AUGMENTED LAGRANGIAN LINE SEARCHES
FOR SUCCESSIVE QUADRATIC PROGRAMMING

by

J.E. Cuthrell *h* L.T. Biegler

December, 1983

DRC-06-43-83

Augmented Lagrangian Line Searches for Successive Quadratic Programming

by

J.E. Cuthrell and L-T. Biegler
Carnegie-Kellon University
Department of Chemical Engineering
Pittsburgh, PA 15213 *

Abstract

Successive Quadratic Programming (SQP) algorithms have been effective and efficient in solving nonlinearly constrained optimization problems. To guarantee global convergence, however, a line search must be performed after solving the quadratic program. The line search terminates when a step size is found that causes a suitable decrease in some merit function. Because of some problems with previously suggested merit functions, a line search that uses an augmented Lagrangian is proposed.

This function follows quite naturally from the derivation of SQP methods and exhibits superior convergence properties compared to an exact penalty function. First, global and local convergence results are presented which are valid for penalty parameters that allow descent directions. We then present an algorithm that chooses penalty parameters that allow good performance. Finally, a numerical study on 15 test problems is presented that compares the proposed line search to existing strategies.

1. INTRODUCTION

The nonlinear programming problem can be written as

$$\text{(1.1)} \qquad \min_{x} \; f(x) \quad \text{s. t.} \quad g(x) \le 0, \quad h(x) = 0$$

for

$$f: R^n \to R$$

$$g: R^n \to R^m$$

$$h: R^n \to R^k$$

To solve this problem, successive quadratic programming (SQP) algorithms were first proposed as the SOLVER methods of Wilson (1963) and Beale (1967). However, these require second derivatives of the constraints and objective function and initial estimates of the Kuhn-Tucker multipliers.

A significant improvement was made with the results o[illegible] Palomares and Mangasarian (1976). Here local convergence pr[illegible] proven even if the Hessian in the quadratic program is approximated by some positive definite updating scheme.

Han (1977) was able to show global convergence properties by conducting an inexact minimization on an exact penalty function, in the search direction found by the quadratic program.

The algorithm proposed by Han and later modified by Powell (1977) is the following:

1) Solve the QP:

$$Q\,(x^i, B^i) \qquad \min_d \ \nabla f^T(x^i)d + \frac{1}{2} d^T B^i d$$

$$\text{s.t.} \quad g(x^i) + \nabla g(x^i)^T d \le 0$$

$$h(x^i) + \nabla h(x^i)^T d = 0$$

where B^i is a positive definite quasi-Newton approximation to $\nabla_{xx} L(x^i, u^i, v^i)$.

Here, $L(x,u,v) = f(x) + u^T g(x) + v^T h(x)$, and the multipliers (u, v)

are found from the QP.

2) If $|\nabla f(x^i)^T d| + |u^{iT} g(x^i)| + |v^{iT} h(x^i)| \le Z$, stop. Here $Z > 0$ is some small Kuhn-Tucker tolerance.

3) Else, find a stepsize λ such that

$$P(x^i + \lambda d) \le P(x^i) + \delta\lambda P'(x^i)$$

where

(1.2) $$P(x) = f(x) + \sum_{j=1}^{m} r_j g_j(x)_+ + \sum_{j=1}^{k} s_j |h_j(x)| ,$$

$$g_j(x)_+ = \max\,(g_j(x), 0) ,$$

$$P'(x^i) = \nabla f(x^i)^T d - \sum_{J=i}^{m} r_j\, g_j\,(x^i)_+ - \sum_{J=i}^{K} s_j\,|h_j\,(x^i)|$$

is an approximation to the directional derivative of $P(x)$,

and $\delta \in (0,1/2)$.

With Han's line search function, the vectors r and s have constant scalar elements, c, given by:

$$c > \| u^*, v^* \|_\infty$$

where u,v are the multipliers at the K-T point of (1.1).

Powell's implementation, which is not as restrictive on the penalty terms, defines the vectors r and s to be

$$r^0 = 0 \quad ; \quad s^0 = 0$$

$$r_j^i = \max \left(u_j^i \, , \, \frac{1}{2} (u_j^i + r_j^{i-1}) \right) \quad , \quad j = 1, m$$

$$s_j^i = \max \left(|v_j^i| \, , \, \frac{1}{2} (|v_j^i| + s_j^{i-1}) \right) \quad , \quad j = 1, k$$

However, Powell's implementation does not have the global convergence properties shown by Han.

Chamberlain (1979) gave two examples where Powell's algorithm cycled between two vertices of the linearized constraints. Continued cycling in the second problem even caused the Hessian approximation matrix to become unbounded. However, Chamberlain et al. (1982) showed that use of Han's penalty function causes convergence to be too slow in some cases.

To resolve some of these difficulties, Chamberlain et al (1982) proposed the watchdog technique. Here the method alternates between an exact penalty line search and full steps in the search direction. If the exact penalty function decreases monotonically, then usually full steps will be chosen. This method is harder to implement because the convergence proof may require a restart from a previous point. Thus, all the information at this point must be stored.

The next section discusses the use of the augmented Lagrangian line search function. To motivate the discussion we show that this follows quite naturally from the development of earlier quasi-Newton and augmented Lagrangian algorithms. It should be mentioned that our line search function differs from the recent work of Yamashita (1982) and Schittkowski (1981) because our function is nondifferentiable at certain points and our penalty parameter is chosen adaptively.

The third section presents global convergence results that, parallel the work of Han (1977) and local convergence properties that use the recent work of Schittkowski (1981). The fourth section discusses an adaptive strategy for choosing the penalty parameter and presents the line search algorithm. Other enhancements to the SQP algorithm are also mentioned. This section is followed by an extensive numerical comparison that includes the slow convergence effects of Chamberlain et al (1982), the cycling problems of Chamberlain (1979) as well as numerical results for fifteen well-known test problems.

Finally, we summarize the results of the paper and state conclusions in the last section.

2) AUGMENTED LAGRANGIAN LINE SEARCHES

To motivate the presentation of this line search function, let us first consider the equality constrained problem:

$$(2.1) \qquad \text{Min } f(x) \quad \text{s.t. } h(x) = 0$$

Necessary optimality conditions can be written as:

$$\nabla f(\bar{x}) + \nabla h(\bar{x})\,\bar{v} = 0$$

$$h(\bar{x}) = 0$$

which are simply stationary points with respect to x and v of the Lagrange function:

$$L(x,v) = f(x) + h(x)^T v$$

If we augment this function with a penalty term:

$$(2\text{-}2) \qquad L_a(x,v,\alpha) = f(x) + h(\bar{x})^T v + \frac{\alpha}{2} h(x)^T h(x)$$

we find that the augmented Lagrangian has the same stationary point as the

Lagrange function, regardless of the value of α , the penalty parameter. The stationary point can be found by applying Newton's method to $\nabla L_a(x,v,\alpha) = 0$.

