ON THE ROLE OF SLACK VARIABLES IN QUASI-NEWTON METHODS FOR CONSTRAINED OPTIMIZATION*

DOE/ER/05046-15

MASTER

R. A. Tapia
Department of Mathematical Sciences
Rice University
Houston, Texas

In constrained optimization the technique of converting an inequality constraint into an equality constraint by the addition of a squared slack variable is well known but rarely used. In choosing an active constraint philosophy over the slack variable approach researchers quickly justify their choice with the standard criticisms: The slack variable approach increases the dimension of the problem, is numerically unstable and gives rise to singular systems.

In this paper we show that these criticisms of the slack variable approach need not apply and the two seemingly distinct approaches are actually very closely related. In fact, the squared slack variable formulation can be used to develop a superior and more complete active constraint philosophy.

INTRODUCTION AND PRELIMINARIES

The background material on quasi-Newton methods for constrained optimization is taken from the author's papers, Tapia (1974a),(1974b),(1977) and (1978), and these papers will be referred to often. We first consider three more or less standard approaches for applying quasi-Newton methods to equality constrained optimization.

By a quasi-Newton method for approximating a stationary point x^* of $f:R^n \to R$ we mean the iterative procedure

(1) $$\bar{x} = x - B^{-1}\nabla f(x)$$

(2) $$\bar{B} = \mathcal{B}(x,\bar{x},B)$$

where $\mathcal{B}(x,\bar{x},B)$ is in some sense an approximation to $\nabla^2 f(x^*)$. As examples of quasi-Newton methods we have

Newton's Method:

(3) $$\mathcal{B}(x,\bar{x},B) = \nabla^2 f(\bar{x}) .$$

Discrete Newton's Method:

(4) $$\mathcal{B}(x,\bar{x},B) = \left(\frac{1}{h}\left[\frac{\partial f(\bar{x}+he_j)}{\partial x_i} - \frac{\partial f(\bar{x})}{\partial x_i}\right]\right)$$

*This work was supported by the United States Department of Energy under Contract EY-76-S-05-5046.

where $e_1,\ldots,e_n$ are the natural basis vectors for R^n, h is a small positive scalar (ideally somewhere near the square root of the machine tolerance of the particular computer system being used) and (a_{ij}) denotes the matrix whose i,j-th component is a_{ij}.

Secant Methods:

(5) $$\mathcal{B}(x,\bar{x},B) = \mathcal{B}_S(s,y,B)$$

where $s = \bar{x} - x$, $y = \nabla f(\bar{x}) - \nabla f(x)$ and $\mathcal{B}_S$ satisfies the secant equation

(6) $$\mathcal{B}_S(s,y,B)s = y .$$

Several of the more popular secant updates (choices of $\mathcal{B}_S$) are the so-called Broyden, PSB, DFP and BFGS secant updates and can be found (along with their inverse updates) in Tapia (1977) and in greater detail in Dennis and Moré (1977).

Now, let us consider the equality constrained optimization problem

(7) $$\text{minimize } f(x); \quad \text{subject to} \quad g(x) = 0$$

where $f:R^n \to R$ and $g:R^n \to R^m$ $(m \leq n)$. Corresponding to problem (1) we have the Lagrangian

(8) $$L(x,\lambda) = f(x) + \lambda^T g(x)$$

Observe that $L:R^n \times R^m \to R$. The well-known classical Lagrange multiplier theory says that (under mild conditions) a necessary condition that x^* solves problem (7) is that there exists a corresponding multiplier λ^* such that (x^*,λ^*) is a solution of

(9) $$\nabla_x L(x,\lambda) = 0 \quad \text{and} \quad g(x) = 0 .$$

Notice that (9) can be written

(10) $$\nabla L(x,\lambda) = 0 .$$

In the following, we use x^* to denote a solution of problem (7) and λ^* to denote the corresponding multiplier, i.e.,

(11) $$\nabla L(x^*,\lambda^*) = 0 .$$

By the extended problem corresponding to problem (7) we mean problem (10). The motivation for the use of the terminology extended should be clear from the fact that the dimension of problem (10) is actually $n+m$.

