TECHNICAL RESEARCH REPORT

S Y S T E M S R E S E A R C H C E N T E R

Supported by the
National Science Foundation
Engineering Research Center
Program (NSFD CD 8803012),
Industry and the University

# A Note on the Positive Definiteness of BFGS Update in Constrained Optimization

by J. Zhou

TR 90-75

A Note on the Positive Definiteness of BFGS Update in Constrained Optimization[1]

*Jian Zhou*

Systems Research Center

and

Department of Electrical Engineering

University of Maryland, College Park, MD 20742

**Abstract**

This note reviews a few existing methods to maintain the positive definiteness of BFGS in constrained optimization, and their impacts on both global and local convergence. The boundedness of the matrix from above is also briefly addressed. Some new strategies are proposed. Convergence analysis and numerical examples are not included.

**Key words:** Positive definiteness, BFGS update, constrained optimization, global and local convergence.

**1. Introduction.** Consider the problem

$$(P) \qquad \min_{x \in R^n} f(x) \quad \text{s.t.} \quad x \in X$$

where $X$ is defined as $X = \{x : g_i(x) \leq 0 \; i = 1, \ldots, m; \; g_i(x) = 0 \; i = m+1, \ldots, m'\}$ and assumed compact. We assume that $f : R^n \to R$ and $g_i : R^n \to R$ for $i = 1, \ldots, m'$ are three times continuously differentiable. The classical sequential quadratic programming (SQP) method solves $(P)$ by iteratively solving the following model problem (possibly with some

[1]This research was supported by NFS grant No. DMC-88-15966.

variations):

$$
(M)\qquad \begin{array}{ll} \min\limits_{d\in R^n} & \frac{1}{2}\langle d, H_k d\rangle + \langle \nabla f(x_k), d\rangle \\ \text{s.t.} & g_i(x_k) + \langle \nabla g_i(x_k), d\rangle \le 0, \quad i = 1, \ldots, m \\ & g_i(x_k) + \langle \nabla g_i(x_k), d\rangle = 0, \quad i = m+1, \ldots, m' \end{array}
$$

where $H_k$ is some positive definite matrix. And $x_{k+1} = x_k + t_k d_k$ for some $t_k \in (0,\ 1]$. For simplicity, we omit the subscript $k$ for any quantity at iteration $k$ and add an upper bar for the quantity at iteration $k+1$. If the Lagrange function associated with $(P)$ is defined by

$$
L(x, \lambda) = f(x) + \sum_{1}^{m} \lambda_i g_i(x) + \sum_{m+1}^{m'} \mu_i g_i(x), \tag{1}
$$

then it is usually desired that in the tangential subspace defined by the active constraints $I(x) = \{i \in \{1, \ldots, m\} :\ g_i(x) = 0\} \cup \{m+1, \ldots, m'\}$, $H$ approximate the Hessian matrix of $L(x, \lambda)$ with respect to $x$. Among many formulae which update $H > 0$ based on the first order information from previous iterations is the well known BFGS update which in its original form is given by

$$
\bar{H} = H - \frac{(Hd)(Hd)^T}{d^T H d} + \frac{\eta\eta^T}{\eta^T d} \tag{2}
$$

with $d = \bar{x} - x$ and

$$
\eta = \nabla_x L(\bar{x}, \lambda) - \nabla_x L(x, \lambda) \tag{3}
$$

where $\lambda$ is associated with $(M)$. In some multiplier updating methods, $\lambda$ involved above could be $\bar{\lambda}$ updated in a certain way. A basic and essential assumption is $\nabla_{xx} L(x, \lambda) > 0$, or so at least in the tangential subspace at a local solution, for many established theorems on superlinear convergence with both full Hessian updating and projected Hessian updating (e.g. [1], [2], [3], [4]). This legitimates the procedure of keeping $\bar{H} > 0$ if $H > 0$. It can be shown, due to

$$
\det[H - \frac{(Hd)(Hd)^T}{d^T H d}] = 0
$$

and the eigenvalue interlocking property, that a sufficient and necessary condition for $\bar{H} > 0$ is $\eta^T d > 0$. As is well known, this condition can be always achieved (though possibly very costly) in unconstrained optimization by means of Wolfe's line search [5]. Yet it is not true that we can always do so in constrained situation. In fact this can only be possibly done if the neighborhood of a strong minimizer (satisfying 2nd order sufficiency conditions) is approached since then we have that