Expanding $L_a(x,v,\alpha)$ formally in a Taylor series with respect to x and v about a point $(x^i, v^i) \in R^{n+k}$ yields:

$$L_a(x,v,\alpha) = L_a(x^i, v^i, \alpha) + \nabla L_a(x^i, v^i, \alpha)^T \begin{bmatrix} x-x^i \\ v-v^i \end{bmatrix}$$

$$+ \frac{1}{2} \begin{bmatrix} x-x^i \\ v-v^i \end{bmatrix}^T \nabla^2 L_a(x^i, v^i, \alpha) \begin{bmatrix} x-x^i \\ v-v^i \end{bmatrix}$$

$$+ 0 \left(\left\| \begin{bmatrix} x-x^i \\ v-v^i \end{bmatrix} \right\|^3 \right)$$

Truncating the series after three terms and finding a stationary point with respect to x and v, yields:

$$\nabla L_a(x^i, v^i, \alpha) + \nabla^2 L_a(x^i, v^i, \alpha) \begin{bmatrix} \bar{x}-x^i \\ \bar{v}-v^i \end{bmatrix} = 0 \tag{2.3}$$

Let $(\bar{x}-x^i)$ be defined by the vector d. Since

$$\nabla L_a(x,v,\alpha) = \begin{bmatrix} \nabla f(x) + \nabla h(x)v + \alpha \nabla h(x) \; h(x) \\ h(x) \end{bmatrix} ,$$

$$\nabla^2 L_a(x,v,\alpha) = \begin{bmatrix} \nabla_{xx} L_a(x,v,\alpha) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{bmatrix}, \quad \text{and}$$

$$\nabla_{xx} L_a(x,v,\alpha) = \nabla^2 f(x) + \nabla^2 h(x) v + \alpha \nabla h(x) \nabla h(x)^T$$
$$+ \alpha \nabla^2 h(x)\, h(x) ,$$

we can simplify (2.3) to:

$$\nabla_x L_a(x^i,\bar{v},\alpha) + \nabla_{xx} L_a(x^i,v^i,\alpha)^T d = 0$$

$$h(x^i) + \nabla h(x^i)^T d = 0$$

If the constraints are not highly nonlinear we can neglect $\alpha \nabla^2 h(x)h(x)$, the last term in $\nabla_{xx} L_a(x,v,\alpha)$. The above equations can then be expanded to:

(2.4)
$$\nabla f(x^i) + \nabla h(x^i)\bar{v} + \alpha \nabla h(x^i)\, h(x^i)$$
$$+ \left(\nabla^2 f(x^i) + \nabla^2 h(x^i)\, v + \alpha \nabla h(x^i) \nabla h(x^i)^T \right) d = 0$$

(2.5)
$$h(x^i) + \nabla h(x^i)^T d = 0$$

Because of equation (2.5) it is clearly seen that the truncated Newton step for $L_a(x,v,\alpha)$ can be found by solving the following quadratic program:

(2.6)
$$\min_{d \in R^n} \quad \nabla f(x^i)^T d + \tfrac{1}{2} d^T \nabla_{xx} L(x^i, v^i) d$$
$$\text{s.t.} \quad h(x^i) + \nabla h(x^i)^T d = 0$$

and obtaining d and $\bar{v}$.

Note that the solution to (2.6) is independent of *of* (Han 1978),

FJetcher (1974)). Since $\nabla_{xx} M x^1, v^1)$ involves calculation of second

derivatives we approximate this matrix by B_f which is constructed by quasi-Newton updates to $\nabla_{xx} L(x,v)$.

The result is the familiar SQP algorithm for equality constrained problems. Since this method follows from minimization of an augmented Lagrangian function, it is quite natural to choose this function to determine the stepsize X along the direction, $\begin{bmatrix} d \\ \bar{v} - v \end{bmatrix}$ chosen by the QP.

Inequalities can be included by allowing the QP to determine the active set from linearizations of all the constraints. After solving the QP:

$$Q(x^i,B) \qquad \text{Kin} \quad \nabla f(x^i)^T d + \frac{1}{2} d^T B d$$

$$g(x^i) + \nabla g(x^i)^T d \leq 0$$

$$h(x^i) + \nabla h(x^i)^T d = 0$$

the stepsize along $\begin{bmatrix} d \\ \bar{u} - u \\ \bar{v} - v \end{bmatrix}$ can be found by minimizing a modified augmented Lagrangian function:

$$(2.7) \qquad L^*(x,u,v,\sigma) = f(x) + u^T g(x)_+ + v^T h(x)$$

$$+ \frac{\sigma}{2} \left\| g(x)_+ , h(x) \right\|^2$$

where

$$g_j(x)_+ = \max(0, g_j(x))$$

u,v - multipliers for g and h, respectively

$\|\cdot\|$ - the Euclidean norm

The form of (2.7) ignores the inequality constraints unless they are violated during the line search.

(2.7) is similar to classical differentiate augmented Lagrange functions (BertseKas,1976):

$$f(x) + \frac{1}{2\alpha} \sum_{j=1}^{m} \left[(\alpha g_j(x) + u_j)_+^2 - u_j^2 \right]$$

$$+ v^T h(x) + f \, \| h(x) \|^2$$

However, classical techniques or multiplier methods are less efficient because they generally involve two nested iterations. The inner iteration minimizes the augmented Lagra.ngian for x with $u_f v$, and or fixed_r while the outer iteration updates u and v to maximize the function. The penalty parameter is increased in the outer iteration only if there is no decrease in the magnitude of the constraint violation.
The classical augmented Lagrangian function was also used for an SQP line search in the recent work of Yamashita (198Z) and Schittkowski (1981). Though they have substantially different implementations, both authors present desirable convergence properties.

Although our line search function (2.7) is not everywhere differentiable, it has several advantages over the functions of Han and Powell. In addition to promoting convergence, the line search function should help maintain or approach feasibility. The function used by Han and especially the one used by Powell may not suitably penalize constraint infeasibilities. The vectors r and s are determined directly from the Kuhn-Tucker multipliers of a quadratic program that handles linearized constraints. Thus a violation of a nonlinear constraint at x+d may be ignored if the quadratic program does not make this constraint active. Two examples of this are given by Chamberlain (1979). The augmented Lagrangian has similar multiplier-related terms but also contains a squared penalty term that emphasizes all of the constraint infeasibilities. Because the QP solution is independent of cy , we can adjust this penalty parameter as needed to approach or maintain feasibility in the search.

Another important feature is the number of derivative discontinuities in the line search function. With $P(x, r_f s)$ each active constraint has a dis-

continuous derivative at g(x) or h(x)=O. With L only the $u^x g(x)_+$ term contains derivative discontinuities. During the line search, the stepsize can be found efficiently by minimizing a quadratic function fitted by values of the line search function at the two end points and the directional derivative. If fewer derivative discontinuities are present, the quadratic fit and the choice of stepsize will be more accurate. For the augmented Lagrangian function, this is especialy true if equality constraints are present.

3) CONVERGENCE OF AUGMENTED LAGRANGIAN LINE SEARCHES

We begin by showing that the search direction found by the quadratic program is a descent direction of the augmented Lagrangian function if α is sufficiently large. This will be used later for the global and local convergence proofs.

