International Journal of Inventive Engineering and Sciences (IJIES)
ISSN: 2319–9598, Volume-4 Issue-5, March 2017

Stability and Efficiency of the Positive Definite Quadratic Programming Algorithms

Awatif M .A .Elsiddieg

Abstract: ***In this paper we introduce some stable and efficiency algorithms for the positive definite quadratic programming . Sections (1), introduce matrix factorizations QR factorization ,orthogonal transformation using Householder matrices , which leads to our main work. In section(2) general consideration is given. In section (3) we introduce the basic concepts methods linear equality and inequality constraints that leads to our methods. In section (4) we give some of the stable and efficiency algorithms for positive quadratic programming only using KKT-conditions. We conclude our paper by showing that there are stable and efficient methods for indefinite programming as the extended Dantzig Wolfe method[20].***

Keywords: KKT-conditions, QR factorization, active set methods, penalty and barrier functions, complementrity.

I. INTRODUCTION

Section (1):

A. Basic Concepts:

In this section we lay down the theoretical background needed in presenting the work of the following sections. We introduce some basic matrix factorizations that we use in our methods.

B. Matrix Factorizations:

Definition: (Positive Definite Matrices**):** A symmetric matrix A is positive definite if and only if

$$\underline{x}^T A \underline{x} > 0$$

For all non-zero vectors $\underline{x}$. This can be a difficult condition to verify, but there are equivalent definitions that are sometimes more practical. For example, A will be positive definite if all of the eigen-values are positive. Also if Gaussian elimination is applied to A without pivoting to transform A to upper triangular form:

$$A \to U = \begin{pmatrix} u_{1,1} & & u_{1,n} \\ & u_{2,2} & u_{2,n} \\ \bigcirc & & u_{n,n} \end{pmatrix}$$

and if $u_{i,i} > 0$ for all i, then A is positive definite similarly a symmetric matrix A is:

* Positive semi-definite if $\underline{x}^T A \underline{x} \geq 0$ for all $\underline{x} \neq \underline{0}$ (or equivalently, all the eigen values of A are non-negative).

* Negative definite if $\underline{x}^T A \underline{x} < 0$ for all $\underline{x}$ (all the eigenvalues of A are negative)

* Negative semi-definite if $\underline{x}^T A \underline{x} \leq 0$ for all $\underline{x}$ (all the eigen-values of A are non-positive).

* indefinite of $\underline{x}^T A \underline{x}$ can take both positive and negative values (A has both positive and negative eigenvalues).

Example: Consider the matrix:

$$A = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 5 & 2 \\ 0 & 2 & 6 \end{pmatrix}$$

Then

$$\underline{x}^T A \underline{x} = 4x_1^2 + 2x_1x_2 + 5x_2^2 + 4x_2x_3 + 6x_3^2$$

$$= 3x_1^2 + 3x_2^2 + 2x_3^2(x_1 + x_2)^2 + (x_2 + 2x_3)^2 > 0$$

If $\underline{x} \neq \underline{0}$. So A is positive definite, if the Gaussian elimination is up lied to A, then:

$$A \to U = \begin{pmatrix} 4 & 1 & 0 \\ 0 & 4.75 & 2 \\ 0 & 0 & 5.1579 \end{pmatrix}$$

All the diagonal elements are positive the eigen values of A;

(2.8549, 4.4760, 7.669) are all positive

The matrix: $B = \begin{pmatrix} 4 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 4/5 \end{pmatrix}$

is positive semi-definite but not positive definite

$$\underline{x}^T B \underline{x} = 4x_1^2 - 4x_1x_2 + 6x_2^2 + 4x_2x_3 + \frac{4}{5}x_3^2$$

$$= (2x_1 - x_2)^2 + 5\left(x_2 + \frac{2}{5}x_3\right)^2 \geq 0$$

If $\underline{x} = \left(\frac{1}{5}, -\frac{2}{5}, 1\right)^T$, then $\underline{x}^T B \underline{x} = 0$

When Gaussian elimination[25] is applied to B

Revised Version Manuscript Received on March 15, 2017.

Awatif M. A. Elsiddieg, Department of Mathematical, Faculty of Science& Humanities Studies, Prince Sattam Bin Abdul-Aziz University, Kingdom of Saudi Arabia. E-mail: wigdan1965@hotmail.com

$$B \to U = \begin{pmatrix} 4 & 2 & 0 \\ 0 & 5 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$ and the diagonal entries in U are the nonnegative $U_{3,3} = 0$

The eigen values of B are: (0, 3.1284, 7.6716)

The matrix C: $$C = \begin{pmatrix} 3 & 5 & 0 \\ 5 & 4 & 7 \\ 0 & 7 & 2 \end{pmatrix}$$

is indefinite, For $\underline{x} = (1,0,0)^T$, $\underline{x}^T C \underline{x} = 3 > 0$, but for $\underline{x} = (1,-1,0)^T$, $\underline{x}^T C \underline{x} = -3 < 0$ Gaussian elimination applied to C produces

$$C \to U == \begin{pmatrix} 3 & 5 & 0 \\ 0 & -4.333 & 7 \\ 0 & 0 & 13.3077 \end{pmatrix}$$ and the matrix U

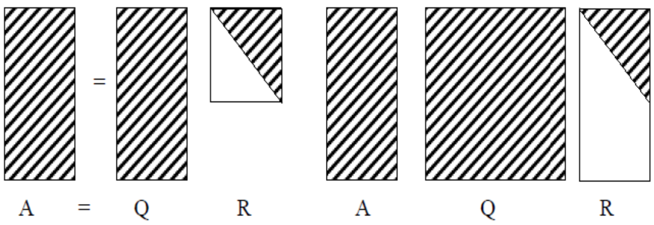
has both positive and negative diagonal entries. The eigen values of C are also both positive and negative (-5.4885, 2.6662, 11.8223).

II. AN ORTHOGONAL MATRIX FACTORIZATION:

The matrix factorization is based on entirely different principles, and applies in a quite different setting. It uses orthogonal transformations and can be applied to any matrix. The matrix A need not have an inverse; in fact, the matrix need not be square. It is typically used when A is rectangular, in particular, when A has more rows than columns. Its most common usage is in the context of least-squares problems, but it is frequently used to represent the linear constraints in an optimization problem. It is called a QR factorization because it represent A as: A = QR

where Q is an orthogonal matrix (i.e. $Q^TQ =$ I) and R is an upper triangular (or "right" triangular) matrix. There is a slight ambiguity in the definition of the QR factorization.

The first chooses Q to be the same size as A and then R is *nxn* ; the second chooses R to be the same size as A and, then Q is *mxm* the figure



The first form is all that normally required to solve least squares problems, the second form is often more useful when solving constrained optimization problems. Orthogonal matrix factorizations are most often used when solving least squares problems. To see why, consider a least squares problem written in the form:

$$\min imize \ \|A\underline{x} - \underline{b}\|_2^2$$

where A is an mxn matrix, $m \geq n$. least squares problems cannot be solved by applying elimination to the matrix A. When linear systems of equations are solved, the techniques of elimination result in a sequence of equivalent linear systems. Elimination is not applied to least squares problems because the techniques of elimination do not leave the least-squares problem un changed. However, if *P* is an orthogonal matrix so that $P^T P = I$, then

$$\|Py\|_2^2 = (Py)^T (Py) = y^T P^T py = yy^T = \|y\|_2^2$$

That is an orthogonal transformation does not effect the 2-norm of a vector. Hence

$$\|A\underline{x} - b\|_2^2 = \|P(A\underline{x} - \underline{b})\|_2^2$$

and so orthogonal transformation can be used to generate a sequence of equivalent least squares problems.

Example: $\min\limits_{\underline{x}} imize \ \ \|A\underline{x} - \underline{b}\|_2^2$

With

$$A = \begin{bmatrix} 5 & 2 & 1 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} 13 \\ 17 \\ 4 \\ 12 \end{bmatrix}$$

This is equivalent to solving

$$\min\limits_{\underline{x}} imize \ (5x_1 + 2x_2 + x_3 - 13)^2 + (3x_2 + 4x_3 - 17)^2 + (2x_3 - 4)^2 + (12)^2$$

The solution can be obtained via back substitution. This is a sum of squared terms and the smallest value each term can achieve is zero. The term $(12)^2$ remains unchanged no matter what the values of the x_i are.

The other three terms can be made equal zero by solving the triangular system of equations,

$$5x_1 + 2x_2 + x_3 = 13$$

$$3x_2 + 4x_3 = 17$$

$$2x_3 = 4$$

The solution is $\underline{x} = (x_1, x_2, x_3)^T = (1,3,2)^T$ if the matrix A has been factored as $A = QR$

Where $Q^T Q = I$ and R is an upper triangular, then

$$\|A\underline{x} - \underline{b}\|_2^2 = \|QR\underline{x} - \underline{b}\|_2^2$$

$$= \|R\underline{x} - Q^T \underline{b}\|_2^2$$

Hence the QR factorization allows us to transform a general least-squares problems to a triangular – least – squares problems that can be solved via back substitution.[15]

Example: (Generating A Basis Matrix Using the QR Factorization)

Consider the matrix: $A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

An orthogonal factorization of A^T yields:

$$Q = \begin{bmatrix} -\sqrt{2/2} & 0 & -\frac{1}{2} & -\frac{1}{2} \\ \sqrt{2/2} & 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -\sqrt{2/2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\sqrt{2/2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$R = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & -\sqrt{2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Hence $$Q_2 = Z = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

is a basis for the null space of A .

The QR factorization method has the important advantage that the basis Z can be formed in a numerically stable manner. Moreover, computation performed with respect to the resulting basis Z are numerically stable.

The LDL^T and Cholesky Factorization:-

These two matrix factorizations are primarily of interest when factoring a positive definite matrix A, although variants of the LDL^T factorization can be applied to more general symmetric matrices. They are used in nonlinear optimization problems to represent the Hessian matrix of the objective function. If A is symmetric and positive definite, it can be shown that Gaussian elimination can always be applied without partial pivoting, with no danger of the method trying to divide by zero, and with no danger of near-zero pivots that can lead to numerical difficulties.

If no row interchanges are used the LU factorization takes the form: $A = LU$

The first two factorizations are obtained by manipulating this formula. Let D be the diagonal matrix whose entries are the diagonal entries of $U : d_{i,i} = u_{i,i}$. Then define $\hat{U} = D^{-1}U$ so that $D\hat{U} = U$. hence $A = LD\hat{U}$. If A is positive definite, then it is also symmetric, so that:

$$A^T = \hat{U}^T D^T L^T = \hat{U}^T D L^T = A = LD\hat{U}$$

It is then easy to verify that $\hat{U} = L^T$, so that

$$A = LDL^T$$

This is the first of the new factorizations, a factorization of A into the product of a lower triangular matrix, a diagonal matrix, and the transpose of the lower triangular matrix.

Slightly more can be deduced. If A is positive definite then $\underline{x}^T A \underline{x} > 0$ for all $\underline{x} \neq 0$.

Using the factorization:

$$0 < \underline{x}^T A \underline{x} = \underline{x}^T LDL^T \underline{x} = \left(L^T \underline{x}\right)^T D \left(L^T \underline{x}\right)$$

$$\equiv y^T D y$$

where $y = L^T\underline{x}$. Since L is nonsingular (it is triangular and all of its diagonal entries are equal to 1), $y \neq 0$ if and only if $\underline{x} \neq 0$

Hence $0 < y^T D y = \sum_i d_{i,i} y_i^2$

for all $y \neq 0$. this can only happen if $d_{i,i} > 0$ for all i. Hence D is a diagonal matrix with positive diagonal entries. It should be noted that the reverse is also true. i.e. if A can be represented as $A = LDL^T$ where D has positive diagonal entries, then A must be symmetric and positive definite. If we discover that $d_{i,i} \leq 0$ at some stage during the computation of the factorization, then A is not positive definite. This property will be useful when we apply Newton's method to multidimensional optimization problems.

Example: $A = LDL^T$. To illustrate this factorization.

Consider the positive definite matrix

$$A = \begin{pmatrix} 1.4 & -0.2 & 0.1 \\ -0.2 & 1.5 & -0.3 \\ 0.1 & -0.3 & 1.8 \end{pmatrix}$$

Then A can be represented as LDL^T where

$$L = \begin{pmatrix} 1 & 0 & 0 \\ -0.1429 & 1 & 0 \\ 0.0714 & -0.1942 & 1 \end{pmatrix}$$ and

$$D=\begin{pmatrix}1.4000 & 0 & 0\\ 0 & 1.4714 & 0\\ 0 & 0 & 1.7374\end{pmatrix}$$

The second factorization is obtained easily from the first. Since D has positive diagonal entries we can write

$$D=\hat{D}\hat{D}$$

Where $\hat{D}$ is a diagonal matrix with $\hat{d}_{i,i}=\sqrt{d_{i,i}}$. If we then define: $\hat{L}=L\hat{D}$

Then $\hat{L}$ is also a lower triangular matrix and $A=\hat{L}\hat{L}^T$

The $\hat{D}$ is often omitted, and we simply write A = LL^T. This is the second of the factorization, also referred to as a Cholesky factorization.

Orthogonal Transformations Using Householder Matrices:-

Introduction:-

We look at a special class of orthogonal matrices known as a Householder matrices. We show how to construct a Householder matrix that will transform a given vector to a simpler form. With this construction as a tool, we look at a transformation of a given $m\times n$ matrix A, ($m\geq n$) has rank n.