$$\langle d, \bar{H}d\rangle = \langle \eta, d\rangle \approx \langle d, \nabla_{xx}L(x,\lambda)d\rangle > 0, \tag{4}$$

assuming that in the limit $d$ is driven close to the tangential space. So far it is still an open problem to steer the iterates to a local strong minimizer (if any). Most methods existing in the literature just solve the first order necessary conditions! Hence Wolfe's line search won't work in nonlinear constrained optimization. Even in the context of projected Hessian quasi-Newton method, there is no guarantee that the curvature is nonnegative when $x$ is far away from a local minimizer (this is also true in unconstrained optimization). Thus we are faced with the same problem in both full Hessian updating and projected Hessian updating.

Now that $\eta^T d$ cannot be always positive, an important issue is how to maintain the positive definiteness of $H$ without destroying the convergence properties of BFGS if the curvature is negative along a given direction. Another issue we are going to touch is the boundedness of $H$. This problem does not seem to give rise to wide attention in the field of constrained optimization. We will review what people have done recently about this and then discuss it in detail more later.

In the rest of this paper, some existing methods are reviewed first. Then a few new approaches are proposed. We are going to restrict ourselves to full Hessian updating.

**2. Existing Mechanisms.** Roughly three kinds of remedy can be categorized for maintaining $H > 0$ in the literature and one for the boundedness from above. They are reviewed as follows.

**(i). Natural Skipping.** The first is that $H$ simply is kept unchanged until the curvature becomes positive. This seems reasonable because as far as global convergence is concerned, $H$ could be any positive definite matrix. It also preserves the established convergence rate of a given algorithm under some mild conditions. To make this clear, let us assume that the underlying algorithm generates an infinite sequence which converges to a local minimizer that satisfies the second order sufficiency conditions for any given bounded positive definite matrix $H$. Let $H = I$, the identity matrix at the beginning. The worst case is that $H \equiv I$ until the sequence has reached the neighborhood of a local solution. Therefore the algorithm works just like generating a steepest descent direction. When the neighborhood is approached (thus the curvature will be presumably positive along the given direction), $H$ is updated and fast local convergence will be recovered. A "good" scheme to determine if the neighborhood is reached is essential to obtain fast convergence. But we have no loss of global convergence at all even we delay the updating of $H$ because the steepest descent direction method is globally convergent. This justifies the procedure that $H$ is updated whenever the curvature is positive. A slightly related issue is to stablize an algorithm with fast local convergence rate. Essentially it combines together a first order globally convergent algorithm and an algorithm of fast local convergence. The key point is still to devise a scheme to determine if the neighborhood of some solution has been reached. Usually in nonlinear constrained optimization, faster convergence than that of steepest descent methods can be expected even during early iterations. A simple skipping does not provide us satisfactory result. The following two methods deal with this issue in order to obtain fast local convergence rate and to preserve global convergence at the same time.

**(ii). Powell's method.** The second one is Powell's modification [6]. Powell's formula proves very encouraging in theory [2] and mostly gives superlinear convergence in practice [7] [8]. But it is really not clear how it helps in principle when the curvature is negative except for making $\bar{H}$ positive definite. In his modification, Powell defines a new vector

$$\gamma_0 = \theta\eta + (1-\theta)Hd \tag{5}$$

with $\eta$ as in (3) and replace $\eta$ by $\gamma_0$ in (2). The parameter $\theta \in (0, 1]$ is given by

$$\theta = \begin{cases} 1 & \eta^T d \geq 0.2\langle d, Hd\rangle \\ \frac{0.8\langle d.Hd\rangle}{\langle d,Hd\rangle - \eta^T d} & \text{otherwise} \end{cases}$$

where 0.2 could be any positive number which gives sufficient positiveness. A little thought reveals that this formula can be viewed as a special result of the following quadratic problem

$$\begin{array}{ll} \min\limits_{\theta \in R^1} & \frac{1}{2}(1-\theta)^2 \\ \text{s.t.} & \theta \langle \nabla_x L(\bar{x}, \lambda), d \rangle - \langle \nabla_x L(x, \lambda), d \rangle \geq \epsilon \end{array} \tag{6}$$

