THE GENERALIZED INVERSE DUAL FOR QUADRATICALLY CONSTRAINED QUADRATIC PROGRAMS

By

WILLIAM DAVID RANDOLPH

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1974

To Alicia

ACKNOWLEDGMENTS

The author wishes to take this opportunity to express his gratitude to all those who lent support in this research effort. Thanks are especially due Dr. D. W. Hearn for his suggestion of the research topic and his continued support in the pursuit of meaningful results. The author would also like to thank Dr. M. E. Thomas, Dr. H. D. Ratliff, Dr. R. L. Francis and Dr. N. Passell for their advice and helpful suggestions as members of his supervisory committee.

The author wishes to express his appreciation also to the many faculty and fellow students whom he studied with at Rose Polytechnic Institute, The George Washington University, and the University of Florida.

Finally, thanks go to his wife, Alicia, for her confidence, encouragement, and support.

This research was supported in part by U. S. Army Contract No. DA-ARO-D-31-124-72-G123.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Abstract of Dissertation Presented to the
Graduate Council of the University of Florida in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy

THE GENERALIZED INVERSE DUAL FOR
QUADRATICALLY CONSTRAINED QUADRATIC PROGRAMS

By

William David Randolph

June, 1974

Chairman: Dr. D. W. Hearn
Major Department: Industrial and Systems Engineering

The generalized inverse dual (GID) and its properties are developed for convex quadratically constrained quadratic programs, based on the Moore-Penrose generalized inverse of a matrix. The resulting dual problem has a concave objective function and the constraints are shown to be effectively linear. The closure of the dual constraint set is characterized by an index maximal dual vector.

It is also shown that the GID is equivalent to the conjugate function dual (CFD), developed through conjugate function theory or generalized geometric inequalities, and has significantly fewer variables.

Comparison of the two duals, GID and CFD, for a projected gradient algorithm, demonstrates that for strictly convex problems the GID is favored over the CFD; some experimental results for the GID are given. For nonstrictly convex problems some advantages of the GID are discussed, e.g., determining an initial dual feasible vector.

Five areas of application where the GID could prove to have significant computational advantages are discussed: multifacility Euclidean distance location problem, stochastic programming, the general Fermat problem, portfolio selection, and general convex programming.

CHAPTER 1

INTRODUCTION

This dissertation is concerned with convex quadratically constrained quadratic programs.

Various examples and references to such programs have occurred in the mathematical programming literature. John [37] in his classic paper of 1948 presented as an example a quadratically constrained program with a linear objective function. The special structure of quadratically constrained quadratic programs was also recognized by Kuhn and Tucker [40] and used as an example saddle point problem.

Other examples of quadratically constrained quadratic programs are found in the field of chance-constrained programming; see Charnes and Cooper [11], Euclidean distance location problems, see Elzinga and Hearn [25], Hearn [34], and Francis [29,30,31].

Some authors have addressed specific quadratic problems. For example, Bellman [3] discusses a solution technique for a quadratic program with one quadratic constraint. The procedure is based on simultaneous diagonalization of the two Hessians and a change of variables. van de Panne [69] presents a finite algorithm for a linear objective function with linear constraints and one quadratic constraint.

Major developments in this area have been the 1967 generalized geometric inequality dual of Peterson and Ecker [51,52,53], and the conjugate function duality of Rockafellar [57]. The two developments have one point in common; neither appears to relate to the duality of

Wolfe [71] which has been a powerful tool in mathematical programming.

A more recent work is that of Baron [2]. Baron discusses and develops computational procedures for the class of quadratically constrained quadratic programs using two forms of the Lagrangian function. The duality theory employed was first shown by Falk [26] for strictly convex programs. Baron uses the well-known convex programming algorithm of Dantzig [20] for solution of these Lagrangian dual problems.

The research reported herein develops a new dual form for the class of convex quadratically constrained quadratic programs.

In Chapter 2 some extensions to the theory of the Moore-Penrose [47,50] generalized inverse for sums of positive semidefinite matrices are developed. In particular, general theorems are stated which are germane to the developments of Chapter 3.

In Chapter 3 the generalized inverse dual and its properties for convex quadratically constrained quadratic programs are developed. The only theoretical basis required for the generalized inverse dual is linear algebra and Wolfe [71] duality. It is also shown that the corresponding duals of Peterson and Ecker and Rockafellar are derivable from the generalized inverse dual and conversely.

Chapter 4 is a discussion of a feasible direction algorithm for the generalized inverse dual and the conjugate function dual of Peterson and Ecker and Rockafellar. It is concluded that for a particular subclass of convex quadratically constrained quadratic programs the generalized inverse dual algorithm should prove more favorable than that for the conjugate function dual.

Chapter 5 presents five applications where the generalized inverse dual leads to more computationally tractable mathematical programming problems.

Chapter 6 sets out specific areas of future research based on the results reported herein.

CHAPTER 2

THE MOORE-PENROSE GENERALIZED INVERSE OF POSITIVE SEMIDEFINITE MATRICES

2.1 Introduction

The development of the duality theory for convex quadratically constrained quadratic programs encompasses specializations of the following extensions to the theory of the Moore-Penrose generalized inverse. These extensions are developed and presented at this time due to their importance and the fact that they are not known to exist elsewhere in the literature.

The theory of this chapter is restricted to sums of symmetric positive semidefinite matrices. The reason for this restriction will be evident in Chapter 3.

2.2 Extensions

Notationally, let Q_j, $j = 1,2,\ldots,m$, be symmetric positive semidefinite matrices of order nxn and rank ρ_j. Let

$$Q = \sum_{j=1}^{m} Q_j.$$

Then Q is a symmetric positive semidefinite matrix.

Lemma 2.1. Let Q^- be the generalized inverse of Q, then

$$Q^-Q = QQ^-.$$

Proof of Lemma 2.1. By the properties of the generalized inverse Q^-Q is symmetric,

$$Q^- Q = (Q^- Q)^t$$
$$= Q^t Q^{-t}.$$

From Corollary A.3.1, $Q = Q^t$ implies that $(Q^t)^- = Q^-$, hence,

$$Q^- Q = QQ^-.$$

Q.E.D.

Theorem 2.1. Let Q^- be the generalized inverse of Q, then

$$Q_j(I-Q^-Q) = 0,$$
$$Q_j(I-QQ^-) = 0,$$
$$(I-QQ^-)Q_j = 0,$$
$$(I-Q^-Q)Q_j = 0.$$

Proof of Theorem 2.1. By definition

$$QQ^-Q = Q,$$
$$Q(I-Q^-Q) = 0,$$
$$\sum_{j=1}^{m} Q_j(I-Q^-Q) = 0. \tag{2.2.1}$$

From Theorem B.2, the trace of the lefthand side matrix of (2.2.1) is

$$\mathrm{tr}[\sum_{j=1}^{m} Q_j(I-Q^-Q)] = \sum_{j=1}^{m} \mathrm{tr}[Q_j(I-Q^-Q)] = 0.$$

Now Q_j is symmetric positive semidefinite and by Theorem A.10, $(I-Q^-Q)$ is positive semidefinite; hence, by Corollary B.4.1,

$$\mathrm{tr}[Q_j(I-Q^-Q)] \geq 0, \quad \text{for } j = 1,2,\ldots,m.$$

For a sum of nonnegative terms to add to zero they must all equal zero. Therefore,

$$\mathrm{tr}(Q_j(I-Q^-Q)) = 0 \text{ and } Q_j(I-Q^-Q) = 0$$

by Theorem B.4.

The other forms follow in the same manner using Lemma 2.1 and the alternate form of (2.2.1)

$$\sum_{j=1}^{m} (I-QQ^-)Q_j = 0.$$

Q.E.D.

A notational device for weighted sums of matrices will be $\hat{Q} = \hat{Q}(y) = \sum_{j=1}^{m} y_jQ_j$, $y_j \in E^1$. If there is more than one set of weights, the distinguishing character will be adjoined to $\hat{Q}$, e.g., y_j' and $\hat{Q}'$. For emphasis the functional notation $\hat{Q}(y)$ will be retained.

Corollary 2.1.1. Let $\hat{Q} = \sum_{j=1}^{m} y_jQ_j$, $y_j > 0$, $j = 1,2,\ldots,m$, then

$$Q_j(I-\hat{Q}^-\hat{Q}) = 0,$$

$$Q_j(I-\hat{Q}\hat{Q}^-) = 0,$$

$$(I-\hat{Q}\hat{Q}^-)Q_j = 0,$$

$$(I-\hat{Q}^-\hat{Q})Q_j = 0.$$

Corollary 2.1.2. Corollary 2.1.1 holds when Q_j is replaced by Q_j^-, the generalized inverse of Q_j.

Proof of Corollary 2.1.2. It will suffice to show that $(I-\hat{Q}\hat{Q}^-)Q_j^- = 0$ since the same approach is used in all cases.

$$(I-\hat{Q}\hat{Q}^-)Q_j^- = (I-\hat{Q}\hat{Q}^-)Q_j^-Q_jQ_j^-$$

since Q_j is symmetric, by Lemma 2.1,

$$(I-\hat{Q}\hat{Q}^-)Q_j^- = (I-\hat{Q}\hat{Q}^-)Q_jQ_j^-Q_j^- = 0.$$

Q.E.D.

Theorem 2.2. Let $\hat{Q} = \hat{Q}(y) = \sum_{j=1}^{m} y_jQ_j$ and $\hat{Q}' = \hat{Q}(y') = \sum_{j=1}^{m} y_j'Q_j$, where $y_j > 0$, $y_j' > 0$ for $j = 1,2,\ldots,m$, and $y_j' \neq y_j$. Then $\hat{Q}^-\hat{Q} = \hat{Q}'^-\hat{Q}'$.

Proof of Theorem 2.2. The proof will be to show $(I-\hat{Q}^-\hat{Q})$ and $(I-\hat{Q}'^-\hat{Q}')$ are both generalized inverses of $(I-\hat{Q}^-\hat{Q})$ which by the uniqueness of the generalized inverse, Theorem A.2, implies

$$(I-\hat{Q}^{-}\hat{Q}) = (I-\hat{Q}'^{-}\hat{Q}').$$

Corollary A.8.2 and Theorems A.9 and A.10 show the generalized inverse of $(I-\hat{Q}^{-}Q)$ as being $(I-\hat{Q}^{-}\hat{Q})$. Now, using Lemma 2.1 and Corollary 2.1.1, it is easily shown that $(I-\hat{Q}'^{-}\hat{Q}')$ satisfies the four properties of the generalized inverse, e.g.,

$$
\begin{aligned}
(i)\quad (I-\hat{Q}^{-}\hat{Q})(I-\hat{Q}'^{-}\hat{Q}') &= (I-\hat{Q}^{-}\hat{Q}) - (I-\hat{Q}^{-}\hat{Q})\hat{Q}'^{-}\hat{Q}' \\
&= (I-\hat{Q}^{-}\hat{Q}) - (I-\hat{Q}^{-}\hat{Q})\hat{Q}'\hat{Q}'^{-} \\
&= (I-\hat{Q}^{-}\hat{Q}) - \sum_{j=1}^{m} y_j'(I-\hat{Q}^{-}\hat{Q})Q_j\hat{Q}'^{-} \\
&= (I-\hat{Q}^{-}\hat{Q}), \text{ hence symmetric}
\end{aligned}
$$

and therefore $(I-\hat{Q}'^{-}\hat{Q}')$ is also a generalized inverse of $(I-\hat{Q}^{-}\hat{Q})$. Thus,

$$(I-\hat{Q}^{-}\hat{Q}) = (I-\hat{Q}'^{-}\hat{Q}')$$

and

$$\hat{Q}^{-}\hat{Q} = \hat{Q}'^{-}\hat{Q}'.$$

Q.E.D.

Corollary 2.2.1.

$$\hat{Q}^{-}\hat{Q} = \hat{Q}'\hat{Q}'^{-}$$

$$\hat{Q}\hat{Q}^{-} = \hat{Q}'^{-}\hat{Q}'$$

$$\hat{Q}\hat{Q}^{-} = \hat{Q}'\hat{Q}'^{-}.$$

Proof of Corollary 2.2.1. By Lemma 2.1,

$$\hat{Q}^{-}\hat{Q} = \hat{Q}\hat{Q}^{-},$$

$$\hat{Q}'^{-}\hat{Q}' = \hat{Q}'\hat{Q}'^{-}$$

and the result is immediate.

Q.E.D.

Another important corollary to Theorem 2.2 deals with the column space and null space of a matrix. The following two definitions are stated for emphasis.

Definition. Let A be an nxm matrix; denote the m columns of A as vectors in E^n, so that $A = [a_1, a_2, \ldots, a_m]$. The vector space spanned by these m column vectors of A is defined as the column space of the matrix A.

Definition. Let A be an nxm matrix. The null space of the matrix A is defined to be the set of vectors S where

$$S = \{y \mid Ay = 0;\ y \in E^m\}.$$

If the matrix A is symmetric, the following lemma establishes the equivalence between its null space and orthogonal complement of the column space.

Lemma 2.2. Let A be an nxm matrix. The null space of A^t and the orthogonal complement of the column space of A are the same.

A proof of this lemma is found in Graybill [32].

Corollary 2.2.2. Given $\hat{Q} = \sum_{j=1}^{m} y_j Q_j$ and $\hat{Q}' = \sum_{j=1}^{m} y_j' Q_j$; y_j, $y_j' > 0$ for $j = 1,2,\ldots,m$. Then the column vectors of $\hat{Q}$ span the same space as do those of $\hat{Q}'$; i.e., $\hat{Q}$ and $\hat{Q}'$ have the same column space.

Proof of Corollary 2.2.2. By Lemma 2.2 and Corollary A.8.1 the null space of $\hat{Q}$ is spanned by $(I-\hat{Q}^-\hat{Q})$ and that of $\hat{Q}'$ by $(I-\hat{Q}'^-\hat{Q}')$. Let b be in the column space of $\hat{Q}$. Then,

$$(I-\hat{Q}^-\hat{Q})b = 0,$$

$$\hat{Q}^-\hat{Q}b = b,$$

but by Theorem 2.2

$$\hat{Q}^-\hat{Q} = \hat{Q}'^-\hat{Q}'$$

and

$$(I-\hat{Q}'^-\hat{Q}')b = 0.$$

Hence, b is in the column space of $\hat{Q}'$. Likewise, a vector b' in the column space of $\hat{Q}'$ is in the column space of $\hat{Q}$. Therefore, $\hat{Q}$ and $\hat{Q}'$ have the same column space.

Q.E.D.

Corollary 2.2.3. Given $\hat{Q} = \sum_{j=1}^{m} y_j Q_j$ and for some index set $I = \{i_1, i_2, \ldots, i_{m_1}\}$, $m_1 \leq m$, such that $I \subseteq \{1,2,\ldots,m\}$ let $\hat{Q}' = \sum_{j\varepsilon I} y_j' Q_j$, $y_j' > 0$ and $y_j' \neq y_j$, then the column space of $\hat{Q}'$ is contained in or equal to the column space of $\hat{Q}$.

Proof of Corollary 2.2.3. Take b' in the column space of $\hat{Q}'$ then

$$\hat{Q}'\hat{Q}'^{-}b' = b',$$

premultiplying by $\hat{Q}\hat{Q}^{-}$ and applying Corollary 2.1.1 it is seen that

$$\hat{Q}\hat{Q}^{-}\hat{Q}'\hat{Q}'^{-}b' = \hat{Q}'\hat{Q}'^{-}b' = \hat{Q}\hat{Q}^{-}b' = b'.$$

Hence, the column space of $\hat{Q}'$ is contained in the column space of $\hat{Q}$. Now take b in the column space of $\hat{Q}$,

$$\hat{Q}\hat{Q}^{-}b = b.$$

Premultiply by $(I-\hat{Q}'\hat{Q}'^{-})$; i.e., if b is in the column space of $\hat{Q}'$ it will have zero component in the null space of $\hat{Q}'$.

$$(I-\hat{Q}'\hat{Q}'^{-})\hat{Q}\hat{Q}^{-}b = (I-\hat{Q}'\hat{Q}'^{-})b,$$

$$\sum_{j\notin I} y_j(I-\hat{Q}'\hat{Q}'^{-})Q_j\hat{Q}^{-}b = (I-\hat{Q}'\hat{Q}'^{-})b,$$

by Corollary 2.1.1 and clearly only if Q_j, for $j\notin I$, is in the column space of $\hat{Q}'$ will b have zero component in the null space of $\hat{Q}'$. Hence, in general, the column space of $\hat{Q}$ will not be properly included in the column space of $\hat{Q}'$.

Q.E.D.

An alternative and useful way of describing the matrix $\hat{Q}$, through decomposition, is based on the following lemma.

Lemma 2.3. An nxn symmetric positive semidefinite matrix Q can be factored as $Q = B^tB$, where B is an nxn upper triangular matrix. For Q of rank ρ, B is of the form

$$B = [\bar{B}^t, 0]^t,$$

$\bar{B}$ being of order ρxn.

Proof of Lemma 2.3. It is well known [7] that a symmetric positive semidefinite matrix Q of rank ρ can be factored into $\bar{B}^t\bar{B}$ where $\bar{B}$ is an upper triangular matrix of order ρxn. Form the nxn matrix $B = [\bar{B}^t, 0]^t$, 0 being a zero matrix of order nx(n-ρ), then

$$B^tB = [\bar{B}^t, 0] \begin{bmatrix} \bar{B} \\ \\ 0^t \end{bmatrix} = \bar{B}^t\bar{B} + 0 = Q,$$

conformability requirements being satisfied.

Q.E.D.

The matrix $\hat{Q}$, which is a sum of symmetric positive semidefinite matrices Q_j, j = 1,2,...,m, can now be expressed as

$$\hat{Q} = \sum_{j=1}^{m} y_j B_j^t B_j.$$

which can be put into a matrix product form,

$$\hat{Q} = \bar{B}^t Y^2 \bar{B}$$

where

$$\bar{B}^t = [B_1^t, B_2^t, \ldots, B_m^t]$$

and

$$Y^2 = I x \begin{bmatrix} y_1 & 0 & \ldots & 0 \\ 0 & y_2 & \ldots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \ldots & y_m \end{bmatrix}$$, I the nxn identity matrix.

Y^2 is defined by a Kronecker product; see Appendix A for definition.

A further simplification of this notational form can be realized by noting that Y is defined as having diagonal elements $\sqrt{y_j}$, then

$$\hat{Q} = \hat{B}^t\hat{B}$$

with

$$\hat{B} = \hat{B}(y) = Y\bar{B}.$$

<u>Theorem 2.3.</u>

$$\hat{Q}\hat{Q}^- = \hat{B}^t\hat{B}^{t-} = \hat{B}^-\hat{B}.$$

<u>Proof of Theorem 2.3.</u>

$$\hat{Q}\hat{Q}^- = \hat{B}^t\hat{B}(\hat{B}^t\hat{B})^-$$

By Corollary A.3.2, $(\hat{B}^t\hat{B})^- = \hat{B}^-\hat{B}^{t-}$, then

$$\begin{aligned}\hat{Q}\hat{Q}^- &= \hat{B}^t\hat{B}\hat{B}^-\hat{B}^{t-}\\ &= \hat{B}^t\hat{B}^{t-}\hat{B}^t\hat{B}^{t-}\\ &= \hat{B}^t\hat{B}^{t-}\\ &= \hat{B}^-\hat{B}.\end{aligned}$$

Q.E.D.

<u>Corollary 2.3.1.</u>

$$\hat{B}^t\hat{B}^{t-}Q_j = \hat{B}^-\hat{B}Q_j = Q_j,$$

$$Q_j\hat{B}^t\hat{B}^{t-} = Q_j\hat{B}^-\hat{B} = Q_j.$$

<u>Proof of Corollary 2.3.1.</u> By Corollary 2.1.1,

$$\hat{Q}\hat{Q}^-Q_j = Q_j,$$

$$Q_j\hat{Q}\hat{Q}^- = Q_j,$$

and the result is immediate.

Q.E.D.

2.3 <u>Differential Forms</u>

In order to express the differential of the generalized inverse of $\hat{Q} = \sum_{j=1}^{m} y_jQ_j$, $y_j > 0$ for $j = 1,2,\ldots,m$, it will be useful to state the following lemma.

The lemma, while well known, is stated and proved to emphasize the correspondence between the partial derivative forms for the nonsingular inverse and the generalized inverse. It is assumed that $\hat{Q}$ is a function of y_j with the Q_j fixed for $j = 1,2,\ldots,m$.

Lemma 2.4. Given $\hat{Q} = \sum_{j=1}^{m} y_j Q_j$, $y_j > 0$ for $j = 1,2,\ldots,m$ and $\hat{Q}$ is nonsingular, then

$$\frac{\partial \hat{Q}^{-1}}{\partial y_k} = -\hat{Q}^{-1} Q_k \hat{Q}^{-1}$$

Proof of Lemma 2.4.

$$\hat{Q}^{-1}\hat{Q} = I$$

$$\frac{\partial \hat{Q}^{-1}}{\partial y_k} \hat{Q} + \hat{Q}^{-1} Q_k = 0$$

$$\frac{\partial \hat{Q}^{-1}}{\partial y_k} = -\hat{Q}^{-1} Q_k \hat{Q}^{-1}.$$

Q.E.D.

Using the alternate definition of the generalized inverse the partial derivative of $\hat{Q}^-$ can be given for $\hat{Q}$ singular. It is seen that while the proof is more involved a similar form is derived as for the nonsingular $\hat{Q}$ and as expected the generalized form is equivalent to that of Lemma 2.4 for $\hat{Q}$ nonsingular.

Theorem 2.4. Given $\hat{Q} = \sum_{j=1}^{m} y_j Q_j$, $y_j > 0$ for $j = 1,2,\ldots,m$, then

$$\frac{\partial \hat{Q}^-}{\partial y_k} = -\hat{Q}^- Q_k \hat{Q}^-.$$

Proof of Theorem 2.4. The alternate definition of the generalized inverse of $\hat{Q}$ is

$$\hat{Q}^- = \lim_{\delta \to 0} (\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q}$$

where $(\hat{Q}\hat{Q} + \delta^2 I)$ is a nonsingular symmetric matrix for all $\delta > 0$. The partial derivative of $\hat{Q}^-$ in respect to y_k is then

$$\frac{\partial \hat{Q}^-}{\partial y_k} = \lim_{\delta \downarrow 0} \frac{\partial}{\partial y_k} (\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q} + (\hat{Q}\hat{Q} + \delta^2 I)^{-1} Q_k. \tag{2.3.1}$$

By Lemma 2.4,

$$\frac{\partial}{\partial y_k} (\hat{Q}\hat{Q} + \delta^2 I)^{-1} = -(\hat{Q}\hat{Q} + \delta^2 I)^{-1}(Q_k\hat{Q} + \hat{Q}Q_k)(\hat{Q}\hat{Q} + \delta^2 I)^{-1},$$

which when substituted into (2.3.1) results in,

$$\begin{aligned} \frac{\partial \hat{Q}^-}{\partial y_i} = & \lim_{\delta \downarrow 0} (\hat{Q}\hat{Q} + \delta^2 I)^{-1} Q_k \\ & -\lim_{\delta \downarrow 0} (\hat{Q}\hat{Q} + \delta^2 I)^{-1} Q_k \hat{Q}(\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q} \\ & -\lim_{\delta \downarrow 0} (\hat{Q}\hat{Q} + \delta^2 I)^{-1} \hat{Q}Q_k(\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q}. \end{aligned}$$

Noting that $\lim_{\delta \downarrow 0} (\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q} = \hat{Q}^-$

then

$$\lim_{\delta \downarrow 0} \hat{Q}(\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q} = \hat{Q}\hat{Q}^- = \hat{Q}^-\hat{Q}.$$

Recalling Corollary 2.1.1, it is seen that,

$$\lim_{\delta \downarrow 0} (\hat{Q}\hat{Q} + \delta^2 I)^{-1} Q_k \hat{Q}(\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q} = \lim_{\delta \downarrow 0} (\hat{Q}\hat{Q} + \delta^2 I)^{-1} Q_k.$$

Therefore,

$$\frac{\partial \hat{Q}^-}{\partial y_k} = -\lim_{\delta \downarrow 0} (\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q}Q_k(\hat{Q}\hat{Q} + \delta^2 I)^{-1}\hat{Q}$$

which, by the alternate definition of the generalized inverse is equivalent to

$$\frac{\partial \hat{Q}^-}{\partial y_k} = -\hat{Q}^- Q_k \hat{Q}^-.$$

Q.E.D.

