LINEARLY CONSTRAINED OPTIMISATION

By DIONYSIOS KARABALIS
M. Sc. (Essex), Bank of Greece.

I. INTRODUCTION

One of the basic problems in numerical optimisation is that of minimising a general function subject to a number of linear equality constraints. A class of methods to solve this problem is by the introduction of the classical Lagrangian function, and the problem reduced to the solution of a system of non-linear equations.

The Jacobian of the constraints is constant, and the system Jacobian has a well known special form

$$\left| \begin{array}{c|c} G & J^T \\ \hline J & 0 \end{array} \right| = \left| \begin{array}{c|c} K & J^T \\ \hline J & 0 \end{array} \right|$$

with the following properties :

a) G, and therefore the whole matrix, is symmetric.

b) J is the constraint Jacobian.

c) O is an m by m null matrix.

Two methods are described which update the system Jacobian, while maintaining the same form. One of them (method 2) preserves all the properties of the matrix, and the other (method 1) loses the symmetry.

II. SPECIAL QUASI-NEWTON METHODS FOR LINEAR CONSTRAINTS

1. Quasi-Newton Methods

1.1 Preliminaries, General Properties

The simplest form of Newton's method for solving the set of nonlinear equations

$$f(x) = 0 \quad x = (x_i)\ i = 1, \ldots, \tag{1a}$$

is

$$x^{(k+1)} = x^{(k)} - (J^{(k)})\ f^{(k)}\ a^{(k)} \tag{1b}$$

where $x^{(k)}$ is the kth approximation to the solution, $f^{(k)}$ denotes $f(x^{(k)}$, $j^{(k)}$ denotes the Jacobian of $f^{(x)}$ evalutated at $x^{(k)}$, and $a^{(k)} = 1$ for all k.

The advantages of (1b) are that if it works at all then it works extremely well, convergence is rapid, and, if a sufficiently good initial estimate of the solution can be determined, it is probably the best available method.

The most serious charge levelled against Newton's method is that it often fails to converge to a solution from a poor initial estimate. The second disadvantage of the above form is the difficulty of evaluating J(x) if f(x) is a complicated function of x. The third disadvantage of Newton's method is the necessity of solving a set of linear equations at each iteration.

The Quasi - Newton methods have been motivated to overcome the second disadvantage of Newton's method, that is they generate approximations to $J^{(k)}$ with no additional function evaluations. Suppose that at the kth iteration an approximation $B^{(k)}$ to $J^{(k)}$ is given, $B^{(k+1)}$ is forced to satisfly the Quasi - Newton equation

$$B^{(k+1)} \rho^{(k)} = y^{(k)} \tag{2a}$$

where

$$\rho^{(k)} = x^{(k+1)} - x^{(k)} \tag{2b}$$

$$y^{(k)} = f^{(k+1)} - f^{(k)}. \tag{2c}$$

$B^{(k+1)}$ is regarded as «corrected» version of $B^{(k)}$ that is

$$B^{(k+1)} = f^{(k)} + D$$

where D is the correction matrix (update) usually of low rank.

In the casr where the Jacobian matrix is square (and non singular) instead of storing and modifying an approximation $B^{(k)}$ to the Jacobian it is sufficient to store and modify an approximation to the inverse Jacobian. If this is denoted by $H^{(k)}$ equation (1b) becomes

$$x^{(k+1)} = x^{(k)} - H^{(k)} f^{(k)} a^{(k)} \tag{3}$$

when seeking an Quasi - Newton algorithm to solve a given problem we would take into consideration the following

(1) the algorithm should not converge to an incorrect solution. An example of this would be convergence to a saddle point.

(2) Premature or false convergence should be avoided. It sometimes occurs that the matrix H becomes singular or nearly so, and the resulting step length $||p||$ becomes in consequence negligible. If one terminated the iteration by testing $||y||$ alone, then one could have false convergence, but this may be prevented by testing also $||g||$ which is the norm of the gradient of the function F(x) under minimisation.

(3) Algorithms should not fail catastrophically when updating the matrix H. Another feature that affects the overall speed of an algorithm is the amount

of work required during each iteration. If $a^{(k)} = 1$ in (1b) and (3) then the amount of work is minimal. If $H^{(k)}$ is positive definite and $x^{(k+1)}$ is given by (3) then, for $a^{(k)}$ sufficiently small and positive, we have $F^{(k+1)} < F^{(k)}$.