where $\epsilon$ is prescribed somehow to ensure a positive definite $\bar{H}$. $\epsilon = 0.2 d^T H d$ in Powell's formula. Therefore $\gamma_0$ is kind of the closest vector to $\eta$ which gives $\gamma_0^T d > 0$. He could establish $r$-superlinear convergence of the variables, if it converges, under standard assumptions and some additional assumption on the Hessian matrices. Yet there is no local convergence established. This is the main theoretical drawback. The significance is that he provided the basis for not using any penalty functions (e.g., augmented Lagrange function method which will be addressed shortly) in order to obtain superlinear convergence when the second derivative matrix of the Lagrange function has some negative eigenvalues. This avoids a lot of practical difficulties. Powell even conjectured that his approach to maintain positive definite Hessian matrices even far away from a solution is actually $q$-superlinear convergence. It would only be a matter of proving it in his opinion. The simplicity of this method and satisfactory numerical results make it almost dominate the field of variable metric methods for constrained optimization. It is only recently challenged by the third method described below. We will come back to this method in next section to motivate some new practical methods. This is because we can find many other closest vectors in many different senses which achieve the same goal.

**(iii). Augmented Lagrange Method.** This method is best understood with only equality constraints. We define a new augmented Lagrange function as

$$L_a(x, \mu, \rho) = L(x, \lambda) + \frac{1}{2}\rho \sum_1^{m'} g_i^2(x)$$

with $L(x, \lambda)$ defined in (1). Its Hessian matrix with respect to $x$ is given by

$$\begin{aligned} H(x, \mu, \rho) &= \nabla_{xx} L_a(x, \lambda) = \nabla_{xx} f(x) + \textstyle\sum_1^{m'} (\mu_i + \rho g_i) \nabla_{xx} g_i(x) + \rho \nabla_x g(x) \nabla_x g(x)^T \\ &= \nabla_{xx} L(x, \mu, \rho) + \rho \nabla_x g(x) \nabla_x g(x)^T. \end{aligned} \tag{7}$$

It is well known that under the second order sufficiency conditions for any augmentation parameter $\rho$ greater than a threshold value $\bar{\rho}$, $H(x, \mu, \rho)$ is positive definite at a local solution of $(P)$. Therefore if we define

$$\eta_1 = \nabla_x L_a(\bar{x}, \mu, \rho) - \nabla_x L_a(x, \mu, \rho) \tag{8}$$

instead of (3), we can guarantee that near a solution the curvature of the augmented Lagrange function is positive definite for sufficiently large $\rho$. Then it makes sense to preserve the positive definiteness of $H$ even when the curvature is negative during early iterations. But a serious practical problem is that $\bar{\rho}$ is not known *a priori* for a given problem and it changes for different problems. Also too large a $\rho$ may give rise to some numerical difficulties as discussed by Tapia [9] and Nocedal and Overton [3]. It may happen that far away from a solution, it is impossible to make $\eta_1^T d > 0$ no matter how big $\rho$ is chosen. This method recently has been significantly improved toward practical implementation by Tapia [10], Byrd and Tapia and Zhang [11]. There is no question to date about global convergence. The principal issue here is still how to choose $\rho$. Choosing proper $\rho$ to achieve faster local rate of convergence is the challenge. Tapia proved the corresponding SQP with augmented Lagrange function gave local and $q$-superlinear convergence in the variable under fairly standard assumptions and the assumption that the augmentation parameter $\rho$ was greater than a threshold value $\bar{\rho}$. Byrd et al. investigated further the methods proposed by Tapia and developed some effective guidelines and heuristics for choosing the augmentation parameter in Tapia's BFGS structured augmented Lagrange secant algorithm (SALSA). This choice produces globally an $\eta_1$ such that the positive definiteness of the appropriate Hessian matrices will be maintained. They also established strong theoretical results concerning local rate of convergence. Specifically under standard assumptions and no assumption on the initial approximation to the Hessian matrix, they showed that if SALSA converges, then the convergence in $x$ is $r$-superlinear. This result is stronger than that obtained by Powell and Tapia since the latter require the assumption on the initial guess of the Hessian matrix. If a value for the augmentation parameter happens to be picked up that is greater than the threshold value, Byrd et al. could prove $x$ will converge $q$-superlinearly, if it converges, under standard assumptions. They performed an extensive test on their implementation of SALSA and numerical results are very encouraging and comparable with (they believe even superior to) the results obtained with Powell's method. They also claimed based on their numerical experiences that SALSA is superior to Powell's in terms of robustness as measured by the number of irregular terminations. To sum up, they actually achieved some profound theoretical progress in establishing the foreground for maintaining the positive definiteness, and at the same time obtaining satisfactory local convergence under certain context. They provide a strong alternative for Powell's method.

