To the Graduate Council:

I am submitting herewith a thesis written by James R. Williams entitled "The Convex Quadratic Programming Problem." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mathematics.

Yueh-er Kuo

Yueh-er Kuo, Major Professor

We have read this thesis and recommend its acceptance:

Jeremy Haefner

[illegible]

Accepted for the Council:

[illegible]

Vice Provost
and Dean of The Graduate School

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at The University of Tennessee, Knoxville, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of the source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in her absence, by the Head of Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature James R Williams

Date 4-25-89

THE CONVEX QUADRATIC PROGRAMMING PROBLEM

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

James R. Williams

May 1989

ACKNOWLEDGMENTS

The author is extremely grateful to Dr. Yueh-er Kuo for her generous help and guidance in preparing this thesis.

Appreciation is also expressed to Dr. Balram S. Rajput and Dr. Jeremy Haefner for their time reading this thesis.

ABSTRACT

The purpose of this paper is to investigate the convex quadratic programming problem and its applications. The emphasis is on methods for solving the convex quadratic programming problem.

Chapter I discusses the general background of the problem and developes necessary and sufficient conditions for a solution of the problem.

Chapter II presents Wolfe's algorithm for solving convex quadratic programs. Chapter III investigates Lemke's algorithm for solving this problem.

Chapter IV discusses a recent barrier function method for solving the convex quadratic programming problem. Chapter V introduces an active set method for solving the problem.

Chapter VI discusses applications of quadratic programming in optimal portfolio analysis and in finding the optimal search direction in a general mathematical programming problem.

TABLE OF CONTENTS

CHAPTER I

CONVEX FUNCTIONS AND THE KUHN-TUCKER CONDITIONS

We will consider the problem

$$\text{minimize} \quad \underline{c}^T\underline{x} + \frac{1}{2}\,\underline{x}^T H\,\underline{x} \tag{1.1}$$

$$\text{subject to:} \quad A\underline{x} \leq \underline{b}$$
$$\underline{x} \geq 0$$

where

$\underline{c}$ is an n-vector,

$\underline{b}$ is an m-vector,

A is an $m \times n$ matrix, and

H is an $n \times n$ symmetric matrix.

This problem is known as the quadratic programming problem. Note that for a function

$$f(\underline{x}) = \Sigma_{i=1}^{n}\,\Sigma_{j=i}^{n}\; h_{ij}\, x_i x_j + \Sigma_{i=1}^{n}\, c_i x_i$$

H is simply the hessian of $f(\underline{x})$. Also note that maximizing $f(\underline{x})$ is the same as minimizing $-f(\underline{x})$ so we consider only the minimization problem.

Throughout this paper capital letters will denote real matrices. A^T will denote the transpose of a real matrix. $\underline{x}$ will denote a real column vector, that is

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} .$$

E_n will denote Euclidean n-space.

The algorithms used to solve nonlinear programming problems normally search for local minima. One of the great difficulties in nonlinear programming is determining whether such a local minimum is also a global minimum. In the first part of this chapter, we show that a local minimum of $f(\underline{x}) = \underline{c}^T\underline{x} + \frac{1}{2}\underline{x}^T H\underline{x}$ is also a global minimum if H is positive semidefinite.

<u>Definition 1.1</u>. A set K is said to be <u>convex</u> if for $\underline{x}_1, \underline{x}_2 \in K$, $\lambda\underline{x}_1 + (1 - \lambda)\underline{x}_2 \in K$ for all $\lambda \in [0, 1]$.

Geometrically, this means that a line segment connecting any two points of a convex set is contained in the set. The set of feasible solutions to Problem 1.1 is an example of a convex set. To see this, suppose that $\underline{x}_1$ and $\underline{x}_2$ are feasible solutions to $A\underline{x} \leq \underline{b}$ and $\underline{x} \geq 0$. Also, let $\lambda \in [0, 1]$. Then $\lambda\underline{x}_1 + (1 - \lambda)\underline{x}_2 \geq 0$ since λ, $1 - \lambda$, $\underline{x}_1$ and $\underline{x}_2$ are all non-negative. In addition, $\lambda A \underline{x}_1 + (1 - \lambda) A \underline{x}_2 \leq \lambda b + (1 - \lambda)b = b$.

Definition 1.2. A function $f(\underline{x})$ defined over a convex set K is called a convex function if for $\underline{x}_1, \underline{x}_2 \in K$ and $\lambda \in [0, 1]$ we have

$$f[\lambda\underline{x}_1 + (1 - \lambda)\underline{x}_2] \leq \lambda f(\underline{x}_1) + (1 - \lambda) f(\underline{x}_2) .$$

If strict inequality holds for distinct $\underline{x}_1$ and $\underline{x}_2$ and $\lambda \in (0, 1)$ then $f(\underline{x})$ is said to be strictly convex. Similarly, $f(\underline{x})$ is said to be concave if $-f(\underline{x})$ is convex.

Basically this means that a function is convex if for all $\underline{x}_1, \underline{x}_2 \in K$ the line segment from $f(\underline{x}_1)$ to $f(\underline{x}_2)$ lies on or above the graph of $f(\underline{x})$. $f(\underline{x})$ is strictly convex if the line segment from $f(\underline{x}_1)$ to $f(\underline{x}_2)$ lies strictly above the graph of $f(x)$ except at its endpoints.

Theorem 1.1 [1] . Let S be a nonempty convex set in E_n and $f: S \to E_1$. Consider the problem to minimize $f(\underline{x})$ subject to $\underline{x} \in S$. Suppose that $\underline{y} \in S$ is a local minimum to the problem.

1. If f is convex, then $\underline{y}$ is a global minimum.
2. If f is strictly convex, then $\underline{y}$ is the unique global optimal solution.

Definition 1.3. A matrix H is said to be positive definite if $\underline{x}^T H\underline{x} > 0$ for all $\underline{x} \neq 0$. It is said to be positive semidefinite if $\underline{x}^T H\underline{x} \geq 0$ for all $\underline{x}$.

There are several ways of testing whether a matrix is positive definite. It is well known that a matrix is positive definite (semidefinite) if and only if its eigenvalues are positive (non-negative). Also, H is positive definite if and only if its principal minors are all strictly positive. For $n = 4$ this means that

$$|h_{11}| > 0 \qquad \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} > 0 \qquad \begin{vmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{vmatrix} > 0 \qquad |H| > 0$$

<u>Theorem 1.2 [1]</u>. Let S be a non-empty open convex set in E_n and let $f\colon S \to E_1$ be twice differentiable on S. Then f is convex if and only if the Hessian matrix is positive semidefinite at each point in S.

<u>Corollary 1.1</u>. $f(\underline{x}) = \frac{1}{2}\,\underline{x}^T H\underline{x} + \underline{c}^T\underline{x}$ is convex if and only if H is positive semidefinite.

<u>Proof</u>. $\nabla f(\underline{x}) = \underline{c} + H\underline{x}$.

$$\nabla^2 f(\underline{x}) = H .$$

The result then follows from the theorem.

<u>Theorem 1.3</u>. The function $f(\underline{x}) = \underline{c}^T\underline{x} + \frac{1}{2}\,\underline{x}^T H\underline{x}$ is strictly convex if and only if H is positive definite.

<u>Proof</u>. [1] proves that a function is strictly convex if its hessian is positive definite.

Conversely, suppose that $f(\underline{x})$ is strictly convex. Let $\underline{x} \in E_n$, $\underline{x} \neq 0$. Then

$$f(\tfrac{1}{2}\underline{x} + \tfrac{1}{2}(0)) < \tfrac{1}{2} f(\underline{x}) + \tfrac{1}{2} f(0) .$$

Therefore

$$\underline{c}^T(\tfrac{1}{2}\underline{x}) + \tfrac{1}{2}(\tfrac{1}{2}\underline{x})^T H(\tfrac{1}{2}\underline{x}) < \tfrac{1}{2}\underline{c}^T\underline{x} + \tfrac{1}{4}\underline{x}^T H\underline{x}$$

and

$$\tfrac{1}{8}\underline{x}^T H\underline{x} < \tfrac{1}{4}\underline{x}^T H\underline{x} .$$

This implies that $\underline{x}^T H\underline{x} > 0$ for all $\underline{x} \neq 0$. Therefore H is positive definite.

Example 1.1. Consider the function

$$f(\underline{x}) = 2x_1 - 3x_2 + x_1^2 + x_2^2 - x_1x_2 .$$

We can write

$$f(\underline{x}) = \underline{c}^T\underline{x} + \tfrac{1}{2}\underline{x}^T H\underline{x}$$

where

$$c = \begin{bmatrix} 2 \\ -3 \end{bmatrix} \text{ and } H = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} .$$

To show that $f(\underline{x})$ is strictly convex, we find the eigenvalues of H. The characteristic polynomial is

$$f(t) = \begin{vmatrix} t - 2 & 1 \\ 1 & t - 2 \end{vmatrix} = t^2 - 4t + 3 .$$

The eigenvalues are $t_1 = 3$ and $t_2 = 1$. These are positive so H is positive definite and $f(\underline{x})$ is strictly convex.

Unfortunately, the algorithms we will use to solve the quadratic programming problem only work if the objective function is convex (meaning H is positive semidefinite). From this point on, we assume that this is the case. With this additional restriction, problem (1.1) is known as the convex quadratic programming problem.

We now turn our attention to developing necessary and sufficient conditions for a vector to solve the quadratic programming problem. These conditions are the well-known Kuhn-Tucker conditions. We will prove a specialized version of these conditions that applies to the quadratic programming problem and then state without proof a more general result.

Lemma 1.1 [8]. $\underline{x}_0$ minimizes a convex continuously differentiable function over a convex set K if and only if $(\underline{x} - \underline{x}_0)^T \nabla f(\underline{x}_0) \geq 0$ for all $\underline{x} \in K$.

Theorem 1.4 [2] (The theorem of the separating hyperplane). For any $m \times n$ matrix A and any m-vector $\underline{b}$ either

$$A\underline{x} = \underline{b} \ , \ \underline{x} \geq 0$$

has a solution or

$$A^T\underline{y} \geq 0 \ , \ \underline{b}^T\underline{y} < 0$$

has a solution, but not both.

Geometrically, the theorem says that either $\underline{b}$ is a vector which lies within the cone spanned by the non-negative linear combinations of the columns of A ; or else there exists a vector $\underline{y}$ which makes non-obtuse angles with the columns of A but an obtuse angle with $\underline{b}$.

Theorem 1.5 [2] (Kuhn-Tucker). A vector $\underline{x}$ solves the convex quadratic programming problem if and only if there exist vectors $\underline{y}$, $\underline{u}$, and $\underline{v}$ such that:

$$A\underline{x} + \underline{y} = \underline{b} \tag{1.2}$$

$$\underline{v} = H\underline{x} + A^T\underline{u} + \underline{c} \tag{1.3}$$

$$\underline{x}^T\underline{v} = 0 \ , \ \underline{u}^T\underline{y} = 0 \tag{1.4}$$

$$\underline{x}, \underline{y}, \underline{u}, \underline{v} \geq 0 \tag{1.5}$$

Proof. Suppose that vectors $\underline{x}$, $\underline{u}$, $\underline{v}$, and $\underline{y}$ which satisfy the above conditions have been found. Also, let $\hat{\underline{x}}$ be another feasible solution. Then

$$\begin{aligned} f(\underline{x}) - f(\hat{\underline{x}}) &= \underline{c}^T\underline{x} + \frac{1}{2}\underline{x}^T H\underline{x} - \underline{c}^T\hat{\underline{x}} - \frac{1}{2}\hat{\underline{x}}^T H\hat{\underline{x}} \\ &= \underline{c}^T(\underline{x} - \hat{\underline{x}}) - \frac{1}{2}(\underline{x} - \hat{\underline{x}})^T H(\underline{x} - \hat{\underline{x}}) + \underline{x}^T H\underline{x} - \underline{x}^T H\hat{\underline{x}} \end{aligned}$$

$$\le (\underline{c}^T + \underline{x}^T H)(\underline{x} - \hat{\underline{x}}) \qquad \text{since } \underline{x}^T H\underline{x} \ge 0 \text{ for all } \underline{x} .$$

$$= (\underline{v} - A^T\underline{u})^T (\underline{x} - \hat{\underline{x}}) \qquad \text{by (1.3)}$$

$$\le -\underline{u}^T A(\underline{x} - \hat{\underline{x}}) \qquad \text{since } \underline{x}^T\underline{v} = 0 \text{ and } \hat{\underline{x}}^T\underline{v} \ge 0 .$$

$$= -\underline{u}^T[\underline{b} - \underline{y} - \underline{b} + \hat{\underline{y}}]$$

$$= -\underline{u}^T \hat{\underline{y}} \qquad \text{since } \underline{u}^T\underline{y} = 0$$

$$\le 0 \qquad \text{since } \underline{u} \ge 0 \text{ and } \hat{\underline{y}} \ge 0 .$$

Therefore,

$$f(\underline{x}) \le f(\hat{\underline{x}}) .$$

Thus $\underline{x}$ is the optimal solution to the problem. This proves the sufficiency of the conditions.

Conversely, suppose that $\hat{\underline{x}}$ minimizes

$$f(\underline{x}) = \underline{c}^T\underline{x} + \frac{1}{2}\underline{x}^T H\underline{x}$$

subject to

$$A\underline{x} \le \underline{b} \text{ and } \underline{x} \ge 0 .$$

Then, we must obviously have

$$A\hat{\underline{x}} + \hat{\underline{y}} = \underline{b} , \hat{\underline{x}} \ge 0 , \hat{\underline{y}} \ge 0 . \tag{1.6}$$

We want to prove that when $\hat{\underline{x}}$ is the solution vector, an n-vector $\hat{\underline{v}}$ and an m-vector $\hat{\underline{u}}$ exist such that

$$\hat{\underline{v}} = H\hat{\underline{x}} + \underline{A}^T\hat{\underline{u}} + \underline{c} \geq 0$$

$$\hat{\underline{x}}^T\hat{\underline{v}} = 0 \ , \ \hat{\underline{u}}^T\hat{\underline{y}} = 0$$

Now let us partition the non-negativity constraints into two groups indexed by L and M and the inequality constraints into two groups indexed by H and J depending on whether strict equality or strict inequality holds. We also renumber the variables so that all the zero elements of $\hat{\underline{x}}$ come first and reorder the inequalities so that those which strictly hold appear on top. Then we have

$$\hat{\underline{x}} = [\hat{\underline{x}}_L \ \hat{\underline{x}}_M]^T \quad \text{where} \quad \hat{\underline{x}}_L = 0 \quad \text{and} \quad \hat{\underline{x}}_M > 0 \ ,$$

$$A_H\hat{\underline{x}} = \underline{b}_H \ , \quad \text{and}$$

$$A_J\hat{\underline{x}} < \underline{b}_J \ .$$

Clearly, $L \cap M = \emptyset$, $L \cup M = \{1,\ldots, n\}$, $H \cap J = \emptyset$, and $H \cup J = \{1,\ldots, m\}$. We also have $\hat{\underline{y}}_H = 0$ and $\hat{\underline{y}}_J > 0$ for the slack variable $\hat{\underline{y}}$.