Lemma 3.1 (Dem'yanov and Kalozemov (1974), referred by Han (1977))

If q_i , $i=1,\ldots,I$ are continuously differentiate functions from $R^n \to R$ and

$$\psi(x) = \max_i \{q_i(x)\}$$

then for any direction d, the upper directional derivative $D_d\psi(x)$ exists and

$$D_d \psi(x) = \max_{i \in I(x)} (\nabla q_i(x)^T d)$$

where

$$I(x) = \{i \mid q_i(x) = \psi(x)\}$$

Theorem 3.1

Let f,g and h be continuously differentiable on x and B be a positive definite, symmetric matrix. Here $x \in R^n$, $d \in R^n$, $u \in R^m$, $v \in R^k$ and $B \in R^{n \times n}$. If $(d, \bar{u}, \bar{v})$ is a Kuhn-Tucker triple of $Q(x,B)$, $d \neq 0$ and

$$\alpha > \frac{\nabla f^{iT} d + \left(-\begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} - 2 \begin{bmatrix} u^i \\ v^i \end{bmatrix} \right)^T \begin{bmatrix} g^i_+ \\ h^i \end{bmatrix}}{\| g^i_+ , h^i \|^2}$$

Then $\quad D_p L^* (x,u,v,\alpha) \le 0 \quad$ where $p = \begin{bmatrix} d \\ \bar{u}-u^i \\ \bar{v}-v^i \end{bmatrix}$

Proof

We write the upper directional derivative of $L^*(x,u,v,\alpha)$ as:

$$D_p L^* = \nabla f^{iT} d + u^{iT} \nabla g_+^{iT} d + v^{iT} \nabla h^{iT} d$$

$$+ \alpha (g_+^{iT} \nabla g^{iT} d) + \alpha (h^{iT} \nabla h^{iT} d)$$

$$+ (\bar{U} - U^i)^T (g_+^i) + (\bar{V} - V^i)^T (h^i)$$

Here

$$(\nabla g_+^T d)_j \begin{cases} = 0 & \hat{J} = \{ j : g_j(x) < 0 \} \\ = (\nabla g^T d)_j \quad j \in J & J = \{ j : g_j(x) > 0 \} \\ = \max \{ 0, (\nabla g^T d)_j \} \quad j \in \bar{J} & \bar{J} = \{ j : g_j(x) = 0 \} \end{cases}$$

From the QP we have:

$$-h^i = \nabla h^{iT} d$$

$$-g^i = \nabla g^{iT} d$$

which gives

$$D_p L^* \le \nabla f^{iT} d + \left(\begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} - 2 \begin{bmatrix} u^i \\ v^i \end{bmatrix} \right)^T \begin{bmatrix} g^i{}_+ \\ h^i \end{bmatrix} - \alpha \, || g^i{}_+ \, , h^i ||^2$$

Substituting for α gives $D_p L^* < 0$.

We call the lower bound of α that gives the descent direction, α_{dd}.

Corollary 3.1

If the search direction $p = \begin{bmatrix} d \\ \bar{u}-u^i \\ \bar{v}-v^i \end{bmatrix} \neq 0$

then $\exists$ a $\lambda > 0$ such that

$$L^*(z^i + \lambda p \, , \alpha) \le L^* (x^i , u^i , v^i , \alpha) + \delta \lambda \phi (z^i)$$

where $z^i = \begin{bmatrix} x^i \\ u^i \\ v^i \end{bmatrix}$, $\alpha > \alpha_{dd}$ $\delta \in (0,\frac{1}{2})$ and $\phi(z^i) = D_p L^*(z^i)$

We continue with a perturbation lemma which is due to Daniel (1973). (See also Lemma 3.2, Han (1977))

Lemma 3.2 (Theorem 4.3, Daniel (1973))

Let $\hat{d}$ minimize $q(d) = \frac{1}{2} d^T \hat{B} d + \hat{b}^T d$

$$\text{s.t.} \quad \hat{A} d \le \hat{a}$$
$$\hat{C} d = \hat{c}$$

and let $\bar{d}$ minimize

$$q(d) = \frac{1}{2} d^T\bar{B}d + \bar{b}^T d$$

$$\text{s.t.} \quad \bar{A}d \leq \bar{a}$$

$$\bar{C}d = \bar{c} \quad .$$

Then for any fixed norm $||\cdot||$, there exist $s > 0$ and some $\bar{\epsilon}$ such that

$$|| \hat{d} - \bar{d} || \leq s\epsilon$$

if a) $\epsilon \leq \bar{\epsilon}$

b) B is positive definite

where

$$\epsilon = \max \{|| \hat{B}-\bar{B}|| \, , \, || \hat{A}-\bar{A}|| \, , \, ||\hat{C}-\bar{C}|| \, , \, ||\hat{a}-\bar{a}|| \quad , \\ ||\hat{b}-\bar{b}|| \, , \, ||\hat{c}-\bar{c}|| \quad \}$$

Similarly, we can establish a bound on the multipliers by applying this lemma to the dual quadratic program:

$$\text{Min } \frac{1}{2} (b + A^Tu + C^Tv)^T B^{-1}(b + A^Tu + C^Tv) - a^Tu - c^Tv$$

$$\text{s.t.} \quad u \geq 0$$

The bounds on the multipliers are therefore:

$$\|(\hat{u}, \hat{v}) - (\bar{u}, \bar{v})\| \leq t\varepsilon$$

whenever $\varepsilon \leq \bar{\varepsilon}$ and $\varepsilon = \max\left(\|\hat{H} - \bar{H}\|, \|\hat{r} - \bar{r}\|\right)$

where
$$H = \begin{bmatrix} AB^{-1}A^T & AB^{-1}C^T \\ CB^{-1}A^T & CB^{-1}C^T \end{bmatrix}$$

$$r = \begin{bmatrix} AB^{-1}b - a \\ CB^{-1}b - c \end{bmatrix}$$

$$t > 0 .$$

Let f, g and h be continuously diffcrentiable and assume some $\underline{\beta}, \bar{\beta} > 0$ such that:

$$\underline{\beta} \|x\|^2 \leq x^T B x \leq \bar{\beta} \|x\|^2, \forall x$$

and α is always greater than or, , at each iteration.
dd

Then any sequence $\{x^i, u^i, v^i\}$ with well-defined Quasi-Newton updates $\{B^i\}$ either terminates at a Kuhn-Tucker point of (1.1) or any accumulation point $\{x^*, u^*, v^*\}$ satisfying the Mangasarian-Fromowitz Constraint Qualification (MFCQ):

$$\exists\, w \in R^n \mid \nabla g_k(x^*)^T w < 0,\ k \in K$$

where $K = \{k \mid g_k(x) = 0\}$

$\nabla h(x^*)^T w = 0$

and $\nabla h(x^*)$ has full rank

is a Kuhn-Tucker point of (1.1).

