Discussion Paper No. 221

Optimality Conditions in Generalized Geometric Programming

by

Elmor L. Peterson

May 20, 1976

Optimality Conditions in Generalized Geometric Programming

by

Elmor L. Peterson*

Abstract. Generalizations of the Kuhn-Tucker optimality conditions are given, as are the fundamental theorems having to do with their necessity and/or sufficiency.

CONTENTS

Key words: optimality conditions, geometric programming, nonlinear programming, ordinary programming, Kuhn-Tucker conditions.

*Department of Industrial Engineering/Management Sciences and Department of Mathematics, Northwestern University, Evanston, Illinois 60201.
Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-73-2516.

1. Introduction. Optimization problems from the real world usually possess a linear-algebraic component, either directly in the form of problem 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 following sections present optimality conditions that are tailored to geometric programming and can be viewed as generalizations of the Kuhn-Tucker optimality conditions. Although the fundamental theorems having to do with their necessity and/or sufficiency 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 unconstrained case is treated separately (even though its theory is actually embedded in the theory for the constrained case).

Mathematically, this paper is essentially self-contained.

2. <u>The unconstrained case</u>. Given a nonempty cone $\mathcal{X} \subseteq E_n$ (n-dimensional Euclidean space), and given a function g with a nonempty domain $\mathcal{C} \subseteq E_n$, the resulting "geometric programming problem" $\mathcal{A}$ is then defined in the following way.

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

$$\mathcal{S} \stackrel{\Delta}{=} \mathcal{X} \cap \mathcal{C},$$

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

$$\varphi \stackrel{\Delta}{=} \inf_{x \in \mathcal{S}} g(x)$$

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

$$\mathcal{S}^* \stackrel{\Delta}{=} \{x \in \mathcal{S} \mid g(x) = \varphi\}$$

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

Our optimality conditions for the preceding problem $\mathcal{A}$ utilize the "dual cone"

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

They are stated as part of the following definition.

DEFINITION. A <u>critical solution</u> (stationary solution, equilibrium solution, P solution) for problem $\mathcal{A}$ is any vector x^* that satisfies the following P optimality conditions:

$$x^* \in \mathcal{X} \cap \mathcal{C}, \qquad \nabla g(x^*) \in \mathcal{Y}$$

and

$$0 = \langle x^*, \nabla g(x^*) \rangle .$$

If the cone $\mathcal{X}$ is actually a vector space, then $\mathcal{Y} = \mathcal{X}^{\perp}$ and hence the P optimality condition $0 = \langle x^*, \nabla g(x^*) \rangle$ is redundant and can be deleted. Furthermore, in the ordinary programming case, the vector space $\mathcal{Y} = E_n^{\perp} = \{0\}$, so the remaining P optimality conditions become the (more familiar) "ordinary optimality conditions"

$$x^* \in \mathcal{C} \text{ and } \nabla g(x^*) = 0 .$$

The following theorem gives two convexity conditions that guarantee the necessity and/or sufficiency of the P optimality conditions.

Theorem 1. Under the hypothesis that g is differentiable at x^*,

(i) given that $\mathcal{X}$ is convex, if x^* is an optimal solution to problem $\mathcal{A}$, then x^* is a critical solution for problem $\mathcal{A}$ (but not conversely),

(ii) given that g is convex on $\mathcal{C}$, if x^* is a critical solution for problem $\mathcal{A}$, then x^* is an optimal solution to problem $\mathcal{A}$.

Proof. To prove part (i), first recall that the optimality of x^* implies that $x^* \in \mathcal{X} \cap \mathcal{C}$. Then, notice that the optimality of x^*, the differentiability of g at x^*. and the convexity of $\mathcal{X}$ imply that the directional derivative

$$\langle \nabla g(x^*), x \rangle \geq 0 \qquad \text{for each} \qquad x \in \mathcal{X}.$$

Likewise, the optimality of x^*, the differentiability of g at x^*, and the observation that $x^* + s(-x^*) \in \mathcal{X}$ for $s \leq 1$ imply that the directional derivative

$$\langle \nabla g(x^*), -x^* \rangle \geq 0.$$

Consequently, $\nabla g(x^*) \in \mathcal{Y}$ and $0 = \langle x^*, \nabla g(x^*) \rangle$.

