DISCUSSION PAPER NO. 239

SADDLE POINTS AND DUALITY IN

GENERALIZED GEOMETRIC PROGRAMMING

by

Elmor L. Peterson

August 1976

Saddle Points and Duality in Generalized Geometric Programming

by

Elmor L. Peterson*

Abstract. Extensions of the ordinary Lagrangian are used both in saddle-point characterizations of optimality and in a development of duality theory.

Key words: Saddle points, duality, generalized geometric programming, ordinary Lagrangian, geometric Lagrangian, Kuhn-Tucker vectors, P vectors, extremality conditions.

TABLE OF CONTENTS

1. Introduction. Optimization problems from the real world usually possess a linear-algebraic component, either directly in the form of pro- blem linearities (e.g., those involving the node-arc incidence matrices in network optimization) or indirectly in the more subtle form of certain problem nonlinearities (e.g., those involving the coefficient matrices in quadratic programming or, perhaps more appropriately, those involving the exponent matrices in signomial programming). As demonstrated in the author's recent survey paper [1] and some of the references cited therein, such a component can frequently be exploited by taking a (generalized) geometric programming approach. In fact, geometric programming is primarily a body of techniques and theorems for inducing and exploiting as much linearity as possible.

The duality that exploits such linearity differs considerably from ordinary Lagrangian duality. Since section 33 of Rockafellar's book [2] indicates that any such duality can be viewed as originating from an ap- propriate Lagrangian (or saddle function), it is natural to seek such a Lagrangian for geometric programming. Although such a Lagrangian and the corresponding saddle-point characterizations of optimality have already been described in [1], proofs are given here for the first time.

In geometric programming, problems with only linear constraints are treated in essentially the same way as problems without constraints. Only problems with nonlinear constraints require additional attention and hence are classified as constrained problems.

Since many important problems are unconstrained (e.g., most network optimization problems), and since the theory for the unconstrained case is much simpler than its counterpart for the constrained case, the un- constrained case is treated separately (even though its theory is actually

embedded in the theory for the constrained case).

The only prerequisites for section 2 are the basic facts about the conjugate transformation described in section 3.1.2 of [1] (and established in [2]). An additional prerequisite for section 3 is the geometric inequality established in section 3.3.3 of [1].

2. The unconstrained case. In harmony with sections 2.1 and 3.1 of [1], suppose that $g:\mathcal{C}$ is a (proper) function g with a nonempty (effective) domain $\mathcal{C} \subseteq E_n$ (n-dimensional Euclidean space), and assume that $\mathcal{X}$ is a non-empty cone in E_n. For purposes of easy reference and mathematical precision, the resulting "geometric programming problem" $\mathcal{Q}$ is now given the following formal definition in terms of classical terminology and notation.

PROBLEM $\mathcal{Q}$. Using the feasible solution set

$$\mathcal{J} \overset{\Delta}{=} \mathcal{X} \cap \mathcal{C},$$

calculate both the problem infimum

$$\varphi \overset{\Delta}{=} \inf_{x \in \mathcal{J}} g(x)$$

and the optimal solution set

$$\mathcal{J}^* \overset{\Delta}{=} \{x \in \mathcal{J} \mid g(x) = \varphi\}.$$

Needless to say, the "ordinary programming" case occurs when $\mathcal{X}$ is actually the entire vector space $\ddot{E}_n$.

2.1. Saddle points. Our Lagrangian for problem $\mathcal{Q}$ utilizes the "conjugate transform" $h:\mathcal{P}$ of $g:\mathcal{C}$, whose domain

$$\mathcal{D} \overset{\Delta}{=} \{ y \in E_n \mid \sup_{x \in \mathcal{C}} \; [\langle y, x \rangle - g(x)] \text{ is finite} \}$$

and whose functional value

$$h(y) \overset{\Delta}{=} \sup_{x \in \mathcal{C}} \; [\langle y, x \rangle - g(x)].$$

The corresponding saddle-point characterization of optimality also utilizes the "dual cone"

$$\mathcal{Y} \overset{\Delta}{=} \{ y \in E_n \mid 0 \le \langle x, y \rangle \text{ for each } x \in \mathcal{X} \}.$$

The following definition lays a foundation for that characterization.

DEFINITION. For a consistent problem $\mathcal{A}$ with a finite infimum ω, a P vector is any vector y^* with the two properties

$$y^* \in \mathcal{D}$$

and

$$\varphi = \inf_{x \in \mathcal{X}} L_g(x; y^*),$$

where the (geometric) Lagrangian

$$L_g(x; y) \overset{\Delta}{=} \langle x, y \rangle - h(y).$$

It is worth noting that even in the ordinary programming case the geometric Lagrangian L_g is unlike the ordinary Lagrangian $L_o = g$. In fact, L_g exists only if g has a conjugate transform h.

The following theorem gives a saddle-point characterization of optimal solutions x^* and P vectors y^*. It also provides a basis for other im-

portant characterizations of such vectors.

Theorem 1. <u>Given that $g:\mathcal{C}$ is convex and closed, let $x^* \in \mathcal{X}$ and let $y^* \in \mathcal{D}$. Then x^* is optimal for problem $\mathcal{Q}$ and y^* is a P vector for problem $\mathcal{Q}$ if and only if the ordered pair $(x^*;y^*)$ is a "saddle point" for the Lagrangian L_g, that is,</u>

$$\sup_{y \in \mathcal{D}} L_g(x^*;y) = L_g(x^*;y^*) = \inf_{x \in \mathcal{X}} L_g(x;y^*),$$

<u>in which case</u> L_g <u>has the saddle-point value</u>

$$L_g(x^*;y^*) = g(x^*) = \varphi.$$

<u>Moreover</u>,

$$\sup_{y \in \mathcal{D}} L_g(x^*;y) = L_g(x^*;y^*)$$

<u>if and only if x^* and y^* satisfy both the feasibility condition</u>

$$x^* \in \mathcal{C}$$

<u>and the subgradient condition</u>

$$y^* \in \partial g(x^*),$$

<u>in which case</u>

$$L_g(x^*;y^*) = g(x^*).$$

<u>Futhermore</u>,

$$L_g(x^*;y^*) = \inf_{x \in \mathcal{X}} L_g(x;y^*)$$

if and only if x^* and y^* satisfy both the feasibility condition

$$y^* \in \mathcal{Y}$$

and the orthogonality condition

$$0 = \langle x^*, y^* \rangle,$$

in which case

$$L_g(x^*;y^*) = -h(y^*).$$

Proof. The following two lemmas must be used repetitively.

Lemma 1a. Given that $g:\mathcal{C}$ is convex and closed, a vector x satisfies the restraint $x \in \mathcal{C}$ if and only if $\sup_{y \in \mathcal{D}} L_g(x;y)$ is finite, in which case

$$\sup_{y \in \mathcal{D}} L_g(x;y) = g(x)$$

and

$$\{y \in \mathcal{D} \mid L_g(x;y) = g(x)\} = \partial g(x).$$

Proof. Immediate from the definition of $L_g(x;y)$ and conjugate transform theory.

Lemma 1b. A vector y in $\mathcal{D}$ satisfies the cone condition $y \in \mathcal{Y}$ if and only if $\inf_{x \in \mathcal{X}} L_g(x;y)$ is finite, in which case

$$\inf_{x \in \mathcal{X}} L_g(x;y) = -h(y)$$

and

$$\{x \in \mathcal{X} \mid L_g(x;y) = -h(y)\} = \{x \in \mathcal{X} \mid 0 = \langle x,y \rangle\}.$$

Proof. Immediate from the assumption that $\mathcal{X}$ is a cone; because in that case it is clear that $y \in \mathcal{Y}$ if and only if $\inf_{x \in \mathcal{X}} \langle x,y \rangle$ is finite, in which case $\inf_{x \in \mathcal{X}} \langle x,y \rangle = 0$.

Now, assuming that x^* is optimal for problem $\mathcal{Q}$ and that y^* is a P vector for problem $\mathcal{Q}$, we deduce from Lemma 1a that

$$\sup_{y \in \mathcal{B}} L_g(x^*;y) = g(x^*) = \varphi = \inf_{x \in \mathcal{X}} L_g(x;y^*),$$

by virtue of the defining properties for optimal solutions and P vectors. Notice that the first and second equations show that $L_g(x^*;y^*) \leq g(x^*) = \varphi$, and observe that the second and third equations show that $g(x^*) = \varphi \leq L_g(x^*;y^*)$; so we infer that

$$L_g(x^*;y^*) = g(x^*) = \varphi.$$

Consequently, $(x^*;y^*)$ is a saddle point for L_g.

