v})$ is a feasible solution of (*QGD*). In this case, we call δ_i a *null* variable, and the pair of programs $(QGP)_{x,u}$ and (*QGD*) *degenerate*. Degenerate programs are just the opposite of canonical programs. They have vanishing term(s) in the primal problem and null variable(s) in the dual problem. When every term in the primal objective is a vanishing term, we call such a program *totally degenerate*. Its primal program, if consistent, must have unbounded infimum, and its dual program must be infeasible, because all of its dual variables in [0] are null and hence cannot satisfy the normality condition: $\lambda_0 = 1$. Properties of these programs are particularly simple. If a degenerate program is not totally degenerate, i.e., at least one objective term is not a vanishing term, it is called *mildly degenerate*. Such a program can be reduced to a canonical program by dropping all of its vanishing terms in the primal problem and deleting all of its null variables in the dual problem. It is clear that this reduction procedure will not affect the value of its dual objective function, nor its dual feasibility. On the primal side, however, things are not so simple, and this is where the concept of *sub-consistency* comes into play. However, we do not explore further in this direction here. We end this section by pointing out a strong version of Theorem 4.7.

Theorem 6.3.6 (Strong version of Second Duality theorem of QGP) *Suppose that dual program* (*QGD*) *is canonical and has a finite positive supremum. Then primal program* (*QGP*) *has a minimum solution* $\mathbf{t}^*$ *such that*

$$\tilde{G}_0(\mathbf{t}^*) = \min\left(QGP\right) = \sup\left(QGD\right).$$

That this theorem is stronger than Theorem 4.7 is because its assumption ñ(*QGD*) *is canonical and has a finite positive supremum*ò is weaker than the previous assumption ñ*primal program QGP is consistent and dual program QGD is canonical*ò by the main lemma of *QGP*. This theorem was first proved by [Fang and Rajasekera, 1987].

7.0 DUALITY THEORY OF l_PGP AND QGP

In this chapter we consider an important branch of (CGP), viz., (l_pGP), which arises when each function h_l is defined by $h_l(\xi_l) = \frac{1}{p_l} | \xi_l |^{p_l}$ with $p_l > 1, \forall l \in L$. Of course, when each $p_l = 2$ this program becomes a (QGP) program.

7.1 PROBLEM FORMULATIONS OF l_PGP AND l_PGD

Primal program (l_pGP)

$$(l_p GP) \ \ \inf_{\mathbf{0} < \boldsymbol{t} \in R^m} \tilde{G}_0(\boldsymbol{t}) \ s.t. \ \tilde{G}_k(\boldsymbol{t}) \le 1, \ k \in K,$$

with

$$\tilde{G}_k(\boldsymbol{t}) := \sum_{i \in [k]} \left[U_i(\boldsymbol{t}) \cdot \exp \sum_{l \in <i>} \tfrac{1}{p_l} | \ln V_l(\boldsymbol{t}) |^{p_l} \right] \tag{7.1}$$

where $p_l > 1, \ \forall l \in <i>, \forall i \in I^+$.

Primal program (QGP)

$$(QGP) \ \ \inf_{\mathbf{0} < \boldsymbol{t} \in R^m} \tilde{G}_0(\boldsymbol{t}) \ s.t. \ \tilde{G}_k(\boldsymbol{t}) \le 1, \ k \in K,$$

with

$$\tilde{G}_k(\boldsymbol{t}) := \sum_{i \in [k]} \left[U_i(\boldsymbol{t}) \cdot \exp \sum_{l \in <i>} \tfrac{1}{2} | \ln V_l(\boldsymbol{t}) |^{2} \right] \tag{7.2}$$

If we take the logarithm of all problem functions in (l_pGP) and express them in the logarithm of the design variables t_j, we arrive at an equivalent program:

A convex form of (l_pGP) in the variables $z_j =: \ln t_j,\, j \in J$

$$(l_pGP)_z \ \inf_{z \in R^m} \tilde{g}_0(\boldsymbol{z}) \ s.t.\ \tilde{g}_k(\boldsymbol{z}) \le 0,\ \forall k \in K,$$

where the problem functions

$$(7.3) \qquad \tilde{g}_k(\boldsymbol{z}) := \ln\left\{\sum_{i\in[k]} \exp\left[c_i + \boldsymbol{a}^i \boldsymbol{z} + \sum_{l\in<i>} \tfrac{1}{p_l} |\boldsymbol{b}^l \boldsymbol{z} + d_l|^{p_l}\right]\right\},\ \forall k \in K$$

are clearly convex and differentiable (recall that $U_i = e^{\boldsymbol{a}^i z + c_i}$, $V_l = e^{\boldsymbol{b}^l z + d_l}$, $d_l =: \ln D_l$, and $\boldsymbol{b}^l$ is the l^{th} row of the matrix B).

A convex form of (QGP) in the variables $z_j =: \ln t_j,\, j \in J$

$$(QGP)_z \ \inf_{z \in R^m} \tilde{g}_0(\boldsymbol{z}) \ s.t.\ \tilde{g}_k(\boldsymbol{z}) \le 0,\ \forall k \in K,$$

where the problem functions

$$(7.4) \qquad \tilde{g}_k(\boldsymbol{z}) := \ln\left\{\sum_{i\in[k]} \exp\left[c_i + \boldsymbol{a}^i \boldsymbol{z} + \sum_{l\in<i>} \tfrac{1}{2} |\boldsymbol{b}^l \boldsymbol{z} + d_l|^{2}\right]\right\},\ \forall k \in K$$

It is worth noting that if the coefficient $\frac{1}{p_l}$ that appears in the formula (7.1) for $\tilde{G}_k(\boldsymbol{t})$ or in the formula (7.3) for $\tilde{g}_k(\boldsymbol{z})$ is any other positive constant, say s_l, the program is still convertible to an l_pGP problem simply by multiplying the l^{th} row [b^l d_l] of the data matrix $\begin{bmatrix} B & d \end{bmatrix}$ by $\left(p_l s_l\right)^{1/p_l}$, since $s_l\, |\boldsymbol{b}^l \boldsymbol{z} + d_l|^{p_l} = \frac{1}{p_l}\left|\left(p_l s_l\right)^{1/p_l}\left(\boldsymbol{b}^l \boldsymbol{z} + d_l\right)\right|^{p_l}$.

The above convex form (l_pGP)$_{\mathbf{z}}$ can be turned into a variables-separated convex form by further changes of variables:

A GGP form of (l_pGP) in the variables $x_i =: \boldsymbol{a}^i \boldsymbol{z},\ i \in I$, and $\xi_l =: \boldsymbol{b}^l \boldsymbol{z},\ l \in L$.

$$(l_p GP)_{\boldsymbol{x},\xi} \quad \inf_{(\boldsymbol{x},\xi)\in R^{n+r}} \tilde{g}_0(\boldsymbol{x},\boldsymbol{\xi}) \ s.t.\ \tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi}) \le 0,\ \forall k \in K,\ (\boldsymbol{x},\boldsymbol{\xi}) \in \mathscr{P},$$

where $\mathscr{P}$, the *primal space* of this program, is the column space of the exponent matrix *M*, and the variables in these convex problem functions

$$\text{(7.5)} \qquad \tilde{g}_k(\boldsymbol{z}) = \ln\left\{\sum\nolimits_{i\in[k]} \exp\left[x_i + c_i + \sum\nolimits_{l\in<i>} \tfrac{1}{p_l} |\xi_l + d_l|^{p_l}\right]\right\} = \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]})$$

are separated for different $k \in \breve{K}$.

A GGP form of (QGP) in the variables $x_i =: \boldsymbol{a}^i \boldsymbol{z},\ i \in I$, and $\xi_l =: \boldsymbol{b}^l \boldsymbol{z},\ l \in L$.

$$(QGP)_{\boldsymbol{x},\xi} \quad \inf_{(\boldsymbol{x},\xi)\in R^{n+r}} \tilde{g}_0(\boldsymbol{x},\boldsymbol{\xi}) \ s.t.\ \tilde{g}_k(\boldsymbol{x},\boldsymbol{\xi}) \le 0,\ \forall k \in K,\ (\boldsymbol{x},\boldsymbol{\xi}) \in \mathscr{P},$$

where $\mathscr{P}$, the *primal space* of this program, is the column space of the exponent matrix *M*, and the variables in these convex problem functions

$$\text{(7.6)} \qquad \tilde{g}_k(\boldsymbol{z}) = \ln\left\{\sum\nolimits_{i\in[k]} \exp\left[x_i + c_i + \sum\nolimits_{l\in<i>} \tfrac{1}{2} |\xi_l + d_l|^{2}\right]\right\} = \tilde{g}eo(\boldsymbol{x}^k + \boldsymbol{c}^k, \boldsymbol{\xi}^{[k]} + \boldsymbol{d}^{[k]})$$

are separated for different $k \in \breve{K}$.

For the uni-variate l_p function $h(\xi) := \frac{1}{p}|\xi|^p$ with $p > 1$, we let q be the *conjugate* of *p*, which means $\quad 1/p + 1/q = 1 \ (\Leftrightarrow p + q = pq \Leftrightarrow (p-1)(q-1) = 1)$.

and find its conjugate function $h^*(\eta) := \frac{1}{q}|\eta|^q$, thus for $y \ge 0,\ yh^*(\eta / y) = \dfrac{|\eta|^q}{qy^{q-1}}$.

