QUADRATIC PROGRAMMING (QP) ALGORITHMS
&
MEAN-VARIANCE PORTFOLIO OPTIMISATION

QP is the **optimization of a quadratic function subject to linear equality and inequality constraints**. It arises in least squares (e.g curve fitting) problems under constraints. Similarly, it arises in multiple objective decision making where the departure, or deviation, of the actual decisions from their corresponding ideal, or bliss, value is evaluated using a quadratic distance measure.

In nonlinear programming, sequential quadratic programming algorithms require the solution of **QP** problems to determine the direction of search at each iteration.

QP is also widely used in portfolio optimisation in finance in the formulation of mean-variance optimization of investment decisions under uncertainty. The constraints are linear and the objective function is quadratic (Markowitz, 1959). The decision maker needs to reconcile the conflicting desires of maximizing expected portfolio return, represented by the linear portfolio return term, and minimizing the portfolio risk, represented by the quadratic portfolio variance term, in the objective function.

1. INTRODUCTION

The **QP** multi-objective problem (for v, u $\in \mathbb{R}^n$, $< v, u > \; \equiv \; v^T u = \sum_{i=1}^{n} v^i u^i$):

$$\boxed{\min \left\{ < x - x^d, \; \mathcal{Q} \left(x - x^d \right) > \; \middle| \; \mathcal{H}^T x \le h \right\}}. \tag{1.1}$$

x^d is the **desired** or **bliss value** of x $\in \mathbb{R}^n$, $\mathcal{Q}$ is a symmetric positive definite matrix penalizing the deviations of x from x^d. The matrix $\mathcal{H} \in \mathbb{R}^{n \times i}$ and the column vector h $\in \mathbb{R}^i$ represent the linear inequality constraints with

$$\mathcal{H} = [\, \mathfrak{H}^1 \quad \mathfrak{H}^2 \quad \ldots \quad \mathfrak{H}^{\mathrm{i}} \quad \ldots \quad \mathfrak{H}^{i} \,],$$

where $\mathfrak{H}^{\mathrm{i}}$ is the *n*-dimensional column vector of coefficients for the i-th constraint.

Nonlinear optimization algorithms require the sequential solution of **QP** subproblems of the type

$$\boxed{\min \left\{ < a, x > + \tfrac{1}{2} < x, \mathcal{Q} \, x > \; \middle| \; \mathcal{H}^T x \le h \right\}} \tag{1.2}$$

where a $\in \mathbb{R}^n$ is a constant vector. It is clear that (1.1) is equivalent to (1.2) and that (1.2) can be formulated as (1.1) using $x^d = -\mathcal{Q}^{-1} a$, which is the unconstrained optimum of the objective function of (1.2).

There are numerous methods for solving **QP** problems. The first, and historically most important, class of methods are those relying on an active set strategy. These methods identify a feasible point that lies on some of the inequality constraints. They generate a descent direction that lies in the intersection of all the constraints currently satisfied as equalities. A step along this direction is either restricted by another inequality constraint or eventually reaches the optimum within the intersection of the constraints currently satisfied as equalities. If a constraint is encountered, then the constraint is added to the active set and the optimum is sought in the intersection of the new set of constraints. If the optimum in the intersection of currently active constraints is attained, then the multipliers at this optimum indicate which of the currently active constraints can be dropped from the active set to generate a direction of further progress. A sequence of equality constraints is solved, each considering a subset of the constraints as active, or satisfied as equalities. For non-degenerate problems, each equality constrained problem leads to an improvement in the quadratic objective function. Thus, it is not possible to return to the same active set. As there are a finite number of active sets, convergence is ensured in finite number of steps.

There are two types of active set methods depending on the subspace in which the direction of progress is generated. One is based on the range space of the matrix of active constraints and the other on the null space. We discuss the former in Section 2, the latter in Section 3. The basic operation and progress of both algorithms are similar and are highlighted in Algorithm 2.1. In Section 4, we discuss more recent, interior point, methods based barrier functions. These introduce a logarithmic penalty function to penalize the transgression of the inequality constraints. The algorithms then solve the optimality condition of the overall problem by applying a nonlinear equation solver.

2. ACTIVE SET ALGORITHM: NULL SPACE OF CONSTRAINTS

Let the indices of **inequality constraints active at $\mathbf{x_k}$** be represented by the subset

$$\boxed{\mathcal{I}_k = \left\{ \, i \;\middle|\; < \mathfrak{H}^i \, , \; x_k > \; = \; h^i \right\}} \tag{2.1, a}$$

and in matrix form, we represent these active constraints as

$$\boxed{\mathcal{H}_k^T \, x_k \; = \; h_k}, \tag{2.1, b}$$

$$\mathcal{H}_k = [\, \mathfrak{H}^1 \quad \mathfrak{H}^2 \quad \ldots \quad \mathfrak{H}^i \quad \ldots \quad \mathfrak{H}^{i_k} \,],$$

where $\mathfrak{H}^i$ is the n-dimensional column vector of coefficients for the i-th constraint, i_k is the number of active constraints at x_k, the matrix $\mathcal{H}_k \in \mathbb{R}^{n \times i_k}$ and the column vector $h_k \in \mathbb{R}^{i_k}$. Methods in this class generate a basis of the null space of $\mathcal{H}_k^T$. Starting from a feasible point, the direction of progress towards the solution is then defined as a linear combination of the vectors spanning this null space. This approach was pioneered by Gill and Murray (1978).

Let $Z \in \mathbb{R}^{n \times n - i_k}$ be a matrix whose columns span the subspace orthogonal to the columns of $\mathcal{H}_k$, hence $\mathcal{H}_k^T \, Z = 0$.

Example 1: Consider the QP Problem

$$\min_{x^1,\, x^2} \; (x^1 - 1)^2 + (x^2 - 2.5)^2$$

Subject to:

Constraint 1: $\quad - x^1 + 2x^2 \leq 2;$

Constraint 2: $x^1 + 2x^2 \leq 6;$
Constraint 3: $x^1 - 2x^2 \leq 2;$
Constraint 4: $x^1 \geq 0;$
Constraint 5: $x^2 \geq 0$

Given the initial point $x_0 = \begin{bmatrix} x_0^1 \\ x_0^1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, we have, for constraints 3 and 5:

$$x^1 - 2x^2 = 2 - 2\,(0) = 2$$

$$x^1 = 0,$$

hence $\mathcal{I}_0 = \left\{ 3, 5 \right\}$. □

Example 2: Upper and lower bounds

For upper and lower bound constraints such as $x_l \leq x \leq x_u$, where x_l, x_u are fixed vectors, we define a vector $J(x_k) \in \mathbb{R}^n$ as

$$\boxed{J^i\,(x_k) = \begin{cases} 0, & x_k^i \neq x_l^i \text{ or } x_u^i \\ 1, & x_k^i = x_u^i \\ -1, & x_k^i = x_l^i \end{cases}.}$$

Let i_k denote the nonzero elements of $J(x_k)$. $Z \in \mathbb{R}^{n\times(n-i_k)}$ is the matrix consisting of the non-null columns of the diagonal matrix whose i th diagonal element is

$$1 - \left| J^i\,(x_k) \right| .$$

Thus, Z defines the subspace of the non-fixed (free) variables. □

****** Z MATRIX-DISCUSSION UNTIL JUST AFTER (2.2) BELOW******
IS FOR THE INTERESTED STUDENT

There are two ways of representing the matrix Z in general. The first involves the partitioning of the matrix $\mathcal{H}_k$ as

$$\mathcal{H}_k = \begin{bmatrix} \mathcal{H}_B \\ \mathcal{H}_N \end{bmatrix} \qquad (2.1, z)$$

where $\mathcal{H}_B$ is a $i_k \times i_k$ nonsingular matrix; x is conformably partitioned as $x = [x_B, x_N]$. Then the constraints (2.1) can be written as

$$\mathcal{H}_B^T\, x_B + \mathcal{H}_N^T\, x_N = h_k \qquad (2.2, z)$$

and since $\mathcal{H}_B$ is nonsingular, we have

$$x_B = \mathcal{H}_B^{-T} \left(h_k - \mathcal{H}_N^T\, x_N \right). \qquad (2.3, z)$$

The i_k variables x_B corresponding to the columns of $\mathcal{H}_B^T$ are the endogenous variables dependent on the remaining $n - i_k$ exogenous or independent variables. Thus, for any x_N, x_B given as above satisfy the linear equality constraints $\mathcal{H}_k\, x = h_k$. As we are able to express the variables x_B in terms of x_N, we are able to eliminate all other occurrences of x_B, such as in the objective function, using this relationship. This technique is called the variable reduction method and Z given by

$$Z = \begin{bmatrix} -\mathcal{H}_B^{-T}\mathcal{H}_N^T \\ \cdots \\ I \end{bmatrix} \tag{2.4, z}$$

satisfies the orthogonality condition

$$\mathcal{H}_k^T Z = [\mathcal{H}_B^T \vdots \mathcal{H}_N^T] \begin{bmatrix} -\mathcal{H}_B^{-T}\mathcal{H}_N^T \\ \cdots \\ I \end{bmatrix} = 0. \tag{2.5, z}$$

The second way of obtaining the matrix Z is based on the Q R decomposition of $\mathcal{H}_k$ where $Q \in \mathbb{R}^n \times \mathbb{R}^n$ is an orthonormal matrix such that $Q^TQ = I$ and $R \in \mathbb{R}^{i_k} \times \mathbb{R}^{i_k}$ is a non-singular upper triangular matrix such that

$$\mathcal{H}_k = Q \begin{bmatrix} R \\ 0 \end{bmatrix}. \tag{2.2}$$

Let Q be partitioned into two submatrices

$$Q = [\, Y \vdots Z \,]$$

where $Y \in \mathbb{R}^{n\times i_k}$ and $Z \in \mathbb{R}^{n\times(n - i_k)}$ and $\mathcal{H}_k^T Z = 0$. As Q is orthonormal, we have

$$Z^T Z = I \in \mathbb{R}^{(n - i_k)\times(n - i_k)}.$$

The Q R decomposition of matrix can be obtained using either Householder transformations or Givens rotations.