For the sake of simplicity, in this paper we have elected to work with the Lagrangian instead of the augmented Lagrangian. All our comments and results apply equally well to formulations which use the augmented Lagrangian. However, a key point of Tapia (1977) (see in particular, Corollary 7.1 and Theorem (10.3)) is that the augmented Lagrangian may offer little if anything over the standard Lagrangian formulations. See also Bertocchi, Cavalli and Spedicato (1979).

The Multiplier Extension Quasi-Newton Methods:

By a multiplier extension quasi-Newton method for problem (7) we mean the iterative procedure

(12) $$\begin{pmatrix}\bar{x}\\ \bar{\lambda}\end{pmatrix} = \begin{pmatrix}x\\ \lambda\end{pmatrix} - B^{-1}\nabla L(x,\lambda)$$

(13) $$\bar{B} = \begin{bmatrix}\bar{B}_x & \nabla g(\bar{x})\\ \nabla g(\bar{x})^T & 0\end{bmatrix}$$

where $\bar{B}_x$ is an approximation to $\nabla_x^2 L(x^*,\lambda^*)$.

The multiplier extension secant methods result by choosing

(14) $$\bar{B}_x = \mathcal{B}_S(s,y,B_x)$$

where $s = \bar{x} - x$, $y = \nabla_x L(\bar{x},\bar{\lambda}) - \nabla_x L(x,\bar{\lambda})$, B_x is the current approximation to $\nabla_x^2 L(x^*,\lambda^*)$ and $\mathcal{B}_S$ is one of the popular secant updates. The multiplier extension secant method played an important role in the theory developed in Tapia (1977) and (at present) we have no references to earlier usage.

The Multiplier Update Quasi-Newton Methods:

By a multiplier update quasi-Newton method for problem (7), we mean the iterative procedure

(15) $$\bar{\lambda} = (\nabla g^T B^{-1} \nabla g)^{-1}(g - \nabla g^T B^{-1} \nabla f)$$

(16) $$\bar{x} = x - B^{-1}\nabla_x L(x,\bar{\lambda})$$

(17) $$\bar{B} = \mathcal{B}(x,\bar{x},\lambda,\bar{\lambda},B)$$

where $\mathcal{B}(x,\bar{x},\lambda,\bar{\lambda},B)$ is an approximation to $\nabla_x^2 L(x^*,\lambda^*)$. The multiplier update secant methods result by choosing

(18) $$\mathcal{B}(x,\bar{x},\lambda,\bar{\lambda},B) = B_S(s,y,B)$$

where $s = \bar{x} - x$, $y = \nabla_x L(\bar{x},\bar{\lambda}) - \nabla_x L(x,\bar{\lambda})$ and B_S is one of the popular secant updates.

The multiplier update Newton method was proposed by the author in Tapia (1974a) and extended to inequality constraints in Tapia (1974b). The multiplier update secant methods were proposed by the author in Tapia (1977) and in that paper inequality constraints were handled via a slack variable. Independently, Han (1977) proposed secant methods for problems with equality and inequality constraints which use an intermediate quadratic program to solve for the multipliers. In the case of problem (7) (no inequality constraints) it is a simple matter to show that Han's quadratic program reduces to (15) and hence his algorithm reduces to the multiplier update secant method. Glad (1976) also independently, proposed the multiplier update secant method. He used an active constraint philosophy to handle inequality constraints. All three papers established superlinear convergence.