Conversely, assuming that $(x^*;y^*)$ is a saddle point for L_g, we deduce from Lemma 1a that $x^* \in \mathcal{C}$; so x^* is feasible by virtue of the hypothesis $x^* \in \mathcal{X}$. From Lemma 1a we also infer that $\sup_{y \in \mathcal{B}} L_g(x^*;y) = g(x^*)$; so the saddle-point equations imply that

$$g(x^*) = \inf_{x \in \mathcal{X}} L_g(x;y^*),$$

which in turn means that

$$g(x^*) \leq L_g(x;y^*) \text{ for each } x \in \mathcal{X}.$$

Moreover, Lemma 1a also implies that

$$L_g(x;y^*) \leq g(x) \text{ for each } x \in \mathscr{A},$$

by virtue of the hypothesis $y^* \in \mathscr{B}$. It then follows from these two dis-
played inequalities that $g(x^*) \leq g(x)$ for each $x \in \mathscr{A}$. Thus, x^* is
optimal for problem $\mathcal{A}$, and hence $\varphi = g(x^*)$. This equation and the pre-
ceding displayed equation show that y^* is a P vector for problem $\mathcal{A}$, by
virtue of the hypothesis $y^* \in \mathscr{B}$.

Now, assuming that $\sup_{y \in \mathscr{B}} L_g(x^*;y) = L_g(x^*;y^*)$, we infer from Lemma 1a
that $x^* \in \mathcal{C}$ and that $L_g(x^*;y^*) = g(x^*)$; which in turn imply that $y^* \in \partial g(x^*)$
by virtue of the hypothesis $y^* \in \mathscr{B}$ and Lemma 1a. Conversely, assuming
that $x^* \in \mathcal{C}$ and that $y^* \in \partial g(x^*)$, we infer from Lemma 1a that

$$\sup_{y \in \mathscr{B}} L_g(x^*;y) = g(x^*) = L_g(x^*;y^*).$$

Finally, assuming that $\inf_{x \in \mathcal{X}} L_g(x;y^*) = L_g(x^*;y^*)$, we infer from Lemma
1b that $y^* \in \mathscr{Y}$ and that $L_g(x^*;y^*) = -h(y^*)$, which in turn implies that
$0 = \langle x^*, y^* \rangle$ by virtue of the definition of $L_g(x^*;y^*)$. Conversely, assuming
that $y^* \in \mathscr{Y}$ and that $0 = \langle x^*, y^* \rangle$, we infer from Lemma 1b that

$$\inf_{x \in \mathcal{X}} L_g(x;y^*) = -h(y^*) = L_g(x^*;y^*). \qquad \text{q.e.d.}$$

Since the second assertion of Theorem 1 gives certain conditions that
are equivalent to the first saddle-point equation, and since the third
assertion of Theorem 1 gives certain other conditions that are equivalent
to the second saddle-point equation, Theorem 1 actually provides four dif-
ferent characterizations of all ordered pairs $(x^*;y^*)$ of optimal solutions
x^* and P vectors y^*.

Of course, each of those four characterizations provides a character-
ization of all optimal solutions x^* in terms of a given P vector y^*, as

well as a characterization of all P vectors $y*$ in terms of a given optimal solution $x*$.

Still another characterization of all optimal solutions $x*$ to certain problems $\mathcal{Q}$ has been given by the author [3].

2.2. Duality. Corresponding to problem $\mathcal{Q}$ is the following "geometric dual problem" $\mathcal{B}$.

PROBLEM $\mathcal{B}$. Using the feasible solution set

$$\mathcal{T} \overset{\Delta}{=} \{y \in \mathcal{B} \mid \inf_{x \in \mathcal{X}} L_{\mathcal{Q}}(x;y) \text{ is finite}\}$$

and the objective function

$$\mathcal{V}(y) \overset{\Delta}{=} \inf_{x \in \mathcal{X}} L_{\mathcal{Q}}(x;y),$$

calculate both the problem supremum

$$\Psi \overset{\Delta}{=} \sup_{y \in \mathcal{T}} \mathcal{V}(y)$$

and the optimal solution set

$$\mathcal{T}* \overset{\Delta}{=} \{y \in \mathcal{T} \mid \mathcal{V}(y) = \Psi\}.$$

Even though problem $\mathcal{B}$ is essentially a "maximin problem" -- a type of problem that tends to be relatively difficult to analyse -- the minimization problems that must be solved to obtain the objective function $\mathcal{V}:\mathcal{T}$ have trivial solutions. In particular, Lemma 1b clearly implies that

$$\mathcal{T} = \mathcal{Y} \cap \mathcal{B} \text{ and } \mathcal{V}(y) = -h(y),$$

so problem $\mathcal{B}$ can actually be rephrased in the following more direct way.

PROBLEM $\mathcal{B}$. <u>Using the feasible solution set</u>

$$\mathcal{T} \overset{\Delta}{=} \mathcal{Y} \cap \mathcal{D},$$

<u>calculate both the problem infimum</u>

$$\psi \overset{\Delta}{=} \inf_{y \in \mathcal{T}} h(y) = -\Psi$$

<u>and the optimal solution set</u>

$$\mathcal{T}\ast \overset{\Delta}{=} \{y \in \mathcal{T} \mid h(y) = \psi\}.$$

When phrased in this way, problem $\mathcal{B}$ closely resembles problem $\mathcal{A}$ and is in fact a geometric programming problem. Of course, the geometric dual problem $\mathcal{B}$ can actually be defined in this way, but the preceding derivation serves as an important link between geometric Lagrangians and geometric duality (analogous to the link between ordinary Lagrangians and ordinary duality).

To further strengthen that link, we first need to develop the most basic duality theory -- a theory in which the following definition is almost as important as the definition of the dual problems $\mathcal{A}$ and $\mathcal{B}$.

DEFINITION. The <u>extremality conditions</u> (for unconstrained geometric programming) are:

(I) $\qquad\qquad x \in \mathcal{X}$ and $\qquad\qquad y \in \mathcal{Y}$,

(II) $\qquad\qquad 0 = \langle x, y \rangle$,

(III) $\qquad\qquad y \in \partial g(x)$.

The following "duality theorem" is the basis for many important theorems.

Theorem 2. If x and y <u>are feasible solutions to problems $\mathcal{A}$ and $\mathcal{B}$ re-</u> <u>spectively (in which case the extremality conditions (I) are satisfied),</u> <u>then</u>

$$0 \le g(x) + h(y),$$

<u>with equality holding if and only if the extremality conditions (II) and</u> <u>(III) are satisfied, in which case x and y are optimal solutions to</u> <u>problems $\mathcal{A}$ and $\mathcal{B}$ respectively.</u>

Proof. The following fact is formalized as a lemma in order to facilitate a comparison between the constrained and unconstrained cases.

Lemma 2a. <u>If $x \in \mathcal{C}$ and $y \in \mathcal{D}$, then</u>

$$\langle x, y \rangle \le g(x) + h(y),$$

<u>with equality holding if and only if the extremality condition (III) is</u> <u>satisfied.</u>

Proof. Simply invoke the conjugate inequality presented in section 3.1.2 of [1].

Now, the fact that x and y are in the cone $\mathcal{X}$ and its dual $\mathcal{Y}$ respectively combined with Lemma 2a shows that

$$0 \le \langle x, y \rangle \le g(x) + h(y),$$

with equality holding in both of these inequalities if and only if the equality conditions stated in the theorem are satisfied.   q.e.d.

The following important corollary is an immediate consequence of Theorem 2.

Corollary 2A. If the dual problems $\mathcal{A}$ and $\mathcal{B}$ are both consistent, then

    (i)   the infimum φ for problem $\mathcal{A}$ is finite, and

$$0 \leq \varphi + h(y) \text{ for each } y \in \mathcal{T},$$

    (ii)   the infimum ψ for problem $\mathcal{B}$ is finite, and

$$0 \leq \varphi + \psi$$

The strictness of the inequality in conclusion (ii) plays a crucial role in almost all duality theorems.