****** THE Z-DISCUSSION ABOVE IS FOR THE INTERESTED STUDENT ******

To generate the basic search direction, let x_k be a feasible point, with active constraints (2.1, a), and let the direction be given by $d_k = \bar{x} - x_k$. We consider the first order necessary conditions of the equality constrained problem

$$\boxed{\min \left\{ < a, x > + \tfrac{1}{2} < x, \mathcal{Q} x > \;\middle|\; \mathcal{H}_k^T x = h_k \right\}} \tag{2.3}$$

given by

$$a + \mathcal{Q}\bar{x} + \mathcal{H}_k \mu_{k+1} = 0 \tag{2.4, a}$$

$$\mathcal{H}_k^T \bar{x} = h_k. \tag{2.4, b}$$

These can be written as

$$\begin{aligned} a + \mathcal{Q} x_k + \mathcal{Q} d_k + \mathcal{H}_k \mu_{k+1} &= 0 \\ \mathcal{H}_k^T d_k &= 0 \end{aligned} \tag{2.5}$$

and d_k satisfying $\mathcal{H}_k^T d_k = 0$ can be expressed as a linear combination of the columns of Z. Thus, we have, $d_k = Z\omega$ and $\mathcal{H}_k^T Z\omega = 0$. Hence, premultiplying (2.5) by Z^T yields

$$Z^T \mathcal{Q} Z \omega = - Z^T \left(a + \mathcal{Q} x_k \right) \tag{2.6}$$

[Note: $Z^T \mathcal{Q} Z$: $(n - i_k) \times (n - i_k)$] thus, d_k is given by

$$\boxed{d_k = - Z \left(Z^T \mathcal{Q} Z \right)^{-1} Z^T \left(a + \mathcal{Q} x_k \right)}. \tag{2.7}$$

$Z^T \mathcal{Q} Z$ is referred to as the *projected Hessian*, or the projection of $\mathcal{Q}$ on the intersection of the constraints $\mathcal{H}_k^T x = h_k$. Thus, $Z^T \mathcal{Q} Z$ is positive definite if the Hessian $\mathcal{Q}$ is positive definite in the intersection of the constraints. The positive definiteness of $\mathcal{Q}$ clearly ensures that $Z^T \mathcal{Q} Z$ is positive definite as well. However, $Z^T \mathcal{Q} Z$ may be positive definite even when $\mathcal{Q}$ is not (see Gill, Murray and Wright, 1981). Also, as $Z^T \mathcal{Q} Z \in \mathbb{R}^{(n - i_k) \times (n - i_k)}$, the dimension of the linear equations that need to be solved in (2.6), or equivalently size of the matrix that needs to be inverted to compute d_k, reduces as the number of the active constraints increases.

d_k is a *descent direction*: starting from x_k, any positive step along direction d_k produces a reduction in the quadratic objective function. Consider the objective function at

$$x(\tau) = x_k + \tau d_k, \tag{2.8}$$

where $\tau \in (0, 1]$ is the stepsize is to be determined. From (2.8) we have

$$\begin{aligned} &< a, x(\tau) > + \tfrac{1}{2} < x(\tau) , \mathcal{Q} x(\tau) > \\ &\quad = < a, x_k > + \tfrac{1}{2} < x_k , \mathcal{Q} x_k > + \tau_k < d_k, a + \mathcal{Q} x_k > + \tfrac{\tau_k^2}{2} < d_k, \mathcal{Q} d_k > . \end{aligned}$$

It follows from d_k given by (2.7) that

$$\begin{aligned} < d_k, \mathcal{Q} x_k + a > &= - < \left(a + \mathcal{Q} x_k \right), Z \left(Z^T \mathcal{Q} Z \right)^{-1} Z^T \left(a + \mathcal{Q} x_k \right) > \\ < d_k, \mathcal{Q} d_k > &= \left(a + \mathcal{Q} x_k \right)^T Z \left(Z^T \mathcal{Q} Z \right)^{-1} Z^T \mathcal{Q} Z \left(Z^T \mathcal{Q} Z \right)^{-1} Z^T \left(a + \mathcal{Q} x_k \right) \\ &= < \left(a + \mathcal{Q} x_k \right), Z \left(Z^T \mathcal{Q} Z \right)^{-1} Z^T \left(a + \mathcal{Q} x_k \right) > \end{aligned}$$

As the columns of Z are linearly independent and $Z^T \mathcal{Q} Z$ is positive definite, we have

$$\begin{aligned} < a, x(\tau) > + \tfrac{1}{2} < x(\tau), \mathcal{Q} x(\tau) > &= < a, x_k > + \tfrac{1}{2} < x_k, \mathcal{Q} x_k > \\ &\quad + \left(\tfrac{\tau_k^2}{2} - \tau_k \right) < \left(a + \mathcal{Q} x_k \right), Z \left(Z^T \mathcal{Q} Z \right)^{-1} Z^T \left(a + \mathcal{Q} x_k \right) > \\ &< < a, x_k > + \tfrac{1}{2} < x_k , \mathcal{Q} x_k > \end{aligned} \tag{2.9}$$

since $\left(\frac{\tau_k^2}{2} - \tau_k \right) < 0$ for $\tau_k \in (0, 1]$.

Optimal stepsize τ along d_k (ignoring the constraints not in the current active set): As the objective is quadratic, its unconstrained optimum along d_k, in the intersection of the constraints in the active set $\mathcal{I}_k$, is located where the derivative

$$\frac{d\left(<a, x(\tau)> + \frac{1}{2}<x(\tau), \mathcal{Q}\, x(\tau)>\right)}{d\,\tau} \;=\; (\tau_k - 1) < \left(a + \mathcal{Q}\, x_k\right), Z\left(Z^T \mathcal{Q} Z\right)^{-1} Z^T \left(a + \mathcal{Q}\, x_k\right) > \;=\; 0. \tag{2.10}$$

For d_k given by (2.7), this is achieved at $\tau = 1$. However, proceeding from x_k to

$$\boxed{x(1) \;=\; x_k \;+\; d_k},$$

if a constraint, not in $\mathcal{I}_k$, is encountered, then the stepsize has to be shortened to ensure that this constraint is not transgressed. The distance from x_k, along d_k to constraint j, **not** in $\mathcal{I}_k$ is given by the value of τ that satisfies

$$< \mathfrak{H}^j, x(\tau) > \;=\; < \mathfrak{H}^j, x_k + \tau\, d_k > \;=\; h^j,\; j \notin \mathcal{I}_k.$$

To ensure feasibility, the smallest value of τ, over all those constraints, is required and hence

$$\boxed{\tau_k \;=\; \min_{j \notin \mathcal{I}_k} \left\{ 1,\;\; \frac{h^j - <\mathfrak{H}^j, x_k>}{<\mathfrak{H}^j, d_k>} \;\middle|\; < \mathfrak{H}^j, d_k > \; > \; 0 \right\}}. \tag{2.11}$$

The condition $< \mathfrak{H}^j, d_k > \; > \; 0$ simply ensures that only those constraints that are likely to be violated are considered. From (2.11), we can set $x_{k+1} = x_k + \tau_k\, d_k$. If $\tau_k < 1$, the constraints that have been encountered are added to the active set, thus forming $\mathcal{I}_{k+1}$. If further descent in the intersection of the constraints in $\mathcal{I}_{k+1}$ is possible, this is done by generating descent directions similar to (2.7).

If the optimum in the intersection of a given set of active constraints is attained, then the multipliers at that point indicate if one of the currently active constraints can be dropped from $\mathcal{I}_k$ in order to generate a further descent direction.

To establish the basic results for dropping a constraint from the active set, let $\mathcal{I}_k$, $\mathcal{H}_k$, h_k be obtained by *dropping the i th constraint from* $\mathcal{I}_{k-1}$, and hence the i th column and element of $\mathcal{H}_{k-1}$ and h_{k-1} respectively. We write the first order necessary conditions (2.4, a, b) for the equality constrained problem with active set $\mathcal{I}_{k-1}$ as

$$a \;+\; \mathcal{Q}\, x_k \;=\; -\; \mathcal{H}_{k-1}\, \mu_k \tag{2.12, a}$$

$$\mathcal{H}_{k-1}^T\, x_k \;=\; h_{k-1}. \tag{2.12, b}$$

The same conditions for the equality constrained problem (2.3), with active set $\mathcal{I}_k$, are

$$a \;+\; \mathcal{Q}\, x_{k+1} \;=\; -\; \mathcal{H}_k\, \mu_{k+1} \tag{2.12, c}$$

$$\mathcal{H}_k^T\, x_{k+1} \;=\; h_k. \tag{2.12, d}$$

The multipliers μ_{k+1} are evaluated as the least squares solution to (2.5). When the columns of $\mathcal{H}_k$ are linearly independent, this is given by

$$\boxed{\mu_{k+1} \;=\; -\; \left(\mathcal{H}_k^T\, \mathcal{H}_k\right)^{-1} \mathcal{H}_k^T\, [a + \mathcal{Q}\, x_k + \mathcal{Q}\, d_k]}. \tag{2.13}$$

We show in Section 3 below that this choice of μ_{k+1} is consistent with the direction d_k and (2.12, c).

Example 3:

Consider

$$\min_{x^1, x^2, x^3} \; 3\,(x^1)^2 + 2\,x^1x^2 + x^1x^3 + 2\,x^2x^3 + 2.5\,(x^2)^2 + 2\,(x^3)^2 \;-\; 8\,x^1 \;-\; 3\,x^2 \;-\; 3\,x^3$$

subject to

$x^1 + x^3 = 3;$

$x^2 + x^3 = 0.$

We can write this problem in the form (2.3) by defining

$$\mathcal{Q} = \begin{bmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix};\; a = \begin{bmatrix} -8 \\ -3 \\ -3 \end{bmatrix};\; \mathcal{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix};\; h = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

It is shown below that the null space matrix Z is given by

$$Z = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

which ensures $Z^T \mathcal{H} = 0$ and can be used with (2.7) to derive the optimum solution from any feasible initial point, e.g.

$$x_0 = \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix}$$

and the associated multipliers. To derive Z and compute the optimum, note that $\mathcal{H}^T x = b$ is given by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

or,

$$\mathcal{H}_B^T\, x_B + \mathcal{H}_N^T\, x_N = h_k$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x_B + \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_N = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

$$Z = \begin{bmatrix} -\mathcal{H}_B^{-T}\,\mathcal{H}_N^T \\ I \end{bmatrix} = \begin{bmatrix} -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}.$$

Using (2.7), we have

$$d = -\,Z\left(Z^T \mathcal{Q}\, Z\right)^{-1} Z^T\left(a + \mathcal{Q}\, x_0\right)$$

$$= -\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}\left[[-1 \;\; -1 \;\; 1] \begin{bmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}\right]^{-1} [-1 \;\; -1 \;\; 1]\left(a + \mathcal{Q}x_0\right)$$

where

$$\left(a + \mathcal{Q}\, x_0\right) = \begin{bmatrix} -8 \\ -3 \\ -3 \end{bmatrix} + \begin{bmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix} = \begin{bmatrix} -8 \\ -3 \\ -3 \end{bmatrix} + \begin{bmatrix} -9 \\ -9 \\ 6 \end{bmatrix} = \begin{bmatrix} -11 \\ -12 \\ 3 \end{bmatrix}$$

and thus

$$d = -\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} [13]^{-1} \begin{bmatrix} -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} -11 \\ -12 \\ 3 \end{bmatrix} = -\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} [13]^{-1} [26] = \begin{bmatrix} 2 \\ 2 \\ -2 \end{bmatrix}$$

Thus, the solution is

$$x_* = x_0 + d = \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

with the multipliers given by:

$$x_* = \; ; \quad \mu_* = \begin{bmatrix} 3 \\ -2 \end{bmatrix}. \qquad \square$$

The following is a summary of the basic null space algorithm. The statement of the range space algorithm discussed in Section 3 is basically the same, except that the direction d_k and the multiplier μ_{k+1} are evaluated using an alternative formulation.

Algorithm 2.1

Step 0: Given an initial feasible point x_0, identify the active set $\mathcal{I}_0$. Set k =0

Step 1: Solve the equality constrained **QP** problem (2.3) to obtain $d_k = \bar{x} - x_k$.