The Quadratic Programming Quasi-Newton Methods:

By a quadratic programming quasi-Newton method for problem (7) we mean the iterative procedure

(19) $$\bar{x} = x + \Delta x$$

(20) $$\bar{B} = \mathcal{B}(x,\bar{x},B)$$

where $\mathcal{B}(x,\bar{x},B)$ is an approximation to $\nabla_x^2 L(x^*,\lambda^*)$ and Δx is a solution of the quadratic program

(21) $$\min_{\Delta x} q(\Delta x) = f(x) + \langle \nabla f(x), \Delta x\rangle + \tfrac{1}{2}\langle B\Delta x, \Delta x\rangle$$

$$\text{subject to } \nabla g(x)^T \Delta x + g(x) = 0 .$$

The quadratic programming secant methods result by choosing

(22) $$\mathcal{B}(x,\bar{x},B) = \mathcal{B}_S(s,y,B)$$

where $s = \bar{x} - x$, $y = \nabla_x L(\bar{x},\lambda_{QP}) - \nabla_x L(x,\lambda_{QP})$, $\mathcal{B}_S$ is one of the popular secant updates and λ_{QP} is the multiplier obtained in the solution of the quadratic program (21).

This form of the quadratic programming quasi-Newton method was introduced by Garcia Palomares and Mangasarian (1976). Han (1976) added some analysis and specific secant updates and Powell (1972),(1978) added further refinements and analysis.

When they were first proposed, many thought that these three approaches were distinct. However, as the following theorem from Tapia (1978) shows, they are equivalent.

Theorem 1. The multiplier extension secant method, the multiplier update secant method, and the quadratic programming secant method generate identical (x,λ) iterates and are locally Q-superlinearly convergent.

Proof. For the proof of this theorem and other details see Tapia (1977),(1978).

INEQUALITY CONSTRAINTS.

Consider the extension of problem (7) given by

(23) $$\text{minimize } f(x), \text{ subject to } \begin{array}{ll} g_i(x) = 0 & i = 1,\dots,m \\ g_i(x) \le 0 & i = m+1,\dots,p . \end{array}$$

where

$$f, g_i : R^n \to R^1 .$$

For $x \in R^n$, let

$$B(x) = \{i : 1 \le i \le m \text{ or } g_i(x) = 0,\ i = m+1,\dots,p\}$$

In the case of problem (7) we let $B(x) = \{1,\dots,m\}$. A point $x \in R^n$ is said to be a regular point of problem (7) or problem (23) if $\{\nabla g_i(x) : i \in B(x)\}$ is a linearly independent set. Well known necessary conditions for problem (23) are that, if the regular point x^* is a solution of problem (23), then there exist Lagrange multipliers $\lambda^* = (\lambda_i^*)$, $i = 1,\dots,p$, such that (x^*,λ^*) is a solution of the following system of equations and inequalities:

$$
(24)\qquad
\begin{aligned}
\nabla f(x) + \lambda_1 \nabla g_1(x) + \ldots + \lambda_p \nabla g_p(x) &= 0, \\
\lambda_i g_i(x) &= 0, \quad i = m+1,\ldots,p \\
g_i(x) &= 0, \quad i = 1,\ldots,m \\
g_i(x) &\le 0, \quad i = m+1,\ldots,p \\
\lambda_i &\ge 0, \quad i = m+1,\ldots,p
\end{aligned}
$$

It is not obvious how one extends the multiplier extension and the multiplier update philosophies to handle inequality constraints. Clearly, one way of including inequality constraints in the quadratic programming philosophy is obvious. Specifically one merely carries them along as linearized inequalities in the quadratic program. On the surface this seems to be an advantage of the quadratic programming approach. However, mathematically the situation is not well-defined since the theory for handling the inequalities will now depend on the particular quadratic programming code employed in the implementation and can vary significantly. In many ways this approach "sweeps the dirt under the rug". Moreover, Chamberlain (1978) recently demonstrated that this approach can lead to cycling.

Active Constraint Philosophy:

The active constraint philosophy consists of ignoring certain inequality constraints and treating the remaining inequality constraints as equality constraints at each stage of the iterative process. In its purest form one merely ignores inequality constraints which are satisfied and treats inequality constraints which are violated as equality constraints.

Locally (or perhaps better said, asymptotically), the active constraint philosophy is optimal. If one merely knew which constraints were active or binding at the solution, then the inequality constrained problem could be handled as an equality constrained problem with the minimal number of constraints. Indeed, the knowledge that a particular constraint is active at the solution can be expected only when the iterates are near the solution. Far from the solution the information that a constraint is either satisfied or violated would be a poor indicator of the properties of this constraint at the solution.

