# AU3:IE'ITEO L1GSVMUN LINE SEARCHES FOR SUCCESSIVE 3UADHATIC PROGRA,4;1TrIG

by

Lor?nz T. Biogl^r

Dj^iinb^r, 1932

DRC-35^1-82

# AUGMENTED LAGRANGIAN LINE SEARCHES FOR SUCCESSIVE QUADRATIC PROGRAMMING

Lorenz T. Biegler

Department of Chemical Engineering
Carnegie-Mellon University
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 which causes a suitable decrease in an exact penalty or Lagrangian function. Because of some problems with these functions, a line search that uses an augmented Lagrangian is proposed.

This function follows quite naturally from the derivation of SQP methods and has superior properties when compared to an exact penalty or Lagrangian function. The augmented Lagrangian is applied successfully to three example problems that present difficulties for current line search techniques. Global and local convergence results are also presented which are valid for all positive penalty parameters in the augmented Lagrangian.

## 1• INTRODUCTION

The nonlinear programming problem can be written as

(1.1) **Mln** f(x)
x

8.t. g(x) £ 0

h(x) - 0

**for** $f:R^n$ •+ R

$g:R^n$ -• $R^m$

h:R -• R

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 of Garcia-Palomares and Mangasarian (1976). Here local convergence properties are proved 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: Min $\nabla f^T(x^i)d + \frac{1}{2} d \nabla d$

QCx1^1) $d$

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

$h(x^*) + \nabla hCx^1)^1\wedge - 0$

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

Here, $L(x,u,v) \bullet f(x) + u^T g(x) + v^T h(x)$ and the multipliers $(u,v)$ are found from the QP.

2) If IvfCx1)^! + |u4 g{x1)\ + jv1 hCx1)! $ € , stop. Here c > 0 is some small Kuhn-Tucker tolerance.

3) Else, find a stepsize X such that

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

vhere

$$(1.2) \qquad P(x) = f(x) + X^r j^8 j^{(x)} + {}^+ I^8 j^{\prime h} j^{(x)} l \quad f$$

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

P^x1) « Vf(x)Td - ^rjgj(x)+ - J Sjlh^x)! is an approximatlon to the directional derivative of P(x),

and 6 c (0,l)

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

$$r] - s] > | |\bar{u},\bar{v}| |.$$

where $\bar{u},\bar{v}$ are the multipliers at the K-T point of (!•!).

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

$$r^{\circ} - a^{0} - 0$$

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

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

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

Chamber1in (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 matrix to become unbounded. However, Chamberlin et al. (1979) showed that use of Han's penalty function causes convergence to be. too slow in some cases.

To resolve some of these difficulties, Chamberlin et al. (1979) proposed the watchdog technique. Here, an improvement in either the Lagrangian or exact penalty function is usually accepted during the line search. However, if the exact penalty function does not decrease monotonically, only the exact penalty is used in succeeding line searches. The authors present superlinear convergence results for this method and show the advantages of this method over Han's function on a small example problem. The watchdog technique, however, is harder to implement than the Han or Powell algorithms and requires more work, in general, during the line search step.

The next section discusses the use of the augmented Lagrangian function for the line search step. We will show that this follows quite naturally from the development of earlier quasi-Newton and augmented Lagrangian algorithms. Desirable properties, especially in the equality constrained case, are also discussed. The third section presents global and local convergence results for this function by paralleling the global results of Han and recent local results of Boggs et al. (1982). In the fourth section, the three above-mentioned problems of Chamberlin (1979) and Chamberlin et al. (1979) are solved to illustrate the advantages of augmented Lagrangian over the Han and Powell implementations. The last section summarizes the results of the previous sections and suggests further work in choosing the penalty parameter for the augmented Lagrangian function.

## 2. AUGMENTED LAGRANGIAN LINE SEARCHES

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

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

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:

$$(2.2) \qquad L_a(x,v,\alpha) = f(x) + h(x)^T v + \frac{\alpha}{2}\, h(x)^T h(x)$$

we find that the augmented Lagrangian has the same stationary point as the Lagrange function, regardless of the value of $\alpha$, the penalty parameter.

This stationary point can be found by applying a Newton method to $\nabla L_a(x,v,\alpha)$.

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:

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

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

••(IfcaiD

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

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

Let $(x-x^i)$ be defined as a search direction 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:

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

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

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

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

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

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

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

Note that the solution to (2.6) is independent of a (Han (1978), Fletcher (1974)). Since $\nabla_{xx}L(x^i,v^i)$ involves calculation of second derivatives we approximate this matrix by B, 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 stepwise X along the direction, d, 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:

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

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

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

the stepsize along d can be found by minimizing a modified augmented Lagrangian function:

$$(2.7) \qquad L(x,u,v,\alpha) = f(x) + u^T g(x)_+ + v^T h(x) + \frac{\alpha}{2} \|g(x)_+, h(x)\|^2$$

where

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

$u, v$ - multipliers for $g$ and $h$, respectively, found from $Q(x^i, B)$

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

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

It is similar to classical augmented Lagrange functions such as:

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

(see Bertsekas (1974)). However, classical techniques or multiplier methods are less efficient because they generally involve two nested iterations. The inner iteration minimizes the augmented Lagrangian for $x$ with $u$, $v$ and $\sigma$ fixed, 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 constraint violations.

On the other hand, by substituting (2.7) for the line search functions of Han and Powell, we have a multiplier method that simultaneously updates $x$, $u$ and $v$.