The function $h: R \to R$ is strictly convex, and continuously differentiable with derivative function $h'(\xi) = |\xi|^{p-1} \operatorname{sgn} \xi$ (strictly) increasing from *R* onto *R*, and thus possesses a continuous (strictly) increasing inverse function from *R* onto *R*. In other words, the derivative function h' is a bi-continuous bijection (i.e. a homeomorphism) between *R* and itself; likewise, its conjugate function $h^*: R \to R$, being a function of the same type as h with $q > 1$, and its derivative function $h^*'(\eta) = |\eta|^{q-1} \operatorname{sgn} \eta$ have exactly the same properties mentioned above. This pair of functions h and h^* of course are *lsc*, proper convex. They satisfy a conjugate inequality:

$$\frac{1}{p}|\xi|^p + \frac{1}{q}|\eta|^q \geq \xi\eta, \ \forall \xi, \eta \in R \tag{7.7}$$

where equality holds exactly when one of the following 3 equivalent conditions holds true:

$$\begin{aligned} &(a)\ \eta = h'(\xi) = |\xi|^{p-1} \operatorname{sgn} \xi \\ &(b)\ \xi = h^*'(\eta) = |\eta|^{q-1} \operatorname{sgn} \eta \\ &(c)\ \xi\eta = |\xi|^p = |\eta|^q \end{aligned} \tag{7.8}$$

Since $(a) \Leftrightarrow (b)$, h' and h^*' are each other's inverse functions.

From (5.6) we can derive

Dual program (l_pGD) (*Convention*: $|\eta_l|^{q_l} / (q_l y_i^{q_l - 1}) = 0$ for $y_i = 0 = \eta_l$ and $= \infty$ for $y_i = 0 \neq \eta_l$)

$$(l_pGD) \quad \begin{cases} \sup\limits_{(\mathbf{y},\boldsymbol{\eta}) \in R^n \times R^r} \tilde{V}(\boldsymbol{y}, \eta) := \prod\limits_{k \in \breve{K}} \prod\limits_{i \in [k]} \left(\dfrac{C_i \lambda_k}{y_i} \right)^{y_i} \cdot \exp\left[\sum\limits_{i \in I^+} \sum\limits_{l \in <i>} \left(d_l \eta_l - \dfrac{|\eta_l|^{q_l}}{q_l y_i^{q_l - 1}} \right) \right] \\ s.t. \qquad \lambda_0 = 1, \qquad \text{(Normality Condition)} \\ \quad \sum\limits_{i=1}^{n} a_{ij} y_i + \sum\limits_{l=1}^{r} b_{lj} \eta_l = 0, \ j \in J \quad \text{(Orthogonality Conditions)} \\ \quad y_i \geq 0, \forall i \in I, \ \sum\limits_{i \in [k]} y_i = \lambda_k, \ \forall k \in \breve{K} \\ \quad y_i = 0 \Rightarrow \eta_l = 0, \ \forall l \in <i>, \ \forall i \in I^+ \qquad (*) \end{cases} \tag{7.9}$$

Dual program (QGD) (*Convention*: $|\eta_l|^2 / (2y_i) = 0$ for $y_i = 0 = \eta_l$ and $= \infty$ for $y_i = 0 \neq \eta_l$)

$$(QGD) \quad \begin{cases} \sup\limits_{(\mathbf{y},\boldsymbol{\eta}) \in R^n \times R^r} \tilde{V}(\boldsymbol{y}, \eta) := \prod\limits_{k \in \breve{K}} \prod\limits_{i \in [k]} \left(\dfrac{C_i \lambda_k}{y_i} \right)^{y_i} \cdot \exp\left[\sum\limits_{i \in I^+} \sum\limits_{l \in <i>} \left(d_l \eta_l - \dfrac{|\eta_l|^2}{2 y_i} \right) \right] \\ s.t. \quad \text{exactly the same constraints as in the above } (l_p\text{GD}) \end{cases} \tag{7.10}$$

7.2 MAIN LEMMA OF l_pGP

The main lemma of (l_pGP) follows from the main lemma of CGP (Lemma 5.2.1).

Lemma 7.2.1 (Main Lemma of l_pGP) *If* $\boldsymbol{t}$ *is feasible for primal program* (l_pGP) *and* $(\boldsymbol{y},\eta)$ *is feasible for dual program* (l_pGD)*, then* $\tilde{G}_0(\boldsymbol{t}) \geq \tilde{V}(\boldsymbol{y},\eta)$

Moreover, under the same conditions, $\tilde{G}_0(\boldsymbol{t}) = \tilde{V}(\boldsymbol{y},\eta)$ *if, and only if, one of the following two sets of equivalent conditions holds:*

$$(a)\begin{cases}\tilde{G}_k(\boldsymbol{t})^{\lambda_k} = 1,\ k \in K, \\ y_i\tilde{G}_k(\boldsymbol{t}) = \lambda_k\tilde{U}_i(\boldsymbol{t}),\ i \in [k],\ k \in K\end{cases} \tag{7.11}$$

(Extremality condition 1)

$$(b)\ \eta_l = y_i\,|\ln V_l(\boldsymbol{t})|^{p_l-1}\,\mathrm{sgn}(\ln V_l(\boldsymbol{t})),\ l \in <i>,\ i \in I^+$$

$$(c)\ y_i = \begin{cases}\tilde{U}_i(\boldsymbol{t})/\tilde{G}_0(\boldsymbol{t}), & i \in [0] \\ \lambda_k\tilde{U}_i(\boldsymbol{t}), & i \in [k],\ k \in K\end{cases} \tag{7.12}$$

(Extremality condition 2)

$$(d)\ \eta_l = y_i\,|\ln V_l(\boldsymbol{t})|^{p_l-1}\,\mathrm{sgn}(\ln V_l(\boldsymbol{t})),\ l \in <i>,\ i \in I^+$$

(*Condition* (b) *is equivalent to* $(b)'$ $\ln V_l(\boldsymbol{t}) = \mathrm{sgn}\,\eta_l \cdot |\eta_l / y_i|^{q_l-1}$, if $y_i > 0,\ \forall l \in <i>,\ i \in I^+$) *in which case* $\boldsymbol{t}$ *is optimal for primal program* (l_pGP) *and* $(\boldsymbol{y},\eta)$ *is optimal for dual program* (l_pGD). (Recall from Lemma 2.1.0 that the above conditions (a) and (c) are equivalent.)

Lemma 7.2.2 (Main Lemma of QGP) *If* $\boldsymbol{t}$ *is feasible for primal program* (QGP) *and* $(\boldsymbol{y},\eta)$ *is feasible for dual program* (QGD)*, then* $\tilde{G}_0(\boldsymbol{t}) \geq \tilde{V}(\boldsymbol{y},\eta)$

Moreover, under the same conditions, $\tilde{G}_0(\boldsymbol{t}) = \tilde{V}(\boldsymbol{y},\eta)$ *if and only if one of the following two sets of equivalent conditions holds:*

(7.13) $(a)\begin{cases}\tilde{G}_k(\boldsymbol{t})^{\lambda_k} = 1, \ k \in K, \\ y_i\tilde{G}_k(\boldsymbol{t}) = \lambda_k \tilde{U}_i(\boldsymbol{t}), \ i \in [k], k \in K\end{cases}$ (Extremality condition 1)

$(b) \ \eta_l = y_i \ln V_l(\boldsymbol{t}), \ l \in \langle i \rangle, \ i \in I^+$

(7.14) $(c) \ y_i = \begin{cases}\tilde{U}_i(\boldsymbol{t})/\tilde{G}_0(\boldsymbol{t}), & i \in [0] \\ \lambda_k \tilde{U}_i(\boldsymbol{t}), & i \in [k], \ k \in K\end{cases}$ (Extremality condition 2)

$(d) \ \eta_l = y_i \ln V_l(\boldsymbol{t}), \ l \in \langle i \rangle, \ i \in I^+$

(*Condition* (b) *is equivalent to* $(b)'$ $\ln V_l(\boldsymbol{t}) = \eta_l / y_i$, if $y_i > 0, \ \forall l \in \langle i \rangle, \ i \in I^+$) *in which case **t** is optimal for primal program* (QGP) *and* $(\boldsymbol{y}, \eta)$ *is optimal for dual program* (QGD). (Recall from Lemma 2.1.0 that the above conditions (a) and (c) are equivalent.)

Corollary 7.2.3 (Weak Duality Theorem for (l_pGP))

3) *Always* $\infty \geq \inf l_p GP \geq \sup l_p GD \geq 0$

4) *When* (l_p*GP) and* (l_p*GD*) *are feasible,* $\infty > \inf l_p GP \geq \sup l_p GD > 0$.

Corollary 7.2.4 (Weak Duality Theorem for (QGP))

5) *Always* $\infty \geq \inf QGP \geq \sup QGD \geq 0$

6) *When* (QGP*) and* (QGD) *are feasible,* $\infty > \inf QGP \geq \sup QGD > 0$.

7.3 FIRST AND SECOND DUALITY THEOREMS OF l_PGP

The first duality theorem of (l_pGP) follows from the First Duality Theorem of CGP (Theorem 5.3.1). The statement for the first duality theorem of (QGP) is almost exactly the same, so it is not repeated here.

Theorem 7.3.1 (First Duality Theorem of l_pGP) *Suppose that primal program* (l_pGP) *is superconsistent. Then the following 3 conditions are equivalent*:

1) ***t'*** *is a minimal solution to* (l_pGP).

2) *There exists a vector* $\lambda' \in R_+^p$ *for* ***z'*** (*where* ***z'****= ln* ***t'***) *such that* $(\boldsymbol{z}', \lambda')$ *forms a saddle point of* $\tilde{l}(\boldsymbol{z}, \lambda)$.