DEFINITION. Consistent dual problems $\mathcal{A}$ and $\mathcal{B}$ for which

$$0 < \varphi + \psi$$

have a duality gap of $\varphi + \psi$.

A much more thorough discussion of duality theory and the role played by duality gaps is given in [1] and some of the references cited therein.

    The link between geometric Lagrangians and geometric duality can now be further strengthened by the following tie between dual problem $\mathcal{B}$ and the P vectors for problem $\mathcal{A}$ defined in section 2.1.

Theorem 3. Given that problem $\mathcal{A}$ is consistent with a finite infimum φ,

    (1)   if problem $\mathcal{A}$ has a P vector, then problem $\mathcal{B}$ is consistent and

$$0 = \varphi + \psi,$$

(2) if problem $\mathcal{B}$ is consistent and $0 = \varphi + \psi$, then

$$\{y^* \mid y^* \text{ is a } P \text{ vector for problem } \mathcal{A}\} = \mathcal{T}^*.$$

Proof. If y^* is a P vector for problem $\mathcal{A}$, then Lemma 1b implies that y^* is feasible for problem $\mathcal{B}$ and that $\varphi = -h(y^*)$, which in turn implies that $\psi = h(y^*)$ by virtue of conclusion (i) to Corollary 2A. Consequently, $0 = \varphi + \psi$ and $y^* \in \mathcal{T}^*$.

On the other hand, if problem $\mathcal{B}$ is consistent and $0 = \varphi + \psi$, then each vector y^* in $\mathcal{T}^*$ has the property $\varphi = -h(y^*)$, and hence each such vector y^* is a P vector for problem $\mathcal{A}$ by virtue of the first two paragraphs of this subsection.   q.e.d.

An important consequence of Theorem 3 is that, when they exist, all P vectors for problem $\mathcal{A}$ can be obtained simply by computing the dual optimal solution set $\mathcal{T}^*$. However, there are cases in which the vectors in $\mathcal{T}^*$ are not P vectors for problem $\mathcal{A}$, though such cases can occur only when $0 < \varphi + \psi$, in which event Theorem 3 implies that there can be no P vectors for problem $\mathcal{A}$.

3. The constrained case. In harmony with sections 2.2 and 3.3 of [1], we introduce two nonintersecting (possibly empty) positive-integer index sets I and J, with finite cardinality o(I) and o(J) respectively. In terms of these index sets I and J we also introduce the following notation and hypotheses:

(Ia) For each $k \in \{0\} \cup I \cup J$ suppose that $g_k : C_k$ is a (proper) function g_k with a nonempty (effective) domain $C_k \subseteq E_{n_k}$ (n_k - dimensional Euclidean

space), and for each $j \in J$ let D_j be the (effective) domain of the "conjugate transform" $h_j : D_j$ of $g_j : C_j$.

(IIa) For each $k \in \{0\} \cup I \cup J$ let x^k be an independent vector variable in E_{n_k}, and let κ be an independent vector variable with components κ_j for each $j \in J$.

(IIIa) Denote the cartesian product of the vector variables x^i, $i \in I$ by the symbol x^I, and denote the cartesian product of the vector variables x^j, $j \in J$ by the symbol x^J. Then the cartesian product $(x^0, x^I, x^J) \overset{\Delta}{=} x$ is a vector variable in E_n, where

$$n \overset{\Delta}{=} n_0 + \sum_I n_i + \sum_J n_j.$$

(IVa) Assume that X is a nonempty cone in E_n.

For purposes of easy reference and mathematical precision, the resulting "geometric programming problem" A is now given the following formal definition in terms of classical terminology and notation.

PROBLEM A. <u>Consider the objective function</u> G <u>whose domain</u>

$$C \overset{\Delta}{=} \{(x, \kappa) \mid x^k \in C_k, \ k \in \{0\} \cup I, \text{ and } (x^j, \kappa_j) \in C_j^+, \ j \in J\}$$

<u>and whose functional value</u>

$$G(x, \kappa) \overset{\Delta}{=} g_0(x^0) + \sum_J g_j^+(x^j, \kappa_j),$$

<u>where</u>

$$C_j^+ \overset{\Delta}{=} \{(x^j, \kappa_j) \mid \underline{\text{either}} \ \kappa_j = 0 \ \underline{\text{and}} \ \sup_{d^j \in D_j} \langle x^j, d^j \rangle < +\infty, \ \underline{\text{or}} \ \kappa_j > 0 \ \underline{\text{and}} \ x^j \in \kappa_j C_j\}$$

<u>and</u>

$$
g_j^+(x^j, \kappa_j) \overset{\Delta}{=} \begin{cases} \sup_{d^j \in D_j} \langle x^j, d^j \rangle & \underline{\text{if}} \ \kappa_j = 0 \ \underline{\text{and}} \ \sup_{d^j \in D_j} \langle x^j, d^j \rangle < + \infty \\ \\ \kappa_j g_j(x^j/\kappa_j) & \underline{\text{if}} \ \kappa_j > 0 \ \underline{\text{and}} \ x^j \in \kappa_j C_j. \end{cases}
$$

Using the feasible solution set

$$
S \overset{\Delta}{=} \{ (x, \kappa) \in C \mid x \in X, \ \text{and} \ g_i(x^i) \leq 0, \ i \in I \},
$$

calculate both the problem infimum

$$
\omega \overset{\Delta}{=} \inf_{(x, \kappa) \in S} G(x, \kappa)
$$

and the optimal solution set

$$
S* \overset{\Delta}{=} \{ (x, \kappa) \in S \mid G(x, \kappa) = \omega \}.
$$

Needless to say, the unconstrained case occurs when $I = J = \emptyset$, $g_0 : C_0 \overset{\Delta}{=} g : \mathcal{C}$, and $X \overset{\Delta}{=} \mathcal{X}$. On the other hand, the "ordinary programming" case occurs when

$$
J = \emptyset,
$$

$$
n_k = m \ \text{and} \ C_k \overset{\Delta}{=} C_0 \ \text{for some set} \ C_0 \subseteq E_m \qquad\qquad k \in \{0\} \cup I,
$$

and

$$
X \overset{\Delta}{=} \text{column space of} \begin{bmatrix} U \\ U \\ \cdot \\ \cdot \\ \cdot \\ U \end{bmatrix}
$$

where there is a total of $1 + o(I)$ identity matrices U that are $m \times m$.

In particular, an explicit elimination of the vector space condition $x \in X$
by the linear transformation

$$\begin{pmatrix} x^0 \\ x^I \end{pmatrix} = \begin{bmatrix} U \\ U \\ \cdot \\ \cdot \\ \cdot \\ U \end{bmatrix} \dot{z}$$

shows that problem A is then equivalent to the very general ordinary pro-
gramming problem

$$\underline{\text{Minimize }} g_0(z) \underline{\text{ subject to}}$$

$$g_i(z) \leq 0 \qquad i \in I$$

$$z \in C_o.$$

3.1. <u>Saddle points</u>. Our Lagrangian for problem A utilizes the
"conjugate transform" $h_k : D_k$ of $g_k : C_k$, whose domain

$$D_k \overset{\Delta}{=} \{y^k \in E_{n_k} \mid \sup_{x^k \in C_k} [\langle y^k, x^k \rangle - g_k(x^k)] \text{ is finite}\}$$

and whose functional value

$$h_k(y^k) \overset{\Delta}{=} \sup_{x^k \in C_k} [\langle y^k, x^k \rangle - g_k(x^k)].$$

Notationally, it involves both the vector variable (x,κ) and an analogous
vector variable (y,λ), where the vector variable y has the same cartesian-
product structure as the vector variable x and where the vector variable
λ has components λ_i for each $i \in I$. The corresponding saddle-point character-
ization of optimality also utilizes the "dual cone"

$$Y \overset{\Delta}{=} \{y \in E_n \mid 0 \leq \langle x, y \rangle \text{ for each } x \in X\}.$$

The following definition lays a foundation for that characterization.