1.2. Quasi - Newton methods for Equality Constraints

Assuming that the first partial derivates of the objective function and the constraints are explicitly available, with initial estimates of x and λ, then Quasi - Newton iterations can be constructed for their improvement estimates. The characteristic feature of such methods is that they generate, from information readily available, an approximation to the Jacobian of the non - linear system under consideration. The exact Jacobian of the system, when partitioned into four sub-matrices has the form

$$\left[\begin{array}{c|c} G & J^T \\ \hline J & 0 \end{array}\right]$$

where G is the Hessian of $\varphi(x,\lambda) = F(x) + \lambda^T C(x)$. The proposed iterations are various combinations of the following properties of the matrix.

A. G, and therefore the whole matrix, is symmetric.

B. J is the constraint Jacobian which will be evaluated at each iteration in order to calculate the residuals.

C. O is an m by m null matrix.

In principle then the only unknown part of this Jacobian is the submatrix G, and a Newton - like method it appears that only this submatrix need be estimated. However as mentioned in section 1.1. the efficiency of the Quasi - Newton methods is partly due to the direct generation of approximations to the inverse, thereby avoiding the numerical solution of a linear system at each step.

A well - known formula gives the inverse of a matrix partitioned in the above manner as

$$\left[\begin{array}{c|c} G & J^T \\ \hline J & 0 \end{array}\right]^{-1} = \left[\begin{array}{c|c} X & -G^{-1} J^T W \\ \hline -W J G^{-1} & W \end{array}\right] \qquad (4a)$$

where

$$W = -(JG^{-1} J^T)^{-1} \qquad (4b)$$

$$X = G^{-1} + G^{-1} J^T W J G^{-1} \qquad (4c)$$

It is possible therefore to create direct updating procedures for the inverse. However the validity of this inversion depends upon the sub - matrix G being non - singular, and this cannot be guaranteed. Hence, methods which overcome this problem are considered.

1.3. Quasi - Newton Iterations

Suppose that in the solution of (1) estimates of x and B^{-1} are available, where B is an approximation of the system Jacobian. A Quasi - Newton iteration updating x to x_1 and B^{-1} to B_1^{-1} then takes the form similar to (2)

$$\rho = -B^{-1} f \qquad (5a)$$

$$x_1 = x + \rho \qquad (5b)$$

$$B_1^{-1} = = B^{-1} + D \qquad (5c)$$

Quasi - Newton methods are characterised by choosing D so that the matrix B_1 satisfies the Quasi - Newton equation

$$B_1\rho = f_1 - f = y. \qquad (6)$$

This relationship does not of course determine B_1 uniquely and different methods are derived from different choices of the n - m degrees of freedom.

One such method, which has some satisfactory theoretical properties, in addition to justifying itself in practice is that proposed by Broyden 1965

$$B_1 = B + (y - B\rho)\bar{\rho}^T \qquad (7a)$$

where

$$\bar{\rho}^T = \rho^T / \rho^T \rho \qquad (7b)$$

Of course a routine use of Sherman - Morrison formula allows (7) to be written as a correction to B^{-1} with no reference to B itself. This procedure is followed for the construction of the new methods.

In view of the symmetry of the exact Jacobian for the system we have in mind, a symmetric analogue of this update, given by Powell 1970, would seem a useful choice to make. Assuming that B is symmetric, they Powell's formula is

$$B_1 = B + (y - B\rho)\,\bar{\rho}^T + \bar{\rho}\,(y - B\rho)^T - 2\theta\,\bar{\rho}\,\bar{\rho}^T \qquad (8a)$$

where

$$2\theta = \rho^T y - \rho^T B\rho, \qquad (8b)$$

Both of these methods fall within the methods satisfying all the properties mentioned in 1.1.

The update given by the formula (8) is not the only one which preserves symmetry, for example, the simple rank update

$$B_1 = B + (y - B\rho)\,(y - B\rho)^T / a \qquad (9a)$$

where

$$a = (y - B\rho)^T \rho \qquad (9b)$$

also preserves symmetry. However, the denominator a for this update could eb possibly zero, another source of unreliability; this is rejected therefore in favour of formula (8).

Having established means of handling symmetry we now consider the thrid of the listed properties of the system Jacobian, and assume that the approximations B, B_1 take the forms

$$\left[\begin{array}{c|c} K & N \\ \hline M & 0 \end{array}\right] \quad \text{and} \quad \left[\begin{array}{c|c} K_1 & N_1 \\ \hline M_1 & 0 \end{array}\right]$$

respectively. The partitions M and N are not now equal to J and J^T respectively, despite the assumption that these latter matrices are known exactly.