We may now partition the matrices $\underline{c}$, H, and A as

$$\underline{c} = \begin{bmatrix} c_L \\ c_M \end{bmatrix} \quad H = \begin{bmatrix} H_{LL} & H_{LM} \\ H_{ML} & H_{MM} \end{bmatrix} \quad A = \begin{bmatrix} A_{HL} & A_{HM} \\ A_{JL} & A_{JM} \end{bmatrix} = \begin{bmatrix} A_H \\ A_J \end{bmatrix} \ .$$

When we introduce the vectors $\hat{\underline{v}}$ and $\hat{\underline{u}}$, we will partition them in a similar manner. Note that we need the vectors $\hat{\underline{v}}$ and $\hat{\underline{u}}$ to satisfy

$$\hat{\underline{x}}_L^T \, \hat{\underline{v}}_L + \hat{\underline{x}}_M^T \, \hat{\underline{v}}_M = 0$$

and

$$\hat{\underline{y}}_H^T \, \hat{\underline{u}}_H + \hat{\underline{y}}_J^T \, \hat{\underline{u}}_J = 0$$

Since $\hat{\underline{x}}_L = 0$ and $\hat{\underline{y}}_H = 0$, this becomes the requirement that $\hat{\underline{v}}_M = 0$ and $\hat{\underline{u}}_J = 0$.

Now consider an infinitesimal change $\underline{\partial x}$ to the vector $\hat{\underline{x}}$ such that $\hat{\underline{x}} + \underline{\partial x}$ results. We consider only changes such that

$$\underline{\partial x}_L \geq 0 \tag{1.7}$$

and

$$A_H \, \underline{\partial x} \leq 0 \quad \text{or} \quad -A_H \, \underline{\partial x} \geq 0 \; .$$

This insures that $\hat{\underline{x}} + \underline{\partial x}$ is a feasible solution since

$$A(\hat{\underline{x}} + \underline{\partial x}) = \begin{bmatrix} A_{HL} & A_{HM} \\ A_{JL} & A_{JM} \end{bmatrix} \begin{bmatrix} \hat{\underline{x}}_L + \underline{\partial x}_L \\ \hat{\underline{x}}_M + \underline{\partial x}_M \end{bmatrix}$$

$$= \begin{bmatrix} A_{HL}\,\hat{\underline{x}}_L + A_{HM}\,\hat{\underline{x}}_M + A_{HL}\,\underline{\partial x}_L + A_{HM}\,\underline{\partial x}_M \\ \\ A_{JL}\,\hat{\underline{x}}_L + A_{JM}\,\hat{\underline{x}}_M + A_{JL}\,\underline{\partial x}_L + A_{JM}\,\underline{\partial x}_M \end{bmatrix}$$

$$= \begin{bmatrix} A_{HM}\,\hat{\underline{x}}_M + A_H\,\underline{\partial x} \\ \\ A_{JM}\,\hat{\underline{x}}_M + A_J\,\underline{\partial x} \end{bmatrix}$$

since $\hat{\underline{x}}_L = 0$. Now, $A_{HM}\,\hat{\underline{x}}_M = \underline{b}_H$ and $A_H\,\underline{\partial x} \leq 0$. Also, $A_{JM}\,\hat{\underline{x}}_M < \underline{b}_H$ and $A_J\,\underline{\partial x}$ is very small so

$$\begin{bmatrix} A_{HM}\,\hat{\underline{x}}_M + A_H\,\underline{\partial x} \\ \\ A_{JM}\,\hat{\underline{x}}_M + A_J\,\underline{\partial x} \end{bmatrix} \leq \begin{bmatrix} \underline{b}_H \\ \\ \underline{b}_J \end{bmatrix}$$

and $\hat{\underline{x}} + \underline{\partial x}$ is feasible.

Also, $\nabla f(\hat{\underline{x}}) = \underline{c} + H\hat{\underline{x}}$ and it follows from Lemma 1.1. that

$$(\underline{c} + H\hat{\underline{x}})^T\,\underline{\partial x} \geq 0 \ . \tag{1.8}$$

We write (1.7) and (1.8) using partitioned matrices and see that there is no solution to the system

$$\begin{bmatrix} I & 0 \\ \\ -A_{HL} & -A_{HM} \end{bmatrix} \begin{bmatrix} \underline{\partial x}_L \\ \\ \underline{\partial x}_M \end{bmatrix} \geq 0$$

and

$$\left(\begin{bmatrix} \underline{c}_L \\ \underline{c}_M \end{bmatrix} + \begin{bmatrix} H_{LL} & H_{LM} \\ H_{ML} & H_{MM} \end{bmatrix} \begin{bmatrix} \hat{\underline{x}}_L \\ \hat{\underline{x}}_M \end{bmatrix} \right)^T \begin{bmatrix} \partial x_L \\ \partial x_M \end{bmatrix} < 0 \ .$$

Hence, by the theorem of the separating hyperplane, there exist non-negative vectors $\hat{\underline{v}}_L$ and $\hat{\underline{u}}_H$ such that

$$\begin{bmatrix} I & -A_{HL}^T \\ 0 & -A_{HM}^T \end{bmatrix} \begin{bmatrix} \hat{\underline{v}}_L \\ \hat{\underline{u}}_H \end{bmatrix} = \begin{bmatrix} \underline{c}_L \\ \underline{c}_M \end{bmatrix} + \begin{bmatrix} H_{LL} & H_{LM} \\ H_{ML} & H_{MM} \end{bmatrix} \begin{bmatrix} \hat{\underline{x}}_L \\ \hat{\underline{x}}_M \end{bmatrix}$$

or

$$\hat{\underline{v}}_L - A_{HL}^T \hat{\underline{u}}_H = \underline{c}_L + H_{LL} \hat{\underline{x}}_L + H_{LM} \hat{\underline{x}}_M \tag{1.9}$$

$$- A_{HM}^T \hat{\underline{u}}_H = \underline{c}_M + H_{ML} \hat{\underline{x}}_L + H_{MM} \hat{\underline{x}}_M \ . \tag{1.10}$$

We now add $\hat{\underline{v}}_M = 0$ to the left hand side of (1.10). Also, let $\hat{\underline{u}}_J = 0$ and subtract from the left hand side of (1.9) and (1.10), respectively, the vectors $A_{JL}^T \hat{\underline{u}}_J$ and $A_{JM}^T \hat{\underline{u}}_J$. We then have

$$\hat{\underline{v}}_L - A_{HL}^T \hat{\underline{u}}_H - A_{JL}^T \hat{\underline{u}}_J = \underline{c}_L + H_{LL} \hat{\underline{x}}_L + H_{LM} \hat{\underline{x}}_M$$

$$\hat{\underline{v}}_M - A_{HM}^T \hat{\underline{u}}_H - A_{JM}^T \hat{\underline{u}}_J = \underline{c}_M + H_{ML} \hat{\underline{x}}_L + H_{MM} \hat{\underline{x}}_M$$

or

$$\underline{\hat{v}} - A^T\underline{\hat{u}} = \underline{c} + H\underline{\hat{x}} \ . \tag{1.11}$$

We also have that $\underline{\hat{x}}_L = 0$, $\underline{\hat{y}}_H = 0$, $\underline{\hat{v}}_M = 0$, and $\underline{\hat{u}}_J = 0$ so that

$$\underline{\hat{x}}^T \underline{\hat{v}} = 0 \quad \text{and} \quad \underline{\hat{u}}^T \underline{\hat{y}} = 0 \tag{1.12}$$

Also recall that $\underline{\hat{u}}_H$ and $\underline{\hat{v}}_L$ are non-negative so

$$\underline{\hat{u}}, \underline{\hat{v}} \geq 0 \tag{1.13}$$

Therefore, given the optimal solution $\underline{\hat{x}}$, there must exist vectors $\underline{\hat{y}}$, $\underline{\hat{v}}$, and $\underline{\hat{u}}$ which satisfy (1.6), (1.11), (1.12) and (1.13). This is what we needed to show to establish the necessity of the conditions in the theorem and so this completes the proof.

The vectors $\underline{u}$ and $\underline{v}$ are called the Lagrangian multiplier vectors associated with $A\underline{x} \leq \underline{b}$ and $x \geq 0$, respectively. The conditions $\underline{x}^T \underline{v} = 0$ and $\underline{u}^T\underline{y} = 0$ are called the complementary slackness constraints.

The previous theorem is actually a special case of a more general result which we present below without proof. It should also be noted that it is possible to relax the convexity and differentiability restrictions in the two theorems below (refer to [1]) .

Theorem 1.6 [1] (Kuhn-Tucker Necessary Conditions). Let X be a nonempty open set in E_n, and let $f: E_n \to E_1$, $g_i: E_n \to E_1$ for $i = 1,\ldots, m$, and $h_i: E_n \to E_1$ for $i = 1,\ldots, \ell$. Consider the problem

$$\begin{array}{lll} \text{Minimize} & f(\underline{x}) & \\ \text{subject to} & g_i(\underline{x}) \leq 0 \quad \text{for} \quad i = 1,\ldots, m & \\ & h_i(\underline{x}) = 0 \quad \text{for} \quad i = 1,\ldots, \ell & \\ & \underline{x} \in X\ . & \end{array} \tag{1.14}$$

Let $\hat{\underline{x}}$ be a feasible solution and suppose that f and g_i for $i = 1,\ldots, m$ are differentiable at $\hat{\underline{x}}$ and that h_i for $i = 1,\ldots, \ell$ are continuously differentiable at $\hat{\underline{x}}$. Also suppose that $\nabla g_i(\hat{\underline{x}})$ for $i \in \{i: g_i(\hat{\underline{x}}) = 0\}$ and $\nabla h_i(\hat{\underline{x}})$ for $i = 1,\ldots, \ell$ are linearly independent. Then, if $\hat{\underline{x}}$ solves problem (1.14) locally, there exist scalars u_i for $i = 1,\ldots, m$ and v_i for $i = 1,\ldots, \ell$ such that

$$\nabla f(\hat{\underline{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\hat{\underline{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\hat{\underline{x}}) = 0 \tag{1.15}$$

$$u_i\, g_i(\hat{\underline{x}}) = 0 \quad \text{for} \quad i = 1,\ldots, m\ . \tag{1.16}$$

$$u_i \geq 0 \quad \text{for} \quad i = 1,\ldots, m\ . \tag{1.17}$$

Theorem 1.7 [1] (Kuhn-Tucker Sufficient Conditions). Let X be a nonempty open set in E_n, and let $f: E_n \to E_1$, $g_i: E_n \to E_1$ for

$i = 1,\ldots,m$ and $h_i\colon E_n \to E_1$ for $i = 1,\ldots,\ell$. Suppose that $\hat{\underline{x}}$ is a feasible solution to (1.14). Also suppose that f, g_i for $i \in \{i\colon g_i(\hat{\underline{x}}) = 0\}$, and h_i for $i = 1,\ldots,\ell$ are convex. If there exist scalars u_i for $i = 1,\ldots,m$ and v_i for $i = 1,\ldots,\ell$ such that $\hat{\underline{x}}$ satisfies the Kuhn-Tucker conditions (1.15), (1.16), and (1.17) then $\hat{\underline{x}}$ is a global optimal solution to problem (1.14).

CHAPTER II

WOLFE'S ALGORITHM

In this chapter, we develop one of the earliest algorithms used to solve the convex quadratic programming problem. This algorithm makes use of the simplex method of linear programming to find a point which satisfies the Kuhn-Tucker conditions. The reader who is not familiar with linear programming may refer to the Appendix for a brief discussion. Initially, we assume that H is positive definite. Later, we will consider the positive semidefinite case.

Consider the system of equations determined by (1.2) and (1.3):

$$\begin{aligned} A\underline{x} + \underline{y} &= \underline{b} \\ -H\underline{x} - A^T\underline{u} + \underline{v} &= \underline{c} \end{aligned} \qquad (2.1)$$

This is a system of $(n + m)$ equations with $2(n + m)$ variables. If we also require $\underline{x}^T\underline{v} = 0$ and $\underline{u}^T\underline{y} = 0$ then only $(n + m)$ variables can be nonzero. This, along with the nonnegativity requirements in (1.5), means that any vector which satisfies the Kuhn-Tucker conditions must be a basic feasible solution to (2.1).

Theorem 2.1 [12]. When H is positive definite, the convex quadratic programming problem cannot have an unbounded solution.

Therefore, if H is positive definite, the quadratic programming problem is either infeasible or it has a unique solution. We will present a two step algorithm which first determines the feasibility of the problem and then finds the optimal solution.

First, we must insure that all the elements in $\underline{b}$ and $\underline{c}$ are non-negative. To do this we multiply by (-1) any row of (2.1) that corresponds to a negative element of $\underline{b}$ or $\underline{c}$. This yields the system

$$A^*\underline{x} + D\underline{y} = \underline{b}^*$$

$$-H^*\underline{x} - A^{*T}\underline{u} + E\underline{v} = \underline{c}^* \qquad (2.2)$$

where D and E are diagonal matrices with diagonal elements equal to either 1 or -1 .

Then we add artificial variables $\underline{\lambda}$ and $\underline{s}$ to (2.2) to get the system

$$A^*\underline{x} + D\underline{y} + \underline{s} = \underline{b}^*$$

$$-H^*\underline{x} - A^{*T}\underline{u} + E\underline{v} + \underline{\lambda} = \underline{c}^* \qquad (2.3)$$

where $\underline{s}$ is an m-vector and $\underline{\lambda}$ is an n-vector. We are now ready to apply Wolfe's algorithm.

Step 1: We minimize $\Sigma_{i=1}^{m} s_i$ subject to (2.3) and $\underline{x}, \underline{y}, \underline{u}, \underline{v}, \underline{s}, \underline{\lambda} \geq 0$ using the simplex method. However, we do not allow $\underline{u}$ or $\underline{v}$ to assume non-zero values. This means that as we proceed only elements of $\underline{x}$ and $\underline{y}$ will be allowed to assume positive values.

If $\underline{s} = 0$ solves the problem just stated then it follows from (2.3) that there must exist vectors $\underline{\hat{x}}$ and $\underline{\hat{y}}$ such that

$A^*\underline{\hat{x}} + D\underline{\hat{y}} = \underline{b}^*$. Therefore the quadratic programming problem is feasible. Conversely, if the quadratic problem is feasible there exist non-negative solutions to the constraint $A^*\underline{x} + D\underline{y} = \underline{b}^*$. In this case, applying the simplex method as described in Step 1 will result in a basic feasible solution to (2.3) with $\underline{s} = 0$. Therefore, the quadratic programming problem is feasible if and only if Step 1 of the algorithm yields a solution with $\underline{s} = 0$.

Step 2: If the problem is feasible, then we apply the simplex method again to minimize $\Sigma_{i=1}^{n} \lambda_i$. We use the final tableau from Step 1 as our initial tableau. Note that this amounts to starting with an initial basis involving terms of $\underline{x}$, $\underline{y}$, and $\underline{\lambda}$. However, we must modify the simplex method slightly to preserve the complementary slackness constraints. We do not allow both x_i and v_i in the basis at the same time. Similarly, we do not allow both y_i and u_i in the same basis.