**(iv). Boundedness From Above.** This notion appears in [12] and [3], used in [13], for specific choices of $\eta$ and $d$ in the context of projected Hessian updating, and probably is not widely used (even noticed) by other people. The idea is as follows. It is not unusual that $\|d\|$ is very small while $\|\eta\|$ is not. If this is the case, BFGS formula (2) becomes very large in norm since we have

$$\det(\bar{H}) = \det(H)\frac{\eta^T d}{d^T H d}.$$

Even Powell's modification or SALSA does not help as we may have $\eta^T d \gg d^T H d$. Coleman and Conn try to ensure $\|d\|$ is not very small compared to $\|\eta\|$ while Nocedal and Overton simply suggest that $\bar{H} \equiv H$ (skipping!) if $\|d\|$ is too small. They can still exhibit the superlinear convergence. The impact of this scheme on the local convergence in the context of full Hessian updating still needs investigating (hopefully their proof could be carried over to full Hessian updating.). It can be induced that this is likely to happen in full Hessian updating only if the active set changes dramatically since all functions involved here are assumed smooth, thus in turn it happens only during early iterations. We certainly do not expect too big $H$ during early iterations as it may result in a search direction with too small norm, yet still far away from a solution. It may then tremendously slow down the global speed of convergence. There are not many reported results which indicate clearly how this notion may contribute to the general theory of convergence. Yet it is a problem from both theoretical and computational points of view. We also observed occasionally that poor convergence happens when the Hessian is too large in norm during early iteration in the case that active set of constraints at the solution is largely different from the set during early direction, as we expected earlier. In what follows, we will try to take this into account in our proposed methods.

**3. New Approaches.** For simplicity we consider only inequality constraints from now on. Let's define the tangential subspace

$$\mathcal{S}(x) = \{d : \langle \nabla g_i(x), d \rangle = 0, \ \forall\, i \in I(x) \ \&\ \lambda_i > 0\}.$$

Then the second order sufficiency conditions associated with a local minimizer $x^*$ are

$$\begin{aligned} &\nabla_x L(x^*, \lambda^*) = 0 \\ &\langle d, \nabla_{xx} L(x^*, \lambda^*) d \rangle > 0, \ \forall\, d \in \mathcal{S}(x^*) \\ &\lambda_i^*(x^*) \geq 0 \ \ \forall\, i = 1, \ldots, m. \end{aligned} \tag{9}$$

The fact that solving $(M)$ will provide us the correct set $I(x^*)$ near a solution $x^*$ makes it plausible to guess the search direction $d$ will eventually approach $\mathcal{S}(x^*)$ in the neighborhood of $x^*$ if $x$ converges to $x^*$. In other words, equation (9) will be probably satisfied at $x^*$ for the $d$ defined by $(M)$ in a long run. This tells us that we would be allowed to define any $\eta$ in (2) such that (4) will be achieved. It is implicitly required that $\eta$ have to asymptotically approach (3) if $x$ is a strong minimizer. Powell's method is probably the simplest one. But if the curvature is negative during early iterations, it has little to do with equation (4). We would like to propose a few alternatives in the sense that we try to force the approximation to the curvature on the tangential subspace $\mathcal{S}(x)$ as early as possible. We will see that all these new methods are associated with quadratic programming. In the following, we will need $\epsilon_1 > 0$ to define sufficient positive definiteness and $\epsilon_2 > 0$ to bound $H$ from above and denote by $\gamma_i$ the new vectors which replace $\eta$ in (2).