3) *There exists a vector* $\lambda' \in R_+^p$ for ***z'*** (*where* ***z'****= ln* ***t'***) *such that* $(\boldsymbol{z}', \lambda')$ *satisfies the KKT conditions for* (l_pGP)$_z$:

$$(a)\ \lambda_k' \geq 0,\ \tilde{g}_k(\boldsymbol{z}') \leq 0,\ \lambda_k' \tilde{g}_k(\boldsymbol{z}') = 0, \forall k \in K$$
$$(b)\ \nabla_z \tilde{l}(\boldsymbol{z}', \lambda') = \sum_{k \in \check{K}} \lambda_k' \nabla \tilde{g}_k(\boldsymbol{z}') = \boldsymbol{0},\ where\ \lambda_0' = 1$$

in which case the set of all such vectors $\boldsymbol{\lambda}'$ *is a non-empty compact convex subset of* R_+^p, *and the dual program* (l_pGD) *also has a maximum solution* $(\boldsymbol{y}', \eta')$ *such that*

$$\min(l_p\text{GP}) = \tilde{G}_0(\boldsymbol{t}') = \tilde{V}(\boldsymbol{y}', \eta') = \max(l_p\text{GD}) \qquad \text{(Perfect duality)}$$

A theorem of the alternatives from linear programming duality theory

Of the following two linear systems exactly one has a solution (*where* $M = \begin{bmatrix} A \\ B \end{bmatrix} : (n+r) \times m$):

(I) *Find* $\boldsymbol{z}$ *with* $0 \neq \boldsymbol{Az} \leq 0,\ \boldsymbol{Bz} = 0$ (II) *Find* $\begin{pmatrix} \boldsymbol{y} \\ \eta \end{pmatrix}$ with $\boldsymbol{y} > \boldsymbol{0}$, and $A^T \boldsymbol{y} + B^T \eta = \boldsymbol{0}$

We say that program (l_pGD) is *canonical* if system (II) has a solution when M is the programôs composite exponent matrix.

Theorem 7.3.2 (Second Duality Theorem of l_pGP) *Suppose that primal program* (l_pGP) *is consistent. Then the minimum set of program* (l_pGP)$_z$ *is non-empty and bounded if and only if its dual program* (l_pGD) *is canonical.*

Proof By assumption, program $(l_p\text{GP})_z$ is consistent. For the function $h(\xi) = \frac{1}{p}|\xi|^p$ with $p > 1$, one has $h^{\infty}(\xi) = \lim_{s\to\infty} s^{p-1}\frac{1}{p}|\xi|^p = i_0(\xi)$. Thus

$$\tilde{g}_k^{\infty}(\boldsymbol{z}) = \max_{i\in[k]}\left[\boldsymbol{a}^{\boldsymbol{i}}\boldsymbol{z} + \sum\nolimits_{l\in<i>} i_0(\boldsymbol{b}^{\boldsymbol{l}}\boldsymbol{z})\right] = \max(\boldsymbol{A}^k\boldsymbol{z}) + i_{\boldsymbol{0}}(\boldsymbol{B}^{[k]}\boldsymbol{z})$$

and system (5.19) becomes

$$\begin{aligned}\tilde{g}_k^{\infty}(\boldsymbol{z}) \le 0, \forall k \in \breve{K} &\Leftrightarrow \max(\boldsymbol{A}^k\boldsymbol{z}) \le 0, \boldsymbol{B}^{[k]}\boldsymbol{z} = \boldsymbol{0}, \forall k \in \breve{K} \\ &\Leftrightarrow \boldsymbol{Az} \le \boldsymbol{0}, \boldsymbol{Bz} = \boldsymbol{0}\end{aligned}$$

Since M is of full column rank, it is one-to-one, hence $\boldsymbol{z} \ne 0 \Rightarrow \boldsymbol{Mz} \ne 0$. Therefore by theorem 2.5.10,

The minimum set of program $(l_p\text{GP})_z$ is non-empty and bounded

$\Leftrightarrow$ The system $\tilde{g}_k^{\infty}(\boldsymbol{z}) \le 0,\ k \in \breve{K}$ has no non-trivial solution in $\boldsymbol{z}$

$\Leftrightarrow$ System (I) above has no solution

$\Leftrightarrow$ System (II) above has a solution, i.e. program (l_pGD) is canonical ■

Theorem 7.3.3 (Second Duality Theorem of QGP) *Suppose that primal program* (QGP) *is consistent. Then the minimum set of program* $(\text{QGP})_z$ *is non-empty and bounded if and only if its dual program* (QGD) *is canonical.*

8.0 CONCLUSION AND SOME DIRECTIONS OF FUTURE RESEARCH

In this chapter we summarize the contributions of this research and suggest some directions for future work.

8.1 SUMMARY OF CONTRIBUTIONS

In this thesis, the traditional GP models have been extended to the more general Exponential GP models, and then to the even more general Composite GP models. The latter generalization includes (QGP), (l_pGP) and (EGP) as important special cases. It was also shown that all of these are special cases of Petersonös GGP models, for which only a Main Lemma is available, and First and Second duality theorem are not defined or proved.

For the Main Lemma, multiple direct proofs have been provided, with a second set of equivalent Extremality Conditions in each extended case.

For the First duality theorem, it was shown that a superconsistent primal program (CGP) has a minimal solution $\boldsymbol{t}'$ if, and only if, there exists a vector $\lambda' \in R_+^p$ such that $(\boldsymbol{z}', \lambda')$ forms a saddle point of the Lagrangian $\tilde{l}(\boldsymbol{z}, \lambda)$ and if, and only if, it satisfies the KKT conditions for $(CGP)_z$. It was also shown that in this case the set of all such Lagrange multiplier vectors $\boldsymbol{\lambda}$ is a non-empty compact convex subset of R_+^p, whereas the original theorem is a ñIfé thenò statement and merely showed the existence of $\boldsymbol{\lambda}$ <u>for the (GP) case.</u>

For the Second duality theorem, it was shown that for the important special cases of (QGP), (l_pGP) and (EGP), *the minimum set of a consistent primal program is non-empty and bounded if, and only if, its dual program is canonical*, whereas the original theorem merely showed the existence of a primal solution <u>for the (GP) case</u>, when its dual program is canonical.

For the Sensitivity Analysis of the extended GP model, it was shown that for an optimal solution $\boldsymbol{z}^*$ of $(\text{CGP})_z$, the following constitiute the sensitivity information:

$$\frac{\partial \tilde{g}_0(\boldsymbol{z}^*)}{\partial c_i} = y_i^*, \quad \frac{\partial \tilde{g}_0(\boldsymbol{z}^*)}{\partial d_l} = \eta_l^*, \frac{\partial \tilde{g}_0(\boldsymbol{z}^*)}{\partial b_k} = -\lambda_k^*,$$
$$\frac{\partial \tilde{g}_0(\boldsymbol{z}^*)}{\partial a_{ij}} = y_i^* z_j^*, \quad \frac{\partial \tilde{g}_0(\boldsymbol{z}^*)}{\partial b_{lj}} = \eta_l^* z_j^*, \text{ where } b_k = \ln B_k,$$

where B_k is the new right hand side replacing 1 for the k^{th} constraint in (CGP).

A computational procedure for the ôDual to primal conversionô of optimal solutions, when there is no duality gap, i.e. $\tilde{G}_0(\boldsymbol{t}) = \tilde{V}(\mathbf{y}, \eta)$ was also derived.

8.2 SOME DIRECTIONS FOR FUTURE RESEARCH

The first extension would be to prove a strong version of the First Duality Theorem of (CGP): *If a primal program* (CGP) *is superconsistent and has a finite infimum*, *then the dual program* (CGD) *has a maximum solution* $(\boldsymbol{y}', \eta')$ *such that* $\inf(\text{CGP}) = \max(\text{CGD}) = \tilde{V}(\boldsymbol{y}', \eta')$.

Following this one could explore the Second duality theorem: *The minimum set of a consistent primal program is non-empty and bounded if and only if its dual program is canonical*, to see if it holds for other branches of CGP. So far, it is only shown to hold for (EGP), (QGP), and (l_pGP). On the other hand, it might be possible to show for the last three cases that, *if the dual program is canonical and has a finite positive supremum, then a primal minimum solution exists and there is no duality gap.*

Finally, it is worth searching for more examples of practical applications of (CGP), and developing computational algorithms exploiting the duality theory.

APPENDIX A

MATHEMATICAL PROOFS OF SOME INEQUALITIES

Geometric Programming duality hinges upon a so-called *geometric inequality*, which is a slight generalization of the classical (weighted) arithmetic-geometric mean (AGM) inequality, whose strict version is equivalent to the strict convexity of the exponential function *exp x* and to the strict concavity of the logarithmic function *ln x*.

A.1 ON STRICTLY CONVEX FUNCTIONS

We say a real-valued function f on a non-empty convex set $C \subset R^n$ is *strictly convex* on C if $\forall x \neq y \in C, \forall t \in (0,1), \quad f[(1-t)x + ty] < (1-t)f(x) + tf(y)$, (Rockafellar, 1970, p.253). A function g is *strictly concave* on C, if ï g is strictly convex on C (i.e. if the above defining inequality holds in reverse direction). Strict convexity is essentially a 1-dimensional property; behavior with respect to line segments is all that matters. *A function f is strictly convex if and only if* (*iff*) *it is strictly convex relative to every line.*

Proposition 1 *A real-valued function f on a non-empty convex set C is strictly convex on C iff*

$\forall x_i \in C, \forall t_i \in (0,1)$, i=1,⋯,n; $n \geq 2$, with $\sum_{i=1}^{n} t_i = 1$, one has

(A1.1) $$f(\sum_{i=1}^{n} t_i x_i) \leq \sum_{i=1}^{n} t_i f(x_i)$$

and equality holds when all the x_i's are equal.