Proof

If $p^i = 0$, then (x^i, u^i, v^i) satisfies the Kuhn-Tucker conditions of (1.1). Suppose $p^i \neq 0$ then $z^{i+1} = z^i + \lambda p^i$ exists for some $\lambda > 0$ where

$$L^*(z^{i+1}, \alpha) \leq L^*(z^i, \alpha) + \delta \lambda f(z^i)$$

Let z^* be an accumulation point of $\{z^i\}$ satisfying tbe constraint qualification. Because B^i is formed by well-defined Quasi-Newton updates, then as

$$z^i \rightarrow z^*, \quad B^i \rightarrow B^*.$$

From the definition of B^i and the MFCQ it follows that $Q(x^*, B^*)$ has a Kuhn-Tucker point, p^*. If $p^* = 0$ then z^* is a Kuhn-Tucfcer point of (1.1) and we are done. Suppose $p^* \neq 0$. By Lemma 3.2 and the solution of $Q(x^*, B^*)$:

$$u^i \to u^*$$
$$v^i \to v^*$$

By Corollary 3.1 we can find $\bar{X}$ such that:

$$L(z + X p, \sigma) - L(z, \sigma) \leq \delta X f(z) < 0$$

Since $p^* \neq 0$, $X^i \to 0$. Here i sufficiently large that for some $Y > 1$ the steplength rule is violated:

$$L^*(z^i + X^i Y p^i, \sigma) - L''(\wedge <^*) > \delta X \wedge f(z^i)$$

The LHS can be expanded in X to yield:

$$YX^i \phi(z^i) + yX \wedge CX^i) > Y X^i \delta f(z^i)$$

and dividing by $Y X^i$ gives:

$$\phi(z^i) + O(X^i) > \delta f(z^i)$$

From Theorem 3.1, $\phi(z^i) < 0$ and we have a contradiction since:

$$0 > (1-\delta)\, \phi(z^i) \wedge (1-\delta) f(z^i) + O(X^i) > 0$$

and $\delta \in (0, 1/2)$.

Theorem 3.3

Based on the assumptions in Theorems 3.1 and 3.2, there exists a finite σ such that $\sigma > \max_i \sigma^i_{dd}$,

Proof

From Theorem 3.1

$$\sigma_{dd} = \frac{\nabla f^{iT} d + u^{iT} \nabla g^{iT} d + v^{iT} \nabla h^{iT} d + L\forall \cdot V^{:L} J \quad Lh\wedge J}{\| g^i_+ , h^i \|^2}$$

From the QP:

$$\nabla f^{iT} d = - d^T B d - \bar{u} \nabla g^{iT} d - \bar{v} \nabla h^{iT} d$$

$$\nabla g(x^i)^T d \leq - g(x^i)_+$$

$$\nabla h(x^i)^T d = - h(x^i)$$

$$\bar{u}^T \nabla g(x^i)_+^T d - \bar{u}^T g(x^i)_+$$

Substituting for α_{dd} gives :

$$\alpha_{dd} \leq \frac{- d^T B d + 2 (\bar{u}-u^i)^T g(x^i)_+ + 2(\bar{v}-v^i)^T h(x^i)}{\| g^i_+ , h^i \|^2}$$

Now, realizing that B is positive definite, we take the norm of the numerator and write:

$$\alpha_{dd} \leq \frac{- \beta \|d\|^2 + 2 \left\| \begin{matrix} \bar{u}-u^i \\ \bar{v}-v^i \end{matrix} \right\| \| g^i_+ , h^i \|}{\| g^i_+ , h^i \|^2}$$

Moreover, since $p = \begin{bmatrix} d \\ \bar{u}-u^i \\ \bar{v}-v^i \end{bmatrix}$ we write:

$$\alpha_{dd} \leq \left(\frac{- \beta \| p \|^2 + 2 \|p\| \| g^i_+ , h^i \|}{\| g^i_+ , h^i \|^2} \right)$$

For $\|p\| \neq 0$, α remains bounded above even if $\| g^i_+ , h^i \| \to 0$.

KMWifII,I1-2;.0>t|lcrc ither:

i) $\|p\| = O\|g_+, h\|$

$$\alpha_{dd} \leq \frac{-p^T H p + 2K\|p\|^2}{K^2\|p\|^2} \leq \frac{2K - \beta}{K^2}$$

where K is a postive constant. Hence α_{dd} has a finite upper bound

ii) $\|g_+, h\| = o\|p\|$. Here $K \to 0$

$$\alpha_{dd} \leq \frac{2K-P}{K^2} \to -\infty$$

and any positive value of a gives a descent direction.

iii) $\|p\| = o\|g_+, h\|$. Here $K \to 0$

$$\alpha_{dd} * \frac{-BK^2\|g_+, h\|^2 + 2K\|g_+, h\|^2}{\|g_+, h\|^2}$$

$$= (2 - \beta K) K \to 0$$

and any positive a gives a descent direction.

Remark 3.1

To show that an accumulation point exists we require that: x and d be bounded, that the gradient matrix of active constraints at each iteration has full rank, and that each QPP is solvable. Thus $(\vec{u},\vec{v})$ is bounded, and because the Armijo inequality is always satisfied for L* (zi), an accumulation point will be found. (see e.g. Ortega & Rheinboldt (1970))

Local Superlinear Convergence

We restrict the analysis to show that the Armijo inequality on this line search function allows full steps in the region about the solution of (1.1). Once this can be shown we invoke the local results of Kan (1976) and Powell (1978). We also assume that the active set has been determined. Mere any inactive constraints may be discarded and the constraint set

$$J = (j \mid g_j(x) > 0)$$

is at worst a subset of the active constraints.

In addition we assume that the Hessian approximation, B, is always positive definite and satisfies the property:

$$d^T(\nabla_{xx} L(z^*) - B)d = o (|| d||^2) .$$

as it converges. This is characteristic of several quasi-Newton updating formulas (see Boggs, et al., 1982),

Theorem 3.4

If $\forall$ i sufficiently large, we have:

1) $\gamma ||p|| \leq d^T Bd \leq \bar{\gamma} ||p||$

for some γ , $\bar{\gamma} > 0$

2) The active set of inequality constraints is determined and

$j \in J = \{ j \mid g_j(x) > 0 \}$ belongs to this set.

3) $\nabla^2 h_j^i$ and $\nabla^2 g_j^i$ are finite, $\forall x^i$

4) $d^T (\nabla_{xx} L(z^*) - B^i)d = o (||d||^2)$ as $i \to \infty$

where z^* is the Kuhn-Tucker triple of (1.1)

5) and $\alpha \geq \dfrac{\left[- (1-2\epsilon)d^T Bd + 2 (\bar{u}-u^i)g(x^i)_+ + 2(\bar{v}-v^i)h(x^i) \right]}{|| g(x^i)_+ , h(x^i) ||^2}$

where $\quad \epsilon > -\min\left(0, \dfrac{\sum_{j \in \hat{J}} \bar{u}_j d^T \nabla^2 g_j(x^i) d}{(1-2\delta) d^T B d}\right)$

$$\hat{J} = \{ j \mid g_j(x) < 0 \}.$$

Then

$$L^*(z^i + p) \le L^*(z^i) + \delta \phi(z^i)$$

and the Armijo inequality is satisfied with a stepsize of unity.