CHAPTER 3

GENERALIZED INVERSE DUAL FOR CONVEX QUADRATICALLY CONSTRAINED QUADRATIC PROGRAMS

3.1 Introduction

The material of this chapter is an extension of duality theory developed from the classical Lagrangian analysis. In particular, the theory is developed from the duality of Wolfe [71].

The philosophy of the analysis parallels that of Falk [26]. Falk's results are for more general problems, but have limited usefulness in algorithmic application. His results for strictly convex problems are extended to convex quadratically constrained quadratic programs.

Other duals for quadratically constrained quadratic programs are those of Peterson and Ecker [51,52,53] and Rockafellar [57]. The Peterson-Ecker dual is established by use of a generalized geometric inequality and the Rockafellar dual is by conjugate function analysis.

The work of Peterson and Ecker and Rockafellar motivated a number of results in this chapter. It is shown that the smaller generalized inverse dual has the same desirable features as the conjugate dual and that the duals are derivable from each other. All proofs are new and require only linear algebra. Furthermore, the generalized inverse dual reveals a useful characterization of the constraint set not available with the conjugate dual form.

3.2 Duality

The following development establishes the generalized inverse dual and its properties for the class of convex quadratically constrained quadratic programs.

The primal problem is stated in the general form,

$$\{P.1\} \quad \underset{x}{\text{minimize}} \quad \Phi_o(x) = \frac{1}{2} x^t Q_o x + h_o^t x + c_o$$

$$\text{subject to:} \quad \Phi_j(x) = \frac{1}{2} x^t Q_j x + h_j^t x + c_j \leq 0$$

$$\text{for } j = 1,2,\ldots,m$$

where Q_j is an nxn real, symmetric, positive semidefinite matrix, $h_j \in E^n$, $c_j \in E^1$ for $j = 0,1,2,\ldots,m$.

Theorem 3.1 is the major result of this chapter and is based on the theory of the Moore-Penrose generalized inverse of a matrix.

<u>Theorem 3.1</u>. The Wolfe dual of {P.1} is equivalent to

$$\{D.1\} \quad \text{maximize} \quad \psi(y) = -\frac{1}{2} y^t H^t \hat{Q}^- Hy + \bar{c}^t y$$

$$\text{subject to:} \quad y = (y_o, y_1, y_2, \ldots, y_m)^t \geq 0$$

$$y_o = 1$$

$$\hat{Q}\hat{Q}^- Hy = Hy$$

where

$$H = (h_o, h_1, h_2, \ldots, h_m),$$

$$\hat{Q} = \hat{Q}(y) = \sum_{j=0}^{m} y_j Q_j,$$

$$\bar{c} = (c_o, c_1, c_2, \ldots, c_m)^t,$$

and $\hat{Q}^-$ is the generalized inverse of $\hat{Q}$.

<u>Proof of Theorem 3.1</u>. The Wolfe dual [71] of {P.1} is

$$\{W.1\} \quad \text{maximize} \quad \psi(x,y) = \frac{1}{2} x^t \hat{Q} x + y^t H^t x + \bar{c}^t y \tag{3.2.1}$$

$$\text{subject to:} \quad y = (y_o, y_1, \ldots, y_m)^t \geq 0$$

$$\hat{Q}x + Hy = 0 \tag{3.2.2}$$

$$y_o = 1$$

H, $\hat{Q}$, and $\bar{c}$ defined as above.

By Theorem A.7, (3.2.2) can be expressed as

$$x = -\hat{Q}^- Hy - (I-\hat{Q}^-\hat{Q})g \tag{3.2.3}$$

for $g \in E^n$. Equation (3.2.2) implies that for a given nonnegative y to be feasible Hy must lie in the column space of $\hat{Q}$. Hence,

$$\hat{Q}\hat{Q}^- Hy = Hy$$

as stated in Theorem A.6. Substitution for x, by (3.2.3) in (3.2.1) results in

$$\psi(y) = -\frac{1}{2} y^t H^t \hat{Q}^-(y) Hy + \bar{c}^t y.$$

Thus, {W.1} and {D.1} are equivalent.

Q.E.D.

The proof that $\psi(y)$ is concave on the feasible set will be deferred until section 3.4. This is done due to the proof being more direct when based on differentiability of $\psi(y)$.

As a result of the equivalence of {D.1} to {W.1} the major duality results of Wolfe can be applied to {D.1}. The weak duality theorem states that $\psi(y) \leqq \Phi_o(x)$ for y feasible to {D.1} and x feasible to {P.1} and by Wolfe's duality theorem $\psi(y^o) = \Phi_o(x^o)$ for y^o and x^o optimal. Also, the unbounded dual theorem and the no primal minimum theorem apply to {P.1} and {D.1}.

<u>Corollary 3.1.1</u>. If each of the Q_j of {P.1} is a diagonal matrix then the dual {D.1} is a fractional programming problem with linear constraints.

<u>Proof of Corollary 3.1.1</u>. Clearly $\hat{Q}$ is a diagonal matrix of rank r, then by a suitable transformation

$$\hat{Q} = (q_{ij}) = (q_{ij}(y))$$

where

q_{ii} = polynomial of degree one in y

$i = 1,2,\ldots,r$

$q_{ij} = 0$ for $i \neq j$

$q_{ii} = 0$ for $i = r+1, r+2, \ldots, n$.

Furthermore, reflecting the above transformation, let

$$Hy = (\xi_1, \xi_2, \ldots, \xi_n)^t$$

where

ξ_i, $i = 1,2,\ldots,n$, is a polynomial of degree one in y.

Using Theorem A.10 and simplifying, {D.1} becomes

{D.2} maximize $\psi(y) = -\frac{1}{2} \sum_{i=1}^{r} \frac{\xi_i^2}{q_{ii}} + \sum_{j=1}^{m} c_j y_j + c_o$

subject to: $y_j \geq 0$, $j = 1,2,\ldots,m$

$\xi_{r+k} = 0$, $k = 1,2,\ldots,n-r$.

Q.E.D.

A subclass of {P.1} for which the only dual constraint is nonnegativity is characterized by $\Phi_o(x)$ being strictly convex. This corresponds, by a well-known theorem, to Q_o being positive definite.

Corollary 3.1.2. Given {P.1} has a strictly convex objective function then {D.1} reduces to

{D.3} maximize $\psi(y) = -\frac{1}{2} y^t H^t \hat{Q}^{-1} Hy + \bar{c}^t y$

subject to: $y = (y_o, y_1, y_2, \ldots, y_m)^t \geq 0$

$y_o = 1$.

Proof of Corollary 3.1.2. Q_o is positive definite and therefore nonsingular. Hence, $\hat{Q}$ is positive definite and nonsingular. The generalized inverse of $\hat{Q}$ is then $\hat{Q}^{-1}$.

Q.E.D.

Note that if $y_j > 0$ for some Q_j positive definite, the same result holds. This corresponds to a strictly convex Φ_j being an active primal constraint so that, via complementary slackness, $y_j > 0$.

Corollary 3.1.3. {D.3} is a fractional programming problem.

Proof of Corollary 3.1.3. Each element of $\hat{Q}^{-1}$ is a polynomial in the y_j divided by a polynomial in y_j, that is, let

$$\hat{Q}^{-1} = (\bar{q}_{ij}) = (\bar{q}_{ij}(y))$$

where $\bar{q}_{ij}$ is determined by the co-factor of the ji^{th} element of $\hat{Q}$ and is a polynomial in the y_j of degree n-1 divided by the determinant of $\hat{Q}$ which is a polynomial in the y_j of degree n.

The vector Hy has elements which are polynomials in the y_j of degree 1. Therefore, $y^tH^t\hat{Q}^{-1}Hy$ will be a polynomial of degree n+1 divided by a polynomial of degree n.

Q.E.D.

{D.1} is a function of dual variables only; the number of dual variables being equal to the number of primal constraints. In this respect, it is a theoretical improvement over previously developed duals for the class {P.1}.

To expand on {D.1} and investigate its potential as a computational tool, the characterization of the constraint set must be simplified. The constraint set will be shown to be convex, but in general it is neither open nor closed. The approach is then to define its closure and relative interior.

It will be seen that there is a unique characterization of the closure and relative interior of the constraint set determined by a specially structured feasible dual vector. In fact, by just knowing the form of this vector the dual problem is significantly simplified in that the constraints reduce to linear equality and inequality forms.

The importance of the closure characterization will be demonstrated by showing that the dual problem solved on the closure of the constraint set results in the same solution as if solved over the original constraints.

3.3 Feasible Set F

The set of feasible points for {D.1} can be expressed by

$$F = \{y \in E_{+}^{m+1} \mid y_o = 1,\ \hat{Q}(y)\hat{Q}^{-}(y)Hy = Hy\}$$

where $y = (y_o, y_1, \ldots, y_m)^t$ and E_{+}^{m+1} is the nonnegative orthant of E^{m+1}.

Theorem 3.2. The feasible set F is convex.

Proof of Theorem 3.2. Let $y^1, y^2 \in F$ and $y^3 = \lambda y^1 + (1-\lambda)y^2$, $0 \leq \lambda \leq 1$. By y^1, y^2 feasible,

$$\hat{Q}(y^1)\hat{Q}^{-}(y^1)Hy^1 = Hy^1,$$

$$\hat{Q}(y^2)\hat{Q}^{-}(y^2)Hy^2 = Hy^2,$$

and $\lambda y^1 + (1-\lambda)y^2 = y^3 \geq 0$ with $y_o^3 = 1$. Furthermore, by Corollary 2.1.1, noting that if $y_j^1 > 0$ then $y_j^3 > 0$ and if $y_k^2 > 0$ then $y_k^3 > 0$,

$$[I-\hat{Q}^{-}(y^3)\hat{Q}(y^3)]\hat{Q}(y^1) = 0,$$

$$[I-\hat{Q}^{-}(y^3)\hat{Q}(y^3)]\hat{Q}(y^2) = 0.$$

Hence,

$$[I-\hat{Q}^{-}(y^3)\hat{Q}(y^3)]\{\lambda\hat{Q}(y^1)\hat{Q}^{-}(y^1)Hy^1 + (1-\lambda)\hat{Q}(y^2)\hat{Q}^{-}(y^2)Hy^2\} = 0$$

and $$[I-\hat{Q}^{-}(y^3)\hat{Q}(y^3)]Hy^3 = 0.$$

Thus, $$y^3 = \lambda y^1 + (1-\lambda)y^2 \in F.$$

Q.E.D.

A new notion will now be introduced which leads to a linear characterization of the closure of F, designated by F^*. The notion is that of an index maximal dual feasible vector. The existence of such a vector is assured for all dual feasible problems. The disadvantage is that it is not easily computed.

To introduce the concept of an index maximal dual feasible vector the index set of feasible dual vectors is defined.

Definition. The index set of feasible dual vectors is

$$N(y) = \{j \mid y_j > 0\}.$$

Note that for every $y \varepsilon F$, $0 \varepsilon N(y)$.

The size of $N(y)$ is defined as the number of elements it contains.

Definition. The size of the index set of feasible dual vectors is

$$S(y) = r$$

where r is the number of dual variables, including y_0, that are positive, e.g., given

$$N(y) = \{0, j_1, j_2, \ldots, j_{r-1}\}$$

then

$$S(y) = r.$$

Definition. If $y^* \varepsilon F$ and $S(y^*) \geq S(y)$ for all $y \varepsilon F$ then y^* is an index maximal dual feasible vector.

It is clear there will exist such a vector for every dual feasible problem.

In referring to such vectors, the shortened term "index maximal" will often be used, but where emphasis is required the full term will be employed.

Theorem 3.3. Given $y \in F$ then

$$(I-\hat{Q}^-(y)\hat{Q}(y))Q_j = Q_j(I-\hat{Q}^-(y)\hat{Q}(y)) = 0$$

for $j \in N(y)$.

Proof of Theorem 3.3. The proof is immediate from Corollary 2.1.1 on noting that $y_j > 0$ for $j \in N(y)$.

Q.E.D.

Corollary 3.3.1. Let $y, y' \in F$ such that $N(y) = N(y')$. Then $\hat{Q}(y)$ and $\hat{Q}(y')$ span the same column space and $\hat{Q}(y)\hat{Q}^-(y) = \hat{Q}(y')\hat{Q}^-(y')$.

Proof of Corollary 3.3.1. The proof follows from Corollary 2.2.2 by again noting that $y_j > 0$ for $j \in N(y)$ and Theorem 2.2.

Q.E.D.

Theorem 3.4. Given $y^* \in F$ is an index maximal dual feasible vector and $y \in F$, then

$$N(y) \subseteq N(y^*).$$

Proof of Theorem 3.4. If F consists of a single point the theorem is trivial; therefore, assume that F contains more than one point.

The proof will be by contradiction. Assume that $N(y) \nsubseteq N(y^*)$. y^* index maximal implies that $N(y^*) \not\subset N(y)$.

Since F is convex there exist $y^2 \in F$ such that

$$y^2 = \lambda y^* + (1-\lambda)y, \text{ for some } 0 < \lambda < 1. \text{ Hence,}$$

$$N(y^2) = N(y^*) \cup N(y)$$

$$= \{j \mid y_j^* > 0 \text{ or } y_j > 0\}.$$

This leads to a contradiction on y^* being index maximal, i.e., $N(y^*) \subset N(y^2)$ implying that $S(y^2) > S(y^*)$.

Q.E.D.

Theorem 3.5. Let

$$F^* = \{y \in E_+^{m+1} \mid y_o = 1, y^* \text{ index maximal, } [I-\hat{Q}^-(y^*)\hat{Q}(y^*)]Hy = 0\},$$

then $F \subseteq F^*$.

Proof of Theorem 3.5. Let y^* be index maximal and $y \in F$, then by Theorem 3.4, $N(y) \subseteq N(y^*)$. $y \in F$ implies $\hat{Q}(y)\hat{Q}^-(y)Hy = Hy$; multiplying by $[I-\hat{Q}^-(y^*)\hat{Q}(y^*)]$,

$$[I-\hat{Q}^-(y^*)\hat{Q}(y^*)]\hat{Q}(y)\hat{Q}^-(y)Hy = [I-\hat{Q}^-(y^*)\hat{Q}(y^*)]Hy.$$

By Theorem 3.3

$$[I-\hat{Q}^-(y^*)\hat{Q}(y^*)]\hat{Q}(y) = 0,$$

therefore,

$$[I-\hat{Q}^-(y^*)\hat{Q}(y^*)]Hy = 0$$

for all $y \in F$ and $F \subseteq F^*$.

Q.E.D.

F^* is closed and convex and if $\Phi_j(x)$ is linear for $j = 1,2,\ldots,m$, $F^* = F$.

A significant consequence of this theorem is that while the concept of an index maximal dual feasible vector implies dual feasibility it is only necessary to know $N(y^*)$. That is, by Corollary 3.3.1 any vector y such that $N(y) = N(y^*)$ will suffice for defining the matrix $(I-\hat{Q}^-(y^*)\hat{Q}(y^*))$. The most logical being that vector which has $y_j = 1$ for $j \in N(y^*)$ and $y_j = 0$ for $j \notin N(y^*)$. Therefore, the set $N(y^*)$ and not the particular y^* defines F^*. For an example of this idea let

$$Q_o = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}, \quad Q_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 5 \\ 0 & 0 & 5 & 5 \end{bmatrix} \quad \text{and}$$

$y^* = (1,3)^t$ resulting in

$$\hat{Q}(y^*) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 13 & 1 & 1 \\ 1 & 1 & 17 & 17 \\ 1 & 1 & 17 & 17 \end{bmatrix} \text{ and } \hat{Q}^-(y^*) = \frac{1}{192}\begin{bmatrix} 220 & -16 & -6 & -6 \\ -16 & 16 & 0 & 0 \\ -6 & 0 & 3 & 3 \\ -6 & 0 & 3 & 3 \end{bmatrix}.$$

Finally,

$$\hat{Q}(y^*)\hat{Q}^-(y^*) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Now let $y = (1,1)^t$, then

$$\hat{Q}(y) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 5 & 1 & 1 \\ 1 & 1 & 7 & 7 \\ 1 & 1 & 7 & 7 \end{bmatrix}, \quad \hat{Q}^-(y) = \frac{1}{24}\begin{bmatrix} 34 & -6 & -2 & -2 \\ -6 & 6 & 0 & 0 \\ -2 & 0 & 1 & 1 \\ -2 & 0 & 1 & 1 \end{bmatrix}$$

and $\hat{Q}(y)\hat{Q}^-(y) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$.

Thus it is seen that $\hat{Q}(y^*)\hat{Q}^-(y^*) = \hat{Q}(y)\hat{Q}^-(y)$ for $N(y^*) = N(y)$.

An example of $F \subset F^*$ is the following. Take the primal problem

{E.1} minimum $\Phi_o(x) = \frac{1}{2} x^t \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}^t x$

subject to: $\Phi_1(x) = \frac{1}{2} x^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}^t x - 5 \leq 0$

$\Phi_2(x) = \frac{1}{2} x^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}^t x - 4 \leq 0$

The primal solution is $x^o = (-1,2,2)$ and $\Phi_o(x^o) = -6.5$. For the dual problem

$$\hat{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & y_1 & 0 \\ 0 & 0 & y_2 \end{bmatrix}, \quad Hy = \begin{bmatrix} 1 \\ -1 + y_2 \\ -2 + y_1 \end{bmatrix}$$

and $\hat{Q}$ is seen to be nonsingular for $y = (1,1,1)^t$, an index maximal dual feasible vector. Hence,

$$[I-\hat{Q}^{-}(y)\hat{Q}(y)] = 0$$

and

$$F^* = \{y_o = 1,\ y_1 \geq 0,\ y_2 \geq 0\}.$$

F is seen to differ from F^* by noting that no $y = (1,0,y_2)^t$, $y_2 \neq 1$ or $y = (1,y_1,0)^t$, $y \neq 2$ is in F. See Figure 3-2(c).

$$F = \{y_o = 1,\ y_1 > 0,\ y_2 > 0\} \cup \{(1,0,1),(1,2,0)\}$$

and $F \subset F^*$, $F \neq F^*$.

Another example is as follows:

{E.2} minimum $\Phi_o(x) = \frac{1}{2} x^t \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} x + \begin{bmatrix} -2 \\ -2 \\ -2 \end{bmatrix}^t x$

subject to: $\Phi_1(x) = \frac{1}{2} x^t \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} x + \begin{bmatrix} -1 \\ +1 \\ 0 \end{bmatrix}^t x - 1 \leq 0$

$\Phi_2(x) = \frac{1}{2} x^t \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} x + \begin{bmatrix} 0 \\ +1 \\ -1 \end{bmatrix}^t x - 1 \leq 0$

The primal solution is $x^o = (1,1,1)^t$ and $\Phi_o(x^o) = -5$. The dual has

$$\hat{Q} = \begin{bmatrix} 1+y_1+y_2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1+y_1+y_2 \end{bmatrix}, \quad Hy = \begin{bmatrix} -2-y_1 \\ -2+y_1+y_2 \\ -2-y_2 \end{bmatrix},$$

$(I-\hat{Q}^{-}\hat{Q}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ for all $y_1,y_2 \geq 0$, and $F = F^*$. An index maximal dual feasible vector will have $y_1,y_2 > 0$ and it is seen that

$$(I-\hat{Q}^{-}\hat{Q})Hy = \begin{bmatrix} 0 \\ -2+y_1+y_2 \\ 0 \end{bmatrix}$$

or $F = F^* = \{y_o,y_1,y_2 \geq 0 \mid y_o = 1, y_1 + y_2 = 2\}$.

Before stating the next three corollaries it will be useful to state two definitions dealing with the topology of convex sets. These concepts and their properties are developed by Fenchel [27] and Rockafellar [58].

Definition. The affine hull of $C \in E^n$, aff C, is the intersection of the collection of affine sets M such that $C \subseteq M$.

Definition. The relative interior of a convex set C, ri C, consists of the points $x \in \text{aff } C$ for which there exists an $\varepsilon > 0$, such that $y \in C$ whenever $y \in \text{aff } C$ and $||x-y|| < \varepsilon$. In other words,

$$\text{ri } C = \{x \in \text{aff } C \mid \varepsilon > 0 \text{ and } (x + \varepsilon B) \cap \text{aff } C \subset C\}$$

where B is the unit ball in E^n, $B = \{x \mid ||x|| \leq 1\}$, and

$$x + \varepsilon B = \{x' \mid \; ||x-x'|| \leq \varepsilon\}.$$

It is evident that for a convex set C

$$\text{ri } C \subseteq C \subseteq \text{cl } C,$$

cl C is the closure of C.

The following corollary of Rockafellar [58, Corollary 6.3.1], relating the closure of two sets given a hypothesis on the relative interiors, is stated without proof as a lemma.

Lemma 3.1. Let C_1 and C_2 be convex sets in E^n. Then cl C_1 = cl C_2 if and only if ri C_1 = ri C_2.

The following is an important consequence of Theorem 3.5.

Lemma 3.2. aff F = aff F^*

Proof of Lemma 3.2. Clearly, if $F = F^*$ the lemma is trivial. Therefore, assume that $F \neq F^*$. Now $F \subset F^*$; therefore, aff $F \subseteq$ aff F^*. Let $y^* \in F \subset F^*$ be index maximal and $y \in F^*$ but $y \notin F$. F^* is convex hence, for $0 < \lambda' < 1$ there exist

$$y' = \lambda' y^* + (1-\lambda')y \in F^*.$$

From this,

$$N(y') \subseteq N(y^*) \subseteq N(y'),$$

thus,

$N(y') = N(y^*)$ by y^* being index maximal and

$\hat{Q}^-(y^*)\hat{Q}(y^*) = \hat{Q}^-(y')\hat{Q}(y')$ by Corollary 3.3.1.

Then $y' \varepsilon F$ and the line passing through y^*, y', y is in the affine hull of F, i.e.,

$$\text{aff } F^* \subseteq \text{aff } F,$$

which combined with $\text{aff } F \subseteq \text{aff } F^*$ results in $\text{aff } F = \text{aff } F^*$.

Q.E.D.

The next theorem and associated corollaries establish the importance of the concept of an index maximal dual feasible vector.

<u>Theorem 3.6</u>. $\text{ri } F = \text{ri } F^*$.

<u>Proof of Theorem 3.6</u>. By Lemma 3.2, $\text{aff } F = \text{aff } F^*$. If $F = F^*$ the result is trivial; hence, assume that $F \subset F^* = \text{cl } F^*$. Also, it is clear that if $y \varepsilon \text{ ri } F$ then $y \varepsilon \text{ ri } F^*$.

Now take $y \varepsilon \text{ ri } F^*$. There exist y', $y^* \varepsilon F^*$ such that y^* is index maximal and because $y \varepsilon \text{ ri } F^*$,

$$y = \lambda y^* + (1-\lambda)y'$$

for some $0 < \lambda < 1$. From this it is seen that y is also index maximal and hence in F, by application of Corollary 3.3.1.

Select $y'' \varepsilon (y + \varepsilon B) \cap \text{aff } F^*$, that is, select y'' in the nonempty ε-neighborhood of y such that $||y-y''|| < \delta = \frac{1}{2}\varepsilon$. Then there exist y''' in the ε-neighborhood such that

$$y'' = \lambda''y + (1-\lambda'')y'''$$

for some $0 < \lambda'' < 1$. Hence, y'' is an index maximal dual feasible vector in F. Thus, $(y + \frac{1}{2}\varepsilon B) \cap \text{aff } F \subset F$, implying $y \varepsilon \text{ ri } F$ and ri $\mathcal{F}$ = ri $\mathcal{F}^*$.

Q.E.D.

Corollary 3.6.1. cl($\mathcal{F}$) = $\mathcal{F}^*$.

Proof of Corollary 3.6.1. By definition $\mathcal{F}^*$ is closed. Using Theorem 3.6 and Lemma 3.1 the result is immediate.

Q.E.D.

Corollary 3.6.2. $y \in \mathcal{F}(\mathcal{F}^*)$ is an index maximal dual feasible vector if and only if $y \in$ ri $\mathcal{F}$(ri $\mathcal{F}^*$).

