Journal of Computational Mathematics, Vol.8, No.3, 1990, 195–201.

A NEW METHOD FOR EQUALITY CONSTRAINED OPTIMIZATION*

Jia Li-xing
(*Department of Mathematics, Shaanxi Normal University, Xi'an, Shaanxi, China*)

Abstract

This paper prensents a detailed derivation and description of a new method for solving equality constrained optimization problem. The new method is based upon the quadratic penalty function, but uses orthogonal transformations, derived from the Jacobian matrix of the constraints, to deal with the numerical ill-conditioning that affects the methods of this type.

At each iteration of the new algorithm, the orthogonal search direction is determined by a quasi-Newton method which can avoid the necessity of solving a set of equations and the step-length is chosen by a Armijo line search. The matrix which approaches the inverse of the projected Hessian of composite function is updated by means of the BFGS formula from iteration to iteration. As the penalty parameter approaches zero, the projected inverse Hessian has special structure which can guarantee us to obtain the search direction accurately even if the Hessian of composite function is ill-conditioned in the former penalty function methods.

§1. Introduction

We consider the problem

$$\text{minimize} \quad F(x), \quad x \in R^n, \tag{1.1a}$$

$$\text{subject to} \quad c(x) = 0, \quad c \in R^m, \quad m \le n \tag{1.1b}$$

where $F(x)$ and $c_i(x)(i = 1(1)m)$ are all twice continuously differentiable functions of x.

The former penalty function method for solving (1.1) is minimizing the composite function

$$\Phi(x, r) = F(x) + c^T c/2r \tag{1.2}$$

where $c^T c/2r$ is the quadratic penalty term and r is the penalty parameter. It is known that if x^* is the solution of (1.1) and $x^*(r)$ is the unconstrained minimum of (1.2), then under mild conditions [4],

$$\lim_{r \to 0} x^*(r) = x^*.$$

Thus, we will deal with the unconstrained problem

$$\text{minimize}\Phi(x, r) \tag{1.3}$$

in stead of (1.1). The method we suggest for solving (1.3) will generate a sequence that converges (as $r \to 0$) to solution x^* of the original problem (1.1).

* Received July 15, 1987.

§2. Basic Philosophy

In order to simplify our discussion, we introduce the following notations:

g : gradient of the objective functions F,
J : Jacobian of the constraints, $J = [\partial c_i/\partial x_j]$, $\operatorname{rank}(J) = m$,
f : gradient of $\Phi(= g + J^T c/r)$,
Q : orthogonal matrix satisfying $JQ^T = [U^T, O]$, where U is upper triangular,
h : projected gradient, $h = Qg$,
H_0 : Hessian of F,
H_i : Hessian of c_i,
H_Φ : $H_\Phi = H_0 + (1/r)\sum_{i=1}^{m} c_i H_i$,
G : projected Hessian, $G = QH_\Phi Q^T$,
$\overline{B}$: an approximation to the inverse of projected Hessian of Φ.

If we solve (1.2) by applying Newton's method to the equation

$$f = g + J^T c/r = 0,$$

to perform one step of Newton's iteration, we must solve the equation

$$(H_\Phi + J^T J/r)p = -f \tag{2.1}$$

to determine the search direction p. The matrix H_Φ is generally well-behaved but for the standard penalty function method, $J^T J$ is of rank m and $\|J^T J/r\| \to \infty$ as $r \to 0$, with singularity occuring in the limit [5], which causes difficulty for solving (2.1). This is a fatal problem if we try to solve equation (2.1) directly. The second difficulty is that, while using Newton's method, we must supply a formula with which the second derivative matrices can be evaluated, and this can be a major disincentive if $F(x)$ and/or $c(x)$ are complicated functions of x. The third and final problem is the necessity of solving a set of equations at each iteration. Therefore, we may need to investigate a new method for solving (2.1).

§3. The New Method

In order to devise a new method, we rewrite equation (2.1) as

$$(H_\Phi + J^T J/r)p = -(g + J^T c/r), \tag{3.1}$$

and impose an orthogonal transformation on J^T, the transpose of the Jacobian of $c(x)$. Then we can obtain an orthogonal matrix Q :

$$QJ^T = \begin{bmatrix} U \\ O \end{bmatrix}$$

where U is upper triangular and nonsingular since J has full row rank. Then we transform equation (3.1) using the orthogonal matrix Q :

$$Q(H_\Phi + J^T J/r)Q^T Qp = -Q(g + J^T c/r),$$

that is,

$$\bar{G}\bar{p} = -\bar{f} \tag{3.2}$$

where

$$\bar{G} = Q(H_\Phi + J^T J/r)Q^T = G + (1/r)\begin{bmatrix} UU^T & O \\ O & O \end{bmatrix} = \begin{bmatrix} G_{11} + (1/r)UU^T & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$$

and

$$G = QH_\Phi Q^T, \bar{p} = Qp, \bar{f} = Q(g + J^T c/r) = \begin{bmatrix} h_1 + Uc/r \\ h_2 \end{bmatrix}$$

where

$$h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = Qg.$$

In implementation of the new algorithm, the following conclusion can be used as a stopping criterion to terminate the algorithm.