We know at this point that the quadratic problem does have an optimal solution (since H is positive definite and the problem is feasible) so there is a basic feasible solution to (2.2). We also know that, in the absence of degeneracy, the simplex method will produce a basic feasible solution. However, it is not immediately clear if our restrictions on basis entry to preserve the complementary slackness constraints will prevent the algorithm from converging. [9] proves that they do not and therefore Step 2 of the algorithm produces the optimal solution to the quadratic programming problem.

Example 2.1. Consider the problem

$$\begin{aligned} \text{Minimize} \quad & f(\underline{x}) = 2x_1 - 3x_2 + x_1^2 + x_2^2 - x_1x_2 \\ \text{subject to} \quad & 2x_1 + 3x_2 \leq 5 \\ & 3x_1 + x_2 \geq 3 \,. \end{aligned} \tag{2.4}$$

Note that $f(\underline{x})$ is the same function considered in Example 1.1 and that it is strictly convex. In this problem

$$c = \begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad H = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad A = \begin{bmatrix} 2 & 3 \\ -3 & -1 \end{bmatrix} \quad \underline{b} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

The Kuhn-Tucker conditions are then

$$\begin{aligned} 2x_1 + 3x_2 + y_1 \qquad\qquad\qquad\qquad\qquad &= 5 \\ -3x_1 - x_2 \qquad + y_2 \qquad\qquad\qquad\qquad &= -3 \\ -2x_1 + x_2 \qquad\qquad - 2u_1 + 3u_2 + v_1 \qquad &= 2 \\ x_1 - 2x_2 \qquad\qquad - 3u_1 + u_2 \qquad + v_2 &= -3 \end{aligned} \tag{2.5}$$

and

$$\underline{x},\ \underline{y},\ \underline{u},\ \underline{v} \geq 0 \ , \ \underline{x}^T\underline{v} = 0 \ , \ \underline{u}^T\underline{y} = 0 \ . \tag{2.6}$$

We make all the right-hand side elements of (2.5) positive and add artificial vectors s_1, λ_1, and λ_2 to get

$$\begin{array}{llllllllr} 2x_1 + 3x_2 + y_1 & & & & & & = 5 & \\ 3x_1 + \ \ x_2 & - y_2 & & & & + s_1 & = 3 & \\ -2x_1 + \ \ x_2 & & - 2u_1 + 3u_2 + v_1 & & & + \lambda_1 & = 2 & \\ - x_1 + 2x_2 & & + 3u_1 - \ \ u_2 & & - v_2 & + \lambda_2 & = 3 & (2.7) \end{array}$$

Notice that we are using y_1 as one of our initial basis vectors.

The first step of the algorithm is to minimize s_1 subject to (2.7) and $\underline{x}, \underline{y}, \underline{u}, \underline{v}, \underline{s}, \underline{z} \geq 0$. We do not allow elements of $\underline{u}$ or $\underline{v}$ to enter the basis.

Iteration 1.

$\underline{B}$	$\underline{c}$	$\underline{b}$	$\underline{x}_1$	$\underline{x}_2$	$\underline{y}_1$	$\underline{y}_2$	$\underline{u}_1$	$\underline{u}_2$	$\underline{v}_1$	$\underline{v}_2$	$\underline{s}_1$	$\underline{\lambda}_1$	$\underline{\lambda}_2$
y_1	0	5	2	3	1	0	0	0	0	0	0	0	0
s_1	1	3	3	1	0	-1	0	0	0	0	1	0	0
λ_1	0	2	-2	1	0	0	-2	3	1	0	0	1	0
λ_2	0	3	-1	2	0	0	3	-1	0	-1	0	0	1
		3	3	1	0	-1	0	0	0	0	0	0	0

Iteration 2:

$\underline{B}$	$\underline{c}$	$\underline{b}$	$\underline{x}_1$	$\underline{x}_2$	$\underline{y}_1$	$\underline{y}_2$	$\underline{u}_1$	$\underline{u}_2$	$\underline{v}_1$	$\underline{v}_2$	$\underline{\lambda}_1$	$\underline{\lambda}_2$
y_1	0	3	0	7/3	1	2/3	0	0	0	0	0	0
x_1	0	1	1	1/3	0	-1/3	0	0	0	0	0	0
λ_1	0	4	0	5/3	0	-2/3	-2	3	1	0	1	0
λ_2	0	4	0	7/3	0	-1/3	3	-1	0	-1	0	1
		0	0	0	0	0	0	0	0	0	0	0

We see that problem (2.4) is feasible and that $(x_1, x_2) = (1, 0)$ is a feasible solution. We now continue with the same tableau and minimize $\lambda_1 + \lambda_2$. Note that if $\underline{x}_1$ were not in the basis we could use $\underline{v}_1$ as a basis vector and we would only need one artificial variable. However, we must preserve the constraint $\underline{x}^T\underline{v} = 0$.

Iteration 3.

$\underline{B}$	$\underline{c}$	$\underline{b}$	$\underline{x}_1$	$\underline{x}_2$	$\underline{y}_1$	$\underline{y}_2$	$\underline{u}_1$	$\underline{u}_2$	$\underline{v}_1$	$\underline{v}_2$	$\underline{\lambda}_1$	$\underline{\lambda}_2$
x_1	0	3	0	7/3	1	2/3	0	0	0	0	0	0
x_2	0	1	1	1/3	0	-1/3	0	0	0	0	0	0
λ_1	1	4	0	5/3	0	-2/3	-2	3	1	0	1	0
λ_2	1	4	0	7/3	0	-1/3	3	-1	0	-1	0	1
		8	0	4	0	-1	1	2	1	-1	0	0

Iteration 4.

B	c	b	x_1	x_2	y_1	y_2	u_1	u_2	v_1	v_2	λ_1	λ_2
λ_2	0	9/7	0	1	3/7	2/7	0	0	0	0	0	0
x_1	0	4/7	1	0	-1/7	-3/7	0	0	0	0	0	0
λ_1	1	13/7	0	0	-5/7	-8/7	-2	3	1	0	1	0
λ_2	1	1	0	0	-1	-1	3	-1	0	-1	0	1
		20/7	0	0	-12/7	-15/7	1	2	1	-1	0	0

Iteration 5.

B	c	b	x_1	x_2	y_1	y_2	u_1	u_2	v_1	v_2	λ_2
x_2	0	9/7	0	1	3/7	2/7	0	0	0	0	0
x_1	0	4/7	1	0	-1/7	-3/7	0	0	0	0	0
u_2	0	13/21	0	0	-5/21	-8/21	-2/3	1	1/3	0	0
λ_2	1	34/21	0	0	-26/21	-29/21	7/3	0	1/3	-1	1
		34/21	0	0	-26/21	-29/21	7/3	0	1/3	-1	0

Iteration 6.

B	c	b	x_1	x_2	y_1	y_2	u_1	u_2	v_1	v_2
x_2	0	9/7	0	1	3/7	2/7	0	0	0	0
x_1	0	4/7	1	0	-1/7	-3/7	0	0	0	0
u_2	0	53/49	0	0	-29/49	-38/49	0	1	3/7	-2/7
u_1	0	34/49	0	0	-26/49	-29/49	1	0	1/7	-3/7
		0	0	0	0	0	0	0	0	0

We see then that

$$x_1 = 4/7 \ , \ x_2 = 9/7$$

is a Kuhn-Tucker point (with $u_1 = 34/49$, $u_2 = 53/49$, and $y_1 = y_2 = u_1 = u_2 = 0$) . This is the optimal solution and the minimum value of $f(\underline{x})$ is $f(4/7, 9/7) = -72/49$.

Recall that we assumed in developing Wolfe's algorithm that H was positive definite. The method will often succeed even if H is only positive semidefinite. If it does not then we perturb H slightly to become $H + \varepsilon I$. This matrix is positive definite and the result should be an acceptable solution to the unperturbed problem if ε is small ([12]) . This is known as Charne's resolution of the semidefinite case. Note that if the original problem has an unbounded solution, this method will result in an incorrect answer. However, most practical problems should not have unbounded solutions.

It is also easy to modify Wolfe's algorithm to handle problems with both inequality and equality constraints. The only major change results from the fact that the Lagrangian multipliers $\hat{\underline{u}}$ associated with the equality constraints need not be positive. In order to apply Wolfe's algorithm we let $\hat{\underline{u}} = \underline{t}_1 - \underline{t}_2$ where $\underline{t}_1 \geq 0$ and $\underline{t}_2 \geq 0$. Note that if we replace any equality constraints $B\underline{x} = \underline{d}$ by $B\underline{x} \leq \underline{d}$ and $B\underline{x} \geq \underline{d}$ to get the constraints

$$\begin{bmatrix} \underline{B} \\ \\ -B \end{bmatrix} \underline{x} \leq \begin{bmatrix} \underline{d} \\ \\ -\underline{d} \end{bmatrix}$$

then the Kuhn-Tucker constraint (1.3) becomes

$$-H\underline{x} - [B \quad -B] \begin{bmatrix} t_1 \\ \\ t_2 \end{bmatrix} + \underline{v} = \underline{c}$$

or

$$-H\underline{x} - B(t_1 - t_2) + \underline{v} = \underline{c} .$$

This, of course, is the same as replacing $\hat{\underline{u}}$ by $\underline{t}_1 - \underline{t}_2$.

CHAPTER III

LEMKE'S COMPLEMENTARY PIVOTING ALGORITHM

In this chapter we will discuss the linear complementary problem and show that the convex quadratic programming problem can be cast as a linear complementary problem. We will then present a method for solving the linear complementary problem. The results given in this chapter may be found in [1] and the theorems are proven there.

Let M be a $p \times p$ matrix and let $\underline{q}$ be a p-vector. The linear complementary problem is to find vectors $\underline{w}$ and $\underline{z}$ such that

$$\underline{w} - M\underline{z} = \underline{q} \tag{3.1}$$

$$\underline{w} \geq \underline{0}\ ,\ \underline{z} \geq \underline{0} \tag{3.2}$$

$$\underline{w}^T\underline{z} = 0\ . \tag{3.3}$$

We call (w_j, z_j) complementary variables and a solution $(\underline{w}, \underline{z})$ to the above system is called a complementary basic feasible solution if $(\underline{w}, \underline{z})$ is a basic feasible solution to (3.1) and (3.2) and if exactly one variable of the pair (w_j, z_j) is basic for $j = 1, \ldots, p$.

Recall the Kuhn-Tucker conditions to the convex quadratic programming problem are

$$A\underline{x} + \underline{y} = \underline{b} \tag{3.4}$$

$$-H\underline{x} - A^T\underline{u} + \underline{v} = \underline{c} \tag{3.5}$$

$$\underline{x}^T\underline{v} = 0 \ , \ \underline{u}^T\underline{y} = 0 \tag{3.6}$$

$$\underline{x}, \underline{y}, \underline{u}, \underline{v} \geq 0 \tag{3.7}$$

Letting

$$M = \begin{bmatrix} 0 & -A \\ A^T & H \end{bmatrix} \qquad \underline{q} = \begin{bmatrix} \underline{b} \\ \underline{c} \end{bmatrix} \qquad \underline{w} = \begin{bmatrix} \underline{y} \\ \underline{v} \end{bmatrix} \quad \text{and} \quad \underline{z} = \begin{bmatrix} \underline{u} \\ \underline{x} \end{bmatrix}$$

we see that (3.4)-(3.7) can be rewritten as

$$\underline{w} - M\underline{z} = \underline{q}$$

$$\underline{w}^T\underline{z} = 0$$

$$\underline{w} \geq 0 \ , \ \underline{z} \geq 0 \ .$$

If we can find vectors $\underline{w}$ and $\underline{z}$ that solve this linear complementary problem then we will have found a Kuhn-Tucker point of the quadratic programming problem.

With this in mind, we now develop an algorithm for solving the linear complementary problem. If $\underline{q} \geq 0$ then the problem is easily solved by letting $\underline{w} = \underline{q}$ and $\underline{z} = 0$. If $\underline{q} \ngeq 0$, we introduce an artificial variable to get the system

$$\underline{w} - M\underline{z} - \underline{1}\, z_0 = \underline{q} \tag{3.8}$$

$$\underline{w},\ \underline{z} \geq 0\ ,\ z_0 \geq 0 \tag{3.9}$$

$$\underline{w}^T\underline{z} = 0 \tag{3.10}$$

where $\underline{1}$ is a vector of 1's . Letting $z_0 = \text{maximum}\{-q_i : 1 \leq i \leq p\}$, $\underline{z} = 0$, and $\underline{w} = \underline{q} + \underline{1}\, z_0$ we obtain a solution to the above system. We will then attempt to drive the artificial variable z_0 to 0 , thus producing a solution to (3.1)-(3.3) .

Definition 3.1. A feasible solution $(\underline{w}, \underline{z}, z_0)$ to (3.8)-(3.10) is called an almost complementary basic feasible solution if:

1. $(\underline{w}, \underline{z}, z_0)$ is a basic feasible solution to (3.8)-(3.10).
2. Neither w_s nor z_s are basic, for some $s \in \{1,\ldots, p\}$.
3. z_0 is basic and exactly one variable of the pair (w_j, z_j) is basic for $j = 1,\ldots, p$ and $j \neq s$.

Definition 3.2. Given an almost complementary basic feasible solution $(\underline{w}, \underline{z}, z_0)$ with w_s and z_s both nonbasic, an adjacent almost complementary basic feasible solution $(\hat{\underline{w}}, \hat{\underline{z}}, z_0)$ is obtained by introducing either w_s or z_s into the basis if pivoting drives a variable other than z_0 from the basis.

Of course, if the pivot mentioned in the previous definition drives z_0 from the basis then it produces a solution to (3.1)-(3.3). The algorithm summarized below moves among adjacent almost complementary basic feasible solutions until either a complementary basic feasible solution is obtained or a direction indicating that the region defined by (3.8)-(3.10) is unbounded is produced.

Initialization. If $\underline{q} \geq 0$, let $(\underline{w}, \underline{z}) = (\underline{q}, 0)$ and stop. Otherwise, display the system (3.8) in tableau format. Let $-q_s = \text{maximum}\{-q_i: 1 \leq i \leq p\}$ and pivot at row s and column z_0 . Thus the basic variables z_0 and w_j for $j = 1, \ldots, p$ and $j \neq s$ are non-negative. Let $y_s = z_s$ and go to the main step.

Main Step. Let $\underline{d}_s$ be the column in the current iteration under the variable y_s . If $\underline{d}_s \leq 0$, then the algorithm stops with ray termination. Let $(\underline{w}, \underline{z}, z_0)$ be the last almost complementary basic feasible solution and let $\underline{d}$ be a vector with a 1 at the row corresponding to y_s , $-\underline{d}_s$ at the rows of the current basic variables, and zero everywhere else. Then a ray $R = \{(\underline{w}, \underline{z}, z_0) + \lambda\underline{d}: \lambda \geq 0\}$ is found such that every point in R satisfies (3.8)-(3.10).