One obvious problem with the active constraint approach is that we can never handle more than n (dimension of x) equality constraints. Hence, it is not clear what one should do when the number of original equality constraints plus the number of violated constraints is greater than n. In this case one must consider notions analogous to "most violated". Moreover, since the iterates generated by the three quasi-Newton methods described above satisfy linear constraints and the empirical fact that linear approximations to nonlinear inequality constraints can be very misleading, we would expect the active constraint philosophy in its purest form to not be immune to cycling.

In summary, we feel that what is needed is a conservative active constraint philosophy. Namely, a strategy which allows for a certain amount of indecisiveness far from the solution, i.e., doesn't force the drastic choice of either ignore or treat as an equality. It is the objective of this paper to argue that the squared slack variable approach which we now describe, can lead to such a strategy.

Squared Slack Variable Philosophy:

If we introduce the slack variables $y_{m+1},\ldots,y_p$ and define $F,G_i:R^{n+p-m}\to R$ by

$$F(x,y) = f(x),$$
$$G_i(x,y) = g_i(x), \qquad i = 1,\ldots,m$$
$$G_i(x,y) = \tfrac{1}{2}y_i^2 + g_i(x), \quad i = m+1,\ldots,p$$

then we may consider the following equality constrained optimization problem:

(25) $$\text{minimize } F(x,y), \quad \text{subject to } G_i(x,y) = 0, \quad i = 1,\ldots,p .$$

The following proposition is not difficult to prove and justifies the use of squared slack variables.

<u>Proposition 1</u>. Suppose $x \in R^n$. Then

(i) x solves problem (23) $\Leftrightarrow$ (x,y) solves problem (25);

(ii) x is a regular solution of problem (23) $\Leftrightarrow$ (x,y) is a regular solution of problem (25);

(iii) x satisfies (24) the necessary conditions for problem (23) $\Leftrightarrow$ (x,y) satisfies the necessary conditions for problem (25) (see (9)), (with $\lambda_i \geq 0$, $i = m+1,\ldots,p$).

In the above proposition, when x is a solution of problem (23) the corresponding y should be interpreted as $y_i = \sqrt{-2g_i(x)}$.

The squared slack variable approach suggested above is rarely used by workers in optimization theory. Its use seems to be restricted to some statisticians and engineers. The idea of replacing a variable which is constrained to be nonnegative by a squared variable is quite common in statistics (see for example, Chapter 4 and Appendix II.2 of Tapia and Thompson (1978)). At any rate, most workers in optimization theory quickly reject the squared slack variable approach with one or more of the following standard criticisms:

(1) Squared slack veriables increase the dimension of the problem, i.e., the dimension of the linear systems that must be solved;

(2) Squared slack variables are less stable than nonsquared slack variables (see Robinson (1976));

(3) Squared slack variables lead to (asymptotic) singularities and in particular singular Hessians.

We will now investigate these three criticisms in the light of our three quasi-Newton methods. We show that, contrary to some authors' biases, the multiplier update quasi-Newton method can be implemented so that it does not suffer from any of the above three criticisms and also leads to a promising active constraint philosophy.

The following qualifications are extremely important. Theorem 1 says that our three approaches are mathematically equivalent in the sense that they produce identical iterates. However, there may be significant differences from a numerical and practical point of view. Hence, in a particular application, if one uses a straightforward implementation one approach may have advantages over another. Of course it could be argued that this straightforward implementation was naive and could be modified so as to take advantage of any positive aspect of one of the other formulations. While this is mathematically true, it is exactly this point with which we are concerned. Namely, the modification could well be of such a nature as to essentially produce the equivalent formulation. Hence it is the straightforward implementation that we are concerned with and it is in this context that we will argue that the use of squared slack variables is natural for the multiplier update quasi-Newton method and leads to an effective active constraint philosophy.