Lemma 3.1. *Let x^* be a solution of* (1.1). *Then the last $(n-m)$ components of $\bar{f}^* = \bar{f}(x^*)$ approach to zeros, i.e., $h_2^* \to 0$.*

Proof. Since x^* is optimal, from the first necessary condition, we have

$$g^* = J^{*^T}\lambda^*$$

where λ^* is the Lagrange multipliers at x^*. Substituting g^* with f^*, we obtain

$$f^* = J^{*^T}\lambda^* + J^{*^T}c^*/r = J^{*T}(\lambda^* + c^*/r). \tag{3.3}$$

Multiplying (3.3) by Q^*, we find that

$$\begin{aligned} \bar{f}^* &= Q^* f^* = Q^* J^{*^T}(\lambda^* + c^*/r) = \begin{bmatrix} U^* \\ O \end{bmatrix}(\lambda^* + c^*/r) \\ &= \begin{bmatrix} U^*(\lambda^* + c^*/r) \\ O \end{bmatrix} = \begin{bmatrix} h_1^* + Uc^*/r \\ h_2^* \end{bmatrix}. \end{aligned} \tag{3.4}$$

Hence, at the solution of x^*, we have zeros in the last $(n-m)$ components of the projected gradient $\bar{f}^*$, which justifies the use of (3.4) as a stopping criterion.

To terminate the algorithm, we use both of the following two stopping criteria

$$\|h_2(x)\| < \varepsilon_1 \tag{3.5}$$

and

$$\|c(x)\| < \varepsilon_2. \tag{3.6}$$

The use of (3.6) is to check whether the constraints are satisfied and to prevent the premature termination.

The following two theorems show the crucial points of the new method compared with the former method.

Theorem 3.1. *Let E be a change in equation* (3.1) *and e be the change in the solution p induced by the change E due to rounding to t-binary digits. Then $\|e\|/\|p\|$ is unbounded as the penalty parameter r approaches zero.*

Proof. Consider the equation

$$(H_\Phi + J^TJ/r + E)(p+e) = -f.$$

Let $A = H_\Phi + J^TJ/r$. If A is nonsingular and $\|A^{-1}\|\cdot\|E\| < 1$, then, we have[9]

$$\frac{\|e\|}{\|p\|} \le \frac{k(A)\dfrac{\|E\|}{\|A\|}}{1-k(A)\dfrac{\|E\|}{\|A\|}} \tag{3.7}$$

where $k(A) = \|A\|\|A^{-1}\|$. Since error E is caused by rounding $A = H_\Phi + J^TJ/r$ in order to store it to t-binary places (assume that H_Φ and J^TJ are exact), then, if $\|\cdot\|$ denotes $\|\cdot\|_1$ or $\|\cdot\|_\infty$, $\|E\| \le 2^{-t}\|A\|$, (3.7) becomes

$$\frac{\|e\|}{\|p\|} \le \frac{2^{-t}k(A)}{1-2^{-t}k(A)}.$$

As $r \to 0$, $k(A) \to \infty$, the bound becomes indefinitely large as $k(A) \to 2^t$.

Theorem 3.2. *Let $\bar{E}$ be a change in equation (3.2) and $\bar{e}$ be the change in the solution $\bar{p}$ induced by the change $\bar{E}$ due to rounding to t-binary digits. Then $\|\bar{e}\|/\|\bar{p}\|$ is bounded as the penalty parameter approaches zero.*