If $\underline{d}_s \not\leq 0$, we use the following minimum ratio test to find the index r :

$$\frac{\bar{q}_r}{d_{rs}} = \underset{1 \leq i \leq p}{\text{minimum}}\{ \frac{\bar{q}_i}{d_{is}} : d_{is} > 0\}$$

where $\underline{\bar{q}}$ is the updated right-hand side vector.

Now let y_s enter the basis by pivoting at row r and column y_s . If z_0 leaves the basis then we have found a complementary basic feasible solution. If not, then the basic variable at row r is either w_ℓ or z_ℓ , for some $\ell \neq s$. If w_ℓ leaves the basis then let $y_s = z_\ell$ and if z_ℓ leaves the basis, let $y_s = w_\ell$. Then repeat the main step.

Before discussing the convergence behavior of the algorithm we give an example.

Example 3.1.

$$\begin{aligned}\text{Minimize}\quad & f(\underline{x}) = 3x_1 - x_2 + x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{subject to}\quad & x_1 + 2x_2 \le 6 \\ & -x_1 + 2x_2 \le 4 \\ & x_1, x_2 \ge 0 .\end{aligned}$$

Then we have:

$$c = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \quad H = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 2 \\ -1 & 2 \end{bmatrix} \quad b = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

It is easy to verify that H is positive definite. Therefore, we know that the problem has a unique solution.

We let

$$M = \begin{bmatrix} 0 & -A \\ A^T & H \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & -2 \\ 0 & 0 & 1 & -2 \\ 1 & -1 & 2 & -2 \\ 2 & 2 & -2 & 4 \end{bmatrix}$$

and

$$\underline{q} = \begin{bmatrix} 6 \\ 4 \\ 3 \\ -1 \end{bmatrix} \qquad \underline{w} = \begin{bmatrix} y_1 \\ y_2 \\ v_1 \\ v_2 \end{bmatrix} \qquad \underline{z} = \begin{bmatrix} u_1 \\ u_2 \\ x_1 \\ x_2 \end{bmatrix}$$

Then the Kuhn-Tucker conditions for the problem can be expressed as the linear complementary problem

$$\underline{w} - M\underline{z} = \underline{q}$$

$$\underline{w}^T\underline{z} = 0$$

$$\underline{w},\ \underline{z} \geq 0\ .$$

We then display the system $\underline{w} - M\underline{z} - \underline{1}\, z_0$ in tableau format and apply Lemke's algorithm.

<u>Initialization Step</u>.

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	q
w_1	1	0	0	0	0	0	1	2	-1	6
w_2	0	1	0	0	0	0	-1	2	-1	4
w_3	0	0	1	0	-1	1	-2	2	-1	3
w_4	0	0	0	1	-2	-2	2	-4	-1	-1

$-q_s = \text{maximum}\{-q_i : 1 \leq i \leq 4\} = -q_4$. So we pivot at row 4 and column z_0 . We let $y_s = z_4$.

Iteration 1:

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	$\bar{q}$
w_1	1	0	0	-1	2	2	-1	6	0	7
w_2	0	1	0	-1	2	2	-3	6	0	5
w_3	0	0	1	-1	1	3	-4	6	0	4
z_0	0	0	0	-1	2	2	-2	4	1	1

Let $d_s = z_4$. Then

$$\frac{\bar{q}_r}{d_{r_s}} = \underset{1 \le i \le p}{\text{minimum}}\left\{\frac{\bar{q}_i}{d_{is}} : d_{is} > 0\right\} = \frac{\bar{q}_4}{d_{4s}} = \frac{1}{4}$$

So we pivot at column z_4 and row 4. This drives the artificial variable from the basis.

Iteration 2:

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	$\bar{q}$
w_1	1	0	0	1/2	-1	-1	2	0	-3/2	11/2
w_2	0	1	0	1/2	-1	-1	0	0	-3/2	7/2
w_3	0	0	1	1/2	-2	0	-1	0	-3/2	5/2
z_4	0	0	0	-1/4	1/2	1/2	-1/2	1	1/4	1/4

Therefore

$$\underline{w} = \begin{bmatrix} 11/2 \\ 7/2 \\ 5/2 \\ 0 \end{bmatrix} , \quad \underline{z} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1/4 \end{bmatrix}$$

is a solution to the linear complementary problem and

$$x_1 = 0 , \; x_2 = 1/4$$

is the solution to the quadratic programming problem.

It is natural to ask if Lemke's algorithm will always produce the optimal solution to a convex quadratic programming problem.

Theorem 3.1 [1] . If each almost complementary basic feasible solution to (3.8)-(3.10) is nondegenerate (that is, each basic variable is positive), then none of the points generated by the algorithm is repeated, and furthermore, the algorithm must stop in a finite number of steps.

Definition 3.1. Let M be a $p \times p$ matrix. Then M is said to be copositive if $\underline{z}^T M\underline{z} \geq 0$ for each $\underline{z} \geq 0$. Furthermore, M is said to be copositive-plus if it is copositive and if $\underline{z} \geq 0$ and $\underline{z}^T M\underline{z} = 0$ imply that $(M + M^T)z = 0$.

Theorem 3.2 [1] . Suppose that each almost complementary basic feasible solution to (3.8)-(3.10) is nondegenerate, and suppose that M is copositive-plus. Then, if the system defined by (3.1) and (3.2) is consistent, the algorithm stops with a complementary basic feasible solution to the problem. If the system (3.1) and (3.2) is inconsistent, the problem stops with ray termination.

Theorem 3.3 [1] . Let A be an $m \times n$ matrix, and let H be an $n \times n$ symmetric matrix. If H is positive semidefinite then

$$M = \begin{bmatrix} 0 & -A \\ A^T & H \end{bmatrix}$$

is copositive-plus.

Theorem 3.4. Suppose we write the Kuhn-Tucker conditions to the convex quadratic programming problem as the linear complementary problem $\underline{w} - M\underline{z} = \underline{q}$, $\underline{w} \geq 0$, $\underline{z} \geq 0$, $\underline{w}^T\underline{z} = 0$ where

$$M = \begin{bmatrix} 0 & -A \\ A^T & H \end{bmatrix} \quad \underline{p} = \begin{bmatrix} \underline{b} \\ \underline{c} \end{bmatrix} \quad \underline{w} = \begin{bmatrix} \underline{y} \\ \underline{v} \end{bmatrix} \quad z = \begin{bmatrix} \underline{u} \\ \underline{z} \end{bmatrix}$$

Then, in the absence of degeneracy, Lemke's algorithm stops in a finite number of iterations with a Kuhn-Tucker point if such a point exists. If the algorithm stops with ray termination the problem is either infeasible or unbounded.

Proof. If a Kuhn-Tucker point exists, then the system

$$\begin{aligned}
&A\underline{x} + \underline{y} = b \\
&-H\underline{x} - A^T\underline{u} + \underline{v} = \underline{c} \\
&\underline{x}^T\underline{v} = 0 \ , \ \underline{y}^T\underline{u} = 0 \\
&\underline{x}, \underline{y}, \underline{u}, \underline{v} \geq 0
\end{aligned} \tag{3.11}$$

is consistent. Therefore, the system

$$\begin{aligned}
&\underline{w} - M\underline{z} = \underline{q} \\
&\underline{w}^T\underline{z} = 0 \\
&\underline{w}, \underline{z} \geq 0
\end{aligned} \tag{3.12}$$

is consistent.

Also, M is copositive-plus by Theorem 3.3. It follows from Theorem 3.2 that Lemke's algorithm produces a complementary basic feasible solution to (3.12) and hence the algorithm produces a Kuhn-Tucker point.

If the algorithm stops with ray termination then Theorem 3.2 implies that the system (3.11) is inconsistent. It follows from Theorem 1.5 that the convex quadratic program is either infeasible or unbounded.

CHAPTER IV

A BARRIER FUNCTION ALGORITHM FOR QUADRATIC PROGRAMMING

In this chapter, we present a barrier function algorithm found in [4] for solving the problem

$$\text{Minimize} \quad \frac{1}{2}\underline{x}^T H\underline{x} + \underline{c}^T\underline{x}$$

$$\text{subject to} \quad A\underline{x} \geq \underline{b} \tag{4.1}$$

where H is symmetric and positive semidefinite. This problem is transformed into a sequence of unconstrained problems whose solutions converge to the solution of (4.1). Instead of actually solving the unconstrained problems, we approximate the solutions to the unconstrained problems in such a way that the approximate solutions still converge to the solution of the constrained problem.

Let

$$S = \{x: \underline{a}_i^T\underline{x} - \underline{b}_i \geq 0 \ , \ i = 1,\ldots, n\}$$

where $\underline{a}_i$ is the ith row of A. The algorithm makes two assumptions about S:

1. int $S \neq \emptyset$, i.e., there exists $\underline{x}_0 \in E_n$ such that $A\underline{x}_0 > \underline{b}$. Note that this excludes the possibility of equality constraints.
2. S is bounded.

We transform the quadratic problem into a sequence of unconstrained problems:

$$P\mu_k: \text{minimize}\{\frac{1}{2}\underline{x}^T H\underline{x} + \underline{c}^T\underline{x} - \mu_k \sum_{i=1}^{n} \ln(\underline{a}_i^T \underline{x} - b_i)\}$$

where $\lim_{k\to\infty} \mu_k = 0$. Intuitively, $-\sum_{i=1}^{n} \ln(\underline{a}_i^T \underline{x} - b_i)$ gets large as $\underline{x}$ approaches the boundary of S and the term $-\mu_k \sum_{i=1}^{n} \ln(\underline{a}_i^T \underline{x} - b_i)$ is a "barrier" which prevents the solution to $P\mu_k$ from being a point outside of S . As $\mu_k \to 0$, the solutions to $P\mu_k$ approach the solution of the quadratic programming problem. For a formal discussion of barrier function methods and for a proof of the convergence of such methods refer to [1].

Let

$$f(\underline{x}, \mu_k) = \frac{1}{2}\underline{x}^T H\underline{x} + \underline{c}^T\underline{x} - \mu_k \sum_{i=1}^{n} \ln(\underline{a}_i^T \underline{x} - b_i)$$

and let $\underline{x}_0^k$ be the value of $\underline{x}$ which minimizes $f(\underline{x}, \mu_k)$. We will write $\underline{x}_0^k \equiv \text{argmin } f(\underline{x}, \mu_k)$. We know that $\lim_{k\to\infty} \underline{x}_0^k = \hat{\underline{x}}$ where $\hat{\underline{x}}$ is the solution to the quadratic problem. Unfortunately, $\underline{x}_0^k$ cannot, in general, be found in polynomial time. Instead we will find an approximation $\underline{x}^k$ to $\underline{x}_0^k$. We define

$$\begin{aligned} F(\underline{x}, \mu_k) &= f(\underline{x}, \mu_k) - f(\underline{x}_0^k, \mu_k) \\ &= \frac{1}{2}(\underline{x}^T H\underline{x} - (\underline{x}_0^k)^T H(\underline{x}_0^k)) + \underline{c}^T(\underline{x} - \underline{x}_0^k) \\ &\qquad - \mu_k \sum_{i=1}^{n} \ln \frac{\underline{a}_i^T \underline{x} - b_i}{\underline{a}_i^T \underline{x}_0^k - b_i} \ . \end{aligned}$$

<u>Definition 4.1</u>. $\underline{x}^k$ is an approximation of $\underline{x}_0^k$ if $F(\underline{x}^k, \mu_k) \leq \varepsilon\mu_k$ for some $\varepsilon > 0$.

Lemma 4.1 [4]. F is strictly convex.

Proof.

$$\nabla F(\underline{x}, \mu) = H\underline{x} + \underline{c} - \mu \sum_{i=1}^{n} \frac{\underline{a}_i^T \underline{x}_0^k - b_i}{\underline{a}_i^T \underline{x} - b_i} \cdot \frac{(\underline{a}_i^T \underline{x}_0^k - b_i)\underline{a}_i}{(\underline{a}_i^T \underline{x}_0^k - b_i)^2}$$

$$= H\underline{x} + \underline{c} - \mu \sum_{i=1}^{n} \frac{\underline{a}_i}{\underline{a}_i^T \underline{x} - b_i}$$

$$= H\underline{x} + \underline{c} - A^T D_\mu^{-1} \underline{e}$$

where D_μ^{-1} is an $m \times m$ diagonal matrix with $D_{ii}^{-1} = \dfrac{\mu}{\underline{a}_i^T \underline{x} - b_i}$ and $\underline{e} = [1, \ldots, 1] \in E_m$.

The Hessian G_k of F is given by

$$\nabla^2 F(\underline{x}, \mu) = H + A^T D_\mu^{-2} A$$

where D_μ^{-2} has diagonals $D_{ii}^{-2} = \dfrac{\mu}{(\underline{a}_i^T \underline{x} - b_i)^2}$ for $i = 1, \ldots, m$. Here the columns of A are assumed to be linearly independent. H is positive semidefinite and D_μ^{-2} is positive definite because $\underline{a}_i^T \underline{x} - b_i > 0$ for all i when $x \in \text{int } S$ and $\mu > 0$. Therefore $A^T D_\mu^{-2} A$ is positive definite and the result follows.

It follows from the lemma that $F(\underline{x}, \mu_k)$ has a unique minimum on S . It is also clear from the definition of $F(\underline{x}, \mu_k)$ that this minimum must be the point $\underline{x}_0^k \equiv \text{argmin } f(\underline{x}, \mu_k)$.

Suppose we have a point $\underline{x}^{k-1}$ such that $F(\underline{x}^{k-1}, \mu_{k-1}) \leq \varepsilon \mu_{k-1}$. We will then find $\underline{x}^k$, the approximation to $\underline{x}_0^k$, by taking a step in the Newton direction $\underline{d}_k$ satisfying

$$G_k \underline{d}_k = -\underline{g}_k$$

where G_k and $\underline{g}_k$ are the Hessian and gradient, respectively, of $F(\underline{x}, \mu_k)$ evaluated at $\underline{x}^{k-1}$. Note that

$$\underline{g}_k = H\underline{x}^{k-1} + \underline{c} - A^T D_\mu^{-1} \underline{e}$$

and

$$G_k = H + A^T D_\mu^{-2} A \ .$$

Computationally, D_μ^{-2} is updated as follows:

$$D_{ii}^{-2} \text{ is reset to } \frac{\mu_k}{(\underline{a}_i^T \underline{x}^k - b_i)^2}$$

$$\text{if} \quad D_{ii}^{-2} \notin [\frac{\mu_k}{1.1\ [(\underline{a}_i^T \underline{x}^k - b_i)^2} \ , \ \frac{1.1\mu_k}{(\underline{a}_i^T \underline{x}^k - b_i)^2}] \qquad (6.2)$$

This leads to a matrix $\overline{D}_\mu^{-2}$ which is an approximation of D_μ^{-2} . Similarly,

$$\overline{G}_k = H + A^T \overline{D}_\mu^{-2} A$$

is an approximation of G_k .

We are now ready to present the algorithm proposed in [4] .