DEFINITION. Consider the function H whose domain

$$D \overset{\Delta}{=} \{(y, \lambda) \mid y^k \in D_k, \ k \in \{0\} \cup J, \text{ and } (y^i, \lambda_i) \in D_i^+, \ i \in I\}$$

and whose functional value

$$H(y, \lambda) \overset{\Delta}{=} h_0(y^0) + \sum_I h_i^+(y^i, \lambda_i),$$

where

$$D_i^+ \overset{\Delta}{=} \{(y^i, \lambda_i) \mid \text{either } \lambda_i = 0 \text{ and } \sup_{c^i \in C_i} \langle c^i, y^i \rangle < +\infty, \text{ or } \lambda_i > 0 \text{ and } y^i \in \lambda_i D_i\}$$

and

$$h_i^+(y^i, \lambda_i) \overset{\Delta}{=} \begin{cases} \sup_{c^i \in C_i} \langle c^i, y^i \rangle & \text{if } \lambda_i = 0 \text{ and } \sup_{c^i \in C_i} \langle c^i, y^i \rangle < +\infty \\ \\ \lambda_i h_i(y^i / \lambda_i) & \text{if } \lambda_i > 0 \text{ and } y^i \in \lambda_i D_i. \end{cases}$$

For a consistent problem A with a finite infimum φ, a _P vector_ is any vector $(y*, \lambda*)$ with the two properties

$$(y*, \lambda*) \in D$$

and

$$\varphi = \inf_{\substack{x \in X \\ \kappa \geq 0}} L_g(x, \kappa; \ y*, \lambda*),$$

where the (geometric) _Lagrangian_

$$L_g(x, \kappa;\ y, \lambda) \overset{\Delta}{=} \langle x, y \rangle - H(y, \lambda) - \sum_J \kappa_j h_j(y^j).$$

Needless to say, even in the ordinary programming case the geometric Lagrangian L_g is unlike the ordinary Lagrangian $L_o \overset{\Delta}{=} g_0 + \sum_I \lambda_i g_i$. In fact, L_g exists only if g_k has a conjugate transform h_k for $k \in \{0\} \cup I \cup J$.

On the other hand, the following important concept from ordinary programming plays a crucial role in the theory to come.

DEFINITION. For a consistent problem A with a finite infimum φ, a _Kuhn-Tucker vector_ is any vector λ^* in $E_{o(I)}$ with the two properties

$$\lambda_i^* \geq 0 \qquad\qquad\qquad i \in I,$$

and

$$\varphi = \inf_{\substack{(x, \kappa) \in C \\ x \in X}} L_o(x, \kappa; \lambda^*),$$

where the (ordinary) _Lagrangian_

$$L_o(x, \kappa; \lambda) \overset{\Delta}{=} G(x, \kappa) + \sum_I \lambda_i g_i(x^i).$$

It is important to realize that the preceding definition of Kuhn-Tucker vectors differs considerably from the widely used definition involving the Kuhn-Tucker optimality conditions. Even in the ordinary convex programming case the two definitions are not equivalent, though it is well-known that the preceding definition simply admits a somewhat larger set of vectors in that case.

The following theorem gives a saddle-point characterization of

optimal solutions (x^*, K^*) and P vectors (y^*, λ^*). It also provides a

basis for other important characterizations of such vectors.

Theorem 4. Given that $g_k : C_k$, $k \in \{0\} \cup I \cup J$ is convex and closed, let

(x^*, K^*) be such that $x^* \in X$ and $K^* \geq 0$, and let $(y^*, \lambda^*) \in D$. Then (x^*, K^*)

is optimal for problem A and (y^*, λ^*) is a P vector for problem A if and

only if the ordered pair $(x^*, K^*; y^*, \lambda^*)$ is a "saddle point" for the

Lagrangian L_g, that is,

$$\sup_{\substack{(y, \lambda) \in D}} L_g(x^*, K^*; y, \lambda) = L_g(x^*, K^*; y^*, \lambda^*) = \inf_{\substack{x \in X \\ K \geq 0}} L_g(x, K; y^*, \lambda^*),$$

in which case L_g has the saddle-point value

$$L_g(x^*, K^*; y^*, \lambda^*) = G(x^*, K^*) = \wp.$$

Moreover,

$$\sup_{\substack{(y, \lambda) \in D}} L_g(x^*, K^*; y, \lambda) = L_g(x^*, K^*; y^*, \lambda^*)$$

if and only if (x^*, K^*) and (y^*, λ^*) satisfy both the feasibility conditions

$$(x^*, K^*) \in C,$$

$$g_i(x^{*i}) \leq 0 \qquad\qquad i \in I,$$

and the subgradient and "complementary slackness" conditions

$$y^{*0} \in \partial g_0(x^{*0}),$$

either $\lambda_i^* = 0$ and $\langle x^{*i}, y^{*i} \rangle = \sup_{c^i \in C_i} \langle c^i, y^{*i} \rangle$, or $\lambda_i^* > 0$ and $y^{*i} \in \lambda_i^* \partial g_i(x^{*i})$, $i \in I$,

$$\lambda_i^* g_i(x^{*i}) = 0 \qquad\qquad i \in I,$$

<u>either</u> $\kappa_j^* = 0$ <u>and</u> $\langle x^{*j}, y^{*j} \rangle = \sup\limits_{d^j \in D_j} \langle x^{*j}, d^j \rangle$, <u>or</u> $\kappa_j^* > 0$ <u>and</u> $y^{*j} \in \partial g_j(x^{*j}/\kappa_j^*)$, $j \in J$,

<u>in which case</u>

$$L_g(x^*, \kappa^*; y^*, \lambda^*) = G(x^*, \kappa^*).$$

<u>Futhermore</u>,

$$L_g(x^*, \kappa^*; y^*, \lambda^*) = \inf\limits_{\substack{x \in X \\ \kappa \geq 0}} L_g(x, \kappa; y*, \lambda*)$$

<u>if and only if</u> (x^*, κ^*) <u>and</u> (y^*, λ^*) <u>satisfy both the feasibility conditions</u>

$$y^* \in Y,$$

$$h_j(y^{*j}) \leq 0 \qquad\qquad\qquad j \in J,$$

<u>and the orthogonality and complementary slackness conditions</u>

$$0 = \langle x*, y* \rangle,$$

$$\kappa_j^* h_j(y^{*j}) = 0 \qquad\qquad\qquad j \in J,$$

<u>in which case</u>

$$L_g(x^*, \kappa^*; y^*, \lambda^*) = -H(y^*, \lambda^*).$$

Proof. The following lemmas must be used repetitively.

<u>Lemma 4a</u>. <u>Given that</u> $g_k : C_k$, $k \in \{0\} \cup I \cup J$ <u>is convex and closed</u>, <u>a vector</u> (x, κ) <u>with</u> $\kappa \geq 0$ <u>satisfies the restraint</u> $(x, \kappa) \in C$ <u>and the constraints</u> $g_i(x^i) \leq 0$, $i \in I$ <u>if and only if</u> $\sup\limits_{(y, \lambda) \in D} L_g(x, \kappa; y, \lambda)$ <u>is finite</u>, <u>in which</u> <u>case</u>

$$\sup\limits_{(y, \lambda) \in D} L_g(x, \kappa; y, \lambda) = G(x, \kappa)$$

<u>and</u>

$$\{(y,\lambda) \in D \mid L_g(x,\kappa;\ y,\lambda) = G(x,\kappa)\}$$

$$= \{(y,\lambda) \mid y^0 \in \partial g_0(x^0); \ \underline{\text{either}} \ \lambda_i = 0 \ \underline{\text{and}} \ \langle x^i, y^i \rangle = \sup_{c^i \in C_i} \langle c^i, y^i \rangle, \ \underline{\text{or}}$$

$$\lambda_i > 0 \ \underline{\text{and}} \ y^i \in \lambda_i \partial g_i(x^i), \ i \in I; \ \lambda_i g_i(x^i) = 0, \ i \in I;$$

$$\underline{\text{either}} \ \kappa_j = 0 \ \underline{\text{and}} \ \langle x^j, y^j \rangle = \sup_{d^j \in D_j} \langle x^j, d^j \rangle, \ \underline{\text{or}}$$

$$\kappa_j > 0 \ \underline{\text{and}} \ y^j \in \partial g_j(x^j/\kappa_j), \ j \in J\}.$$