If $d_k = x_{k+1} - x_k = 0$, go to **Step 3**.

Step 2: Set $x_{k+1} = x_k + \tau_k d_k$, where τ_k is determined by (2.11).

If $\tau_k = 1$, go to **Step 3**.

Otherwise, add the new constraint determined by (2.11) to the active set $\mathcal{I}_k$ to form $\mathcal{I}_{k+1}$.

Set k = k + 1 and go to **Step 1**.

Step 3: Compute the multipliers μ_{k+1} using (2.13).

Determine the minimum element of μ_{k+1}.

If this element $\mu_k^i < 0$, drop the i th constraint from active set $\mathcal{I}_k$ to obtain $\mathcal{I}_{k+1}$, set k = k + 1, go to **Step 1**. If $\mu_k^i \geq 0, \forall\, i \in \mathcal{I}_k$, x_{k+1} is optimal; stop.

Clearly, $d_k = 0$ occurs when Algorithm 2.1 has reached the optimum $\bar{x} = x_k$ in the intersection of the currently active constraints. The multipliers need to be checked to see if, by dropping any of the currently active constraints, a further feasible descent direction can be generated. Finally, as the active set changes, it is necessary to compute Z as well as other matrices involved in the computation of d_k and μ_{k+1}. Gill et al (1974) discuss numerically stable methods for modifying these matrices.

Proposition 2.1

Let $\mathcal{Q}$ be positive definite. If $\exists$ a negative element of μ_k , say $\mu_k^i < 0$, and $\mathcal{I}_{k+1}$, $\mathcal{H}_{k+1}$, h_{k+1} are obtained by deleting corresponding constraint i from active set, then

- (i) $d_k = x_{k+1} - x_k$ is a descent direction at x_k; hence $< d_k, \left(a + \mathcal{Q}\, x_k\right) > \; < 0$;
- (ii) direction d_k satisfies the i th constraint just dropped from $\mathcal{I}_k$ and hence x_{k+1} is feasible for the i th constraint; and,
- (iii) if μ_k does not have a negative element, then x_k satisfies the first order necessary conditions of the **QP** problem.

Proof

To establish (i), we write

$$
\begin{aligned}
< d_k, \left(a+ \mathcal{Q}\, x_k\right) > \;&=\; < d_k, \left(a + \mathcal{Q}\, x_k \;-\; \mathcal{Q}\, x_{k+1} \;+\; \mathcal{Q}\, x_{k+1}\right) > \\
&=\; < d_k, \left(\left(a + \mathcal{Q}\, x_{k+1}\right) \;-\; \mathcal{Q}\left(x_{k+1} \;-\; x_k\right) \right) > \\
&=\; < d_k, \left(\left(a + \mathcal{Q}\, x_{k+1}\right) \;-\; \mathcal{Q}\, d_k \right) > \\
&<\; < d_k, \left(a + \mathcal{Q}\, x_{k+1}\right) > && (2.14) \\
&=\; - < d_k, \mathcal{H}_k\, \mu_{k+1} > \\
&=\; 0. && (2.15)
\end{aligned}
$$

Where the last equalities are due to (2.12, c, d).

To establish (ii), consider descent property (2.14) and (2.12, a)

$$0 \;>\; < d_k, \left(a \;+\; \mathcal{Q}\, x_k\right) > \;=\; - < d_k, \mathcal{H}_{k-1}\, \mu_k > . \tag{2.16}$$

From (2.12, b, d), we write $\mathcal{H}_{k-1}$ and $\mathcal{H}_k$ as follows

$$\mathcal{H}_{k-1} = [\, \mathfrak{H}^1 \quad \mathfrak{H}^2 \quad \ldots \quad \mathfrak{H}^{i} \quad \ldots \quad \mathfrak{H}^{i_{k-1}}\,]$$

$$\mathcal{H}_k = [\, \mathfrak{H}^1 \quad \mathfrak{H}^2 \quad \ldots \quad \mathfrak{H}^{i-1} \quad \mathfrak{H}^{i+1} \quad \ldots \quad \ldots \quad \mathfrak{H}^{i_k}\,].$$

Compared with $\mathcal{H}_k$, $\mathcal{H}_{k-1}$ only has one additional column, corresponding to the i th constraint, dropped in $\mathcal{H}_k$. Since $\mathcal{H}_k^T\, d_k = 0$, we have

$$\sum_{j \in \mathcal{I}_k} < d_k, \mathfrak{H}^j > \; \mu_k^j$$

$$= \; < d_k, [\, \mathfrak{H}^1 \quad \mathfrak{H}^2 \quad \ldots \quad \mathfrak{H}^{i-1} \quad \mathfrak{H}^{i+1} \quad \ldots \quad \ldots \quad \mathfrak{H}^{i_k}\,] \begin{bmatrix} \mu_k^1 \\ \mu_k^2 \\ \vdots \\ \mu_k^{i-1} \\ \mu_k^{i+1} \\ \vdots \\ \mu_k^{i_k} \end{bmatrix} > \; = \; < d_k, \mathcal{H}_k \begin{bmatrix} \mu_k^1 \\ \mu_k^2 \\ \vdots \\ \mu_k^{i-1} \\ \mu_k^{i+1} \\ \vdots \\ \mu_k^{i_k} \end{bmatrix} > \; = 0$$

Clearly, i $\notin \mathcal{I}_k$ and (2.16) becomes

$$0 > - < d_k, \mathcal{H}_{k-1}\, \mu_k > = -\sum_{j \in \mathcal{I}_k} < d_k, \mathfrak{H}^j > \mu_k^j - < \mathfrak{H}^i, d_k > \mu_k^i = - < \mathfrak{H}^i, d_k > \mu_k^i \tag{2.17}$$

and to ensure (2.17), with $\mu_k^i < 0$, we must have $< \mathfrak{H}^i, d_k > < 0$.

When μ_k does not have any negative elements, hence $\mu_k \geq 0$, we note from the optimality conditions of the **QP** problem (i.e. first order necessary conditions for a general nonlinear programming problem which become necessary and sufficient for this convex objective and convex constrained problem) that these are satisfied at x_k, μ_k. □

Corollary

Let $\mathcal{Q}$ positive semi-definite. Then,

(i) in general, the strict descent condition is replaced by $< d_k, \left(a + \mathcal{Q} x_k\right) > \leq 0$.

(ii) if $Z^T \mathcal{Q}\, Z$ is positive definite strict descent is ensured.

Proof

For (i), inequality (2.14) becomes

$$< d_k, \left(a + \mathcal{Q}\, x_k\right) > \leq < d_k, \left(a + \mathcal{Q}\, x_{k+1}\right) > = - < d_k, \mathcal{H}_k\, \mu_{k+1} > = 0.$$

For (ii), $d_k = Z\,\omega$ and $< d_k, \mathcal{Q}\, d_k > = <\omega, Z^T \mathcal{Q}\, Z\, \omega > > 0$. Thus, we have the strict inequality in (2.14)

$$< d_k, \left(a + \mathcal{Q}\, x_k\right) > < 0.$$ □

In practice, it is usual to choose the constraint to be dropped from $\mathcal{I}_k$ as the one corresponding to the negative multiplier with largest magnitude. This ensures the largest decrease in the objective function per unit step along d_k.

$\mathcal{I}_k$ is characterized by a set of constraints which are linearly independent from each other. Barring degeneracy, the algorithm proceeds from active set to active set, by either adding further constraints or dropping a constraint. As the objective function is reduced by every such move, return to the same active set is not possible. The number of constraints is finite and therefore so is the number of possible active sets. Thus, the algorithm converges in a finite number of steps.

3. ACTIVE SET ALGORITHM: RANGE SPACE OF CONSTRAINTS

If the number of linear constraints is large, then the null space of $\mathcal{H}_k^T$ is relatively small and thus characterizing the solution to the problem in terms of this null space is of special importance. This is discussed in Section 2. If the constraints are relatively few, then it is convenient to use the range space of $\mathcal{H}_k^T$ to construct the solution of the equality constrained problem. To do this, we write the optimality conditions (2.4) as

$$\mathcal{Q}\, d_k + \mathcal{H}_k\, \mu_{k+1} = - \left(a + \mathcal{Q}\, x_k\right) \tag{3.1}$$

$$\mathcal{H}_k^T\, d_k = 0. \tag{3.2}$$

If $\mathcal{Q}$ is positive definite, then d_k in (3.1) is given by

$$d_k \quad = \quad - \mathcal{Q}^{-1} \left[\mathcal{H}_k \, \mu_{k+1} \; + \; \left(a \; + \; \mathcal{Q} \; x_k \right) \right]. \tag{3.3, a}$$

Moreover, if the columns of $\mathcal{H}_k$ are linearly independent, using (3.2, b), μ_{k+1} is given by

$$\mu_{k+1} \quad = \quad - \left(\mathcal{H}_k^T \; \mathcal{Q}^{-1} \; \mathcal{H}_k \right)^{-1} \mathcal{H}_k^T \; \mathcal{Q}^{-1} \left(a \; + \; \mathcal{Q} \; x_k \right). \tag{3.3, b}$$

and hence (substituting (3.3, b) into (3.3, a))

$$d_k \quad = \quad \mathcal{Q}^{-1} \left[\mathcal{H}_k \left(\left(\mathcal{H}_k^T \; \mathcal{Q}^{-1} \; \mathcal{H}_k \right)^{-1} \mathcal{H}_k^T \; \mathcal{Q}^{-1} \left(a \; + \; \mathcal{Q} \; x_k \right) \right) - \left(a \; + \; \mathcal{Q} \; x_k \right) \right]$$

$$= \quad - \left[\mathcal{Q}^{-1} \; - \; \mathcal{Q}^{-1} \mathcal{H}_k \left(\mathcal{H}_k^T \; \mathcal{Q}^{-1} \; \mathcal{H}_k \right)^{-1} \mathcal{H}_k^T \; \mathcal{Q}^{-1} \right] \left(a \; + \; \mathcal{Q} \; x_k \right) \tag{3.3, c}$$

Expression (3.3) can also be obtained by considering (3.2) as a system

$$\begin{bmatrix} \mathcal{Q} & \mathcal{H}_k \\ \mathcal{H}_k^T & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \mu_{k+1} \end{bmatrix} \quad = \quad - \begin{bmatrix} \left(a + \mathcal{Q} \; x_k \right) \\ 0 \end{bmatrix}$$

with the inverse of the coefficient matrix on the left given by

$$\begin{bmatrix} \mathcal{Q}^{-1} - \mathcal{Q}^{-1} \mathcal{H}_k \left(\mathcal{H}_k^T \; \mathcal{Q}^{-1} \; \mathcal{H}_k \right)^{-1} \mathcal{H}_k^T \, \mathcal{Q}^{-1} & \mathcal{Q}^{-1} \mathcal{H}_k \left(\mathcal{H}_k^T \; \mathcal{Q}^{-1} \; \mathcal{H}_k \right)^{-1} \\ \left(\mathcal{H}_k^T \; \mathcal{Q}^{-1} \; \mathcal{H}_k \right)^{-1} \mathcal{H}_k^T \; \mathcal{Q}^{-1} & \left(\mathcal{H}_k^T \; \mathcal{Q}^{-1} \; \mathcal{H}_k \right)^{-1} \end{bmatrix} .$$