<u>Initialization</u>. Let $\underline{x}^0 \in \text{int } S$ and $\mu_0 > 0$ be given such that $F(\underline{x}^0, \mu_0) \leq \varepsilon\, \mu_0$. Let

$$\varepsilon = .036\ , \quad \delta = .04\ , \ \varepsilon_1 = .051 \ \text{ and } \ \varepsilon^t > 0\ .$$

<u>Main Step</u>

1. Solve $\overline{G}_k\, \underline{d}_k = -\underline{g}_k$ to find the Newton direction.
2. Select r such that $.45 \leq r\sqrt{\mu_k} \leq .95$.
3. Let $\lambda_k = \dfrac{r\, \mu_k\, \sqrt{\varepsilon_1}}{\sqrt{\underline{d}_k^T\, \overline{G}_k\, \underline{d}_k}}$ be the steplength .
4. Let $\underline{x}^{k+1} = \underline{x}^k + \lambda_k\, \underline{d}^k$.
5. Let $\mu_{k+1} = (1 - \delta/\sqrt{m})\mu_k$ where A is an $m \times n$ matrix.
6. Update D_{ii}^{-2} as described in (4.2) .
7. If $\mu_k < \varepsilon^t$ then the algorithm terminates; otherwise return to step 1 .

<u>Example 4.1</u>.

Consider the problem

$$\begin{array}{ll} \text{Minimize} & x_1^2 - x_1x_2 + x_2^2 - 3x_1 \\ \text{subject to} & x_1 \geq 0 \\ & x_2 \geq 0 \\ & -x_1 - x_2 \geq -2\ . \end{array}$$

We will illustrate the algorithm by showing one iteration. Suppose $\underline{x}^k = (1, .5)$ and $\mu = 1$.

We have

$$H = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \underline{c} = \begin{bmatrix} -3 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}, \text{ and } \quad \underline{b} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}$$

In this case,

$$D_\mu^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \text{and} \quad D_\mu^{-2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Also,

$$G_k = H + A^T D_\mu^{-2} A = \begin{bmatrix} 7 & 3 \\ 3 & 10 \end{bmatrix}$$

and

$$\underline{g}_k = H\underline{x} + \underline{c} - A^T D_\mu^{-1} \underline{e} = \begin{bmatrix} -.5 \\ 0 \end{bmatrix} .$$

Solving the system $G_k \underline{d} = -\underline{g}_k$ yields

$$\underline{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \cong \begin{bmatrix} .078 \\ -.023 \end{bmatrix}$$

Let $r = .95$. Note that $\underline{d}^T G_0 \underline{d} \cong .037$ and

$$\lambda = \frac{.95\sqrt{.051}}{\sqrt{.037}} \cong 1.12 .$$

Then,

$$\underline{x}^{k+1} = \begin{bmatrix} 1 \\ -5 \end{bmatrix} + 1.12 \begin{bmatrix} .078 \\ -.023 \end{bmatrix} = \begin{bmatrix} 1.09 \\ .47 \end{bmatrix} .$$

In the next chapter, we will see that the solution to this problem is $(x_1, x_2) = (3/2 , 1/2)$.

The convergence of the algorithm is proven in three steps. First, it is shown that $\underline{x}^k$ lies on the boundary of

$$E^k(r) = \{\underline{x}: (\underline{x} - \underline{x}^{k-1})^T \bar{G}_k(\underline{x} - \underline{x}^{k-1}) \leq r^2 \varepsilon_1(\mu_k)^2\}$$

and that $E^k(r) \subseteq S$. It follows that the points generated by the algorithm are feasible.

To complete the proof, it is then necessary to show that $\underline{x}^k$ is an approximation of $\underline{x}_0^k$. [4] does this by showing first that

$$F(\underline{x}^{k-1}, \mu_k) \leq \varepsilon_1 \mu_k .$$

Finally, it is shown that

$$F(\underline{x}^k, \mu_k) \le \gamma\, F(\underline{x}^{k-1}, \mu_k)\ .$$

The parameters are chosen so that $\gamma\, \varepsilon_1 \le \varepsilon$ and it follows that

$$F(\underline{x}^k, \mu_k) \le \varepsilon\, \mu_k\ .$$

Lemma 4.2. $\underline{x}^k$ lies on the boundary of $E^k(r)$.

Proof.

$$(\underline{x}^k - \underline{x}^{k-1})^T\, \bar{G}_k(\underline{x}^k - \underline{x}^{k-1})$$

$$= (\lambda_k\, \underline{d}_k)^T\, \bar{G}_k(\lambda_k\, \underline{d}_k)$$

$$= \left(\frac{r\, \mu_k\sqrt{\varepsilon_1}}{\sqrt{\underline{d}_k^T\, \bar{G}_k\, \underline{d}_k}}\, \underline{d}_k\right)^T \quad G_k\left(\frac{r\, \mu_k\, \sqrt{\varepsilon_1}}{\sqrt{\underline{d}_k^T\, \bar{G}_k\, \underline{d}_k}}\, \underline{d}_k\right)$$

$$= r^2\, \varepsilon_1(\mu_k)^2\ .$$

Theorem 4.2 [4] . $E^k(r) \subseteq S$.

Let $\mu_k = (1 - \delta/\sqrt{m})\mu_{k-1}$, $\delta \in (0,\, 1)$, $\varepsilon = \frac{\alpha^2}{2} - \frac{\alpha^3}{3}$ and $\alpha \in (0,\, 1)$. We assume that $F(\underline{x}^{k-1}, \mu_{k-1}) \le \varepsilon\, \mu_k$. Under these assumptions the following results are true.

Lemma 4.3 [4] . $\dfrac{\delta}{\sqrt{m}}\ \mu_{k-1} \sum_{i=1}^{n} \ln \dfrac{\underline{a}_i^T\, \underline{x}^{k-1} - b_i}{\underline{a}_i^T\, \underline{x}_0^{k-1} - b_i} \le \alpha\mu_{k-1}\sqrt{m}$.

Lemma 4.4 [4] . $F(\underline{x}_0^{k-1}\ ,\, \mu_k) \le \delta^2\, \mu_{k-1}$.

<u>Theorem 4.2 [4]</u> . $F(\underline{x}^{k-1}, \mu_k) \leq \varepsilon_1 \mu_k$ where

$$\varepsilon_1 = \frac{\varepsilon + \delta^2 + \delta\alpha}{1 - \delta/\sqrt{m}} .$$

<u>Proof</u>.

$$F(\underline{x}^{k-1}, \mu_{k-1}) = \tfrac{1}{2}[(\underline{x}^{k-1})^T H\underline{x}^{k-1} - (\underline{x}_0^{k-1})^T H\underline{x}_0^{k-1}]$$

$$+ \underline{c}^T(\underline{x}^{k-1} - \underline{x}_0^{k-1}) - \mu_{k-1} \sum_{i=1}^{m} \ln \frac{\underline{a}_i^T \underline{x}^{k-1} - b_i}{\underline{a}_i^T \underline{x}_0^{k-1} - b_i} .$$

$$F(\underline{x}_0^{k-1}, \mu_k) = \frac{1}{2} [(\underline{x}_0^{k-1})^T H \underline{x}_0^{k-1} - (\underline{x}_0^k)^T H \underline{x}_0^k]$$

$$+ \underline{c}^T(\underline{x}_0^{k-1} - \underline{x}_0^k) - \mu_k \sum_{i=1}^{m} \ln \frac{\underline{a}_i^T \underline{x}_0^{k-1} - b_i}{\underline{a}_i^T \underline{x}_0^k - b_i} .$$

Adding and letting $\mu_k = (1 - \delta/\sqrt{m}) \mu_{k-1}$, $\delta \in (0, 1)$ yields

$$F(\underline{x}^{k-1}, \mu_{k-1}) + F(\underline{x}_0^{k-1}, \mu_k) = F(\underline{x}^{k-1}, \mu_k)$$

$$- \frac{\delta}{\sqrt{m}} \mu_{k-1} \sum_{i=1}^{n} \ln \frac{\underline{a}_i^T \underline{x}^{k-1} - b_i}{\underline{a}_i^T \underline{x}_0^{k-1} - b_i}$$

It follows that

$$F(\underline{x}^{k-1}, \mu_k) = F(\underline{x}^{k-1}, \mu_{k-1}) + F(\underline{x}_0^{k-1}, \mu_k)$$

$$+ \frac{\delta}{\sqrt{m}} \mu_{k-1} \sum_{i=1}^{m} \ln \frac{\underline{a}_i^T \underline{x}^{k-1} - b_i}{\underline{a}_i^T \underline{x}_0^{k-1} - b_i} .$$

Using the two previous lemmas and the assumption that $F(\underline{x}^{k-1}, \mu_{k-1}) \leq \varepsilon\, \mu_{k-1}$ yields

$$F(\underline{x}^{k-1}, \mu_k) \leq \varepsilon\, \mu_{k-1} + \delta^2\, \mu_{k-1} + \delta\alpha\, \mu_{k-1}$$

$$= \frac{\varepsilon + \delta^2 + \delta\alpha}{1 - \delta/\sqrt{m}}\, \mu_k$$

$$= \varepsilon_1 \mu_k \;.$$

Lemma 4.5 [4] . $\underline{g}_k^T(\underline{x}^k - \underline{x}^{k-1}) \leq -\theta r \;\sqrt{\varepsilon\varepsilon_1}\, (\mu_k)^{3/2}$ where

$$\theta = \frac{\sqrt{1 - \alpha}}{\sqrt{1.1 \sum_{j=0}^{\infty} \frac{\alpha^j}{(j+1)(j+2)}}} \;.$$

Lemma 4.6 [4] . $\frac{1}{2}\, (\underline{x}^k - \underline{x}^{k-1})^T\, G_k(\underline{x}^k - \underline{x}^{k-1}) \leq .55\, r^2\, \varepsilon_1(\mu_k)^2$.

Proof. $\underline{x}^k \in E^k(r)$ so

$$(\underline{x}^k - \underline{x}^{k-1})^T\, \overline{G}_k(\underline{x}^k - \underline{x}^{k-1}) \leq r^2 \varepsilon_1(\mu_k)^2 \;.$$

Then,

$$\frac{1}{2}(\underline{x}^k - \underline{x}^{k-1})^T\, G_k(\underline{x}^k - \underline{x}^{k-1})$$

$$= \frac{1}{2}\, (\underline{x}^k - \underline{x}^{k-1})^T(H + \sum_{i=1}^{n} \mu_k \frac{\underline{a}_i\, \underline{a}_i^T}{(\underline{a}_i^T\, \underline{x}^{k-1} - b_i)^2})\; (\underline{x}^k - \underline{x}^{k-1})$$

$$\leq \frac{1.1}{2}\, (\underline{x}^k - \underline{x}^{k-1})^T\; (H + A^T\, D_\mu^{-2}\, A)\; (\underline{x}^k - \underline{x}^{k-1})$$

$$= .55(\underline{x}^k - \underline{x}^{k-1})^T \bar{G}_k(\underline{x}^k - \underline{x}^{k-1})$$

$$\leq .55\ r^2 \varepsilon_1 (\mu_k)^2 .$$

Lemma 4.7 [4] .

$$\left| \sum_{j=3}^{\infty} \frac{(-1)^j}{j} \sum_{i=1}^{m} \frac{\mu_k [\underline{a}_i^T (\underline{x}^k - \underline{x}^{k-1})]^j}{(\underline{a}_i^T \underline{x}^{k-1} - b_i)^j} \right| \leq q r^2 \varepsilon_1 (\mu_k)^2$$

where $q = \dfrac{(1.1)^{3/2}\ r\ \sqrt{\varepsilon_1 \mu_k}}{3[1 - r\ \sqrt{1.1\varepsilon_1 \mu_k}]}$.

Theorem 4.3 [4] . Suppose that $F(\underline{x}^{k-1}, \mu_{k-1}) \leq \varepsilon\mu_{k-1}$. Let $\underline{x}^k$ be the point on the boundary of $E^k(r)$ along the Newton direction from $\underline{x}^{k-1}$. Then,

$$F(\underline{x}^k, \mu_k) \leq (1 - \theta r \sqrt{\frac{\varepsilon}{\varepsilon_1}}\ \sqrt{\mu_k} + r^2 \mu_k\ (.55 + q))\varepsilon_1 \mu_k$$

where θ and q are as defined in Lemma 4.5 and Lemma 4.7.

Proof. Using the Taylor series expansion of $F(\underline{x}, \mu_k)$ at $\underline{x}^{k-1}$ we write

$$F(\underline{x}^{k-1} + \lambda\underline{d}, \mu_k) = F(\underline{x}^{k-1}, \mu_k) + \lambda\ \underline{g}_k^T\ d + \frac{\lambda^2}{2} \underline{d}^T\ G_k\ d$$

$$+ \sum_{j=3}^{\infty} \frac{(-1)^j\ \lambda^j}{j} \sum_{i=1}^{m} \mu_k \frac{(\underline{a}_i^T\ \underline{d})^j}{(\underline{a}_i^T\ \underline{x}^{k-1} - b_i)^j} .$$

Letting $\underline{x}_k = \underline{x}_{k-1} + \lambda \underline{d}$ this becomes

$$F(\underline{x}^k, \mu_k) = F(\underline{x}^{k-1}, \mu_k) + \underline{g}_k^T(\underline{x}^k - \underline{x}^{k-1})$$

$$+ \frac{1}{2} (\underline{x}^k - \underline{x}^{k-1})^T G_k(\underline{x}^k - \underline{x}^{k-1})$$

$$+ \sum_{j=3}^{\infty} \frac{(-1)^j}{j} \sum_{i=1}^{n} \mu_k \frac{[\underline{a}_i^T(\underline{x}^k - \underline{x}^{k-1})]^j}{(\underline{a}_i^T \underline{x}^{k-1} - b_i)^j} .$$

Using the previous lemmas this becomes

$$F(\underline{x}^k, \mu_k) \leq \varepsilon_1\mu_k - \theta r \sqrt{\varepsilon\varepsilon_1}(\mu_k)^{3/2}$$

$$+ .55r^2 \varepsilon_1(\mu_k)^2 + qr^2\varepsilon_1(\mu_k)^2$$

$$= (1 - \theta r \sqrt{\varepsilon\varepsilon_1} \sqrt{\mu_k} + r^2\mu_k(.55 + q)) \varepsilon_1\mu_k .$$

<u>Theorem 4.4 [4]</u>. Let $\alpha \leq .3$ and $0 \leq \delta \leq .04$. Then for $m \geq 3$ and $.45 \leq r \sqrt{\mu_k} \leq .95$,

$$F(\underline{x}^k, \mu_k) \leq \varepsilon \mu_k .$$

<u>Proof</u>. Let $\gamma = 1 - \theta r \sqrt{\frac{\varepsilon}{\varepsilon_1}} \sqrt{\mu_k} + (.55 + \frac{1.1}{3} \frac{r \sqrt{1.1\varepsilon_1\mu_k}}{1 - r\sqrt{1.1 \varepsilon_1\mu_k}}) r^2\mu_k$

since $\alpha \leq .3$ and $\delta \leq .4$, $\varepsilon = \frac{\alpha^2}{2} - \frac{\alpha^3}{3} \leq .036$. Also, $\theta > 1$ and $\varepsilon_1 = \frac{\varepsilon + \delta^2 + \delta\alpha}{1 - \delta/\delta\sqrt{m}} \leq .051$. Then, for $.45 \leq r\sqrt{\mu_k} \leq .95$, $\gamma \leq \varepsilon/\varepsilon_1$.