Proof. First, observe that

$$\sup_{(y,\lambda) \in D} L_g(x,\kappa;\ y,\lambda) = \sup_{(y,\lambda) \in D} [\langle x,y \rangle - H(y,\lambda) - \sum_J \kappa_j h_j(y^j)]$$

$$= \sup_{(y,\lambda) \in D} [\langle x^0, y^0 \rangle + \sum_I \langle x^i, y^i \rangle + \sum_J \langle x^j, y^j \rangle - h_0(y^0) - \sum_I h_i^+(y^i, \lambda_i)$$

$$- \sum_J \kappa_j h_j(y^j)]$$

$$= \sup_{y^0 \in D_0} [\langle x^0, y^0 \rangle - h_0(y^0)] + \sum_I \sup_{(y^i, \lambda_i) \in D_i^+} [\langle x^i, y^i \rangle + 0\lambda_i - h_i^+(y^i, \lambda_i)]$$

$$+ \sum_J \sup_{y^j \in D_j} [\langle x^j, y^j \rangle - \kappa_j h_j(y^j)].$$

From conjugate transform theory we know that the preceding supremum with index 0 is finite if and only if $x^0 \in C_0$, in which case

$$\sup_{y^0 \in D_0} [\langle x^0, y^0 \rangle - h_0(y^0)] = g_0(x^0)$$

and

$$\{y^0 \in D_0 \mid \langle x^0, y^0 \rangle - h_0(y^0) = g_0(x^0)\} = \partial g_0(x^0).$$

Analogous information about the preceding suprema with indices i and j is provided by the following two sublemmas, which collectively complete the proof of Lemma 4a.

<u>Sublemma 4ai</u>. <u>Given that</u> $g_i : C_i$ <u>is convex and closed, the</u>

$\sup\limits_{(y^i, \lambda_i) \in D_i^+} [\langle x^i, y^i \rangle + \alpha_i \lambda_i - h_i^+(y^i, \lambda_i)]$ <u>is finite if and only if both</u>

$x^i \in C_i$ <u>and</u> $g_i(x^i) + \alpha_i \le 0$, <u>in which case</u>

$$\sup\limits_{(y^i, \lambda_i) \in D_i^+} [\langle x^i, y^i \rangle + \alpha_i \lambda_i - h_i^+(y^i, \lambda_i)] = 0$$

<u>and</u>

$\{ (y^i, \lambda_i) \in D_i^+ \, | \, \langle x^i, y^i \rangle + \alpha_i \lambda_i - h_i^+(y^i, \lambda_i) = 0 \}$

$\quad = \{ (y^i, \lambda_i) \in D_i^+ \, | \; \underline{\text{either}} \; \lambda_i = 0 \; \underline{\text{and}} \; \langle x^i, y^i \rangle =$

$\qquad \sup\limits_{c^i \in C_i} \langle c^i, y^i \rangle, \; \underline{\text{or}} \; \lambda_i > 0 \; \underline{\text{and}} \; y^i \in \lambda_i \partial g_i(x^i); \; \lambda_i [g_i(x^i) + \alpha_i] = 0 \}.$

<u>Proof</u>. First, observe that

$\sup\limits_{(y^i, \lambda_i) \in D_i^+} [\langle x^i, y^i \rangle + \alpha_i \lambda_i - h_i^+(y^i, \lambda_i)]$

$\quad = \sup\limits_{\lambda_i \ge 0} [\sup\limits_{y^i} \{ \langle x^i, y^i \rangle + \alpha_i \lambda_i - h_i^+(y^i, \lambda_i) \, | \, (y^i, \lambda_i) \in D_i^+ \}]$

$\quad = \sup\limits_{\lambda_i \ge 0} [\alpha_i \lambda_i + \sup\limits_{y^i} \{ \langle x^i, y^i \rangle - h_i^+(y^i, \lambda_i) \, | \, (y^i, \lambda_i) \in D_i^+ \}]$

$\quad = \sup\limits_{\lambda_i \ge 0} [\alpha_i \lambda_i + \left\{ \begin{array}{ll} \sup\limits_{y^i} \{ \langle x^i, y^i \rangle - \sup\limits_{c^i \in C_i} \langle c^i, y^i \rangle \, | \, \sup\limits_{c^i \in C_i} \langle c^i, y^i \rangle < + \infty \} & \\[2em] & \text{if } \lambda_i = 0 \\[3em] \sup\limits_{y^i} \{ \langle x^i, y^i \rangle - \lambda_i h_i(y^i/\lambda_i) \, | \, y^i/\lambda_i \in D_i \} & \text{if } \lambda_i > 0 \end{array} \right.$

$$= \sup_{\lambda_i \geq 0} [\alpha_i \lambda_i + \begin{pmatrix} 0 & \text{if } \lambda_i = 0 \text{ and } x^i \in \bar{C}_i \\ +\infty & \text{if } \lambda_i = 0 \text{ and } x^i \notin \bar{C}_i \\ +\infty & \text{if } \lambda_i > 0 \text{ and } x^i \notin C_i \\ \lambda_i g_i(x^i) & \text{if } \lambda_i > 0 \text{ and } x^i \in C_i \end{pmatrix}],$$

where the final step makes use of the fact that the zero function with domain $\bar{C}_i$ (the topological closure of C_i) is the conjugate transform of the conjugate transform of the zero function with domain C_i. Now, note that the last expression is finite only if $x^i \in C_i$, in which case the last expression clearly

$$= \sup_{\lambda_i \geq 0} [\alpha_i \lambda_i + \lambda_i g_i(x^i)].$$

But this expression is obviously finite if and only if $g_i(x^i) + \alpha_i \leq 0$, in which case this expression is clearly zero.

Finally, given that $x^i \in C_i$, if $(y^i, \lambda_i) \in D_i^+$, then the geometric inequality established in section 3.3.3 of [1] implies that

$$\langle x^i, y^i \rangle + \alpha_i \lambda_i - h_i^+(y^i, \lambda_i) \leq \lambda_i [g_i(x^i) + \alpha_i],$$

with equality holding if and only if

either $\lambda_i = 0$ and $\langle x^i, y^i \rangle = \sup_{c^i \in C_i} \langle c^i, y^i \rangle$, or $\lambda_i > 0$ and $y^i \in \lambda_i \partial g_i(x^i)$.

Moreover, given that $g_i(x^i) + \alpha_i \leq 0$, if $(y^i, \lambda_i) \in D_i^+$, then the fact that $\lambda_i \geq 0$ implies that

$$\lambda_i [g_i(x^i) + \alpha_i] \leq 0,$$

with equality holding if and only if

$$\lambda_i [g_i (x^i) + \alpha_i] = 0.$$

Taken together, these two inequalities and the corresponding characterizations of equality clearly imply the last equation of Sublemma 4ai.

Sublemma 4aj. <u>Given that</u> $g_j : C_j$ <u>is convex and closed and given that</u> $\kappa_j \geq 0$, <u>the</u> $\sup_{y^j \in D_j} [\langle x^j, y^j \rangle - \kappa_j h_j (y^j)]$ <u>is finite if and only if</u> $(x^j, \kappa_j) \in C_j^+$, <u>in</u> <u>which case</u>

$$\sup_{y^j \in D_j} [\langle x^j, y^j \rangle - \kappa_j h_j (y^j)] = g_j^+ (x^j, \kappa_j)$$

and

$$\{y^j \in D_j \,|\, \langle x^j, y^j \rangle - \kappa_j h_j (y^j) = g_j^+ (x^j, \kappa_j)\}$$

$$= \{y^j \in D_j \,|\, \underline{\text{either}} \;\; \kappa_j = 0 \;\; \underline{\text{and}} \;\; \langle x^j, y^j \rangle = \sup_{d^j \in D_j} \langle x^j, d^j \rangle,$$

$$\underline{\text{or}} \;\; \kappa_j > 0 \;\; \underline{\text{and}} \;\; y^j \in \partial g_j (x^j / \kappa_j)\}.$$

Proof. First, observe that

$$\sup_{y^j \in D_j} [\langle x^j, y^j \rangle - \kappa_j h_j (y^j)] = \begin{cases} \sup_{y^j \in D_j} \langle x^j, y^j \rangle & \text{if } \kappa_j = 0 \\[2ex] \kappa_j g_j (x^j / \kappa_j) & \text{if } \kappa_j > 0 \text{ and } x^j \in \kappa_j C_j \\[2ex] + \infty & \text{if } \kappa_j > 0 \text{ and } x^j \notin \kappa_j C_j. \end{cases}$$

Now, note that this expression is finite only if $(x^j, \kappa_j) \in C_j^+$, in which case this expression is clearly $g_j^+ (x^j, \kappa_j)$.