Example 4: The active set strategy (see Figure 1)

Consider the problem in Example 1. Since x_0 lies on a vertex (the intersection of constraints 3 and 5 is a vertex in this two dimensional problem), it is obviously a minimiser with respect to the constraints in $\mathcal{I}_0$. That is, the solution of (3.3, c) at k = 0 yields $d_0 = 0$.

$$\left[\mathcal{Q} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} ; \; a = \begin{bmatrix} -2 \\ -5 \end{bmatrix} ; \; \mathcal{H}_0 = \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix} ; \; h_0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} ; x_0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \right]$$

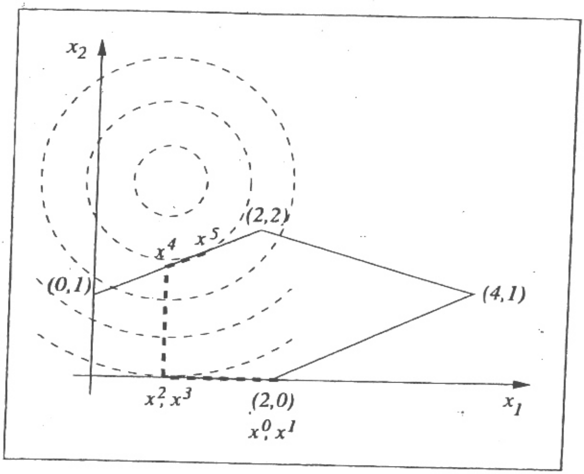
From (3.3,c), we have:

$$d_0 \quad = - \left[\mathcal{Q}^{-1} \; - \; \mathcal{Q}^{-1} \mathcal{H}_0 \left(\mathcal{H}_0^T \; \mathcal{Q}^{-1} \; \mathcal{H}_0 \right)^{-1} \mathcal{H}_0^T \; \mathcal{Q}^{-1} \right] \left(a \; + \; \mathcal{Q} \; x_0 \right)$$

$$= - \left[\mathcal{Q}^{-1} \; - \; \mathcal{Q}^{-1} \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}^T \mathcal{Q}^{-1} \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}^T \mathcal{Q}^{-1} \right] \left(a \; + \; \mathcal{Q} \; x_0 \right)$$

$$= - \left[\mathcal{Q}^{-1} \; - \; \mathcal{Q}^{-1} \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}^{-1} \mathcal{Q} \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}^{-T} \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}^T \mathcal{Q}^{-1} \right] \left(a \; + \; \mathcal{Q} \; x_0 \right)$$

$$= \; - [\; \mathcal{Q}^{-1} \; - \; \mathcal{Q}^{-1} \,] \left(a \; + \; \mathcal{Q} \; x_0 \right)$$



Iterates of the active-set method.

Example 1 - Section 2
Example 4 - Section 3

$= \ 0.$

We can then use (3.3, b) or (3.1) to obtain the corresponding multipliers μ^3 and μ^5 associated with the active constraints:

$$\mathcal{Q}\ d_0 \ + \ \mathcal{H}_0\ \mu_1 = -\left(a \ + \ \mathcal{Q}\ x_0\right)$$

as $d_0 = 0$, we have

$$\begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} \mu_1^3 \\ \mu_1^5 \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \end{bmatrix} \quad \Rightarrow \quad \mu^3 = -2, \ \mu^5 = -1 \ .$$

We remove constraint 3 from the active set $\mathcal{I}_0$ as it has the most negative multiplier (Hence, removing this will result in the largest reduction in the objective function per unit distance travelled in d_1 to be determined next. This is a greedy approach. We could have equally chosen constraint 5 to drop which would also have resulted in a reduction, but a smaller decrease per unit step travelled. As we cannot forecast the future path to be taken by either strategy, we simply choose the greedy approach which will do the best at the given stage.) The active set is thus $\mathcal{I}_1 = \{5\}$. We begin iteration 1 by computing (3.3, a or c)

$$d_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

The steplength formula leads to $\tau = 1$ and the new iterate becomes

$$x_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

There are no constrains blocking x_2 so that $\mathcal{I}_2 = \mathcal{I}_1 = \{5\}$. We find that at the start of iteration 2 that the solution (3.3, a or c) again yields $d_2 = 0$. From (3.3, b) or (3.1) we deduce that the multiplier of the active constraint 5 is $\mu_3^5 = -5$. So we drop this constraint to obtain $\mathcal{I}_3 = \emptyset$.

Iteration 3 starts by solving the unconstrained problem to yield

$$d_3 = -\mathcal{Q}^{-1}\left(a \ + \ \mathcal{Q}\ x_2\right) = -\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\left(\begin{bmatrix} -2 \\ -5 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 2.5 \end{bmatrix}.$$

The steplength formula yields $\tau_3 = 0.6$ (verify) and thus

$$x_4 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.6 \begin{bmatrix} 0 \\ 2.5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}.$$

There is an active constraint at x_4: $\mathcal{I}_4 = \{1\}$.

Iteration 4 starts by solving (3.3)

$$d_4 = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} \ ; \ \tau_4 = 1$$

Thus,, the new iterate is

$$x_5 = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix} + 1 \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 1.4 \\ 1.7 \end{bmatrix}.$$

and there are no new active constraints so $\mathcal{I}_5 = \mathcal{I}_4 = \{1\}$. Finally we solve (3.3) for k = 5 and obtain $d_5 = 0$. The multiplier $\mu_6^1 = 1.25$ so the solution is given by x_5 and the algorithm terminates. □

*****THE DISCUSSION UP TO (3.4 a, b) IS FOR THE INTERESTED STUDENT*****

In practice, the orthogonal factorization of $\mathcal{H}_k$ can be used to devise a version of the range space algorithm that involves orthogonal matrices, thereby injecting an element of numerical stability (Stewart, 1973; Golub and van Loan, 1983). Consider the QR factorization of $\mathcal{H}_k$ given by (2.2). Considering the submatrices of Q,

$$\mathcal{H}_k = [\, Y \,\vdots\, Z \,] \begin{bmatrix} R \\ 0 \end{bmatrix} = Y R$$

and substituting this in (3.3), we have

$$d_k = - \mathcal{Q}^{-1} \left[Y\theta + \left(a + \mathcal{Q} x_k \right) \right]. \qquad (3.4, a)$$

$$\theta = - \left(Y^T \mathcal{Q}^{-1} Y \right)^{-1} Y^T \mathcal{Q}^{-1} \left(a + \mathcal{Q} x_k \right). \qquad (3.4, b)$$

The multiplier μ_{k+1} can be recovered using the relation $R\,\mu_{k+1} = \theta$.

We now establish the equivalence of d_k given by (2.7) and (3.3, a) and μ_{k+1} given by (2.13) and (3.3, b).

Proposition 3.1

Let $\mathcal{Q}$ be positive definite and the columns of $\mathcal{H}_k$ be linearly independent. Then d_k given by (2.7) and (3.3, a) and μ_{k+1} given by (2.13) and (3.3, b) are equivalent.

Proof

Since d_k given by (3.3, a) satisfies $\mathcal{H}_k^T d_k = 0$, it follows that d_k can be expressed as a linear combination of the columns of Z, which form a basis of the null space of $\mathcal{H}_k^T$. Thus, $d_k = Z\,\omega$, for some ω. Then, using (3.3, a), we have

$$\mathcal{Q}\, d_k = - \left[\mathcal{H}_k \mu_{k+1} + \left(a + \mathcal{Q} x_k \right) \right]$$

$$\begin{aligned} Z^T \mathcal{Q}\, d_k &= - Z^T \left[\mathcal{H}_k \mu_{k+1} + \left(a + \mathcal{Q} x_k \right) \right] \\ &= - Z^T \left(a + \mathcal{Q} x_k \right) \\ Z^T \mathcal{Q}\, d_k &= Z^T \mathcal{Q} Z\, \omega \\ &= - Z^T \left(a + \mathcal{Q} x_k \right) \end{aligned}$$

$$\omega \quad = \quad -\left(Z^{T}\, \mathcal{Q}\, Z\right)^{-1} Z^{T}\left(a + \mathcal{Q}\, x_k\right)$$

and thus

$$\begin{aligned} d_k \quad &= \quad Z\,\omega \\ &= \quad -\, Z\left(Z^{T}\, \mathcal{Q}\, Z\right)^{-1} Z^{T}\left(a + \mathcal{Q}\, x_k\right) \end{aligned}$$

which establishes the equivalence with (2.7).

Having established the equivalence of d_k, we use d_k given by (3.3, a), which satisfies the optimality condition (3.1), with μ_{k+1} given by (3.3, b), to show that this multiplier is equivalent to the one given by (2.13). We start with μ_{k+1} given by (2.13) and use d_k given by (3.3, a) to establish equivalence with (2.13)

$$\begin{aligned} \mu_{k+1} \;&=\; -\left(\mathcal{H}_k^{T}\, \mathcal{H}_k\right)^{-1} \mathcal{H}_k^{T}\left[a + \mathcal{Q}\, x_k + \mathcal{Q}\, d_k\right] \\ &=\; \left(\mathcal{H}_k^{T}\, \mathcal{H}_k\right)^{-1} \mathcal{H}_k^{T}\left[\mathcal{H}_k\, \mu_{k+1}\right] \;=\; \mu_{k+1}. \end{aligned} \qquad \square$$

The basic algorithm, due to Fletcher (1971) and Goldfarb (1972), is the same as the null space method, discussed in Algorithm 2.1, with the exception that d_k, μ_{k+1} are evaluated using (3.3) or (3.4).

4. INTERIOR POINT ALGORITHM FOR QP

The active set strategies discussed in Sections 2 and 3 are essentially similar to the strategy employed by the simplex algorithm, developed by G. Dantzig in 1947, for linear programming (Dantzig, 1963). These algorithms proceed on the surface of the polygon described by the linear constraints, generating a sequence of active sets until the active set corresponding to the optimal solution is reached. The main difference between linear and **QP** problems is that the optimal solution of the latter may lie in the interior of the linear inequalities, in which case the solution is the unconstrained optimum of the quadratic function, or on one of the sides of the constraint polygon. By contrast, the solution of a linear programming problem is generally at a vertex. Although the simplex algorithm performs quite satisfactorily in solving linear programming problems, there is the theoretical danger that in searching for the optimum vertex, it may visit all, or almost all feasible vertices of the problem. Indeed, Klee and Minty (1972) construct examples in which this occurs. In such cases, the number of vertices visited, consequently the time taken, by the algorithm grows exponentially with the number of variables and constraints. Attempts to design a linear programming algorithm to solve the problem in polynomial time (i.e. the solution time for any linear programming problem bounded above by a polynomial function of the size of the problem) have resulted in Karmarkar's algorithm (1984). In this section we discuss a simple extension of Karmarkar's algorithm to **QP** problems.