By Theorem 4.2,

$$F(\underline{x}^k, \mu_k) \leq \gamma\, \varepsilon_1 \mu_k \ .$$

The result then follows since $\gamma \leq \varepsilon/\varepsilon_1$.

It follows that the algorithm converges. It can also be shown that the algorithm converges in polynomial time (see [4]) .

CHAPTER V

AN ACTIVE SET METHOD FOR QUADRATIC PROGRAMMING

Suppose we have a convex quadratic programming problem with only equality constraints. That is, we have the problem

$$\begin{aligned} &\text{Minimize} && \frac{1}{2}\underline{x}^T H\underline{x} + \underline{c}^T\underline{x} \\ &\text{subject to} && A\underline{x} = \underline{b} \end{aligned} \tag{5.1}$$

where H is positive definite. The Kuhn-Tucker conditions for this problem are

$$\begin{aligned} H\underline{x} + A^T\underline{\lambda} &= -\underline{c} \\ A\underline{x} \qquad &= \underline{b} \end{aligned} \tag{5.2}$$

Therefore solving (5.1) is simply a matter of solving the system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \underline{x} \\ \underline{\lambda} \end{bmatrix} = \begin{bmatrix} -\underline{c} \\ \underline{b} \end{bmatrix} \tag{5.3}$$

The matrix $\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}$, known as the Lagrangian matrix, is symmetric but not necessarily positive definite. [7] and [9] discuss efficient ways of solving such systems.

Now consider the more general problem

$$\text{Minimize} \qquad f(\underline{x}) = \frac{1}{2}\,\underline{x}^T H\underline{x} + \underline{c}^T\underline{x}$$

$$\text{subject to} \qquad A\underline{x} \le \underline{b} \qquad (5.4)$$

where H is symmetric and positive definite. Suppose $\underline{x}$ is a feasible solution and we wish to move to a feasible solution $\hat{\underline{x}}$ such that $f(\hat{\underline{x}}) < f(\underline{x})$. The only constraints which effect the direction we can move are those constraints for which

$$\underline{a}_i^T\,\underline{x} = b_i$$

where $\underline{a}_i$ is the transpose of the ith row of A . We will index the constraints by the set I and will call them active constraints. The basic idea behind active set methods is to solve a problem

$$\text{Minimize} \qquad \frac{1}{2}\,\underline{x}^T H\underline{x} + \underline{c}^T\underline{x}$$

$$\text{subject to} \qquad a_i^T x = b_i \;,\; i \in I \qquad (5.5)$$

at each iteration. Constraints are added or deleted from the set of active constraints as needed. Eventually the solution to one of the equality problems such as (5.5) is a solution to problem (5.4). We will describe one active set method which is presented in [7] .

Assume that a feasible point $x^{(1)}$ is known and let $I^{(1)}$ index the constraints which are equality satisfied at $x^{(1)}$. Set $k = 1$. The algorithm then proceeds as follows:

1. If $\underline{s} = 0$ does not solve the problem

$$\text{Minimize} \quad \frac{1}{2}\underline{s}^T H\underline{s} + \underline{s}^T \underline{c}^{(k)}$$

$$\text{subject to} \quad \underline{a}_i^T \underline{s} = 0 \ , \ i \in I^{(k)} \ , \tag{5.6}$$

where $\underline{c}^{(k)} = \underline{c} + H\underline{x}^{(k)} = \nabla f(\underline{x}^{(k)})$, then go to step 3 .

2. Solve the system

$$H\underline{x}^{(k)} + \bar{A}^T \underline{\lambda} = -\underline{c} \ ,$$

where $\bar{A}$ denotes the matrix of active constraints, to find $\underline{\lambda}^{(k)}$. If $\underline{\lambda}^{(k)} \geq 0$ then terminate the algorithm. Otherwise select q to solve $\min\limits_{i \in I \cap F} \lambda_i^{(k)}$ where $F = \{i: \lambda_i^{(k)} < 0\}$. Remove q from $I^{(k)}$.

3. Solve (5.6) to find $\underline{s}^{(k)}$.
4. Let

$$\alpha^{(k)} = \min(1, \min_{\substack{i \notin I \\ \underline{a}_i^T \underline{s}^{(k)} > 0}} \frac{b_i - \underline{a}_i^T \underline{x}^{(k)}}{a_i^T s^{(k)}}) \ . \tag{5.7}$$

5. Let $\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha^{(k)} \underline{s}^{(k)}$.
6. If $\alpha^{(k)} < 1$ let p to $I^{(k)}$ where p is the index at which (5.7) achieves its minimum.
7. Let $k = k + 1$ and let $I^{(k+1)} = I^{(k)}$. Go to step 1.

<u>Example 5.1</u>. Consider the problem

$$\begin{array}{lll} \text{Minimize} & x_1^2 - x_1 x_2 + x_2^2 - 3x_1 & \\ \text{subject to} & x_1 \geq 0 & (1) \\ & x_2 \geq 0 & (2) \\ & x_1 + x_2 \leq 2 \; . & (3) \end{array}$$

Let $\underline{x}^{(1)} = [0 \quad 0]^T$. In this case,

$$H = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad \text{and} \quad \underline{c} = \begin{bmatrix} -3 \\ 0 \end{bmatrix} .$$

$I^{(1)} = \{1, 2\}$ since the first two constraints are equality satisfied. We first consider the problem

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2} \underline{s}^T H\underline{s} + \underline{s}^T \underline{c}^{(1)} \\ \text{subject to} & \underline{s}_1 = 0 \\ & \underline{s}_2 = 0 \; . \end{array}$$

Clearly $\underline{s} = [0 \quad 0]^T$ is the solution to this problem.

We then solve

$$H\underline{x} + \bar{A}^T \underline{\lambda} = -\underline{c}$$

where

$$\overline{A} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

We find that $\lambda_1 = -3$ and $\lambda_2 = 0$. We then let $I^{(1)} = \{2\}$ since $\lambda_1 < 0$.

We then consider the problem

$$\text{Minimize} \qquad \frac{1}{2}\, \underline{s}^T H\underline{s} + \underline{s}^T\underline{c}^{(1)}$$

$$\text{subject to} \qquad s_2 = 0 \;.$$

This simplifies to the problem

$$\text{Minimize} \quad s_1^2 - 3\, s_1 \;.$$

Therefore $\underline{s}^T = [3/2 \quad 0]$ and $\underline{x}^{(2)} = \underline{x}^{(1)} + \underline{s}^{(k)} = [3/2 \quad 0]^T$.

Let $I^{(2)} = \{2\}$ and consider the problem

$$\text{Minimize} \qquad \frac{1}{2}\, \underline{s}^T H\underline{s} + \underline{s}^T\underline{c}^{(2)}$$

$$\text{subject to} \qquad s_2 = 0 \;.$$

This becomes

$$\text{minimize} \quad s_1^2 \;.$$

Therefore, $\underline{s} = 0$ solves this problem. Solving the system

$$H\underline{x}^{(2)} + \overline{A}^T \underline{\lambda} = -c$$

we find

$$\lambda_2 = -\frac{3}{2} .$$

Therefore, we delete constraint 2 . Now $I^{(2)} = \emptyset$.

We then solve the unconstrained problem

$$\text{Min } s_1^2 - s_1 s_2 + s_2^2 - \frac{3}{2} s_2$$

to find $\underline{s} = [\frac{1}{2} \quad 1]^T$. Note that $\underline{a}_3^T \underline{s} > 0$ and

$$\alpha = \min(1, \frac{2 - 3/2}{3/2}) = 1/3 .$$

Then

$$\underline{x}^{(3)} = \begin{bmatrix} 3/2 \\ 0 \end{bmatrix} + \frac{1}{3} \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10/6 \\ 1/3 \end{bmatrix} .$$

The third constraint is added to the active set so $I^{(3)} = \{3\}$.

To begin the next iteration we solve the problem

$$\begin{aligned} &\text{Minimize} \quad && s_1^2 - s_1 s_2 + s_2^2 - s_2 \\ &\text{subject to} \quad && s_1 + s_2 = 0 \end{aligned}$$

and find $\underline{s} = [-1/6 \quad 1/6]^T$. We note that $a_1^T \underline{s} > 0$ but

$$\alpha = \min(1, \frac{0 + 10/6}{1/6}) = 1 .$$

Therefore,

$$\underline{x}^{(4)} = \begin{bmatrix} 10/6 \\ \\ 1/3 \end{bmatrix} + \begin{bmatrix} -1/6 \\ \\ 1/6 \end{bmatrix} = \begin{bmatrix} 3/2 \\ \\ 1/2 \end{bmatrix}$$

At the next iteration, we solve the problem

$$\text{Maximize} \quad s_1^2 - s_1 s_2 + s_2^2 - \frac{1}{2} s_1 - \frac{1}{2} s_2$$

$$\text{subject to} \qquad s_1 + s_2 = 0$$

and find $\underline{s} = [0 \quad 0]^T$.

Solving

$$H\underline{x}^{(4)} + \bar{A}^T \lambda = -\underline{c}$$

we find that

$$\lambda = 1/2 \quad .$$

Therefore, the algorithm terminates and the optimal solution is

$$\underline{x} = [3/2 \quad 1/2]^T .$$

The following results deal with how and why this algorithm works.

Theorem 5.1. The vectors $\underline{a}_i$, $i \in I^{(k)}$ are linearly independent if the vectors $\underline{a}_i$, $i \in I^{(1)}$ are linearly independent.

Proof. The initial vectors are linearly independent. Deleting vectors from I obviously does not change this. If a vector $\underline{a}_p$ is added to the set of active constants then by (5.7) $\underline{a}_p^T \underline{s}^{(k)} > 0$. Suppose that

$$\underline{a}_p = c_1\underline{a}_1 + \dots + c_\ell\underline{a}_\ell$$

where $\underline{a}_1,\dots, \underline{a}_\ell$ are the active constraints. Then

$$\underline{a}_p^T \underline{s}^{(k)} = (c_1\underline{a}_1 + \dots + c_\ell a_\ell)^T \underline{s}^k = 0$$

since $\underline{a}_i^T \underline{s}^{(k)} = 0$ for $i \in I$. This contradicts $\underline{a}_p^T \underline{s}^{(k)} > 0$ so we conclude that the vectors $\underline{a}_1,\dots, \underline{a}_\ell$, $\underline{a}_p$ are linearly independent.

Theorem 5.2. Each iterate $x^{(k)}$ is a feasible solution to problem (5.4) and $x^{(k)}$ equality satisfies all the active constraints.

Proof. The result holds for $\underline{x}^{(1)}$. Suppose it is true for $\underline{x}^{(k)}$. If the result then holds for $\underline{x}^{(k+1)}$ the theorem follows by induction.

We let $\bar{A}$ denote the matrix of active constraints. Then

$$\bar{A}(\underline{x}^{(k+1)}) = \bar{A}(\underline{x}^{(k)} + \alpha^{(k)}\ \underline{s}^{(k)})$$

$$= \bar{A}\ \underline{x}^{(k)} + \alpha^{(k)}\ \bar{A}\ \underline{s}^{(k)}$$

$$= \underline{b}^{(k)}$$

where $\underline{b}^{(k)}$ are the components of $\underline{b}$ associated with the active constraints.

Now suppose $p \notin I$. First consider the case $\underline{a}_p^T\ \underline{s}^{(k)} \leq 0$. Then

$$\underline{a}_p^T\ (\underline{x}^{(k+1)}) = \underline{a}_p^T\ (\underline{x}^k + \alpha^{(k)}\ \underline{s}^{(k)})$$

$$\leq b_p$$

since $\underline{a}_p^T\ \underline{x}^k \leq b_k$ by the induction hypothesis and $\alpha^{(k)}\ \underline{a}_p^T\ \underline{s}^{(k)} \leq 0$. (Note that $\alpha^{(k)} \in [0, 1]$).

Now consider the case $\underline{a}_p^T\ \underline{s}^{(k)} > 0$. In this case note that

$$\alpha^{(k)} \leq \frac{b_p - \underline{a}_p^T\ \underline{x}^{(k)}}{\underline{a}_i^T\ \underline{s}^{(k)}} \quad .$$

Therefore

$$\underline{a}_p^T (x^{k+1}) = \underline{a}_p^T (\underline{x}^{(k)} + \alpha^{(k)} \underline{s}^{(k)})$$

$$\leq \underline{a}_p^T \underline{x}^{(k)} + \underline{a}_p^T \frac{b_p - \underline{a}_p^T \underline{x}^{(k)}}{\underline{a}_p^T \underline{s}^k} \underline{s}^{(k)}$$

$$= \underline{a}_p^T \underline{x}^{(k)} + \underline{b}_p - \underline{a}_p^T \underline{x}^{(k)}$$

$$= \underline{b}_p \ .$$

Note that equality holds if and only if p is the index at which (5.7) achieves its minimum. In this case, p is added to I so $\underline{x}^{(x+1)}$ satisfies the new set of active constraints. This completes the proof.

Theorem 5.3. $\underline{s} = 0$ solves problem (5.6) if and only if $\underline{x}^{(k)}$ solves

$$\begin{array}{lll} \text{Minimize} & \frac{1}{2} \underline{x}^T H\underline{x} + \underline{c}^T \underline{x} & \\ \text{subject to} & \underline{a}_i^T \underline{x} = b_i \ , \quad i \in I^{(k)} \ . & \end{array} \tag{5.8}$$

Proof. Note that it follows from Theorem 5.1 that (5.6) and (5.8) are well-defined.

Suppose $\underline{s} = 0$ solves (5.6). Then it follows from (5.2) that

$$H\underline{s} + \bar{A}^T \underline{\lambda} = -\underline{c} - H\underline{x}^{(k)}$$

since $\underline{s} = 0$,

$$\bar{A}^T \underline{\lambda} = -\underline{c} - H\, \underline{x}^{(k)} .$$

Also it follows from the previous theorem that

$$\bar{A}\, \underline{x}^{(k)} = \underline{b}^{(k)} .$$

It then follows from (5.2) that $\underline{x}^{(k)}$ solves problem (5.8).

Conversely, suppose $\underline{x}^{(k)}$ solves (5.8). Again, we have

$$H\underline{x}^{(k)} + \bar{A}^T \underline{\lambda} = -\underline{c} .$$

Therefore, if $\underline{s} = 0$,

$$H\underline{s} + \bar{A}^T \underline{\lambda} = -\underline{c} - H\underline{x}^{(k)} = -\underline{c}^{(k)}$$

and

$$\bar{A}\, \underline{s} = 0 .$$

It follows from (5.2) that $\underline{s} = 0$ solves problem (5.6).