Proof of Corollary 3.6.2. Assume $\mathcal{F}$ consists of more than one point, otherwise the corollary is immediate.

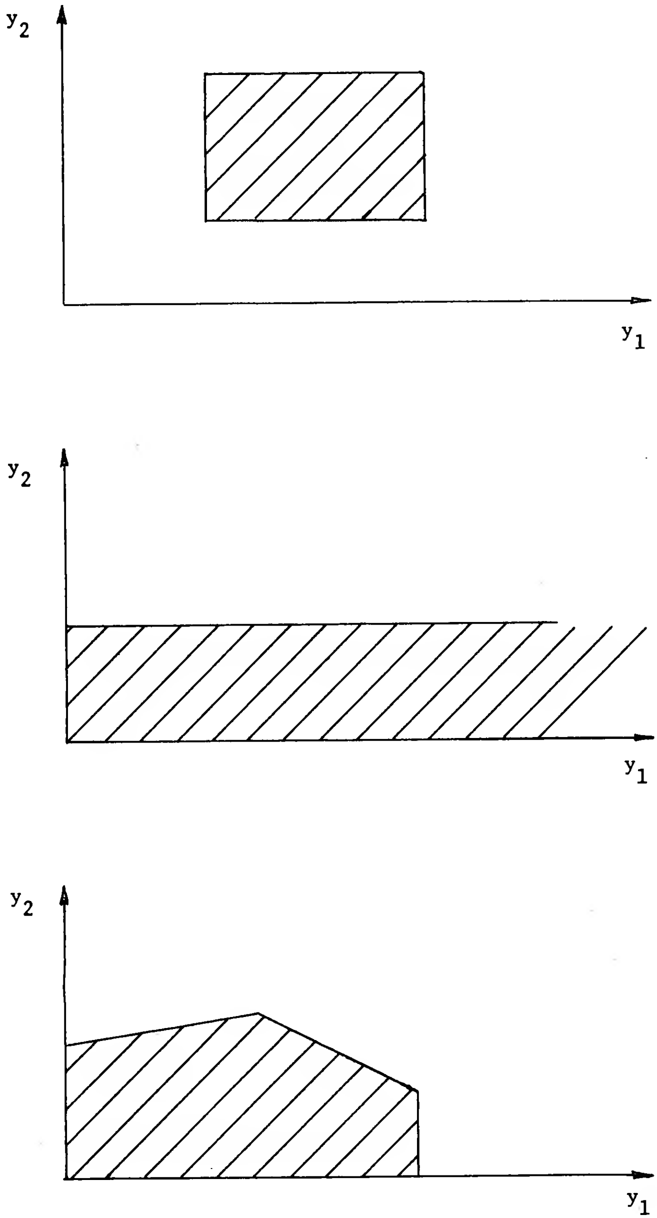
Take $y \in$ ri $\mathcal{F}$, then there exists points y^*, $y' \in \mathcal{F}^*$ such that y^* is index maximal and $y = \lambda y^* + (1-\lambda)y'$ for some $0 < \lambda < 1$. Hence, $N(y) = N(y^*)$ and y is index maximal.

Now take y index maximal in $\mathcal{F}^*$ and let $\xi = \min_{j \in N(y)} \{y_j\}$. Then any $y' \in (y + \frac{1}{2}\xi B) \cap \text{aff } \mathcal{F}^*$ is in $\mathcal{F}^*$. Therefore, $y \in$ ri $\mathcal{F}^*$(ri $\mathcal{F}$).

Q.E.D.

The implications of the preceding corollary are significant. First, there can be no dual of {P.1} which exhibits a closed $\mathcal{F}$ properly contained in the interior of E_+^{m+1}. Furthermore, relative boundary points of $\mathcal{F}$ are always on at least one of the coordinate hyperplanes of E^{m+1}, $\mathcal{F}$ is assumed to consist of more than one point.

Figure 3-1 consists of three convex sets, m = 2, which are not permissible as $\mathcal{F}$; while Figure 3-2 consists of three examples of $\mathcal{F}$, m = 2 again.



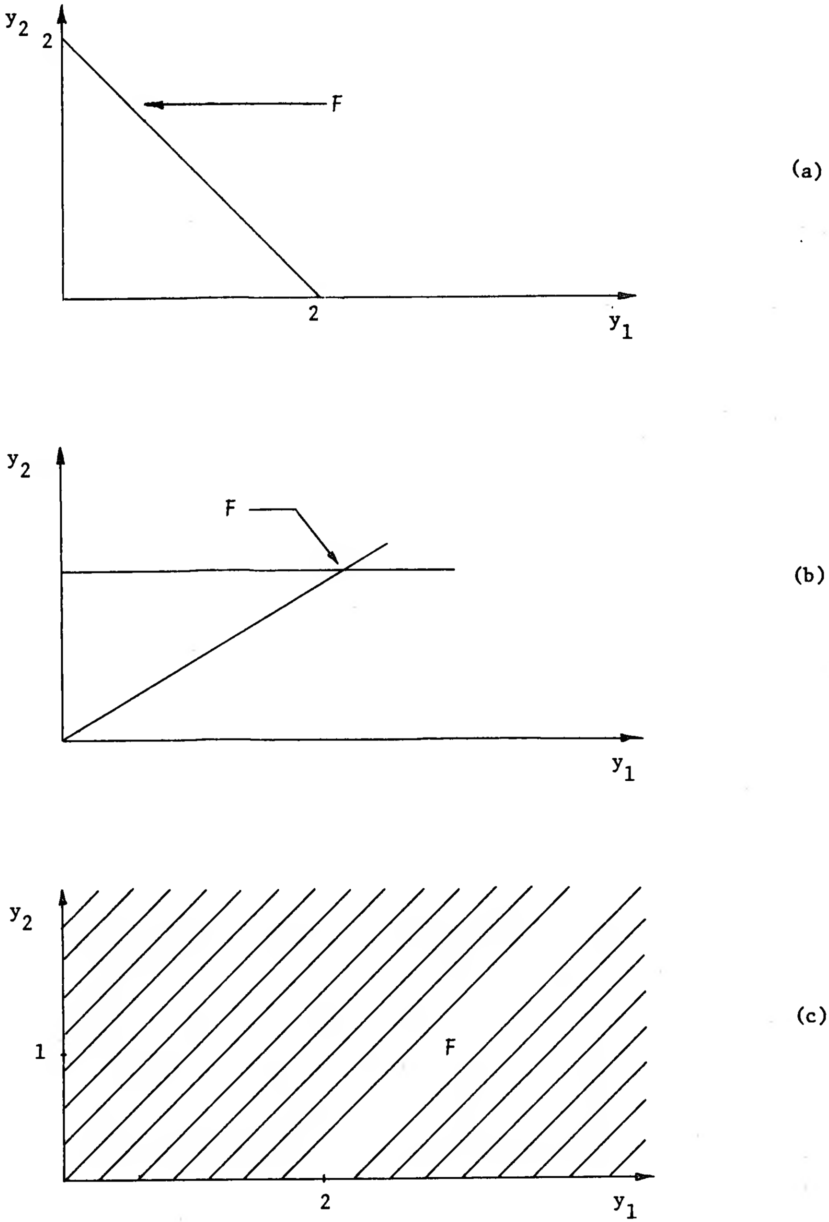
(a)

(b)

(c)

(y_0 = 1, not shown)

Figure 3-1



(y_0 = 1, not shown)

Figure 3-2

3.4 Objective Function $\psi(y)$

Having seen that F can be characterized by F^*, i.e., ri F = ri F^* and cl $F = F^*$, the next step is to investigate $\psi(y)$ on F^*. This is done by initially looking at the vector valued function $f(y)$ over F;

$$f(y) = \hat{Q}^-(y)Hy.$$

It will be shown that $\psi(y)$ is concave, continuous, and twice differentiable on F. Also, $\psi(y)$ approaches negative infinity as a relative boundary point of F, not in F, is approached.

The definition of $\psi(y)$ will be extended to F^* and $\psi(y)$ will be shown to be upper semicontinuous on F^*.

Lemma 3.3. For $y^* \in$ ri F^*, $y \in F^*$, and $0 < \lambda < 1$

$$\begin{aligned} f(\lambda) &= f(\lambda y^* + (1-\lambda)y) \\ &= [\lambda\hat{Q}(y^*) + (1-\lambda)\hat{Q}(y)]^- H(\lambda y^* + (1-\lambda)y) \end{aligned}$$

$$\begin{aligned} f(\lambda) = {} & (R^tR)^- Hy^* \\ & + (I-R^-\hat{B}(y^*))\hat{Q}^-(y)(I-R^-\hat{B}(y^*))^t Hy \\ & + \left(\frac{\lambda}{1-\lambda}\right)(I-R^-\hat{B}(y^*))\hat{Q}^-(y)(I-R^-\hat{B}(y^*))^t Hy^* \\ & - \left(\frac{\lambda}{1-\lambda}\right)(I-R^-\hat{B}(y^*))\hat{Q}^-(y)\hat{B}^t(y^*)GM(\lambda)G\hat{B}(y^*)\hat{Q}^-(y)(I-R^-\hat{B}(y^*))^t Hy^* \\ & - \left(\frac{\lambda}{1-\lambda}\right)^2 (I-R^-\hat{B}(y^*))\hat{Q}^-(y)\hat{B}^t(y^*)GM(\lambda)G\hat{B}(y^*)\hat{Q}^-(y)(I-R^-\hat{B}(y^*))^t Hy \\ & + \left(\frac{1-\lambda}{\lambda}\right)(R^tR)^- Hy \end{aligned}$$

where

$$R = \hat{B}(y^*)(I-\hat{Q}^-(y)\hat{Q}(y))$$

$$G = I-RR^-$$

$$M(\lambda) = \{I + \left(\frac{\lambda}{1-\lambda}\right)^2 G\hat{B}(y^*)\hat{Q}^-(y)\hat{B}^t(y^*)G\}^{-1}$$

Proof of Lemma 3.3.

$$[\lambda\hat{Q}(y^*) + (1-\lambda)\hat{Q}(y)]^- = \frac{1}{\lambda}[\hat{Q}(y^*) + \left(\frac{1-\lambda}{\lambda}\right)\hat{Q}(y)]^-.$$

Letting $U = \hat{B}^t(y)/\left(\frac{\lambda}{1-\lambda}\right)^{1/2}$, $V = \hat{B}^t(y^*)$ and applying Cline's result, Theorem A.14, the result is obtained.

Q.E.D.

The term $M(\lambda)$ can readily be shown to exhibit the property

$$\lim_{\lambda \downarrow 0} M(\lambda) = I + O(\lambda^2)$$

where a matrix valued function is $O(\lambda^2)$ if each element is $O(\lambda^2)$. The definition employed for $O(\lambda^2)$ is the following:

Definition. A scalar function $T(\cdot)$, of a real variable λ, is said to be $O(\lambda^n)$ as $\lambda \to 0$ if $T(\lambda)/\lambda^n$ is bounded as $\lambda \to 0$.

The next lemma establishes a useful relationship for some matrix products inherent in $f(\lambda)$.

Lemma 3.4. $R^{t-}\hat{Q}^-(y) = 0$ and $R^{t-}(I-\hat{Q}^-(y)\hat{Q}(y)) = R^{t-}$.

Proof of Lemma 3.4. Noting that

$$\hat{Q}^-(y)\hat{Q}(y)R^tR\hat{Q}(y) = \hat{Q}^-(y)\hat{Q}(y)(I-\hat{Q}^-(y)\hat{Q}(y))\hat{B}^t(y^*)R\hat{Q}(y) = 0,$$

$$R^tR\hat{Q}(y) = R^t\hat{B}(y^*)(I-\hat{Q}^-(y)\hat{Q}(y))\hat{Q}(y) = 0,$$

$$R^tR^{t-}\hat{Q}(y)\hat{Q}(y)R^t = R^tR^{t-}\hat{Q}(y)(I-\hat{Q}^-(y)\hat{Q}(y))\hat{B}^t(y^*) = 0,$$

$$\hat{Q}(y)\hat{Q}(y)R^t = \hat{Q}(y)\hat{Q}(y)(I-\hat{Q}^-(y)\hat{Q}(y))\hat{B}^t(y^*) = 0,$$

the hypothesis of Theorem A.16 is satisfied for $[\hat{Q}(y)R^t]^-$. Hence,

$$[\hat{Q}(y)R^t]^- = R^{t-}\hat{Q}^-(y)$$

but $\quad \hat{Q}(y)R^t = 0.$

$$R^{t-} = R^{t-}-R^{t-}\hat{Q}^-(y)\hat{Q}(y) = R^{t-}(I-\hat{Q}^-(y)\hat{Q}(y)).$$

Q.E.D.

There is a direct connection between the matrix R and the set of dual feasible vectors F. This is exhibited in the following.

Lemma 3.5. If $y^* \in \text{ri } F^*$, $y \in F^*$ and $R = \hat{B}(y^*)(I-\hat{Q}^-(y)\hat{Q}(y)) = 0$ then $y \in F$.

Proof of Lemma 3.5. By definition of F^*, $\hat{Q}(y^*)\hat{Q}^-(y^*)Hy = Hy$. Premultiplying by $(I-\hat{Q}^-(y)\hat{Q}(y))$ results in

$$(I-\hat{Q}^-(y)\hat{Q}(y))\hat{Q}(y^*)\hat{Q}^-(y^*)Hy = (I-\hat{Q}^-(y)\hat{Q}(y))Hy,$$

but by Theorem 2.3 $\hat{Q}(y^*)\hat{Q}^-(y^*) = \hat{B}^t(y^*)\hat{B}^{t-}(y^*)$, therefore by hypothesis $(I-\hat{Q}^-(y)\hat{Q}(y))Hy = 0$ and $y \in F$.

Q.E.D.

Theorem 3.7. If $y^* \in \text{ri } F^*$, $y \in F^*$ then

(a) $\lim\limits_{\lambda \downarrow 0} \psi(\lambda y^* + (1-\lambda)y) = -\infty$ if $y \notin F$,

(b) $\lim\limits_{\lambda \downarrow 0} \psi(\lambda y^* + (1-\lambda)y) = \psi(y)$ if $y \in F$,

hence $\psi(y)$ is continuous over F.

Proof of Theorem 3.7. First it is noted that

$$\psi(\lambda y^* + (1-\lambda)y) = -\frac{1}{2} f^t(\lambda)[\lambda\hat{Q}(y^*) + (1-\lambda)\hat{Q}(y)]f(\lambda)$$
$$+ \bar{c}^t(\lambda y^* + (1-\lambda)y)$$

and on substitution of $f(\lambda)$ from Lemma 3.3, it is seen that for:

(a) if $F = F^*$ the result is satisfied vacuously; otherwise one term of $\lambda f^t(\lambda)\hat{Q}(y^*)f(\lambda)$ is

$$\lambda\left(\frac{1-\lambda}{\lambda}\right)^2 y^tH^t(R^tR)^-\hat{Q}(y^*)(R^tR)^-Hy$$
$$= \lambda\left(\frac{1-\lambda}{\lambda}\right)^2 y^tH^tR^-R^{t-}(I-\hat{Q}^-(y)\hat{Q}(y))\hat{Q}(y^*)(I-\hat{Q}^-(y)\hat{Q}(y))R^-R^{t-}Hy$$
$$= \lambda\left(\frac{1-\lambda}{\lambda}\right)^2 y^tH^t(R^tR)^-Hy = \lambda\left(\frac{1-\lambda}{\lambda}\right)^2 [R^{t-}Hy]^t[R^{t-}Hy].$$

The first equivalence follows from Corollary A.3.2 and Lemma 3.4. By Lemma 3.5, $R \neq 0$ and $y \notin F$ implies that $R^{t-}Hy \neq 0$. Hence, the term is positive and as λ approaches zero, ψ approaches negative infinity, all other terms being finite.

(b) $y \in F$ implies $Hy = \hat{Q}^-(y)\hat{Q}(y)Hy$, Corollary A.3.2 implies $(R^tR)^- = R^-R^{t-}$ and using Lemma 3.4 $(R^tR)^-Hy = 0$. The only term remaining in the limit is

$$y^tH^t\hat{Q}^-(y)(I-R^-\hat{B}(y^*))^t\hat{Q}(y)(I-R^-\hat{B}(y^*))\hat{Q}^-(y)Hy$$

which on application of Lemma 3.4 reduces to $y^tH^t\hat{Q}^-(y)\hat{Q}(y)\hat{Q}^-(y)Hy$ and $\lim_{\lambda\downarrow 0} \psi(\lambda y^* + (1-\lambda)y) = \psi(y)$ for $y \in F$.

The continuity of $\psi(y)$ on F is immediate from part (b).

Q.E.D.

The directional derivative of $\psi(y)$ also exhibits a useful property as relative boundary points are approached.

Corollary 3.7.1. If $y^* \in \text{ri } F$, $y \in F^*$ then

$$\lim_{\lambda\downarrow 0} \frac{d}{d\lambda} \psi(\lambda y^* + (1-\lambda)y) = +\infty \text{ for } y \notin F.$$

Proof of Corollary 3.7.1.

$$\psi(\lambda y^* + (1-\lambda)y) = -\frac{1}{2} f^t(\lambda)[\lambda\hat{Q}(y^*) + (1-\lambda)\hat{Q}(y)]f(\lambda) + \bar{c}^t(\lambda y^* + (1-\lambda)y)$$

where $f(\lambda)$ is as developed in Lemma 3.3. From this, it can be shown that,

$$\lim_{\lambda\downarrow 0} \frac{d}{d\lambda} \psi(\lambda y^* + (1-\lambda)y) = K(y^*,y) + \lim_{\lambda\downarrow 0} \frac{1}{2}\left(\frac{1-\lambda}{\lambda}\right) y^tH^t(R^tR)^-Hy,$$

$y^tH^t(R^tR)^-Hy > 0$ by hypothesis with

$$\begin{aligned} K(y^*,y) = &-\frac{1}{2} y^{*t}H^t(R^tR)^-H(y^*-2y) \\ &-y^{*t}H^t(R^tR)^-\hat{Q}(y^*)D_1Hy \\ &-\frac{1}{2} y^tH^tD_1^t(\hat{Q}(y^*)-\hat{Q}(y))D_1Hy \\ &+(y-y^*)^tH^tD_1^t\hat{Q}(y^*)(R^tR)^-Hy \\ &+y^{*t}H^tD_2^t\hat{Q}(y^*)(R^tR)^-Hy \\ &-y^tH^tD_1^t\hat{Q}(y)D_1Hy^* \end{aligned}$$

$$+ y^t H^t D_1^t \hat{Q}(y) D_2 H y^*$$

$$+ \bar{c}^t (y^*-y)$$

where $$D_1 = (I-R^- \hat{B}(y^*))\hat{Q}^-(y)(I-\hat{B}^t(y^*)R^{t-})$$

$$D_2 = (I-R^- \hat{B}(y^*))\hat{Q}^-(y)\hat{Q}(y^*)D_1.$$

$K(y^*,y)$ is constant.

Thus it is seen that

$$\lim_{\lambda \downarrow 0} \frac{d}{d\lambda} \psi(\lambda y^* + (1-\lambda)y) = +\infty \text{ for } y \notin F.$$

Q.E.D.

Up to this point $\psi(y)$ has only been defined on F and not on F^*. Extending the domain of $\psi(y)$ to F^* can be accomplished by use of Theorem A.13. That is, for $y \in F^*$, $y \notin F$ accept the best approximate solution to the system $\hat{Q}(y)x + Hy = 0$, namely, $x = -\hat{Q}^-(y)Hy$. Then $\psi(y)$ has the same form as on F and is finite for such boundary points. Hereafter, $\psi(y)$ will be used to represent the function on both F and F^*.

In order to characterize $\psi(y)$ on F^* the definition for an upper semicontinuous function is stated for emphasis.

Definition. A real valued function f defined on a normed space E^n is said to be upper semicontinuous at x_o if, given $\varepsilon > 0$, there is a $\delta > 0$ such that $f(x)-f(x_o) < \varepsilon$ for $||x-x_o|| < \delta$.

From the continuity of $\psi(y)$ on F, the definition of $\psi(y)$ over F^*, and Theorem 3.7 the following is immediate.

Corollary 3.7.2. $\psi(y)$ is an upper semicontinuous function on F^*.

The form of the gradient and the Hessian of $\psi(y)$ will now be developed. To simplify the proof of the continuity of the gradient and the Hessian, the next lemma is developed.

Stewart [66] has shown that a necessary and sufficient condition for a generalized inverse to be continuous is that its rank be constant over its domain. This is an extension of the work of Ben-Israel [4] who first addressed the question.

The relative interior of F is a set of index maximal dual feasible vectors, Corollary 3.6.2, which by Corollary 3.3.1 imples that the rank of Q is constant over ri F. Therefore, applying Stewart's result the lemma is stated as;

Lemma 3.6. $f(y) = \hat{Q}^{-}(y)Hy$ is continuous over the relative interior of F.

Theorem 3.8. The gradient of $\psi(y)$, $y \in F$, is

$$\nabla\psi(y) = [\Phi_1(-\hat{Q}^{-}(y)Hy),\ \Phi_2(-\hat{Q}^{-}(y)Hy), \dots,\ \Phi_m(-\hat{Q}^{-}(y)Hy)]^t$$

and is continuous on ri F.

Proof of Theorem 3.8. The k^{th} component of $\nabla\psi(y)$ is

$$\frac{\partial\psi(y)}{\partial y_k} = - h_k^t\hat{Q}^{-} Hy + c_k - \frac{1}{2} y^t H^t \frac{\partial\hat{Q}^{-}}{\partial y_k} Hy$$

which by Theorem 2.4, for $k \in N(y)$, reduces to

$$\frac{\partial\psi(y)}{\partial y_k} = \Phi_k(-\hat{Q}^{-}(y)Hy).$$

For $k \notin N(y)$, modification of the proof of Theorem 2.4 gives

$$\frac{\partial\hat{Q}^{-}}{\partial y_k} = \lim_{\delta\downarrow 0} (\hat{Q}\hat{Q} + \delta^2 I)(I-\hat{Q}^{-}\hat{Q})-\hat{Q}^{-}\hat{Q}_k\hat{Q}^{-}$$

and

$$\frac{\partial\hat{Q}^{-}}{\partial y_k} Hy = \frac{\partial\hat{Q}^{-}}{\partial y_k} \hat{Q}^{-}\hat{Q}Hy = -\hat{Q}^{-}\hat{Q}_k\hat{Q}^{-} Hy.$$

Then

$\dfrac{\partial\psi(y)}{\partial y_k} = \Phi_k(-\hat{Q}^{-}(y)Hy)$ for $k \notin N(y)$ and the form of the gradient is defined.

The continuity of $\nabla\psi(y)$ follows from Lemma 3.6, noting that

$$\frac{\partial\psi(y)}{\partial y_k} = \Phi_k(-f(y)).$$

Q.E.D.

This is an extension of the result of Falk [26] where it is seen that $\nabla\psi(y)$ is defined at a particular point, i.e., $x = -\hat{Q}^-(y)Hy$, instead of any $x = -\hat{Q}^-(y)Hy-(I-\hat{Q}^-(y)\hat{Q}(y))g$.

<u>Corollary 3.8.1</u>. The jk^{th} element of the Hessian of $\psi(y)$, for $y \in \text{ri } F$, is

$$\hat{H}_{jk} = \hat{H}_{jk}(y) = -y^tH^t\hat{Q}^-Q_j\hat{Q}^-Q_k\hat{Q}^-Hy + h_j^t\hat{Q}^-Q_k\hat{Q}^-Hy + h_k^t\hat{Q}^-Q_j\hat{Q}^-Hy - h_j^t\hat{Q}^-h_k.$$

The Hessian exist for each $y \in \text{ri } F$ and is continuous on ri F.

<u>Proof of Corollary 3.8.1</u>. The proof follows from

$$H_{jk} = \frac{\partial^2}{\partial y_j \partial y_k}\ \psi(y),$$

Theorem 2.5 and Theorem 3.8. The continuity of the Hessian follows from the generalized inverse being continuous on ri F and Lemma 3.6.

Q.E.D.

The next corollary will be used in the proof of the concavity of $\psi(y)$ on F.