Karmarkar's algorithm is shown by Gill et al (1986) to be an application of a penalty function method to linear (and quadratic) programming. The inequality constrained problem is reformulated as an optimization of a penalty function constrained by equality constraints only. The optimality conditions of the reformulation can be interpreted as the perturbed optimality conditions of the original problem. The Karmarkar algorithm can then be seen as the application of the Newton algorithm to the system of nonlinear equations formed by the perturbed optimality conditions. The perturbation is gradually allowed to disappear, yielding the solution of the original optimality conditions.

Let

$$q(x) = \langle a, x \rangle + \tfrac{1}{2} \langle x, \mathcal{Q}\, x \rangle \; ; \; \mathcal{G} \in \mathbb{R}^{n \times e} \; ; \; g \in \mathbb{R}^{e} \; ; \; e < n,$$

and consider the **QP**

$$\boxed{\min \left\{ q(x) \;\middle|\; \mathcal{G}^{T} x = g \; ; \; x \geq 0 \right\}}. \tag{4.1}$$

The original **QP** problem (1.2) can be reformulated as (4.1) by simply introducing nonnegative slack variables to transform the inequalities in (1.2) to equalities. For some $\eta \geq 0$, we define the penalty function

$$\ell(x, \eta) = q(x) - \eta \sum_{j=1}^{n} \ln x^{j} \tag{4.2}$$

and the equality constrained optimization problem

$$\boxed{\min \left\{ \langle a, x \rangle + \tfrac{1}{2} \langle x, \mathcal{Q}\, x \rangle - \eta \sum_{j=1}^{n} \ln x^{j} \;\middle|\; \mathcal{G}^{T} x = g \right\}}. \tag{4.3}$$

The function $\ell(x, \eta)$ is well defined for $x^{j} > 0$. It is referred to as a *barrier function* since it has a positive singularity at the constraint boundary (as x^{j} approaches the boundary of the feasible region). Many barrier functions have been proposed. We consider the logarithmic barrier function, first suggested by Frisch (1955). As the weight η assigned to the singularities approaches zero, the solution of (4.3) approaches to the solution of the original problem (4.1). Methods based on barrier functions are referred to as *interior point methods* since they require a strictly feasible initial point for each minimization, and generate a sequence of strictly feasible iterates. The strict feasibility requirement is relaxed later in this section. We introduce the properties of barrier method in this section. Detailed discussions of the general method are given by Fiacco and McCormick (1968) and Fiacco (1979).

Example 5: Upper and lower bounds

The barrier transformation of problem (4.1) and the associated solution algorithm can also be defined for **QP**s with upper and lower bounds on the variables, of the form:

$$\min \left\{ q(x) \;\middle|\; \mathcal{G}^{T} x = g \; ; \; l \leq x \leq u \right\}.$$

The problem analogous to (4.3) is

$$\boxed{\min \left\{ \langle a, x \rangle + \tfrac{1}{2} \langle x, \mathcal{Q} x \rangle - \eta \sum_{j=1}^{n} \ln\left(x^{j} - l^{j}\right) - \eta \sum_{j=1}^{n} \ln\left(u^{j} - x^{j}\right) \;\middle|\; \mathcal{G}^{T} x = g \right\}}$$

(e.g. Gill et al, 1986; Lustig, Marsten and Shanno, 1989; 1992). □

To ensure the existence of a global constrained minimum for (4.2), we assume that $\mathcal{Q}$ is positive semi-definite on the null space of $\mathcal{G}^{T}$. In this case, (4.3) is a convex problem and, by invoking the basic convexity result in the beginning of the course, the first order conditions become both necessary and sufficient for optimality. Convexity also ensures that we are concerned with global optima of ℓ.

The main question and the underlying motivation for interior point methods is the convergence to the solution of the original problem as the penalty parameter approaches zero. We use the SUMT framework introduced earlier as motivation. Let $\{\eta_k\}$ be a non-negative sequence, strictly decreasing to zero. Define $x_k = x(\eta_k)$ such that

$$\ell(x_k, \eta_k) = \min \left\{ \ell(x, \eta_k) \;\middle|\; \mathcal{G}^T x = g \right\}. \tag{4.4}$$

The existence of x_k is guaranteed by the continuity of the quadratic objective and the barrier function for $x > 0$. Also, each $x_k > 0$ and satisfies the equality constraints in (4.4). Let x_* denote a limit point of the uniformly bounded sequence $\{x_k\}$. Clearly, we have

$$x_* \in \left\{ x \;\middle|\; \mathcal{G}^T x = g;\; x \geq 0 \right\}.$$

Let

$$q_* = \min \left\{ q(x) \;\middle|\; \mathcal{G}^T x = g \;;\; x \geq 0 \right\}.$$

and suppose that x_* is *not* the solution of (4.1). Then by the basic property of q_*, $q(x_*) > q_*$. There is a x_0 such that

$$x_0 \in \left\{ x \;\middle|\; \mathcal{G}^T x = g;\; x > 0 \right\}$$

where

$$q_* < q(x_0) < q(x_*).$$

Then

$$\lim \inf_{k\to\infty} \ell(x_k, \eta_k) \geq q(x_*) > q(x_0) = \lim_{k\to\infty} \ell(x_0, \eta_k).$$

This contradicts the assumption that x_k minimizes ℓ, subject to the equality constraints, for large k. **Hence, $\lim_{k\to\infty} q(x_k) = q_*$ and x_* solves (4.1)**.

Example 6:

Consider the problem

$$\min \left\{ (x^1 - 1)^2 + (x^2 - 1)^2 \;\middle|\; x^1 + x^2 \leq 1 \right\}$$

and the barrier function formulation

$$\min \left\{ (x^1 - 1)^2 + (x^2 - 1)^2 - \eta_k \ln(-x^1 - x^2 + 1) \right\}.$$

The first order conditions for an unconstrained minimum of the barrier function yield

$$(x^1 - 1) + \frac{\eta_k}{-x^1 - x^2 + 1} = 0 \;; \quad (x^2 - 1) + \frac{\eta_k}{-x^1 - x^2 + 1} = 0.$$

Hence, we have

$$\boxed{x^1(\eta_k) = x^2(\eta_k) = \frac{3 \pm \sqrt{9 - 8(1 - \eta_k)}}{4}.}$$

The negative term corresponds to the minimum. As $\eta_k \to 0$, the constrained minimum $x^1 = x^2 = \frac{1}{2}$ is achieved. □

It can also be shown that $\{q(x_k)\}$ is a monotonically *decreasing* and $\left\{-\eta\sum_{j=1}^{n}\ln(x^j)\right\}$ is a monotonically *increasing* sequence. Let $q(x_k) = q(x(\eta_k)) = q_k$. Because each $x_k = x(\eta_k)$ is a global minimum of ℓ,

$$q_k - \eta_k \sum_{j=1}^{n} \ln x_k^j \leq q_{k+1} - \eta_k \sum_{j=1}^{n} \ln x_{k+1}^j \tag{4.5}$$

$$q_{k+1} - \eta_{k+1} \sum_{j=1}^{n} \ln x_{k+1}^j \leq q_k - \eta_{k+1} \sum_{j=1}^{n} \ln x_k^j .$$

Multiplying (4.5) by η_{k+1}/η_k, adding to the second inequality yields

$$\left(1 - \frac{\eta_{k+1}}{\eta_k}\right) q_{k+1} \leq \left(1 - \frac{\eta_{k+1}}{\eta_k}\right) q_k .$$

Since $\eta_{k+1} < \eta_k$, the monotonic *decrease* of $\{q(x_k)\}$ follows. Using this result and (4.5) we have

$$- \eta_k \sum_{j=1}^{n} \ln x_k^j \leq q_k - q_{k+1} - \eta_k \sum_{j=1}^{n} \ln x_k^j \leq - \eta_k \sum_{j=1}^{n} \ln x_{k+1}^j$$

and hence the monotonic *increase* of $\left\{-\eta\sum_{j=1}^{n}\ln x^j\right\}$.

Interior point methods for **QP** depart slightly from the above framework by applying the Newton algorithm for solving the optimality condition of (4.3) while allowing η_k to decrease. The Newton algorithm is a general method for solving systems of nonlinear equations in many variables. We consider the application of the algorithm to the particular problem of solving the optimality conditions of (4.3).

The interior point algorithm is based on the solution of the optimality conditions of (4.3). Consider the Lagrangian

$$< a, x > + \tfrac{1}{2} < x, \mathcal{Q} x > - \eta_k \sum_{j=1}^{n} \ln x^j + < \mathcal{G}^T x - g, \lambda >$$

and the first order optimality conditions

$$a + \mathcal{Q} x - \eta_k X^{-1}e + \mathcal{G} \lambda = 0$$

$$\mathcal{G}^T x - g = 0$$

where λ are the dual variables, $X = \begin{bmatrix} x^1 & & & & & \\ & x^2 & & & & \\ & & \ddots & & & \\ & & & x^i & & \\ & & & & \ddots & \\ & & & & & x^n \end{bmatrix}$ and $e \in \mathbb{R}^n$ is the vector with all unit elements. It is clear that the term X^{-1} will lead to problems as any search for the solution approaches the boundary of $x \geq 0$. To overcome this difficulty, we introduce the transformation $y = \eta X^{-1}e$. The transformation can be equivalently represented as

$$X Y e = \eta_k e$$

where,

$$Y = \begin{bmatrix} y^1 & & & & & \\ & y^2 & & & & \\ & & \ddots & & & \\ & & & y^i & & \\ & & & & \ddots & \\ & & & & & y^n \end{bmatrix}$$

Hence, the first order conditions of (4.3) are given by the nonlinear system

$$\boxed{\begin{bmatrix} \mathcal{G}^T x - g \\ \mathcal{G}\,\lambda + a + \mathcal{Q}\,x - y \\ X\,Y\,e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \eta_k e \end{bmatrix}} \qquad (4.6)$$

With x $\geq$ 0, y $\geq$ 0 and η_k = 0, (4.6) also coincides with the optimality condition for the original QP (4.1) where the vector y corresponds to the multipliers of the constraints x $\geq$ 0.

We can eliminate the dual variables λ from (4.6) by premultiplying the second equation by the *nonsingular* matrix $[\,\mathcal{G} \vdots Z\,]^T$. As in Section 2.1, we let the columns of Z $\in \mathbb{R}^{n\times(n-e)}$ form a basis for the null space of $\mathcal{G}^T$. Noting that $Z^T\,\mathcal{G} = 0$, we obtain

$$0 = \begin{bmatrix} \mathcal{G}^T \\ Z^T \end{bmatrix} [\,\mathcal{G}\,\lambda + a + \mathcal{Q}\,x - y\,] = \begin{bmatrix} \mathcal{G}^T\mathcal{G}\,\lambda + \mathcal{G}^T(a + \mathcal{Q}\,x - y) \\ Z^T\,(a + \mathcal{Q}\,x - y) \end{bmatrix}.$$

Since $\mathcal{G}^T\mathcal{G}$ is nonsingular, once x and y are known, λ is uniquely determined by

$$\boxed{\lambda = -(\mathcal{G}^T\mathcal{G})^{-1}[\,\mathcal{G}^T(a + \mathcal{Q}\,x - y)\,]}.$$

****** YOU ARE NOT RESPONSIBLE FOR (4.7) - (4.8) BELOW ******

Removing the equation for λ, we have the following 2 n-dimensional nonlinear system with nonnegativity constraints for x, y

$$\begin{bmatrix} \mathcal{G}^T x - g \\ Z^T\,(a + \mathcal{Q}\,x - y) \\ X\,Ye \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \eta_k e \end{bmatrix}, \quad x \geq 0, y \geq 0. \qquad (4.7)$$

We note that

$$\boxed{F(x, y) \equiv \begin{bmatrix} \mathcal{G}^T x - g \\ Z^T\,(a + \mathcal{Q}\,x - y) \\ X\,Ye \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad x \geq 0, \ \ y \geq 0,} \qquad (4.8)$$

denotes the first order optimality conditions of the original **QP** and, as $\{\eta_k\} \rightarrow 0$, the solution of (4.7) approaches to that of (4.8).