Counterexamples to the converse of part (i) are numerous and easy to construct. In fact, the reader is probably already familiar with counterexamples from the ordinary programming case.

To prove part (ii), first recall that the convexity of g and the differentiability of g at x^* imply that

$$g(x) - g(x^*) \geq \langle \nabla g(x^*), x - x^* \rangle \qquad \text{for each} \qquad x \in \mathcal{C}.$$

Then, notice that the assumptions $0 = \langle x^*, \nabla g(x^*) \rangle$ and $\nabla g(x^*) \in \mathcal{Y}$ imply that

$$\langle \nabla g(x^*), x - x^* \rangle = \langle \nabla g(x^*), x \rangle \geq 0 \qquad \text{for each} \qquad x \in \mathcal{X}.$$

From the preceding displayed relations we see that $g(x) - g(x^*) \geq 0$ for each $x \in \mathcal{X} \cap \mathcal{C}$. Consequently, the assumption $x^* \in \mathcal{X} \cap \mathcal{C}$ shows that x^* is optimal for problem $\mathcal{A}$. q.e.d.

It is worth noting that g is differentiable everywhere for most of the examples given in section 2.1 of [1]. Moreover, $\mathcal{X}$ is polyhedral and hence convex for each of those examples, and g is convex for important special cases of each of those examples. Consequently, the P optimality conditions frequently characterize the optimal solution set $\mathcal{S}^*$ for problem $\mathcal{A}$.

3. <u>The constrained case</u>. To extend geometric programming by the explicit inclusion of (generally nonlinear) constraint functions, 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:

(1a) For each $k \in \{0\} \cup I \cup J$ there is a function g_k with a nonempty domain $C_k \subseteq E_{n_k}$, and there is a nonempty set $D_j \subseteq E_{n_j}$ for each $j \in J$.

(2a) For each $k \in \{0\} \cup I \cup J$ there is an independent vector variable x^k in E_{n_k}, and there is an independent vector variable κ with components κ_j for each $j \in J$.

(3a) x^I denotes the cartesian product of the vector variables x^i, $i \in I$, and x^J denotes the cartesian product of the vector variables x^j, $j \in J$. Hence, the cartesian product $(x^0, x^I, x^J) \stackrel{\Delta}{=} x$ is an independent vector variable in E_n, where

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

(4a) There is a nonempty cone $X \subseteq E_n$.

The resulting "geometric programming problem" A is then defined in the following way.

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

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

<u>and whose functional value</u>

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

where

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

$$\kappa_j > 0\ \underline{\text{and}}\ x^j \in \kappa_j C_j\}$$

and

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

Using the feasible solution set

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

calculate both the problem infimum

$$\varphi \stackrel{\Delta}{=} \inf_{(x,\kappa) \in S} G(x,\kappa)$$

and the optimal solution set

$$S^* \stackrel{\Delta}{=} \{(x,\kappa) \in S \mid G(x,\kappa) = \varphi\}.$$

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

$$J = \emptyset,$$

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

and

$$X \overset{\Delta}{=} \text{column space of} \begin{bmatrix} U \\ U \\ \cdot \\ \cdot \\ \cdot \\ U \end{bmatrix} \text{ where there is a total of } 1 + o(I) \text{ identity matrices } U \text{ that are } m \times m.$$

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

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

shows that problem A is then equivalent to the very general ordinary programming problem

<u>Minimize</u> $g_0(z)$ <u>subject to</u>

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

$$z \in C_o .$$

Our optimality conditions for the preceding problem A utilize the dual cone

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

whose vector variable y has the same cartesian-product structure as the vector variable x. They are stated as part of the following definition.