Theorem 5.4. If $\underline{s} = 0$ solves (5.6) and $\underline{\lambda}^{(k)} \geq 0$ then $\underline{x}^{(k)}$ solves problem (5.4).

Proof. The Kuhn-Tucker conditions for problem (5.4) are

$$\begin{aligned} H\underline{x} + \underline{c} + A^T\underline{\lambda} &= 0 \\ A\underline{x} + \underline{y} &= \underline{b} \\ \underline{\lambda}^T \underline{y} &= 0 \\ \underline{\lambda} &\geq 0 \,. \end{aligned}$$

Let $\lambda_i = 0$ if $i \notin I$. It then follows form the previous theorem that

$$H\underline{x}^{(k)} + A^T \underline{\lambda} + \underline{c} = 0 \,.$$

Let $\lambda_i = 0$ if $i \notin I$ and $y_i = 0$ if $i \in I$ so

$$\underline{\lambda}^T\underline{y} = 0 \,.$$

Theorem (5.2) assures us that $\underline{x}^{(k)}$ is feasible so

$$A\underline{x} + \underline{y} = \underline{b}$$

Therefore, if $\underline{\lambda} \geq 0$, the Kuhn-Tucker conditions to the convex quadratic program hold and $\underline{x}^{(k)}$ is the optimal solution to (5.4).

Theorem 5.5. If $\alpha(k) = 1$, then $\underline{x}^{(k+1)}$ solves the problem

$$\begin{aligned} &\text{Minimize} && \frac{1}{2}\underline{x}^T H\underline{x} + \underline{c}^T\underline{x} \\ &\text{subject to} && \underline{a}_i^T \underline{x} = b_i \,,\ i \in I^{(k+1)} \,. \end{aligned} \tag{5.9}$$

Proof. We have already shown that

$$\bar{A}\underline{x}^{(k+1)} = \underline{b}^{(k+1)}$$

Also, $H\underline{x}^{(k+1)} = H(\underline{x}^{(k)} + \underline{s}^{(k)})$. Since $\underline{s}$ solves (5.6), we know that

$$H\underline{s}^{(k)} + \bar{A}^T \underline{\lambda} = -\underline{c} - H\underline{x}^{(k)} .$$

It follows that

$$H\underline{x}^{(k+1)} + \bar{A}^T \underline{\lambda} = -\underline{c} .$$

It then follows from (5.2) that $\underline{x}^{(k+1)}$ solves problem (5.9).

Theorem 5.6. If $\alpha(k) > 0$ then the iterates $x^{(k)}, x^{(k+1)}, \ldots, x^*$ are such that $f(\underline{x}) = \frac{1}{2} x^T H\underline{x} + \underline{c}^T\underline{x}$ is monotonically decreasing.

Proof. We will show that the directional derivative of f at $x^{(k)}$ in the direction $s^{(k)}$, $\underline{s}^T \nabla f(\underline{x}^{(k)})$, is negative.

$$\begin{aligned}
\underline{s}^T \nabla f(\underline{x}^{(k)}) &= \underline{s}^T (H\underline{x}^{(k)} + \underline{c}) \\
&= \underline{s}^T H\underline{x}^{(k)} + \underline{s}^T\underline{c} \\
&= -\underline{s}^T(-H\underline{x}^{(k)} - \underline{c}) \\
&= -\underline{s}^T(H\underline{s} + \bar{A}^T\underline{\lambda})
\end{aligned}$$

$$= -\underline{s}^T H\underline{s} - \underline{s}^T \bar{A}^T \underline{\lambda}$$

$$= -\underline{s}^T H\underline{s} - (\bar{A}\underline{s})^T \underline{\lambda}$$

$$= -\underline{s}^T H\underline{s}$$

$$< 0$$

if $s \neq 0$, since H is positive definite.

Suppose instead that $\underline{s} = 0$. If $\underline{\lambda} \geq 0$ then $\underline{x}^{(k)}$ is the optimal solution. If it is not then a constraint q is deleted from $I^{(k)}$. Suppose $\underline{s} = 0$ still solves the problem

$$\begin{aligned} &\text{Minimize} && \tfrac{1}{2}\underline{s}^T H\underline{s} + \underline{s}^T \underline{c}^{(k)} \\ &\text{subject to} && \underline{a}_i^T \underline{s} = 0 \ , \ i \in I^{(k)} \ . \end{aligned} \tag{5.10}$$

Then $x^{(k)}$ solves the problem

$$\begin{aligned} &\text{Minimize} && \tfrac{1}{2}\underline{x}^T H\underline{x} + \underline{c}^T\underline{x} \\ &\text{subject to} && A^* \underline{x} = \underline{b}^* \end{aligned}$$

where A^* and $\underline{b}^*$ are associated with the active constraints after constraint q is deleted.

Let $\bar{A}$ be the matrix of active constraints before q is deleted. We know from (5.2) that

$$H\underline{x}^{(k)} + \bar{A}^T \underline{\lambda} = -\underline{c}$$

and

$$H\underline{x}^{(k)} + A^{*T} \underline{\lambda}^* = -\underline{c} .$$

Therefore,

$$\bar{A}^T \underline{\lambda} = A^{*T} \underline{\lambda}^* .$$

This implies that

$$\lambda_1\underline{a}_1 + \ldots + \lambda_q\underline{a}_q + \ldots \lambda_\ell\underline{a}_\ell$$

$$= \lambda_1^* \underline{a}_1 + \ldots + \lambda_{q-1}^* \underline{a}_{q-1} + \lambda_{q+1}^* \underline{a}_{q+1} + \ldots \lambda_\ell^* \underline{a}_\ell .$$

Noting that $\lambda_q < 0$ and $a_q \neq 0$, this implies that $\underline{a}_q$ is a linear combination of the other active constraints. This contradicts Theorem 5.1 so we conclude that $\underline{s} = 0$ does not solve problem (5.10).

Therefore $\underline{s}^{(k)}$ in step 5 of the algorithm is nonzero and it follows from the argument at the beginning of the proof that $f(\underline{x})$ is monotonically decreasing.

Theorem 5.7 [7]. If $\alpha(k) > 0$ at each iteration, the active set algorithm converges to the unique optimal solution of problem (5.4).

Proof. There is a subsequence $\{x^{(k)}\}$ which solves problem (5.8). Only when $\alpha(k) < 1$ can $\underline{x}^{(k+1)}$ not solve such an equality problem. In this case a constraint is added to the active set. There are only a finite number of constraints so eventually we must run out of constraints to add to the active set. Examination of (5.7) shows that in this case $\alpha(k) = 1$ and the next iterate must then solve an equality problem.

The number of equality problems (and sets of active constraints) is finite. Obviously, the optimal solution to (5.4) must solve one of these equality problems (including possibly the problem with no constraints). $f(\underline{x})$ is monotonically decreasing so the algorithm never returns to an equality problem (or set of active constraints) that has already been solved. It follows that termination must occur.

Note that this proof breaks down when $\alpha(k) = 0$ and $f(\underline{x})$ does not decrease. In this case, the algorithm can cycle by returning to a previous active set. This is a result of degeneracy in the constraints. Theoretically, cycling can be avoided but, according to [7], there are some practical difficulties and most algorithms ignore degeneracy.

CHAPTER VI

APPLICATIONS OF QUADRATIC PROGRAMMING

This chapter presents two applications of quadratic programming. The first application shows how quadratic programming can be used to find the optimal search direction for a general programming problem. The second application shows how quadratic programming can be applied to optimal portfolio analysis.

A General Programming Algorithm [6]

Consider the general programming problem

$$\text{Minimize} \quad f(\underline{x})$$

$$\text{subject to} \quad g_i(\underline{x}) \geq 0 \ , \quad i = 1, 2, \ldots, n \ . \tag{6.1}$$

One approach to solving this problem is to employ the following iterative algorithm.

<u>Initialization</u>. Set $k = 0$. Select $\underline{x}_0$, the initial point, and λ_0, the initial stepsize.

<u>Main Step</u>

1. Let $\underline{x}_{k+1} = \underline{x}_k + \lambda_k \underline{u}_k$ where $\lambda_k \geq 0$ is the stepsize and $\underline{u}_k$ is the direction of steepest descent. By this we mean that moving in the direction $\underline{u}_k$ decreases $f(\underline{x})$ at the largest possible rate while still insuring that $\underline{x}_{k+1} = \underline{x}_k + \lambda_k \underline{u}_k$ is feasible if λ_k is small enough.
2. Correct $\underline{x}_{k+1}$ so that all the constraints are satisfied.

3. Check to see if the termination criteria is satisfied.
4. Adjust stepsize: $\lambda_k \to \lambda_{k+1}$.
5. $k \to k + 1$ and go to step 1.

Our primary interest in this type of algoritym is that $\underline{u}_k$, the direction of steepest descent, can be found by solving a quadratic programming problem.

Suppose that we wish to find the direction of steepest descent at a point $\underline{x}$. Let

$$\underline{t} = -\nabla f(\underline{x}) .$$

If there were no constraints, $\underline{t}$ would be the direction of steepest descent. Also, let $A_i = \nabla g_i(\underline{x})$ and note that only those constraints with $g_i(\underline{x}) = 0$, called binding constraints, actually effect the direction in which we may move.

If A_i is the gradient associated with one of the binding constraints, then A_i is normal to the tangent plane of the constraint surface at $\underline{x}$. If $\underline{x}$ lies on a constraint surface, the direction of steepest descent, $\underline{u}$, may lie on the constraint surface or it may point into the interior of the set. That is,

$$A_i^T \underline{u} \geq 0 . \tag{6.2}$$

The direction of steepest descent is given by the unit vector $\underline{u}$ which satisfies (6.2) and has the largest component along $\underline{t}$. To find $\underline{u}$, we must solve the problem

$$\begin{aligned} &\text{Maximize} && \underline{t}^T \underline{u} \\ &\text{subject to} && \underline{u}^T \underline{u} = 1 \\ & && A^T \underline{u} \geq 0 \end{aligned} \qquad (6.3)$$

where the columns of A are the gradients of the binding constraints evaluated at $\underline{x}$.

The Kuhn-Tucker conditions for this problem are

$$\begin{aligned} \underline{t} + A\underline{v} - \delta\underline{u} &= 0 \\ \underline{u}^T\underline{u} &= 1 \\ A^T\underline{u} &\geq 0 \\ \underline{v} &\geq 0 \\ \underline{v}^T A^T\underline{u} &= 0 \end{aligned} \qquad (6.4)$$

Assuming that the columns of A are linearly independent, these are necessary conditions for a vector $\underline{u}$ to solve (6.3). If the columns of A are not linearly independent, then we simply remove the redundent columns. Also note that $\underline{t}^T\underline{u}$, $A_i^T \underline{u}$, and $\underline{u}^T\underline{u} - 1$ are all convex and so conditions (6.4) are also sufficient.

Suppose $\underline{u}$ is the solution to (6.3). If $\underline{t}^T\underline{u} \leq 0$ then taking a step in the direction $\underline{u}$ does not decrease the value of $f(\underline{x})$ and it follows that $\underline{x}$ is a local minimum of problem (6.1). If $\underline{u}$ is the solution to (6.3) then $\underline{v}$ and δ exist which satisfy (6.4). Multiplying the first equation of (6.4) by $\underline{u}^T$ yields

$$\delta \underline{u}^T \underline{u} = \underline{t}^T \underline{u} + \underline{u}^T A \underline{v}$$

But $\underline{u}^T \underline{u} = 1$ and $\underline{u}^T A \underline{v} = 0$ so

$$\delta = \underline{t}^T \underline{u} .$$

Note from (6.4) that $\underline{u}$ is a vector with direction $\underline{t} + A\underline{v}$. Therefore, $\underline{t}^T \underline{t} + \underline{t}^T A \underline{v}$ has the same sign as $\delta = \underline{t}^T \underline{u}$.

Now we simplify the conditions in (6.4). The first equation becomes $\delta \underline{u} = \underline{t} + A\underline{v}$. Therefore, $A^T \underline{u} \geq 0$ becomes

$$\frac{1}{\delta} A^T \underline{t} + \frac{1}{\delta} A^T Av \geq 0 .$$

If $\delta > 0$, this becomes

$$A^T t \geq -M\underline{v}$$

where $M = A^T A$. It follows that

$$A^T \underline{t} = -M\underline{v} + \underline{z}$$

$$\underline{z} \geq 0 , \underline{v} \geq 0 . \qquad (6.5)$$

Also,

$$A^T \underline{u} = \frac{1}{\delta} A^T (\underline{t} + A\underline{v})$$

$$= \frac{1}{\delta} [A^T \underline{t} + M\underline{v}]$$

$$= \frac{1}{\delta} \underline{z} \ .$$

Therefore $\underline{v}^T A^T \underline{u} = 0$ implies

$$\underline{v}^T \underline{z} = 0 \ . \tag{6.6}$$

Notice that $\underline{x}^T M\underline{x} = \underline{x}^T A^T A\underline{x} = (A\underline{x})^T A\underline{x} \geq 0$. Therefore M is positive semidefinite. In addition, if the columns of A are linearly independent then M is positive definite.

Notice also that (6.5) and (6.6) are simply the Kuhn-Tucker conditions for the problem

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2} \underline{v}^T M\underline{v} + t^T A\underline{v} \\ &\text{subject to} \quad \underline{v} \geq 0 \ . \end{aligned} \tag{6.7}$$

Since M is positive definite, this problem has a unique solution which can be found by any of the methods previously described.

Suppose that (δ, u, v) is a solution to (6.4) with $\underline{\delta} = \underline{t}^T \underline{u} > 0$. Then $\underline{u}$ is the direction of steepest descent. But then $\underline{v}$ and $\underline{z} = A^T \underline{t} + M\underline{v}$ are a solution of (6.5) and (6.6). Therefore, $\underline{v}$ is the unique optimal solution to (6.7).

Therefore, if we wish to find $\underline{u}$, we solve problem (6.7) to find $\underline{v}$. If $\underline{t}^T \underline{t} + \underline{t}^T A\underline{v} \leq 0$, then $\underline{t}^T \underline{u} \leq 0$ and $\underline{x}$ must be a local minimum. If $\underline{t}^T \underline{t} + \underline{t}^T A\underline{v} > 0$, then the unit vector $\underline{u} = \frac{1}{\delta}(\underline{t} + A\underline{v})$ satisfies $\underline{t}^T \underline{u} > 0$. It is clear that $\underline{t} + A\underline{v} - \delta\underline{u} = 0$, $\underline{u}^T \underline{u} = 1$, and $\underline{v} \geq 0$. In addition,

$$A^T\underline{u} = \frac{1}{\delta}(A^T\underline{t} + M\underline{v}) = \underline{z}/\delta \geq 0$$

since $\underline{z} \geq 0$ and $\delta > 0$. Finally,

$$\underline{v}^T A^T\underline{u} = \frac{1}{\delta}(\underline{v}T\ A^T\underline{t} + \underline{v}^T M\underline{v})$$

$$= \underline{v}^T\underline{z}$$

$$= 0 \ .$$

Therefore, $\underline{u} = \underline{t} + A\underline{v}$ is the direction of steepest descent.