<u>Corollary 3.8.2</u>.

$$\nabla\psi(y)y = \psi(y) - \Phi_o(-\hat{Q}^-Hy)$$

<u>Proof of Corollary 3.8.2</u>.

$$\begin{aligned}\nabla\psi(y)y &= \sum_{j=1}^{m}[\tfrac{1}{2}\, y^tH^t\hat{Q}^-Q_j\hat{Q}^-Hy - h_j^t\hat{Q}^-Hy + c_j]y_j \\ &= \tfrac{1}{2}\, y^tH^t\hat{Q}^-Hy - \tfrac{1}{2}\, y^tH^t\hat{Q}^-Q_o\hat{Q}^-Hy - y^tH^t\hat{Q}^-Hy \\ &\quad + h_o^t\hat{Q}^-Hy + \bar{c}^ty - c_o \\ &= \psi(y) - \Phi_o(-\hat{Q}^-Hy)\end{aligned}$$

Q.E.D.

$\psi(y)$ can now be proved to be concave based on the following result found in Mangasarian [43],

$\psi(y)$ is concave on F if and only if

$\psi(y^2) - \psi(y^1) \leq \nabla\psi(y^1)(y^2-y^1)$ for each $y^1, y^2 \in F$.

Theorem 3.9. $\psi(y)$ is concave on F.

Proof of Theorem 3.9. Take $y^1, y^2 \in F$, then $\hat{Q}(y^2) = \sum_{j \in N(y^2)} y_j^2 Q_j$ is positive semidefinite and

$$0 \leq \frac{1}{2}\,[y^{2t}H^t\hat{Q}^-(y^2) - y^{1t}H^t\hat{Q}^-(y^1)]\hat{Q}(y^2)[\hat{Q}^-(y^2)Hy^2 - \hat{Q}^-(y^1)Hy^1]$$

$$0 \leq \frac{1}{2}\,y^{2t}H^t\hat{Q}^-(y^2)Hy^2 + \frac{1}{2}\,y^{1t}H^t\hat{Q}^-(y^1)\hat{Q}(y^2)\hat{Q}^-(y^1)Hy^1$$

$$- \frac{1}{2}\,y^{2t}H^t\hat{Q}^-(y^2)\hat{Q}(y^2)\hat{Q}^-(y^1)Hy^1 - \frac{1}{2}\,y^{1t}H^t\hat{Q}^-(y^1)\hat{Q}(y^2)\hat{Q}^-(y^2)Hy^2$$

Now $y^2 \in F$ implies that $\hat{Q}(y^2)\hat{Q}^-(y^2)Hy^2 = Hy^2$. This, and adding $\bar{c}^t y^2$ to the inequality;

$$\psi(y^2) \leq \frac{1}{2}\,y^{1t}H^t\hat{Q}^-(y^1)\hat{Q}(y^2)\hat{Q}^-(y^1)Hy^1 - y^{2t}H^t\hat{Q}^-(y^1)Hy^1 + \bar{c}^t y^2$$

$$\psi(y^2) \leq \sum_{j=0}^{m} [\frac{1}{2}\,y^{1t}H^t\hat{Q}^-(y^1)Q_j\hat{Q}^-(y^1)Hy^1 - h_j^t\hat{Q}^-(y^1)Hy^1 + c_j]y_j^2$$

But this is recognized by Corollary 3.8.2 to be

$$\psi(y^2) \leq \nabla\psi(y^1)y^2 + \Phi_o(-\hat{Q}^-(y^1)Hy^1)$$

adding $-\psi(y^1)$ and using Corollary 3.8.2 again, the inequality reduces to

$\psi(y^2) - \psi(y^1) \leq \nabla\psi(y^1)(y^2-y^1)$ and $\psi(y)$ is concave on F.

Q.E.D.

3.5 Optimal Primal Solutions

Wolfe duality theory applied to {D.1} insures that the solution of the dual problem characterizes the primal solution of {P.1}. This is certainly true in the sense that the optimal objective function value of {P.1} is determined from the optimal dual solution, yet all optimal primal variables are not explicitly defined. This is due to the primal solution being defined in two parts, explicitly on the range space and implicitly on the null space of $\hat{Q}$. Only that part defined explicitly on the range space of $\hat{Q}$ is immediately determined from the optimal dual solution. This should not be construed to imply that an optimal dual solution does not return the optimal primal vector; it does.

Assume {D.1} has been solved and the optimal dual vector is y^o. Then $y^o \in F$ and the optimal primal vector is given by

$$x^o = -\hat{Q}^-(y^o)Hy^o - (I-\hat{Q}^-(y^o)\hat{Q}(y^o))g$$

for $g \in E^n$. The optimal primal vector will be unique only if $\hat{Q}(y^o)$ has rank n, i.e., $I = \hat{Q}^-(y^o)\hat{Q}(y^o)$. Therefore, to establish x^o, g must be determined.

Substituting x^o into {P.1} results in the following;

$$\{P^o.1\} \quad \text{minimize} \quad \Phi_o(g) = -h_o^t(I-\hat{Q}^-\hat{Q})g + \Phi_o(-\hat{Q}^-Hy^o)$$

$$\text{subject to:} \quad \Phi_j(g) = -h_j^t(I-\hat{Q}^-\hat{Q})g + \Phi_j(-\hat{Q}^-Hy^o) = 0$$

$$\text{for } j \in N(y^o)$$

$$\Phi_j(g) = \frac{1}{2}\, g^t(I-\hat{Q}^-\hat{Q})Q_j(I-\hat{Q}^-\hat{Q})g$$

$$-(h_j^t - y^{ot}H^t\hat{Q}^-Q_j)(I-\hat{Q}^-\hat{Q})g$$

$$+ \Phi_j(-\hat{Q}^-Hy^o) \leq 0$$

$$\text{for } j \notin N(y^o)$$

where $\hat{Q}$ is understood to be a function of y^o and the order of the constraints has been arranged so that the first $S(y^o)-1$ constraints are active, that is, $N(y^o) - \{0\} = \{1,2,\ldots,S(y^o)-1\}$.

By complementary slackness, $\Phi_j(g)$ for $j \in N(y^o) - \{0\}$ is zero. This results in a consistent set of linear equalities.

Let $H_o = (h_1, h_2, \ldots, h_{S(y^o)-1})$, an $n\times(S(y^o)-1)$ matrix, and Φ^o be the $S(y^o)-1$ vector

$$\Phi^o = [\Phi_1(-\hat{Q}^- Hy^o),\ \Phi_2(-\hat{Q}^- Hy^o), \ldots, \Phi_{S(y^o)-1}(-\hat{Q}^- Hy^o)]^t.$$

The first $S(y^o)-1$ constraints of $\{P^o.1\}$ are expressable as

$$H_o^t(I-\hat{Q}^-(y^o)\hat{Q}(y^o))g = \Phi^o.$$

Solving for $(I-\hat{Q}^-\hat{Q})g$ results in

$$(I-\hat{Q}^-\hat{Q})g = H_o^{t-}\ \Phi^o + (I-H_o^{t-}H_o^t)g^r \tag{3.5.1}$$

for $g^r \varepsilon E^n$.

<u>Theorem 3.10</u>. $\Phi_o(g)$ is a constant for y^o. That is, the primal objective function value is fully determined by the optimal dual vector.

<u>Proof of Theorem 3.10</u>.

$\Phi_o(g) = -h_o^t(I-\hat{Q}^-\hat{Q})g + \Phi_o(-\hat{Q}^- Hy^o)$ and by the constraint set of $\{D.1\}$

$$-h_o^t(I-\hat{Q}^-\hat{Q}) = [y_1^o h_1, y_2^o h_2, \ldots, y_m^o h_m]^t(I-\hat{Q}^-\hat{Q})$$

which can be written as

$$-h_o^t(I-\hat{Q}^-\hat{Q}) = \bar{y}^{ot}H_o^t(I-\hat{Q}^-\hat{Q})$$

where $\bar{y}^o = (y_1^o, y_2^o, \ldots, y_{S(y^o)-1}^o)^t$ by noting $y_k^o = 0$ for $k \geq S(y^o)$.

Now,

$$\Phi_o(g) = \bar{y}^{ot}H_o^t(I-\hat{Q}^-\hat{Q})g + \Phi_o(-\hat{Q}^- Hy^o)$$

and on substitution for $(I-\hat{Q}^-\hat{Q})g$ by (3.7.1)

$$\Phi_o(x^o) = \Phi_o(g) = \bar{y}^{ot}H_o^t H_o^{t-}\Phi^o + \Phi_o(-\hat{Q}^-(y^o)Hy^o).$$

Q.E.D.

$\{P^o.1\}$ is now characterized as the minimization of a constant subject to $m-S(y^o)+1$ quadratic constraints. The following theorem defines those instances when it is unnecessary to solve $\{P^o.1\}$.

<u>Theorem 3.11</u>. The optimal primal solution is fully determined by the optimal dual solution provided the $n\times(S(y^o)-1)$ matrix H_o has rank n

and the optimal primal solution is

$$x^o = -\hat{Q}^-(y^o)Hy^o-(H_oH_o^t)^{-1}H_o\phi^o.$$

Proof of Theorem 3.11. If H_o has rank n, then by Theorem A.6, $H_o^{t-} = (H_oH_o^t)^-H_o$ and $(I-H_o^{t-}H_o^t) = (I-(H_oH_o^t)^-H_oH_o^t) = 0$. Therefore, $(I-\hat{Q}^-\hat{Q})g = H_o^{t-}\phi^o = (H_oH_o^t)^-H_o\phi^o$ and $x^o = -\hat{Q}^-(y^o)Hy^o-(H_oH_o^t)^-H_o\phi^o$ for $H_o = (h_1,h_2,\ldots,h_{S(y^o)-1})$, $(H_oH_o^t)^- = (H_oH_o^t)^{-1}$.

Q.E.D.

When H_o does not have rank n, $\{P^o.1\}$ is a special form of $\{P.1\}$ which can readily be solved. Substitution by (3.5.1) in $\{P^o.1\}$, noting that it is only required to determine a g^r which satisfies the $m-S(y^o)+1$ constraints, and solving the following strictly convex auxiliary primal problem defines the component of x^o which is in the null space of $\hat{Q}(y^o)$.

$\{P^o.2\}$ minimize $\frac{1}{2}\, g^{rt}Ig^r$

subject to: $\phi_j(g^r) \leq 0$, for $j \notin N(y^o)$

where

$$\begin{aligned}\phi_j(g^r) = {} & \frac{1}{2}\, g^{rt}(I-H_oH_o^-)Q_j(I-H_oH_o^-)g^r \\ & + [\phi^oH_o^{t-}Q_j-h_j^t + y^{ot}H^t\hat{Q}^-Q_j](I-H_oH_o^-)g^r \\ & + [\tfrac{1}{2}\, \phi^{ot}H_o^{t-}Q_j-h_j^t + y^{ot}H^t\hat{Q}^-Q_j]H_o^{t-}\phi^o \\ & + \phi_j(-\hat{Q}^-(y^o)Hy^o)\end{aligned}$$

Solving $\{P^o.2\}$ by the dual $\{D.1\}$ will return a unique g^r, $\hat{Q}$ will be of rank n. Therefore, at most, two dual problems will require solution and at least one will be a strictly convex problem permitting the use of Corollary 3.1.2.

3.6 Linear Constraints

The class of convex quadratically constrained quadratic programs {P.1} can be expanded to admit problems with linear constraints.

{P.1'} minimize $\Phi_o(x) = \frac{1}{2} x^t Q_o x + h_o^t x + c_o$

subject to: $\Phi_j(x) = \frac{1}{2} x^t Q_j x + h_j^t x + c_j \leq 0$

for $j = 1,2,\ldots,m_1$

$a_k^t x + c_k \leq 0$

for $k = m_1 + 1, m_1 + 2,\ldots,m$

where $m - m_1 = m_2$

By treating linear inequality constraints as quadratic constraints, with zero quadratic terms, {P.1'} is merely a special form of {P.1} and the theory of Section 3.2 applies.

The more interesting case is when {P.1'} has linear equality constraints. Let the set of linear equality constraints be given by

$$Ax = c \tag{3.6.1}$$

where A is an $m_2$xn coefficient matrix and c is an m_2x1 constant vector.

Given that the primal problem is feasible there exist solutions to (3.6.1). Using the generalized inverse and the assumption of feasibility,

$$x = A^- c + (I - A^- A)g \tag{3.6.2}$$

for all $g \in E^n$. Clearly any solution to {P.1'} is of the form (3.6.2) and the primal problem can be transformed to one in g.

Substituting (3.6.2) into {P.1'} results in the following

{P.2'} minimize $\Phi_o'(g) = \frac{1}{2} g^t (I-A^-A) Q_o (I-A^-A) g + (h_o^t + c^t A^{-t})(I-A^-A)g$

$+ (c_o + \frac{1}{2} c^t A^{-t} Q_o A^- c + h_o^t A^- c)$

$$\text{subject to: } \Phi_j'(g) = \frac{1}{2} g^t(I-A^-A)Q_j(I-A^-A)g + (h_j^t + c^tA^{-t})(I-A^-A)g$$

$$+ (c_j + \frac{1}{2} c^tA^{-t}Q_jA^-c + h_j^tA^-c) \leq 0$$

for $j = 1,2,\ldots,m$

This transforms {P.1'} with linear equality constraints to the form {P.1}.

The advantage of the conversion is that the dual problem {D.1} is reduced from a problem in m dual variables to one in only m_1 dual variables.

It is possible to express a set of linear inequality constraints as a set of equality constraints by the use of slack variables. This would be highly desirable based on the potential reduction of the size of the dual problem. Unfortunately, the use of slack variables will also increase the size of the set of nonnegativity constraints and the dual problem size remains effectively unchanged.

In conclusion, it can be said that linear constraints cause no difficulty in applying the dual {D.1}. While linear equality constraints permit the formation of an equivalent reduced problem, linear inequality constraints are treated as special quadratic form constraints.

3.7 Linear and Quadratic Programming

Linear and convex quadratic programs are proper subsets of the class {P.1}. The dual {D.1}, therefore, should admit of special forms for these two subclasses.

For linear programs, {D.1} results in the unsymmetric dual. This is seen by noting that $\hat{Q} = 0$ and the constraint set F is given by

$$F = \{y_j \geq 0,\ j = 1,2,\ldots,m \mid \sum_{j=1}^{m} h_jy_j = -h_o\},$$

the objective function is

$$\psi(y) = c_o + \sum_{j=1}^{m} c_j y_j .$$

The application of {D.1} to convex quadratic programs results in a nonsymmetric dual which is again a quadratic program.

Define the general convex quadratic program as

$$\{Q.1\} \qquad \text{minimize} \qquad \Phi_o(x) = \frac{1}{2} x^t Q_o x + h_o^t x + c_o$$

$$\text{subject to:} \quad \Phi_j(x) = h_j^t x + c_j \leq 0$$

$$\text{for } j = 1,2,\ldots,m$$

where Q_o is symmetric and positive semidefinite. {D.1} reduces to

$$\{D.4\} \qquad \text{maximize} \qquad \psi(y) = -\frac{1}{2} y^t \bar{H}^t Q_o^- \bar{H} y$$

$$+ (\bar{c}^t - h_o^t Q_o^- \bar{H}) y$$

$$+ (-\frac{1}{2} h_o^t Q_o^- h_o + c_o)$$

$$\text{subject to:} \quad y = (y_1, y_2, \ldots, y_m)^t \geq 0$$

$$(I - Q_o^- Q_o)\bar{H} y = -(I - Q_o^- Q_o) h_o$$

where $\bar{H} = (h_1, h_2, \ldots, h_m)$ and $\bar{c} = (c_1, c_2, \ldots, c_m)^t$.

{D.4} reflects a notational variation from {D.1} in that $\hat{Q} = Q_o$ and is not a function of the y_j. This permits the implicit representation of $y_o = 1$ in {D.4}. It is clear that $\psi(y)$ is quadratic and that the constraint set is linear. For Q_o of rank r, there are n-r linear constraints in {D.4}. This is seen by noting that there exists an orthogonal matrix P such that

$$P^t Q_o P = D_o,$$

D_o is a diagonal matrix which has r nonzero diagonal elements equal to the eigenvalues of Q_o. The constraint set is thereby reduced to

$$(I-D_o^- D_o)P^t Hy = -(I-D_o^- D_o)P^t h_o.$$

$(I-D_o^- D_o)$ is a diagonal matrix with r zero and n-r nonzero diagonal elements. Hence, there will be n-r linear equality constraints for {D.4}.

If Q_o is nonsingular, {D.4} has only nonnegativity constraints and is the dual form addressed by Lemke [41].

3.8 Equivalence of Dual Forms

In sections 3.2 through 3.6 the generalized inverse dual {D.1} and its properties were developed for convex quadratically constrained quadratic programs. {D.1} was developed from the Wolfe dual, {W.1}, of {P.1}. Three of the significant advantages of {D.1} over {W.1} are (1) {D.1} is an m variable dual while {W.1} has m+n variables, (2) the objective function of {D.1} is concave whereas that of {W.1} is not, and (3) the constraint set of {D.1} is also characterized by a linear system.

There are two other dual forms of {P.1}. One is the generalized geometric inequality dual of Peterson and Ecker [51,52,53] and the conjugate function dual of Rockafellar [57].

The dual of Peterson and Ecker is applicable to all convex ℓp-constrained ℓp-programs and {P.1} in particular while Rockafellar's dual is applicable to the class of faithfully convex programs. Faithful convexity being defined as:

Definition. A function f is faithfully convex if it is not affine (simultaneously convex and concave) along any line segment, unless f is affine along the entire line extending the line segment.

Because Rockafellar [57] has shown that his dual and that of Peterson and Ecker are equivalent for {P.1} the designation {CD.1} will be employed to refer to both.

The following form of {CD.1} will be used.

{CD.1} maximize $T(y,z) = -\frac{1}{2} \sum_{j\varepsilon N(y)} \frac{1}{y_j} z_j^t z_j + \bar{c}^t y$

subject to: $y_j \geq 0,\ j = 1,2,\ldots,m$

$$y_o = 1$$

$$\sum_{j=0}^{m} B_j^t z_j = \sum_{j=0}^{m} h_j y_j$$

$$z_j = 0 \text{ if } y_j = 0$$

$$z_j \ \varepsilon\ E^n,\ y_j \ \varepsilon\ E^1$$

{CD.1} is stated explicitly in terms of the dual variables y_j, z_j for $j = 0,1,2,\ldots,m$. For notational convenience the number of dual variables is exhibited as $(m+1)(n+1)$ but in applications the number of dual variables will be $\sum_{j=0}^{m} \rho_j + m$ where ρ_j is the rank of Q_j.

Both Peterson and Ecker and Rockafellar have developed properties of {CD.1} comparable to those developed for {D.1}. Three differences between {CD.1} and {D.1} are seen to be the number of variables, $\sum_{j=0}^{m} \rho_j + m$ compared to m, the form of the constraint set and the form of the objective function.

To compare the duals {D.1} and {CD.1} it is first noted that there is an alternate characterization of the system

$$\hat{Q}(y)\hat{Q}^-(y)Hy = Hy$$

resulting in a consistent system of equations in an expanded space.

<u>Theorem 3.12</u>. Given that $\hat{Q}x = -Hy$ is consistent, i.e., $\hat{Q}\hat{Q}^{-}Hy = Hy$, then there exists a vector of m+1 n-vectors

$$z = (z_o^t, z_1^t, z_2^t, \ldots, z_m^t)^t;$$

$$z_j = z_j' \sqrt{y_j}\ , \ z_j' \in E^n,$$

such that for $Q_j = B_j^t B_j$, $y_o = 1$, $y_j \geq 0$, $j = 0,1,2,\ldots,m$

$$\sum_{j=0}^{m} B_j^t z_j = - \sum_{j=0}^{m} h_j y_j,$$

is consistent. The converse holds if $z_j = 0$ when $y_j = 0$.

<u>Proof of Theorem 3.12</u>. By Lemma 2.3, $\hat{Q}$ can be factored into

$$\hat{Q} = \hat{B}^t\hat{B}$$
$$= \bar{B}^t YY\bar{B}$$

where $\bar{B}^t = (B_o^t, B_1^t, B_2^t, \ldots, B_m^t)$ for $Q_j = B_j^t B_j$, $j = 0,1,2,\ldots,m$,

B_j of order nxn for all j,

and

$$Y = I \times \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & \sqrt{y_1} & & & \vdots \\ \vdots & & \sqrt{y_2} & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & & 0 & \sqrt{y_m} \end{bmatrix} \quad \text{of order } n(m+1)\mathrm{x}n(m+1),$$

a Kronecker product.

Using this factorization and Theorem 2.3, $\hat{Q}\hat{Q}^{-}Hy = Hy$ reduces to

$$\hat{B}^t\hat{B}^{t-}Hy = Hy.$$

Hence, there exists a vector $z' = (z_o'^t, z_1'^t, \ldots, z_m'^t)^t$ such that

$$\hat{B}^t z' = \pm Hy$$

or

$$\bar{B}^t Y z' = \pm Hy.$$

By defining the m+1 n-vector $z = (z_o^t, z_1^t, \ldots, z_m^t)^t$,

$$z_j = z'_j \sqrt{y_j}$$

it is seen that

$$\sum_{j=0}^{m} B_j^t z_j = \pm \sum_{j=0}^{m} h_j y_j .$$

The converse is obtained by reversing the above sequence of arguments with $z_j = 0$ when $y_j = 0$.

Q.E.D.

The expression for the feasible set F in the expanded dual space is now

$$F = \{y \in E_+^{m+1} | y_o = 1 \text{ and } \quad z_j \in E^n, \; j = 0,1,\ldots,m \\ \text{solving } \sum_{j=0}^{m} B_j^t z_j = \sum_{j=0}^{m} h_j y_j \text{ and } y_j = 0 \text{ implies} \\ z_j = 0\}.$$

<u>Theorem 3.13</u>. There exists a one to one correspondence between the feasible points of {D.1} and {CD.1}. Furthermore, for such points the associated objective function values are equal.

<u>Proof of Theorem 3.13</u>. The proof will be by demonstrating that {CD.1} can be derived from {D.1} and vice versa.

From {D.1},

$$\psi(y) = -\frac{1}{2} y^t H^t \hat{Q}^-(y)\hat{Q}(y)\hat{Q}^-(y)Hy + \bar{c}^t y$$

which by the factorization of $\hat{Q}(y) = \hat{B}^t(y)\hat{B}(y)$ is equivalent to

$$\psi(y) = -\frac{1}{2} [\hat{B}^{t-}(y)Hy]^t[\hat{B}^{t-}(y)Hy] + \bar{c}^t y$$

where $\hat{B}(y) = \bar{B}Y$ as defined for Theorem 3.12.

Define

$$z = Y\hat{B}^{t-}(y)Hy = Y\hat{B}\hat{Q}^- Hy, \tag{3.8.1}$$

z is a vector of $m+1$ n-vectors and the j^{th} n-vector is zero if y_j is zero.

Multiply (3.8.1) by Y^-,

$$Y^-z = Y^-Y\hat{B}\hat{Q}^-Hy = (Y^-Y)Y\hat{B}\hat{Q}^-Hy = Y\bar{B}\hat{Q}^-Hy$$

or

$$Y^-z = \hat{B}(y)\hat{Q}^-(y)Hy = \hat{B}^{t-}(y)Hy. \qquad (3.8.2)$$

Substituting (3.8.2) into $\psi(y)$ results in

$$T(y,z) = -\frac{1}{2}(Y^-z)^t(Y^-z) + \bar{c}^ty. \qquad (3.8.3)$$

Multiply (3.8.2) by $\hat{B}^t(y)$,

$$\hat{B}^t(y)Y^-z = \bar{B}^tYY^-z = \hat{B}^t(y)\hat{B}^{t-}(y)Hy = Hy$$

by the constraints of {D.1} and Theorem 3.12.

By Theorem A.11, YY^- is a diagonal matrix with $y_j > 0$ implying a diagonal element of 1 and $y_j = 0$ implying 0, resulting in $YY^-z = z$ by definition. Then by Theorem 3.12 the constraints of {D.1} can be stated as:

$$\bar{B}^tz = Hy$$

$$z_j = 0 \text{ if } y_j = 0$$

$$y_j \geq 0 \text{ for } j = 1,2,\ldots,m \text{ and } y_o = 1$$

and the objective function by (3.8.3), hence {D.1} implies {CD.1}.