Householder Transformations:-

Let $\underline{u}\in IR^n$ be non-zero. An $n\times n$ matrix H of the form:

$$H=I-\frac{2\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}} \qquad (1.1)$$

is known as a Householder transformation. When a vector $\underline{x}$ is multiplied by H, it is reflected in the hyper-plane span $\{\underline{u}\}^{\perp}$. Householder matrices are symmetric and orthogonal. They are important because of their ability to zero specified entries in a matrix or vector. In particular, given any non-zero $\underline{x}\in IR^n$, it is easy to construct in (1.1) such that $H\underline{x}$ is a multiple of $\underline{e}_1$, the first column of I. Noting that:

$$H\underline{x}=\left(I-\frac{2\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}\right)\underline{x}=\underline{x}-\frac{2\underline{u}^T x\underline{u}}{\underline{u}^T\underline{u}},\ \underline{u}^T\underline{u}=I \qquad (1.2)$$

Householder Reflection:-

Definition: If $\underline{u}\in IR^n$ and $\underline{u}\neq\underline{0}$, then the matrix of the form: $H=I-\frac{2\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}$ is called the Householder matrix or Householder reflection and the vector $\underline{u}$ is called the Householder vector.

Proposition:

The Householder matrix is (1) symmetric and (2) orthogonal (3) the Householder reflection reflects every vector $\underline{x}\in IR^n$ in the hyper-plane span $\{\underline{u}\}^{\perp}$

Proof: Indeed,

1) $H^T=I^T-2\frac{\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}=I-2\frac{\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}=H$

2) $HH^T=H^TH=\left(I-2\frac{\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}\right)^2$

$$=I-4\frac{\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}+\frac{4\underline{u}\left(\underline{u}^T\underline{u}\right)\underline{u}^T}{\left(\underline{u}^T\underline{u}\right)\left(\underline{u}^T\underline{u}\right)}=I$$

To prove the third part of the assertion, we choose on the hyper plane span $\{\underline{u}\}^{\perp}$ an orthogonal basis $\{\underline{a}_1,\ldots,\underline{a}_{n-1}\}$. Hence $\underline{u}\perp\underline{a}_i$, *(i = 1; n – 1)*

$\underline{u}^T\underline{a}_i=\underline{0}$, *(i = 1; n – 1)* and

$\underline{x}=\beta\underline{u}+\beta_1\underline{a}_1+....+\beta_{n-1}\underline{a}_{n-1}$, then

$$H\underline{x}=H(\beta\underline{u})+H(\beta_1\underline{a}_1)+\ldots+H(\beta_{n-1}\underline{a}_{n-1})$$

$$=\beta\left(I-2\frac{\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}\right)\underline{u}+\beta_1\left(I\frac{-2\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}\right)\underline{a}_1+\ldots+\beta_{n-1}\left(I-\frac{2\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}\right)\underline{a}_{n-1}$$

$$=\beta\left(\underline{u}\frac{-2\underline{u}\left(\underline{u}^T\underline{u}\right)}{\underline{u}^T\underline{u}}\right)+\beta_1\left(\underline{a}_1\frac{-2\underline{u}(\underline{u}^T\underline{a}_1)}{\underline{u}^T\underline{u}}\right)+\ldots+\beta_{n-1}\left(\underline{a}_1\frac{-2\underline{u}(\underline{u}^T\underline{a}_1}{\underline{u}^T\underline{u}}\right)$$

$$=-\beta\underline{u}+\beta_1\underline{a}_1+\ldots+\beta_{n-1}a_{n-1}$$

i.e., the vectors $\underline{x}$ and $H\underline{x}$, have onto the hyper-plane span $\{\underline{u}\}^{\perp}$ the same orthogonal projection $\beta_1\underline{a}_1+....+\beta_n\underline{a}_{n-1}$

Proposition: If $\underline{x}\in IR^n$ and $\underline{u}=\underline{x}\pm\|\underline{x}\|_2\underline{e}_1$, then the vector $H\underline{x}$, where H is the Householder matrix denoted by (1), has the same direction as $\underline{e}_1$ i.e., the Householder reflection H applied to the vector $\underline{x}$ annihilates all but the first component of the vector $\underline{x}$.

Proof: Our aim is to determine for a nonzero vector $\underline{x}$ the Householder vector $\underline{u}$ so that $H\underline{x} \in$ span $\{\underline{e}_1\}$ since

$$H\underline{x} = \left(I - \frac{2\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}}\right)\underline{x} = \underline{x} - \frac{2\underline{u}\left(\underline{u}^T\underline{x}\right)}{\underline{u}^T\underline{u}} = \underline{x} - \frac{2\underline{u}^T\underline{x}}{\underline{u}^T\underline{u}}\underline{u}$$

and $H\underline{x} \in$ span $\{\underline{e}_1\}$, then $\underline{u} \in$ span $\{\underline{x}, \underline{e}_1\}$. by choosing

$$\underline{u} = \underline{x} + \alpha\underline{e}_1$$

Obtain that:

$$\underline{u}^T\underline{x} = x^T\underline{x} + \alpha\underline{e}_1\underline{x} = \underline{x}^T\underline{x} + \alpha\Im_1$$

and $H\underline{x} = \underline{x} - 2\frac{\underline{u}^T\underline{x}}{\underline{u}^T\underline{u}}\underline{u} = \underline{x} - \frac{2\underline{x}^T\underline{x} + 2\alpha\Im_1}{\underline{x}^T\underline{x} + 2\alpha\Im_1 + \alpha^2}(\underline{x} + \alpha\underline{e}_1)$

$$= 1 - \frac{2\underline{x}^T\underline{x} + 2\alpha\Im_1}{\underline{x}^T\underline{x} + 2\alpha\Im_1 + \alpha^2} = 0$$

$$\Leftrightarrow \underline{x}^T\underline{x} + 2\alpha\Im + \alpha^2 - 2\underline{x}^T\underline{x} - 2\alpha\Im_1 = 0$$

$$\Leftrightarrow \underline{-x}^Tx + \alpha^2 \Leftrightarrow \|\underline{x}\|_2^2 = \alpha$$

For this choose $= \pm\|\underline{x}\|_2$ we have $\underline{u} = \underline{x} \pm \|\underline{x}\|_2\underline{e}_1$

and: $$H\underline{x} = -2\alpha\frac{\underline{u}^T\underline{x}}{\underline{u}^T\underline{u}}\underline{e}_1 = \frac{-2\alpha\underline{x}^T\underline{x} \pm \alpha\Im_1}{\underline{x}^T\underline{x} \pm 2\alpha\Im_1 + \underline{x}^T\underline{x}}\underline{e}_1$$

$$= -\alpha\underline{e}_1 = \pm\|\underline{x}\|_2\underline{e}_1$$

Example: Let $\underline{x} = [2\,6 - 3]^T$. Find the Householder vector $\underline{u}$ and according to it the Householder transformation annihilates the two last coordinates of the vector $\underline{x}$.

Solution: By proposition above we compute:

$\underline{u} = \underline{x} \pm \|\underline{x}\|_2 \quad \underline{e}_1 = [2\ 6\ -3]^T \pm 7\underline{e}_1$ Choose the sign plus for coefficient of $\underline{e}_1$ and we obtain $\underline{u} = [9\,6 - 3]^T$. Find the Householder matrix H that depends only on the direction of $\underline{u}$

$$H = I\frac{-2\underline{u}\underline{u}^T}{\underline{u}^T\underline{u}} = I - \frac{2}{14}\begin{bmatrix}3\\2\\-1\end{bmatrix}\begin{bmatrix}3 & 2 & -1\end{bmatrix}$$

$$= I - \frac{1}{7}\begin{bmatrix}9 & 6 & -3\\6 & 4 & -2\\-3 & -2 & 1\end{bmatrix} = \frac{1}{7}\begin{bmatrix}-2 & -6 & 3\\-6 & 3 & 2\\3 & 2 & 6\end{bmatrix}$$

check, $$H\underline{x} = \frac{1}{7}\begin{bmatrix}-2 & -6 & 3\\-6 & 3 & 2\\3 & 2 & 6\end{bmatrix}\begin{bmatrix}2\\6\\-3\end{bmatrix} = \begin{bmatrix}-7\\0\\0\end{bmatrix}$$

The QR-Factorization with Householder Reflection

Introduction:

The idea is similar to the Gauss elimination in the LU-factorization language. Given an $n \times m$ matrix A (could be rectangular as well), we bring it into an upper triangular form (R) by multiplying it from the left appropriately by chosen Householder matrices. We can assume that non of the columns of A is fully zero (such a column just corresponds to fully zero column in R, hence and then can be put back at the end).

In the first step we eliminate all but the top entry in the first column of A. We can do it by one single Householder matrix namely $H(\underline{a}_1)$ is the first column of A. The result:

$$A_1 = H(\underline{a}_1)A = \begin{bmatrix}x & x & x & \ldots. & x\\0 & x & x & & x\\0 & x & x & & x\\\vdots & \vdots & \vdots & & \vdots\\0 & x & x & & x\end{bmatrix}$$

where x denote a generic (usually non-zero). Let $H_1 = H(\underline{a}_1)$ for brievely. Next we look at the second column of the matrix: $A_1 = H(\underline{a}_1)A$

Cut off the first entry of the second column (since we do not want to change the first row any more) i.e., consider the vector $\tilde{a}_2$ of size (n-1) formed from the underlined elements. If this vector is zero, its only nonzero element is on the top, then already the second column is in upper triangular form, so we can proceed to the next column.

If this cut off vector $\tilde{a}_2$ has non zero elements below the top entry, then we use a Householder reflection $H(\tilde{a}_2)$ in the place of "cut off" vectors. In the original space this means a
Multiplication by the matrix:

$$H_2 = \begin{bmatrix}1 & 0 & 0\ldots\ldots\ldots\ldots & 0\\0 & & & \\\vdots & & H(\tilde{a}_2) & \\0 & & & \end{bmatrix}$$

The result is: $A_2 = H_2 A_1 = \begin{bmatrix} x & x & x & .... & x \\ 0 & x & x & .... & x \\ \vdots & 0 & \underline{x} & .... & x \\ \vdots & \vdots & \vdots & .... & x \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \underline{x} & & x \end{bmatrix}$

In the next step we consider the next column of A_2 which has nonzero elements below the third row. Again cutoff the top two entries of this vector and consider the vector $\tilde{a}_3$ of size (n-2) (underlined elements in A_2). We can find a Householder reflection $H(\tilde{a}_3)$ in IR^{n-2}, and if we multiply A_2 from the left with

$$H_3 = \begin{bmatrix} 1 & 0 & 0 & 0......... & 0 \\ 0 & 1 & 0 & 0......... & 0 \\ 0 & 0 & & & \\ \vdots & & & H(\tilde{a}_2) & \\ 0 & 0 & & & \end{bmatrix}$$

then the result is:

$$A_2 = H_2 A_1 = \begin{bmatrix} x & x & x & .... & x \\ 0 & x & x & & x \\ 0 & 0 & x & & x \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & x \end{bmatrix}$$

After at most (m-1) step, we clearly arrive at an upper triangular matrix R, hence, we have:

$$H_{m-1} H_{m-2} ....H_2 H_1 A = R$$

it is clear that all H_i matrices are orthogonal (they are Householder matrices on a subspace and identity on the component of that subspace)

i.e. A= QR, with

$$Q = H_1 H_2 ....H_{m-2} H_{m-1}$$

(recall that $H_i^{-1} = H_i^t = H_i$) Notice that we always multiply by orthogonal matrices, which is a stable operation.

Problem: Find the QR – factorization of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 2 \\ 2 & 0 & 1 \end{bmatrix}$$

With the Householder reflections

Solution: The first column vector is $\underline{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ the corresponding Householder vector is: $\underline{\tilde{u}} = \begin{pmatrix} 1+\sqrt{5} \\ 0 \\ 2 \end{pmatrix}$

with norm square $\|\underline{\tilde{u}}\|_2^2 = 10 + 2\sqrt{5}$, hence :

$$H_1 = H(\underline{a}_1) = I \frac{-2\underline{\tilde{u}}\underline{\tilde{u}}^T}{10+2\sqrt{5}} = \begin{pmatrix} -\frac{1}{\sqrt{5}} & 0 & \frac{-2}{\sqrt{5}} \\ 0 & 1 & 0 \\ \frac{-2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{pmatrix}$$

and

$$A_1 = H_1 A = \begin{pmatrix} -\sqrt{5} & \frac{-2}{\sqrt{5}} & -\sqrt{5} \\ 0 & 3 & 2 \\ 0 & \frac{-4}{\sqrt{5}} & -\sqrt{5} \end{pmatrix} = \begin{bmatrix} -2.236 & -0.894 & -2.236 \\ 0 & 3 & 2 \\ 0 & -1.788 & -2.236 \end{bmatrix}$$

The next cutoff column vector is: $\underline{\tilde{a}}_2 = \begin{pmatrix} 3 \\ \frac{-4}{\sqrt{5}} \end{pmatrix}$,

its norm is $\|\underline{\tilde{a}}_2\|_2 = \sqrt{\frac{61}{5}}$ and the corresponding Householder vector is: $\underline{\tilde{u}} = \begin{bmatrix} 3+\sqrt{\frac{61}{5}} \\ \frac{-4}{\sqrt{5}} \end{bmatrix} = \begin{pmatrix} 6.492 \\ -1.788 \end{pmatrix}$

Hence its norm space is $\|\underline{\tilde{a}}_2\|_2^2 = 45.346$, so

$$H(\underline{\tilde{a}}_2) = I - \frac{2\underline{\tilde{u}}\underline{\tilde{u}}^T}{45.346} = \begin{pmatrix} -0.859 & 0.512 \\ 0.512 & 0.859 \end{pmatrix}$$

From this we form the corresponding Householder matrix:

$$H_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.859 & 0.512 \\ 0 & 0.512 & 0.859 \end{bmatrix}$$

and $A_2 = H_2 A_1 = \begin{bmatrix} -2.236 & -0.894 & -2.236 \\ 0 & -3.492 & -2.862 \\ 0 & 0 & -0.896 \end{bmatrix}$

This is the R matrix in the QR-decomposition To obtain Q we compute

$$Q = H_1 H_2 = \begin{bmatrix} -.447 & -.458 & -0.768 \\ 0 & -.859 & 0.51 \\ -.894 & .229 & .384 \end{bmatrix}$$

Hence

$$A = \begin{bmatrix} -.447 & .458 & -.768 \\ 0 & -.859 & .512 \\ -.894 & .229 & .384 \end{bmatrix} \begin{bmatrix} -2.236 & -0.894 & -2.236 \\ 0 & -3.492 & -2.861 \\ 0 & 0 & -0.896 \end{bmatrix}$$

Is the QR-decomposition.