Finally, given that $(x^j, \kappa_j) \in c_j^+$, if $y^j \in D_j$, then the geometric inequality established in section 3.3.3 of [1] asserts that

$$\langle x^j, y^j \rangle - \kappa_j h_j(y^j) \le g_j^+(x^j, \kappa_j),$$

with equality holding if and only if

either $\kappa_j = 0$ and $\langle x^j, y^j \rangle = \sup_{d^j \in D_j} \langle x^j, d^j \rangle$, or $\kappa_j > 0$ and $x^j \in \kappa_j \partial h_j(y^j)$.

This inequality and the corresponding characterization of equality imply the last equation of Sublemma 4aj, because the relation $x^j \in \kappa_j \partial h_j(y^j)$ is equivalent to the relation $y^j \in \partial g_j(x^j/\kappa_j)$ when $\kappa_j > 0$ (by virtue of conjugate transform theory).

The proof of Lemma 4a is now complete.

Lemma 4b. <u>A vector (y, λ) in D satisfies the cone condition $y \in Y$ and the constraints $h_j(y^j) \le 0$, $j \in J$ if and only if $\inf_{\substack{x \in X \\ \kappa \ge 0}} L_g(x, \kappa; y, \lambda)$ is finite, in</u> which case

$$\inf_{\substack{x \in X \\ \kappa \ge 0}} L_g(x, \kappa; y, \lambda) = -H(y, \lambda)$$

and

$$\{(x, \kappa) \mid x \in X, \ \kappa \ge 0; \ \underline{and} \ L_g(x, \kappa; y, \lambda) = -H(y, \lambda)\}$$

$$= \{(x, \kappa) \mid x \in X, \ \kappa \ge 0; \ 0 = \langle x, y \rangle; \ \underline{and} \ \kappa_j h_j(y^j) = 0, \ j \in J\}.$$

Proof. First, observe that

$$\inf_{\substack{x \in X \\ \kappa \geq 0}} L_g(x,\kappa;y,\lambda) = \inf_{\substack{x \in X \\ \kappa \geq 0}} \left[\langle x,y \rangle - H(y,\lambda) - \sum_J \kappa_j h_j(y^j) \right]$$

$$= \inf_{x \in X} \langle x,y \rangle + \inf_{\substack{\kappa \geq 0}} \sum_J \kappa_j [-h_j(y^j)] - H(y,\lambda).$$

The proof of Lemma 4b is now immediate from the assumption that X is a cone and elementary considerations.

Now, assuming that (x^*,κ^*) is optimal for problem A and that (y^*,λ^*) is a P vector for problem A, we deduce from Lemma 4a that

$$\sup_{(y,\lambda) \in D} L_g(x^*,\kappa^*;y,\lambda) = G(x^*,\kappa^*) = \varphi = \inf_{\substack{x \in X \\ \kappa \geq 0}} L_g(x,\kappa;y^*,\lambda^*),$$

by virtue of the defining properties for optimal solutions and P vectors. Notice that the first and second equations show that $L_g(x^*,\kappa^*; y^*,\lambda^*) \leq G(x^*,\kappa^*) = \varphi$, and observe that the second and third equations show that $G(x^*,\kappa^*) = \varphi \leq L_g(x^*,\kappa^*; y^*,\lambda^*)$; so we infer that

$$L_g(x^*,\kappa^*; y^*,\lambda^*) = G(x^*,\kappa^*) = \varphi.$$

Consequently, $(x^*,\kappa^*; y^*,\lambda^*)$ is a saddle point for L_g.

Conversely, assuming that $(x^*,\kappa^*; y^*,\lambda^*)$ is a saddle point for L_g, we deduce from Lemma 4a that $(x^*,\kappa^*) \in C$ and $g_i(x^{*i}) \leq 0$, $i \in I$; so (x^*,κ^*) is feasible by virtue of the hypotheses $x^* \in X$ and $\kappa^* \geq 0$. From Lemma 4a we also infer that $\sup_{(y,\lambda) \in D} L_g(x^*,\kappa^*; y,\lambda) = G(x^*,\kappa^*)$; so the saddle-point equations imply that

$$G(x^*,\kappa^*) = \inf_{\substack{x \in X \\ \kappa \geq 0}} L_g(x,\kappa;y^*,\lambda^*),$$

which in turn means that

$$G(x^*, \kappa^*) \le L_g(x, \kappa; y^*, \lambda^*) \text{ for each } x \in X \text{ and each } \kappa \ge 0.$$

Moreover, Lemma 4a also implies that

$$L_g(x, \kappa; y^*, \lambda^*) \le G(x, \kappa) \text{ for each } (x, \kappa) \in S,$$

by virtue of the hypothesis $(y^*, \lambda^*) \in D$. It then follows from these two displayed inequalities that $G(x^*, \kappa^*) \le G(x, \kappa)$ for each $(x, \kappa) \in S$. Thus, (x^*, κ^*) is optimal for problem A, and hence $\varphi = G(x^*, \kappa^*)$. This equation and the preceding displayed equation show that (y^*, λ^*) is a P vector for problem A, by virtue of the hypothesis $(y^*, \lambda^*) \in D$.

Now, assuming that $\sup_{(y, \lambda) \in D} L_g(x^*, \kappa^*; y, \lambda) = L_g(x^*, \kappa^*; y^*, \lambda^*)$, we

infer from Lemma 4a that $(x^*, \kappa^*) \in C$ and $g_i(x^{*i}) \le 0$, $i \in I$, and that $L_g(x^*, \kappa^*; y^*, \lambda^*) = G(x^*, \kappa^*)$. This in turn implies that

$$y^{*0} \in \partial g_0(x^{*0}),$$

either $\lambda_i^* = 0$ and $\langle x^{*i}, y^{*i} \rangle = \sup_{c^i \in C_i} \langle c^i, y^{*i} \rangle$, or $\lambda_i^* > 0$ and $y^{*i} \in \lambda_i^* \partial g_i(x^{*i})$,
$$i \in I,$$

$$\lambda_i^* g_i(x^{*i}) = 0 \qquad i \in I,$$

either $\kappa_j^* = 0$ and $\langle x^{*j}, y^{*j} \rangle = \sup_{d^j \in D} \langle x^{*j}, d^j \rangle$, or $\kappa_j^* > 0$ and $y^{*j} \in \partial g_j(x^{*j}/\kappa_j^*)$,
$$j \in J,$$

by virtue of the hypothesis $(y^*, \lambda^*) \in D$ and Lemma 4a. Conversely, assuming that $(x^*, \kappa^*) \in C$ and $g_i(x^{*i}) \le 0$, $i \in I$, and that (y^*, λ^*) satisfies the preceding displayed relations, we infer from Lemma 4a that

$$\sup_{(y, \lambda) \in D} L_g(x^*, \kappa^*; y, \lambda) = G(x^*, \kappa^*) = L_g(x^*, \kappa^*; y^*, \lambda^*).$$

Finally, assuming that $L_g(x^*, \kappa^*; y^*, \lambda^*) = \inf_{\substack{x \in X \\ \kappa \ge 0}} L_g(x, \kappa; y^*, \lambda^*)$, we

infer from Lemma 4b that $y* \in Y$ and $h_j(y^{*j}) \leq 0$, $j \in J$, and that

$L_g(x*, \kappa*; y*, \lambda*) = -H(y*, \lambda*)$. This in turn implies that

$$0 = \langle x*, y* \rangle$$

and

$$\kappa_j^* h_j(y^{*j}) = 0 \qquad j \in J,$$

by virtue of the hypotheses $x* \in X$ and $\kappa* \geq 0$. Conversely, assuming that $y* \in Y$ and $h_j(y^{*j}) \leq 0$, $j \in J$, and that $(x*, \kappa*)$ satisfies the preceding displayed relations, we infer from Lemma 4b that $\inf\limits_{\substack{x \in X \\ \kappa \geq 0}} L_g(x, \kappa; y*, \lambda*) =$

$-H(y*, \lambda*) = L_g(x*, \kappa*; y*, \lambda*)$.   q.e.d.

Since the second assertion of Theorem 4 gives certain conditions that are equivalent to the first saddle-point equation, and since the third assertion of Theorem 4 gives certain other conditions that are equivalent to the second saddle-point equation, Theorem 4 actually provides four different characterizations of all ordered pairs $(x*, \kappa*; y*, \lambda*)$ of optimal solutions $(x*, \kappa*)$ and P vectors $(y*, \lambda*)$.