Algorithm 4.1 is essentially aimed at solving (4.8) by successive steps taken using (4.7) while η_k is allowed to approach zero. It can be stated in compact form as follows:

Algorithm 4.1

Step 0: Given a pair $(x_0, y_0) > 0$, set k =0.

Step 1: Choose $\sigma_k \in [0, 1)$, $\tau_k \in (0, 1)$ and set $\eta_k = \frac{\sigma_k <x_k, y_k>}{n}$.

Step 2: Solve for d_k^x, d_k^y the system

$$\nabla F(x_k, y_k) \begin{bmatrix} d^x \\ d^y \end{bmatrix} = - F(x_k, y_k) + \begin{bmatrix} 0 \\ \eta_k e \end{bmatrix}. \tag{4.9}$$

Step 3: Compute the step length:

$$\alpha_k = \frac{-\tau_k}{\min\left\{ X_k^{-1} d^x,\ Y_k^{-1} d^y, -\tau_k \right\}}.$$

Step 4: Set

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} + \alpha_k \begin{bmatrix} d_k^x \\ d_k^y \end{bmatrix},$$

k = k + 1 and go to Step 1.

****** COMPLEXITY DISCUSSION (UP TO END OF SECTION 4) ******
FOR INTERESTED STUDENTS ONLY

The above algorithm is due to Zhang, Tapia and Potra (1993). Among many variants of interior point methods, it is one of the simplest to describe. For a survey of this area, we refer the reader to Ye (1990; 1991). Algorithm 4.1 depends on the specification of the sequences $\{\tau_k\}$ and $\{\sigma_k\}$. It is shown in Zhang, Tapia and Potra (1993; Theorem 3.1) that provided $\{\tau_k\} \to 1$ and $\{\sigma_k\} \to 0$, the sequence $\{F(x_k, y_k)\}$ is convergent at a Q-superlinear rate, i.e.

$$\boxed{\lim_{k\to\infty} \sup \frac{\|F(x_{k+1}, y_{k+1})\|}{\|F(x_k, y_k)\|} = 0; \quad \text{and that} \quad \lim_{k\to\infty} \alpha_k = 1.}$$

Without the perturbation term η_k e on the right-hand side of (4.9), the search direction (d^x, d^y) is the Newton step for solving F(x, y) = 0. The starting point (x_0, y_0) is not required to be feasible. Also, $\alpha_k \in (0, 1]$ and $\alpha_k = 1$ if and only if $\min\left\{ X_k^{-1} d^x,\ Y_k^{-1} d^y \right\} \geq -\tau_k$. In most cases, it is expected that $\alpha_k < 1$. The choice $\tau_k = 1$ corresponds to allowing steps to the boundary of the bounds $x \geq 0$ and thus to a loss of strict feasibility. Using a similar argument as in (2.11), we see that the maximum step Algorithm 4.1 may take, from a feasible point (x_k, y_k), to reach the boundary of the inequalities $x \geq 0$, $y \geq 0$, is given by

$$\boxed{\frac{1}{\min\left\{ X_k^{-1} d^x,\ Y_k^{-1} d^y \right\}}.}$$

Thus, if (x_k, y_k) is strictly feasible, so is (x_{k+1}, y_{k+1}).

The above algorithm is closely related to a number of interior point algorithms, including Kojima, Meggido and Noma (1991); Kojima, Mizuno and Noma (1989); Kojima, Mizuno and Yoishe (1989); Monteiro and Adler (1989, a, b). The strategies defining the stepsize and the penalty parameter are particular characteristics that very across algorithms. For example, Monteiro and Adler (1989, b) use an alternative stepsize strategy and the penalty parameter defined by $\eta_{k+1} = \eta_k \left(1 - \frac{\delta}{\sqrt{n}}\right)$ for some δ (set at the value $\delta \equiv 0.1$). In other studies, including the special case of linear programming (i.e. $\mathcal{Q} \equiv 0$), varying stepsize strategies have been used to define x_{k+1}, y_{k+1}, with η depending on $< x_{k+1}, y_{k+1} > /n$ (e.g. Ye and Anstreicher, 1993; Ye, Güler, Tapia, Zhang, 1993; see also Lustig, Marsten Shanno, 1989; 1992).

The exciting aspect of interior point algorithms is their polynomial complexity: that they terminate in a number of iterations that can be bounded by a polynomial depending on the data used to describe the problem. To give an example for the quantities involved, we assume that the entries of the vectors a, c and the matrices $\mathcal{Q}$, $\mathcal{G}$ are integral and define the size of the **QP** problem (4.1) as

$$L(\mathcal{Q}, \mathcal{G}, \mathrm{a}, \mathrm{g}) = \left\lceil \log \left[\begin{array}{c} \text{largest absolute value of determinant of any} \\ \text{square submatrix of} \begin{bmatrix} \mathcal{Q} & \mathcal{G} \\ \mathcal{G}^{\mathrm{T}} & 0 \end{bmatrix} \end{array} + 1 \right] \right\rceil$$

$$+ \left\lceil \log \left(1 + \max_{i} \; | a^i | \right) \right\rceil + \left\lceil \log \left(1 + \max_{j} \; | h^j | \right) \right\rceil + \left\lceil \log \left(n + e \right) \right\rceil .$$

By starting from an appropriate initial position satisfying $< x_0, y_0 > \;\leq 2^{\nu L}$, where ν is a positive constant, independent of n, Zhang and Tapia (1993; Theorem 6.1, Corollary 6.2) show that they ensure sufficient finite progress at each iteration to ensure convergence in at most O ($n\ L$) iterations for an interior point linear programming algorithm. Monteiro and Adler (1989, b) show that if $\eta_0 = 2^{O(L)}$, then their **QP** algorithm solves the problem in at most $O(\sqrt{n}\ L)$ iterations. Actually, these results are based on interior point algorithms solving the system F(x, y) = 0 such that $< \hat{x}, \hat{y} > \;\leq 2^{-O(L)}$. At such a point, it can be shown that the exact solution of the system would require O (n^3) iterations, using conventional solvers (see Papadimitriou and Steiglitz, 1982; Lemma 8.7).

Other related interior point algorithms include Kojima, Mizuno, Yoshise (1991) who consider the reduction of the potential function

$$f(x, y) = \eta \ln < x, y > \; - \sum_{j=1}^{n} \ln x^j y^j$$

where $\eta = n + \sqrt{n}$ (see also Kojima, Megiddo, Noma, Yoshise, 1991). Starting from a feasible point, the algorithm generates a sequence $\{x_k, y_k\}$ that converges in $O(\sqrt{n}\ L)$ steps using a stepsize τ_k that ensures the feasibility of

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} + \tau \begin{bmatrix} d_k^x \\ d_k^y \end{bmatrix}$$

and

$$f(x_k + \tau\, d_k^x,\ \ y + d_k^y) - f(x_k,\ y_k) \leq \beta$$

for $\beta > 0$ (chosen as $\beta = 0.2$).

Another algorithm, similar in spirit to that of Kojima, Mizuno, Yoshise (1991), is due to Goldfarb and Liu (1993) who consider the minimisation of the potential function

$$\Phi(x, y) = \eta \ln \langle x, y \rangle - \sum_{j=1}^{n} \ln x^j - \sum_{j=1}^{n} \ln y^j$$

where $\eta = n + \sigma \sqrt{n}$ with $\sigma \geq 1$. The method generates a sequence $\{x_k, y_k, \lambda_k\}$ that converges in $O(n^3 L)$ steps using a stepsize τ_k determined (as also in Anstreicher and Bosch, 1992) using an Armijo (1966) type strategy that satisfies

$$\Phi(x_k + \tau d_k^x, y + d_k^y) - \Phi(x_k, y_k) \leq \rho\, \tau \left(\langle \nabla_x \Phi(x_k, y_k), d_k^x \rangle + \langle \nabla_y \Phi(x_k, y_k), d_k^y \rangle \right)$$

for $\rho \in (0, 1)$. Furthermore, a potential function reduction requirement is satisfied by τ_k

$$\Phi(x_k + \tau d_k^x, y + d_k^y) - \Phi(x_k, y_k) \leq -\beta$$

for $\beta > 0$.

5. MEAN-VARIANCE PORTFOLIO OPTIMIZATION

Decision making under uncertainty may be formulated as the simultaneous optimization of the expected (mean value) of the total return and the associated risk, measured in terms of the variance of the return. There is clearly a trade-off in emphasising the maximization of the return and the minimization of the risk. The more risk-seeking the decision maker, the greater will the emphasis be on the expected return. Similarly, as the decision maker considers more risk-averse, or cautious, policies, the importance of minimizing the variance of the returns will increase.

We introduce the mean-variance approach. This is the classical quadratic mean-variance portfolio optimization problem (Markowitz, 1959; Sharpe, 1970; Elton and Gruber, 1991).