Householder QR Factorization:-

Apply the Householder reflection to the matrix $A \in IR^{m\times n} (m \geq n)$ to obtain the QR-factorization.

Example: Suppose $A \in IR^{5\times 4}$ and assume that the Householder matrices H_1 and H_2 have been computed so that:

$$H_2 H_1 A = \begin{bmatrix} x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & \otimes & x \\ 0 & 0 & \otimes & x \\ 0 & 0 & \otimes & x \end{bmatrix}$$

Concentrating on the highlighted vector $\begin{bmatrix} \otimes \\ \otimes \\ \otimes \end{bmatrix}$, we determine a Householder matrix $\tilde{H}_3$ such that:

$$\tilde{H}_3 \begin{bmatrix} \otimes \\ \otimes \\ \otimes \end{bmatrix} = \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix}$$

Choosing H_3 = diag $\left(I_2, \tilde{H}_3\right)$, we get

$$H_3 H_2 H_1 A = \begin{bmatrix} x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \\ 0 & 0 & 0 & \otimes \\ 0 & 0 & 0 & \otimes \end{bmatrix}$$

Next consider the highlighted vector $\begin{bmatrix} \otimes \\ \otimes \end{bmatrix}$ and determine

$\tilde{H}_4$ such that: $\tilde{H}_4 \begin{bmatrix} \otimes \\ \otimes \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}$

Choosing H_4 = diag $\left(I_3, \tilde{H}_4\right)$, we get:

$$H_4 H_3 H_2 H_1 A = \begin{bmatrix} x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \\ 0 & 0 & 0 & x \\ 0 & 0 & 0 & 0 \end{bmatrix} = R$$

y setting $Q = H_1 H_2 H_3 H_4$, we obtain

$$QR = \underbrace{H_1 H_2 H_3 H_4}_{Q} \; \underbrace{H_4 H_3 H_2 H_1}_{Q^T} A = A$$

Proposition:- If $A \in IR^{m\times n} (m \geq n)$, then there exist the Householder matrices H_i such that

$$Q = \begin{cases} H_1 \ldots\ldots H_{n-1}, & m > n \\ H_1 \ldots\ldots H_{n-1}, & m = n \end{cases}$$

$$R = \begin{cases} H_1 \ldots\ldots H_{n-1} A, & m > n \\ H_1 \ldots\ldots H_{n-1} A, & m = n \end{cases}$$

and A = QR, where $Q \in IR^{m\times m}$ is orthogonal and $R \in IR^{m\times n}$ is upper triangular

Example: Find the Householder QR-factorization for

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 6 & 2 & 0 \\ -3 & -1 & -1 \end{bmatrix}$$

In example (1) there has been found the Householder matrix for the transformation of the first column vector $[2, 6, -3]^T$ of A:

$$H_1 = \frac{1}{7} \begin{bmatrix} -2 & -6 & 3 \\ -6 & 3 & 2 \\ 3 & 2 & 6 \end{bmatrix}$$

Find that:

$$H_1 A = \frac{1}{7} \begin{bmatrix} -2 & -6 & 3 \\ -6 & 3 & 2 \\ 3 & 2 & 6 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ 6 & 2 & 0 \\ -3 & -1 & -1 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} -49 & -15 & -5 \\ 0 & 4 & -8 \\ 0 & -2 & -3 \end{bmatrix}$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

To find $\tilde{H}_2$, we compute the according Householder vector

$$\tilde{\underline{u}} = \begin{bmatrix}4\\-2\end{bmatrix} - \sqrt{20}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}4-\sqrt{20}\\-2\end{bmatrix}$$

Hence:

$$\tilde{H}_2 = I\frac{-2\tilde{\underline{u}}\tilde{\underline{u}}^T}{\tilde{\underline{u}}^T\tilde{\underline{u}}} = ... = \frac{\sqrt{5}}{5}\begin{bmatrix}2 & -1\\-1 & -2\end{bmatrix}$$

and $H_2 = diag\left(I_1, \tilde{H}_2\right) = \begin{bmatrix}1 & 0 & 0\\0 & \frac{2\sqrt{5}}{5} & \frac{-\sqrt{5}}{5}\\0 & \frac{-\sqrt{5}}{5} & \frac{-2\sqrt{5}}{5}\end{bmatrix}$

and also:

$$R = H_2H_1A = \frac{1}{7}\begin{bmatrix}1 & 0 & 0\\0 & \frac{2\sqrt{5}}{5} & \frac{-\sqrt{5}}{5}\\0 & \frac{-\sqrt{5}}{5} & \frac{-2\sqrt{5}}{5}\end{bmatrix}\begin{bmatrix}-49 & -15 & -5\\0 & 4 & -8\\0 & -2 & -3\end{bmatrix}$$

$$= \begin{bmatrix}-7 & -\frac{15}{7} & \frac{-5}{7}\\0 & \frac{2\sqrt{5}}{7} & \frac{-13\sqrt{5}}{35}\\0 & 0 & \frac{2\sqrt{5}}{5}\end{bmatrix}$$

Find also the orthogonal matrix:

$$Q = H_1H_2 = \frac{1}{7}\begin{bmatrix}1 & 0 & 0\\0 & \frac{2\sqrt{5}}{5} & \frac{-\sqrt{5}}{5}\\0 & -\frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5}\end{bmatrix}\begin{bmatrix}-2 & -6 & 3\\-6 & 3 & 2\\3 & 2 & 6\end{bmatrix}$$

$$Q = \frac{\sqrt{5}}{35}\begin{bmatrix}-2\sqrt{5} & -15 & 0\\-6\sqrt{5} & 4 & -7\\3\sqrt{5} & -2 & -14\end{bmatrix}$$

and check the result:

$$QR = \frac{\sqrt{5}}{35}\begin{bmatrix}-2\sqrt{5} & -15 & 0\\-6\sqrt{5} & 4 & -7\\3\sqrt{5} & -2 & -14\end{bmatrix}\begin{bmatrix}-7 & \frac{-15}{7} & \frac{-5}{7}\\0 & \frac{2\sqrt{5}}{7} & \frac{-13\sqrt{5}}{35}\\0 & 0 & \frac{2\sqrt{5}}{5}\end{bmatrix} = A$$

Hence: $\tilde{H}_2 = I\frac{-2\tilde{u}\tilde{u}^T}{\tilde{u}_T\tilde{u}} = I - 2\frac{\begin{bmatrix}4\\-2\end{bmatrix}\begin{bmatrix}4-2\end{bmatrix}}{\begin{bmatrix}4-2\end{bmatrix}\begin{bmatrix}4\\-2\end{bmatrix}} = I\frac{-2}{10}\begin{bmatrix}16 & -8\\-8 & 4\end{bmatrix}$

$$= I - \frac{1}{5}\begin{bmatrix}16 & -8\\-8 & 4\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} + \begin{bmatrix}\frac{-16}{5} & \frac{8}{5}\\\frac{8}{5} & \frac{-4}{5}\end{bmatrix}$$

$$= \begin{bmatrix}1-\frac{16}{5} & 0-\frac{8}{5}\\0-\frac{8}{5} & 1-\frac{4}{5}\end{bmatrix}$$

Section (2):

III. GENERAL CONSIDERATIONS:

The structure of most constrained optimization problems is essentially contained in the following

minimize $f(\underline{x})$

$$\underline{x} \in IR^n$$

subject to: $C_i(\underline{x}) = 0 \quad , \quad i \in E$

$$C_i(\underline{x}) \geq 0 \quad , \quad i \in I \qquad (2.1)$$

Here $f(\underline{x})$ is called the objective function.

$c_i(\underline{x}) \quad , \quad i = 1,2,...,m$ are the constraint functions. E is the index set of equality constraints, I is the index set of inequality constraints. Any point $\underline{x}$ which satisfies all the constraints in (2.1) is said to be a feasible point and the set of all such points is referred to as the feasible region R. We say that $\underline{x}^*$ is a constrained local minimizer if $\underline{x}^*$ is feasible and satisfies $f(\underline{x}^*) \leq f(\underline{x})$, for all feasible $\underline{x}$ sufficiently close to $\underline{x}^*$. A constraints C_i is said to be active at $\underline{x}^*$ if $C_i = C_i(\underline{x}^*) = 0$. This means that all equality constraints are active.

Active constraints at any point $\underline{x}^*$ are defined by the set

$$\eta = \eta(\underline{x}) = \{i : C_i(\underline{x}) = 0\}$$

clearly $\eta \supset E$ .The set η^* of active constraints at the solution of (2.1) is of some importance. If this set is known, then the remaining constraints can be ignored (locally) and the problem can be treated as equality constraints problem with $E = \eta^*$. This is because constraint with $E = \eta^*$ can be perturbed by small amounts without affecting the local solution where as this is not usually true for an active constraint.

An example is given by

Minimize $f(\underline{x}) = -x_1 - x_2$

Subject to

$$c_1(\underline{x}) = x_2 - x_1^2 \geq 0, \quad c_2(\underline{x}) = 1 - x_1^2 - x_2^2 \geq 0.$$

Clearly the solution is achieved at $\underline{x}^* = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ when the contour of $f(\underline{x})$ is tangent to the unit circle. Thus $\eta^* = \{2\}$ and the unit circle constraint $C_2(\underline{x})$ is active[18]. The parabola constraint $c_1(\underline{x})$ is inactive and can be perturbed or removed from the problem without changing $\underline{x}^*$.It is assumed that the functions $C_i(\underline{x})$ are continuous which implies that R is closed. It is also assumed that $f(\underline{x})$ is continuous for all $\underline{x} \in R$ and preferably for all $\underline{x} \in IR^n$. If in addition the feasible region is nonempty and bounded then it follows that a solution $\underline{x}^*$ exists. In fact most practical methods require the stronger assumption that the objective and the constraint functions are also smooth is that their first and often second continuous derivatives exist. Methods of solution of (2.1) are usually iterative [27],[28] ,so that a sequence $\underline{x}^{(1)}, \underline{x}^{(2)}, \underline{x}^{(3)}, \ldots;$ say is generated from a given point $\underline{x}^{(1)}$, hopefully converging to or terminating at $\underline{x}^*$.

Linearly Constrained Optimization:-

Here we examine ways of representing linear constraints. The goal is to write the constraints in a form that makes it easy to move from one feasible point to another. The constraints specify interrelationships among the variables so that, for example, if we increase the first variable retaining feasibility might require making a complicated sequence of changes to all the other variables. In the general case these constraints may either be equalities or in-equalities. Since any in equality of the "less than or equal type may be transformed to an equivalent constraint of the "greater than or equal" type, any problem with linear constraints may be written as follows:

minimize $f(\underline{x})$

$$\text{subject to} \quad \begin{cases} \underline{a}_i^T \underline{x} = \underline{b}_i, & i \in E \\ \underline{a}_i^T \underline{x} \geq \underline{b}_i, & i \in I \end{cases}$$

Each a_i here is a vector of length n and each b_i is a scalar. E is an index set for the equality constraints and I is an index set of the in-equality constraints. We denote by A the matrix whose rows are the vectors a_i^T , and denote by $\underline{b}$ the vector of right hand side coefficients b_i. An example is given by a problem with linear constraints.

Consider the problem:-

minimize $f(\underline{x}) = x_1^2 + x_2^3 x_3^4$

subject $x_1 + x_2 + x_3 = 6$

for this example $E = \{1\}, I = \{2,3,4\}$. The vectors $\{a_i\}$ that determine the constraints are

$$\underline{a}_1 = (1\ \ 2\ \ 3)^T, \quad \underline{a}_2 = (1\ \ 0\ \ 0)^T$$

$$\underline{a}_3 = (0\ \ 1\ \ 0)^T, \quad \underline{a}_4 = (0\ \ 0\ \ 1)^T$$

and the right hand sides are

$$b_1 = 6, \quad b_2 = 0, \quad b_3 = 0, \quad b_4 = 0,$$

In the following sections we will be discussing the optimality conditions. In section (3) we derive optimality conditions when only equality constraints are present, i.e $I = \phi$. In section (4) the conditions are derived for inequality constraints only i.e. $E = \phi$. The case when both equality and inequality constraints are present will be assumed up in section 1.6. There we conclude by introducing KKT conditions. In section 1.7 we introduce the first order necessary conditions.

Section (3):

IV. LINEAR EQUALITY AND INEQUALITY CONSTRAINTS:-

In this section introduce the optimality conditions for linear equality and inequality constraints.

1- Linear Equality constraints:

Here we discuss the optimality conditions for nonlinear problems, where all constraints are linear equalities

minimize $f(\underline{x})$, $\underline{x} \in IR^n$

subject to $A^T \underline{x} = \underline{b}$ (3.1) and A is an $m \times n$ matrix. We assume that f is twice continuously differentiable over the feasible region. We also assume that the rows of A are linearly independent, that is, A has full row rank. The main idea is to transform this constrained problem into an equivalent unconstrained problem. The theory and methods for unconstrained optimization can then be applied to the new problem to demonstrate the approach. Any problem with linear equality constraints $A\underline{x} = \underline{b}$ can be recast as an equivalent unconstrained problem. Suppose we have a feasible direction. Any feasible direction must lie in the null space of A, the set of vectors p satisfying A$\underline{p}$ = $\underline{0}$. Denoting the null space of A^T and denoting *N(A)*, the feasible region can be described by $\{\underline{x} = \underline{x} + p\,,\ \ p \in N(A)\}$. Let Z be an $n \times r$ null space matrix for A (with $r \geq n - m$). Then the feasible region is given by:

$$\{\underline{x} : \underline{x} + Zv\} \text{ where } v \in IR^r \qquad (3.2)$$

the null space of A^T must have a basis. Let the columns of the matrix Z form such a basis, then $A^T Z = 0$, and every feasible direction can be written as a linear combination of the columns of Z. Therefore if f satisfies (3.2), $\underline{P}$ can be written as $Z\underline{P}_Z$ for some vector $\underline{P}_Z$ .In order to determine the optimality condition of a given feasible point x^*, along a feasible direction $\underline{P}(\underline{P} = Z\underline{P}_Z)$

$$f(\underline{x}^* + \varepsilon Z\underline{P}_Z) = f(\underline{x}^*) + \varepsilon \underline{P}_Z^T Z^T G\left(\underline{x}^* + \varepsilon\theta\underline{P}\right) Z\underline{P}_Z \qquad (3.3)$$

where θ satisfies $0 \leq \theta \leq 1$ and ε is taken, without loss of generality as a positive scalar. Now (3.3) shows that if $\underline{P}_z^T Zg(\underline{x}^*)$ is negative then every neighborhood of $\underline{x}^*$ will contain feasible points with a strictly lower function value. Thus a necessary condition for $\underline{x}^*$ to be a local minimum of (3.1) is that $\underline{P}_z^T Zg(\underline{x}^*)$ must vanish for every $\underline{P}z$, which implies that:

$$Z^T g(\underline{x}^*) = \underline{0} \qquad (3.4)$$

The vector $Z^T g(\underline{x}^*)$ is termed the projected gradient of *f* at $\underline{x}^*$. Any point at which the projected gradient vanishes is termed a constrained stationary point.