First Proof of Necessity: The ōifô clause is trivial. No suppose *f* is strictly convex on C; we shall show (A1.1) by induction on *n*. The case for *n*=2 is the definition of strict convexity of *f*. Assume that it is true for some $n = k \geq 2$. For the case $n = k+1$, observe that

$$f(\sum_{i=1}^{k+1} t_i x_i) = f(\tilde{t}_k \tilde{x}_k + t_{k+1} x_{k+1}), \text{ where } \tilde{t}_k =: \sum_{i=1}^{k} t_i,\ \tilde{x}_k =: \sum_{i=1}^{k} (t_i / \tilde{t}_k) x_i$$

$$\leq \tilde{t}_k f(\tilde{x}_k) + t_{k+1} f(x_{k+1}), \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (a)$$

$$\leq \tilde{t}_k \sum_{i=1}^{k} (t_i / \tilde{t}_k) f(x_i) + t_{k+1} f(x_{k+1}), \cdots\cdots\cdots\cdots (b)$$

$$= \sum_{i=1}^{k+1} t_i f(x_i),$$

whence $f(\sum_{i=1}^{k+1} t_i x_i) \leq \sum_{i=1}^{k+1} t_i f(x_i) \cdots\cdots\cdots\cdots\cdots (c)$ follows.

Itōs trivial that when the x_iōs are all equal, equality holds in (A1.1). On the other hand, under the condition that $x_1, \cdots, x_k, x_{k+1}$ are not all equal, if $\tilde{x}_k \neq x_{k+1}$, then inequality (a) is strict; if $\tilde{x}_k = x_{k+1}$, then $x_1, \cdots, x_k$ are not all equal, inequality (b) must be strict. Thus inequality (c) must be strict in either case. ■

Second Proof of Necessity: Let *f* be strictly convex on C, we shall show (A1.1) by induction on *n*. The case for *n*=2 is the definition of strict convexity of *f*. Assume that it is true for some $n = k \geq 2$. For the case $n = k+1$, observe that

$$f(\sum_{i=1}^{k+1} t_i x_i) = f(\sum_{i=1}^{k-1} t_i x_i + s_k y_k), \text{ where } s_k =: t_k + t_{k+1},\ y_k =: (t_k x_k + t_{k+1} x_{k+1})/s_k$$

$$\le \sum_{i=1}^{k-1} t_i f(x_i) + s_k f(y_k), \text{ with equality holding iff } x_1 = ... = x_{k-1} = y_k$$

$$\le \sum_{i=1}^{k-1} t_i f(x_i) + s_k[(\tfrac{t_k}{s_k}) f(x_k) + (\tfrac{t_{k+1}}{s_k}) f(x_{k+1})], \text{ with equality holding iff } x_k = x_{k+1}$$

$$= \sum_{i=1}^{k+1} t_i f(x_i),$$

whence $f(\sum_{i=1}^{k+1} t_i x_i) \le \sum_{i=1}^{k+1} t_i f(x_i)$ follows, and equality holds iff $x_1 = ... = x_k = x_{k+1}$.

(since $x_1 = ... = x_{k-1} = y_k$, and $x_k = x_{k+1} \Leftrightarrow x_1 = ... = x_{k-1} = x_k = x_{k+1}$) ■

If we relax the weight requirements in the above proposition from $t_i > 0$ to $\delta_i \ge 0$, then the strict convexity of *f* can be rephrased as follows. A real-valued function *f* defined on a non-empty convex set $C \subset R^n$ is strictly convex iff $\forall x_i \in C, \forall \delta_i \ge 0,\ i = 1, \cdots, n,$ with $\sum_{i=1}^{n} \delta_i = 1,$ where $2 \le n \in N$, one has

$$f(\sum_{i=1}^{n} \delta_i x_i) \le \sum_{i=1}^{n} \delta_i f(x_i),$$

and equality holds exactly when all the $x_i$ôs *for which* $\delta_i > 0$ are equal. Geometrically, this implies that if $\bar{x} =: \sum_{i=1}^{n} \delta_i x_i \in \text{co}\{x_1, \ldots, x_n\}$, the convex hull of a set of *n distinct* points $x_1, \ldots, x_n$, but $\bar{x}$ is not any of the generating points $x_1, \ldots, x_n$, then $\bar{x}$ is in the convex hull of a set of at least two *distinct* points among $x_1, \ldots, x_n$, and $f(\bar{x}) < \sum_{i=1}^{n} \delta_i f(x_i)$ [Duffin et.al. 1967, p. 56, exercise 2(a)].

It is easy to identify strictly convex functions when they are also differentiable [Rockafellar and Wets, 2004, p.46].

Theorem 2 (1-dimensional derivative tests). *For a differentiable function f on an open interval* $I \subset R$*, each of the following conditions is equivalent to f being strictly convex on I*:

(a) *f*Nj *is increasing on I: any points* $x_0 < x_1$ *in I satisfy* $f'(x_0) < f'(x_1)$.

(b) $f(y) - f(x) > f'(x)(y-x), \ \forall x, y \in I$ with $x \neq y$.

The following condition is only sufficient, but not necessary for f being strictly convex (*The strictly convex function f(x) =*x^4*, on R with* $f''(0) = 0$ *serves as a counter-example*.)

(c) $f''(x) > 0, \ \forall x \in I$ (*Assuming twice differentiability*).

Example 1

- $f(x) = e^{ax}$ is strictly convex on R when $a \neq 0$.
- $f(x) = \ln(bx)$ is strictly concave on $(0,\infty)$ when $b > 0$.
- $f(x) = x^r$ on $(0,\infty)$ is $\begin{cases} \text{strictly convex, if } r > 1 \text{ or } r < 0. & (\text{e.g. } x\sqrt{x},\ 1/\sqrt{x}) \\ \text{strictly concave, if } 0 < r < 1. & (\text{e.g. } \sqrt{x}) \end{cases}$

Solution. Since $f'(x) = rx^{r-1} = re^{(r-1)\ln x}$ is $\begin{cases} \text{increasing on } (0,\infty), & \text{if } r > 1 \text{ or } r < 0. \\ \text{decreasing on } (0,\infty), & \text{if } 0 < r < 1. \end{cases}$ ■

Example 2 [Duffin et.al. 1967, p. 56, exercise 2(b)]

$\forall x, y, z \in R, \left|\frac{x}{2} + \frac{y}{3} + \frac{z}{6}\right| < \sqrt{\ln(\frac{1}{2}e^{x^2} + \frac{1}{3}e^{y^2} + \frac{1}{6}e^{z^2})}$, unless *x*=*y*=*z*.

Solution Since the function $f(t) =: e^{t^2}$ has second derivative $f''(t) = 2e^{t^2}(1+2t^2) > 0, \ \forall t,$ it is strictly convex; hence $\forall x, y, z \in R, \ \ \exp(\frac{x}{2} + \frac{y}{3} + \frac{z}{6})^2 < \frac{1}{2}e^{x^2} + \frac{1}{3}e^{y^2} + \frac{1}{6}e^{z^2}$, unless *x*=*y*=*z*; i.e.

$$\left|\frac{x}{2} + \frac{y}{3} + \frac{z}{6}\right| < \sqrt{\ln(\frac{1}{2}e^{x^2} + \frac{1}{3}e^{y^2} + \frac{1}{6}e^{z^2})}, \text{ unless } x=y=z.$$ ■

For any two functions *f* and *g* we define their maximum and minimum as follows:

$$(f \vee g)(x) := \max[f(x), g(x)], \qquad (f \wedge g)(x) := \min[f(x), g(x)]$$

Theorem 3 *If f and g are strictly convex, so is f*∨*g.*

A.2 THE AGM INEQUALITY AND THE GEOMETRIC INEQUALITY

The classical inequality relating the arithmetic mean to the geometric is stated and proved in any book on inequalities, e.g. [Beckenbach and Bellman, 1965], and [Hardy, Littlewood, and Polya, 1952, pp.17-18]. We show below that the strict version of this classical (weighted) AGM inequality is actually equivalent to the strict convexity of the exponential function *exp x* and to the strict concavity of the logarithmic function *ln x*.

Theorem 3 (Strict AGM Inequality)

$\forall T_i > 0, \forall \delta_i > 0,\ i = 1, \cdots, n$, with $\sum_{i=1}^{n} \delta_i = \lambda$, where $2 \le n \in N$,

$$\left(\sum_{i=1}^{n} \delta_i T_i \right)^{\lambda} \ge \prod_{i=1}^{n} T_i^{\delta_i} \cdot \lambda^{\lambda},$$

where equality holds exactly when all the T_i's are equal. When $\lambda = 1$, this reduces to the more familiar equivalent form **[**Duffin and Peterson 1966, p.1316, Lemma 0]:

$$\sum_{i=1}^{n} \delta_i T_i \ge \prod_{i=1}^{n} T_i^{\delta_i},$$

where equality holds exactly when all the T_i's are equal.