Now taking {CD.1}; it is seen by Theorem 3.12 that the constraint set of {CD.1} is equivalent to that of {D.1}. Furthermore,

$$\bar{B}^tz = \bar{B}^tYY^-z = \hat{B}^tY^-z = Hy$$

by definition of Y and z for {CD.1}. This equation is consistent; therefore,

$$Y^-z = \hat{B}^{t-}Hy + (I-\hat{B}^{t-}\hat{B}^t)g$$

for $g \in E^{n(m+1)}$. Substituting this into the objective function, equation (3.8.3), results in

$$T(y,g) = -\frac{1}{2}[(\hat{B}^{t-}Hy)^t(\hat{B}^{t-}Hy) + g^t(I-\hat{B}^{t-}\hat{B}^t)g] + \bar{c}^ty.$$

The matrix $(I-\hat{B}^{t-}\hat{B}^{t})$ is a symmetric positive semidefinite matrix and maximizing $-\frac{1}{2}\, g^{t}(I-\hat{B}^{t-}\hat{B}^{t})g$ implies that g is in the range space of $\hat{B}$. In particular, $g = 0$ is in the range space of $\hat{B}$ and so without loss of generality $T(y,g)$ can be simplified to

$$\psi(y) = -\frac{1}{2}\,(\hat{B}^{t-}Hy)^{t}(\hat{B}^{t-}Hy) + \bar{c}^{t}y.$$

Hence, {CD.1} implies {D.1}.

Q.E.D.

The question of which dual form would be more advantageous for a particular application will be addressed in Chapter 4.

CHAPTER 4

DUAL ALGORITHMIC CONSIDERATIONS

4.1 Introduction

In Chapter 3 three dual forms were introduced for {P.1}; the Wolfe dual {W.1}, the generalized inverse dual {D.1}, and the conjugate dual {CD.1}.

Since {W.1} lacks many desirable properties which {D.1} and {CD.1} possess the analysis of this chapter will address the question of whether {D.1} or {CD.1} offers computational advantages for either or both of the two subclasses of {P.1}.

The two subclasses of {P.1} will be defined as definite problems and semidefinite problems. Definite problems will be those for which $\Phi_o(x)$ is strictly convex or those which have a known active strictly convex constraint. Semidefinite problems will be those which are not definite.

The comparison analysis will be subjective in nature, but will catalog those critical elements of algorithmic development which may likely prove to be advantageous or not computationally.

To this date there has been one reported algorithm developed for {CD.1}. This work, done by Ecker and Niemi [24], reported one experimental result based on a definite problem, but offered no comparison information with other algorithmic approaches.

For intermediate scale definite problems, primal problems of 40 variables and 30 constraints, a projected gradient algorithm was developed for {D.3} and reported on by Hearn and Randolph [35].

Experimental results, CPU time, obtained with this algorithm are catalogued in Table 1. The problems were randomly generated, had 100 percent dense matrices with positive components between zero and ten, and strictly convex objective functions. For each n,m three problems were generated and solved. Also, a general primal algorithm, the Sequential Unconstrained Minimization Technique (SUMT) of Fiacco and McCormick [28] as implemented by Mylander, Holmes and McCormick [48], was applied to three of the primal problems. The first problem $n = 5$, $m = 3$ had a CPU time of 2.6 seconds; the second $n = 10$, $m = 5$ had a CPU time of 11.9 seconds and a third problem, $n = 15$, $m = 10$, was attempted, but terminated after 120 seconds.

4.2 Algorithm

To compare the duals, {D.1} and {CD.1}, the projected gradient algorithm will be used as the standard for investigation. A discussion of the projected gradient algorithm is found in Rosen [59 ,60]. A discussion of feasible direction approaches, which includes gradient projection, is found in Zoutendijk [72].

The basic form of the algorithm is as follows, stated for {D.1}:

(a) At a feasible point y^k compute the gradient of the objective function, $\nabla\psi(y^k)$.

(b) Determine the matrix T, defining the support of the constraint set at y^k. The rows of T are the gradients of the constraints active at y^k (or a subset of the active constraints).

TABLE 1

DUAL EXPERIMENTAL RESULTS* - DEFINITE PROBLEMS

Number of Primal Variables	Number of Dual Variables	Dual CPU Time (seconds**)
5	3	0.3
5	5	0.6
10	5	2.9
15	10	4.7
20	15	14.2
40	15	123.3
40	30	70.7

*Runs made on the IBM 370/165 at the University of Florida Computing Center with a projected gradient dual algorithm.

**Stopping rule based on objective function value accuracy of 10^{-3} resulting in relative precision of 10^{-4} or better for all problems except for the sixth problem which had a relative precision of 10^{-2}.

(c) Compute the projection matrix P, which projects vectors onto the null space of T.

(d) Compute λ^k, $y^{k+1} = y^k + \lambda^k P \nabla\psi(y^k)$, such that $\psi(y^{k+1})$ is maximized, for $P\nabla\psi(y^k) \neq 0$, and go to (b). If $P\nabla\psi(y^k) = 0$ and the orthogonal projection of $\nabla\psi(y^k)$ onto T, the null space of P, results in nonpositive components then y^k is optimal; otherwise select the largest positive component of this projection and remove the associated equality constraint from the set of active constraints and go to (b).

Rosen, in defining the support matrix, required the selection of a linearly independent set of active constraint gradients for the rows of T. Other authors have used the definition of a regular point to meet this requirement, for example, see Luenberger [42]. This particular concept, while theoretically acceptable, creates computational difficulties. The solution to this difficulty is obviated by the use of the generalized inverse of T. That is, by Corollary A.8.1 the projection onto the null space of T is defined by the matrix $(I-T^{-}T)$ and the projection onto the null space of P by $T^{-}T$.

Therefore, in comparing the two dual forms the generalized inverse of the support matrix will be derived.

Other necessary computations for the algorithm involve finding gradients of the objective functions and solving the one-dimensional maximization problem.

An important consequence of both {D.1} and {CD.1} with respect to feasible direction algorithms is that initiating such an algorithm at a dual feasible point will be sufficient to insure that no infeasible dual solutions are generated. For {D.1}, Theorem 3.8 and Corollary 3.8.1

are seen to prove this statement.

Therefore, use of the projected gradient algorithm generates feasible directions and the one-dimensional maximization will result in dual feasible solutions.

4.3 Projection Matrices

The support matrices of the two dual problems will be determined and their null space projectors defined. To distinguish between the two, a sub-D will designate matrices for {D.1} and a sub-c those corresponding to {CD.1}.

To simplify the notation assume that the dual variables, $y_1, y_2, \ldots, y_s$, $s < m$, are zero and the active constraints {D.1} are arranged as

$$
\begin{aligned}
y_1 &= 0 \\
y_2 &= 0 \\
&\vdots \\
y_s &= 0 \\
(I - \hat{Q}^{-}(y)\hat{Q}(y))Hy &= 0,
\end{aligned}
$$

where $y \in \bar{F}$. The matrix defining the support manifold at the intersection of these active constraints is given by

$$T_D = \left[\begin{array}{c|c} I & 0 \\ \hline \multicolumn{2}{c}{(I-\hat{Q}^{-}\hat{Q})H} \end{array}\right]$$

where $I \in E^{sxs}$, $0 \in E^{sx(m-s)}$, $(I-\hat{Q}^{-}\hat{Q})H \in E^{nxm}$, and it is understood that $H = (h_1, h_2, \ldots, h_m)$, the h_o column being removed due to y_o being constant for all $y \in F$. T_D is seen to be an $(s+n)xm$ matrix and its null space projection matrix is defined as

$$P_D = (I - T_D^{-}T_D).$$

By Theorem A.15

$$T_D^t T_D^{t-} = UU^- + CC^-$$

where $U = \left[\frac{I}{0}\right] \in E^{mxs}$, $V = H^t(I-\hat{Q}^-\hat{Q}) \in E^{mxn}$, and $C = (I-UU^-)V$.

By Corollary A.3.4 $U^- = [I,0]$ and $UU^- = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right] \in E^{mxm}$, $I \in E^{sxs}$.

Therefore,

$$C = (I-UU^-)V = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I \end{array}\right] H^t(I-\hat{Q}^-\hat{Q}),\ I \in E^{(m-s)x(m-s)},$$ and

partitioning H, $H = (H_s, \bar{H}_s)$ where $H_s = (h_1,h_2,\ldots,h_s)$ and $\bar{H}_s = (h_{s+1}, h_{s+2}, \ldots, h_m)$ it is seen that

$$C = \left[\frac{0}{\bar{H}_s^t(I-\hat{Q}^-\hat{Q})}\right] \in E^{mxn}\quad,\ 0 \in E^{sxn}.$$

Again applying Corollary A.3.4 results in

$$CC^- = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \bar{H}_s^t(I-\hat{Q}^-\hat{Q})[\bar{H}_s^t(I-\hat{Q}^-\hat{Q})]^- \end{array}\right] \in E^{mxm}$$

and

$$\bar{H}_s^t(I-\hat{Q}^-\hat{Q})[\bar{H}_s^t(I-\hat{Q}^-\hat{Q})]^- = [(I-\hat{Q}^-\hat{Q})\bar{H}_s]^-\ \bar{H}_s - [(I-\hat{Q}^-\hat{Q})\bar{H}_s]^-\hat{Q}^-Q\bar{H}_s$$

which reduces to

$$[(I-\hat{Q}^-\hat{Q})\bar{H}_s]^-\ \bar{H}_s$$

since, by Theorem A.16, $[(I-\hat{Q}^-\hat{Q})\bar{H}_s]^-\hat{Q}^- = [\hat{Q}(I-\hat{Q}^-\hat{Q})\bar{H}_s]^- = 0$.

Combining the above,

$$P_D = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I-[(I-\hat{Q}^-\hat{Q})\bar{H}_s]^-\bar{H}_s \end{array}\right] \in E^{mxm}$$

and $\{I-[(I-\hat{Q}^-\hat{Q})\bar{H}_s]^-\bar{H}_s\} \in E^{(m-s)x(m-s)}$.

Hence, applying the projected gradient algorithm to {D.1} it is only necessary to work with the dual variables which are positive, i.e., y_j such that $j \in N(y)$.

Now turning to {CD.1}, again assume that $y_1,y_2,\ldots,y_s = 0$ and the set of active constraints is

$$y_1 = 0$$
$$y_2 = 0$$
$$\vdots$$
$$y_s = 0$$
$$\sum_{j=0}^{m} B_j^t z_j - \sum_{j=0}^{m} h_j y_j = 0.$$

From this the matrix defining the support manifold of active constraints at the point y is

$$T_c = \left[\begin{array}{c|c} I & 0 \\ \hline -H & \bar{B}^t \end{array}\right] \varepsilon\ E^{(s+n)x(m+\hat{m})}$$

where $I \varepsilon E^{sxs}$, $\hat{m} = (m+1)n$, $\bar{B}^t = (B_0^t, B_1^t, \ldots, B_m^t)$, and $H = (h_1, h_2, \ldots, h_m)$.

The null space projection matrix for {CD.1} is then

$$P_c = (I - T_c^- T_c).$$

Applying Theorem A.15,

$$T_c^t T_c^{t-} = UU^- + CC^-$$

with $U = \left[\begin{array}{c} I \\ \hline 0 \end{array}\right]$ and $C = (I - UU^-) \left[\begin{array}{c} -H^t \\ \hline \bar{B} \end{array}\right]$.

Corollary A.3.4 gives

$$UU^- = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right]$$ resulting in

$$C = \left[\begin{array}{c} 0 \\ \hline -\bar{H}_s^t \\ \hline \bar{B} \end{array}\right], \ 0 \varepsilon\ E^{sxn},$$

and $$CC^- = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \left[\begin{array}{c} -\bar{H}_s^t \\ \bar{B} \end{array}\right] \left[\begin{array}{c} -\bar{H}_s^t \\ \bar{B} \end{array}\right]^- \end{array}\right].$$

Therefore,

$$P_c = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I-(-\bar{H}_s,\bar{B}^t)^-(-\bar{H}_s,\bar{B}^t) \end{array}\right] \in E^{(m+\hat{m})x(m+\hat{m})}$$

where $[I-(-\bar{H}_s,\bar{B}^t)^-(-\bar{H}_s,\bar{B}^t)] \in E^{(\hat{m}+m-s)x(\hat{m}+m-s)}$. Further use of Theorem A.15 will not simplify this form, but again it is seen that the projection matrix P_c depends on the positive y_j and not on the dual variable vectors z_j.

4.4 Objective Function Gradient

The gradient of $\psi(y)$ is given in Theorem 3.8. For {CD.1}, the objective function is

$$T(y,z) = -\frac{1}{2} \sum_{j\in N(y)} \frac{1}{y_j} z_j^t z_j + \bar{c}^t y$$

and its gradient is

$$\nabla T(y,z) = \begin{bmatrix} \frac{\partial T}{\partial y} \\ \frac{\partial T}{\partial y} \\ \vdots \\ \frac{\partial T}{\partial y_m} \\ \nabla_{z_o} T \\ \nabla_{z_1} T \\ \vdots \\ \nabla_{z_m} T \end{bmatrix}$$

where $\frac{\partial T}{\partial y_k} = \frac{1}{2}\frac{1}{y_k^2} z_k^t z_k + c_k$ for $k = 1,2,\ldots,m$,

$\nabla_{z_k} T = -\frac{1}{y_k} z_k$ for $k = 0,1,\ldots,m$, defined only for $y_k > 0$.

4.5 Definite Problems

By definition of definite problems $\hat{Q}(y)$ is nonsingular and by Corollary 3.1.2 {D.1} specializes to {D.3}.

It is clear that for {D.3}; $F = F^*$, the selection of an initial $y \in \mathrm{ri}\, F$ is trivial, and the optimal dual solution completely defines the optimal primal solution.

For {CD.1} the major benefit of the definite problem is in the determination of an initial feasible solution point. That is, say Q_o is positive definite, then

$$z_o = -\left(B_o^t\right)^{-1} \sum_{j=0}^{m} [h_j y_j + B_j^t z_j],$$

arbitrarily selecting $y_j, z_j > 0$ for $j = 1,2,\ldots,m$, computing z_o, and recalling that $y_o = 1$, results in an initial feasible dual solution.

In terms of applying a projected gradient algorithm it is seen that for {D.1} the projection matrix reduces to

$$P_D = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I \end{array}\right] \in E^{mxm},$$

where $I \in E^{sxs}$ but in contrast, for {CD.1}, using Theorem A.6,

$$P_c = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I - \begin{bmatrix} -\bar{H}_s^t \\ \bar{B} \end{bmatrix} (\bar{H}_s \bar{H}_s^t + \hat{Q}(1))^{-1} (-\bar{H}_s, \bar{B}^t) \end{array}\right]$$

where $\hat{Q}(1) = \sum_{j=0}^{m} Q_j$.

One means of comparing algorithms is to calculate the number of multiplications required. This basis of comparison, while not foolproof, can be used as an indicator of relative computing requirements.

Table 2 details the multiplications for {D.1} and {CD.1} applied to definite problems. For convenience it has been assumed that $y_1, y_2, \ldots, y_s = 0$ for both {D.1} and {CD.1} and that the Choleski method of matrix inversion was employed (see Clasen [15] and Westlake [70]). Also, it is assumed that for a given primal problem {D.1} and {CD.1} will be solved in the same number of iterations.

An initial observation to be made is that the determination of P_c and $P_c \nabla T(y,z)$ contributes the majority of the iteration operations to {CD.1}. This implies that if it were known that the optimal solution was in the relative interior of F_c, the feasible region of {CD.1}, then the iteration count for {CD.1} would be drastically reduced and other considerations relative to the two duals would come into play.

Another case where the multiplication counts are significantly different is quadratic programming. Here the iteration counts for both duals are reduced. That is, for {D.1} it is only necessary to compute one inverse at the initiation of the algorithm and for {CD.1} the order of P_c is reduced from $(m+1)n+m-s$ to $n+m-s$. The cumulative result of these reductions though, is that while the order of P_c is reduced it is still necessary to invert an nxn matrix for each iteration not in the relative interior of F_c and {D.1} still offers significantly fewer multiplicative operations.

For the general case, m proper quadratic constraints, it is clear that for noninterior point solutions {D.1} incurs fewer multiplicative operations than does {CD.1}.

Other characteristics to be considered are the relative size of the two duals and associated computer storage requirements. {D.1} has m dual

TABLE 2

DEFINITE PROBLEMS

(Number of Multiplications)

Computations	{D.1}	{CD.1}	Notes
B_j, $j = 0,1,\ldots,m$	-	$\frac{1}{6}(m+1)n^3-6n^2+12n-6)$	Required once per problem
Initial feasible point $y^o,(y^o,z^o)$	-	$\frac{1}{2}n^3+m\,\frac{n(n+1)}{2}+mn+n^2$	Required once per problem
$\psi(y^o),T(y^o,z^o)$	$\frac{1}{2}n^3+(n+1)(m+n-s)$	$(m-s)(3n+1)$	Assumed $y_1^o,y_2^o,\ldots,y_s^o = 0$
$\nabla\psi(y^o),\nabla T(y^o,z^o)$	$(m-s)(n^2+2n)$	$(m-s)n$	
P_D,P_c	-	$\frac{1}{2}n^3+n[(m+1)n+m-s]^2+$ $n^2[(m+1)n+m-s]$	
$P_D\nabla\psi(y)$; $P_c\nabla T(y,z)$	-	$[(m+1)n+m-s]^2$	
Line search	$5[\frac{1}{2}n^3+(n+1)(m+n-s)+$ $m-s]$	$5[(m-s)(3n+2)+(m+1)n]$	Assumed, for purposes of comparison, that 5 functional evaluations are required (based on computational results)
Multiplications per iteration	$3n^3+n^2(m+6-s)+$ $n(6+8(m-s))+7(m-s)$	$n^3(2.5+m^2+3m)+$ $n^2(m^2+5m+3-2(2m+3))+$ $n(m^2+26m+7+s^2-4sm-21s)+$ $(m-s)(11+m-s)$	

variables whereas {CD.1} has $\sum_{j=0}^{m} \rho_j$ + m dual varialbes, ρ_j being the rank of Q_j. For example, if n = 50 and m = 25, assuming the average ρ_j = 25, {D.1} would require 25 dual variables while {CD.1} would require 650 dual variables, 26 times as many. Computer storage requirements will differ in a comparable fashion also. That is, even though {D.1} needs to store the same number of matrices as {CD.1}, the Q_j and B_j, it will be necessary to maintain storage for the matrix P_c, which will be of order 1300 x 1300, and the matrix of z_j dual variables.

In conclusion then for definite problems the generalized inverse dual, {D.1} specialized to {D.3}, should prove to be favorable over the conjugate dual, {CD.1}, computationally.

4.6 Semidefinite Problems

The comparison of {D.1} to {CD.1} for semidefinite problems is not so clear cut. The major difficulty is that of determining an initial feasible dual vector, preferably in the relative interior of F. This problem is common to both dual forms and, therefore, will only be discussed as related to {D.1}.

The ideal initial dual vector for {D.1}, as stated, would be one in the relative interior of F. The following are classifications for which such an initial vector can be readily determined.

(a) For some $k \in \{0,1,2,\ldots,m\}$, Φ_k is strictly convex.

This implies that Q_k is positive definite and F^* is the nonnegative orthant. The initial point, y^1, can be selected such that $y_j^1 > 0$ for $j = 1,2,\ldots,m$, $y^1 \varepsilon$ ri F.

(b) For $j,k = 0,1,2,\ldots,m$, h_j is in the column space of Q_k. Again F^* is the nonnegative orthant and y^1 can be selected such that $y_j^1 > 0$, $j = 1,2,\ldots,m$, $y^1 \varepsilon$ ri F.

The next set of classifications determine an initial dual vector y^1 in F, either a relative boundary or relative interior point.

(c) h_o is in the column space of Q_o, then $y_j^1 = 0$ for $j = 1,2,\ldots,m$ will result in $y^1 \in F$. Also, if there is an index set $I_r = \{0,j_1,j_2,\ldots,j_r\}$, such that for $k \in I_r$, h_k is in the column space of some Q_j, $j \in I_r$, then $y_k^1 > 0$ for $k \in I_r$, $y_k^1 = 0$ for $k \notin I_r$ results in $y^1 \in F$.

(d) There exist $y^1 \geq 0$ such that $Hy^1 = 0$. This clearly results in $y^1 \in F$. $Hy^1 = 0$ and $y^1 \geq 0$ can be expressed by the system

$$\sum_{j=1}^{m} h_j y_j = -h_o, \tag{4.6.1}$$

$$y_j \geq 0 \tag{4.6.2}$$

To facilitate recognition of this characterization it is noted that by Farkas' theorem [43] there will either exist a solution to the system (4.6.1), (4.6.2) or to

$$h_j^t x \leq 0 \text{ for } j = 1,2,\ldots,m \tag{4.6.3}$$

$$-h_o^t x < 0 \tag{4.6.4}$$

but not to both. The system (4.6.3), (4.6.4) is readily recognized through a linear programming formulation. By solving

$$\{LP.1\} \quad \text{minimize} \quad \chi(x) = -h_o^t x$$

$$\text{subject to: } h_j^t x \leq 0 \quad \text{for } j = 1,2,\ldots,m$$

a solution to the system (4.6.3), (4.6.4) is determined to exist or not. That is, if the optimal objective function value for {LP.1} is nonnegative then (4.6.1), (4.6.2) has a solution which is in F.

Obviously, not all problems will fall into the preceding categories and even if they did the recognition of a particular category is not especially straightforward. The introduction of the following technique is intended to offer one approach to alleviating the difficulty by using an extension of the 'Big M' method, see Charnes [12], of linear programming which was also designed for determining initial feasible solutions.

Taking {P.1}, add the $m+1^{st}$ constraint

$$\Phi_{m+1}(s) = \frac{1}{2} x^t I x - M \leq 0$$

with M a large positive number. {P.1} with this constraint will be referred to as the auxiliary problem, {AP.1}. Clearly, $\Phi_{m+1}(x)$ is strictly convex and thereby {AP.1} falls into classification (a) for selecting an initial feasible solution to the dual auxiliary problem, {AD.1}. The objective function of {AD.1} is

$$\psi'(y) = -\frac{1}{2} y^t H^t \hat{Q}^-(y) Hy + \bar{c}y$$

where

$$y^t = (y_o, y\ , \ldots, y_m, y_{m+1}),$$

$$H = (h_o, h\ , \ldots, h_m, 0),$$

$$\hat{Q} = Q_o + \sum_{j=1}^{m} y_j Q_j + y_{m+1} I,$$

and $\bar{c}^t = (c_o, c_1, \ldots, c_m, -M)$. Maximizing $\psi'(y)$ using an initial feasible solution containing $y^1_{m+1} > 0$ will lead to a feasible solution of {D.1}, driving y_{m+1} to zero, if such a feasible solution exists. As in the two phase method of linear programming, see Dantzig [20], when y_{m+1} is reduced to zero continued iterations should be performed using {D.1} instead of {AD.1}, since y_{m+1} will never be made positive again due to the influence of -M on the objective function.

The above technique, while not particularly elegant, is simple to implement and will determine a feasible solution to {D.1} or indicate that {D.1} is infeasible.

Another potentially useful technique for solving semidefinite problems is again based on an auxiliary problem which is definite. The theoretical basis of the approach is found in Fiacco and McCormick [28, Theorem 37]. In essence, the proof is that if a sequence of {P.1} problems is solved with objective functions

$$\Phi_o(\varepsilon_i, x_i) = \frac{1}{2} x_i^t (Q_o + \varepsilon_i I) x_i + h_o^t x + c_o,$$

$$\varepsilon_i > 0 \text{ and } \lim_{i \to \infty} \varepsilon_i = 0$$

then the solution to the original {P.1} is the limit x_i^o as $i \to \infty$. This approach is computationally advantageous in that each problem in the sequence is definite and can be solved using {D.1} with no additional dual variables required. The disadvantage is the requirement to solve a sequence of problems.

For semidefinite problems both {D.1} and {CD.1} involve the computation of generalized inverses. In recent years, a sizable literature has developed relating to generalized inverse computational methods, but experimental results have yet to indicate preferable methods for classes of matrices. Rust, Burrus and Schneeberger [62] have developed a FORTRAN program based on Gramm-Schmidt orthogonalization while Greville [33] and Tewarson [67] have developed methods for special partitioned matrices. Computational methods based on Gauss-Jordan elimination have been proposed by Ben-Israel and Wersan [6] and Noble [49]; whereas Pyle [56] has a method employing a gradient projection method.