Proof

We first relate the modified augmented Lagrangian functions to the Lagrangian.

$$L^*(z^i) = f(x^i) + u^{iT} g(x^i)_+ + v^{iT} h(x^i)$$

$$+ \frac{\alpha}{2} \| g(x^i)_+ , h(x^i) \|^2$$

$$= L(z^i) - u^{iT} g(x^i)_- + \frac{\alpha}{2} \| g(x^i)_+ , h(x^i) \|^2$$

where

$$u^{iT} g(x^i)_- = \sum_{j \in \hat{J}} u_j^i g_j(x^i)$$

$$\hat{J} = \{ j \mid g_j(x^i) < 0 \}$$

$$L^*(z^i + p) = f(x^i + d) + \bar{u}^T g(x^i + d)_+ + \bar{v}^T h(x^i + d)$$

$$+ \frac{\alpha}{2} \| g(x^i + d)_+ , h(x^i + d) \|^2$$

$$= L(z^i + p) - \bar{u}^T g(x^i + d)_- + \frac{\alpha}{2} \| g(x^i + d)_+ , h(x^i + d) \|^2$$

$$\phi(z^i) = \nabla f(x^i)^T d + u^{iT} \nabla g(x^i)_+^T d + v^{iT} \nabla h(x^i)^T d$$

$$+ (\bar{u} - u^i)^T g(x^i)_+ + (\bar{v} - v^i)^T h(x^i)$$

$$- \alpha \| g(x^i)_+ , h(x^i) \|^2$$

$$= \nabla L(z^i)^T p - u^{iT} \nabla g(x^i)_-^T d - (\bar{u} - u^i)^T g(x^i)_-$$

$$- \alpha \| g(x^i)_+ , h(x^i) \|^2$$

but since all constraints are active (by 2))

$$\phi(z^{i}) = \nabla L(z^{i})^{T} p + (2u^{i}-\bar{u})^{T} g(x^{i})_{-} - \sigma \| g(x^{i})_{+}, h(x^{i}) \|^{2}$$

Now we need to show that:

$$\text{(AI)} \quad L^{*}(z^{i} + p) - L^{*}(z^{i}) - \delta\, \phi(z^{i})$$

is nonpositive. Substituting the above quantities in (AI) gives:

$$\Big[L(z^{i}+p) - L(z^{i}) - \delta \nabla L(z^{i})^{T} p \Big]$$
$$- \bar{u}^{T} g(x^{i}+d)_{-} + \frac{1}{2}\sigma \| g(x^{i}+d)_{+}, h(x^{i}+d) \|^{2}$$
$$+ u^{iT} g(x^{i})_{-} - \frac{1}{2}\sigma \| g(x^{i})_{+}, h(x^{i}) \|^{2} - \delta (2u^{i}-\bar{u})^{T} g(x^{i})_{-}$$
$$+ \delta \sigma \| g(x^{i})_{+}, h(x^{i}) \|^{2}$$

From a Taylor series expansion and the solution of the QP we know:

$$g_j(x^{i} + d) = g_j(x^{i}) + \nabla g_j(x^{i})^{T} d + \frac{1}{2} d^{T} \nabla^{2} g_j(x^{i}) d + O(\|d\|^{3})$$

$$= \frac{1}{2} d^{T} \nabla^{2} g_j(x^{i}) d + O(\|d\|^{3})$$

$$h_j(x^{i} + d) = h_j(x^{i}) + \nabla h_j(x^{i})^{T} d + \frac{1}{2} d^{T} \nabla^{2} h_j(x^{i}) d + O(\|d\|^{3})$$

$$= \frac{1}{2} d^{T} \nabla^{2} h_j(x^{i}) d + O(\|d\|^{3})$$

so (AI) becomes

$$\Big[L(z^{i}+p) - L(z^{i}) - \delta \nabla L(z^{i})^{T} p \Big] - \frac{1}{2} (\bar{u}^{T} d^{T} \nabla^{2} g(x^{i})_{-} d)$$
$$+ O(\|d\|^{3}) + (1-2\delta) u^{iT} g(x^{i})_{-} - \frac{1}{2}(1-2\delta)\sigma \| g(x^{i})_{+}, h(x^{i}) \|^{2}$$
$$+ \delta \bar{u}^{T} g(x^{i})_{-} \quad (\text{by } 3))$$

The bracketed quantity above is:

$$[(1-\dot{6})\ \nabla L(z^1)^T p + \tfrac{1}{2} p^T \nabla W)\ p + 0(\|p\|^3)]$$

where

$$\nabla L(z^1)^T p = \nabla f(x^L)^T d + u^{1T} \nabla g(x^1)^T d + v^{1T} \nabla h(x^1)^T d$$

$$+ (\bar{u}-u^i)^T g(x^i) + (\bar{v}-v^i)^T h(x^i)$$

$$= -d^T B d + 2(\bar{u}-u^1)^T g(x^1) + 2(\bar{v}-v^1)^T h(x^i)$$

(by the QP)

$$p^T \nabla^2 M(z^1) p = d^T \nabla^2 L(z^1)\ d + 2(\bar{u}-u^L)^T \nabla g(x^i)^T d$$

$$+ 2(\bar{v}-v^1)^T \nabla h(x^1)^T d$$

$$= d^T \nabla_{xx} L(z^1) d - 2(\bar{u}-u^i)^T g(x^i)$$

$$- 2(\bar{v}-v^1)^T h(x^1) \qquad \text{(from 2))}$$

Thus (AI) becomes:

$$-(\tfrac{1}{2}-6)\ d^T B d + \tfrac{1}{2} d^T (\nabla_{yx} L(z^1) - B)\ d$$

$$+ (1-26)\ [(\bar{u}-u^1)^T g(x^1) + (\bar{v}-v^1)^T h(x^1) + u^{1T} g(x^1)]$$

$$- f\ (1-26)\| g(x^i)_+ , h(x^1) \|^2 + 6\ G^T g(x^i)$$

$$- \tfrac{1}{2} (\bar{u}\ \nabla v^2 g(x^1).\ d) + 0(\|p\|^3)$$

Now because :

$$d^T \left[\left(\nabla_{xx} L(z^i) - \nabla_{xx} L(z^*) \right) + \left(\nabla_{xx} L(z^*) - B \right) \right] d = o(\|d\|^2)$$

$$\delta < \frac{1}{2} \qquad \bar{u}^T g(x)_- < 0 \quad , \quad u^T g(x) \le u^T g(x)_+$$

and using assumption 4) we have:

$$(AI) \le \left(\frac{1}{2} - \delta\right) \left[-d^T B d + 2(\bar{u}-u^i)^T g(x^i)_+ \right.$$

$$+ 2(\bar{v}-v^i)^T h(x^i) - \alpha \| g(x^i)_+ , h(x^i) \|^2$$

$$\left. - \frac{\bar{u}^T d \; \nabla^2 g(x^i)_- d}{(1-2\delta)} \right] + o(\|p\|^2)$$

Using assumption 1) and 5) gives

$$(AI) \le -\left(\frac{1}{2} - \delta\right) c_\gamma \|p\|^2 + o(\|p\|^2)$$

and for i sufficiently large:

$$(AI) \le -\left(\frac{1}{2} - \delta\right) c_\gamma \|p\|^2 + o(\|p\|^2) \le -\left(\frac{1}{2} - \delta\right) \frac{c_\gamma}{2} \|p\|^2 \le 0$$

Thus the Armijo Inequality is satisfied for a stepsize of one and the theorem is proved.