The result (3.4) implies that $g(\underline{x}^*)$ must be a linear combination of the columns of A; i.e

$$g(\underline{x}^*) \;=\; \sum_{i=1}^{m} a_i \lambda_i \;=\; A\underline{\lambda} \qquad (3.5)$$

for some vector $\underline{\lambda}$, which is termed the vector of Lagrange multipliers. The Lagrange multipliers are unique only if the columns of A are linearly independent.

Condition (3.4) is equivalent to (3.5) because every n-vector can be expressed as a linear combination of the columns of A and the columns of Z and hence $g(\underline{x}^*)$ can be written as $g(\underline{x}^*) = A\lambda + Zgz$ for some vector gz. Pre-multiplying $g(\underline{x}^*)$ by Z^T and using (3.4), it follows that $Z^T Zg_Z = \underline{0}$, since $Z^T Z$ is non singular by definition of a basis, this will be true only if $\underline{g}_z = \underline{0}$

Since $Z^T g(\underline{x}^*)$, the Taylor series expansion (3.3) becomes:

$$f(\underline{x}^* + \varepsilon Z\underline{P}_Z) = f(\underline{x}^*) + 0.5\varepsilon^2 \underline{P}_Z^T Z^T G\left(\underline{x}^* + \varepsilon\theta\underline{P}\right) Z\underline{P}z \qquad (3.6)$$

(3.6) indicates that if the matrix $Z^T G(\underline{x})Z$ is indefinite, every neighbourhood of $\underline{x}^*$ contains feasible points with strictly lower value of *f*. Therefore a second order necessary condition for $\underline{x}^*$ to be optimal for (3.1) is that the matrix $Z^T G(\underline{x})Z$ which is termed the projected Hessian matrix, must be positive semi-definite. We summarize in the following theorem.

Theorem (1): The necessary conditions for $\underline{x}^*$ to be a local minimum of (3.1) are the following:

1) $A^T \underline{x}^* = \underline{b}$
2) $Z^T g(\underline{x}^*) = 0$ or equivalently $g(\underline{x}^*) = A\underline{\lambda}$, and
3) $Z^T G(\underline{x}^*)Z$ is positive semi-definite

Now, if $Z^T G(\underline{x}^*)Z$ is positive definite, by continuity $Z^T G(\underline{x})Z$ is positive definite for all points in some neighbourhood of $\underline{x}^*$. If ε is small enough then $\underline{x}^T + \varepsilon Z\underline{P}_Z$ will be made inside that neighbourhood. Hence, for all such ε it holds that $P_Z^T Z^T G(\underline{x}^* + \varepsilon\theta\underline{P})Z\underline{P}_z > 0$.

From (3.6) this implies that $f\left(\underline{x}^*\right)$ is strictly less than the value of *f* for all points in some neighbourhood of $\underline{x}^*$. Thus we summarize in the following theorem.

Theorem (2): Sufficient conditions for $\underline{x}^*$ to be a local minimum for (3.1) are:

1) $A^T \underline{x}^* = \underline{b}$
2) $Z^T g(\underline{x}^*) = 0$ or equivalently: $g(\underline{x}^*) = A\lambda$, and
3) $Z^T G(\underline{x}^*)Z$ is positive definite

For example: minimize

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$f(\underline{x}) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3$ subject to $x_1 - x_2 + 2x_3 = 2$

since $\nabla f(\underline{x}) = \left(2x_1 - 2, 2x_2, -2x_3 + 4\right)^T$, then at the feasible point $\underline{x}^* = (2.5, -1.5, -1)$ the gradient of f is $(3, -3, 6)^T$. The matrix $A = (1, -1, 2)$, is easily verified that $Z^T g(\underline{x}^*) = (0,0)^T$. Thus, the reduced gradient vanishes at $\underline{x}^*$ and the first – order necessary condition for a local minimum is satisfied at this point. Checking the reduced Hessian matrix, we find that

$$Z^T G(\underline{x}^*) Z = \begin{bmatrix} 1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ -4 & 6 \end{bmatrix}$$

is positive definite at $\underline{x}^*$. Hence the second-order sufficiency conditions are satisfied, and $\underline{x}^*$ is a strict local minimizer of *f* Note that G(x*) itself is not positive definite. At any feasible point, the variable x_1 can be expressed in terms of x_2 and x_3 using $x_1 = 2 + x_2 - 2x_3$. Substituting this into the formula for $f(\underline{x})$, we obtain the equivalent unconstrained[16].

minimize $2x_2^2 + 3x_3^2 - 4x_2 x_3 + 2x_2$

The number of variables has been reduced from three to two. It is easy to verify that a strict local minimizer to the unconstrained problem is $x_2 = -1.5, x_3 = -1$. The solution to the original problem is $\underline{x}^* = (2.5, -1.5, -1)^T$ with an optimal objective value of $f(\underline{x}^*) = -1.5$.

Consider the problem: minimize

$$f(\underline{x}) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3$$

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

select $Z = \begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$, as a null space matrix for the constraint matrix A = (1, -1, 2). Using the (arbitrary) feasible point $\bar{x} = (2,0,0)^T$, any feasible point can be written as :

$$\underline{x} = \underline{\bar{x}} + Z\underline{v} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \underline{v},$$

for some $\underline{v} = (v_1, v_2)^T$. Substituting into f we obtain the reduced function $\phi(\underline{v}) = 2v_1^2 + 3v_2^2 - 4v_1 v_2 + 2v_1$. This is the same reduced function as before, except that now the variables are called v_1 and v_2 rather than x_2 and x_3.

2- Linear Inequality Constraints:

Consider the following problem: minimize

$$f(\underline{x})\,; \quad x \in IR^n \qquad (3.7)$$

subject to $A^T \underline{x}^* \geq \underline{b}$

We first try to derive a characterization of the feasible point in the neighbourhood of a possible solution. If the j^{th} constraint is active at the feasible point $\underline{x}^*$, it is possible to move a non-zero distance from $\underline{x}^\backslash$ in any direction without violating that constraint; i.e. for any vector $\underline{P}$, $\underline{x}^\backslash + \varepsilon \underline{P}$ will be feasible with respect to an inactive constraint if ε is small enough. On the other hand, an active constraint restricts feasible perturbations in every neighbourhood of a feasible point. Suppose that the i^{th} constraint is active at $\underline{x}^\backslash$, so that $a_i^T \underline{x}^\backslash = b_i$. There are two categories of feasible directions with respect to an active inequality constraint. Firstly, if $\underline{P}$ satisfies : $a_i^T \underline{P} = 0$.The direction $\underline{P}$ is termed a binding perturbation with respect to the i^{th} constraint, since the i^{th} constraint remains active at all points $\underline{x}^\backslash + \alpha \underline{P}$ for any α. Secondly, if $\underline{P}$ satisfies $\underline{a}_i^T \underline{P} > 0$. $\underline{P}$ is thus termed a non-binding perturbation with respect to the i^{th} constraint. Since $\underline{a}_i^T (\underline{x} + \alpha \underline{P}) = b_i + \alpha a_i^T \underline{P} > b_i$, if $\alpha > 0$ the i^{th} constraint becomes inactive at the perturbed point $\underline{x}^\backslash + \alpha \underline{P}$.In order to determine whether the feasible point $\underline{x}^*$ is optimal for (3.7), it is necessary to identify the active constraints. Let the t columns of the matrix $\hat{A}$ contain the coefficients of constraints active at $\underline{x}^*$, with a similar convention for the vector $\hat{\underline{b}}$, so that

$\hat{A}^T \underline{x}^* = \hat{\underline{b}}$.The renumbering of columns of $\hat{A}$ corresponds to the order of the active constraints; so that $\hat{a}_1$ contains the coefficients of "first" active constraint. For simplicity in the proof we assume that the columns of $\hat{A}$ are linearly independent; however, the derived conditions hold even when $\hat{A}$ does not have full column rank. Let Z be a matrix whose columns form a basis for the set of vectors orthogonal to the columns of $\hat{A}$. Every $\underline{P}$ satisfying $\hat{A}\underline{P}^T = \underline{0}$ can therefore be written as a linear combination of the columns of Z. Consider the Taylor series expansion[24] of *f* about $\underline{x}^*$ along a binding perturbation

$$\underline{P}(\underline{P} = Z\underline{P}_Z)$$

$$, f\left(\underline{x}^* + \varepsilon Z \underline{P}_Z\right) = f\left(\underline{x}^*\right) + \varepsilon P_Z^T Z^T g\left(\underline{x}^*\right)$$

$$+0.5\varepsilon^2 \underline{P}_Z^T Z^T G\left(\underline{x}^* + \varepsilon\theta \underline{P}\right) Z \underline{P}_Z \qquad (3.8)$$

where θ satisfies $0 \le \theta \le 1$, and ε is taken without loss of generality to be positive. As in the equality constraint case, (3.8) shows that if $\underline{P}_Z^T Z^T g(\underline{x}^*)$ is non zero for any $\underline{P}_Z$; then $\underline{x}^*$ cannot be a local minimum. Thus a necessary condition for optimality of $\underline{x}^*$ is that $Z^T g\left(\underline{x}^*\right) = \underline{0}$; or, equivalently, that

$$g\left(\underline{x}^*\right) = \hat{A}\underline{\lambda} \qquad (3.9)$$

The above condition ensures that *f* is stationary along all binding perturbations from $\underline{x}^*$, are also feasible directions with respect to the active inequality constraints, the point $\underline{x}^*$ will not be optimal if there exists any non-binding perturbation $\underline{P}$ that is a descent direction for *f* ($\underline{p}$) is a descent direction for f at $\underline{x}$ if $f\left(\underline{x} + \varepsilon\underline{P}\right) < f\left(\underline{x}\right)$, $0 < \varepsilon < 1$ for some $\varepsilon^\backslash > 0$). If such a direction exist, a sufficiently small positive step a long it will remain feasible and produce a strict decrease in *f*. To avoid this possibility, we seek a condition to ensure that for all $\underline{P}$ satisfying $\hat{A}^T \underline{P} \ge \underline{0}$, it holds that $g\left(\underline{x}^*\right)^T \underline{P} \ge 0$. Since we know already from (3.8) that $g\left(\underline{x}^*\right)$ is a linear combination of the columns of $\hat{A}$, the desired condition is that:

$$g\left(\underline{x}^*\right)^T \underline{P} = \lambda_1 \underline{\hat{a}}_1 \underline{P} + \ldots + \lambda_i \underline{\hat{a}}_i \underline{P} \ge 0 \qquad (3.10)$$

where $\underline{\hat{a}}_i^T \ge 0$, $i = 1, \ldots, t$

The condition (3.10) will hold only if $\lambda_i \ge 0$, $i = 1, \ldots, t$ i.e $\underline{x}^*$ will not be optimal if there are any negative multipliers. To see why, assume that $\underline{x}^*$ a local minimum (so that (3.9) must hold), but that $\lambda_j < 0$ for some j. Because the columns of $\hat{A}$ are linearly independent, corresponding to such a value of j there must exist a non-binding perturbation $\underline{P}$ such that

$$\underline{\hat{a}}_j^T \underline{P} = 1 \quad ; \quad \underline{\hat{a}}_j^T \underline{P} = 0, \quad i \ne j$$

for such a $\underline{P}$,

$$g(\underline{x}^*)^T \underline{P} = \lambda_j \underline{\hat{a}}_j^T \underline{P} = \lambda_j \underline{\hat{a}}_j^T \underline{P} = \lambda_j < 0$$

and hence $\underline{P}$ is a feasible descent direction,[31],[12] which contradicts the optimality of $\underline{x}^*$. Thus a necessary condition for a solution of (3.7) is that all the Lagrange multipliers must be non negative. By considering the Taylor-series expansion of f about $\underline{x}^*$ along binding perturbation, we can derive the second order necessary condition that projected Hessian matrix $Z^T G\left(\underline{x}^*\right) Z$ must be positive semi-definite. This condition is precisely analogous to the second order necessary condition for the equality constrained problem.

Theorem (3): If $\underline{x}^*$ is a local minimizer of f over the set $\{\underline{x} : A\underline{x} \ge \underline{b}\}$, then for some vector $\underline{\lambda}^*$ of Lagrange multipliers

1) $g(\underline{x}^*) = A^T \underline{\lambda}^*$ or equivalently $Z^T g(\underline{x}^*) = \underline{0}$

2) $\underline{\lambda}^* \ge \underline{0}$

3) $\underline{\lambda}^{*T}(A\underline{x}^* - \underline{b}) = 0$ and

4) $Z^T G(\underline{x}^*) Z$ is positive semi-definite

where Z is a null space matrix for the matrix of active constraints at $\underline{x}^*$. We also develop sufficiency conditions that guarantee that a stationary point of *f* is indeed a local minimizer. For the sake of clarity, we only consider the case when Z is a basis matrix for the null space of $\hat{A}$. To illustrate that the positive-definiteness of the projected Hessian does not suffice to ensure optimality when there are zero Lagrange multipliers, consider the two dimensional example of minimizing $x_1^2 - x_2^2$ subject to $x_2 \ge 0$. Here we have:

$$g\left(\underline{x}\right) = \begin{pmatrix} 2x_1 \\ -2x_2 \end{pmatrix}, \qquad G\left(\underline{x}\right) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

let $\underline{x}^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, so $\hat{A} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$; and $Z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$,

therefore $g\left(\underline{x}^*\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \hat{A}\lambda$ with $\lambda = 0$.