**(i).** The first approach is defined by the solution of the following minimization

$$\begin{array}{ll} \min\limits_{\theta \in R^1} & \frac{1}{2}(1-\theta)^2 \\ \text{s.t.} & \epsilon_2 \geq \theta\langle \nabla_x L(\bar{x}, \lambda), d\rangle - \langle \nabla_x L(x, \lambda), d\rangle \geq \epsilon_1 \end{array} \tag{10}$$

This is simply an extension of Powell's method (c.f. (6)). The best is that $\theta = 1$ and the constraint lies strictly in between the bounds. But usually we may achieve either the upper bound or the lower bound of the constraint because of the monotonicity in $\theta$. We still define

$$\gamma_0 = \theta\eta + (1-\theta)Hd. \tag{11}$$

It is easy to obtain the following expression

$$\theta = \begin{cases} \frac{d^T Hd - \epsilon_1}{d^T Hd - \eta^T d} & \text{if } \ \eta^T d < \epsilon_1 \\ \quad 1 & \text{if } \ \epsilon_2 \geq \eta^T d \geq \epsilon_1 \\ \frac{\epsilon_2 - d^T Hd}{\eta^T d - d^T Hd} & \text{if } \ \epsilon_2 \leq \eta^T d \end{cases} \tag{12}$$

which is as simple as Powell's original version. Then $\theta \in (0,\ 1)$ if $\epsilon_1 \leq d^T Hd$ and $\epsilon_2 \geq d^T Hd$ which are reasonable requirements since eventually we need to have (4). In a practical problem where we met difficulty, the number of constraints is about 100 and the number of variables is 3. The number of active constraints at a solution is less than 10. Functions involved are very nonlinear. We chose $\epsilon_2 = 5d^T Hd$ and drastic improvement was obtained. We had a hard time even to converge without bounding $H$ above. And this choice has no significant impact on the set of tested results we obtained in [8].

**(ii).** The second one is to obtain a better estimate of the Lagrange multipliers based on $\bar{x}$, and subject to some constraints. Specifically we solve

$$\begin{array}{ll} \min\limits_{\bar{\lambda} \in R^m} & \frac{1}{2}\|\nabla_x L(\bar{x}, \bar{\lambda})\|^2 \\ \text{s.t.} & \bar{\lambda}_i \geq 0 \ \ \forall\, i = 1, \ldots, m \\ & \epsilon_2 \geq \langle \nabla_x L(\bar{x}, \bar{\lambda}) - \nabla_x L(x, \lambda), d\rangle \geq \epsilon_1. \end{array} \tag{13}$$

where $d$ and $\lambda$ are the solution from $(M)$. The QP can be always solved since we could let $\bar{\lambda} = \theta\lambda$ as in (6). And the answer then would be $\bar{\lambda} = \theta = 0$. This also indicates the big difference of this method and Powell's. We then define

$$\gamma_1 = \nabla_x L(\bar{x}, \bar{\lambda}) - \nabla_x L(x, \lambda). \tag{14}$$

This can be considered as a generalization of some scheme, say in [9], which updates the Lagrange multiplier vector based on $\bar{x}$. Potentially the solution is not unique since the coefficient matrix of $\bar{\lambda}$ is not full rank in general. In this case we need to choose the solution of minimum Euclidean norm. Here we not only try to find a least square solution for the multipliers but also let it subject to some constraints which we hope to achieve near the solution. Adding $\bar{\lambda} \geq 0$ is just due to intuition. We can expect better result than just from least square solution if $\epsilon_1$ and $\epsilon_2$ are properly chosen. It seems more reasonable that $\nabla_x L(x, \lambda)$ should be replaced by $\nabla_x L(x, \bar{\lambda})$ in (13). But then the solvability of the QP is in question. Also the resulting multiplier vector may deviate too far away from the original one. This may imply that the active set of constraints are quite different from what we get from solving $(M)$, which we want to avoid. There is no simple expression for $\bar{\lambda}$ but the QP should not be expensive. If the number of constraints is too large, we could delete the constraint $\bar{\lambda} \geq 0$.