Of course, each of those four characterizations provides a characterization of all optimal solutions $(x*, \kappa*)$ in terms of a given P vector $(y*, \lambda*)$, as well as a characterization of all P vectors $(y*, \lambda*)$ in terms of a given optimal solution $(x*, \kappa*)$.

Still another characterization of all optimal solutions $(x*, \kappa*)$ to certain problems A has been given by the author [3].

3.2. Duality. Corresponding to problem A is the following "geometric dual problem" B.

PROBLEM B. <u>Using the feasible solution set</u>

$$T \overset{\Delta}{=} \{(y,\lambda) \in D \mid \inf_{\substack{x \in X \\ \kappa \geq 0}} L_g(x,\kappa;\ y,\lambda) \ \underline{\text{is finite}}\}$$

<u>and the objective function</u>

$$\mathcal{N}(y,\lambda) \overset{\Delta}{=} \inf_{\substack{x \in X \\ \kappa \geq 0}} L_g(x,\kappa;\ y,\lambda),$$

<u>calculate both the problem supremum</u>

$$\Psi \overset{\Delta}{=} \sup_{(y,\lambda) \in T} \mathcal{N}(y,\lambda)$$

<u>and the optimal solution set</u>

$$T^* \overset{\Delta}{=} \{(y,\lambda) \in T \mid \mathcal{N}(y,\lambda) = \Psi\}.$$

Even though problem B is essentially a "maximin problem" -- a type
of problem that tends to be relatively difficult to analyse -- the minimi-
zation problems that must be solved to obtain the objective function $\mathcal{N}:T$
have trivial solutions. In particular, Lemma 4b clearly implies that

$$T = \{(y,\lambda) \in D \mid y \in Y, \text{ and } h_j(y^j) \leq 0, \ j \in J\} \text{ and } \mathcal{N}(y,\lambda) = -H(y,\lambda),$$

so problem B can actually be rephrased in the following more direct way.

PROBLEM B. <u>Consider the objective function H whose domain</u>

$$D \overset{\Delta}{=} \{(y,\lambda) \mid y^k \in D_k, \ k \in \{0\} \cup J, \ \underline{\text{and}} \ (y^i,\lambda_i) \in D_i^+, \ i \in I\}$$

<u>and whose functional value</u>

$$H(y,\lambda) \overset{\Delta}{=} h_0(y^0) + \sum_I h_i^+(y^i,\lambda_i),$$

<u>where</u>

$$D_i^+ \overset{\Delta}{=} \{(y^i, \lambda_i) \mid \underline{\text{either}} \ \lambda_i = 0 \ \underline{\text{and}} \ \sup_{c^i \in C_i} \langle c^i, y^i \rangle < +\infty, \ \underline{\text{or}} \ \lambda_i > 0 \ \underline{\text{and}} \ y^i \in \lambda_i D_i\}$$

<u>and</u>

$$h_i^+(y^i, \lambda_i) \overset{\Delta}{=} \begin{cases} \sup_{c^i \in C_i} \langle c^i, y^i \rangle & \underline{\text{if}} \ \lambda_i = 0 \ \underline{\text{and}} \ \sup_{c^i \in C_i} \langle c^i, y^i \rangle < +\infty \\[3mm] \lambda_i h_i (y^i / \lambda_i) & \underline{\text{if}} \ \lambda_i > 0 \ \underline{\text{and}} \ y^i \in \lambda_i D_i . \end{cases}$$

<u>Using the feasible solution set</u>

$$T \overset{\Delta}{=} \{(y, \lambda) \in D \mid y \in Y, \ \underline{\text{and}} \ h_j(y^j) \le 0, \ j \in J\},$$

<u>calculate both the problem infimum</u>

$$\psi \overset{\Delta}{=} \inf_{(y, \lambda) \in T} H(y, \lambda) = -\Psi$$

<u>and the optimal solution set</u>

$$T^* \overset{\Delta}{=} \{(y, \lambda) \in T \mid H(y, \lambda) = \psi\}.$$

When phrased in this way, problem B closely resembles problem A and is in fact a geometric programming problem. Of course, the geometric dual problem B can actually be defined in this way, but the preceding derivation serves as an important link between geometric Lagrangians and geometric duality (analogous to the link between ordinary Lagrangians and ordinary duality).

To further strengthen that link, we first need to develop the most basic duality theory -- a theory in which the following definition is

almost as important as the definition of the dual problems A and B.

DEFINITION. The <u>extremality conditions</u> (for constrained geometric

programming) are:

(I) $\qquad\qquad x \in X \qquad\qquad$ and $\qquad\qquad\qquad y \in Y$,

(II) $\qquad g_i(x^i) \leq 0, \ i \in I \qquad$ and $\qquad\qquad h_j(y^j) \leq 0, \ j \in J$,

(III) $\qquad\qquad\qquad\qquad 0 = \langle x, y \rangle$,

(IV) $\qquad\qquad\qquad\qquad y^0 \in \partial g_0(x^0)$,

(V) $\quad$ either $\lambda_i = 0$ and $\langle x^i, y^i \rangle = \sup\limits_{c^i \in C_i} \langle c^i, y^i \rangle$ or $\lambda_i > 0$ and $y^i \in \lambda_i \, \partial g_i(x^i)$,

$$i \in I,$$

(VI) $\quad$ either $\kappa_j = 0$ and $\langle x^j, y^j \rangle = \sup\limits_{d^j \in D_j} \langle x^j, d^j \rangle$ or $\kappa_j > 0$ and $y^j \in \partial g_j(x^j / \kappa_j)$,

$$j \in J,$$

(VII) $\quad \lambda_i g_i(x^i) = 0, \ i \in I, \qquad$ and $\qquad \kappa_j h_j(y^j) = 0, \ j \in J$.

$\qquad$ The following "duality theorem" is the basis for many important

theorems.

<u>Theorem</u> 5. $\ $ <u>If</u> (x, κ) <u>and</u> (y, λ) <u>are feasible solutions to problems</u> A <u>and</u>

B <u>respectively (in which case the extremality conditions</u> (I) <u>and</u> (II)

<u>are satisfied), then</u>

$$0 \leq G(x, \kappa) + H(y, \lambda),$$

<u>with equality holding if and only if the extremality conditions</u> (III)

<u>through</u> (VII) <u>are satisfied, in which case</u> (x, κ) <u>and</u> (y, λ) <u>are optimal</u>

<u>solutions to problems</u> A <u>and</u> B <u>respectively.</u>

Proof. The following lemma will also be used in the proof of other theorems.

Lemma 5a. If $(x, \kappa) \in C$ and $(y, \lambda) \in D$, then

$$\langle x, y \rangle \le G(x, \kappa) + \sum_I \lambda_i g_i(x^i) + H(y, \lambda) + \sum_J \kappa_j h_j(y^j),$$

with equality holding if and only if the extremality conditions (IV) through (VI) are satisfied. Moreover, if $h_j(y^j) \le 0$, $j \in J$ (i.e. the second part of extremality condition (II) is satisfied), then

$$G(x, \kappa) + \sum_I \lambda_i g_i(x^i) + H(y, \lambda) + \sum_J \kappa_j h_j(y^j) \le G(x, \kappa) + \sum_I \lambda_i g_i(x^i) + H(y, \lambda),$$

with equality holding if and only if the second part of extremality condition (VII) is satisfied. Furthermore, if $g_i(x^i) \le 0$, $i \in I$ (i.e. the first part of extremality condition (II) is satisfied), then

$$G(x, \kappa) + \sum_I \lambda_i g_i(x^i) + H(y, \lambda) \le G(x, \kappa) + H(y, \lambda),$$

with equality holding if and only if the first part of extremality condition (VII) is satisfied.