Proof The convexity of the exponential function e^x applied to the points $x_i = \ln T_i$ with weights $t_i = \delta_i / \lambda$ means:

$$\exp\left(\frac{1}{\lambda}\sum_{i=1}^{n}\delta_i \ln T_i\right) \le \frac{1}{\lambda}\sum_{i=1}^{n}\delta_i \exp(\ln T_i)$$

$$\Leftrightarrow \prod_{i=1}^{n} T_i^{\delta_i/\lambda} \le \frac{1}{\lambda}\sum_{i=1}^{n}\delta_i T_i \Leftrightarrow \prod_{i=1}^{n} T_i^{\delta_i}\cdot\lambda^{\lambda} \le \left(\sum_{i=1}^{n}\delta_i T_i\right)^{\lambda}.$$

The concavity of the logarithmic function $ln\, x$ applied to the points $x_i = T_i$ with weights $t_i = \delta_i / \lambda$ means:

$$ln\left(\frac{1}{\lambda}\sum_{i=1}^{n}\delta_i T_i\right) \ge \frac{1}{\lambda}\sum_{i=1}^{n}\delta_i ln T_i$$

$$\Leftrightarrow \frac{1}{\lambda}\sum_{i=1}^{n}\delta_i T_i \ge \prod_{i=1}^{n} T_i^{\delta_i/\lambda} \Leftrightarrow \left(\sum_{i=1}^{n}\delta_i T_i\right)^{\lambda} \ge \prod_{i=1}^{n} T_i^{\delta_i}\cdot\lambda^{\lambda}$$

The condition for equality in both cases is the same: all the T_i*'s* are equal. ■

Remark If the assumption $T_i > 0$ in the above theorem is relaxed to $T_i \ge 0$, the conclusion still holds. Furthermore, one can also relax the assumption $\delta_i > 0$ to $\delta_i \ge 0$ and obtain:

Theorem 4 (Relaxed AGM Inequality)

$\forall T_i \ge 0, \forall \delta_i \ge 0,\ i = 1,\cdots,n, with \sum_{i=1}^{n}\delta_i = \lambda,\ \ where\ 2 \le n \in N$

$$\left(\sum_{i=1}^{n}\delta_i T_i\right)^{\lambda} \ge \prod_{i=1}^{n} T_i^{\delta_i}\cdot\lambda^{\lambda},$$

where equality holds exactly when all the T_i's for which $\delta_i > 0$ are equal. For $\lambda = 1$, this reduces to the more familiar form:

$$\sum_{i=1}^{n}\delta_i T_i \ge \prod_{i=1}^{n} T_i^{\delta_i}$$

Convention: Since $\lim_{a \downarrow 0} a^a = 1$, we define $0^0 = 1$, and thus $(u/0)^0 = u^0 = 1$, for $u \geq 0$.

Proof: (*Case* 1) When $\lambda = 0$, $\delta_i = 0, \forall i \in I := \{1, ..., n\}$, and the inequality holds trivially as ñ1=1ò, and the condition for equality also holds trivially since there is no positive δ_i.

(*Case* 2) When $\lambda > 0$, by applying Theorem 3 and the above remark to the index set $I^+ =: \{i \in I | \delta_i > 0\} \neq \phi$, one obtains

$$\left(\sum_{i=1}^{n} \delta_i T_i\right)^\lambda = \left(\sum_{i \in I^+} \delta_i T_i\right)^\lambda \geq \prod_{i \in I^+} T_i^{\delta_i} \cdot \lambda^\lambda = \prod_{i=1}^{n} T_i^{\delta_i} \cdot \lambda^\lambda$$

Thus the desired inequality follows. The condition for equality is: all the $T_i$ôs for which $i \in I^+$ (i.e., $\delta_i > 0$) are equal. ■

We can slightly generalize the above weighted AGM Inequality to obtain the following

Theorem 5 (Geometric Inequality)

$\forall U_i \geq 0, \forall \delta_i \geq 0, i = 1, \cdots, n,$ *with* $\sum_{i=1}^{n} U_i = G,\ \sum_{i=1}^{n} \delta_i = \lambda,$ *one has*

$$\left(\sum_{i=1}^{n} U_i\right)^\lambda \geq \lambda^\lambda \cdot \prod_{i=1}^{n} (U_i/\delta_i)^{\delta_i}$$

where equality holds iff $U_i \lambda = \delta_i G, \forall i = 1, \cdots, n$ *, i.e., iff* U_i's and δ_i's *are in proportion.*

When $\lambda = 1$*, this becomes*

$$\sum_{i=1}^{n} U_i \geq \prod_{i=1}^{n} (U_i/\delta_i)^{\delta_i}.$$

<u>Note</u>: The original version in [Duffin and Peterson 1966, p.1316, Lemma 1] requires $U_i > 0, \forall i$.

Proof: (*Case* 1) When $\lambda = 0$, $\delta_i = 0, \forall i \in I := \{1, ..., n\}$, and the inequality holds trivially as ñ1=1ò, and the condition for equality also holds trivially as ñ0=0ò.

(*Case* 2) When $\lambda > 0$, define $I^+ =: \{i \in I | \delta_i > 0\} \neq \phi$, and $T_i =: U_i/\delta_i, \forall i \in I^+$. Observe that

$$\left(\sum_{i\in I} U_i\right)^{\lambda} \geq \left(\sum_{i\in I^+} U_i\right)^{\lambda} = \left(\sum_{i\in I^+} \delta_i T_i\right)^{\lambda} \geq \prod_{i\in I^+} T_i^{\delta_i} \cdot \lambda^{\lambda}, \text{ by (1.16),}$$

$$= \prod_{i\in I^+} (U_i/\delta_i)^{\delta_i} \cdot \lambda^{\lambda} = \prod_{i\in I} (U_i/\delta_i)^{\delta_i} \cdot \lambda^{\lambda}$$

Hence, one has $\left(\sum_{i\in I} U_i\right)^{\lambda} \geq \prod_{i\in I} (U_i/\delta_i)^{\delta_i} \cdot \lambda^{\lambda}$ and this is an equality iff

$$\begin{cases} U_i = 0, \ \forall i \notin I^+ \\ \exists T \to T_i = T, \ \forall i \in I^+ \end{cases} \Leftrightarrow \exists T \to \begin{cases} U_i = \delta_i T, \ \forall i \in I \\ G = \lambda T \end{cases} \Leftrightarrow \ U_i \lambda = \delta_i G, \forall i \in I, \text{ since } \lambda > 0. \qquad \blacksquare$$

APPENDIX B

THE STIMULANT OF THE BIRTH OF GP – ZENER'S DISCOVERY

The root of GP lies in some rudimentary work by Zener (1961) for optimizing a posynomial. He first observed the following fact.

Lemma 1 *Let* t^* *be a local minimum (or local maximum) point of the posynomial* $G(t)$. *Then the following linear system*

(B1.1)
$$\begin{cases} \sum_{i=1}^{n} a_{ij}\delta_i = 0, \ \forall j = 1, \cdots, m & \text{(Orthogonality conditions)} \\ \sum_{i=1}^{n} \delta_i = 1, & \text{(Normality condition)} \end{cases}$$

has a positive solution, namely, δ^* *defined by*

(B1.2)
$$\delta_i^* =: U_i(t^*) / G(t^*), \ \forall i .$$

Together, t^* *and* δ^* *satisfy*:

(B1.3)
$$G(t^*) = V(\delta^*), \text{ where } V(\delta^*) =: \prod_{i=1}^{n} \left(\frac{C_i}{\delta_i^*} \right)^{\delta_i^*}$$

and

(B1.4)
$$\sum_{j=1}^{m} a_{ij} ln\, t_j^* = ln(\frac{U_i(t^*)}{C_i}) = ln(\frac{\delta_i^* V(\delta^*)}{C_i}), \ \forall \text{i}=1, \cdots, n. \text{ (dual-to-primal converson)}$$

Proof 1) The orthogonality condition is simply a necessary condition for local optimality of $G(t^*)$ at t^* (It is in fact also a sufficient condition for global minimality, since optimizing a posynomial $G(t)$ amounts to optimizing a certain convex function $g(z)$):

$$\frac{\partial G(t^*)}{\partial t_j}=0 \Rightarrow \frac{\partial G(t^*)}{\partial t_j}\cdot\frac{t_j^*}{G(t^*)}=0 \Rightarrow \sum_{i=1}^{n}\frac{a_{ij}U_i(t^*)}{G(t^*)}=0, \text{ by (1.1.5)} \Rightarrow \sum_{i=1}^{n}a_{ij}\delta_i^*=0, \ \forall j=1,\cdots,m,$$

It can be written more compactly in matrix form as $A^t\delta^*=0$, and geometrically, this means the vector δ^* is orthogonal to all the columns of the matrix A. The normality condition is obvious from the definition of δ^*. This proves (B1.1).

2) (B1.3) follows from (B1.1):

$$G(t^*)=\prod_{i=1}^{n}[G(t^*)]^{\delta_i^*}=\prod_{i=1}^{n}\left(\frac{U_i(t^*)}{\delta_i^*}\right)^{\delta_i^*}=\prod_{i=1}^{n}\left(\frac{C_i}{\delta_i^*}\right)^{\delta_i^*}\cdot\prod_{i=1}^{n}(\prod_{j=1}^{m}t_j^{*a_{ij}})^{\delta_i^*}=V(\delta^*)\cdot\prod_{j=1}^{m}(t_j^*)^{\sum_{i=1}^{n}a_{ij}\delta_i^*}=V(\delta^*)$$

3) (B1.4) follows from the definitions of $U_i(t^*)$, δ_i^* and (B1.3). ■

In essence, Zener discovered that at an optimizing point ***t****** for $G(t)$, the vector *δ*** defined in Theorem 1 is a positive solution to system (B1.1) and together with ***t****, satisfy (B1.3) and (B1.4). *Conversely*, a positive solution to (1) does not necessarily satisfy (0) for any given ***t****, and hence does not necessarily satisfy (2) and (3). A later duality result (*Main Lemma for Unconstrained GP*) by Duffin showed that for a positive vector ***t****, and a nonnegative solution *δ*** to (1), (0) and (2) are actually equivalent. *However,* when *n=m+1* <u>and the linear system (1) is non-singular</u>, it has a unique solution, <u>and if it happens to be positive</u>, it must be the solution defined by (0), and hence must satisfy (2) and (3). Thus *Zener found the optimum value* $G(t^*)$ *may be expressed by a simple formula (2) in this special case n-m-1=0, and the optimizing point* t^* *may also be calculated by solving another system of linear equations* (3), whose solution is unique iff A has full column rank (this is certainly the case if the system (1) is non-singular). In GP literature the number, *n-m-1*, is called *the degree of difficulty* of the GP problem.