We first establish the following notation

(26) $$y = (0,\ldots,0,y_{m+1},\ldots,y_p)^T,$$

(27) $$\lambda = (\lambda_1,\ldots,\lambda_m,\lambda_{m+1},\ldots,\lambda_p)^T,$$

(28) $$Y = \mathrm{diag}(0,\ldots,0,y_{m+1},\ldots,y_p)^T,$$

(29) $$\Lambda = \mathrm{diag}(1,\ldots,1,\lambda_{m+1},\ldots,\lambda_p)^T,$$

(30) $$\ell(x,\lambda) = f(x) + \lambda^T g(x) \ ,$$

and

(31) $$L(x,y,\lambda) = F(x,y) + \lambda^T G(x,y) \ ,$$

where f, F, g and G are as in (23) and (25). We will also partition the vector $u \in R^p$ and the $p \times p$ diagonal matrix $U = \mathrm{diag}(u)$ into a part corresponding to the equality constraints and a part corresponding to the inequality constraints. These parts will be subscripted with E and I respectively, i.e.,

$$u_E = (u_1,\ldots,u_m)^T \ ,$$

$$u_I = (u_{m+1},\ldots,u_p)^T \ ,$$

$$U_E = \mathrm{diag}(u_1,\ldots,u_m),$$

and

$$U_I = \mathrm{diag}(u_{m+1},\ldots,u_p) \ .$$

It is a straightforward matter to show that

(32) $$\nabla_{(x,y)} L(x,y,\lambda) = \begin{pmatrix} \nabla_x \ell(x,\lambda) \\ \Lambda_I y_I \end{pmatrix} ,$$

(33) $$\nabla_{(x,y,\lambda)} L(x,y,\lambda) = \begin{pmatrix} \nabla_x \ell(x,\lambda) \\ \Lambda_I y_I \\ G(x,y) \end{pmatrix} ,$$

(34) $$\nabla^2_{(x,y)} L(x,y,\lambda) = \begin{pmatrix} \nabla^2_x \ell(x,\lambda) & 0 \\ 0 & \Lambda_I \end{pmatrix}$$

and

(35) $$\nabla^2_{(x,y,\lambda)} L(x,y,\lambda) = \begin{pmatrix} \nabla^2_x \ell(x,\lambda) & 0 & \nabla g_E(x) & \nabla g_I(x) \\ 0 & \Lambda_I & 0 & Y_I \\ \nabla g_E(x)^T & 0 & 0 & 0 \\ \nabla g_I(x)^T & Y_I & 0 & 0 \end{pmatrix}.$$

Squared Slack Variables and the Quadratic Programming Approach:

It is not difficult to see that a straightforward application of the quadratic programming approach to problem (25) would lead to an algorithm which suffers from all three of the above criticisms. Specifically, we would be approximating the Hessian matrix given by (34) which is necessarily singular at the solution (except of course in the unlikely situation that all inequality constraints are active at the solution).

Since a linearization of the constraint $G_i(x,y) = g_i(x) + \frac{1}{2}y_i^2$ amounts to

(36) $$\nabla g_i(x)^T \Delta x_i + y_i \Delta y_i$$

we see that the slack variables will not appear in squared form only. Hence, we cannot replace y_i^2 with z_i and work with a nonsquared slack variable. Finally, the dimension of the quadratic program is increased by $p-m$, the number of slack variables.

Squared Slack Variables and the Multiplier Extension Approach:

It is not difficult to see that only the first two criticisms apply in this case. Namely, the Hessian matrix given by (35) is not singular at the solution as long as we have strict complementarity. Strict complementarity means that at the solution not both λ_i and y_i are zero. This assumption is both standard and mild. Again we see that y_i appears in nonsquared form both in (33) and (34).