Although the projected Hessian is positive definite, since $Z^T G\left(\underline{x}^*\right) Z = 2$, $\underline{x}^*$ is not a local minimum, since any positive step along the vector $(0,1)^T$ is a feasible perturbation that reduces *f*.

The origin is not optimal because, although f is stationary along binding perturbation. One means of avoiding such complications is to include the condition that all the Lagrange multipliers be strictly positive, which guarantees that *f* will have a strict increase for any non-binding perturbation with this approach we get:

Theorem (4): The following conditions are sufficient for $\underline{x}^*$ to be a strong local minimum for (3.7)

1) $A^T \underline{x}^* \geq \underline{b}$, with $A^T \underline{x}^* = \hat{\underline{b}}$

2) $Z^T g(\underline{x}^*) = 0$; or equivalently $g(\underline{x}^*) = \hat{A}\underline{\lambda}$;

3) $\lambda_i \geq 0$ $i = 1, \ldots, t$; and

4) $Z^T G(\underline{x}^*) Z$ is positive semi-definite

It is possible to state sufficient conditions even when zero multipliers are present, by including extra restrictions on the projected Hessian matrix to ensure that *f* has positive curvature for all constraints with positive Lagrange multipliers, but may be binding or non-binding for constraints with zero Lagrange multipliers. Let $\hat{A}_+$ contain the coefficients of the active constraints with positive Lagrange multipliers, and let *Z* be a matrix whose columns span the null space of $\hat{A}_+^T$. In this case we can state the following theorem.

Theorem (5): Sufficient Conditions for $\underline{x}^*$ to be a strong local minimum of (3.1) are:

(1) $A^T \underline{x}^* \geq \underline{b}$ with $\hat{A}\underline{x}^* = \hat{\underline{b}}$

2) $Z^T g(\underline{x}^*) = \underline{0}$; or equivalently $g(\underline{x}^*) = \hat{A}\underline{\lambda}$;

3) $\lambda_i \geq 0$ $i = 1, \ldots, t$; and

4) $Z^T G(\underline{x}^*) Z$ is positive definite.

Consider the problem:

$$\text{minimize} \quad f(\underline{x}) = x_1^3 + x_2^2$$

$$\text{subject to} \quad -1 \leq x_1 \leq 0$$

At the point $x^\backslash = (0,0)^T$ the active set consists of the upper bound constraint on *x*. Writing this constraint as $-x_1 \geq 0$ we find that $\hat{A} = (-1, 0)$ is the matrix of the active constraints. Since $g(\underline{x}) = (0,0)$, then for $g(\underline{x}) = \hat{A}^T \hat{\underline{\lambda}}$ and the first order necessary condition is satisfied at this point. We now examine the second-order necessary conditions, using $Z = (0,1)^T$ as a basis matrix for the null space of $\hat{A}$. Then

$$Z^T G(\underline{x}) Z = (0,1)\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 2 > 0$$

So the reduced Hessian matrix is positive definite at $\underline{x}^\backslash$. The point $\underline{x}^\backslash$ is not optimal, however, since any nearby point of the from $(-\varepsilon,0)^T$ (with ε small and positive) has a lower objective value. In this problem $\hat{A}_+$ is an empty matrix. Hence $Z_+ = I$ and $Z_+^T G(\underline{x}) Z_{+\_} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$ is not positive definite, the second-order sufficiency conditions are not satisfied at $\underline{x}^\backslash$.

Section(4):

V. INTRODUCTION

In this section we introduce the stability and efficiency of general quadratic programming algorithms ,we introduce the active set methods, penalty and barrier functions, complementary pivoting methods, the interior point methods ,the gradient projection methods ,Givens transformations using Householder matrices .

Kuhn-Tucker Conditions:

When a linearly constrained problem includes both equalities and inequalities, (4.1), the optimality conditions are combination of those given for the separate cases. In a mixed problem, the active constraints include all the equalities as well as binding equalities, and there is no sign restriction on the Lagrange multiplier corresponding to equality constraints. In this chapter we introduce what is called Kuhn-Tucker (KKT) -conditions for linearly constrained problems (Kuhn and Tucker 1951). They combine the first order conditions for both equality and inequality problems. Let us now define what is known as a Lagrange function. The Lagrange function $l(\underline{x},\underline{\lambda})$ of problem (3.1) is defined by

$$L(\underline{x},\underline{\lambda}) = f(\underline{x}) - \sum \lambda_i C_i(\underline{x}) \qquad (4.2)$$

The KKT - conditions for linearly constrained problems are described as follows. If $\underline{x}^*$ is a local minimizer of problem (3.1), then there exist Lagrange multipliers $\underline{\lambda}$ such that $\underline{x}^*, \underline{\lambda}$,

Satisfy the following system

* $\lambda_i C_i(\underline{x}^*) = 0$, $i \in E$

* $C_i(\underline{x}^* \geq 0$, $i \in I$

$* \quad \nabla x\, l\left(\underline{x}^*, \underline{\lambda}\right) = 0 \qquad (4.3)$

$* \lambda_i \geq 0, \quad i \in I$

$* \quad \lambda_i C_i(\underline{x}^*) = 0, \quad i \in EUI$

Where ∇x is the gradient vector taken with respect to $\underline{x}$ only. The point $\underline{x}^*$ which satisfies the condition is sometimes referred to as a KKT- point. The final condition $\lambda_i C_i(\underline{x}^*) = 0$ is referred to as the complementarily

condition and states that both λ_i and $C_i(\underline{x}^*)$ cannot be non-zero or equivalently, that in active constraints have zero multiplier. In the next chapter we introduce the class of complementary pivoting methods[31], [30] ,[29]. They basically solve KKT- conditions (4.3).

Theorem(6) : (First Order Necessary Conditions) Suppose that x^* is a local solution to

$$l(\underline{x}, \underline{\lambda}) = f(\underline{x}) - \sum_{i \in EVI} \lambda_i C_i(\underline{x}) \qquad (4.4)$$

the function f and C_i in (1.7.1) are continuously differentiable at x^*. Then there is a Lagrange multiplier vector λ^* with components such that the following conditions are satisfied at $\left(x^*, \lambda^*\right)$

1) $\nabla_x l\ (x^*, \lambda^*)$ =0 (4.4. a)

2) $C_i\left(\underline{x}^*\right) = 0, \quad \forall i \in E$ (4.4. b)

3) $C_i\left(\underline{x}^*\right) = 0, \quad \forall i \in I$ (4.4. c)

4) $\lambda_i^* \geq 0, \quad \forall i \in I$ (4.4. d)

5) $\lambda_i^* C_i\left(\underline{x}^*\right) = 0, \quad \forall i \in EUI$ (4.4. e)

The conditions (4.4) are often known as the karush-Kuhn-Tucker conditions, or KKT- conditions. The conditions (4.4e) are complementarity conditions, they imply that either constraint i is active or $\lambda_i^* = 0$, or possibly both. In particular, the Lagrange multipliers corresponding to in active inequality constraints are zero, we can omit the terms for indice from and rewrite $i \in A(x^*)$. From (4.4) this condition[23] ,[7] as:

$$0 = \nabla x l\left(\underline{x}^*, \underline{\lambda}^*\right) = \nabla f\left(\underline{x}^*\right) - \sum_{i \in A(x^*)} \lambda_i^* \nabla C_i(\underline{x}^*) \qquad (4.5)$$

A special case of complementarity is important and deserves its own definition.

Definition: (Strict complementarity):

Given a local solution x^* of

$$\min_{x \in \Re^n} f(\underline{x})$$

Subject to $c_i(\underline{x}) = 0, \quad i \in E$

$c_i(\underline{x}) \geq 0, \quad i \in I$

and a vector $\underline{\lambda}^*$ satisfying (4.4), we say that the strict complementarity condition hold if exactly one of $\underline{\lambda}^*$ and $c_i(x^*)$ is zero for each index $i \in I$. in other words, we have $\lambda_i^* > 0$ for each $i \in I \cap A(\underline{x}^*)$.Quadratic programming[24] represents a special class of non linear programming in which the objective function is quadratic and the constraints are linear. That name was restricted to the specific problem of minimizing a convex quadratic function of several variables subject to linear constraints. When the function to be minimized is convex, the problem is well understood both theoretically and computationally [10]. In the case of convex problems -under certain assumptions- various methods are equivalent([13],[23] are realization of a general one and hence are equivalent.) In solving the general case when the function is non convex some of the above mentioned methods can still locate a local minimum , others can be modified to terminate successfully [13] , did modify, in a stable way the active set method [10], for some discussion. There are also other methods designed for the general case [21].Other work on the problem exists, some of which is included in the references.

We state the QP problem by :

$$\text{minimize} \quad f(\underline{x}) = 0.5\underline{x}^T G \underline{x} + g^T \underline{x}$$

$$\text{subject to} \quad \underline{a}_i^T \underline{x} = b_i \quad i \in E \qquad (4.6)$$

$$\underline{a}_i^T \underline{x} \geq b_i \quad , \ i \in I$$

Where G is a symmetric $n \times n$ matrix E,I are finite sets of indices g, x and $\{a_i\}$, $i \in E \cup I$ are vectors in $\Re^n$. The problem may be infeasible or the solution may be unbounded, however these possibilities are readily detected in the algorithm, so we assume that a solution $\underline{x}^*$ exists. If the Hessian $G\left(\cong \nabla^2 f\right)$ is positive semi-definite, $\underline{x}^*$ is unique Fletcher [10]. When the Hessian G is indefinite local solutions which are not global can occur.

Equality Constraints: The vast majority of methods for solving problems with linear equality constraints are feasible point methods: they start from a feasible point and move along feasible descent directions to consecutively better feasible points.

In this section we study how to find a solution to equality constrains problem in the form

$$\text{Minimize} \ f(\underline{x}) = 0.5\underline{x}^T G \underline{x} + g^T \underline{x} \qquad (4.7)$$

$$\text{subject to} \quad A^T \underline{x} = \underline{b}$$

It is assumed that A is $m \times n$ $(m \le n)$ matrix of full rank, $b \in IR^m$. If the objective function is convex, this point will be a global minimizer of *f*. In the more general case, there is no guarantee that the point will be a global minimizer. In some cases it may not even be a local minimizer. The assumption that A has full rank ensures that unique Lagrange multipliers λ exist. Here we describe a generalized elimination to solve (4.7) [10].

Let S and Z be $n \times m$ and $n \times (n-m)$ matrices respectively such that [S: Z] is non-singular, and in addition let $A^T S = I$ and $A^T Z = 0$. S^T can be regarded as a

left generalized inverse for A so that a solution of $A^T \underline{x} = \underline{b}$ is given by $\underline{x} = S\underline{b}$, this solution is not unique in general and other feasible points are given by $\underline{x} = S\underline{b} + \underline{\delta}$ where $\underline{\delta}$ belongs to the linear space,

$$\Delta = \left\{ \underline{\delta} : A^T \underline{\delta} = \underline{0} \right\} \qquad (4.8)$$

which has dimension n – m. the columns, $\underline{Z}_1, \ldots, \underline{Z}_n$ of $\underline{Z}$ act as basis vectors for Δ so at any feasible point $\underline{x}$ any feasible direction $\underline{\delta}$ can be written as

$$\underline{\delta} = Z\underline{y} \quad , \quad \underline{y} \in IR^{n-m} \qquad (4.9)$$

(Note: Z here is the same as described in section (3), so any feasible point $\underline{x}$ can be written as:

$$\underline{x} = S\underline{b} + Z\underline{y} \qquad (4.10)$$

substituting into $f(\underline{x})$ gives the reduced quadratic function

$$q(\underline{y}) = \frac{1}{2}\underline{y}^T \left(Z^T G Z\right)\underline{y} + \left(g + GS\underline{b}\right)^T Z\underline{y} + \frac{1}{2}\left(g + GS\underline{b}\right)^T S\underline{b} \qquad (4.11)$$

If $Z^T G Z$ is positive definite then a unique minimizer $\underline{y}^*$ exists which (from $\nabla q(\underline{y}) = 0$) solves the linear system

$$\left(Z^T G Z\right)\underline{y} = -Z^T\left(g + GS\underline{b}\right) \qquad (4.12)$$

$\underline{x}^*$ is then obtained by substituting into (4.10) the matrix $Z^T G Z$ in (4.12), is often referred to as the projected Hessian matrix and denoted by G_A or sometimes G_Z.

To obtain $\underline{\lambda}$ we pre-multiply $G\underline{x}^* + g = A\underline{\lambda}$, (see theorem (1)) by S^T to get: $\underline{\lambda} = S^T\left(G\underline{x}^* + \underline{g}\right)$ (4.13)

Depending on the choice of *S* and *Z* a number of methods exists. One choice of particular importance is obtained by using the QR- factorization of the matrix A. This can be written

$$A = Q\begin{bmatrix} R \\ 0 \end{bmatrix} = \left[Q_1 : Q_2\right]\begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1 R \qquad (4.14)$$

where *Q* is $n \times n$ and orthogonal, *R* is $m \times m$ upper triangular, and Q_1, Q_2 are $n \times m$ and $n \times (n-m)$ respectively. The choices:

$$S = Q_1 R^{-T}, \qquad Z = Q_2 \qquad (4.15)$$

are readily observed to have the correct properties this scheme is due to [12], [10], refers to it as the orthogonal factorization method. The orthogonal factorization method is advantageous in that calculating *Z* and *S* involves operations with elementary orthogonal matrices which are very stable numerically. Also this choice ($Z = Q_2$) gives the best possible bound:

$$K\left(Z^T G Z\right) \le K(G) \qquad (4.16)$$

on the condition number $K(Z^T G Z)$.