There is one linear relation among the equations in (3), though. To see this, we first write (3) in matrix form $Az=b^*$, where $A=(a_{ij})$, $z_j = \ln t_j, b_i^* =: \ln\left(\delta_i^* / C_i\right) + v^*$, and $v^* =: \ln V(\delta^*)$, then any solution δ^* given by (1) satisfies

$$\delta^* \cdot Az = A^t \delta^* \cdot z = 0 \cdot z = 0, \text{ and } \delta^* \cdot b^* = \sum_{i=1}^n \delta_i^* \ln(\delta_i^* / C_i) + (\sum_{i=1}^n \delta_i^*) v^* = -v^* + v^* = 0.$$

Thus $\delta^* \cdot (Az - b^*) = 0$, for a nonzero δ^*, which means at least one linear equation is a consequence of the remaining linear equations.

Remarkably, the optimum value $G(\boldsymbol{t}^*)$ was found in (2) before the optimizing point $\boldsymbol{t}^*$ was determined in (3) (On one rare occasion when n=2, m=1, however, this order can be reversed, as we see in Chapter 2.) It is also obvious that in this case the relative contribution δ_i^* of each term does not depend on the cost coefficients C_i, since δ^* is uniquely determined by the matrix A alone.

Letôs illustrate Zenerôs procedure with a three-term two-variable unconstrained GP:

Example 1 $\inf_{t>0} U_1(t) + U_2(t) + U_3(t)$, where $U_i(t) = C_i t_1^{a_{i1}} t_2^{a_{i2}}$, i=1,2,3.

1) In the following, define D_1, D_2, D_3, where $D(\cdot,\cdot)$ denotes the determinant function of row vectors,

$$D_1 = \begin{vmatrix} 1 & 1 & 1 \\ 0 & a_{21} & a_{31} \\ 0 & a_{22} & a_{32} \end{vmatrix} = \begin{vmatrix} a_{21} & a_{31} \\ a_{22} & a_{32} \end{vmatrix} = D(a^2, a^3), \quad D_2 = \begin{vmatrix} 1 & 1 & 1 \\ a_{11} & 0 & a_{31} \\ a_{12} & 0 & a_{32} \end{vmatrix} = \begin{vmatrix} a_{31} & a_{11} \\ a_{32} & a_{12} \end{vmatrix} = D(a^3, a^1),$$

$$D_3 = \begin{vmatrix} 1 & 1 & 1 \\ a_{11} & a_{21} & 0 \\ a_{12} & a_{22} & 0 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} = D(a^1, a^2), \quad D = \begin{vmatrix} 1 & 1 & 1 \\ a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{vmatrix} = D_1 + D_2 + D_3$$

We assume that all D_i>0 orall D_i<0 so that, by Cramerῶs rule, the following system

$$\begin{cases} \delta_1 + \quad \delta_2 + \quad \delta_3 = 1 \\ a_{11}\delta_1 + a_{21}\delta_2 + a_{31}\delta_3 = 0 \\ a_{12}\delta_1 + a_{22}\delta_2 + a_{32}\delta_3 = 0 \end{cases} \quad \text{has a unique solution} \quad \delta^* = \frac{(D_1, D_2, D_3)}{D} > 0$$

2) Compute the dual function value $V(\delta^*)$ which equals the minimum value $G(t^*)$

$$G^* =: G(t^*) = V(\delta^*) =: \prod_{i=1}^{3}\left(\frac{C_i}{\delta_i^*}\right)^{\delta_i^*} = \prod_{i=1}^{3}\left(\frac{C_i}{D_i/D}\right)^{D_i/D} = \prod_{i=1}^{3}\left(\frac{C_i}{D_i}\right)^{D_i/D} D := V^*$$

Calculate $U_i^* = D_i V^* / D,$ and $x_i^* =: \ln(U_i^* / C_i), \forall i = 1,2,3.$

3) Solve $\begin{cases} a_{11}z_1^* + a_{12}z_2^* = x_1^* \\ a_{21}z_1^* + a_{22}z_2^* = x_2^* \end{cases}$

to get $(z_1^*, z_2^*) = \dfrac{1}{D_3}\left(\begin{vmatrix} x_1^* & a_{12} \\ x_2^* & a_{22} \end{vmatrix}, \begin{vmatrix} a_{11} & x_1^* \\ a_{21} & x_2^* \end{vmatrix}\right) = \dfrac{1}{D_3}\left(a_{22}x_1^* - a_{12}x_2^*, a_{11}x_2^* - a_{21}x_1^*\right)$

and $(t_1^*, t_2^*) = (e^{(a_{22}x_1^* - a_{12}x_2^*)/D_3}, e^{(a_{11}x_2^* - a_{21}x_1^*)/D_3}) = \left(\left[\left(\dfrac{U_1^*}{C_1}\right)^{a_{22}}\left(\dfrac{U_2^*}{C_2}\right)^{-a_{12}}\right]^{1/D_3}, \left[\left(\dfrac{U_1^*}{C_1}\right)^{-a_{21}}\left(\dfrac{U_2^*}{C_2}\right)^{a_{11}}\right]^{1/D_3}\right)$

APPENDIX C

EXTENDED REAL-VALUED FUNCTIONS AND SEQUENCES

Notation and terminology

When dealing with optimization and duality, we often find it more convenient to work with functions that are defined on the entire space R^n, and are extended-real-valued, that is: $f : R^n \to \overline{R}$, where $\overline{R} := R \cup \{\pm\infty\} = [-\infty, +\infty]$ is the familiar linearly ordered extended real number system where any two elements x, y are comparable, i.e. $x \le y$ or $y \le x$, together with 2 additional conventions: $\infty - \infty = \infty = -\infty + \infty$, $0 \cdot (\pm\infty) = 0 = (\pm\infty) \cdot 0$. The first rule, termed the *inf-addition* rule, is not symmetric, because we orient toward minimization. The following are immediate consequences of this rule:

Fact 1 *Suppose* $S = a_1 + a_2 + ... + a_n$, *where each* $a_i \in R^*$. *Then*

a) *If one of the* a_i *equals* + ∞, *then S* =+ ∞.

b) *If no* a_i *equals* + ∞, *but some* a_j *equals* -∞, *then S*= -∞.

c) *Otherwise, S is finite.*

Fact 2 *For all* $a, b \in R^*$, *the condition a* ≤ *b is equivalent to each of the following statements*:

1. *b – a* Ó0, 2. *– a* Ó- *b*, 3. *a*+*c* ≤ *b*+*c*, ∀ *c*∈*R,*

4. $\lambda a \le \lambda b$, ∀ $\lambda \in (0,\infty)$.

Note that for $c = \pm\infty$, $c - c$ is not 0, but equals $+\infty$.

Max and Min

A number a in R^* is said to be a *lower bound* for a subset C in R^*, if $a \le x,\ \forall x \in C$. The *infimum* of C is the greatest lower bound (glb) for C. By this definition, ī Ð is a lower bound for any subset C in R^*, and any number a in R^* is a lower bound for the empty set ϕ, so $\inf \phi = +\infty$.

Fact 3 *For any subset C of R*,*

$$\inf(C \setminus \{\infty\}) = \inf C = \inf(C \cup \{\infty\}).$$

We say that a set C in R^* is *bounded below* if it has a finite lower bound. For example, C is not bounded below if $-\infty \in C$. The set C is *bounded* if it is both bounded below and above.

Fact 4 *For any subset C of R*,*

a) $\inf C = \infty$ *iff* $C \subset \{\infty\}$ b) $\inf C = -\infty$ *iff C is not bounded below*

c) inf *C is finite iff C is bounded below and* $C \cap R \neq \infty$.

For any function $f : R^n \to R^*$ and any subset A of R^*, the e*ffective domain* of f is $F = \operatorname{dom} f := \{x \in R^n \mid f(x) < \infty\}$, the *infimum* of f over A is $\alpha = \inf_A f := \inf f(A)$, and the *minimum set* of f over A is $\arg\min f_A := \{x \in A \cap F \mid f(x) = \alpha\}$. By Fact 3 above, we have

Fact 5 *For any function* $f : R^n \to R^*$ *and any subset A of R*,*

$$\inf_A f = \inf f(A \cap F) = \inf\{f(x) \mid x \in A, f(x) < \infty\}.$$

We say that a function f is *bounded below* on a set A if its image set $f(A)$ is bounded below. For example, f is not bounded below on A if $f(x) = -\infty$ for some x in A. The function f is *bounded* on A if its image set $f(A)$ is bounded. By Fact 4 above, we have

Fact 6 *For any function* $f : R^n \to R^*$ *and any subset A of R*,*

a) $\inf_A f = \infty$ *iff* $f \equiv \infty$ *on A* b) $\inf_A f = -\infty$ *iff f is not bounded below on A*

c) $\inf_A f$ *is finite iff f is bounded below on A and* $f(x) \in R$ for some x in A.