Mean-Variance Optimization within a Single Currency

Consider the return vector $r \in \mathbb{R}^n$ on n investments being considered

$$r = \mathcal{E}[r] + \rho$$

where $\mathcal{E}$ denotes expectation, $\rho \sim \mathcal{N}(0, \Lambda_{\rho\rho})$, $\Lambda_{\rho\rho} = \mathcal{E}[(\rho)(\rho)^T]$. There are many ways of estimating $r = \mathcal{E}[r] + \rho$ and we refer the interested reader to the literature on finance for further information on this subject (e.g. Elton and Gruber (1991). To introduce the portfolio problem in a single currency, let $\omega \in \mathbb{R}^n$ denote the weights, in terms of the proportion of the total budget, to be attached to each investment. The expected return on the portfolio is then given by

$$\mathcal{E}[\langle \omega, r \rangle] = \mathcal{E}[\langle \omega, \mathcal{E}[r] + \rho \rangle] = \langle \mathcal{E}[r], \omega \rangle .$$

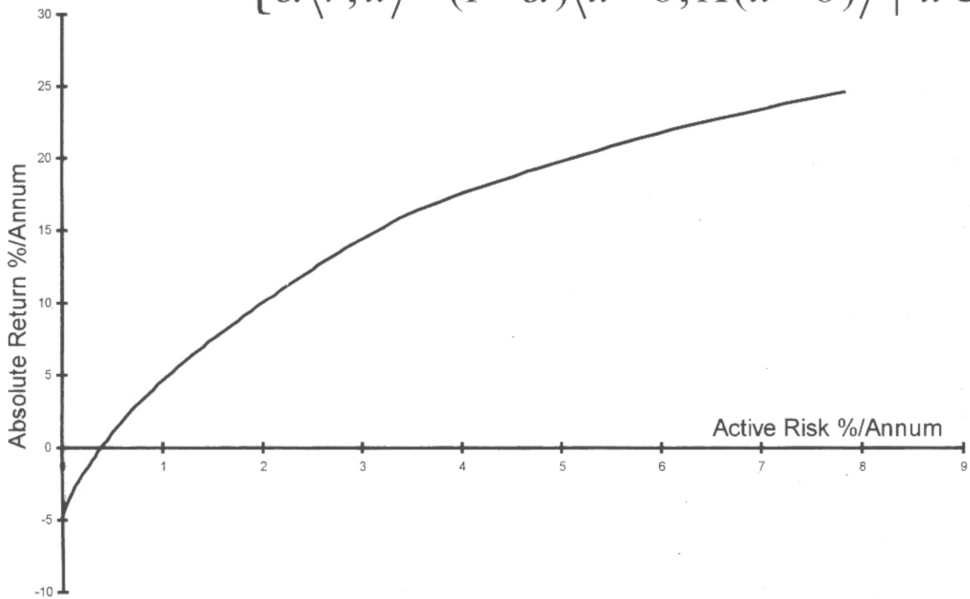
Risk is measured in terms of the statistical variance of the portfolio return, given by

$$\begin{aligned} \text{var} \langle \omega, r \rangle &= \mathcal{E}[\langle \omega, r \rangle - \mathcal{E}[\langle \omega, r \rangle]]^2 \\ &= \mathcal{E}[\langle \omega, r \rangle]^2 \\ &\quad - 2\,\mathcal{E}[\langle \omega, r \rangle \mathcal{E}[\langle \omega, r \rangle]] + [\mathcal{E}[\langle \omega, r \rangle]]^2 \end{aligned}$$

risk-return frontier: markowitz

$$\text{maximise} \quad \alpha \boxed{\begin{matrix}\text{expected}\\ \text{return}\end{matrix}} - (1-\alpha) \boxed{\begin{matrix}\text{expected}\\ \text{risk}\end{matrix}}$$

$$\text{maximise} \left\{ \alpha \langle r, x \rangle - (1-\alpha) \langle x - b, \Lambda (x - b) \rangle \;\middle|\; x \in \Omega \right\}$$



$$= \mathcal{E}\,[<\ \omega, \mathrm{r}\ >]^2 - [\mathcal{E}\,[<\ \omega, \mathrm{r}\ >]]^2$$

$$= \mathcal{E}\,[<\omega, [\mathcal{E}\,[\,\mathrm{r}\,] + \rho]>\ <[\mathcal{E}\,[\,\mathrm{r}\,] + \rho], \omega>] \\ \qquad - [<\ \omega, \mathcal{E}\,[\,\mathrm{r}\,]>]^2$$

$$= [<\ \omega, \mathcal{E}\,[\,\mathrm{r}\,]\ >]^2 + 2\,\mathcal{E}\,[<\omega, \mathcal{E}\,[\,\mathrm{r}\,]>\ <\rho, \omega>] \\ \qquad + \mathcal{E}\,[\ <\omega, \rho>\ <\rho, \omega>] - [<\ \omega, \mathcal{E}\,[\,\mathrm{r}\,]\ >]^2$$

$$= <\omega, \mathcal{E}\,[\rho\,\rho^{\mathrm{T}}]\,\omega> .$$

The decision maker's attitude to risk determines the importance attached to the maximization of the portfolio return vs the minimization of risk, measured by the above variance expression. Let $\alpha \in [0, 1]$ be a fixed scalar and let $<\mathrm{u}, \mathrm{v}> \;\equiv\; \mathrm{u}^{\mathrm{T}}\mathrm{v}$. An optimal mean-variance portfolio is given by a combination of the mean of the portfolio return and its variance (see Figure 2)

$$\max \left\{ \alpha < \mathcal{E}\,[\,\mathrm{r}\,], \omega\ > \ - \ (1 - \alpha)\ <\omega, \Lambda_{\rho\rho}\,\omega\ > \ \middle|\ \ <\boldsymbol{1}, \omega\ > \ = 1, \omega \ \geq \ 0 \right\}$$

where $\boldsymbol{1}$ is the vector with all unit elements. Here, the portfolio return is being maximized simultaneously with the minimization of the risk associated with it. The discussion below does not depend on the nature of the constraints on ω. We have assumed that the budget is restricted and the weights attached to each investment, as a proportion of the budget, are nonnegative. These constraints can be altered, as required, to allow for various trading strategies.

Example 7: An investor has £1000 to invest in three stocks. Let S^i be the (random) variable representing the annual return on £1 invested in stock i (ie if $S^i = 0.12$, £1 invested in stock i at the beginning of the year is worth £1.12 at the end of the year). You are given the following information: $E(S^1) = 0.14$, $E(S^2) = 0.11$, $E(S^3) = 0.10$, $\mathrm{var}\, S^1 = 0.20$, $\mathrm{var}\, S^2 = 0.08$, $\mathrm{var}\, S^3 = 0.18$, $\mathrm{cov}(S^1, S^2) = 0.05$, $\mathrm{cov}(S^1, S^3) = 0.02$, $\mathrm{cov}(S^2, S^3) = 0.03$. Formulate a QP that can, be used to find the minimum risk (variance) portfolio that attains an expected return of at least 12%.

x^j : £'s invested in stock j (j = 1, 2, 3). The annual return on the portfolio is

$$\frac{(x^1S^1 + x^2S^2 + x^3S^3)}{1000}$$

and the expected annual return on the portfolio is

$$\frac{(x^1E(S^1) + x^2E(S^2) + x^3E(S^3))}{1000}$$

To ensure the portfolio has an expected return of at least 12%, we must include the constraint

$$\boxed{\frac{(0.14\,x^1 + 0.11\,x^2 + 0.10\,x^3)}{1000} \ \geq \ 0.12}$$

or

$$\boxed{0.14\,x^1 + 0.11\,x^2 + 0.10\,x^3 \ \geq \ 120.}$$

Of course the wealth constraint needs to be included: $x^1 + x^2 + x^3 = 1000$ and the amount invested in each stock must be nonnegative $x^1, x^2, x^3 \geq 0$. The **objective is to minimise the variance (risk) of the portfolio** and is given by

$$\text{var}\,(x^1S^1 + x^2S^2 + x^3S^3) = \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix}^T \begin{bmatrix} \text{var } S^1 & \text{cov } (S^1S^2) & \text{cov } (S^1S^2) \\ \text{cov } (S^2S^1) & \text{var } S^2 & \text{cov } (S^2S^3) \\ \text{cov } (S^3S^1) & \text{cov } (S^3S^2) & \text{var } S^3 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix}$$

Hence, we have the QP problem:

$$\min \left\{ \text{var}\,(x^1S^1 + x^2S^2 + x^3S^3) \;\middle|\; 0.14\, x^1 + 0.11\, x^2 + 0.10\, x^3 \;\geq\; 120; \right.$$

$$\left. x^1 + x^2 + x^3 = 1000;\; x^1, x^2, x^3 \;\geq\; 0 \right\}.$$

Remark: If, instead of searching for the portfolio ensuring an expected return of 12% (represented by constraint $0.14\, x^1 + 0.11\, x^2 + 0.10\, x^3 \;\geq\; 120$) the requirement is to consider the family of risk averse portfolios (maximising return and minimising risk), then the QP problem would be

$$\min \left\{ \alpha\, \text{var}\,(x^1S^1 + x^2S^2 + x^3S^3) \;-\; (1-\alpha) \left(0.14\, x^1 + 0.11\, x^2 + 0.10\, x^3 \right) \;\middle|\; x^1 + x^2 + x^3 = 1000;\; x^1, x^2, x^3 \;\geq\; 0 \right\}.$$

This problem would be solved for given values of $\alpha \in [0, 1]$. The results would be shown to the decision maker for her/him to choose the value of α corresponding to the portfolio representing her/his attitude to risk. □

Exercise: In the preceding formulation, if the decision maker's attitude to risk was $\alpha = 0$, what would be the optimal portfolio?

****** REVISION EXAMPLES - THE PART RELATING TO ******
THE DERIVATION OF Z IS FOR THE INTERESTED STUDENT ONLY

Example 8: a null-space algorithm application

Consider the active set (2.1, a) and the corresponding system of equalities (2.1, b) with $Z \in \mathbb{R}^n \times \mathbb{R}^{n-i_k}$; $\mathcal{H}_k^T Z = 0$. There are two ways of representing Z. The first involves (2.1, z)-(2.4, z) such that (2.5, z) is satisfied. Thus, d_k is given by (2.7), τ_k by (2.11) and μ_{k+1} by (2.13). We apply this choice of Z to the following problem

$$\text{minimise } (x^1)^2 + (x^2)^2 + (x^3)^2$$
$$\text{subject to}$$
$$x^1 + 2\, x^2 - x^3 = 4; \qquad x^1 - x^2 + x^3 = -2$$

evaluating

$$x_B = \mathcal{H}_B^{-T} \left(h_k - \mathcal{H}_N^T\; x_N \right):$$

$$x_B = \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}^{-1} \left[\begin{bmatrix} 4 \\ -2 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} x^3 \right] = \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1/3 \\ -2/3 \end{bmatrix} x^3$$

substituting in the original quadratic: $14/9\,(x^3)^2 + 8/3\,x^3 + 4$ which is minimised at

$$x^3 = -6/7 \quad \text{and} \quad x_B = \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 10/7 \end{bmatrix}$$

the multipliers satisfy

$$-\,2/7 \begin{bmatrix} 2 \\ 10 \\ -6 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \mu^1 \\ \mu^2 \end{bmatrix}$$

and the first two rows give $\begin{bmatrix} \mu^1 \\ \mu^2 \end{bmatrix} = \begin{bmatrix} 8/7 \\ -4/7 \end{bmatrix}$ which is consistent with the third row. Now you know the result, start the algorithm at a feasible point by assigning any value to x^3 and follow the algorithm using the above Z. You may use MATLAB or any other system to do your matrix inversion etc. □

Example 9: Range Space calculations

Apply the relationships given by (3.3) to the problem below.

$$\text{minimise } (x^1 - 1)^2 + (x^2 - 3)^2$$

subject to

Constraint 1: $x^1 + x^2 \leq 1$;

Constraint 2: $-\, x^1 \leq 0$;

Constraint 3: $-\, x^2 \leq 0$

Let the set initial feasible point be (0, 0). $\mathcal{I}_0 = \{2, 3\}$ the solution in the intersection of these two (nonnegativity) constraints remains (0, 0) check the multipliers

$$\begin{bmatrix} \mu^1 \\ \mu^2 \end{bmatrix}_k = -\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ -6 \end{bmatrix} = \begin{bmatrix} -2 \\ -6 \end{bmatrix}$$

we could drop either constraint: drop $-\, x^2 \leq 0$,

$$d_k = -\left[\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} - \mathcal{Q}^{-1} \mathcal{H}_k \left(\mathcal{H}_k^T\ \mathcal{Q}^{-1}\ \mathcal{H}_k \right)^{-1} \mathcal{H}_k^T\ \mathcal{Q}^{-1} \right] \begin{bmatrix} -2 \\ -6 \end{bmatrix}$$

$$d_k = -\tfrac{1}{2} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \end{bmatrix} [1] \begin{bmatrix} -1 & 0 \end{bmatrix} \right) \begin{bmatrix} -2 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$\text{confirm feasibility:} \quad [-1 \,\vdots\, 0]\, d_k = [-1 \,\vdots\, 0] \begin{bmatrix} 0 \\ 3 \end{bmatrix} = 0$$

however, proceeding along d_k will transgress $x^1 + x^2 \leq 1$

$\tau = 1/3$:

$$\begin{bmatrix} x^1 \\ x^2 \end{bmatrix}_{k+1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 1/3 \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

add $x^1 + x^2 \leq 1$ to $-x^1 \leq 0$ to form new active set. We know the solution is (0, 1) for both constraints. Can we improve by dropping one ($-x^1 \leq 0$) of the constraints?