We conclude this by introducing the method of Lagrange multipliers. The *KKT-* conditions (4.3) applied to (4.7) (Note : $I = \phi$) yields the equations:

$$\begin{aligned} G\underline{x} + \underline{g} - A\underline{\lambda} &= \underline{0} \\ A\underline{x} - \underline{b} &= \underline{0} \end{aligned} \qquad (4.17)$$

which can be rearranged to give the linear system

$$\begin{bmatrix} G & -A \\ -A^T & 0 \end{bmatrix}\begin{bmatrix} \underline{x} \\ \underline{\lambda} \end{bmatrix} = -\begin{bmatrix} \underline{g} \\ \underline{b} \end{bmatrix} \qquad (4.18)$$

The coefficient matrix is referred to as the Lagrangian matrix and is symmetric but not positive definite. If the inverse exists and is expressed as

$$\begin{bmatrix} G & -A \\ -A^T & 0 \end{bmatrix}^{-1} = \begin{bmatrix} H & -T \\ -T^T & U \end{bmatrix} \qquad (4.19)$$

then the solution to (4.18) can be written

$$\underline{x}^* = -H\underline{g} + T\underline{b} \qquad (4.20)$$

$$\underline{\lambda}^* = T^T\underline{g} - U\underline{b} \qquad (4.21)$$

Using *S* and *Z* defined at the beginning of this section we get the following representation of the inverse Lagrangian matrix [10],

$$H = Z\left(Z^T GZ\right)^{-1} Z^T$$

$$T = S - Z\left(Z^T GZ\right)^{-1} Z^T GS \qquad (4.22)$$

$$U = S^T GZ\left(Z^T GZ\right)^{-1} Z^T GS - S^T GS$$

Equations (4.22) prove that the Lagrangian matrix is non singular if and only if there exists Z such that the matrix $Z^T GZ$ is non singular. Furthermore $\underline{x}^*$ is a local minimizer if and only if $Z^T GZ$ is positive definite.

4 - The Active Set Methods:

Most QP problems involve inequality constraints[11]. Here we consider only inequality constraints. Modification to include equality constraints is straight forward. The general form of the problem is

Minimize $f(\underline{x}) = 0.5\underline{x}^T G\underline{x} + g^T \underline{x}$ (4.23)

subject to $\underline{a}_i^T \underline{x} \geq b_i \quad , \quad i \in I$

where G is $n \times n$ and symmetric positive definite.

The active set method is an iterative method. It generates a sequence of feasible points $\underline{x}^{(1)}, \underline{x}^{(2)}, \ldots$ which terminates in a finite number of steps at the solution $\underline{x}^*$. On iteration k a feasible point $\underline{x}^{(k)}$, is known which satisfies the active constraints as equalities. that is : $\underline{a}_i^T \underline{x}^{(K)} = b_i \quad , \quad i \in \eta$. Also except in degenerate cases $\underline{a}_i^T \underline{x}^{(K)} = b_i \quad , \quad i \notin \eta$. Each iteration attempts to locate the solution to an equality problem in which only the active constraints occur. $\underline{x}^{(K+1)}$ is obtained as follows:-

Let $\underline{\delta}^{(K)}$ solve:

$$\text{minimize} \quad 0.5\left(\underline{x}^{(K)} + \underline{\delta}\right)^T G\left(\underline{x}^{(K)} + \underline{\delta}\right) + \underline{g}^T\left(\underline{x}^{(K)} + \underline{\delta}\right)$$

$$\text{subject to} \quad \underline{a}_i^T\left(\underline{x}^{(K)} + \underline{\delta}\right) = b_i \quad , \quad i \in \eta$$

which is equivalent to

$$\text{minimize} \quad 0.5\underline{\delta}^T G\underline{\delta} + \underline{g}^{(x)^T} \underline{\delta} \qquad (4.24)$$

$$\text{subject to} \quad \underline{a}_i^T \underline{\delta} = 0, \quad i \in \eta$$

where $\underline{g}^{(K)}$ is defined by $\underline{g}^{(K)} = \underline{g} + G\underline{x}^{(K)}$. If $\underline{\delta}^{(K)}$ is feasible with respect to the constraints not in η, then the next iterate is taken as $\underline{x}^{(K+1)} = \underline{x}^{(K)} + \underline{\delta}^{(K)}$. If not then a line search is made in the direction of $\underline{\delta}^{(K)}$ to find the best feasible point. This can be expressed by defining the solution of (4.24) as a search direction $\underline{\delta}^{(K)}$ and choosing the step length $\alpha^{(K)}$ to solve

$$\alpha^{(K)} = \min\left(1, \min_{i \in \eta} \frac{b_i - \underline{a}_i^T \underline{x}^{(K)}}{\underline{a}_i^T \underline{S}^{(K)}}\right) \qquad (4.25)$$

so that $\underline{x}^{(K+1)} = \underline{x}^{(k)} + \alpha^{(k)}\underline{S}^{(K)}$, If $\alpha \prec 1$ in (4.24) then a constraint becomes active, defined by the index (P say) which achieves the minimum in (4.25) and this index is added to the index set η.If $\underline{x}^{(K)}$ (that is $\underline{\delta}^{(K)} = 0$) solves the current problem (4.23), then it is possible to compute multipliers ($\lambda^{(K)}$ say) for the active constraints as described in section (2), the vectors $\underline{x}^{(K)}$ and $\underline{\lambda}^{(K)}$, then satisfy all the first order conditions (Theorem(3)) except possibly the condition $\underline{\lambda}_i \geq \underline{0}$. Thus a test is made to determine whether $\underline{\lambda}_i^{(K)} \geq \underline{0} \quad \forall i \in \eta$, if so then the first order conditions are satisfied and these are sufficient to ensure a global solution. Otherwise there exists an index, q say, $q \in \eta$, such that $\lambda_q^{(K)} < 0$. In this case, it is possible to reduce $f(\underline{x})$ by allowing the constraint q to become inactive. Thus q is removed from η and the algorithm continues as before by solving the resulting problem (4.24). If there is more than one index for which $\lambda_q^{(K)} < 0$, then it is usual to select q to solve:

$$\min_{i \in \eta} \quad \lambda_i^{(K)} \qquad (4.26)$$

To summarize the algorithm, therefore, if $\underline{x}^{(K)}$ is a given feasible point and η is the corresponding active set, then the active set method is defined as follows:

(a) Given $\underline{x}^{(1)}$ and η, set $K = 1$

(b) If $\underline{\delta} = \underline{0}$ does not solve (4.26) go to (d)

(c) Compute Lagrange multipliers $\underline{\lambda}^{(K)}$ and solve (4.26). If $\underline{\lambda}_{-q}^{(K)} \geq 0$ terminate with $\underline{x}^* = \underline{x}^{(K)}$, otherwise remove q from η.

(d) Solve (4.24) for $\underline{S}^{(K)}$

(e) Find $\alpha^{(K)}$ to solve (4.25) and set $\underline{x}^{(K+1)} = \underline{x}^{(K)} + \alpha^{(K)}\underline{S}^{(K)}$

(f) If $\alpha^{(K)} < 1$ add P to η

(g) set $K = K + 1$ and go to (b)

The main effort in active set methods devoted to solve (4.24), which is an equality constraints problem. If $A^{(K)}$ is the matrix whose columns are the vectors $a_i, i \in \eta$, then (4.24) becomes:

$$\text{minimize} \qquad 0.5\underline{\delta}^T G\underline{\delta} + g^{(K)}\underline{\delta}$$

$$\text{subject to} \qquad A^{(K)T}\underline{\delta} = \underline{0} \qquad (4.27)$$

One way to solve (4.19) is to use the generalized elimination method described in section (2) so $\underline{\delta}^{(K)}$ is obtained as follows: Solve the system:

$$\left(Z^{(K)}GZ^{(K)}\right)^T \underline{\delta}_A^{(K)} = -Z^{(K)T}\underline{g}^{(K)} \qquad (4.28)$$

Where $Z^{(K)T} A^{(K)} = 0$. $\underline{\delta}_A^{(K)}$ is thus obtained by:

$$\underline{\delta}^{(K)} = Z^{(K)} \quad \underline{\delta}_A^{(K)} \qquad (4.29)$$

(4.28) and (4.29) are special cases of (4.12) and (4.10) respectively. Lagrange multipliers are recovered by

$$\underline{\lambda}^{(K)} = S^T \underline{g}^{(K)}.$$

Finally we add that Gill and Murray extended the above algorithm to cover the case when G is indefinite [13].

Example: Consider minimizing

$$f(x) = \frac{1}{2}\left[(x_1 - 2)^2 + (x_2 + 1)^2 + (x_3 + 2)^2\right]$$

Subject to non-negativity constraints (x$\geq$ 0). The gradient for this function is:

$$\nabla f(x) = (x_1 - 2, x_2 + 1, x_3 + 2)^T$$

Starting at $x_0 = (1,1,1)^T$ we have $\nabla f_0 = (-1,2,3)$, hence $p_0 = (1,-2,-3)^T$. The break points are, therefore, $\bar{\tau} = (\infty, 1/2, 1/3)$ and they are reordered into the sequence

$$\tau_0 = 0, \quad \tau_1 = 1/3, \quad \tau_2 = 1/2, \quad \tau_3 = \infty.$$

On $[0,1/3]$,

$$q(\alpha) = f\left((1,1,1)^T + \alpha(1,-2,-3)^T\right) = \frac{1}{2}\left[(-1+\alpha)^2 + (2-2\alpha)^2 + (3-3\alpha)^2\right]$$

The minimum is obviously at $\alpha = 1$, which is outside the current subinterval. So we set $x_3 = 0$ and continue.

On $[1/3, 1/2]$,

$$q(\alpha) = \frac{1}{2}\left[(-1+\alpha)^2 + (2-2\alpha)^2 + donwt1\right]$$

The minimum is again at $\alpha = 1$ which is still outside the current subinterval. So we set $x_2 = 0$ and continue.

On $[1/2, \infty)$, $q(\alpha) = \frac{1}{2}\left[(-1+\alpha)^2 + \cos t_2\right]$

The minimum is again at $\alpha = 1$, which is in the current subinterval. So we set $x_1 = 1 + \alpha = 2$. The resulting next point $\underline{x}_1 = (2,0,0)^T$ turns out to be optimal.

With the Armijo backtracking procedure outlined in $x_{k+1} = p(x_k - x_k\nabla f_k)$ we would try $\alpha = 1$ first. This yields

$$f(P((2,-1,-2)^T)) = f((2,0,0)^T) = (1+4)/2 < (1+4+9)/2 = f(x_0).$$

Lemma: Let A has full row rank, and assume that the reduced Hessian matrix $Z^T GZ$ is positive definite. Then the KKT- matrix: $\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}$ is nonsingular, and hence there is a unique vector pair (x^*, λ^*) satisfying

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -g \\ b \end{bmatrix} \qquad (4.30)$$

Proof: Suppose there are vectors w and v such that

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} w \\ v \end{bmatrix} = 0 \qquad (4.31)$$

Since $Aw = 0$, we have the form (4.31) that

$$0 = \begin{bmatrix} w \\ v \end{bmatrix}^T \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} w \\ v \end{bmatrix} = w^T GW \qquad (4.32)$$

Since w lies in the null space of A, it can be written as:

w = Zu for some vector $u \in IR^{n-m}$. Therefore, we have

$$0 = w^T GW = u^T Z^T GZu \qquad (4.33)$$

Which by positive definiteness of $Z^T GZ$ implies that $u = 0$. Therefore, $w = 0$, and by (4.31) $A^T v = 0$. Full row rank of A then implies that $v = 0$ We conclude that equation(4.31) is satisfied only if $w = 0$ and $v = 0$, so the matrix is nonsingular as claimed.

Example: Consider the quadratic programming problem

Minimize

$$q(\underline{x}) = 3x_1^2 + 2x_1x_2 + x_1x_3 + 2.5x_2^2 + 2x_2x_3 + 2x_3^2 - 8x_1 - 3x_2 - 3x_3$$

$$\text{subject to :} \quad \begin{aligned} x_1 + x_3 &= 3 \\ x_2 + x_3 &= 0 \end{aligned}$$

We can write this problem in the form

$$\text{minimize} \quad q(x) = \frac{1}{2} x^T G \underline{x} + \underline{x}^T c$$

$$\text{Subject to} \quad A\underline{x} = \underline{b}$$

By defining

$$G = \begin{bmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix}, C = \begin{bmatrix} -8 \\ -3 \\ -3 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad , \quad b = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

The solution $\underline{x}^*$ and optimal Lagrange multiplier vector $\underline{\lambda}^*$ are given by: $\underline{x}^* = (2,-1,1)^T$, $\underline{\lambda}^* = (3,-2)^T$

In this example, the matrix G is positive definite, and the null-space basis matrix can be defined as $Z = (-1,-1,1)^T$

In section (2) it is shown that how the QR factorization is used, in a generalized elimination method, to deal with equality constraints QP problems (I = ϕ), also the active set method is described to solve inequality constraints QP problems.