Proof. For simplicity, we assume that, in $\bar{G}, G_{12}, G_{21}$ and G_{22} are exact. Consider the equation

$$\begin{bmatrix} G_{11}+UU^T/r+\bar{E}_1 & G_{12} \\ G_{21} & G_{22}\end{bmatrix}(\bar{p}+\bar{e}) = -\begin{bmatrix} h_1+Uc/r \\ h_2\end{bmatrix}$$

and let $\bar{p} = [\bar{p}_1, \bar{p}_2]^T$, $\bar{e} = [\bar{e}_1, \bar{e}_2]^T$. Then we have

$$(G_{11}+UU^T/r+\bar{E}_1)(\bar{p}_1+\bar{e}_1) + G_{12}(\bar{p}_2+\bar{e}_2) = -h_1 - Uc/r \tag{3.8}$$

and

$$G_{21}(\bar{p}_1+\bar{e}_1) + G_{22}(\bar{p}_2+\bar{e}_2) = -h_2 \tag{3.9}$$

or

$$G_{22}(\bar{p}_2+\bar{e}_2) = -h_2 - G_{21}(\bar{p}_1+\bar{e}_1). \tag{3.9'}$$

Calculating (3.8) and (3.9) while thinking of (3.2), we obtain

$$(G_{11}+UU^T/r+\bar{E}_1)\bar{e}_1 + G_{12}\bar{e}_2 = -\bar{E}_1\bar{p}_1, \tag{3.10}$$

$$G_{21}\bar{e}_1 + G_{22}\bar{e}_2 = 0. \tag{3.11}$$

From (3.11), we have

$$\bar{e}_2 = -G_{22}^{-1}G_{21}\bar{e}_1. \tag{3.12}$$

Substituting (3.12) into (3.10), we have

$$(UU^T/r + G_{11} + \bar{E}_1 - G_{12}G_{22}^{-1}G_{21})\bar{e}_1 = -\bar{E}_1\bar{p}_1.$$

Multiplying r to the above equation, as $r \to 0$, we obtain

$$(UU^T + r\bar{E}_1)\bar{e}_1 = -r\bar{E}_1\bar{p}_1 \tag{3.13}$$

which yields

$$\frac{\|\bar{e}_1\|}{\|\bar{p}_1\|} \le \frac{k(UU^T)\dfrac{\|r\bar{E}_1\|}{\|UU^T\|}}{1 - k(UU^T)\dfrac{\|r\bar{E}_1\|}{\|UU^T\|}}. \tag{3.14}$$

If $\|\bar{E}_1\| \le 2^{-t}\|UU^T\|/r$, (3.14) becomes

$$\frac{\|\bar{e}_1\|}{\|\bar{p}_1\|} \le \frac{2^{-t}k(UU^T)}{1 - 2^{-t}k(UU^T)}. \tag{3.15}$$

The R.H.S. of (3.15) is independent of the penalty parameter r, and UU^T is nonsingular. Hence, despite $r \to 0$, (3.15) is bounded. On the other hand, from Lemma 1, since $h_2 \to 0$ at the solution, (3.9′) becomes

$$G_{22}(\bar{p}_2 + \bar{e}_2) = -G_{21}(\bar{p}_1 + \bar{e}_1)$$

which yields[9]

$$\frac{\|\bar{e}_2\|}{\|\bar{p}_2\|} \le k(G_{22})\frac{\|G_{21}\bar{e}_1\|}{\|G_{21}\bar{p}_1\|}$$

If $0 < m \le \|G_{21}\| \le M$, then

$$\frac{\|\bar{e}_2\|}{\|\bar{p}_2\|} \le \frac{M}{m}k(G_{22})\frac{\|\bar{e}_1\|}{\|\bar{p}_1\|}. \tag{3.16}$$

Since G_{22} is nonsingular and $\|\bar{e}_1\|/\|\bar{p}_1\|$ is bounded, $\|\bar{e}_2\|/\|\bar{p}_2\|$ is also bounded.

In practice, the orthogonal search direction $\bar{p}$ is obtained by using a quasi-Newton method, i.e. $\bar{p} = -\bar{B}\bar{f}$, where $\bar{B}$ is an approximation to the inverse of $\bar{G}$.

Thus, the new algorithm may be stated as:

1. Choose an initial value of x and a fixed sequence $\{r_k\} \to 0$; set the intial $\bar{B} = I$.

2. For each r_k, find a local minimiser, say, $x^*(r_k)$ to minimise $\Phi(x, r_k)$. At each iteration, the algorithm proceeds as follows:

   1) compute f and J;
   2) compute Q and U;
   3) compute $\bar{f} = Qf$;
   4) compute $\bar{p} = -\bar{B}\bar{f}$;
   5) compute $p = Q^T\bar{p}$;
   6) compute $x \leftarrow x + tp$, where t is chosen to reduce Φ by the Armijo line search[7];
   7) update $\bar{B}$ by using the BFGS formula[8];
   8) repeat from 1) until $\|p\|/\|x\| < \varepsilon$, where ε is a pre-set tolerance.