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

All multipliers are positive so, we terminate at the **optimum**. Suppose, alternatively, at (0, 0) we drop $-x^1 \leq 0$ (instead of $-x^2 \leq 0$). Let

$$\boxed{d_k = \hat{x}_{k+1} - x_k}$$

$$\hat{x}_{k+1} = \begin{bmatrix} \hat{x}^1 \\ x^2 \end{bmatrix}_{k+1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \tfrac{1}{2} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 \\ -1 \end{bmatrix} [1] \begin{bmatrix} 0 & -1 \end{bmatrix} \right) \begin{bmatrix} -2 \\ -6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$[\, 0 \,\vdots\, -1 \,] \begin{bmatrix} \hat{x}^1 \\ x^2 \end{bmatrix}_{k+1} = 0$$

$x^1 + x^2 \leq 1$ is now active and is added to $-x^2 \leq 0$. As we have two constraints in two unknowns, we know the solution is (1, 0) in this intersection. Check the multipliers

$$\mathcal{H} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} \mu^1 \\ \mu^2 \end{bmatrix} = -\begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ -6 \end{bmatrix}.$$

We can drop the nonnegativity constraint and project again with $\mathcal{H} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $h = 1$

$$\begin{bmatrix} \hat{x}^1 \\ x^2 \end{bmatrix}_{k+1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \tfrac{1}{2} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} [2]^{-1} \begin{bmatrix} 1 & 1 \end{bmatrix} \right) \begin{bmatrix} 0 \\ -6 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 3/2 \end{bmatrix}$$

$$\mathcal{H}^T \begin{bmatrix} x^1 \\ x^2 \end{bmatrix}_{k+1} = [\, 1 \,\vdots\, 1 \,] \begin{bmatrix} x^1 \\ x^2 \end{bmatrix}_{k+1} = 1$$

which, of course, transgresses $-x^1 \leq 0$ which is added to the active set

$$\tau_k = \frac{0 - \begin{bmatrix} -1 \\ 0 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} -1 \\ 0 \end{bmatrix}^T \left[\begin{bmatrix} -1/2 \\ 3/2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right]} = 2/3$$

so

$$\begin{bmatrix} x^1 \\ x^2 \end{bmatrix}_{k+1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 2/3 \left[\begin{bmatrix} -1/2 \\ 3/2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right] = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

where the active set is $-x^1 \le 0$; $x^1 + x^2 \le 1$. This is the set that corresponds to the optimum that we computed earlier. □

Exercise: Consider the QP

$$\min \left\{ x^1 + 2\,(x^2)^2 \;\middle|\; x^1 + x^2 \le 2;\;\; 2\,x^1 + x^2 \le 3;\;\; x^1, x^2 \ge 0 \right\}$$

Write the KKT conditions for this QP. Solve the QP.

REFERENCES

K.M. Anstreicher and R.A. Bosh (1992). “Long Steps in a O ($n^3\, L$) Algorithm for Linear Programming", *Mathematical Programming,* 54, 251-265.

Armijo, L. (1966). “Minimization of Functions Having Lipschitz-continuous First Partial Derivatives", *Pacific J. Math.*, 16, 1-3.

Dantzig, G. (1963). *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey.

Den Hertog, D. (1994). *Interior Point Approach to Linear, Quadratic and Convex Programming*, Kluwer Publishers, Dordrecht.

El-Bakry, A.S., R.A. Tapia, T. Tsuchiya and Y. Zhang (1996). “On the Formulation of the Newton Interior-Point Method for Nonlinear Programming", *JOTA*, 89, 507-541.

El-Bakry, A.S., R.A. Tapia and Y. Zhang (1994). “A Study of Indicators for Identifying Zero Variables in Interior Point Methods", *SIAM Review*, 36, 45-72.

Elton, E.J. and M.J. Gruber (1991). *Modern Portfolio Theory and Investment Analysis,* John Wiley, New York.

Fiacco, A.V. (1979). “Barrier Methods for Nonlinear Programming", in *Operations Research Support Methodology*, A. Holzman ed., Marcel Dekker, New York.

Fiacco, A.V. and G.P. McCormick (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley, New York.

Fletcher, R. (1971). “A General Quadratic Programming Algorithm", *J. Institute of Mathematics and its Applications*, 7, 76-91.

Frisch, R. (1955). “Logarithmic Potential Method of Convex Programming", Institute of Economics, University of Oslo.

Gill, P.E. , G.H. Golub, W. Murray, M.A. Saunders (1974). “Methods for Modifying Matrix Factorizations", *Mathematics of Computation*, 28, 505-535.

Gill, P.E, W. Murray, M.A. Saunders, J.A. Tomlin and M.H. Wright (1986). “On Projected Newton Barrier Methods for Linear Programming and an Equivalence to Karmarkar's Projective Method", *Mathematical Programming*, 36, 183-209.

Gill, P.E. and W. Murray (1976). “Minimization Subject to Bounds on the Variables", *Report NAC 72*, National Physical Laboratory, England.

Gill, P.E. and W. Murray (1978). “Numerically Stable Methods for Quadratic Programming", *Mathematical Programming*, 14, 349-372.

Gill, P.E. , W. Murray, M.A. Saunders, J.A. Tomlin, M.H. Wright (1986). “On Projected Newton Barrier Methods for Linear Programming and an Equivalence to Karmarkar's Projective Method", *Mathematical Programming*, 36, 183-209.

Gill, P.E. , W. Murray, M.H. Wright (1981). *Practical Optimization*, Academic Press, New York.

Goldfarb, D. (1972). “Extensions of Newton's Method and Simplex Methods for Solving Quadratic Programs", in F.A. Lootsma (Ed.), *Numerical Methods for Non-linear Optimization*, Academic Press, London and New York.

Goldfarb, D. and S. Liu (1993). “An O (n^3 L) Primal-Dual Potential Reduction Algorithm for Solving Convex Quadratic Programs", *Mathematical Programming,* 61, 161-170.

Golub, G.H and C. F. van Loan (1983). *Matrix Computations*, North Oxford Academic, Oxford.

Kojima, M., N. Meggido and T. Noma (1991).“Homotopy Continuation Methods for Complimentarity Problems", *Math. Oper. Res.*, 16, 754-774.

Kojima, M., N. Meggido, T. Noma and A. Yoshise (1991). *A Unified Approach to Interior Point Algorithms for Linear Complimentarity Problems*, Springer-Verlag, Berlin.

Kojima, M., S. Mizuno and T. Noma (1989). “A New Continuation Mathod for Complimentarity Problems with Uniform p-functions", *Mathematical Programming*, 43, 107-113.

Kojima, M., S. Mizuno and A. Yoshise (1989). “A Primal-dual algorithm for a Class of Linear Complimentarity Problems", *Mathematical Programming*, 44, 1-26.

Kojima, M., S. Mizuno and A. Yoshise (1991). “An O($\sqrt{n}$ L) Iteration Potential Reduction Algorithm for Linear Complimentarity Problems", *Mathematical Programming*, 50, 331-342.

Klee, V. and G. Minty (1972). “How Good is the Simplex Algorithm?", In *Inequalities-III*, ed. O. Shisha, Academic Press, New York.

Karmarkar, N. (1984). “A New Polynomial-Time Algorithm for Linear Programming", *Combinatorica*, 4, 373-395.

Khachiyan, L.G. (1979). “A Polynomial Algorithm in Linear Programming", *Soviet Mathematics Doklady*, 20, 191-194.

Lustig, I.J., R.E. Marsten and D.F. Shanno (1989). “Computational Experience with a Primal-dual Interior Point Method for Linear Programming", *Linear Algebra Appl.*, 152, 191-122.

Lustig, I.J., R.E. Marsten and D.F. Shanno (1992). “On Implementing Mehrotra's Predictor-Corrector Interior Point Method for Linear Programming", *SIAM J. Optimization*, 2, 435-449.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investment*, Wiley, New York.

Monteiro, R.C. and I. Adler (1989, a). “Interior Path-following Primal-Dual Algorithms. Part I: Linear Programming, *Mathematical Programming*, 44, 27-41.

Monteiro, R.C. and I. Adler (1989, b). “Interior Path-following Primal-Dual Algorithms. Part II: Convex Quadratic Programming, *Mathematical Programming*, 44, 43-66.

Papadimitriou, C.R. and K. Steiglitz (1982). *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, Englewood Cliffs, New Jersey.

Sharpe, W. (1970). *Portfolio Theory and Capital Markets*, McGraw-Hill, New York.

Stewart, G.W. (1973). *Introduction to Matrix Computations*, Academic Press, London and New York.

Smale, S. (1983). "On the Average Number of Steps of the Simplex Method of Linear Programming", *Mathematical Programming*, 27, 241-262.

Ye, Y. (1990). "Interior Point Methods for Global Optimization", *Annals of Operations Research*, 25, 59-74.

Ye, Y. (1991). "Interior Point Algorithms for Quadratic Programming", *Recent Developments in Mathematical Programming*, S. Kumar, ed., Gordon Breach Scientific Publishers, New York.

Ye., Y. and K. Anstreicher (1993). "On Quadratic and O($\sqrt{n}$ L) Convergence of a Predictor-corrector Algorithm for LCP, *Mathematical Programming*, 62, 537-551.

Ye, Y., O. Güler, R.A. Tapia, Y. Zhang (1993). "A Quadratically Convergent O($\sqrt{n}$ L)-iteration Algorithm for Linear Programming, *Mathematical Programming*, 59, 151-162.

Zhang, Y., R.A. Tapia, J.E. Dennis (1992). "On the Superlinear and Quadratic Convergence of Primal-dual Interior Point Linear Programming Algorithms", *SIAM J. Optimization*, 2, 304-324.

Zhang, Y., R.A. Tapia (1993). "A Superlinearly Convergent Polynomial Primal-dual Interior Point Linear Programming Algorithm for Linear Programming", *SIAM J. Optimization*, 3, 118-133.

Zhang, Y., R.A. Tapia, F. Potra (1993). "On the Superlinear Convergence of Interior Point Algorithms for a General Class of Problems", *SIAM J. Optimization*, 3, 413-422.