**(iii).** The third one is given by

$$\begin{array}{ll} \min\limits_{y \in R^n} & \frac{1}{2}\|y - \nabla_x L(\bar{x}, \lambda)\|^2 \\ \text{s.t.} & \epsilon_2 \geq \langle y - \nabla_x L(x, \lambda), d \rangle \geq \epsilon_1. \end{array} \tag{15}$$

Clearly, we can define

$$\gamma_2 = y - \nabla_x L(x, \lambda).$$

Note that $y - \nabla_x L(\bar{x}, \lambda) = \gamma_2 - \eta$. Now we are trying to find a closest vector to $\nabla_x L(\bar{x}, \lambda)$ such that $\bar{H} > 0$. In other words, we want to keep as much new information as possible. The keen reader may notice the similarity between formula (15) and (10). In fact (15) will become (10) if we set $y = \theta \nabla_x L(\bar{x}, \lambda)$. But geometrically (15) considers phase changing as well as magnitude changing, as opposed to (6) or (10) which only shrinks $\nabla_x L(\bar{x}, \lambda)$ in the given direction. This QP can also be formed as a more restricted minimization by

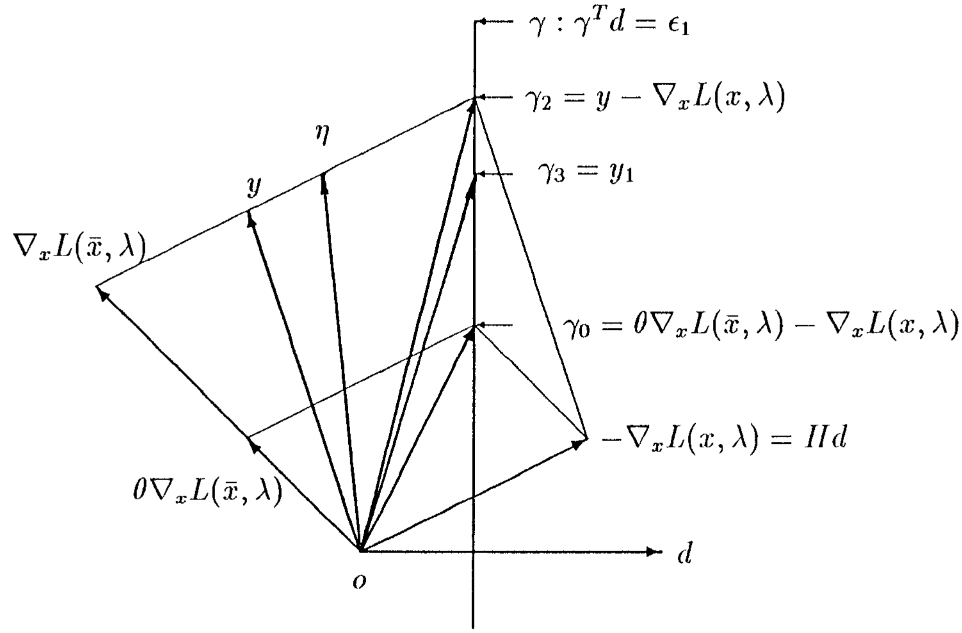


Figure 1: Graphical illustration for (i) and (iii)

making $y = \nabla_x L(\bar{x}, \bar{\lambda})$ and doing the minimization over $\bar{\lambda} \in R^m$. Again it goes back to (10) if $\bar{\lambda} = \theta\lambda$. It is obvious that $y$ or $\nabla_x L(\bar{x}, \bar{\lambda})$ approximates $\nabla_x L(\bar{x}, \lambda)$ better in both phase and magnitude than $\nabla_x L(\bar{x}, \theta\lambda)$ does, measured by the Euclidean distance. Hopefully formula (15) will give better approximation than (10). The graph in Figure 1 gives us better idea about this. We only depict the case that lower bound is to be satisfied. It is similar for the upper bound case. Definitely $y$ contains more new information than $\theta \nabla_x L(\bar{x}, \lambda)$.

**(iv).** Probably the following one gives the formula most significantly different from above ones:

$$\begin{array}{ll} \min\limits_{y_1 \in R^n} & \frac{1}{2}\|y_1 - \eta\|^2 \\ \text{s.t.} & \epsilon_2 \geq \langle y_1, d \rangle \geq \epsilon_1. \end{array} \tag{16}$$