Kcmark 3.2

In practice, assumption]) of Theorem 3.4 is not really required. If for example, $||d|| = o(H\hat{}-u^1)$, $(v-v^1)!!) = o(||p||)$, then the algorithm still converges superlinearly if $K(||p|j^2) / ||g_+, h||^2$ (for some K > O) is added to the right hand side of assumption 5). As will be seen in the next section, our algorithm covers this case as well. Note that even here $a \hat{} a_{da}$.

4. An algorithm for choosing the penalty parameter.

From the convergence results of the previous section we can state the following conditions for the penalty parameter,

1) If *a* is large enough to give a descent direction at every iteration, then an SQP algorithm with this augmented Lagrangian line search is globally convergent.

2) If *a* is larger than Q^ and large enough to compensate for any active inequality constraints that remain feasible as the algorithm converges, (since $d^T \nabla^2 g\, d < 0$, these will be locally concave), then the algorithm has local superlinear convergence.

These two statements suggest an adaptive strategy for updating *a*. This contrasts our work from the a priori updating algorithms of SchittkowsXi (1981) and Yamashita (1982).

Here, it is straightforward to calculate the lower bound a_{bb} and (upper or lower) bounds on *a* that satisfy the Armijo inequality for a given stepsize. Thus, instead of explicitly fixing or for the line, search, we merely determine if a region exists for *a* that satisfies the Armijo inequality and the descent condition. This prevents any unnecessary stepsize restriction during the line search.

We consider two quantities at each iteration

$$\alpha_{dd} = \frac{\nabla f(x^i)^T d + (\bar{u} - 2u^i)^T g(x^i)_+ + (\bar{v} - 2v^i)^T h(x^i)}{||g(x^i)_+ , h(x^\ell)||^2}$$

$$\alpha_{LS} = - \frac{(L_+(z^i) + \lambda \delta \nabla L_+(z^i)^T p) + L_+(z^i + \lambda p)}{\psi(X)}$$

where

$$L_+(z) = f(x) + u^T g(x)_+ + v^T h(x)$$

$$\nabla L_+(z^i)^T p = \nabla f(x^i)^T d + (\bar{u} - 2u^i)^T g(x^i)_+ + (\bar{v} - 2v^i)^T h(x^i)$$

$$\psi(\lambda) = -\frac{1}{2} ||g(x^i \cdot X_d)_+ , h(x^i + \lambda d)||^2 + (\frac{1}{2} - \delta\lambda)||g(x^i)_+ , h(x^i)||^2$$

The above expression for α_{LS} is derived from rearranging the Armijo inequality (AI). Here α_{LS} can represent either an upper or lower bound which α must satisfy for any given λ. If $\psi(\lambda)$ is positive then any $\alpha \geq \alpha_{LS}$ will automatically satisfy the (AI); conversely if $\psi(\lambda) < 0$ then any $\alpha \leq \alpha_{LS}$ satisfies the (AI).

We can now replace step 3 of the algorithm presented in section 1 with the following L.S. procedure

a) Set $\lambda = 1$

b) $\alpha^i = \max(\alpha^{i-1}, 1.1\,\alpha_{dd})$

c) calculate α_{LS}

d) If $\alpha_{LS} < \max(0, \alpha_{dd})$ or $\psi(\lambda) = 0$, go to e.)

If $\alpha_{LS} > \max\limits_j(\alpha^j_{dd})$ set $\alpha^i = \alpha^{LS}$

(the (AI) is automatically satisfied at the current λ go to f.)

e) Test the (AI) for α^i from b)

$$L^*(z^i + \lambda p, \alpha^i) - L^*(z^i, \alpha^i) \leq \delta \lambda \phi(z^i, \alpha^i)$$

(At this point we know that

1) since $\alpha_{LS} < \alpha_{dd} < \alpha^i$

α_{LS} can not be used as a basis for changing α^i because this will destroy the certainty of a descent direction

2) if α_{LS} happens to be a lower bound, either postive or negative, then since $\alpha^i \geq \alpha_{LS}$ the (AI) will be satisfied on the first try. This will be the case if $\psi(\lambda) > 0$

3) if $\psi(\lambda)$ is non-positive then (AI) is not satisfied for $\psi(\lambda) < 0$, (AI) does not depend on α for $\psi(\lambda) = 0$)

If now the (AI) is satisfied go to f.). Otherwise reduce λ using quadratic interpolation of $L^*(z)$ and go to c.)

f. update according to

$$x^{i+1} = x^i + \lambda d$$

$$u^{i+1} = u^i + \lambda (\bar{u} - u^i)$$

$$v^{i+1} = v^i + \lambda (\bar{v} - v^i)$$

$$i = i + 1$$

Go to Step 1) of the algorithm in section 1

The next section presents a comparison of our stepsize strategy with those of both Powell (1977) and Schittkowski (1981) on fifteen well-known test problems. It will be seen that our strategy never required more function evaluations than the other two and sometimes required less. Schittkowski's strategy is less efficient because α increases monotonically and may even be unbounded. Also, Schittkowski's line search function may reward feasibility rather than just penalizing infeasibility. This function contains the terms $[\alpha g_j + u_j]_+$ where: g_j is the constraint value, u_j the Lagrange multiplier and α the penalty term. Notice here that, for any $g_j < 0$ (feasible), the term $[\alpha g_j + u_j]_+$ will be positive, unless $\alpha > -\dfrac{u_j}{g_j}$, thereby providing a contribution to the penalty function for a feasible constraint. No provisions are made to prevent this condition in Schittkowski's algorithm although in practice α does tend to become very large.

Initially we set $(u^o, v^o) = 0$ and $\alpha^o = \xi$ (some small positive number, e.g. 10^{-3}).

Again, we note that as long as α_{dd} gives a descent direction for all iterations j, the algorithm is globally convergent. Since, from Theorem 3.3f we know this quantity is bounded, we use the (heuristic) updating strategy in step d) to determine a - From Theorem 3.4 we know that near the solution 3 cy ^ a_{dd} that allows full steps along the search direction generated by the QP. This a is given by a_{LS} in step d) of the algorithm. Thus we have an adaptive strategy for choosing c* that exploits the properties determined in Section 3.

5. NUMERICAL RESULTS

This section is divided into two parts. In the first, we illustrate hew two shortcomings of the line search strategy of Powell (1977) are remedied by the approach described in the previous section. Here we present detailed results of our strategy and explain why the slow convergence and cycling problems encountered with Powell's line search are avoided.

A. Illustrations of the Line Search

i) Slow convergence with exact penalty function line search (see Chamberlain et al. (1982)].

Consider the problem Min $F(x) = -x_1 + j\ (x_1^{-} + x_2^{-} - 1)$

$$\text{s.t.}\quad c_1(x) = x_1^2 + x_2^2 - 1 = 0$$

This is similar to the example used by Maratos(1978) to discuss the slow convergence effect.