A sequence $\{x_i\}_{i\in N}$ in $\bar{R}$ is a function from $N = \{1,2,3,\cdots\}$ to $\bar{R}$. It is said to *converge* to *a number c in* $\bar{R}$ if $i \to \infty$ implies $x_i \to c$, with the obvious interpretation being made when $c = \infty$ or -∞. A vector sequence $\{\boldsymbol{x}^i\}_{i\in N}$ in R^n is a function from N to R^n. A subsequential limit of this sequence is called a *cluster point*. This sequence converges to a point $\boldsymbol{x}$ in R^n iff it is bounded and has $\boldsymbol{x}$ as its only cluster point. A function $f : R^n \to \bar{R}$ is said to *converge* to a limit α in $\bar{R}$ *at infinity* if $|x| \to \infty \Rightarrow f(x) \to \alpha$ (where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^n$).

Lower and Upper Limits of *R-valued Sequences**

Let $\{x_i\}_{i\in N}$ be a sequence in R^*, i.e. a function $x : i \to x_i = x(i)$ from $N = \{1,2,3,\cdots\}$ to R^*. Construct two new sequences: $\{y_k = \inf_{i\geq k} x_i\}$ and $\{z_k = \sup_{i\geq k} x_i\}$. Then $\{y_k\}$ is *non-decreasing*: $y_k \leq y_{k+1},\ \forall k$, $\{z_k\}$ is *non-increasing*: $z_k \geq z_{k+1},\ \forall k$, and $y_k \leq x_k \leq z_k,\ \forall k$ so that

$$y_1 \leq y_i \leq y_{i\vee j} \leq z_{i\vee j} \leq z_j \leq z_1,\ \forall i, \forall j \text{, where } i \vee j := \max\{i, j\}.$$

Hence

$$y_1 \leq \sup_i y_i \leq \inf_j z_j \leq z_1 .$$

Define the *lower* and *upper limits* of $\{x_i\}$ respectively as:

$$\underline{\lim}_{i\to\infty} x_i := \sup_k y_k = \sup_{k\in N} \inf_{i\geq k} x_i \qquad \text{and} \qquad \overline{\lim}_{i\to\infty} x_i := \inf z_k = \inf_{k\in N} \sup_{i\geq k} x_i .$$

Fact 7 *For any sequence* $\{x_i\}_{i\in N}$ *in* R^*,

$$\inf_i x_i \leq \underline{\lim}_{i\to\infty} x_i \leq \overline{\lim}_{i\to\infty} x_i \leq \sup_i x_i$$

Definition *We say that a number c in* R^* *is a limit of a sequence* $\{x_i\}_{i\in N}$ *in* R^* *and write*

$$\lim_i x_i = c \text{ or } x_i \to c\text{, } \textit{if } \underline{\lim}_{i\to\infty} x_i = c = \overline{\lim}_{i\to\infty} x_i .$$

A sequence in R^* is said to be *monotone* if it is either non-decreasing or non-increasing. Any monotone sequence in R^* has its limit in R^*. Specifically,

Fact 8 *For any sequence* $\{x_i\}_{i\in N}$ *in* R^*,

a) $x_i \to \sup_i x_i$, *if* $\{x_i\}$ *is non-decreasing.*

b) $x_i \to \inf_i x_i$, *if* $\{x_i\}$ *is non-increasing.*

Fact 9 *For any sequence* $\{x_i\}_{i\in N}$ *in* R^*,

a) $x_i \to \infty$ iff $\forall M \in R,\ \exists k \in N,\ i \geq k \Rightarrow x_i \geq M$

b) $x_i \to -\infty$ iff $\forall M \in R,\ \exists k \in N,\ i \geq k \Rightarrow x_i \leq M$

c) $x_i \to c \in R$ iff $\forall \varepsilon > 0,\ \exists k \in N,\ i \geq k \Rightarrow | x_i - c | \leq \varepsilon$ (hence $x_i \in R$).

Fact 10

a) Let $\varnothing \neq C \subset R^*$. *If* $\alpha = \inf C$, *then* $\exists \{x_i\}$ in C, $x_i \to \alpha$.

b) *Let* $f : R^n \to R^*$ *and* $\varnothing \neq A \subset R^n$. *If* $\alpha = \inf_A f$, *then* $\exists \{x^i\}$ in A, $f(x^i) \to \alpha$.

The sequence $\{x^i\}$ that appeared in (b) is called a *minimizing sequence* for the function *f* over the set *A*.

Fact 11 *For any sequences* $\{a^i\}$ and $\{b^i\}$ in $\overline{R}$ *and t>0 one has*

$$\underline{\lim}\ ta^i = t\ \underline{\lim}\ a^i,\ \ \overline{\lim}\ ta^i = t\ \overline{\lim}\ a^i,\ \ \underline{\lim}\ \text{-}a^i = -\overline{\lim}\ a^i,\ \ \overline{\lim}\ \text{-}a^i = -\underline{\lim}\ a^i,$$

and (*under the convention that* $\infty - \infty = \infty$ and *if the following sums on the right are not* $\infty - \infty$)

$$\underline{\lim}\ (a^i + b^i) \geq \underline{\lim}\ a^i + \underline{\lim}\ b^i,\quad \overline{\lim}\ (a^i + b^i) \leq \overline{\lim}\ a^i + \overline{\lim}\ b^i .$$

Fact 12 *For any sequence* $\{a^i\}_{i\in N}$ in $\overline{R}$ *the set of all of its cluster points,* $\mathrm{Lim}\,(\mathrm{a}) := \{\alpha \in \overline{R} \mid \exists \text{ a subsequence } \mathrm{a}^{k_i} \to \alpha\}$, *is a closed set in* $\overline{R}$ *with* $\underline{\lim}\, a^i = \min \mathrm{Lim}\,(\mathrm{a})$, and $\overline{\lim}\, a^i = \max \mathrm{Lim}\,(\mathrm{a})$. *Moreover,* $a^i \to \alpha \in \overline{R}$ *if and only if* Lim (a) = $\{\alpha\}$, in which case $\underline{\lim}\, a^i = \overline{\lim}\, a^i = \alpha$.

BIBLIOGRAPHY

1. Abrams, R.A. [1975], *Projections of Convex Programs with Unattained Infima* SIAM J. Control, Vol. 13, **3**, pp. 706-718.

2. ---[1976], C*onsistency, Super-consistency, and Dual Degeneracy in Posynomial Geometric Programming*, Operations Research, Vol. 24, No. 2, pp.325-335.

3. Avriel, M. (Ed) [1980], *Advances in Geometric Programming*, Plenum Press, New York and London.

4. Avriel, M., Rijckaert, M.J. and Wilde, D.J. (Eds.) [1973], Optimization and Design, *The Impact of Optimization Theory on Engineering Design*, Prentice Hall, Englewood Cliffs, New Jersey, pp.3-6.

5. Beckenbach, E.F. and Bellman R.[1965], *Inequalities*, 2nd revised printing, Springer-Verlag.

6. Beightler, C. S. and Phillips, D. T. [1976], *Applied Geometric Programming*, John Wiley, New York.

7. Ben-Tal and A. Nemirovski [2001], *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, xii, MPS/SIAM Series on Optimization.

8. Berberian, S.K. [1999], *Fundamentals of Real Analysis*, Springer-Verlag New York, Inc.

9. Berkovitz, L.D. [2002], *Convexity and Optimization in* R^n, John Wiley & Sons, Inc., New York, NY.

10. Bertsekas, D.P. [2009], *Convex Optimization Theory*, Athena Scientific.

11. Bertsekas, D.P., Nedic, A., Ozdaglar, A.E. [2003], *Convex Analysis and Optimization*, Athena Scientific.

12. Borwein, J.M., and Lewis, A.S. [2006], C*onvex Analysis and Nonlinear Optimization*, 2e, Canadian Mathematical Society.

13. Boyd and Vandenberghe [2004], *Convex Optimization*, Cambridge University Press.

14. Boyd, S., Kim, S.-J., Vandenberghe, L., and Hassibi, A. [2007], *A tutorial on geometric programming*, Optimization and Engineering, Vol. 8, No.1/March, pp. 67-127. http://www.springerlink.com/content/9941458381562w82/

15. Bronshtein, I.N.; Semendyayev, K.A.; Musiol, G.; Muehling, H. [2004], *Handbook of Mathematics,* 4th edition, Springer ï Verlag. (p.623-624)

16. Chiang, M [2005]. *Geometric programming for communication systems*, Foundations and Trends in Communications and Information Theory, vol. 2, no. 1-2, pp. 1-154.

17. Colding, B.N. and Konig, W. [1971], *Validity of the Taylor Equation in Metal Cutting*, Annals of the CIRP, Vol.19, pp.793-812.

18. Colding, B.N. [1959], *A 3-dimensional Tool-Life Equation-Machining Economics*, Transactions ASME, **B** 81, pp.239-250.

19. Colding, B.N. [1969], *Machining Economics and Industrial Data Manuals*, Annals of the CIRP, Vol.17, pp.279-288.

20. Cover, T.M., and Thomas, J., *Elements of Information Theory*, New York, Wiley, 1991.

21. Dantzig, G.B. [1963], *Linear programming and extensions*, Princeton University Press, Princeton, N.J.

22. Dantzig, George B. and Thapa, M.N. [1997] *Linear Programming 1: Introduction.*http://site.ebrary.com/lib/pitt/docDetail.action?docID=10015675, Springer-Verlag, NY.

23. Dantzig, G.B. and Thapa, M.N. [2003] *Linear programming 2: Theory and Extensions*, New York, Springer.

24. Dembo, R.S. [1978a], *Current state of the art of algorithms and computer software for geometric programming*, Advances in Geometric Programming, JOTA, pp.27-261.