The tehnique is used to preserve this structure is to update the three partitions separately, and for this an analogue of (6) for each partition is necessary. For this case, equations (5a) and (5b) and they become

$$\begin{bmatrix} K & N \\ M & 0 \end{bmatrix} \left[\begin{array}{c} \rho \\ \hline \\ \end{array}\right] = \left[\begin{array}{c} -d \\ \hline -C \end{array}\right] \tag{10a}$$

$$\left[\begin{array}{c} x_1 \\ \hline \lambda_1 \end{array}\right] = \left[\begin{array}{c} x \\ \hline \lambda \end{array}\right] + \left[\begin{array}{c} \rho \\ \hline q \end{array}\right] \tag{10b}$$

and the Quasi - Newton equation (6) becomes

$$K_1 \rho + N_1 q = d_1 - d \tag{11a}$$

$$M_1 \rho = c_1 - c. \tag{11b}$$

The latter equations are in general satisfied exactly by the true system Jacobian only if the system is linear. Since it is assumed that the exact system Jacobian is available at each iteration, we can also consider modifications of these equations which are less restricted to the linearity assumption.

In fact the vector $J_1 p$ and $J_1^T q$ are easily evaluated, and since M_1, N_1 are approximations to J_1, J_1^T respectively, we may impose the linear relationships

$$M_1 \rho = J_1 \rho \tag{12a}$$

$$N_1 q = J_1^T q. \tag{12b}$$

Finally, in order to maintain the symmetry we may set

$$N_1 = M_1^T. \tag{12c}$$

Last equations represent some of the known properties of the exact Jacobian, and there can be used to update the partitions of the approximate Jacobian. From the numerous possibilities the following are chosen in order to restrict the correction to rank two.

(i) M_1 is derived from the Broyden formula (7)

$$M_1 = M + (z - M\rho)\bar{\rho}^T \tag{13a}$$

where

$$\text{either} \quad z = c_1 - c \tag{13b}$$

$$\text{or} \quad z = J_1 \rho \tag{13c}$$

(ii) N_1 is derived from M_1 using (12c), or from a direct application of Broyden's formula (7) using (12b).
The latter formula is

$$N_1 = N + (w - Nq)q^T \tag{14a}$$

where

$$w = J_1^T q\,. \tag{14b}$$

(iii) K_1 is derived either from Broyden's formula, giving

$$K_1 = K + (y - K\rho)\bar{\rho}^T \tag{15}$$

or from Powell's formula if symmetry is to be maintained giving where

$$K_1 = K + (y - K\bar{\rho})\ \bar{\rho}^T + \bar{\rho}\,(y - K\rho)^T - 2\theta\,\bar{\rho}\ \bar{\rho}^T \tag{16a}$$

where

$$2\theta = \rho^T y - \rho^T K \rho\,. \tag{16b}$$

Here the vector defining the Quasi - Newton equation is taken to be

$$\text{either} \quad y = d_1 - d - N_1 q \tag{17a}$$

$$\text{or} \quad y = d_1 - d - J_1^T q \tag{17b}$$

2. Special Updates for Linear - Constraints

2.1. Preliminaries

If the constraints are all linear in the variables x, then the constraint Jacobian is a constant m by n matrix J, and two Quasi - Newton updates can be constructed which allow for this. If we suppose that the approximation of the system Jacobian at the beginning of an iteration has the form

$$B = \left[\begin{array}{c|c} K & J^T \\ \hline J & 0 \end{array}\right] \tag{18}$$

Then there will be two updates, based upon Broyden's update, which can be used to update this matrix while maintaining this form.

It will be assumed that at the beginning of the iteration process, the initial Jacobian inverse is consistent with a matrix of the above form. Two updates can be constructed, one which preserves all the properties of B, and the other one which loses the symmetry of the upper left - hand partition K.

The methods are on the update of the matrix K only, using but part of the Quasi - Newton equation (11) that is the formula

$$K\rho = d_1 - d - N_1 \tag{19a}$$

or by (12b)

$$K\rho = d_1 - d - J^T q . \tag{19b}$$

The vector $J^T q$ can of course be calculated, and in some respect it resembles the «modified» Quasi - Newton equation.