It is more appropriate to let $y_1 = \nabla_x L(\bar{x}, \mu) - \nabla_x L(x, \mu)$ and require $\mu_i \geq 0 \ \forall\, i = 1, \ldots, m$. But we need to investigate the solvability. This QP looks the same as (15). Careful check reveals that they are entirely different from geometric point of view. It is clear that we can define

$$\gamma_3 = y_1$$

which is also marked in Figure 1. It seems from the depict this choice would be the best compared with (i) and (iii), since it is $\gamma$ or $\eta$ not $y$ or $y_1$ that directly influences the approximation of Hessian matrix. Intuitively we should approximate $\eta$. And also we can observe that Powell's modification makes the least use of new information at $\bar{x}$. It is easy to check that the solution $y_1$ actually is the projection of $\eta$ to the plane defined by $\{z : \ z^T d = \epsilon_i\}$. Consequently we have the following formula

$$y_1 = \begin{cases} \eta + \frac{\epsilon_1 - \eta^T d}{d^T d} d & \text{if } \ \eta^T d < \epsilon_1 \\ \eta & \text{if } \ \epsilon_2 \geq \eta^T d \geq \epsilon_1 \\ \eta - \frac{\eta^T d - \epsilon_2}{d^T d} d & \text{if } \ \epsilon_2 \leq \eta^T d \end{cases} \tag{17}$$

which is still easy to implement.

Note that in above QP's, we are bounding $H$ from above as well as from below. The concern of $\eta^T d \gg \langle d, Hd \rangle$ can be relaxed to some extent. This would be very significant during early iterations. We would like to point out that the parameters $\epsilon_i$ need not be constant throughout the optimization process. They could be some sequences which play more role during early iterations than near a solution. And the choice of $\epsilon_2$ is critical since a small value may impede the superlinear convergence to take place. But so far we cannot say more about this since no numerical testing has been done yet.

**4. Remarks.** We reviewed and proposed a few alternatives for maintaining the positive definiteness of Hessian matrix in the variable metric constrained optimization. All of them are based on heuristics and intuition. What is lacking are theoretical analysis and numerical experiments on these new methods. It would be worth investigating along these lines, and their implementation should not be difficult and also is cheap. Hopefully the numerical testing will be done in the near future.

**5. Acknowledgement.** The author is indebted to his advisor Professor André Tits for invaluable discussion and critic suggestion which lead to definite improvement of this note.

## References

[1] S.P. Han, "Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems," *Mathematical Programming* 11 (1976), 263–282.

[2] M.J.D. Powell, "The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations," in *Nonlinear Programming 3*, O.L. Mangasarian, R.R. Meyer & S.M. Robinson, eds., Academic Press, New York, 1978, 27–63.

[3] J. Nocedal & M.L. Overton, "Projected Hessian Updating Algorithms for Nonlinearly Constrained Optimization," *SIAM J. Numer. Anal.* 22 (October, 1985).

[4] R.H. Byrd & J. Nocedal, "An Analysis of Reduced Hessian Methods for Constrained Optimization," Boulder, University of Colorado, Technical Report CU-CS-389-88, August 1988.

[5] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass., 1973.

[6] M.J.D. Powell, "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," in *Numerical Analysis, Dundee, 1977, Lecture Notes in Mathematics 630*, G.A. Watson, ed., Springer-Verlag, 1978, 144–157.

[7] W. Hock & K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems (187), Springer Verlag, 1981.

[8] J. Zhou & A.L. Tits, "User's Guide for FSQP Version 2.0 – A Fortran Code for Solving Optimization Problems, Possibly Minimax, with General Inequality Constraints and Linear Equality Constraints, Generating Feasible Iterates," Systems Research Center, University of Maryland, SRC TR-90-60, College Park, MD 20742, 1990.

[9] R. A. Tapia, "Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization," *J. Optim. Theory Appl.* 22 (1977), 135–194.

[10] R. A. Tapia, "On secant updates for use in general constrained optimization," *Math. Comp.* (1988).

[11] R.H. Byrd, R.A. Tapia & Yin Zhang, "An SQP Augmented Lagrangian BFGS Algorithm for Constrained Optimization," Department of Mathematical Sciences, Houston, Texas, 1989.

[12] T.F. Coleman & A.T. Conn, "On the local convergence of a quasi-Newton method for the nonlinear programming problem," *SIAM J. Numer. Anal.* 22 (October, 1985), 821–850.

[13] R. Fletcher & E. Sainz de la Maza, "Nonlinear Programming and Nonsmooth Optimization by Successive Linear Programming," *Mathematical Programming* 43 (1989), 235–256.