DEFINITION: A <u>critical solution</u> (stationary solution, equilibrium solution, P solution) for problem A is any vector (x^*, κ^*) for which there is a vector λ^* in $E_{o(I)}$ such that (x^*, κ^*) and λ^* jointly satisfy the following P <u>optimality conditions</u>:

$$x^* \in X,$$

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

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

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

$$y^* \in Y,$$

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

and

$$\langle x^{*j}, y^{*j} \rangle = g_j^+(x^{*j}, \kappa_j^*) \qquad j \in J,$$

where

$$y^{*0} \triangleq \nabla g_0(x^{*0}),$$

$$y^{*i} \triangleq \lambda_i^* \nabla g_i(x^{*i}) \qquad i \in I,$$

and

$$y^{*j} \triangleq \nabla g_j(x^{*j}/\kappa_j^*) \qquad j \in J.$$

Needless to say, if the cone X is actually a vector space, then $Y = X^{\perp}$ and hence the P optimality condition $0 = \langle x^*, y^* \rangle$ is redundant and can be deleted. Furthermore, in the ordinary programming case, the vector space

$$Y = \{y \in E_n \mid y^0 + \sum_I y^i = 0\},$$

so the remaining P optimality conditions are essentially the (more familiar) "Kuhn-Tucker optimality conditions"

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

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

$$\lambda_i^* g_i(z^*) = 0 \qquad i \in I,$$

and

$$\nabla g_0(z^*) + \sum_I \lambda_i^* \nabla g_i(z^*) = 0 .$$

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

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

and

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

where the (ordinary) Lagrangian

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

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 two convexity conditions that guarantee the necessity and/or sufficiency of the P optimality conditions.

Theorem 2. Under the hypotheses that g_k is differentiable at x^{*k}, $k \in \{0\} \cup I \cup J$ and that g_j^+ is differentiable at (x^{*j}, κ_j^*), $j \in J$,

(i) given that X is convex, if (x^*, κ^*) is an optimal solution to problem A and if λ^* is a Kuhn-Tucker vector for problem A, then (x^*, κ^*) is a critical solution for problem A relative to λ^* (but not conversely),

(ii) given that g_k is convex on C_k, $k \in \{0\} \cup I \cup J$, if (x^*, κ^*) is a critical solution for problem A relative to λ^*, then (x^*, κ^*) is an optimal solution to problem A and λ^* is a Kuhn-Tucker vector for problem A.

Proof. To prove part (i), first recall that the optimality of (x^*, κ^*) implies that $x^* \in X$ and that $g_i(x^{*i}) \leq 0$, $i \in I$. Then, note that the defining properties for a Kuhn-Tucker vector λ^* assert that $\lambda_i^* \geq 0$, $i \in I$ and that $\varphi \leq G(x^*, \kappa^*) + \sum_I \lambda_i^* g_i(x^{*i})$. Since $\varphi = G(x^*, \kappa^*)$, the preceding inequalities collectively imply that $\lambda_i^* g_i(x^{*i}) = 0$, $i \in I$; from which we infer that

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

Since our hypotheses clearly imply that $L_o(\cdot, \kappa; \lambda^*)$ is differentiable at (x^*, κ^*), the preceding equation and the convexity of X imply that the directional derivative

$$\langle \nabla_x L_o(x^*, \kappa^*; \lambda^*),\ x \rangle \geq 0 \qquad \text{for each} \qquad x \in X.$$

Likewise, the differentiability of $L_o(\cdot, \kappa; \lambda^*)$ at (x^*, κ^*), the preceding equation, and the observation that $x^* + s(-x^*) \in X$ for $s \leq 1$ imply that the directional derivative

$$\langle \nabla_x L_o(x^*, \kappa^*; \lambda^*), - x^* \rangle \geq 0.$$

Consequently, $\nabla_x L_o(x^*, \kappa^*; \lambda^*) \in Y$ and $0 = \langle x^*, \nabla_x L_o(x^*, \kappa^*; \lambda^*) \rangle$, which means that $y^* \in Y$ and $0 = \langle x^*, y^* \rangle$. Finally, since our hypothesis that g_j^+ is differentiable at (x^{*j}, κ_j^*), $j \in J$ clearly implies that $\kappa^* > 0$, the differentiability of $L_o(\cdot, \chi; \lambda^*)$ at (x^*, κ^*) and the preceding displayed equation imply that

$$\nabla_\kappa L_o(x^*, \kappa^*; \lambda^*) = 0,$$

which means that $\langle x^{*j}, y^{*j} \rangle = g_j^+(x^{*j}, \kappa_j^*)$, $j \in J$.