Also, Hestenes' [36] biorthogonalization technique has been extended, see Boullion and Odell [8], to computing generalized inverses and Decell [23] and Ben-Israel and Charnes [5] have an approach which is based on the Cayley-Hamilton theorem.

The question of what technique might prove best for evaluating the objective of {D.1} is a numerical one which requires further research.

CHAPTER 5

APPLICATIONS

5.1 Introduction

The intent of this chapter is to present some areas of application in which the generalized inverse dual, {D.1}, should prove useful.

It is demonstrated that the generalized inverse dual of some dual forms results in the solution of the primal problem which was nondifferentiable on the primal feasible space. Also, it is seen that such problems evidence other simplifications in the generalized inverse dual form.

To simplify the notation for presentation, some assumptions have been made which could be relaxed for discussing these applications in detail.

5.2 Multifacility Euclidean Distance Location Problem

This problem is one of locating in E^n N new facilities such that the maximum weighted Euclidean distance from M existing facilities is minimized.

The problem is a minimax version of the well-known Fermat problem which is addressed by Kuhn [39]. (For additional references, see Cabot, Francis, and Stary [10].) Other distance measures for the problem have also been considered; most notably the rectilinear measure considered by Dearing and Francis [22].

The primal Euclidean distance minimax multifacility location problem is:

$$\{M.1\} \quad \underset{x_1,x_2,\ldots,x_N}{\text{minimize}} \quad \text{maximum} \begin{Bmatrix} w_{ji}||x_j-a_i||^2 \text{ for all } j \text{ and } i; \\ u_{jk}||x_j-x_k||^2 \text{ for } 1 \le j < k \le N \end{Bmatrix}$$

where:

$a_i \in E^n$ represents existing facility site for $i = 1,2,\ldots,M$.

$x_j \in E^n$ new facility site for $j = 1,2,\ldots,N$.

$w_{ji} \in E$ given nonnegative weight (squared) representing interaction between new facility x_j and existing facility a_i for all j and i.

$u_{jk} \in E$ given nonnegative weight (squared) representing interaction between new facilities x_j and x_k for $1 \le j < k \le N$.

$||\cdot||$ Euclidean norm function.

{M.1} can be written as a constrained nonlinear programming problem by letting the variable s represent the minimand;

$$\{M.2\} \quad \text{minimize} \quad s$$

$$\text{subject to:} \quad w_{ji}||x_j-a_i||^2-s \le 0$$
$$\text{for } j = 1,2,\ldots,N \text{ and } i = 1,2,\ldots,M,$$
$$u_{jk}||x_j-x_k||^2-s \le 0$$
$$\text{for } 1 \le j < k \le N$$

To facilitate the presentation it will be assumed that w_{ji}, $j = 1, 2,\ldots,N$ and $i = 1,2,\ldots,M$, are all positive and that u_{jk}, for all $1 \le j < k \le N$ are also positive. This assumption can be replaced by a chaining assumption, see Cabot and Francis [9] for discussion, which insures the problem being well formulated.

Using the Euclidean norm squared leads to the following interpretation of {M.2},

{M.3} minimize $p_o^t x$

subject to: $\frac{1}{2} x^t D_j x + h_{ji}^t x + c_i \leq 0$ (5.2.1)

for $j = 1,2,\ldots,N$ and $i = 1,2,\ldots,M$,

$\frac{1}{2} x^t D_{jk} x + p_{jk}^t x \leq 0$ (5.2.2)

for $1 \leq j < k \leq N$

where now $x = (x_1^t, x_2^t, \ldots, x_N^t, s)^t \in E^{nN+1}$,

$x_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^t \in E^n$,

$h_{ji} = (0_1^t, 0_2^t, \ldots, 0_{j-1}^t, -2a_i^t, 0_{j+1}^t, \ldots, 0_N^t, -\frac{1}{w_{ji}})^t \in E^{nN+1}$

with 0_j the zero vector in E^n,

$p_{jk} = (0_1^t, 0_2^t, \ldots, 0_N^t, -\frac{1}{u_{jk}})^t \in E^{nN+1}$,

$p_o = (0_1^t, 0_2^t, \ldots, 0_N^t, 1)^t$,

$c_i = a_i^t a_i \in E^1$, D_j and D_{jk} are matrices.

Let D be an (nN+1)x(nN+1) matrix represented by

$$D = \begin{bmatrix} d_{11}I & d_{12}I & \cdots & d_{1N}I & 0 \\ d_{21}I & d_{22}I & \cdots & d_{2N}I & 0 \\ \vdots & \vdots & & \vdots & \\ d_{N1}I & d_{N2}I & \cdots & d_{NN}I & 0 \\ 0^t & 0^t & \cdots & 0^t & 0 \end{bmatrix}$$

where $d_{ij} \in E^1$, $I \in E^{nxn}$ and $0 \in E^n$. The matrices D_j and D_{jk} are defined as follows;

$D_j = D$ such that $d_{jj} = 2$ and all other $d_{ij} = 0$,

$D_{jk} = D$ such that $d_{jj} = d_{kk} = 2$, $d_{jk} = d_{kj} = -2$ and all other $d_{ij} = 0$.

Clearly, {M.3} is a quadratically constrained quadratic programming problem and, by definition of the Euclidean norm, D_j and D_{jk} for all j and k are positive semidefinite.

Let the dual variables associated with the primal constraints (5.2.1) be designated by y_{ji} and those associated with (5.2.2) be designated by v_{jk}.

Hence, by Theorem 3.1, the dual of {M.3} is

$$\{MD.1\} \quad \text{maximize} \quad \psi(\bar{y}) = -\frac{1}{2}\bar{y}^t H^t \hat{Q}^- H\bar{y} + \sum_{j=1}^{N}\sum_{i=1}^{M} y_{ji}c_i$$

$$\text{subject to:} \quad \bar{y} \geq 0,\ \bar{y} \in E^{NM+[N(N-1)/2]+1}$$

$$\bar{y} = (y_o, y^t, v^t)^t$$

$$y_o = 1$$

$$y = (y_{11}, y_{12}, \ldots, y_{1M}, \ldots, y_{N1}, \ldots, y_{NM})^t$$

$$v = (v_{12}, v_{13}, \ldots, v_{1N}, v_{23}, \ldots, v_{2N}, \ldots, v_{N-1,N})^t$$

$$(I-\hat{Q}\hat{Q}^-)H\bar{y} = 0$$

$$\text{where} \quad \hat{Q} = \sum_{j=1}^{N}\sum_{i=1}^{M} y_{ji}D_j + \sum_{j=1}^{N-1}\sum_{k=j+1}^{N} v_{jk}D_{jk},$$

$$H = (p_o, h_{11}, h_{12}, \ldots, h_{NM}, p_{12}, \ldots, p_{1N}, p_{23}, \ldots, p_{N-1,n}).$$

It is noted that

$$H\bar{y} = \begin{bmatrix} -\sum_{i=1}^{M} y_{1i}a_i \\ -\sum_{i=1}^{M} y_{2i}a_i \\ \vdots \\ -\sum_{i=1}^{M} y_{Ni}a_i \\ 1 - \sum_{j=1}^{N}\sum_{i=1}^{M} \frac{y_{ji}}{w_{ji}} - \sum_{j=1}^{N-1}\sum_{k=j+1}^{N} \frac{v_{jk}}{u_{jk}} \end{bmatrix}$$

Also, by defining

$$\alpha_{jj} = \sum_{i=1}^{M} y_{ji} + \sum_{k=j+1}^{N} v_{jk} + \sum_{m=1}^{j-1} v_{mj} \tag{5.2.3}$$

for $j = 1,2,\ldots,N$, using the convention $\sum_{k=j+1}^{i} = 0$;

and

$$\alpha_{ij} = -v_{ij} \text{ for } 1 \leq i < j \leq N, \tag{5.2.4}$$

$$\alpha_{ji} = \alpha_{ij} \tag{5.2.5}$$

it is seen that

$$\hat{Q} = \left[\begin{array}{c|c} \hat{Q}_1 & 0 \\ \hline 0 & 0 \end{array}\right]$$

where $\hat{Q}_1 = \text{Ix2A}$, $A = (\alpha_{ij}) \varepsilon E^{NxN}$, $I \varepsilon E^{nxn}$, is defined as a Kronecker product. By Corollary A.3.4,

$$\hat{Q}^- = \left[\begin{array}{c|c} \hat{Q}_1^- & 0 \\ \hline 0 & 0 \end{array}\right]$$

and by Theorem A.12

$$\hat{Q}_1^- = I \times \frac{1}{2} A^- .$$

Therefore,

$$F = \{\bar{y} \geq 0 | y_o = 1, \left[\begin{array}{c|c} I-IxAA^- & 0 \\ \hline 0 & 1 \end{array}\right] H\bar{y} = 0\}.$$

Looking at the matrix A it is noted that by setting all dual variables positive and applying the following known theorem of linear algebra, stated here as a lemma with proof omitted, it is seen that A has rank N.

Lemma 5.1. If A is a symmetric NxN matrix and if $\alpha_{ii} > \sum_{\substack{j=1 \\ j\neq i}}^{N} |\alpha_{ij}|$ for $i = 1,2,\ldots,N$, then A is a positive definite matrix.

Therefore, A is nonsingular and

$$F^* = \{\bar{y} \geq 0 \mid y_o = 1, \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{y_{ji}}{w_{ji}} + \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} \frac{v_{jk}}{u_{jk}} = 1\}$$

with an index maximal dual feasible vector being one which has all components positive. The selection of an initial solution in ri F^* = ri F is trivial.

Furthermore, for any feasible vector $\bar{y}$ it is seen by (5.2.3), (5.2.4) and (5.2.5) that if α_{jj} is zero for a particular j then the elements of the associated row and column are zero. Therefore, by an elementary, orthogonal, row and column transformation, E, A can be expressed as,

$$A = E^t \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & 0 \end{array}\right] E$$

where A_{11} has positive diagonal elements.

It can then be shown that

$$A^- = E^t \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & 0 \end{array}\right]^- E$$

and by Corollary A.3.4

$$A^- = E^t \left[\begin{array}{c|c} A_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] E.$$

Often it will hold that A_{11} is also positive definite, in which case, $A_{11}^- = A_{11}^{-1}$ and the dual is further simplified.

Thus, it is seen that if $\psi(\bar{y})$, a concave objective function, is maximized by application of a feasible direction algorithm (feasible directions to F) over F^*, a linearly constrained region, then such a solution also solves {MD.1}. The computational difficulties of solving this dual are of size (dimension of $\hat{Q}$ and H) and not complexity as the foregoing demonstrates.

The analysis of this section can also be extended to include multi-facility problems with constraints. This would encompass some type of constraint on the weighted distance between any two facilities, either existing or new. For a discussion of this problem see Dearing [21].

5.3 Sinha Duality

Another application of {D.1} is in solving Sinha's [65] problem. Sinha has established a dual for one formulation of the stochastic programming problem. The primal and dual forms of Sinha are

{SP.1} maximize $\Phi_o(x) = d^t x - \sum_{j=1}^{m} (x^t Q_j x)^{1/2}$

subject to: $Ax \leq b$

$x \geq 0$

where $d, x \in E^n$, $Q_j \in E^{nxn}$ and symmetric positive semidefinite, $A \in E^{sxn}$, $b \in E^s$

{SD.1} minimize $\Omega(z) = b^t z$

subject to: $A^t z + \sum_{j=1}^{m} Q_j w^j \geq d$

$w^{j^t} Q_j w^j \leq 1$, for $j = 1,2,\ldots,m$

$z \geq 0$

where $w^j \in E^n$ for $j = 1,2,\ldots,m$ and $z \in E^s$.

Making the following notational definitions,

$w^t = (z^t, w^{1^t}, w^{2^t}, \ldots, w^{m^t}) \in E^{s+mn}$,

$\bar{b}^t = (b^t, 0, 0, \ldots, 0) \in E^{s+mn}$

$\bar{Q}^t = (A^t, Q_1^t, Q_2^t, \ldots, Q_m^t) \in E^{nx(s+mn)}$

$\bar{Q}_j \in E^{(s+mn)x(s+mn)}$ an m+1 block diagonal matrix such that there is first an sxs zero matrix and then m nxn matrices, the $j+1^{st}$ matrix on the diagonal being Q_j, all others zero.

$$\bar{Q}_o^t = (I,0,0,\ldots,0) \in E^{s \times (s+mn)}.$$

Sinha's dual is now given by,

$$\{SD.1'\} \quad \text{minimize} \quad \Omega(w) = \bar{b}^t w$$

$$\text{subject to:} \quad \frac{1}{2} w^t \bar{Q}_j w - \frac{1}{2} \leq 0$$

$$\text{for } j = 1,2,\ldots,m$$

$$-\bar{Q}^t w + d \leq 0$$

$$-\bar{Q}_o^t w \leq 0$$

Defining $y^t = (y_o, y_m^t, y_n^t, y_s^t) \in E^{m+n+s+1}$ where $y_o = 1$, $y_m \in E^m$, $y_n \in E^n$, and $y_s \in E^s$ results in the following generalized inverse dual.

$$\{D.5\} \quad \text{maximize} \quad \psi(y) = -\frac{1}{2}\{\sum_{j=1}^{m} [y_n^t Q_j (y_{m_j} Q_j)^- Q_j y_n + y_{m_j}]\} + d^t y_n$$

$$\text{subject to:} \quad A^t y_n + I_{s \times s} y_s = b$$

$$[I-(y_{m_j} Q_j)^-(y_{m_j} Q_j)] Q_j y_n = 0$$

$$\text{for } j = 1,2,\ldots,m.$$

$$y_m \geq 0$$

$$y_n \geq 0$$

$$y_s \geq 0$$

where it is seen that the vector y_s can be dropped and the first linear equality converted to a linear inequality, $A^t y_n \leq b$.

It is first observed that while Sinha's dual has $s+mn$ variables, {D.5} has only $m+n$. Furthermore, $\psi(y)$ can be rewritten, based on the properties of the generalized inverse and Corollary A.3.5, as

$$\psi(y) = -\frac{1}{2}\{\sum_{j \in N(y_m)} [\frac{y_n^t Q_j y_n}{y_{m_j}} + y_{m_j}]\} + d^t y_n$$

where $N(y_m) = \{j | y_{m_j} > 0\}$, with $Q_j y_n = 0$ for $j \notin N(y_m)$.

The gradient of $\psi(y)$ for any feasible point is likewise available by Theorem 3.8 and recalling that if $(y_{m_j} Q_j)$ is the zero matrix, then its generalized inverse is also a zero matrix. An initial feasible point for {D.5} can be derived by solving $A^t y_n \leq b$, $y_n \geq 0$, and setting $y_{m_j} = 1$ for all $j = 1,2,\ldots,m$. Hence, solution procedures for {D.5} will require no matrix inversion computations and therefore is computationally tractable.

The merit of {D.5} is exemplified in the next theorem.

<u>Theorem 5.1</u>. Given that $(y_n^{ot}, y_m^{ot})^t$ solves {D.5} then $x^o = y_n^o$ solves {SP.1}. Also, if x^o solves {SP.1} then (x^{ot}, y_m^{ot}) solves {D.5} where $y_{m_j}^o = (x^{ot} Q_j x^o)^{1/2}$.

<u>Proof of Theorem 5.1</u>. Clearly $x^o = y_n^o$ is feasible to {SP.1}. Also, for $j \varepsilon N(y_m^o)$

$$\frac{\partial\psi(y)}{\partial y_{m_j}} = \frac{1}{2}\left\{ \frac{y_n^{ot} Q_j y_n^o}{y_{m_j}^{o2}} - 1\right\} = 0$$

and $$y_{m_j}^o = (y_n^{ot} Q_j y_n^o)^{1/2}.$$

Substituting into $\psi(y)$,

$$\psi(y^o) = \sum_{j\varepsilon N(y_m^o)} (y_n^{ot} Q_j y_n^o)^{1/2} + d^t y_n^o = \Phi_o(x{=}y_n^o)$$

where it is recalled that if $j \not\varepsilon N(y_m^o)$, $Q_j y_n^o = 0$. Then $\Phi_o(x{=}y_n^o) \geq \Phi_o(x)$ for all feasible x to {SP.1}; assuming otherwise leads to a contradiction on y_n^o being optimal to {D.5}.

Now, assuming that x^o is a solution to {SP.1} it is seen that by letting $y_{m_j}^o = (x^{ot} Q_j x^o)^{1/2}$ that $(x^{ot}, y_m^t)^t$ is feasible to {D.5} and $\psi(x^o, y_m^o) = \Phi_o(x^o)$. Also, $\psi(x^o, y_m^o) \geq \psi(y_n, y_m)$ for $(y_n^t, y_m^t)^t$ feasible to {D.5} as otherwise a contradiction results to x^o being optimal to {SP.1}.

Q.E.D.

By using {D.5} instead of {SP.1} the difficulty of the nondifferentiability of $\phi_o(x)$ is avoided by expanding the problem dimensionality to n+m. This, opposed to {SD.1} having dimensionality of s+mn, should offer significant advantages in developing efficient computational procedures. Also, solutions to {D.5} define solutions to {SP.1} by the above theorem and it is unnecessary to invoke the added complexity of computing primal variables from dual solutions to {SD.1} which as Sinha discusses will often involve solution of a linear program.

Obvious application of {D.5} is to stochastic programming problems, but another application is found in the work of Cabot and Francis [9] who used a formulation of Sinha's dual in investigating a multifacility location problem involving Euclidean distances. Again the difficulty of the primal problem involves the nondifferentiability of the objective function at feasible points.

Cabot and Francis [9] employ an equality constrained form of Sinha's primal and develop an equality constrained dual. In handling this through {D.5} there are two alternatives. First, as per section 3.6, the linear equality constraint could be used to convert the Sinha dual to an equivalent form with no equality constraints and then formulate {D.5}. Alternatively, the equality constraint could be retained, resulting in y_n being unrestricted in {D.5}. Because of the necessity to compute the generalized inverse of an nx(mn+5) matrix with the first alternative and, by Theorem 5.1, the fact that the second alternative results in the primal solution, the second alternative should prove preferable even at the cost of more variables.

5.4 General Fermat Problem

The Fermat problem dates from the 17^{th} Century and was originally stated as: Given three points in the plane, find a fourth point such that the sum of its distances to the three given points is a minimum. This form of the problem has been addressed by several authors; see Kuhn [39] for a historical sketch.

The general Fermat problem is: given m points in the plane, find the point which minimizes the sum of positively weighted distances to the m points.

$$\{GF.1\} \quad \underset{x}{\text{minimize}} \quad \Phi_o(x) = \sum_{j=1}^{m} w_j ||x-p_j||$$

for $w_j > 0$, p_j the given j^{th} point and $||x-y||$ the Euclidean distance between points x and y.

From the preceding section this is recognized as a special case of the Sinha primal. If one dualizes the corresponding Sinha dual, as in section 5.3, a new version of {GF.1} results, namely

$$\{D.6\} \quad \text{minimize} \quad \psi'(x,y) = \frac{1}{2} \sum_{j=1}^{m} [(x-p_j)^t (y_j I)^- (x-p_j) + y_j w_j^2]$$

$$\text{subject to:} \quad y_j \geq 0$$

$$[I - (y_j I)^- (y_j I)](x-p_j) = 0$$

$$\text{for } j = 1,2,\ldots,m$$

It is clear that $F^* = \{(x,y) \in E^{m+n} | y \geq 0\}$ and an algorithm can readily be initiated at a relative interior point of F. Use of a projected gradient algorithm for {D.5} is easily implemented by noting that the projection matrix is diagonal with zero diagonal elements associated with $y_j = 0$ and $x = p_j$ and unity elsewhere.

Note that at the optimal point, for $y_j > 0$,

$$\frac{\partial\psi'(x,y)}{\partial y_j} = -\frac{(x^o-p_j)^t(x^o-p_j)}{y_j^{o2}} + w_j^2 = 0$$

which implies that $y_j^o = w_j^{-1}||x^o-p_j||$ and as in Theorem 5.1, at optimality $\psi'(x^o,y^o) = \phi_o(x^o)$.

By the theory of Chapter 3 the function $\psi'(x,y)$ is convex and twice continuously differentiable over $y > 0$, x unrestricted, the relative interior of F. Furthermore, if some $y_j \to 0$, $\psi'(x,y) \to +\infty$ unless $x \to p_j$. This is seen by noting that the numerator $||x-p_j||^2$ goes to zero faster than y_j. It is noted that this notion applies to the more general multifacility location problem of Cabot and Francis [9].

5.5 Portfolio Selection

There exist several forms of this problem; examples are found in Saaty and Bram [63], Markowitz [45], Mao and Särndal [44], Roy [61] and Sharpe [64].

The problem addressed here is

{PS.1} maximize $\sum_{j=1}^{n} \mu_j x_j$ $\qquad$ (minimize $-\sum_{j=1}^{n} \mu_j x_j$)

subject to: $\sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j \leq \sigma$

$$\sum_{j=1}^{n} x_j = c$$

$$x_j \geq 0$$

where a given sum of money c is to be allocated to n securities, μ_j is the expected rate of return on the j^{th} security, the nonzero symmetric matrix (a_{ij}) is the covariance matrix of the random variable r_j, the return rate on the j^{th} security, and σ is a constraint on the variance allowable.

The equality constraint is equivalent to

$$x = (h^t)^- c + (I-(h^t)^- h^t)g, \; g \in E^n$$

where $h^t = (1,1,\ldots,1) \in E^n$ each element being unity. It is easily shown that $(h^t)^- = \frac{1}{n} h$, $(h^t)(h^t)^- = 1$, and

$$J = [I-(h^t)^-(h^t)] = \frac{1}{n} \begin{bmatrix} n-1 & -1 & \ldots & -1 \\ -1 & n-1 & \ldots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \ldots & n-1 \end{bmatrix} \quad \text{has rank } n-1.$$

Hence,

$$x = \frac{c}{n} h + Jg$$

and {PS.1} is transformed to determining g.

{PS.2} minimize $\quad -\frac{c}{n} u^t h - u^t Jg$

subject to: $\frac{1}{2} g^t JQJg + \frac{c}{n} h^t QJg - (\sigma - \frac{1}{2} (\frac{c}{n})^2 h^t Qh) \leq 0$

$-Jg - \frac{c}{n} h \leq 0$

where $Q = 2A = 2(a_{ij}) \in E^{nxn}$,

$u^t = (\mu_1, \mu_2, \ldots, \mu_n) \in E^n$ and $g \in E^n$.

{PS.2} has a linear objective function, one quadratic constraint and n linear inequality constraints.

The generalized inverse dual of {PS.2} is

{D.7} maximize $\quad \psi(y) = -\frac{1}{2} y^t H^t (y_1 Q')^- Hy + \bar{c}^t y$

subject to: $y = (y_o, y_1, \ldots, y_{n+1})^t \geq 0$

$y_o = 1$

$Cy = 0$

where $H = (u, -\frac{c}{n} Qh, I)$, $Q' = JQJ$,

$C = [I-(y_1 JQJ)^-(y_1 JQJ)]H,$

and

$$\bar{c}^t = (-\frac{c}{n} u^t h, -\sigma + (\frac{c}{n})^2 h^t Ah, -\frac{c}{n}, \ldots, -\frac{c}{n}) \in E^{n+2}.$$

$\psi(y)$ has been simplified by the readily proved identity,

$(JQJ)^- = J(JQJ)^- J.$

The form of {D.7} can possibly be further simplifed depending on the probability distribution assumed for the random variables r_j; that is, if a multivariate normal distribution is assumed, A will be non-singular and Cy = 0 will reduce to a single linear constraint for $y > 0$.

{D.7} has a concave objective function, linear constraints and offers computational advantages over {PS.1}. The feasible set of {D.7} permits the straightforward application of Rosen's gradient projection algorithm and it is only necessary to compute the generalized inverse of $Q' = JQJ$ once.

5.6 Convex Programming

Topkis and Veinott [68] have stated conditions on the directions and step sizes which assure convergence to a stationary point in feasible direction algorithms for minimizing a real-valued continuous function on a closed set. In particular, for constrained problems, they state second order methods which require the solution of quadratically constrained quadratic programs for the determination of feasible directions. See Topkis and Veinott, Theorem 3 and Lemmas 4 and 5.