25. ----- [1978b], *Dual to Primal Conversion in Geometric Programming*, Advances in Geometric Programming, JOTA, pp. 333-342.

26. ----- [1982], *Sensitivity Analysis in Geometric Programming*, JOTA, **37**, No. 1, May.

27. Dinkel, J.J., Kochenberger, G.A., and Wong, S.N. [1977], *Entropy Optimization and Geometric Programming*, Environment and Planning, Vol. 9, pp. 419-427.

28. Duffin, R. J. [1962a], *Dual programs and minimum cost*, SIAM J. Appl. Math., 10, pp.119.

29. ----- [1962b], *Cost minimization problems treated by geometric means,* ORSA J., 10, pp.668-675.

30. Duffin, R. J. and Peterson, E. L. [1966], *Duality theory for geometric programming*, SIAM J. Appl. Math., 14, pp.1307-1349.

31. Duffin, R. J., Peterson, E. L. and Zener C. [1967], Geometric Programming - Theory and Application, John Wiley, New York.

32. Eiselt, H.A. and Sandblom, C.-L. [2007] *Linear Programming and its Applications*, Springer-Verlag Berlin Heidelberg.

33. Fang, S.C., and Rajasekera, J.R. [1987], *A perturbational approach to the main duality theorem of quadratic geometric programming*, Zeitschrift fur Operations Research, Series A (Theorie), 31 (3), pp. 103-118.

34. Fang, S.C., Peterson, E.L., and Rajasekera, J.R.[1988], *Controlled Dual Perturbations for Posynomial Programs*, European Journal of Operational Research, 35-1 (1988), pp.111 ï 117.

35. Fang, S.C., Rajasekera, J.R., Tsao, H.-S.J., *Entropy Optimization and Mathematical Programming*, Kluwer Academic Publishers, 1997.

36. Gale, D., Kuhn, H.W. and Tucker, A. W. [1951], *Linear programming and the theory of game*s, Activity Analysis of Production and Allocation (T. C. Koopmans, Editor), Wiley, New York, pp.317-329.

37. Goh, C.J., and Yang, X.Q. [2002], *Duality in Optimization and Variational Inequalities*, Taylor and Francis.

38. Hardy, G.H., Littlewood, J.E., and Polya, G., *Inequalities*, Cambridge University Press, 1952.

39. Hiriart-Urruty, J-B, and Lemarechal, C. [1966a], *Convex Analysis and Minimization Algorithms I*, Springer-Verlag.

40. ----- [1966b], *Convex Analysis and Minimization Algorithms II*, Springer-Verlag.

41. Hough, C.L. [1978], *Optimization of the Second Order Logarithmic Machining Economics Problem by Extended Geometric Programming*, Ph.D. Dissertation, Texas A & M University.

42. Hough, C.L., Jr. and Goforth, R.E. [1981a], *Quadratic Posylognomials: an Extension of Posynomial Geometric Programming*, AIIE Transactions, Vol.13, No. **1**, pp.47-54.

43. ----- [1981b], Optimization of the Second Order Logarithmic Machining Economics Problem by Extended Geometric Programming, **Part I**: Unconstrained, AIIE Transactions, Vol.13, No.**2**, pp.151-159.

44. ----- [1981c], Optimization of the Second Order Logarithmic Machining Economics Problem by Extended Geometric Programming: **Part II**: Posynomial Constraints, AIIE Transactions, Vol.13, No.**3**, pp.234-242.

45. Hough, C.L., Jr., and Chang, Y. [1998], *Theory and general case (Constrained Cutting Rate-Tool Life Characteristics Curve, part 1),* Journal of Manufacturing Science and Engineering, 120.n1: 156(4).

46. Jefferson, T. R. and Scott, C. H. [1985], *Quadratic Geometric Programming with Application to Machining Economics*, Mathematical Programming, Vol. 31, pp.137-152.

47. Jefferson, T.R., Wang, Y.P., and Scott, C.H. [1990], *Composite geometric programming*, Journal of Optimization Theory and Applications, pp. 101-18.

48. Karmarkar N. [1984], *A new polynomial time algorithm for linear programming*, Combinatorica, 4(4):373-395.

49. Kortanek, K.O., Xu, X., and Ye, Y., *An Infeasible Interior-Point Algorithm for Solving Primal and Dual Geometric Programs*, Mathematical Programming, Vol. 76, pp. 155-181, 1996.

50. Kuhn, H.W. [1956], *Solvability and Consistency for Linear Equations and Inequalities*, The American Mathematical Monthly 63, No. 4, 217ï 232.

51. Kuhn, H.W. and Tucker, A.W. [1950], *Nonlinear Programming*, in J. Neyman (ed.), Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, California, 481ï 492.

52. Lidor, G., *Chemical Equilibrium Problems Treated by Geometric and Transcendental Programming*, Stanford University, Technical Report No. 75-8, Department of Operations Research, 1975.

53. Lidor, G. and Wilde, D.J. [1978], *Transcendental geometric programs*, Journal of Optimization Theory and Applications, 1978-0926:1, pp.77-96.

54. Luenberger, D. G. [1984], *Linear and Nonlinear Programming*, 2e, Addison-Wesley.

55. Nahmias, S. [1993], *Production and Operations Analysis*, 2e, Irwin.

56. Nesterov Y. and A. Nemirovski [1994], *Interior-point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia.

57. Peressini, A.L., Sullivan, F.E., Uhl, J.J., Jr. [1988], *The Mathematics of Nonlinear Programming,* Springer-Verlag.

58. Peterson, E.L. [1970], *Symmetric duality for generalized unconstrained geometric programming*, SIAM J. of Applied. Math., 19, pp.487-526.

59. ----- [1976], *Geometric Programming - A Survey*, SIAM Review, Vol.18, pp.1-51.

60. ----- [2001a], *The Origins of Geometric Programming*, pp.15-19, Annals of Operations Research, Vol. 105, No. 1-4.

61. -----[2001b], *The fundamental relations between geometric programming duality, parametric programming duality, and ordinary Lagrangian duality*, pp.109-153, Annals of Operations Research, Vol. 105, No. 1-4.

62. Rajgopal, J. and Bricker, D.L. [1992] *On Subsidiary Problems in Geometric Programming*, European Journal of Operational Research, 63, 102-113.

63. ----- [2002] *Solving Posynomial Geometric Programming Problems via Generalized Linear Programming*, Computational Optimization and Applications, 21, 95-109.

64. Rockafellar, R. T [1970a], *Convex Analysis*, Princeton, N.J., Princeton University Press.

65. ----- [1970b], *Some convex programs whose duals are linearly constrained*, Nonlinear Programming, J. B. Rosen and O. L. Mangasarian (eds.), Academic Press, 1970, 293-322.

66. ----- [1971], *Ordinary convex programs without a duality gap*, J. Opt. Theory Appl. 7 (1971), 143-148.

67. Rockafellar, R.T. and Wets, R.J-B. [2004], *VARIATIONAL ANALYSIS*, Grundlehren der Mathematischen Wissenschaften 317, Springer-Verlag 1997 (733 pages). Second printing 2004, third printing 2009. Corrections/additions: to the first (incorporated in the second), to the second (incorporated in the third), and to the third (ongoing).

68. Ross, M. S., *Introduction to Probability and Statistics for Engineers and Scientists*, 3e, Elsevier, Academic Press, 2004.

69. Scott, C.H., and Fang, S.C.(Eds.), *Geometric Programming, Entropy Optimization and Generalizations*, Special issue on Annals of Operations Research, Vol. 105, No. 1-4, July 2001.

70. Scott, C.H., and Jefferson, T.R. [1991], *Composite convex programs*, International J. Systems Sci., Vol. 22, No.12, 2691-2696.

71. Scott, C.H., Jefferson, T.R., and Lee, A [1996]. *Stochastic tool management via composite geometric programming*, Optimization, pp.59-74.

72. Stoer, J. and Witzgall, C. [1970], *Convexity and Optimization in Finite Dimensions I*, Springer-Verlag Berlin.

73. Taylor, F.W. [1907], *On the Art of Cutting Metals,* Transactions of the American Society of Mechanical Engineers (ASME), Vol. 28.

74. Terlaky, T. [1985], *On l_p Programming*, European Journal of Operational Research 22, 70-100.

75. Tsai, P.F [1986], *An optimization algorithm and economic analysis for a constrained machining model*. Ph.D. dissertation. West Virginia University, Morgantown, West Virginia.

76. Tucker, A.W.[1956], *Dual systems of homogeneous linear relations,* in: H.W. Kuhn and A.W. Tucker (eds.), Linear Inequalities and Relate Systems, Princeton University Press, NJ.

77. Varian, Hal R. [1978], *Microeconomic Analysis*, W.W. Norton and Cowman, Inc., New York.

78. Wilde, D. J., *Globally Optimal Design*, John Wiley, New York, 1978.

79. Wu, S.M. [1964], Tool life Testing by Response Surface Methodology - Parts 1 and 2, Transactions ASME, **B** 86, pp.105-116.

80. [2002] www.mosek.com, MOSEK Optimization toolbox.

81. Ye, Y. [1997] *Interior Point Algorithms: Theory and Analysis*, John Wiley, New York.

82. Yeung, R.W. [2002], *A First Course in Information Theory*, Springer, USA.

83. Zener, C. [1961], A mathematical aid in optimizing engineering designs, Proc. Nat. Acad. Sci. U.S.A., 47, pp.537-539.

84. ----- [1962], A further mathematical aid in optimizing engineering designs, Proc. Nat. Acad. Sci. U.S.A., 48, pp.518-522.

85. ----- [1971], Engineering Design by Geometric Programming, John Wiley, NY.