Counterexamples to the converse of part (i) are numerous and easy to construct. In fact, the reader is probably already familiar with counterexamples from the ordinary programming case.

To prove part (ii), first observe that $L_o(\cdot, \chi; \lambda^*)$ is convex on C and that $L_o(\cdot, \chi; \lambda^*)$ is differentiable at (x^*, κ^*). These two observations together imply that

$$L_o(x, \kappa; \lambda^*) - L_o(x^*, \kappa^*; \lambda^*) \geq \langle \nabla_{(x,\kappa)} L_o(x^*, \kappa^*; \lambda^*), (x, \kappa) - (x^*, \kappa^*) \rangle \quad \text{for each } (x, \kappa) \in C.$$

Since the assumption that $\langle x^{*j}, y^{*j} \rangle = g_j^+(x^{*j}, \kappa_j^*)$, $j \in J$ simply means that $\nabla_\kappa L_o(x^*, \kappa^*; \lambda^*) = 0$, elementary linear algebra shows that

$$\langle \nabla_{(x,\kappa)} L_o(x^*, \kappa^*; \lambda^*), (x, \kappa) - (x^*, \kappa^*) \rangle = \langle \nabla_x L_o(x^*, \kappa^*; \lambda^*), x - x^* \rangle \text{ for each } (x, \kappa).$$

Since it is clear that $\nabla_x L_o(x^*, \kappa^*; \lambda^*) = y^*$, the assumptions that $0 = \langle x^*, y^* \rangle$

and $y^* \in Y$ imply that

$$\langle \nabla_x L_o(x^*, \kappa^*; \lambda^*),\ x - x^* \rangle = \langle \nabla_x L_o(x^*, \kappa^*; \lambda^*),\ x \rangle \geq 0 \text{ for each } (x, \kappa) \text{ for which } x \in \mathbf{X}.$$

From the preceding displayed relations we infer that

$$L_o(x, \kappa; \lambda^*) - L_o(x^*, \kappa^*; \lambda^*) \geq 0 \text{ for each } (x, \kappa) \in C \text{ for which } x \in X.$$

Using this inequality and the assumption that $\lambda_i^* g_i(x^{*i}) = 0$, $i \in I$, we see that

$$G(x^*, \kappa^*) \leq G(x, \kappa) + \sum_I \lambda_i^* g_i(x^i) \text{ for each } (x, \kappa) \in C \text{ for which } x \in \mathbf{X}.$$

On the other hand, the assumption that $\lambda_i^* \geq 0$, $i \in I$ guarantees that

$$G(x, \kappa) + \sum_I \lambda_i^* g_i(x^i) \leq G(x, \kappa) \text{ for each } (x, \kappa) \in C \text{ for which } g_i(x^i) \leq 0,\ i \in I.$$

From the preceding two displayed inequalities we infer that

$$G(x^*, \kappa^*) \leq G(x, \kappa) \text{ for each } (x, \kappa) \in S.$$

Consequently, the assumptions that $g_i(x^{*i}) \leq 0$, $i \in I$ and that $x^* \in \mathbf{X}$ imply that (x^*, κ^*) is optimal for problem A, which means of course that $\varphi = G(x^*, \kappa^*)$. Using these facts and the assumption that $\lambda_i^* g_i(x^{*i}) = 0$, $i \in I$, we infer from the last displayed inequality involving L_o that λ^* is a Kuhn-Tucker vector for problem A. q.e.d.

It is worth noting that for most of the examples given or alluded to in section 2.2 of [1] the g_k, $k \in \{0\} \cup I \cup J$ are differentiable everywhere while either J is empty or the g_j^+, $j \in J$ are differentiable everywhere except at the origin. Moreover, X is polyhedral and hence convex for each of those examples, and the g_k, $k \in \{0\} \cup I \cup J$ are convex for important special cases of each of those examples. Consequently, the P optimality conditions frequently characterize the optimal solution set S^* for problem A.

Characterizations of S^* that do not require differentiability of the g_k, $k \in \{0\} \cup I \cup J$ and the g_j^+, $j \in J$, but do require conjugate transform theory are given in [2].

References

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

2. ________________, "Lagrangian Saddle Points and Duality in Generalized Geometric Programming", in preparation.