Given the mathematical program;

$$\begin{aligned} &\text{minimize} \quad \Phi_o(x) \\ &\text{subject to:} \quad \Phi_j(x) \leq 0 \\ &\qquad\qquad\quad \text{for } j = 1,2,\ldots,m \end{aligned}$$

where the $\Phi_j(x)$, $j = 0,1,2,\ldots,m$ are convex and twice differentiable.

The basic second order method algorithm is stated as follows for the k+1st solution given the kth feasible solution x^k:

(1) Is x^k optimal? If yes, stop, otherwise continue.

(2) Compute feasible direction d^k by solution of the following program,

$$\begin{aligned} &\text{minimize} \quad s \\ &\text{subject to:} \quad [\nabla_x \Phi_o(x^k)]^t d + \frac{1}{2} d^t H(\Phi_o(x^k))d - s \leq 0 \\ &\qquad\qquad\quad \Phi_j(x^k) + [\nabla_x \Phi_j(x^k)]^t d + \frac{1}{2} d^t H(\Phi_j(x^k))d - s \leq 0 \\ &\qquad\qquad\quad \text{for } j = 1,2,\ldots,m \\ &\qquad\qquad\quad d^t d \leq 1 \end{aligned}$$

using the generalized inverse dual, {D.1}.

(3) Solve the following for λ^k;

$$\begin{aligned} &\underset{\lambda \geq 0}{\text{minimize}} \quad \Phi_o(x^k + \lambda d^k) \\ &\text{subject to:} \quad \Phi_j(x^k + \lambda d^k) \leq 0 \\ &\qquad\qquad\quad \text{for } j = 1,2,\ldots,m \end{aligned}$$

Put $x^{k+1} = x^k + \lambda^k d^k$ and return to Step (1).

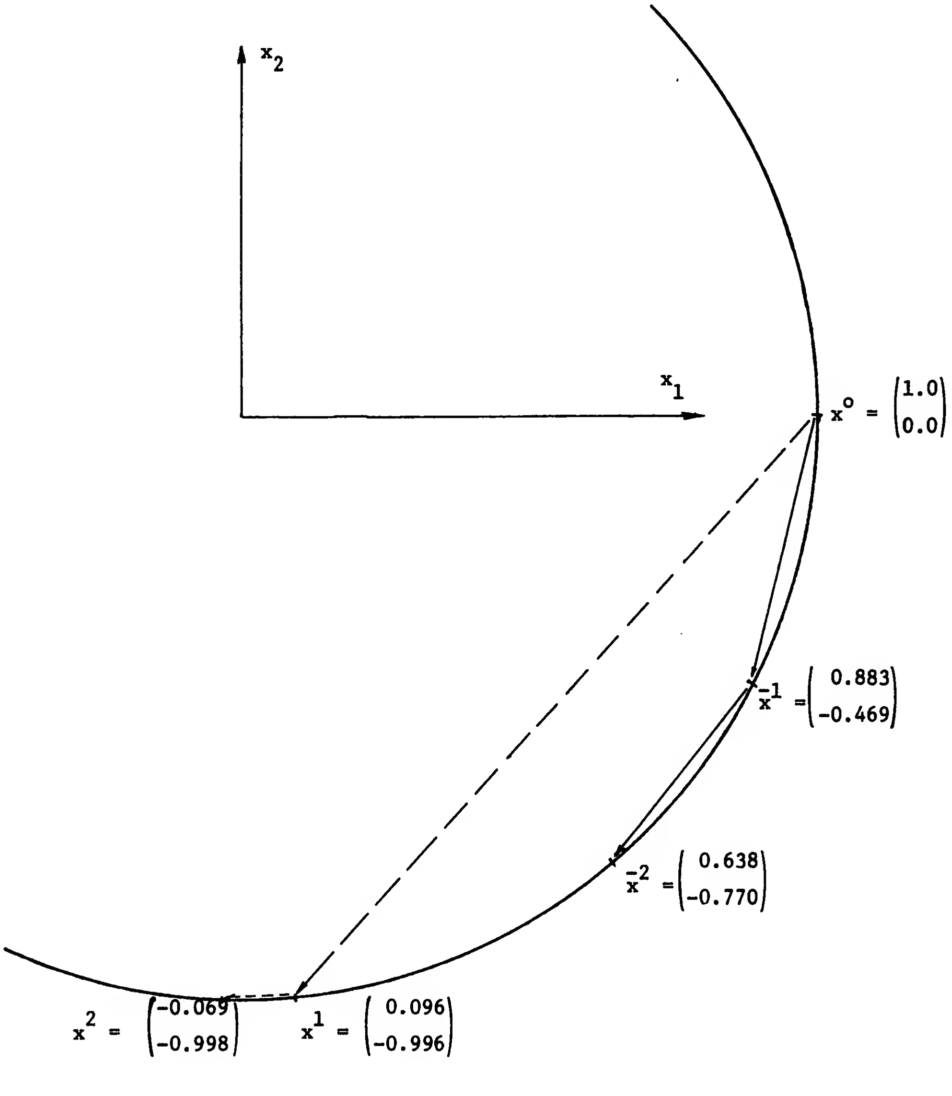
Topkis and Veinott, among others, point out that such an algorithm could prove superior to available first order methods. To see this and gain an appreciation of the potential of the algorithm the following example is offered.

{E.3} minimize x_2

subject to: $(x_1^2 + x_2^2)^2 \leq 1$

The result of two iterations for this problem using the proposed second order method is seen in Figure 5-1 as the dashed line. The second step estimates the minimum at $x_1 = -0.069$ and $x_2 = -0.998$ where the actual values for the minimum are $x_1 = 0$ and $x_2 = -1.000$. The corresponding results for the first-order method (In a first order method d^k is obtained in Step (2) as the solution of a linear program.) appear in Figure 5-1 as the solid line.

The generalized inverse dual, {D.1}, of Chapter 3 offers a computationally tractable method of determining second order feasible directions which has been the major drawback to the algorithm.



x_2
x_1
$x^o = \begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix}$
$\bar{x}^1 = \begin{pmatrix} 0.883 \\ -0.469 \end{pmatrix}$
$\bar{x}^2 = \begin{pmatrix} 0.638 \\ -0.770 \end{pmatrix}$
$x^2 = \begin{pmatrix} -0.069 \\ -0.998 \end{pmatrix}$
$x^1 = \begin{pmatrix} 0.096 \\ -0.996 \end{pmatrix}$

Figure 5-1

CHAPTER 6

FUTURE RESEARCH

In the foregoing, a viable duality theory has been developed for convex quadratically constrained quadratic programs. The results of this research have exposed other pertinent areas in which further investigation should prove fruitful. These areas can be broadly categorized as theory, algorithm design, and numerical theory associated with algorithm implementation.

Attempts at generalizing the primal problem convex quadratic functions to pseudoconvex and quasiconvex quadratic functions has proved to be unsatisfactory. That is, applying the results of Martos [46] and Cottle and Ferland [19] it can readily be shown that the Lagrangian function does not exhibit any usable mathematical structure. This is not totally unexpected because, in general, sums of pseudoconvex or quasiconvex functions are not pseudoconvex or quasiconvex. The significance, though, lies in the fact that while convex quadratic functions impart to the Lagrangian their mathematical structure, a generalization of convexity for quadratic functions negates this structural transfer.

The research into which, if any, subclasses of the generalized convex quadratically constrained quadratic programs evidence a viable generalized inverse dual should continue. The duality theorems of Karamardian [38], for Lagrangian functions which are strictly quasiconvex in the primal variables and strictly quasiconcave in the dual variables, could prove to be a starting point for defining these subclasses. As noted, investigations to date have been based on structural properties of the primal problem and not on the dual form. Hence, one must determine if a recognizable mathematical structure on the dual form dictates structural properties for the primal problem.

Another area for continued research is the extension of the concept of an index maximal dual feasible vector as introduced in Chapter 3. It was seen how this concept led to some very interesting theoretical results for the class {P.1} and it is possible that a broader interpretation and use can be established.

A prelude to the above could be the examination of minimal dimension dual spaces. This idea is prompted by the demonstrated equivalence between the generalized inverse dual and the conjugate function dual. That is, the generalized inverse dual has but one dual variable associated with each inequality constraint of the primal problem, whereas the conjugate function dual associates n+1 dual variables with each inequality constraint and the generalized inverse dual space is smaller than that of the conjugate dual space.

In a sense, the property of taking an n variable, m constraint problem and developing an m variable, n constraint dual has been extended to convex quadratically constrained quadratic programs. Can this extension be carried further to other subclasses of convex programs, the dimension of the dual space being determined by the number of inequalities of the primal problem? If so, what advantages are there?

Another area of particular interest is the investigation of the conjugate function in relation to the generalized inverse. Again this research is prompted by the equivalence between the generalized inverse dual and the conjugate function dual. What is the detailed relationship between the two concepts for quadratic functions and can this relationship be extended to other classes of functions?

Finally, the question of second-order feasible direction methods should be investigated further in light of the generalized inverse dual. The research of this topic falls into all three categories: theory, algorithm design, and numerical theory.

With respect to algorithm design, there are two general areas for future research, the first being design of an algorithm based on the generalized inverse dual for semidefinite problems. Two possible approaches to such an algorithm would be to design around the generalized inverse dual exclusively or to use the generalized inverse dual as the primary form and the conjugate function dual for unidirectional searches, because of its simpler objective function.

The second area of algorithm design is based on the results reported in Chapter 5. Specialized algorithms for the general Fermat problem and extensions to the multifacility problem of Cabot and Francis could be significant improvements over currently available algorithms.

A direct descendant of algorithm design is research into the numerical analysis of algorithm implementation. It is first noted that the existing implemented algorithm for definite problems does not take full advantage of the state of the art. That is, an open question remains as to how efficiently can definite problems be solved by use of the generalized inverse dual.

For semidefinite problems many questions of implementation require research. For example, would it be better to compute $\hat{Q}^{-}(y)$ or solve $\hat{Q}(y)x = -Hy$? Also, with respect to an algorithm design combining the generalized inverse and conjugate function dual, what is required in terms of computation to transfer from one dual form to the other?

In conclusion, it can be seen from this partial listing that several interesting and potentially beneficial areas of research have been opened by the introduction of the generalized inverse dual.

APPENDICES

APPENDIX A

THE MOORE-PENROSE GENERALIZED INVERSE

Ordinarily matrices are thought of as being either nonsingular or singular and consequently having or not having an inverse. In 1920 Moore [47] published a paper on the reciprocal of a general matrix, but it went unnoticed until 1955 when Penrose [50], unaware of Moore's work, published a paper on the generalized inverse of a matrix. Penrose's paper created the impetus for further research and several authors have now added to the theory of what is known as the Moore-Penrose generalized inverse. Because there is yet no standard nomenclature, some authors refer to the Moore-Penrose pseudo-inverse, pseudo-inverse, general reciprocal, Moore-Penrose generalized inverse, generalized inverse, while others refer to various other matrix inverse forms by these and other names. The recent work of Boullion and Odell [8] has an extensive bibliography on the subject. Herein, Moore-Penrose generalized inverse and generalized inverse will be synonymous.

The major advantage of the Moore-Penrose generalized inverse is that it exists and is unique for any matrix. This permits a unified theoretical treatment of matrix calculus and greater flexibility in the application of matrix methods to other fields.

As is often the case in matrix calculus, initial efforts at developing the theory of the generalized inverse were carried on in the field of statistics. A representative bibliography of these works is found in Boullion and Odell.

The application of the generalized inverse theory to operations research has been almost exclusively in the area of linear programming and carried out by Cline [18], Pyle [55], and Charnes, Cooper and Thompson [13]. Charnes and Kirby [14] have addressed the generalized inverse applied to a convex programming problem.

This appendix is included to support the self-contained nature of the dissertation and presents a compilation of the Moore-Penrose generalized inverse theory which is germane to the extensions of Chapter 2, the duality theory of Chapter 3, and the equivalence concepts of Chapter 4. The theory herein presented is for real matrices, but the extension to matrices over the complex field is immediate.

The first definition is due to Penrose [50]. The second and alternate definition, due to Moore [47], is used for theoretical development of derivative forms.

<u>Definition</u>. Let A be an nxm matrix. A matrix A^- which has the following properties is called a <u>generalized inverse of A</u>:

(*i*) AA^- is symmetric

(*ii*) A^-A is symmetric

(*iii*) $AA^-A = A$

(*iv*) $A^-AA^- = A^-$

<u>Definition</u>. For any nxm matrix A, the generalized inverse is defined as

$$A^- = \lim_{\delta \downarrow 0} (A^tA + \delta^2 I)^{-1} A^t$$

$$= \lim_{\delta \downarrow 0} A^t(AA^t + \delta^2 I)^{-1}.$$

A discussion and proof of the equivalence of these two definitions is found in Albert [1].

Before proving the existence theorem for generalized inverses, three lemmas are stated and proved. These lemmas are central to the existence theorem proof.

Lemma A.1. Let M be a matrix of order nxm of rank r, then there are matrices P, of order nxr, and Q, of order rxm, both of rank r such that M = PQ.

Proof of Lemma A.1. Suppose that $n \geq m$, there exist nonsingular matrices R and S such that

$$R^{-1}MS^{-1} = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] \qquad \text{or} \qquad M = R \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] S.$$

Put $P = R \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]$ and $Q = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] S$,

then M = PQ and P and Q both have rank r.

Q.E.D.

Lemma A.2. Let M be an mxr matrix with rank r. Then M^tM is nonsingular.

Proof of Lemma A.2. Since M is of rank r, there is a nonsingular matrix P of size rxr such that

$$MP = \begin{bmatrix} I_r \\ 0 \end{bmatrix} ,$$

$$P^tM^t = (I_r, 0) ,$$

and

$$P^tM^tMP = (I_r, 0) \begin{bmatrix} I_r \\ 0 \end{bmatrix} = I_r .$$

Thus, there exists a nonsingular matrix P such that

$$P^tM^tMP = I_r ,$$

M^tM is now equivalent to I_r and hence nonsingular.

Q.E.D.

<u>Lemma A.3</u>. Let M be an rxm matrix of rank r. Then MM^t is nonsingular.

<u>Proof of Lemma A.3</u>. Define $\bar{M} = M^t$ and apply Lemma 2 proving $\bar{M}^t\bar{M}$ is nonsingular, but $\bar{M}^t\bar{M} = MM^t$ is nonsingular.

Q.E.D.

<u>Theorem A.1</u>. The generalized inverse of an nxm matrix A, if it exists, is of order mxn.

<u>Proof of Theorem A.1</u>. By the properties of the generalized inverse and conformability requirements it is seen that for AA^- and A^-A to be symmetric with A of order nxm, A^- must be of order mxn.

Q.E.D.

The existence theorem of the generalized inverse can now be stated.

<u>Theorem A.2</u>. Let A be an nxm matrix. Then A has a generalized inverse.

<u>Proof of Theorem A.2</u>. If A is a zero matrix of order nxm then A^- is a zero matrix of mxn. Assume that A is not a zero matrix. By Lemma A.1, A can be factored in the form

$$A = PQ$$

where P is nxr and Q is rxm matrices of rank r. P^tP and QQ^t are nonsingular by Lemmas A.2 and A.3.

Put $A^- = Q^t(QQ^t)^{-1}(P^tP)^{-1}P^t$, then

(i) $(AA^-)^t = A^{-t}A^t$

$$= [Q^t(QQ^t)^{-1}(P^tP)^{-1}P^t]^t(PQ)^t$$

$$= P\{(P^tP)^{-1}\}^t\{(QQ^t)^{-1}\}^tQQ^tP^t$$

$$= P(P^tP)^{-1}(QQ^t)^{-1}(QQ^t)P^t$$

$$= P(P^tP)^{-1}P^t.$$

Therefore, AA^- is symmetric and satisfies *(i)* of the definition.

(ii) $(A^-A)^t = A^tA^{-t}$

$$= (PQ)^t[Q^t(QQ^t)^{-1}(P^tP)^{-1}P^t]^t$$

$$= Q^tP^t[P\{(P^tP)^{-1}\}^t\{(QQ^t)^{-1}\}^tQ]$$

$$= Q^t(QQ^t)^{-1}Q.$$

A^-A is symmetric and satisfies *(ii)* of the definition.

(iii) $AA^-A = PQ[Q^t(QQ^t)^{-1}(P^tP)^{-1}P^t]PQ$

$$= PQ$$

$$= A,$$

satisfying *(iii)* of the definition.

(iv) $A^-AA^- = [Q^t(QQ^t)^{-1}(P^t)^{-1}P^t]PQ[Q^t(QQ^t)^{-1}(P^tP)^{-1}P^t]$

$$= Q^t(QQ^t)^{-1}(P^tP)^{-1}P^t$$

$$= A^-.$$

The defined $A^- = Q^t(QQ^t)^{-1}(P^tP)^{-1}P^t$ is therefore a generalized inverse of A and always exists.

Q.E.D.

In the following, it is shown that the generalized inverse as defined is unique.

<u>Theorem A.3</u>. Let A be an nxm matrix. The generalized inverse A^- of A is unique.

Proof of Theorem A.3. Suppose that A has two generalized inverses, say A_1^- and A_2^-. It will first be shown that $AA_1^- = AA_2^-$.

$$\begin{aligned} AA_2^- &= (AA_1^-A)A_2^- \\ &= (AA_1^-)(AA_2^-) \\ &= (AA_2^-)^t(AA_1^-)^t \\ &= (AA_2^-A)A_1^- \\ &= AA_1^- \end{aligned}$$

Similarly, $A_2^-A = A_1^-A$. Therefore,

$$\begin{aligned} A_2^- &= A_2^-AA_2^- \\ &= (A_2^-A)A_2^- \\ &= (A_1^-A)A_2^- \\ &= A_1^-(AA_2^-) \\ &= A_1^-AA_1^- \\ &= A_1^-. \end{aligned}$$

Hence, the generalized inverse is unique.

Q.E.D.

Corollary A.3.1. The generalized inverse of A^t is the transpose of the generalized inverse of A; i.e., $(A^t)^- = (A^-)^t$.

Proof of Corollary A.3.1. The proof consists of showing that $(A^-)^t$ is the generalized inverse of A^t and since the generalized inverse is unique by Theorem A.3, it follows that $(A^t)^- = (A^-)^t$. Write A = PQ as in the proof of Theorem A.2,

$$A^- = Q^t(QQ^t)^{-1}(P^tP)^{-1}P^t. \tag{A.1}$$

Also,

$$A^t = Q^tP^t$$

and

$$(A^t)^- = P(P^tP)^{-1}(QQ^t)^{-1}Q.$$

Taking the transpose of (A.1),

$$(A^{-})^{t} = P(P^{t}P)^{-1}(QQ^{t})^{-1}Q.$$

Hence,

$$(A^{t})^{-} = (A^{-})^{t}$$

and by uniqueness of the generalized inverse of A^{t} the corollary is proved.

Q.E.D.

A direct consequence of this corollary for symmetric matrices is the following:

$$\begin{aligned}(A^{t})^{-} &= (A)^{-}\\ &= (A^{-})^{t}\end{aligned}$$

That is, the generalized inverse of a symmetric matrix is symmetric.

<u>Corollary A.3.2</u>. The following hold for any mxn matrix A:

$$(AA^{t})^{-} = (A^{-})^{t}A^{-},$$

$$(A^{t}A)^{-} = A^{-}(A^{-})^{t}.$$

<u>Proof of Corollary A.3.2</u>. Clearly, conformability requirements are satisfied. The corollary is then proved by demonstrating that the four properties of the definition are satisfied.

<u>Corollary A.3.3</u>. The generalized inverse of an mxn matrix A can be expressed as

$$A^{-} = A^{t}(AA^{t})^{-}$$

or

$$A^{-} = (A^{t}A)^{-}A^{t}.$$

<u>Proof of Corollary A.3.3</u>. Proof is by demonstrating that the definition is satisfied. Note that conformability requirements are met.

For the first,

$$(i)\quad (AA^-)^t = (AA^t(AA^t)^-)^t$$

$$= AA^t(AA^t)^-$$

$$= AA^-$$

by noting that $(AA^t)^-$ is by definition the generalized inverse of AA^t.

$$(ii)\quad (A^-A)^t = (A^t(AA^t)^-A)^t$$

$$= A^t(AA^t)^{-t}A$$

$$= A^t(AA^t)^-A$$

$$= A^-A.$$

$$(iii)\quad AA^-A = A(A^t(AA^t)^-)A$$

$$= AA^t(AA^t)^-A$$

$$= AA^t(A^-)^tA^-A$$

$$= A(A^-A)^tA^-A$$

$$= (AA^-A)A^-A$$

$$= AA^-A$$

$$= A.$$

$$(iv)\quad A^-AA^- = A^t(AA^t)^-AA^t(AA^t)^-$$

$$= A^t(A^-)^tA^-AA^t(A^-)^tA^-$$

$$= (A^-A)^tA^-A(A^-A)^tA^-$$

$$= A^-AA^-A(A^-A)A^-$$

$$= A^-AA^-$$

$$= A^-.$$

The proof of the second form follows in the same manner.

Q.E.D.

Corollary A.3.4. The generalized inverse of $A_1 = \left[\begin{array}{c} A \\ \hline 0 \end{array}\right]$ is

$$A_1^- = (A^-, 0).$$

Proof of Corollary A.3.4. The proof is by observing that the four properties of the definition are satisfied.

Corollary A.3.5.

$$(\alpha A)^- = \frac{1}{\alpha} A^- \quad \text{for } \alpha \varepsilon E^1,\ \alpha \neq 0.$$

Proof of Corollary A.3.5. Again, proof is by noting that the definition is satisfied.

Theorem A.4. If A is nonsingular, $A^- = A^{-1}$.

Proof of Theorem A.4. The proof follows by noting that A^{-1} satisfies the definition for the generalized inverse of nonsingular A.

Theorem A.5. Let A be a matrix, then rank (A^-) = rank (A).

Proof of Theorem A.5.

$$\begin{aligned} \text{rank }(A) &= \text{rank }(AA^-A) \leq \text{rank }(A^-) \\ &= \text{rank }(A^-AA^-) \leq \text{rank }(A) \end{aligned}$$

Q.E.D.

Theorem A.6. Let A be an nxm matrix. Then if

(a) rank (A) = n, $\quad A^- = A^t(AA^t)^{-1}$

(b) rank (A) = m, $\quad A^- = (A^tA)^{-1}A^t$.

From (a) $AA^- = I$ and from (b) $A^-A = I$.

Proof of Theorem A.6. The proof of this theorem follows directly from Corollary A.3.4 and Lemmas A.2 and A.3.

Theorem A.7. The system of equations, $Ax = b$, is consistent if and only if $AA^-b = b$; i.e., AA^- is the projector onto R(A), the range space of A, and so b is in the column space of A.

Proof of Theorem A.7. Suppose the system is consistent, then there exists an x such that $Ax = b$, further

$$(AA^{-})Ax = (AA^{-})b$$
$$Ax = AA^{-}b$$
$$b = AA^{-}b.$$

Suppose now that $AA^{-}b = b$. Put $x = A^{-}b$ then $Ax = AA^{-}b = b$. Hence, x satisfies $Ax = b$, which is then consistent.

Q.E.D.

Theorem A.8. Let the system of equations $Ax = b$ be consistent. Then the vector

$$x = A^{-}b + (I-A^{-}A)g$$

is a solution to $Ax = b$ for every choice of g in the proper space. Moreover, every solution to $Ax = b$ has this form.

Proof of Theorem A.8. For every choice of g,

$$\begin{aligned} Ax &= A\{A^{-}b\} + A(I-A^{-}A)g \\ &= AA^{-}b + Ag - AA^{-}Ag \\ &= AA^{-}b + Ag - Ag \\ &= b. \end{aligned}$$

Hence, $x = A^{-}b + (I-A^{-}A)g$ is a solution for every g. Now let x be any solution of $Ax = b$, then

$$A^{-}Ax = A^{-}b$$
$$0 = A^{-}b - A^{-}Ax,$$

adding x to both sides of this equation,

$$x = A^{-}b - A^{-}Ax + x$$
$$x = A^{-}b + (I-A^{-}A)x$$

which is the required form with $g = x$.

Q.E.D.

Corollary A.8.1. If A is any nxm matrix, the following are true.