Squared Slack Variables and the Multiplier Update Approach:

A straightforward application of the multiplier update quasi-Newton method to problem (25) gives the iterative procedure

(37) $$\bar{\lambda} = (\nabla g(x)^T B^{-1} \nabla g(x) + \Lambda^{-1} Y^2)^{-1} [G(x,y) - \nabla g(x)^T B^{-1} \nabla f(x)],$$

(38) $$\bar{x} = x - B^{-1} \nabla_x \ell(x,\bar{\lambda}),$$

(39) $$\bar{y}_i = (1 - \bar{\lambda}_i/\lambda_i) y_i, \qquad i = m+1,\ldots,p$$

(40) $$\bar{B} = \mathcal{B}(x,\bar{x},\bar{\lambda},B).$$

The first thing we observe is that (39) does not require the solution of a linear system, and the linear systems that must be solved in (37) and (38) are not of a larger dimension than we would have without the addition of slack variables. Hence, criticism (1) does not apply. The second thing we observe is that the slack variables appear twice in (37) but in squared form only. Also, while they do not appear in (39) in squared form only, we can obviously square both sides and obtain an equally nice expression with the slack variables in squared form only. These observations allow us to state the complete algorithm in terms of a nonsquared slack variable. It follows that neither criticism (1) nor criticism (2) will apply. Let us therefore consider criticism (3); namely, singularities. In light of the above comments we first introduce the following transformations and notation:

(41) $$z_i = \tfrac{1}{2}y_i^2 \quad , \; i = m+1,\ldots,p$$

(42) $$z = (0,\ldots,0,z_{m+1},\ldots,z_p)^T$$

(43) $$Z = \mathrm{diag}(z).$$

Clearly $\lambda_i = 0$ for any $m < i \leq p$ produces a singularity in (37) and (39). Since $\lambda_i = 0$ is a reasonable value and clearly is acceptable in the equivalent multiplier extension formulation, it must be a removable singularity. In order to obtain the correct interpretation of (37) and (39) we will look at the multiplier extension formulation. In this consideration the choice of B does not matter; so for the sake of simplicity we may as well consider the Newton formulation

(44) $$\nabla^2_{(x,y,\lambda)}L(x,y,\lambda)\begin{pmatrix}\Delta x\\ \Delta y_I\\ \Delta\lambda\end{pmatrix} = -\nabla_{(x,y,\lambda)}L(x,y,\lambda) \; .$$

The linear system (44) can be partitioned into four subsystems of equations as dictated by the form of the Hessian given in (35). We assume that $z_i \neq 0$ and $\lambda_i \neq 0$ for $i = m+1,\ldots,p$ and then look at the limiting behavior as either $\lambda_i \to 0$ or $z_i \to 0$. We can safely assume that we will never encounter the situation when both variables go to zero, since this would contradict our strict complementarity assumption. By combining the information in the second and fourth subsystems of equations in (44) we obtain for $i = m+1,\ldots,p$

(45) $$\bar{\lambda}_i/\lambda_i = (\nabla g(x)^T\Delta x + g(x) + z)_i/z_i$$

where $(u)_i$ denotes the i-th component of the vector u. From (45) we see that for $i=m+1,\ldots,p$ if $\lambda_i=0$, then $\bar{\lambda}_i = 0$ and the quantity $\bar{\lambda}_i/\lambda_i$ can be obtained from the right-hand side of (45), even in the case that $\lambda_i=0$. The second subsystem of (44) (or (39)) shows that if $z_i = 0$, then $\bar{z}_i = 0$. We are now in a position to rewrite our algorithm using the slack variable z and removing the singularities in (37) and (39). Specifically we have

(46) $$\bar{\lambda} = (\nabla g(x)^T B^{-1}\nabla g(x) + 2\Lambda^{-1}Z)^{-1}[g(x) + z - \nabla g(x)^T B^{-1}\nabla f(x)]$$

(if $\lambda_i = 0$ set $\bar{\lambda}_i = 0$ and do not include it in (46), $i=m+1,\ldots,p$)

(47) $$\bar{x} = x - B^{-1}\nabla_x \ell(x,\bar{\lambda}) \; ,$$

(48) $$\bar{z}_i = (\nabla g(x)^T\Delta x + g(x))_i^2/z_i \quad , \quad i = m+1,\ldots,p$$

(if $z_i = 0$ set $\bar{z}_i = 0$ and do not include it in (48)).