2 -Penalty and Barrier Functions:

Here we discuss non linear programming problems with equality and in- equality constraints. They approach used is to convert the problem into an equivalent unconstrained problem or into a problem with simple constraints. However, in practice, a sequence of problems are solved because of computational considerations. Basically, there are two alternative approaches. The first is called the penalty [25],[26], or the exterior penalty function method in which a penalty term is added to the objective function for any violation of the constraints. The second method is called a barrier or interior penalty function method, in which a barrier term that prevents the points generated from leaving the feasible region is added to the objective function. The method generates a sequence of feasible points whose limit is an optimal solution to the original problem. Methods using penalty functions transform a constrained problem into a single un constrained problem or into a sequence of unconstrained problems. The constraints are placed into the objective function. Consider the following problem with the single constraint $h(\underline{x} \geq 0)$

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

$$\text{subject to} \quad h(\underline{x}) = 0$$

suppose that this problem is replaced by the following unconstrained problem, when $\mu > 0$ is a large number

$$\text{minimize} \quad f(\underline{x}) + \mu h(\underline{x})$$

$$\text{subject to} \quad x \in E_1$$

Example: Consider the following problem

$$\text{minimize} \quad x_1^2 + x_2^2$$

$$\text{subject to} \quad x_1 + x_2 - 1 = 0$$

The optimal solution lies at the point $\left(\frac{1}{2}, \frac{1}{2}\right)$ and has objective value $\frac{1}{2}$. Now consider the following penalty problem, where $\mu > 0$ is a large number

$$\text{minimize} \quad x_1^2 + x_2^2 + \mu\,(x_1 + x_2 - 1)^2$$

$$\text{subject to} \quad (x_1, x_2) \in E_1$$

Note that for any $\mu \geq 0$, the objective function is convex. Thus a necessary and sufficient condition for optimality is that the gradient of $x_1^2 + x_2^2 + \mu(x_1 + x_2 - 1)^2$ is equal to zero yielding

$$x_1 + \mu(x_1 + x_2 - 1) = 0$$

$$x_2 + \mu(x_1 + x_2 - 1) = 0$$

Solving these two equations, we get $x_1 = x_2 = \mu/(2\mu + 1)$. Thus, the optimal solution of the penalty problem can be made arbitrarily close to the solution of the original problem by choosing μ sufficiently large.

Example: consider the problem: minimize:

$$f(\underline{x}) = -x_1 x_2$$

$$\text{subject to:} \quad g(\underline{x}) = x_1 + 2x_2 - 4 = 0$$

Suppose that this problem is solved via a penalty method using the quadratic penalty function. Then a sequence of un-constrained minimization[8], [9] ,problems minimize $I\bar{I}(x,\mu) = -x_1x_2 + \frac{1}{2}\mu(x_1 + 2x_2 - 4)^2$ is solved for increasing values of the penalty parameter μ. The necessary conditions for optimality for the unconstrained problem are

$$-x_2 + \mu(x_1 + 2x_2 - 4) = 0$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$-x_1+\mu(x_1+2x_2-4)(2)=0$$

For $\mu>\frac{1}{4}$ this yields the solution

$$x_1(\mu)=x_1=\frac{8\mu}{4\mu-1} \quad , \quad x_2(\mu)=x_2=\frac{4\mu}{4\mu-1}$$

which is a local as well as a global minimizer. (The unconstrained problem has no minimum if $\mu\le\frac{1}{4}$). Note that $x(\mu)$ is infeasible to the original constrained problem, since

$$g(x(\mu))=x_1+2x_2-4=\frac{16\mu}{4\mu-1}-4=\frac{4}{4\mu-1}$$

At any solution $x(\rho)$ we can define a Lagrange multiplier estimate $\lambda=-\mu g(x(\mu))=\frac{-4\mu}{4\mu-1}$

As μ tends to ∞ we obtain

$$\lim_{\mu\to\infty} x_1(\mu)=\lim_{\mu\to\infty}\frac{2}{1-{}^{1}\!/_{4\mu}} = 2,$$

$$\lim_{\mu\to\infty} x_2(\mu)=\lim_{\mu\to\infty}\frac{1}{1-{}^{1}\!/_{4\mu}} = 1$$

and indeed $\underline{x}^*=(2,1)^T$ is the minimizer for the constrained problem. Further

$$\lim_{\mu\to\infty} \lambda(\mu)=\lim_{\mu\to\infty}\frac{-1}{1-{}^{1}\!/_{4\mu}} = -1$$

and indeed $\lambda^*=-1$ is the Lagrange multiplier at $\underline{x}^*$

To demonstrate the Ill conditioning of the penalty function, we compute the Hessian matrix:

$$\nabla_x^2\pi(x(\mu),\mu) = \begin{pmatrix}\mu & 2\mu-1\\ 2\mu-1 & 4\mu\end{pmatrix}$$

It can be shown that its condition number is approximately $25\mu/4$, when μ is large Hessian matrix is ill conditioned.

3- Complementary Pivoting Methods:

Complementary pivoting methods[27],[30] to solve QP problems have been suggested as an extension of the simplex method for linear programming (LP) problems. Here we give a brief description of the method to solve

$$\text{minimize } 0.5\underline{x}^T G\underline{x}+\underline{g}^T\underline{x} \qquad (4.34)$$

$$\text{subject to } A^T\underline{x}\ge\underline{0} \quad , \quad \underline{x}\ge\underline{0}$$

Here G is $n\times n$ and A is $n\times m$

The method merely solves the KKT-conditions for (4.34). Introducing multipliers $\underline{\lambda}$ for the constraints $A^T\underline{x}\ge\underline{b}$ and $\underline{u}$ for the bounds $\underline{x}\ge\underline{0}$. Also define the slack variables $\underline{V}$ by $\underline{V}=A^T\underline{x}-\underline{b}$. The KKT – Conditions become:-

$$\underline{u}-G\underline{x}+A\underline{\lambda}=\underline{g}$$

$$\underline{V}-A^T\underline{x}=-\underline{b} \qquad (4.35)$$

$$\underline{x},\ \underline{\lambda},\ \underline{v} \text{ and } \underline{u}\ge\underline{0}$$

$$\text{and } \underline{v}^T\underline{\lambda}=\underline{u}^T\underline{x}=0$$

The method assumes that G is positive definite which is a sufficient condition that any solution to (4.35) is a solution to (4.34). The system (4.35) can be expressed in the form

$$\underline{W}-M\underline{Z}=\underline{q} \qquad (4.36)$$

$$w_j\ge 0,\quad Z_j\ge 0 \text{ for } j=1,\dots,n$$

$$w_j Z_j=0 \text{ for } j=1,\dots,n$$

Here (W_j,Z_j) is a pair of complementary variables.

A solution $(\underline{W},\underline{Z})$ to the above system is called a complementary feasible solution. Moreover, such a solution is a complementary basic feasible solution where

$$\text{Where } \underline{W}=\begin{pmatrix}\underline{u}\\ \underline{v}\end{pmatrix},\quad \underline{Z}=\begin{pmatrix}\underline{x}\\ \underline{\lambda}\end{pmatrix}$$

$$M=\begin{pmatrix}G & -A\\ +A^T & 0\end{pmatrix},\ \text{and}\ \underline{q}=\begin{bmatrix}\underline{g}\\ -\underline{b}\end{bmatrix}$$

All methods for solving (4.36) have some features common. They carry out row operations on the equations(4.36) in a closely related way to LP. Let $\underline{t}$ be the $2(n+m)$-vector which contains the elements of $\underline{W}$ and $\underline{Z}$ as its components. Partition the matrix $[I:M]$ into : $[I:M]=[M_B:M_N]$ (4.37)

So that $\underline{M}_B$ is $(n+m)\times(n+m)$ and non singular. Partition $\underline{t}$ accordingly into: $\underline{t}=\begin{bmatrix}\underline{t}_B\\ \underline{t}_N\end{bmatrix}$ (4.38)

So that $\underline{t}_B$ contains the components corresponding to the columns $\underline{M}_B$ and $\underline{t}_N$ contains the components corresponding to the columns of M_N. define the two sets B and N by :

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

B = { $i : t_i$ is a component of $\underline{t}_B$ } (4.39)

N = { $i : t_i$ is a component of $\underline{t}_N$ }

At a general stage, a tableau representation

$$[I : M]\begin{bmatrix} \underline{t}_B \\ \underline{t}_N \end{bmatrix} = \underline{q} \qquad (4.40)$$

of equation (4.36) is available, where $M^{\backslash} = M_B^{-1} M_N$ and $\underline{q}^{\backslash} = M_B^{-1}\underline{q}$. The variables have the values $\underline{t}_B = q^{\backslash}$ and $\underline{t}_N = \underline{0}$. $\underline{t}_B$ are called the basic variables and $\underline{t}_N$ are called the non-basic variables. A variable t_r, $r \in N$, is chosen to be increased and effect on $\underline{t}_B$ from (4.40) is given by

$$\underline{t}_B = q^{\backslash} - \underline{m}_r^{\backslash}\, t_r \qquad (4.41)$$

where $m_r^{\backslash}$ is the column of $M^{\backslash}$ corresponding to the variable t_r. An element t_i, $i \in B$, becomes zero, therefore, when $t_r = \frac{q^{\backslash}}{m_{ir}}$. This determines some variable t_P which has become zero, and so *P* and *r* are interchanged between *B* and *N* as in *LP*. The tableau is then rearranged by making row operations which reduces the old t_r column of the tableau to a unit vector. The algorithm terminates when $\underline{q} \geq 0$ and when the solution is complementary, that is $Z_i \in B \Leftrightarrow W_i \in N$ and $Z_i \in N \Leftrightarrow W_i \in B$ for all *i*. Here we give a general description of what is called the principal pivoting method for solving (4.40). The principal pivoting method is initialized by t_B = w, t_N = Z, M' = - M and

q' = *q* this is complementary[27], so the algorithm terminates if $q' \geq 0$. If not there is an element t_s, $S \in B$ for which $q_s^{'} \prec 0$, (the most negative is chosen if more than one exists), the complementary variable $t_r, r \in N$ is then chosen to be increased (t_r and t_s are complementary if (t_r, t_s) can be identified as (w_i, z_i) for some i). The effect on the basic variables by virtue of (4.41). As long as all non negative variables stay non negative, then t_r is increased until $t_s \uparrow 0$, then r and s are interchanged in B and N and the tableau is updated. The resulting tableau is again complementary so the iteration can be repeated. It may be however, that t_r is increased then a basic variable t_p $\downarrow$ 0. In this case a pivot interchange is made to give a non-complementary tableau. On the next iteration the complement of t_p is chosen to be increased. If this causes t_s $\uparrow$ 0 then complementarity is restored and the algorithm can proceed as described above. However a different t_p may become zero in which case the same operation of increasing the complement is repeated. Excluding degeneracy, each iteration increases t_s [6],[28],[29] ,so eventually complementarity is always restored.

In [6] a proof was given that the principal pivoting method solves (4.36) when G is positive definite. In the case when G is positive semi-definite a proof appeared in [2] showing that the method might work. This can be justified due to the fact that the principal pivoting method works when M belongs to special classes of matrices. In fact the resulting M when G is positive definite or positive semi-definite belongs to one of these classes (for more about that see [6] ,[21] constructed an example which shows the failure of the principal pivoting method to solve (4.36) and hence (4.34) when G is indefinite.

4- Interior-point methods:

The interior-point approach can be applied to convex quadratic programs through a simple extension of the linear programming algorithms. The resulting primal-dual algorithms are easy to describe and are quite efficient on many types of problems. Extensions of interior point methods to non-convex problems [24]. For simplicity, we restrict our attention to convex quadratic programs with inequality constraints, which we write as follows: minimize

$$q(\underline{x}) = \frac{1}{2} x^T G \underline{x} + \underline{x}^T c \qquad (4.42)$$

Subject to: $A\underline{x} \geq \underline{b}$

Where G is symmetric and positive semi definite and where the $m \times n$ matrix A and the right-hand side b are defined by :-

$$A = \{a_i\}_{i \in I}, \qquad \underline{b} = \{b_i\}_{i \in I}, \qquad I = \{1,2,...,m\}$$

(If the equality constraints are also present, they can be accommodated with simple extensions to the approaches described below). Rewriting the KKT - conditions in this notation, we obtain:

$$G\underline{x} - A^T \underline{\lambda} + \underline{C} = 0$$

$$A\underline{x} - \underline{b} \geq 0$$

$$(A\underline{x} - \underline{b})_i \lambda_i = 0, \quad i = 1,2,...,m \;\; \underline{\lambda} \geq \underline{0}$$

By introducing the slack vector $y \geq 0$, we can rewrite these conditions as:

$$G\underline{x} - A^T \underline{\lambda} + \underline{C} = 0$$

$$A\underline{x} - \underline{y} - \underline{b} = 0$$

$$y_i \lambda_i = 0, i, 1,2,....,m$$

$$(y, \lambda) \geq 0 \qquad (4.43)$$

Since we assume that G is positive semi definite, these KKT- conditions are not only necessary but also sufficient [24] , so we can solve the convex quadratic program (4.42) by finding solutions of the system (4.43).

Given accurent iterate ($\underline{x}, \underline{y}, \underline{\lambda}$) that satisfies $(y, \lambda) > 0$, we can define a complementary measure μ by

$$\mu = \frac{y^T \lambda}{m} \qquad (4.44)$$

We derive path-following, primal-dual methods by considering the perturbed KKT- conditions given by

$$F(x, y, \lambda; \sigma\mu) = \begin{bmatrix} G\underline{x} - A^T \underline{\lambda} + \bar{c} \\ A\underline{x} - y - b \\ Y\Lambda e - \sigma\mu e \end{bmatrix} = 0 \qquad (4.45)$$

Where Y = diag (y_1, y_2, … , y_m), Λ = diag $(\lambda_1, \lambda_2, \ldots, \lambda_m)$, e = $(1,1,1,\ldots,1)^T$ and $\sigma \in [0,1]$. The solution of (2.7.4) for all positive values of σ and μ define the central path, which is a trajectory that leads to the solution of the quadratic program as $\sigma\mu$ tends to zero.

By fixing μ and applying Newton's method to

(2.7.4) we obtain the linear system:

$$\begin{bmatrix} G & 0 & -A^T \\ A & -I & 0 \\ 0 & \Lambda & Y \end{bmatrix} \begin{bmatrix} \Delta \underline{x} \\ \Delta \underline{y} \\ \Delta \underline{\lambda} \end{bmatrix} = \begin{bmatrix} rd \\ r_p \\ -\Lambda Y e + \sigma\mu e \end{bmatrix} \qquad (4.46)$$

$$\begin{aligned} where \quad r_d &= G\underline{x} - A^T \underline{\lambda} + \underline{C} \\ r_p &= A\underline{x} - \underline{y} + \underline{b} \end{aligned} \qquad (4.47)$$

We obtain the next iterate by setting

$$(\underline{x}^+, \underline{y}^+, \underline{\lambda}^+) = (\underline{x}, \underline{y}, \underline{\lambda}) + \alpha(\Delta x, \quad \Delta y, \quad \Delta\lambda)$$

Where α is chosen to retain the inequality (y^+, λ^+) and possibly to satisfy various other conditions.