The solution can be seen by inspection as: $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Let $x^k = \begin{bmatrix} \cos 9 \\ \sin 8 \end{bmatrix}$ and assume no errors are made in approximating the matrix

$V_{xx}L(x^*, v_y^*) = I$. x^{k+1} is therefore $\begin{bmatrix} \cos 0 + \sin^2 G \\ \\ \sin 9 \ (l-\cos 8) \end{bmatrix} = x^k + ^d>$

As mentioned by Chamberlain et al. (1979), it follows that for any small positive number e ,

$\|x^k - x^*\|$ and the ratio $\|x^{k+1} - x^*\| / \|x^k - x^*\|$

can be made less than € if 9 is chosen sufficiently small.

From the definition of P(x)

$P(x^k) = -\cos Q$

$P(x^{k+1}) = -\cos e \cdot \sin^2 e + (T + s) \sin^2 6$.

But for a reduction to occur in P(x ⁻⁻⁻) .,

(4.1) $T + S < 1$.

Here, since $s \wedge 0$, the algorithm can not take full steps for $7 \geq 1$.

The Lagrangian and augmented Lagrangian functions are given by:

$$L(x^k, v) = -\cos\theta$$

$$L(x^{k+1}, v) = -\cos\theta - \sin^2\theta + \left(\frac{\cos\theta \sin^2\theta}{2}\right)$$

$$L^*(x^k, v, v) = -\cos\theta$$

$$L^*(x^{k+1}, v) = -\cos\theta - \sin^2\theta + \left(\frac{\cos\theta \sin^2\theta}{2}\right) + \frac{\sigma}{2}\sin^4\theta$$

and the multiplier on the constraint c_1 is:

$$v = -\frac{\cos\theta}{2}$$

Also: $\nabla L^{*T}_{,p} = \nabla L^{T}_{p} = -\sin^2\theta$

So a descent direction is guaranteed $\therefore \nabla\sigma > -\infty$

The Armijo inequality for $L(x,v)$ is :

$$-\sin^2\theta + \frac{\cos\theta \sin^2\theta}{2} + \frac{\alpha}{2}(\sin^4\theta) \leq -\delta \sin^2\theta$$

and
$$\alpha_{LS} = \frac{2}{\sin^2\theta}\left[(1-\delta) - \frac{\cos\theta}{2}\right]$$

Here, from the nature of the functions, α_{LS} is bounded by: $(1-2\delta) \leq \alpha_{LS} \leq \infty$, $\forall\ \theta$ and $\delta \in (0, \frac{1}{2})$. Thus the augmented Lagrangian function will take full steps to the solution because $f(1) < 0$ and $\sigma \leq \alpha_{LS}$.

Here, even the Lagrange function takes full steps because $\sigma = 0$.

(ii.) Cycling with Powell's line search function
(Chamberlain (1979))

The Problem Min x_2

s.t. cj (>:) = a(>:x) - $x_2 \le 0$

$c_2(x)$ = ad-^) - $x_2 \le 0$

where $a(x) = 2x^2 - x^3$

cycles between $x^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $x^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ if Powell's line search function is used.

Let

$x^0 = \begin{bmatrix} (0 \\ 0 \end{bmatrix}$ $B^0 = I$

$c_1^0 = o$ £°-0

$u^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $c_2^0 = 1$

The solution to the quadratic program gives TT = I 0 I, d=x^1 = I 1 I
Ur I o J

and $r^0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ $f^1 = o$

$c^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$P(x^0, r^0) = 1$ $\qquad$ $p'(x^\circ, r^0) = -1$

$L(x^\circ, u^\circ) = 0$ $\qquad$ $\nabla L(z)p = l/(z^\circ) = 0$

$L^*(x^0, u^0, a) = \frac{a}{2}$ $\qquad$ $\nabla L^*(z^0)^T p = L^{*\prime}(z^0) = -$ (ff-1)

where $*^f$ is the directional derivative $\nabla \cdot^T d$.

The Armijo inequality holds for $P(x^0, r^0)$ and $L(x^0, u^0)$ for the step-size, $X = l$

$$P(x^1, r^\circ) \leq P(x^\circ, r^\circ) + 6 P'(x^\circ, r^0)$$

$$0 \leq 1 - \delta$$

$$L(x^X, u^\circ) \wedge L(x^\circ, u^\circ) + 6 L^f(x^\circ, u^\circ)$$

$$0 \wedge 0 + 6$$

v:here 6 is set to O.I by Powell (1977). The augmented Lagrangian line search does not satisfy the inequality for $\backslash = 1$. Here :

$$L^*(x^1, \bar{u}, a) \leq L^*(x^\circ, u^\circ, a) + 6 L^{*\prime}(x^\circ, u^\circ, \alpha)_r$$

$\left(\text{which is } \frac{\alpha}{2} \leq | - 6(\alpha - 1)\right)$, is satisfied

only if $\alpha \leq \alpha_{LS} = 1$. However, α must also be greater than $\alpha_{dd} = 1$.

(Note that the Lagrangian function L(x,u) cycles because here $a = 0 < a_{dd}$.)

Setting B^0 to I, u^0 to 0 and x^0 to 0 and using the algorithm in Section 4, with quadratic line search interpolation, this optimization problem converges to the optimum , $x^T = [\ 0.5\ ,\ 0.375\ ]$, in 4 iterations.

B. Numerical Comparison on Test Problems

In this section we compare three line search strategies embodied in the following computer programs:

OPT - uses the line search strategy described in Section 4.

OPTHP - uses the strategy given by Powell (1977)

OPTSCH - uses the stragegy proposed by Schittkowski (1981)

Otherwise, the three codes use the quadratic programming algorithm of Gill and Murray (1978). This program has the desirable feature that the minimum norm of the infeasibilities is returned if linearization of the constraints is inconsistent. If this happens we merely resolve the QP with the constraint tolerance set to 1.01x(min norm).

We find this device safer than the one suggested by Powell. His strategy introduces a new parameter into the QP and transforms it from:

$$\text{(QP1)}\quad \text{Min}\quad a^T d + \frac{1}{2} d^T B d$$

$$\text{s.t.}\quad c + C^T d \le 0$$

$$r + R^T d = 0$$

to:

$$\text{(QP2)}\quad \text{Min}\quad \begin{bmatrix} a \\ \eta \end{bmatrix}^T \begin{bmatrix} d \\ \xi \end{bmatrix} + \frac{1}{2} \begin{bmatrix} d \\ \xi \end{bmatrix}^T \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} d \\ \xi \end{bmatrix}$$

$$\text{s.t.}\quad \begin{bmatrix} c_- \\ 0 \end{bmatrix} + \begin{bmatrix} C_- & 0 \\ C_+^T & c_+ \end{bmatrix}^T \begin{bmatrix} d \\ \xi \end{bmatrix} \le 0$$

$$\begin{bmatrix} R^T & r \end{bmatrix} \begin{bmatrix} d \\ \xi \end{bmatrix} = 0$$

$$0 \le \xi \le 1$$

where $c_- = c_{j\epsilon\hat{J}}$, $J = \{ j \mid c_j < 0 \}$

$c_+ = c_{j\epsilon J}$, $J = \{ j \mid c_j \geq 0 \}$

η is a large negative number (-10^6 in VFØ2AD (1977))

Incorporation of the ξ parameter will help correct the inconsistent linearization. However, there is no guarantee that d from (QP2) is equivalent to d from (QP1) even if the linearized constraints have a feasible region !