(49) $$\bar{B} = \mathcal{B}(x,\bar{x},\bar{\lambda},B) \; .$$

In the following theorem we assume the standard conditions with respect to differentiability, invertibility, regularity and strict complementarity (see Tapia (1977))

Theorem 2. Consider the multiplier update quasi-Newton method given by (46)-(49) for problem (23). Then

(i) The multiplier update Newton method is locally Q-quadratically convergent in the variables (x,z,λ);

(ii) The multiplier update discrete Newton method is locally Q-linearly convergent in the variables (x,z,λ);

(iii) The multiplier update secant method is locally Q-superlinearly convergent in the variables (x,z,λ).

Proof. The proof follows from Theorem 1 and the convergence analysis given in Tapia (1977).

Observe that (39) can be written as

(50) $$\bar{y}_i/y_i = (\lambda_i - \bar{\lambda}_i)/\lambda_i$$

or

(51) $$\bar{\lambda}_i/\lambda_i = (y_i - \bar{y}_i)/y_i \ .$$

This interesting symmetry has several important consequences.

Theorem 3. Suppose that the multiplier update quasi-Newton method (46)-(49) converges. Then we necessarily have Q-superlinear convergence to zero of the multipliers corresponding to inequality constraints which are inactive at the solution and Q-superlinear convergence to zero of the slack variables corresponding to inequality constraints which are active at the solution.

Proof. The proof follows directly from strict complementarity and the expressions (50)-(51).

Corollary 1. Suppose that the multiplier update quasi-Newton method (46)-(49) converges to (x^*,z^*,λ^*). Then for $i = m+1,\dots,p$

(52) $$\bar{\lambda}_i/\lambda_i \to \begin{cases} 1 & \text{if } g_i(x^*) = 0 \\ 0 & \text{if } g_i(x^*) < 0 \end{cases}$$

and

(53) $$\bar{z}_i/z_i \to \begin{cases} 1 & \text{if } g_i(x^*) < 0 \\ 0 & \text{if } g_i(x^*) = 0 \ . \end{cases}$$

The indicator $\bar{\lambda}_i/\lambda_i$ as given by (45) will be useful as a check on the validity of treating a particular constraint as a binding constraint or removing it from the problem.

Alternative Choices for λ and z:

Clearly in (46)-(49) choosing $z_i = 0$ corresponds to treating the i-th constraint as an equality constraint and choosing $\lambda_i = 0$ corresponds to removing the i-th constraint from the calculations. It follows that the active constraint philosophy in its purest form corresponds to the slack variable approach (46)-(49) with the alternative choices $z_i = 0$ if $g_i(x) \geq 0$ and $\lambda_i = 0$ if $g_i(x) < 0$ for $i = m+1,\dots,p$. Near the solution these choices would be optimal. Far from the solution these choices could be very poor or impossible to implement. What is needed is a mechanism for allowing some constraints to be in a third category. Specifically, this category will consist of the constraints that we feel we do not have have enough information on to decide if they are either active or inactive at the solution. It should be noted that asymptotically the squared slack variable approach coincides with the pure active constraint approach.

In (48) the alternative choice

$$\bar{z}_i = -g_i(\bar{x}) \quad \text{if} \quad g_i(\bar{x}) < 0 \tag{54}$$

should have a lot to offer. In particular, it will make the algorithm less sensitive to poor values of z_i by removing the dependence on z_i in the second expression on the right-hand side of (46).

Suppose that $g_i(x) < 0$ and we decide to remove this constraint, i.e., we decide to set $\lambda_i = 0$. Then by letting $z_i = -g_i(x)$ we can use (45) to calculate

$$\bar{\lambda}_i/\lambda_i = (\nabla g(x)^T \Delta x)_i / -g_i(x) \tag{55}$$

which can be used to check our choice. Namely, if $\bar{\lambda}_i/\lambda_i$ is not small, we should question this choice (see (52)). For constraints that have been removed it would be wise to monitor the behavior of $\bar{\lambda}_i/\lambda_i$ as given by (55) as well as the behavior of $g_i(x)$. A constraint which has been removed should be brought back into the calculation as an equality constraint $(z_i = 0)$. This will allow for a fresh calculation of the multiplier associated with this particular constraint.