2.2. Method - 1, for Linear Constraints

Assuming that the approximate Jacobian for the non - linear system has the form (18) we can use a Broyden rank -1 updatd (7) to create a new approximation to the partition K only at each iteration. The exact constraint Jacobian provides a Quasi - Newton equation for the K partition as follows

NEWTON STEP :

$$K\rho + J^T q = -d$$

$$J\rho = -c$$

QUASI - NEWTON EQUATION :

$$K_1 \rho = y = d_1 - d - J^T q$$

BROYDEN RANK - 1 UPDATE FOR K :

$$K_1 = K + (y - K\rho)\,\bar{\rho}^T$$

The complete update then can be written as

$$\left[\begin{array}{c|c} K_1 & J^T \\ \hline J & 0 \end{array}\right] = \left[\begin{array}{c|c} K & J^T \\ \hline J & 0 \end{array}\right] + rs^T \tag{20}$$

where

$$r = \left[\frac{y - K\rho}{0}\right] = \left[\frac{d_1}{0}\right] \tag{21}$$

and

$$s = \left[\frac{\rho}{0}\right]$$

and Applying the Sherman - Morrison formula to the inversion gives

$$H_1 = H - \frac{Hr\, s^T H}{1 + s^T Hr} \tag{22a}$$

which can be evaluated immediately from Hr and $s^T H$,

$$Hr = H\left[\frac{d_1}{0}\right] \tag{22b}$$

$$s^T H = [(\bar{\rho}^T \mid 0)]\, H \tag{22c}$$

2.3. **Method - 2 for Linear Constraints**

Assuming now that the system Jacobian has'the form (18) where K is symmetric, we can apply the Powell's update (8) to the partition K only at each iteration.

NEWTON STEP :

$$K\rho + J^T q = -d$$
$$J^T q = -c$$

QUASI - NEWTON EQUATION :

$$K_1 \rho = y = d_1 - d - J^T q$$

POWELL SYMMETRIC UPDATE ON N :

$$K_1 \rho = K + (y - K\rho)\bar{\rho}^T + \bar{\rho}\,(y - K\rho)^T - 2\theta\,\bar{\rho}\,\bar{\rho}^T$$

$$2\theta = \rho^T y - \rho^T K \rho = \rho^T d_1 .$$

The complete update is

$$\left[\begin{array}{c|c} K_1 & J^T \\ \hline J & 0 \end{array}\right] = \left[\begin{array}{c|c} K & J^T \\ \hline J & 0 \end{array}\right] + r\, s^T + s\, r^T \tag{23a}$$

where

$$r = \left[\frac{y - K\rho - \theta\bar{\rho}}{0}\right] = \left[\frac{d_1 - \theta\bar{\rho}}{0}\right] \tag{23b}$$

and

$$S = \left[\frac{\bar{\rho}}{0}\right| \tag{23c}$$

The Sherman Morrison formoula applied to the iteration gives

$$H_1 = H + \frac{1}{D}\{\,|-(1 + r^T H s)\, H r + (r^T H r)\, H s \mid s^T H +$$

$$+ \;|\,(s^T H s)\, H r - (1 + s^T H r)\, H s \mid r^T H\} \tag{24a}$$

where

$$D = (1 + r^T Hs)\;\;(1 + s^T Hr) - (s^T Hr)\;(r^T Hr). \tag{24b}$$

The relevant vectors and scalar products are given by

$$2\theta = \rho^T d_1$$

$$Hr = H\left[\frac{d_1 - \theta\bar{\rho}}{0}\right], \quad Hs = H\left[\frac{\bar{\rho}}{0}\right]$$

$$\begin{aligned} r^T H s &= (Hr)^T s \;, & r^T\ Hr &= (Hr)^T r \\ s^T H s &= (Hs)^T s \;, & s^T\ Hr &= (Hs)^T r \\ s^T H &= (Hs)^T \;. & r^T\ H &= (Hr)^T \;. \end{aligned} \tag{24c}$$

Thus (24a) can be evaluated from (24c).

III. EXPERIMENTAL RESULTS - CONCLUSION

The two methods dealing with linear constraints have been tested together with other methods on two different problems BVPI and BVP1. The BVP1 is the differential equation :

$$F = Y - (Y')^{**}\,2 + 2 + 3\,T^{**}\,2$$

subject to the boundary conditions

$$Y - Y' = 0$$

$$Y - Y' = 3$$

The BVP2 is the differeutial ebuation :

$$F = -Y^*\,Y - D^*\,D - Y + 1$$

subject to the boundary conditions

$$Y = 0$$

$$Y = 1$$

From the different experimental results taken place we can state that method - 1 has advantages in computation over similar methods for linear and non - linear constraints.

REFERENCES

1. C. T. Broyden, W. H. HART, 1970 «Quasi - Newton Methods for Constraint Optimisation» Presented at the Seventh Mathematical - Programming Symposium.

2. R. Fletcher, 1971, «Methods for the Solution of Optimisation Problems» Presented at the symposium on Computer - aided Engineering, University of Waterloo, Canada.

3. R. Fletcher, 1971, «Minimising General functions subject to Linear - Constraints» Preseuted at an International Conference on Optimisation, University of Dundee, Scotland.