For example, if the QP solution is completely determined by the equality constraints ($d = -(R^T)^{-1}r$), it is easy to show that (QP2) will not find this solution if:

$$\frac{a^T(R^T)^{-1}r - \eta}{r^T R^{-1} B (R^T)^{-1} r} < 1$$

The following table provides a listing of the fifteen problems solved; N represents the number of independent variables, M the total number of constraints and MEQ the number of equality constraints. Each problem is denoted by a letter and number. The letter corresponds to the reference while the number identifies the problem number in the reference.

Table 5.1 shows, firstly, that the augmented Lagrangian strategy in OPT never required more function evaluations than either OPTHP or OPTSCH. In addition, the cycling exhibited on some problems by the OPTHP algorithm is always avoided as was shown by Theorem 3.2 and Remark 3.1.

The implementation of the Schittkowski line search function suffered often from illconditioning. Both the QPP subproblem, where illconditioning was measured by the Hessian condition number, and the line search subproblem, where very large α values (sometimes 10^{15} or more) were calculated, were responsible for the excessive number of function evaluations or failures shown. Attempts were made at restarting the failed problems with an initialization of the Hessian to I, but this did not always prove successful.

Conclusions

Based on theoretical considerations and numerical results we find that our adaptive augmented Lagrangian strategy performs both more effectively and reliably than previously implemented line search strategies. By exploiting the descent property and the Armijo inequality our adaptive procedure provides a more flexible way of choosing the penalty parameter when compared to other methods.

TABLE 5.1: ALGORITHM COMPARISON

PROBLEM DESCRIPTION				OPT	OPTHP	OPTSCH
	N	M	MEQ	#FN EVALS	#FN EVALS	#FN EVALS
A1	2	2	0	4	CYCLES	4
A2	1	2	0	3	CYCLES	3
B1	5	10	0	5	5	5
B3	5	6	0	3	3	3
B4	4	0	0	52	52	52
B6	6	4	4	13	25	>77*
C3	2	3	0	10	10	23*
C5	3	2	2	9	9	15
C13	5	6	0	4	4	38
C24	2	2	0	4	8	16
D4	10	11	3	30	30	>112*
D5	10	3	3	30	30	30
D9	4	4	0	5	5	5
E	2	1	0	10	10	10
F+	4	3	0	12	14	33

A. CHAMBERLAIN(1979)
B. COLVILLE(1968)
C. HIMKELBLAU(1972)
D. BRACKEN & KcC0RMICK(1968)
E. SCHULDT(1975)
F. ROSEN & SUZUKI(1965)

•converged to a tolerance of 10^{-15}; other problems to 10^{-3}
•required resetting the hessian to I

After introducing the current algorithm and describing its features, we motivate the use of augmented Lagrangian line searches in section 2. This section follows a similar derivation for diagonalized multiplier methods by Tapia (1977).

Section 3 presents conditions for the penalty parameter that lead to local superlinear and global convergence. Based on these conditions we present a line search algorithm that adaptively adjusts a by using the Armijo inequality and descent conditions.

Finally, section 5 presents a numerical comparison that demonstrates the benefits of this approach. Here we examine two problems that have plagued previous line searches and compare results from the solution of 15 test problems.

REFERENCES

'Beale, E.M.L., "Numerical Methods," in Nonlinear Programming (J. Abadie, ed.) North-Holland, Amsterdam, p. 189 (1967).

Bertsekas, D.P., "Multiplier Methods: A Survey," Automatica, 12, p. 133 (1976)

Boggs, P.T., J.W. Tolle and P.Wang, "On the local convergence of Quasi-Newton Methods for Constrained Optimization", SIAM J. Control and Opt., 20, 2, p.161 (1982)

Bracken, J. and G.P. McCormick, Selected Applications of Nonlinear Programming, Wiley, New York (1968).

Chamberlain, R.M., "Some Examples of Cycling in Variable Metric Methods for Constrained Minimization," Math. Prog., 16, p. 378 (1979).

Chamberlain, R.M. , C. Lemarechal, H.C. Pedersen, and M.J.D. Powell, "The Watchdog Technique for Forcing Convergence in Algorithms for Unconstrained Optimization," Math Prog. Study , 16, p.l (1982)

Colville, A.R., "A Comparative Study on Nonlinear Progrmming Codes," IBM New York Scientific Center Report No. 320-2949, (1968)

Daniel, J.M., "Stability of the Solution of Definite Quadratic Programs," Math Prog., 5, p. 41 (1973).

Dem'yanov, V.F. and V.N. Malozemov, Introduction to Minimax, Wiley, New York (1974)

Fletcher, R., "Methods Related to Lagrangian Functions," in Numerical Methods for Constrained Optimization (P.E. Gill and W. Murray, ed.), Academic Press, London (1974), p. 219.

Garcia-Palomares, U.M. and O.L. Kangasarian, "Superlinearly Convergent Quasi-Newton Algorithms for Nonlinearly Constrained Optimization, Math. Prog., 11, p. 1 (1976).

Gill, P.E. and W. Murray, "Numerically Stable Methods **for** QP," Math. Prog., 14, p. 349 (1978).

Han, S-P, "Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems," Math. Programming, LI, p. 263 (1976)

Han, S-P, "A Globally Convergent Method for Nonlinear Programming," JOTA, 22, 3, p. 297 (1977).

Han, S-P, "A Hybrid Algorithm for Constrained Nonlinear Programming," in Nonlinear Programming 3, Academic Press, London (1978).

Himmelblau, D.M. Applied Nonlinear Programming, McGraw-Hill, New York (1972).

Maratos, N., "Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems," PhD. Thesis, University of London, (1978).

Ortega, J.M. and Rheinboldt, W.C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, (1970).

Powell, M.J.D., "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," presented at 1977 Dundee Conference on Numerical Analysis (1977).

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

Rosen, E.K. and S. Suzuki, "Construction of Nonlinear Programming Test Problems," Comm. ACM, 8, p. 113 (1965).

Schittkowski, K. "The Nonlinear Programming Algorithm of Wilson, Kan and Powell with an Augmented Lagrangian Type Line Search Function," Num. Math, 38, p. 83 (1981)

Schuldt, S.B., "A Method of Multipliers for Mathematical Programming Problems with Equality and Inequality Constraints," JOTA, 17, p. 155 (1975)

Tapia, R.A., "Diagonalized Multiplier Methods," Math. Prog., 22, p. 135 (1977)

Wilson, R.B., "A Simplicial Algorithm for Concave Programming," Ph.D. Thesis, Harvard University (1963)

Yamashita, H., "A Globally Convergent Constrained Quasi-Newton Method with an Augmented Lagrangian Type Penalty Function," Math. Prog., 23, p. 75 (1982)

Acknowledgements

Financial support from NSF Grant CPE-8204366 is gratefully acknowledged. We also wish to thank the reviewers of the initial version of this paper for their helpful comments.