3. Check whether both of the conditions $\|h_2\| < \varepsilon_1$ and $\|c(x)\| < \varepsilon_2$ are satisfied, where ε_1 and ε_2 are tolerances specified by the user. If they are satisfied, then terminate the algorithm and accept the current solution x_k and $\Phi(x_k, r_k)$. Otherwise, set $r_{k+1} = r_k/\text{const}$, return to step 2 to minimise $\Phi(x, r_{k+1})$.

§4. Numerical Results

In order to evaluate the new algorithm, various examples have been tested. All the examples are chosen from the collection given by Hock and Schittkowski[6] and the tests have been performed on DEC10/20 computer. All the calculations of the test problems are carried out using programs written in ALGOL in single precision with 27 bit mantissa, approximately equal to 9 decimal digits.

In this section, only four examples are selected from the tested problems to compare the efficiency of the new method (NM) with the former method (FM). They are identified with the same number as in [6]. We choose const $= 10^2, \varepsilon = 10^{-5}$ and $\varepsilon_1 = \varepsilon_2 = 10^{-7}$ in the new algorithm.

The following information is given for each example:

(1) Ni: number of iteration,

(2) Nf: number of composite function evaluation,

(3) Ng: number of gradient vector evaluation,

(4) Ex: $Ex = \|x^* - \bar{x}\|_\infty$; $\bar{x}$ is the solution given by [6],

(5) Ef: $Ef = |F^* - \bar{F}|, F^* = F(x^*) = \Phi(x^*), \bar{F} = F(\bar{x})$.

No.	Method	Ni	Nf	Ng	Ex	Ef
39	FM	60	140	97	0.00002904	0.00001035
	NM	44	47	45	0.00000004	0.00000000
48	FM	36	54	54	0.00001994	0.00000062
	NM	26	29	27	0.00000004	0.00000000
77	FM	36	63	50	0.00005300	0.00000009
	NM	35	42	36	0.00000058	0.00000011
78	FM	41	70	48	0.00002276	0.00008152
	NM	20	38	26	0.00000098	0.00000044

§5. Conclusions

The solution of nonlinearly constrained optimization problems using penalty function methods inevitably leads to serious numerical disadvantages unless appropriate precautions are taken. Hence, the former quadratic penalty function method was considered to be unsuitable for solving the equality constrained problems.

In this paper, a new algorithm which utilizes the quadratic penalty function method with orthogonal transformations based on the Jacobian of the constraints shows that the quadratic penalty function method is viable and that the deficiencies generally associated with the former method can be overcome.

The other advantages of the new method are:

1. For the new method, the number of evaluations of the composite function, gradient, Jacobian and constraints vector is less than the number by using the former quadratic penalty function method.

2. The results given in Section 4 shows that the new method can solve the equality constrained problem stably and accurately.

3. For the new algorithm, the computation of the search direction does not require solving any system of linear equations, and can be expected to require much less work than in some other methods.

Acknowledgment. The author would like to thank Prof. C. G. Broyden and Dr. J. A. Ford for their supervision during his study in the Department of Computer Science, University of Essex, England, and also to express his sincere gratitude to the referees for their helpful comments.

References

[1] C. G. Broyden, Quasi–Newton methods, in: M. Murray, Ed., Numerical Methods for Unconstrained Optimization, Academic Press, London & New York, 1972, 86–106.
[2] C. G. Broyden and N. F. Attia, Penalty functions, Newton methods and quadratic programming, A Report on Nonlinear Optimization, University of Essex, England, 1985.
[3] J. E. Dennis and R. B. Schnable, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prence-Hall, Englewood Cliffs, New Jersey, 1983.
[4] A. V. Fiacco and G. P. McCormic, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Willey and Sons, New York, 1986.
[5] P. E. Gill, W. Murray and M. H. Wright, Practical Optimization, Academic Press, London, 1981.
[6] W. Hock and K. Sckittkowski, Test Examples for Nonlinear Programming Codes, Springer Verlag, Berlin-Heidelberg-New York, 1981.
[7] Jia Li–xing, New methods for equality constrained optimization, M. Phil thesis , University of Essex, 1986.
[8] L. E. Scale, Introduction to Nonlinear Optimization, MacMillan, 1985.
[9] J. H. Willknson, Rounding error in algebraic process, Notes on Applied Science, No. 22, Her Magesty's Stationary Office, London, 1963.