Now, if $g_i(x) > 0$ or $|g_i(x)|$ is small and $\bar{\lambda}_i/\lambda_i$ is near one we should treat this constraint as an equality or binding constraint by choosing $z_i = 0$. Again we should monitor the values of $\bar{\lambda}_i/\lambda_i$ (in this case we would not need (45)) as well as those of $g_i(x)$.

The details concerning the choices in the improved active constraint approach cannot be finalized without some numerical experimentation. However, we are quite confident that these observations can be used to construct an approach which is superior to both the pure slack variable approach and the pure active constraint approach.

Summary.

In this paper we have attempted to demonstrate that the squared slack variable approach to inequality constraints need not suffer from the standard criticisms attached to it: increased dimension, numerical instability and presence of singularities. Specifically, it is these removable singularities that eventually leads to a pure active constraint approach. Moreover, the squared slack variable approach can be used to construct what we expect will be a superior active constraing philosophy. We gain some satisfaction in showing that the pure active constraint philosophy and the squared slack variable philosophy are not as dissimilar as many authors believe them to be.

References

Bertocchi, M., Cavalli, E. and Spedicato, E. (1979). Computational Performance of Diagonalized Multiplier Quasi-Newton Methods for Nonlinear Optimization with Equality Constraints, manuscript.

Chamberlain, R. M. (1978). Some Examples of Cycling in Variable Metric Methods for Constrained Minimization, Department of Applied Mathematics and Theoretical Physics, Cambridge University, England.

Dennis, J. E. and Moré, J. J. (1977). Quasi-Newton Methods: Motivation and Theory, SIAM Review, Vol. 10, pp. 46-89.

Garcia Palomares, U. M. and Mangasarian, O. L. (1976). Superlinearly Convergent Quasi-Newton Algorithms for Nonlinearly Constrained Optimization Problems, Mathematical Programming, Vol. 11, pp. 1-13.

Han , S. P. (1976). Dual Variable Metric Algorithms for Constrained Optimization, SIAM J. on Control and Optimization, Vol. 15, pp. 546-565.

Han, S. P. (1977). Dual Variable Metric Algorithms for Constrained Optimization, SIAM J. on Control and Optimization, Vol. 15, pp. 546-565.

Powell, M. J. D. (1977). A Fast Algorithm for Nonlinearly Constrained Optimization Calculations. Report DAMTP 77/HA 2, Cambridge University, England.

Powell, M. J. D. (1978). The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations, in Nonlinear Programming 3, (O. Mangasarian, R. Meyer, and S. Robinson, eds.), pp. 27-63, Academic Press, New York.

Robinson, S. M. (1976). Stability Theory for Systems of Inequalities, Part II: Differentiable Nonlinear Systems, Mathematics Research Center, University of Wisconsin, Madison.

Tapia, R. A. (1974a). Newton's Method for Optimization Problems with Equality Constraints, SIAM Journal on Numerical Analysis, Vol. 11, pp. 874-886.

Tapia, R. A. (1974b). A Stable Approach to Newton's Method for General Mathematical Programming Problems in R^n, Journal of Optimization Theory and Applications, Vol. 14, pp. 453-476.

Tapia, R. A. (1977). Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization, Journal of Optimization Theory and Applications, Vol. 22, pp. 135-194.

Tapia, R. A. (1978). Quasi-Newton Methods for Equality Constrained Optimization: Equivalence of Existing Methods and a New Implementation, in Nonlinear Programming 3, (O. Mangasarian, R. Meyer and S. Robinson, eds.) pp. 125-164, Academic Press, New York.

Tapia, R. A. and Thompson, J. R. (1978). Nonparametric Probability Density Estimation, Johns Hopkins University Press, Baltimore, Maryland.