5- The Gradient Projection Methods:

The active set and working set change slowly, usually by a single index at each iteration [28],[29]. This method may thus require many iterations to converge on large-scale problems. For instance, if the starting point x^o has no active constraints, while 200 constraints are active at the (non degenerate) solution, then at least 200 iterations of the active-set method will be required to reach the solution.

The gradient projection method allows the active set to change rapidly from iteration to iteration. It is most efficient when the constraints are simple in form-in particular when there are only bounds on the variables. Accordingly, we restrict our attention to the following bound-constrained problem:

$$\begin{aligned} \min q(\underline{x}) &= \frac{1}{2} \underline{x}^T G \underline{x} - \underline{x}^T \underline{C} \\ subject\ to \quad & 1 \le \underline{x} \le \underline{u} \end{aligned} \qquad (4.48)$$

Where G is symmetric and l and u are vectors of lower and upper bounds on the components of $\underline{x}$. we do not may any positive definiteness assumptions on G in this section, because the gradient projection approach can be applied to both convex and non convex problems. The feasible region defined by (4,48) is sometimes called a "box" because of its rectangular shape. Some components of $\underline{x}$ may lack an upper or lower bound, we handle these cases formally by setting the appropriate components of l and u to $-\infty$ and $+\infty$ respectively.

Each iteration of the gradient projection algorithm consists of two stages. In the first stage we search along the steepest descent direction from the current point $\underline{x}$, that is, the direction $-\underline{g}$, where $\underline{g} = G\underline{x} + \underline{C}$, whenever abound is encountered, the search direction is "bent" so that it stays feasible. We search along the resulting piecewise-linear path and locate the first local minimizer of q, which we denote by x^c and refer to as the Cauchy point, [24]. The working set is now defined to be the set of bound constraints that are active at the Cauchy point, denoted by $A(x^c)$. In the second stage of each gradient projection iteration, we explore the face of the feasible box on which the Cauchy point lies by solving a sub problem in which the active components x_i for $i \in A(x^c)$ are fixed at the value x_i^c.

Givens Rotations:-

The QR factorization of a matrix[24],[25] can be computed using Givens rotations. A Givens rotation is any matrix of the form:

$$\begin{bmatrix} 1 & & & & \\ & c & & -s & \\ & & 1 & & \\ & s & & c & \\ & & & & 1 \end{bmatrix} \qquad (4.49)$$

Where $c^2 + s^2 = 1$. the i and j subscript in G_{ij} correspond to the row numbers associated with the c's : The first c is in row j and the second c is in row i. Observe that the intersection of row i and row j with column i and column j for G_{ij} is the 2×2 matrix

$$\begin{bmatrix} c & -s \\ s & c \end{bmatrix}.$$

This matrix is orthogonal since multiplication by its transpose yields I:

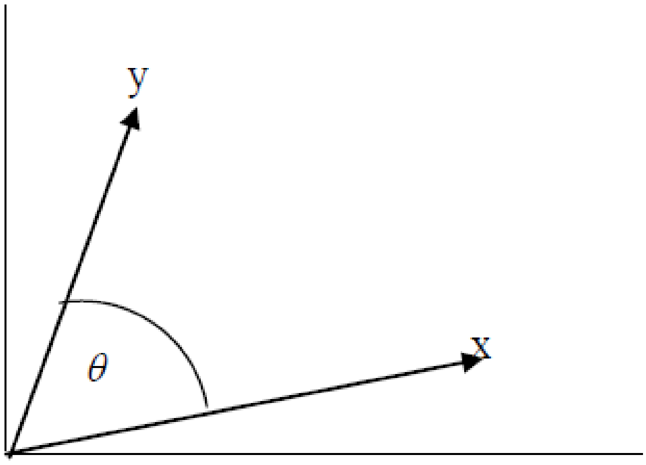
$$\begin{bmatrix} c & -s \\ s & c \end{bmatrix}^T \begin{bmatrix} c & -s \\ s & c \end{bmatrix} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} c & -s \\ s & c \end{bmatrix} = \begin{bmatrix} c^2 + s^2 & 0 \\ 0 & c^2 + s^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

.

Similarly, matrix (4.49) is orthogonal since $G_{ij}^T G_{ij} = I$. Matrix (4.49) is called a rotation for the following reason: If two vectors x and y satisfy the relation

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad (4.50)$$

then y is x rotated by the angle θ (see the Figure)



Consequently, multiplying a vector by the Givens rotation (4.49) is equivalent to rotating two components of the vector through the angle $\theta = \arctan s/c$ while leaving the other components intact .

Given a vector x with two components and defining

$$c = \frac{x_1}{\sqrt{x_1^2 + x_2^2}} \quad \text{and} \quad s = -\frac{x_2}{\sqrt{x_1^2 + x_2^2}}, \qquad (4.51)$$

Observe that $c^2 + s^2 = 1$ and

$$\begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \sqrt{x_1^2 + x_2^2} \\ 0 \end{bmatrix} \qquad (4.52)$$

Utilizing relation (4.52), we can construct a Givens rotation that annihilates a particular matrix element. To QR factor A, we multiply A by a sequence of Givens rotations, annihilating elements beneath the diagonal to obtain the upper triangular matrix R. Finally, Q is the product of rotations.

We illustrate this algorithm with the matrix

$$A = \begin{bmatrix} 90 & -153 & 114 \\ 120 & -79 & -223 \\ 200 & -40 & 395 \end{bmatrix}.$$

The factorization starts with column 1 of A: $\begin{bmatrix} 90 \\ 120 \\ 200 \end{bmatrix}$.

Referring to (4.51) and identifying x_1 with 90 and x_2 with 120, the Givens rotation G_{21} that annihilates the second component of column 1 is:

$$G_{21} = \begin{bmatrix} \frac{90}{150} & \frac{120}{150} & 0 \\ -\frac{120}{150} & \frac{90}{150} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} & 0 \\ -\frac{4}{5} & \frac{3}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Pre-multiplying A by G_{21} yields a matrix that we denote A_{21}:

$$A_{21} = G_{21}A = \begin{bmatrix} .6 & .8 & 0 \\ -.8 & .6 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 90 & -153 & 114 \\ 120 & -79 & -223 \\ 200 & -40 & 395 \end{bmatrix} = \begin{bmatrix} 150 & -155 & -110 \\ 0 & 75 & -225 \\ 200 & -40 & 395 \end{bmatrix} \qquad (4.53)$$

Since each coefficient on the right side of (4.53) is a multiple of 5, we extract the factor 5 to obtain:

$$A_{21} = 5\begin{bmatrix} 30 & -31 & -22 \\ 0 & 15 & -45 \\ 40 & -8 & 79 \end{bmatrix}.$$

Referring to (4.51) and identifying x_1 with 30 and x_2 with 40, the Givens rotation G_{31} that annihilates the third component in column 1 of A_{21} is

$$G_{31} = \begin{bmatrix} .6 & 0 & .8 \\ 0 & 1 & 0 \\ -.8 & 0 & .6 \end{bmatrix}$$

Multiplying A_{21} by G_{31} gives

$$A_{31} = G_{31}A_{21} = 5\begin{bmatrix} .6 & 0 & .8 \\ 0 & 1 & 0 \\ -.8 & 0 & 6 \end{bmatrix} \begin{bmatrix} 30 & -31 & -22 \\ 0 & 15 & -45 \\ 40 & -8 & 79 \end{bmatrix} = 25\begin{bmatrix} 10 & -5 & 10 \\ 0 & 3 & -9 \\ 0 & 4 & 13 \end{bmatrix}.$$

At this point, the sub-diagonal elements in the first column of A_{31} are zero. Proceeding to the second column of A_{31}, the Givens rotation G_{32} that annihilates the third element in the second column is: $G_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & .6 & .8 \\ 0 & -.8 & .6 \end{bmatrix}$

Pre-multiplying A_{31} by G_{32} yields *R*:

$$R = G_{32}A_{31} = 25\begin{bmatrix} 1 & 0 & 0 \\ 0 & .6 & .8 \\ 0 & -.8 & .6 \end{bmatrix} \begin{bmatrix} 10 & -5 & 10 \\ 0 & 3 & -9 \\ 0 & 4 & 13 \end{bmatrix} = 125\begin{bmatrix} 2 & -1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

To summarize,

$$R = G_{32}A_{31} = G_{32}G_{31}A_{21} = G_{32}G_{32}G_{21}A. \qquad (4.54).$$

Since each Givens rotation G_{ij} is an orthogonal matrix, $G_{ij}^T G_{ij} = I$. multiplying the identity (2.9.6) by

$G_{21}^T G_{31}^T G_{32}^T$ gives us the relation: $G_{21}^T G_{31}^T G_{32}^T R = G_{21}^T G_{31}^T G_{32}^T G_{32} G_{31} G_{21} A = A$.

It follows that A = QR, where $Q = G_{21}^T G_{31}^T G_{32}^T$. In the example above, the QR factorization can be expressed:

$$A = \begin{bmatrix} 90 & -153 & 114 \\ 120 & -79 & -223 \\ 200 & -40 & 395 \end{bmatrix} = \left(\frac{1}{125} \begin{bmatrix} 45 & -108 & 44 \\ 60 & -19 & -108 \\ 100 & 60 & 45 \end{bmatrix} \right) \left(125 \begin{bmatrix} 2 & -1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix} \right)$$

where the first factor on the right side is $Q = G_{21}^T G_{31}^T G_{32}^T$, and the second factor is R. For a general $n \times n$ matrix, we annihilate elements 2 through n in the first column using the sequence of Givens rotations $G_{21}, G_{31}, ..., G_{n1}$. then elements 3 through n in the second column are annihilated using the sequence of Givens rotations $G_{32}, G_{42}, ..., G_{n2}$. we continue in this way. The last step annihilates element n of column n – 1 using $G_{n,n-1}$. Finally, Q is the product of the transpose of each rotation. Factoring a matrix using Givens rotations is a little slower than factoring a matrix using Householder reflections since the Givens scheme requires about twice as many multiplications.

VI. CONCLUSION

The work reported in this paper gave no account to the special structures that the matrix of constraints A might have. The work is ideal when A is dense, that is, full of non-zero elements. There are many stable and efficient methods for solving quadratic programming as Steepest decent methods, DFP methods, Dantzig and the extended Dantzig Wolfe method which solve quadratic functions for definite and indefinite Hessian matrices.

REFERENCES

1. Alkhayyal. A. F. (1987) An implicit Enumeration Procedure. For The General Linear Complementary problem .Mathematical programming Study, 31:1-20.
2. Bazaraa, Shetty and Sherali, (1994). Nonlinear Programming: Theory and Application Wiley.
3. Bunday Brian (1984) Basic optimization Methods. Edward Arnold.
4. Burdet G. A. (1971) General quadratic programming. Carriage-Mellon University. Paper W, p. 4 – 71 -2.
5. Coleman and Li,Large Scale (1990). Numerical Optimization. SIAM Books.
6. Cottle. R.W.(1990).The Principle Pivoting method positive visited math program 48, 369-385..
7. Bertsekas D.P. (1991),Linear Network Optimization: Algorithms and codes MIT Press Cambridge, M .A.
8. David G. Luenberger (2003) Linear and nonlinear programming 2nd Edition. Pearson Education, Inc. Publishing as Addison-Wesley.
9. Dennis and Schnabel (1996) Numerical Methods for unconstrained Optimization and nonlinear equation classics in applied Mathematics. SIAM.
10. Fletcher R. (1987) Practical Methods of Optimization, 2nd Ed. John Wiley and Sons.
11. Gill. P. E. and Murray. W. (1975) Numerical Methods for constrained optimization (Eds. P. E. Gill and W. Murray), Academic press, London.
12. Gill. P. E. and Murray. W. (1978) Numerically Stable methods for quadratic programming. Math Prog 14, 349 - 372.
13. Gill. P. E. and Murray. W. (1988) Practical Optimization. Academic Press.
14. Golub G. H. and Van Loan. C. F. (1983) Matrix computations, the John Hofkin University Press.
15. Dennis J. E. and R.B.Schnabel (1983). Numerical Methods for Unconstrained Optimization and Nonlinear Equations,Prentice Hall, Englewood Cliffs,NJ.
16. Kendall, E. (1978). An introduction To Numerical Analysis. John Wiley & Sons.
17. Kunisch K, Rendl F. (2003).An infeasible active set method for quadratic problems with simple bounds.siam journal on optimization volume 14,number1 pp35-52.
18. Karmarkar's (1991).Linear programming algorithm and Newton's method.Mathematical programming 50.291-330.
19. Keller.E.L.(1973).The general quadratic optimization problem. Math. Prog. 3, 311-337.
20. Mohsin H. A. Hashim (1996) An Extension to the Dantzig-Wolfe Method for general quadratic programming Ph.D. thesis, University of Khartoum.
21. M.M . Khader,N.H .Sweilam &A. M.S.Mahdy (2011). An Efficient Numerical Method for Solving the Fractional Diffusion Equation .
22. Byrd R. H, Gillbert J., C., and. Nocedal.J.(2000) A trust region method based on interior point techniques for nonlinear programming, 9, pp. 149 - 185.
23. Boyd .S. and. Vandenberghe .L..(2003). Convex Optimization. Cambridge University.
24. Jorge Nocedal and Stephen Wright books (2006) Springer Series in Operations Research and Financial Engineering. Springer.
25. Stephen G. Nash and Ariela Sofer (1996) . Linear and non -liner programming, McGraw Hill, New York.
26. Stephen G. Nash and Ariela Sofer (2008) . Linear and non -liner programming, McGraw Hill, New York