Proof. From the conjugate inequality presented in section 3.1.2 of [1] we know that

$$\langle x^0, y^0 \rangle \le g_0(x^0) + h_0(y^0),$$

with equality holding if and only if the extremality condition (IV) is satisfied. From the geometric inequality established in section 3.3.3 of [1] we know that

$$\langle x^i, y^i \rangle \leq \lambda_i g_i(x^i) + h_i^+(y^i, \lambda_i),$$

with equality holding if and only if the extremality condition (V) is satisfied. Likewise, we know that

$$\langle x^j, y^j \rangle \leq g_j^+(x^j, \kappa_j) + \kappa_j h_j(y^j),$$

with equality holding if and only if the extremality condition (VI) is satisfied. Adding all $1 + o(I) + o(J)$ of these inequalities and taking account of the defining equations for x, y, G, and H proves the first assertion. The second assertion is an immediate consequence of the fact that $\kappa_j \geq 0$ when $(x^j, \kappa_j) \in C_j^+$, $j \in J$; and the third assertion is an immediate consequence of the fact that $\lambda_i \geq 0$ when $(y^i, \lambda_i) \in D_i^+$, $i \in I$. q.e.d.

Now, the fact that x and y are in the cone X and its dual Y respectively combined with a sequential application of all three assertions of Lemma 5a shows that

$$0 \leq \langle x, y \rangle \leq G(x, \kappa) + \sum_I \lambda_i g_i(x^i) + H(y, \lambda) + \sum_J \kappa_j h_j(y^j)$$

$$\leq G(x, \kappa) + \sum_I \lambda_i g_i(x^i) + H(y, \lambda)$$

$$\leq G(x, \kappa) + H(y, \lambda),$$

with equality holding in all four inequalities if and only if the equality conditions stated in the theorem are satisfied. q.e.d.

The following important corollary is an immediate consequence of Theorem 5.

Corollary 5A. If the dual problems A and B are both consistent, then

   (i) the infimum φ for problem A is finite, and

$$0 \leq \varphi + H(y,\lambda) \text{ for each } (y,\lambda) \in T,$$

   (ii) the infimum ψ for problem B is finite, and

$$0 \leq \varphi + \psi.$$

   The strictness of the inequality in conclusion (ii) plays a crucial role in almost all duality theorems.

DEFINITION. Consistent dual problems A and B for which

$$0 < \varphi + \psi$$

have a duality gap of $\varphi + \psi$.

A much more thorough discussion of duality theory and the role played by duality gaps is given in [1] and some of the references cited therein.

   The link between geometric Lagrangians and geometric duality can now be futher strengthened by the following tie between dual problem B and the P vectors for problem A defined in section 3.1.

Theorem 6. Given that problem A is consistent with a finite infimum φ,

   (1) if problem A has a P vector, then problem B is consistent and
$$0 = \varphi + \psi,$$

   (2) if problem B is consistent and $0 = \varphi + \psi$, then

$$\{(y^*,\lambda^*) \mid (y^*,\lambda^*) \text{ \underline{is a} } P \text{ \underline{vector for problem}} A\} = T^*.$$

Proof. If (y^*,λ^*) is a P vector for problem A, then Lemma 4b

implies that (y^*,λ^*) is feasible for problem B and that $\varphi = -H(y^*,\lambda^*)$,

which in turn implies that $\psi = H(y^*,\lambda^*)$ by virtue of conclusion (i) to

Corollary 5A. Consequently, $0 = \varphi + \psi$ and $(y^*,\lambda^*) \in T^*$.

On the other hand, if problem B is consistent and $0 = \varphi + \psi$, then each

vector (y^*,λ^*) in T^* has the property $\varphi = -H(y^*,\lambda^*)$, and hence each such

vector (y^*,λ^*) is a P vector for problem A by virtue of the first two

paragraphs of this subsection.   q.e.d.

An important consequence of Theorem 6 is that, when they exist, all

P vectors for problem A can be obtained simply by computing the dual

optimal solution set T^*. However, there are cases in which the vectors

in T^* are not P vectors for problem A, though such cases can occur only

when $0 < \varphi + \psi$, in which event Theorem 6 implies that there can be no P

vectors for problem A.

The following theorem provides an important tie between dual problem

B and the Kuhn-Tucker vectors for problem A defined in section 3.1.

Theorem 7. <u>Given that problems A and B are both consistent and that</u>

$0 = \varphi + \psi$, <u>if there is a minimizing sequence</u> $\{(y^q,\lambda^q)\}_1^\infty$ <u>for problem B</u> (i.e.

$(y^q,\lambda^q) \in T$ <u>and</u> $\lim_{q \to +\infty} H(y^q,\lambda^q) = \psi$) <u>such that</u> $\lim_{q \to +\infty} \lambda^q$ <u>exists and is finite,</u>

<u>then</u> $\lambda^* \overset{\Delta}{=} \lim_{q \to +\infty} \lambda^q$ <u>is a Kuhn-Tucker vector for problem</u> A.

Proof. The feasibility of (y^q,λ^q) implies that $\lambda^q \geq 0$ for each q, so

$$\lambda_i^* \geq 0, \qquad\qquad i \in I.$$

Now, the feasibility of (y^q, λ^q) and a sequential application of the first two assertions of Lemma 5a show that

$$\langle x, y^q \rangle \leq G(x, \kappa) + \sum_I \lambda_i^q g_i(x^i) + H(y^q, \lambda^q) + \sum_J \kappa_j h_j(y^{qj})$$

$$\leq G(x, \kappa) + \sum_I \lambda_i^q g_i(x^i) + H(y^q, \lambda^q)$$

for each $(x, \kappa) \in C$ and for each q. Consequently, for each $(x, \kappa) \in C$ such that $x \in X$ we deduce that

$$0 \leq G(x, \kappa) + \sum_I \lambda_i^q g_i(x^i) + H(y^q, \lambda^q)$$

for each q, because $y^q \in Y$. This inequality and the hypotheses

$$\lim_{q \to +\infty} H(y^q, \lambda^q) = \psi \text{ and } \lim_{q \to +\infty} \lambda^q = \lambda* \text{ clearly imply that}$$

$$0 \leq G(x, \kappa) + \sum_I \lambda_i^* g_i(x^i) + \psi$$

for each $(x, \kappa) \in C$ such that $x \in X$. Using the fact that $L_o(x, \kappa; \lambda*) \overset{\Delta}{=} G(x, \kappa) + \sum_I \lambda_i^* g_i(x^i)$ and the hypothesis $0 = \varphi + \psi$, we infer from the preceding inequality that

$$\varphi \leq \inf_{\substack{(x, \kappa) \in C \\ x \in X}} L_o(x, \kappa; \lambda*).$$

Now, choose a minimizing sequence $\{(x^q, \kappa^q)\}_1^\infty$ for problem A, and then observe for each q that

$$L_o(x^q, \kappa^q; \lambda*) \leq G(x^q, \kappa^q),$$

because $\lambda_i^* \geq 0$ and $g_i(x^{qi}) \leq 0$, $i \in I$. From the construction of $\{(x^q, \kappa^q)\}_1^\infty$ we know that $(x^q, \kappa^q) \in C$ and $x^q \in X$ for each q and that $\varphi = \lim_{q \to +\infty} G(x^q, \kappa^q)$; so we conclude from the preceding two displayed inequalities that

$$\phi = \inf_{\substack{(x,\kappa) \in C \\ x \in X}} L_o(x,\kappa; \lambda^*). \qquad\qquad q.e.d.$$

The following corollary ties the dual optimal solution set T* directly to Kuhn-Tucker vectors for problem A.

Corollary 7A. Given that problems A and B are both consistent and that $0 = \phi + \psi$, each $(y^*,\lambda^*) \in T^*$ provides a Kuhn-Tucker vector λ^* for problem A.

The following corollary ties the set of all P vectors (y^*,λ^*) for problem A directly to the set of all Kuhn Tucker vectors λ^* for problem A.

Corollary 7B. Given that problem A is consistent with a finite infimum ϕ, each P vector (y^*,λ^*) for problem A provides a Kuhn-Tucker vector λ^* for problem A.

Proof. Use Theorem 6 along with Corollary 7A.

Finally, it is worth noting that Theorem 7 and its two corollaries also provide certain connections between dual problem B and the "ordinary dual problem" corresponding to problem A. Those connections are left to the reader's imagination while more subtle connections are given in [4].

References

1. Peterson, Elmor L., "Geometric Programming", SIAM Review, 18(1976), 1-51.

2. Rockafellar, R.T., Convex Analysis, Princeton University Press, Princeton, New Jersey (1970).

3. Peterson, Elmor L., "Optimality Conditions in Generalized Geometric Programming", to appear.

4. _____, "Geometric Duality vis-a-vis Ordinary Duality", in preparation.