Example 6.1. Suppose we wish to find $\underline{u}$ when $\underline{x} = [1 \quad 5]^T$ in the problem

$$\begin{aligned} \text{Minimize} \quad & 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \\ \text{subject to} \quad & x_1 + x_2 \leq 6 \\ & 2x_1^2 - x_2 \leq 0 \\ & -x_1 \leq 0 \\ & -x_2 \leq 0 \ . \end{aligned}$$

Then,

$$\nabla f(\underline{x}) = [4x_1 - 2x_2 - 4 \ , \ 4x_2 - 2x_1 - 6]^T \ .$$

So when $\underline{x} = [1 \quad 5]^T$,

$$\underline{t} = [10 \quad -12]^T .$$

Note that when $\underline{x} = [1 \quad 5]^T$, the first constraint is the only binding constraint

$$A_1 = \nabla(x_1 + 5x_2 - 5) = [1 \quad 5]^T .$$

Then,

$$M = [1 \quad 5] \begin{bmatrix} 1 \\ 5 \end{bmatrix} = 26 .$$

In this case, problem (6.7) becomes simply

$$\begin{array}{ll} \text{Minimize} & 13v_1^2 - 50v_1 \\ \text{subject to} & v_1 \geq 0 . \end{array}$$

Then, $v_1 = 25/13 \cong 1.92$. So the direction of steepest descent is

$$\underline{t} + A\underline{v} \cong \begin{bmatrix} 10.92 \\ -2.4 \end{bmatrix} .$$

Portfolio Analysis

Suppose a security portfolio is to be formed by investing in some combination of n different stocks and bonds. Let x_i be the fraction of the total investment to be invested in security i . We will assume that the rate of return on the ith security is normally distributed with rate of return μ_i and variance σ_{ii} . The variance is a measure of the risk involved with security i ; a risky investment would have large variance, while a safe investment would have a small variance.

The portfolio should decrease the risk by "hedging"; that is, trying to offset unfavorable returns on security i by a favorable return on j . The covariance σ_{ij} measures the correlation between the rates of return of securities i and j . We assume that the investor has established values of μ_i, σ_{ii}, and σ_{ij} for all the stocks and bonds. This might be done on the basis of historical data, the beliefs of the investor, or some combination of the two.

Let $\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Then, we know that $\sum_i x_i = 1$. The expected return on the entire portfolio would be

$$\mu = \sum_{i=1}^{n} \mu_i x_i \ .$$

If r_i is the actual return on security i then the variance on the total return is

$$\sigma^2 = E[(\sum_{i=1}^{n} (r_i - \mu_i)x_i)^2]$$

$$= E[\sum_{i=1}^{n} \sum_{j=1}^{n} (r_i - \mu_i)(r_j - \mu_j)\ x_i x_j]$$

$$= \underline{x}^T H\underline{x}$$

where

$$H = \begin{bmatrix} \sigma_{11} & \sigma_{12} & & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & & \sigma_{nn} \end{bmatrix}$$

Note that

$$\begin{aligned} \underline{x}^T H\underline{x} &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} \\ &= E[(\sum_{i=1}^{n} x_i(r_i - \mu_i))^2] \\ &\geq 0 . \end{aligned}$$

Therefore, H is positive semidefinite.

We assume that among alternatives with equal rates of return an investor will prefer the one with the smallest variance. Thus for a given return μ we would like to solve the problem

$$\begin{aligned} \text{Minimize} \quad & \underline{x}^T H\underline{x} = \sigma^2 \\ \text{subject to} \quad & \sum_{i=1}^{n} \mu_i x_i = \mu \\ & \sum_{i=1}^{n} x_i = 1 \\ & \underline{x} \geq 0 . \end{aligned}$$

Note that this is a convex quadratic programming problem.

Suppose we perform this computation and find that the minimum variance is σ_n^2 . Obviously, among all such portfolio with variance σ_n^2 , the investor would prefer the one which maximizes the rate of return. This suggests the concept of efficient pairs (μ, σ^2) . (μ, σ^2) is said to be an efficient pair if σ^2 is the minimum variance given μ and μ is the maximum expected return given σ^2 .

The set of efficient pairs can be found by solving the following convex quadratic programming problem for all $B \geq 0$:

$$\text{Maximize} \quad f(\underline{x}) = \mu - B\sigma^2 = \sum_{i=1}^{n} \mu_i x_i - B\underline{x}^T H\underline{x}$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1$$

$$x \geq 0 \ .$$

The proof of this goes as follows. Let $\beta \geq 0$ be given. Suppose that f is maximized for this given β at the point $\underline{x}^*$. Let $\mu^* = \sum_{i=1}^{n} \mu_i x_i^*$. $f(\underline{x}^*) \geq f(\underline{x})$ for all $\underline{x} \geq 0$, $\sum_{i=1}^{n} x_i = 1$. Therefore, $f(\underline{x}^*) \geq f(\underline{x})$ if we also require $\sum_{i=1}^{n} \mu_i x_i = \mu^*$. Therefore, $\underline{x}^*$ also solves the problem

$$\text{Maximize} \quad \mu^* - \beta\underline{x}^T H\underline{x}$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1$$

$$\underline{x} \geq 0$$

$$\sum_{i=1}^{n} \mu_i x_i = \mu^* \ .$$

Since $\beta \underline{x}^T H\underline{x} \geq 0$, this means that $\underline{x}^*$ also minimizes $\sigma^2 = \underline{x}^T H\underline{x}$ subject to the constraints above.

On the other hand, suppose we set $\sigma^2 = (\underline{x}^*)^T H\underline{x}^*$. Since $f(\underline{x}^*) \geq f(\underline{x})$ for all $\underline{x} \geq 0$, $\sum_{i=1}^{n} x_i = 1$, we also have that $f(\underline{x}^*) \geq f(\underline{x})$ for all $\underline{x} \geq 0$, $\sum_{i=1}^{n} x_i = 1$, and $\underline{x}^T H\underline{x} = \sigma^2$. Therefore,

$$\mu^* - \beta\sigma^2 \geq \sum_{i=1}^{n} \mu_i x_i - \beta\sigma^2 .$$

It follows that $\mu^* \geq \sum_{i=1}^{n} \mu_i x_i$ for all $\underline{x} \geq 0$, $\sum_{i=1}^{n} x_i = 1$. Therefore $(\mu^*, \sigma^2 = (x^*)^T H\underline{x}^*)$ is an efficient pair.

In practice, we would assign a specific value to β and then solve the problem

$$\begin{aligned} \text{Maximize} \quad & \sum_{i=1}^{n} \mu_i x_i - \beta\, \underline{x}^T H\underline{x} \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = 1 \\ & \underline{x} \geq 0 . \end{aligned}$$

Note that β can be interpreted as a measure of the "risk aversion" of the investor. If $\beta = 0$, we simply maximize the expected return and ignore risk. As $\beta \to \infty$, the problem becomes one of minimizing $\underline{x}^T H\underline{x}$ or minimizing the risk.

This approach to portfolio analysis was first suggested by Markowitz [11], [12]. It is also discussed in [3], [9], and [12].

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Bazaraa, Mokhtor S. and C.M. Shetty, Nonlinear Programming Theory and Algorithms. New York: John Wiley & Sons, 1979.

2. Boot, John C.G., Quadratic Programming. Chicago: Rand McNally & Company, 1964.

3. Cooper, Keon, Applied Nonlinear Programming for Engineers and Scientists. New York: Aloray Publishers, 1974.

4. Daya, Mohammad Ben and C.M. Shetty, "Polynomial Barrier Function Algorithms for Convex Quadratic Programming." Report Series J88-5, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

5. _____, "Polynomial Barrier Function Algorithms for Linear Programming." Report Series J88-4, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

6. Dennis, Jack B., Mathematical Programming and Electrical Networks. New York: John Wiley & Son, 1959.

7. Fletcher, R., Practical Methods of Optimization. New York: John Wiley & Son, 1987.

8. Gass, Saul I., Linear Programming. New York: McGraw-Hill, Inc., 1985.

9. Hadley, G., Nonlinear and Dynamic Programmg. Reading, Massachusetts: Wesley Publishing Company, Inc., 1964,

10. Markowitz, H.M., "Portfolio Selection." Journal of Finance, 7(1952), 77-91.

11. _____, Portofolio Selection, Cowles Foundation Monograph, No. 16. New York: John Wiley & Son, Inc., 1959.

12. Walsh, G.R., Methods of Optimizations. New York: John Wiley & Sons, 1975.

APPENDIX

APPENDIX

Linear Programming

In this appendix, we briefly discuss linear programming and the simplex method. For a more comprehensive treatment refer to [8].

The linear programming problem is

$$\begin{aligned} &\text{Minimize} && \underline{c}^T\underline{x} \\ &\text{subject to} && A\underline{x} = \underline{b} \\ & && \underline{x} \geq 0 \end{aligned} \tag{A.1}$$

where $\underline{c}$ and $\underline{x}$ are n-vectors, A is an $m \times n$ matrix and $\underline{b}$ is an m-vector. The set of feasible solution to (A.1) is a convex polyhedron. It is well known that the solution to (A.1) is a vertex of this polyhedron. The simplex method solves the linear programming problem by moving from vertex to vertex until the optimal solution is found.

We assume that there are m columns of A which can be explicitly arranged to form a unit matrix of rank m. Let $P_1, \ldots, P_n$ denote the column vectors of A and let $P_1, \ldots, P_m$ denote the unit vectors previously mentioned. Then

$$P_j = \sum_{i=1}^{m} x_{ij} P_i , \qquad j = 1, \ldots, n .$$

We define

$$z_j = \sum_{i=1}^{m} x_{ij}\, c_i\ , \qquad j = 1,\ldots, n\ .$$

Let the present solution be

$$\underline{x}_0 = (x_{10}, x_{20}, \ldots, x_{m0})\ .$$

Of course,

$$x_{10}P_1 + x_{20}P_2 + \ldots + x_{m0}P_m = \underline{b}\ .$$

The initial tableau is set up as follows:

Basis	$\underline{c}$	P_0	P_1	P_2	$\cdots$	P_m	P_{m+1}	$\cdots$	P_m
P_1	c_1	x_{10}	1	0		0	$x_{1,m+1}$		x_{1m}
P_2	c_2	x_{20}	0	1		0	$x_{2,m+1}$		x_{2m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$
P_m	c_m	x_{m0}	0	0		1	x_{m+m+1}		x_{mm}
		$z_0=\underline{c}^T\underline{x}_0$	0	0		0	$z_{m+1}-c_{m+1}$		z_m-c_m

The simplex method then proceeds:

1. If $z_j - c_j \leq 0$ for $j = 1,\ldots, n$ then the optimal solution is $\hat{\underline{x}}$ where

$$\hat{x}_i = x_{i0}\ , \text{ if } P_i \text{ is in the final basis}$$

and

$$\hat{x}_i = 0\ , \text{ if } P_i \text{ is not in the final basis.}$$

2. If not, let k be the index that corresponds to

$$\max_j (z_j - c_j) .$$

Then P_k will enter the basis.

3. Let

$$\theta = \min_i \{\frac{x_{i0}}{x_{ik}} : x_{ik} > 0\} = \frac{x_{\ell 0}}{x_{\ell k}} .$$

If $x_{ij} \leq 0$ for all i then the solution is unbounded.

4. P_ℓ leaves the basis. The new solution is computed and each non-basic vector is computed in terms of the basic vectors by normal Gauss-Jordan elimination.

5. Return to step 1. Continue until an optimal solution is found.

Example A.1.

Minimize $\quad -2x_1 - 3x_2$

subject to

$$2x_1 + x_2 \leq 2$$

$$3x_1 + 2x_2 \leq 4$$

$$x_1 , x_2 \geq 0 .$$

We introduce slack variables x_3 and x_4 and rewrite the problem as

Minimize $-2x_1 - 3x_2$

subject to $2x_1 + x_2 + x_3 = 2$

$3x_1 + 2x_2 + x_4 = 4$

$x_1, x_2, x_3, x_4 \geq 0$.

Iteration 1.

		$\underline{c}$ =	-2	-3	0	0	
Basis	c	P_0	x_1	x_2	x_3	x_4	θ
x_3	0	2	2	1	1	0	2
x_4	0	4	3	2	0	1	2
		0	2	3	0	0	

Iteration 2.

Basis	c	P_0	x_1	x_2	x_3	x_4
x_2	-3	2	2	1	1	0
x_3	0	0	-1	0	-2	1
		-6	-4	0	-3	0

Therefore,

$x_1 = 0$, $x_2 = 2$

is the optimal solution.

Of course in many problems we will not have such a convenient basis. In this case, we will introduce artificial variables to provide a basis consisting of unit vectors. We assign these variables a large cost, which we call w, to insure that they do not appear in the final answer. In our tableau, we add another row to keep track of the w term in $z_j - c_j$. We look at this row to find $\max_j(z_j - c_j)$ and to determine which vector will enter the basis.

Once an artificial vector leaves the basis, we no longer keep track of it. Once the artificial vectors leave the basis, we continue with the normal simplex procedure. If an artificial variable with a non-zero value remains in the optimal solution, then the problem is not feasible.

<u>Example A.2</u>.

$$\begin{array}{ll} \text{Minimize} & -x_1 - 2x_2 - x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 2 \\ & 2x_1 + x_2 \qquad = 3 \\ & 3x_1 + 2x_2 \qquad = 2 \\ & \quad x_1, x_2, x_3 \geq 0 . \end{array}$$

We introduce artificial variables x_4 and x_5 and proceed as follows.

Iteration 1:

		$\underline{c}$ =	-1	-2	-1	w	w	
Basis	$\underline{c}$	P_0	x_1	x_2	x_3	x_4	x_5	θ
x_3	-1	2	1	1	1	0	0	2
x_4	w	3	2	1	0	1	0	3/2
x_5	w	2	3	2	0	0	1	2/3
		-2	0	0	0	0	0	
		5w	5w	3w	0	0	0	

Iteration 2:

Basis	$\underline{c}$	P_0	x_1	x_2	x_3	x_4
x_3	-1	4/3	0	1/3	1	0
x_4	w	5/3	0	-1/3	0	1
x_5	-1	2/3	1	2/3	0	0
		-2	0	1	0	0
		5/3w	0	-1/3w	0	0

This is the optimal solution. $x_4 = 5/3 > 0$ is still basic so the problem is not feasible.

We call the solution set at each iteration a basic feasible solution. If one of the basic variables has value zero then we say that the solution is degenerate. In this case, it is possible that the simplex method will cycle. This means that it is possible to select a sequence of bases that is repeated without every reaching the optimal solution. There are several techniques for avoiding this difficulty (see [8]) .

VITA

James R. Williams was born in Atlanta, Georgia on November 23, 1965. He was educated in Duluth, Georgia and graduated from Duluth High School in June 1983. He received a Bachelor of Science degree in Mathematics from Shorter College in May 1987. He accepted a teaching assistantship at The University of Tennessee, Knoxville in September 1987 and began study towards a Master of Science degree in Mathematics.