(a) The column spaces of A and AA^- are the same.

(b) The column spaces of A^- and A^-A are the same.

(c) The column space of $(I-AA^-)$ is the orthogonal complement of the column space of A.

(d) The column space of $(I-A^-A)$ is the orthogonal complement of the column space of A^t.

(e) The column space of $(I-A^-A)$ is the same as the null space of A.

(f) The column space of $(I-AA^-)$ is the same as the null space of A^-.

Proof of Corollary A.8.1. Proof follows directly by application of the required definitions.

Definition. An nxn matrix is said to be idempotent if AA = A.

Corollary A.8.2. $(I-A^-A)$ is idempotent.

Proof of Corollary A.8.2.

$$\begin{aligned}(I-A^-A)(I-A^-A) &= (I-A^-A) - (I-A^-A)A^-A \\ &= (I-A^-A) - A^-A + (A^-AA^-)A \\ &= (I-A^-A) - A^-A + A^-A \\ &= (I-A^-A).\end{aligned}$$

Q.E.D.

Theorem A.9. $(I-A^-A)$ is an mxm symmetric, positive semidefinite matrix.

Proof of Theorem A.9.

$$\begin{aligned}(I-A^-A)^t &= I - (A^-A)^t \\ &= I-A^-A,\end{aligned}$$

hence, $(I-A^-A)$ is symmetric and idempotent,

$$x^t(I-A^-A)x = x^t(I-A^-A)^t(I-A^-A)x \geq 0.$$

Therefore, $(I-A^-A)$ is positive semidefinite.

Q.E.D.

Theorem A.10. If A is symmetric idempotent, then $A^- = A$; i.e., if $A = A^t$ and $A = AA$, then $A^- = A$.

Proof of Theorem A.10. The proof is by showing that A satisfies the definition. That is

(i) $AA^- = AA = A$ is symmetric

(ii) $A^-A = AA = A$ is symmetric

(iii) $AA^-A = AA = A$

(iv) $A^-AA^- = AA^- = A$

Q.E.D.

Theorem A.11. Let D be a diagonal matrix, then D^- is a diagonal matrix with diagonal elements the reciprocals of the nonzero diagonal elements of D or zero if the element is zero.

Example:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad D^- = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Proof of Theorem A.11. The proof follows from showing the definition is satisfied.

Definition. Let A be an $m_2 x n_2$ matrix and let B be an $m_1 x n_1$ matrix; then the Kronecker product of A and B, which is written AxB, is a matrix C of order $m_1 m_2 x n_1 n_2$ defined by

$$C = \begin{bmatrix} A\,b_{11} & A\,b_{12} & \cdots & A\,b_{1n_1} \\ A\,b_{21} & A\,b_{22} & \cdots & A\,b_{2n_1} \\ \vdots & \vdots & & \vdots \\ A\,b_{m_1 1} & A\,b_{m_1 2} & \cdots & A\,b_{m_1 n_1} \end{bmatrix}$$

<u>Theorem A.12</u>. Let the matrix A be defined by the Kronecker product

$$A = B \times C.$$

Then the generalized inverse of A is given by

$$A^{-} = B^{-} \times C^{-}.$$

<u>Proof of Theorem A.12</u>. From the theory dealing with the Kronecker product

$$(B \times C)(D \times E) = BD \times CE$$

provided conformability requirements are met, and it can be quickly shown that $A^{-} = B^{-} \times C^{-}$ satisfies the definition.

Q.E.D.

By Theorem A.8 it is seen that for a particular linear system; e.g.,

$$Ax = b,$$

a solution can be found given the system is consistent. Obviously, not all linear systems are consistent and therefore no x can be found which satisfies the system. It is possible, though, to find approximate solutions. The following develops one approach to defining an approximate solution to a linear system that is inconsistent. For further treatment of the topic see Price [54].

<u>Definition</u>. The vector x^o is defined to be the <u>best approximate solution</u> to the system of equations (A is an mxn matrix)

$$Ax-b = e(x)$$

if and only if

(a) for all $x \in E^n$, the relationship $(Ax-b)^t(Ax-b) \geq (Ax^o-b)^t(Ax^o-b)$ holds;

(b) and for those $x \neq x^o$ such that $(Ax-b)^t(Ax-b) = (Ax^o-b)^t(Ax^o-b)$, the relationship $x^tx > x^{ot}x^o$ holds.

Based on the above definition of what will be termed the best approximate solution, the following application of the generalized inverse results.

Theorem A.13. The best approximate solution to the system of equations $Ax = b$ is x^o where

$$x^o = A^-b.$$

Proof of Theorem A.13.

$$\begin{aligned}(Ax-b)^t(Ax-b) &= (Ax-AA^-b + AA^-b-b)^t(Ax-AA^-b + AA^-b-b)\\ &= [A(x-A^-b) + (AA^--I)b]^t[A(x-A^-b) + (AA^--I)b]\\ &= [A(x-A^-b)]^t[A(x-A^-b)] + [(AA^--I)b]^t[(AA^--I)b]\\ &\geq [(AA^--I)b]^t[(AA^--I)b]\end{aligned}$$

where the cross product terms are seen to be zero. This inequality holds for all $x \in E^n$. Letting $x^o = A^-b$,

$$(Ax-b)^t(Ax-b) \geq (Ax^o-b)^t(Ax^o-b)$$

for all $x \in E^n$. Equality holds if and only if

$$[A(x-A^-b)]^t[A(x-A^-b)] = 0,$$

i.e., if and only if

$$Ax = AA^-b.$$

Now it must be shown that for all x such that $Ax = AA^-b$, the relationship $x^tx > x^{ot}x^o$ holds. For all $x \in E^n$ such that $Ax = AA^-b$,

$$A^-b = A^-Ax \text{ and } x = A^-b-A^-Ax + x,$$

hence,

$$\begin{aligned}x^tx &= (A^-b + (I-A^-A)x)^t(A^-b + (I-A^-A)x)\\ &= (A^-b)^t(A^-b) + [(I-A^-A)x]^t[(I-A^-A)x],\end{aligned}$$

or

$$x^t x = (A^- b)^t (A^- b) + (x - A^- b)^t (x - A^- b)$$

and if $x \neq x^o$, $x^o = A^- b$,

$$x^t x > x^{ot} x^o.$$

Q.E.D.

Corollary A.13.1. The best approximate solution always exists and is unique.

Proof of Corollary A.13.1. The best approximate solution is defined as $x^o = A^- b$ and by Theorems A.2 and A.3 always exists and is unique.

Q.E.D.

The following theorems and corollary are due to Cline [16,17]. Complete proofs are available in Cline's paper and will not be included here. Theorem A.15 was first constructively proved by Greville [33]. The statements here given implicitly reflect real matrices, but the theory holds for complex matrices using conjugate transpose.

Theorem A.14. For any matrices, U and V, the generalized inverse of the sum $UU^t + VV^t$ can be written in the form

$$[UU^t + VV^t]^- = (I - C^{-t} V^t) U^{-t} [I - U^- V (I - C^- C) K V^t U^{-t}] U^- (I - VC^-) + C^{-t} C^-.$$

where

$$C = (I - UU^-) V$$

and

$$K = [I + (I - C^- C) V^t U^{-t} U^- V (I - C^- C)]^{-1}.$$

Corollary A.14.1.

$$[UU^t + VV^t]^- = U^{-t} U^- - U^{-t} U^- V K V^t U^{-t} U^-$$

if and only if

$$C = 0 \text{ with } K = [I + V^t U^{-t} U^- V]^{-1}.$$

Theorem A.15. $B = [U,V]$ then

$$B^- = \left[\begin{array}{c} U^-(I-VJ) \\ \hdashline J \end{array}\right]$$

where

$$J = C^- + (I-C^-C)KV^tU^{-t}U^-(I-VC),$$

$$C = (I-UU^-)V,$$

and

$$K = \{I + [U-V(I-C^-C)]^t[U^-V(I-C^-C)]\}^{-1}.$$

One significant disadvantage in the use of the generalized inverse is the fact that $(AB)^- \neq B^-A^-$ in many cases. There are instances though when the form is applicable as defined in the following theorem. The proof has not been included, but can be found in Boullion and Odell [8] or Albert [1].

Theorem A.16. $(AB)^- = B^-A^-$ if and only if $A^-ABB^tA^t = BB^tA^t$ and $BB^-A^tAB = A^tAB$.

APPENDIX B

THE TRACE FUNCTION

A well-known function of square matrices is the trace function, tr,

$$tr : E^{nxn} \rightarrow E^1.$$

In order to facilitate the theoretical developments in Chapter 2, the following known theorems and proofs are presented.

<u>Definition</u>. The trace of an nxn matrix A, which is designated by tr(A), is defined to be the sum of the diagonal elements of A; that is

$$tr(A) = \sum_{i=1}^{n} a_{ii}$$

<u>Theorem B.1</u>. Let A and B be nxn matrices, then

$$tr(AB) = tr(BA).$$

<u>Proof of Theorem B.1</u>. Put AB = C, then

$$c_{pq} = \sum_{j=1}^{n} a_{pj} b_{jq}.$$

Now put BA = G, then

$$g_{rs} = \sum_{i=1}^{n} b_{ri} a_{is}.$$

By definition

$$tr(AB) = tr(C) = \sum_{p=1}^{n} c_{pp} = \sum_{p=1}^{n} \sum_{j=1}^{n} a_{pj} b_{jp}.$$

Also,

$$tr(BA) = tr(G) = \sum_{r=1}^{n} g_{rr} = \sum_{r=1}^{n} \sum_{i=1}^{n} b_{ri} a_{ir}.$$

Thus, it is seen that tr(AB) = tr(BA).

Q.E.D.

Another well-known theorem on the trace function is the following.

Theorem B.2. If A_j, $j = 1,2,\ldots,m$ are nxn matrices and α_j is scalar, then

$$\mathrm{tr}\,[\sum_{j=1}^{m} \alpha_j A_j] = \sum_{j=1}^{m} \alpha_j \mathrm{tr}(A_j).$$

The remaining theorems are not as well known and due to their importance proofs will be given.

Theorem B.3. Let A be an mxn matrix; then $\mathrm{tr}(A^tA) = 0$ if and only if $A = 0$, the zero matrix.

Proof of Theorem B.3. If $A = 0$, then $A^tA = 0$ and certainly $\mathrm{tr}(A^tA) = 0$. Assume now that $\mathrm{tr}(A^tA) = 0$. A^tA is a positive semidefinite matrix and by $\mathrm{tr}(A^tA) = 0$ each diagonal element is zero. By definition of the j diagonal elements of A^tA,

$$\sum_{i=1}^{n} a_{ij}^2 = 0, \text{ for } j = 1,2,\ldots,n,$$

it is clear that $a_{ij} = 0$. Hence $A = 0$.

Q.E.D.

Theorem B.4. Let A and B be nxn positive semidefinite matrices; then

$$\mathrm{tr}(AB) = 0, \text{ if and only if } AB = 0, \text{ the zero matrix.}$$

Proof of Theorem B.4. Clearly, if $AB = 0$, then $\mathrm{tr}(AB) = 0$. To prove the "only if," note that since A and B are positive semidefinite, by Lemma 2.3, there exist matrices U and V such that $A = U^tU$ and $B = VV^t$,

$$\mathrm{tr}(AB) = \mathrm{tr}(U^tUVV^t) = \mathrm{tr}(V^tU^tUV).$$

The last equality follows from Theorem B.1. By the hypothesis, $\mathrm{tr}(AB) = 0$, and $\mathrm{tr}(AB) = \mathrm{tr}[(UV)^t(UV)]$, it follows from Theorem B.3 that

$$UV = 0.$$

Premultiplying by U^t and postmultiplying by V^t results in

$$U^tUVV^t = AB = 0.$$

Q.E.D.

Corollary B.4.1. Let A and B be nxn positive semidefinite matrices; then

$$tr(AB) \geq 0.$$

Proof of Corollary B.4.1. As in the proof of Theorem B.4,

$$tr(AB) = tr[(UV)^t(UV)].$$

Let UV = C, then

$$tr(AB) = tr(C^tC) = \sum_{j=1}^{n} \sum_{i=1}^{n} c_{ij} \geq 0.$$

Q.E.D.

REFERENCES

[1] Arthur Albert, Regression and the Moore-Penrose Pseudoinverse (Academic Press, New York 1972).

[2] D. P. Baron, "Quadratic Programming with Quadratic Constraints," Naval Research Logistics Quarterly 19(1972)253-260.

[3] R. E. Bellman, Introduction to Matrix Analysis (McGraw-Hill, New York 1960).

[4] A. Ben-Israel, "On Error Bounds for the Generalized Inverse," SIAM Journal of Numerical Analysis 3(1966)585-592.

[5] A. Ben-Israel and A. Charnes, "Contributions to the Theory of Generalized Inverses," SIAM Journal 11(1963)667-699.

[6] A. Ben-Israel and S. J. Wersan, "An Elimination Method for Computing the Generalized Inverse of an Arbitrary Complex Matrix," Journal of the Association for Computing Machinery 10(1963)532-537.

[7] E. Bodewig, Matrix Calculus (North-Holland Publishing Company, Amsterdam 1959).

[8] T. L. Boullion and P. L. Odell, Generalized Inverse Matrices (Wiley-Interscience, New York 1971).

[9] A. V. Cabot and R. L. Francis, "Properties of a Minimax Location Problem Involving Euclidean Distances," Naval Research Logistics Quarterly 19(1972)335-353.

[10] A. V. Cabot, R. L. Francis, and M. A. Stary, "A Network Flow Solution to a Rectilinear Distance Facility Location Problem," AIIE Transactions 2(1970)132-141.

[11] A. Charnes and W. W. Cooper, "Chance-Constrained Programming," Management Science 6(1959)73-79.

[12] A. Charnes, W. W. Cooper, and A. Henderson, An Introduction to Linear Programming (John Wiley and Sons, New York 1953).

[13] A. Charnes, W. W. Cooper, and G. L. Thompson, "Constrained Generalized Medians and Hypermedians as Deterministic Equivalents for Two Stage Linear Programs Under Uncertainty," Management Science 12(1965)83-112.

[14] A. Charnes and M. Kirby, "Modular Design, Generalized Inverse, and Convex Programming," Operations Research 13(1965)836-847.

[15] R. J. Clasen, "Numerical Methods for Inverting Positive Definite Matrices," Memorandum RM-4952-PR, The RAND Corporation, Santa Monica, California, 1966.

[16] R. E. Cline, "Representations for the Generalized Inverse of a Partitioned Matrix," SIAM Journal 12(1964)588-600.

[17] R. E. Cline, "Representations for the Generalized Inverse of Sums of Matrices," SIAM Journal of Numerical Analysis 2(1965)99-114.

[18] R. E. Cline, "Representations for the Generalized Inverse of Matrices with Applications in Linear Programming," Dissertation, Purdue University, Lafayette, Indiana, 1963.

[19] R. W. Cottle and J. A. Ferland, "Matrix-Theoretic Criteria for the Quasi-Convexity and Pseudo-Convexity of Quadratic Functions," Technical Report No. 70-6, Operations Research House, Stanford University, California, 1970.

[20] G. B. Dantzig, Linear Programming and Extensions (Princeton University Press, Princeton, New Jersey 1963).

[21] P. M. Dearing, "On Some Minimax Location Problems Using Rectilinear Distance," Dissertation, University of Florida, Gainesville, 1972.

[22] P. M. Dearing and R. L. Francis, "A Network Flow Solution to a Multifacility Location Problem Using Rectilinear Distances," presented at the Joint ORSA-TIMS-AIIE National Meeting, Atlantic City, New Jersey, 1972.

[23] H. P. Decell, Jr., "An Application of the Cayley-Hamilton Theorem to Generalized Matrix Inversion," SIAM Review 7(1965)526-528.

[24] J. G. Ecker and R. D. Niemi, "A Dual Method for Quadratic Programs with Quadratic Constraints," presented at the VIII International Symposium on Mathematical Programming, Stanford University, California, 1973.

[25] J. Elzinga and D. W. Hearn, "Minimax Multifacility Locations with Euclidean Distances," presented at the 41st National Meeting of the Operations Research Society of America, New Orleans, Louisiana 1972.

[26] J. E. Falk, "Lagrange Multipliers and Nonlinear Programming," Journal of Mathematical Analysis and Applications 19(1967)141-159.

[27] W. Fenchel, "Convex Cones, Sets and Functions," Lecture Notes (Princeton University, 1953).

[28] A. V. Fiacco and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Technique (John Wiley and Sons, New York 1968).

[29] R. L. Francis, "A Note on the Optimum Location of New Machines in Existing Plant Layouts," The Journal of Industrial Engineering 14(1963)57-59.

[30] R. L. Francis, "On the Location of Multiple New Facilities with Respect to Existing Facilities," The Journal of Industrial Engineering 15(1964)106-107.

[31] R. L. Francis, "Some Aspects of a Minimax Location Problem," Operations Research 15(1967)1163-1169.

[32] F. A. Graybill, Introduction to Matrices with Applications in Statistics (Wadsworth Publishing Company, Inc., Belmont, California 1969).

[33] T. N. Greville, "Some Applications of the Pseudoinverse of a Matrix," SIAM Review 2(1960)15-22.

[34] D. W. Hearn, "Minimum Covering Spheres," Dissertation, The Johns Hopkins University, 1971.

[35] D. W. Hearn and W. D. Randolph, "Computational Aspects of a Dual Algorithm for Quadratically Constrained Quadratic Programs," in Proceedings of the 1973 Annual Conference of the Association for Computing Machinery (Association for Computing Machinery, New York 1973)48-50.

[36] M. R. Hestenes, "Inversion of Matrices by Biorthogonalization and Related Topics," SIAM Journal 6(1958)51-90.

[37] F. John, "Extremum Problems with Inequalities as Subsidiary Conditions," in Studies and Essays, Courant Anniversary Volume (Interscience Publishers, Inc., New York 1948)pp. 187-204.

[38] S. Karamardian, "Strictly Quasi-Convex (Concave) Functions and Duality in Mathematical Programming," Journal of Mathematical Analysis and Applications 20(1967)344-358.

[39] H. W. Kuhn, "Location Problems and Mathematical Programming," in Separatum-Colloquium on the Application of Mathematics to Economics, Budapest, 1963(Publishing House of the Hungarian Academy of Sciences, Budapest 1965).

[40] H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Ed. Jerzy Neyman (University of California Press, 1950) pp. 481-492.

[41] C. E. Lemke, "A Method of Solution for Quadratic Programs," Management Science 8(1962)442-453.

[42] D. G. Luenberger, Introduction to Linear and Nonlinear Programming (Addison-Wesley, Reading, Mass., 1973).

[43] O. L. Mangasarian, Nonlinear Programming (McGraw-Hill, New York 1969).

[44] C. T. Mao and C. E. Särndal, "A Decision Theory Approach to Portfolio Selection," Management Science 12(1966)B-323-B-333.

[45] H. M. Markowitz, Portfolio Selection, Efficient Diversification of Investments (John Wiley and Sons, New York 1959).

[46] B. Martos, "Subdefinite Matrices and Quadratic Forms," SIAM Journal of Applied Mathematics 17(1969)1215-1223.

[47] E. H. Moore, "On the Reciprocal of the General Algebraic Matrix," Bulletin of the American Mathematical Society 26(1919-1920)394-395.

[48] N. C. Mylander, R. L. Holmes, and G. P. McCormick, A Guide to SUMT-Version 4: The Computer Program Implementing the Sequential Unconstrained Minimization Technique for Nonlinear Programming, RAC-P-63 (Research Analysis Corporation, McLean, Virginia 1971).

[49] B. Noble, "A Method for Computing the Generalized Inverse of a Matrix," SIAM Journal of Numerical Analysis 3(1966)582-584.

[50] R. Penrose, "A Generalized Inverse for Matrices," Proceedings of the Cambridge Philosophical Society 51(1955)406-413.

[51] E. L. Peterson and J. G. Ecker, "Geometric Programming: Duality in Quadratic Programming and ℓ_p-Approximation I," in Proceedings of the Princeton Symposium on Mathematical Programming, Ed. H. W. Kuhn (Princeton University Press, Princeton, New Jersey 1970).

[52] E. L. Peterson and J. G. Ecker, "Geometric Programming: Duality in Quadratic Programming and ℓ_p-Approximation II (Canonical Programs)," SIAM Journal of Applied Mathematics 17(1969)317-340.

[53] E. L. Peterson and J. G. Ecker, "Geometric Programming: Duality in Quadratic Programming and ℓ_p-Approximation III (Degenerate Programs)," Journal of Mathematical Analysis and Applications 29(1970)365-383.

[54] C. M. Price, "The Matrix Pseudoinverse and Minimal Variance Estimates," SIAM Review 6(1964)115-120.

[55] L. D. Pyle, "The Generalized Inverse in Linear Programming," Dissertation, Purdue University, Lafayette, Indiana, 1960.

[56] L. D. Pyle, "Generalized Inverse Computations Using the Gradient Projection Method," Journal of the Association for Computing Machinery 11(1964)422-429.

[57] R. T. Rockafellar, "Some Convex Programs Whose Duals Are Linearly Constrained," in Nonlinear Programming, Ed. J. Abadie (Academic Press, Inc., New York 1970)pp. 293-322.

[58] R. T. Rockafellar, Convex Analysis (Princeton University Press, 1970).

[59] J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming, Part I. Linear Constraints," SIAM Journal of Applied Mathematics 8(1960)181-217.

[60] J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming, Part II. Nonlinear Constraints," SIAM Journal of Applied Mathematics 9(1961)514-532.

[61] A. D. Roy, "Safety First in the Holding of Assets," Econometrica 20(1952)431-449.

[62] G. Rust, W. R. Burrus, and C. Schneeberger, "A Simple Algorithm for Computing the Generalized Inverse of a Matrix," Communications of the Association for Computing Machinery 9(1966)381-387.

[63] T. L. Saaty and J. Bram, Nonlinear Mathematics (McGraw-Hill, New York 1964).

[64] W. F. Sharpe, "A Simplified Model for Portfolio Analysis," Management Science 9(1963)277-293.

[65] S. M. Sinha, "A Duality Theorem for Nonlinear Programming," Management Science 12(1966)385-390.

[66] G. W. Stewart, "On the Continuity of the Generalized Inverse," SIAM Journal of Applied Mathematics 17(1969)33-45.

[67] R. P. Tewarson, "A Direct Method for Generalized Matrix Inversion," SIAM Journal of Numerical Analysis 4(1967)499-507.

[68] D. M. Topkis and A. F. Veinott, "On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming," SIAM Journal of Control 5(1967)268-279.

[69] C. van de Panne, "Programming with a Quadratic Constraint," Management Science 12(1966)798-815.

[70] J. R. Westlake, A Handbook of Numerical Matrix Inversion and Solution of Linear Equations (John Wiley and Sons, New York, 1968).

[71] P. Wolfe, "A Duality Theorem for Nonlinear Programming," Quarterly of Applied Mathematics 19(1961)239-244.

[72] G. Zoutendijk, Methods of Feasible Directions (Elsevier Publishing Company, Amsterdam 1960).

BIOGRAPHICAL SKETCH

William David Randolph was born January 24, 1941 in Eddyville, Illinois and received his primary and secondary education in Indiana, graduating in May 1959 from Edinburg High School. He then entered Rose Polytechnic Institute (now Rose-Hulman Institute of Technology), Terre Haute, Indiana graduating with a B.S. in Mathematics in June, 1963. For the next seven years he worked in the Washington, D. C. metropolitan area as an operations analyst/researcher for the Johns Hopkins University/ Applied Physics Laboratory, the John D. Kettelle Corporation and the Northrop Corporation. During this period of time he continued his education with graduate courses at the American University and the George Washington University, earning an M.S. in Operations Research from the latter in September, 1970. He enrolled in the Graduate School of the University of Florida, Industrial and Systems Engineering Department in September, 1970 and was awarded an M.E. in Systems Engineering in June, 1972.

He is married to the former Alicia Lee Gehringer.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

D. W. Hearn, Chairman
Assistant Professor of Industrial and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

M. E. Thomas
Professor of Industrial and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

R. L. Francis
Professor of Industrial and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

H. D. Ratliff
Associate Professor of Industrial and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

N. Passell
Assistant Professor of Mathematics

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

June, 1974

Dean, College of Engineering

Dean, Graduate School