The new line search function has several other advantages as well. A major reason for performing a line search is to 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 in Section 4. 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 $\sigma$, we can make this penalty parameter as high as needed to approach or maintain feasibility in the line search.

Another important feature is the number of derivative discontinuities in the line search function. With $P(x,r,s)$ each active constraint has a discontinuous derivative at $g(x)$ or $h(x)=0$. With $L^*$ only the $u^Tg(x)+$

tern 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 especially true if equality

## 3. CONVERGENCE OF AUGMENTED LAGRANGIAN LINE SEARCHES

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

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

If $q_i$, 1=1,...**t** are continuously differentiable functions from $R^n$ • R and

$$\bullet(x) - \max_{i} Cq.(x)\}^{*}$$

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

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

**where**

$$I(x) = \{i | q(x) = \phi(x)\}$$

Theorem 3.1

Let f, g and h be continuously differentiable at x and B be a positive definite and symmetric nxn matrix. If (d,u,v) is a Kuhn-Tucker **triple** of Q(x,B) with d *4* 0, then:

$$D_d L^*(x, u, v, or) < 0$$

**Proof**

**Let**

$$I - (i:g(x) > 0\}$$
$$I - (i:g(x) - 0\}$$
$$\hat{I} - \{i:g(x) < 0)$$

**We write the directional derivative of $L^*$ as**

$$(3.1)\qquad D_d L^*(x,u,v;\alpha) = \nabla f(x)^T d + v^T \nabla h(x)^T d + \alpha\,(h^T \nabla h(x)^T d) + \sum_{i\in I} u_i \nabla g_i(x)^T d + \sum_{i\in \bar I} u_i \nabla g_i(x)_+^T d + \alpha \sum_{i\in I} g_i(x) \nabla g_i(x)^T d + \alpha \sum_{i\in \bar I} g_i(x)_+ \nabla g_i(x)_+^T d$$

For the set $\bar I$

$$g_i(x)_+ = 0 \quad , \quad \text{so} \quad \alpha \sum g_i(x)_+ \nabla g_i(x)_+^T d = 0$$

From the quadratic program

$$(3.2)\qquad B^T d + \nabla f(x) + \nabla g(x)\, u + \nabla h(x) v = 0$$

$$(3.3)\qquad u_i(g_i(x) + \nabla g_i(x)^T d) = 0$$

$$u \ge 0$$

$$(3.4)\qquad g(x) + \nabla g(x)^T d \le 0$$

$$(3.5)\qquad h(x) + \nabla h(x)^T d = 0$$

For the set $I$

$$u_i g_i(x) = -u_i \nabla g_i(x)^T d$$

but $$g_i(x) - g_i(x)_+ = 0$$

Hence $$u_i(\nabla g_i(x)^T d) = 0$$

Since
$$(\nabla g_i(x)_+^T d) = \max\,(0,\ \nabla g_i(x)^T d)$$
$$\nabla g_i(x)_+^T d = 0 \quad \text{from} \quad (3.4)$$

Equation (3.1) can be simplified to:

$$D_d L^*(x,u,v,\alpha) = \nabla f(x)^T d + v^T \nabla h(x)^T d + \alpha\,(h^T \nabla h(x)^T d) + \sum_{i\in I} u_i\ \nabla g_i(x)^T d + \alpha \sum_{i\in I} g_i(x) \nabla g_i(x)^T d$$

Substituting (3.2) into (3.1) yields

$$D_d L^*(x,u,v,\alpha) = -d^T B d - \sum_{i\in \hat{I}} u_i\ \nabla g_i(x)^T d + \alpha\,(h^T \nabla h(x)^T d) + \alpha \sum_{i\in I} g_i\ \nabla g_i(x)^T d$$

Substituting (3.3) and (3.5) into the above equation gives the inequality

$$D_d L^*(x,u,v,\alpha) \leqq -d^T B d + \sum_{i\in \hat{I}} u_i\ g_i(x) - \alpha\ ||h(x)||^2 - \alpha \sum_{i\in I} g_i(x)^2$$

Since B is positive definite $u \geq 0$, $d \neq 0$ and $g_i(x) < 0$, $i\epsilon\hat{I}$

$$D_d L^* \leqq -d^T B d + \sum_{i\in \hat{I}} u_i\ g_i(x) - \alpha\ ||g(x)_+,\ h(x)||^2 < 0$$

and the theorem is proved.

Corollary 3.1

If the search direction $d^i \neq 0$ then $\exists$ a $\lambda > 0$ such that

$$L^*(x^i + \lambda d^i, u^i, v^i, \sigma) - L^*(x^i, u^i, v^i, \sigma) \leq -\delta\lambda\Theta(x^i)$$

where we set $\delta = 0.1$

and $\Theta(x^i) = d^{iT} B d^i + \sigma \, \| g(x^i)_+, h(x^i) \|^2$

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

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

and let $\bar{d}$ minimize

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

$$\text{s.t.} \quad \bar{A}d \leq \bar{a}, \quad \bar{C}d = \bar{c}$$

Then for any fixed norm $\|\cdot\|$, there exist $s > 0$ and some $\gamma$ such that

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

If i) $\epsilon \leq \gamma$

ii) $B$ is positive definite

where

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

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

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

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

The bounds on the multipliers are therefore:

$$||(\hat{u}, \hat{v}) - (\bar{u},\bar{v})|| \le t\epsilon$$

whenever $\epsilon \le \bar{\epsilon}$ and $\epsilon = \max\{||\hat{H}-\bar{H}||,\ ||\hat{r}-\bar{r}||\}$

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

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

$$t > 0$$

Theorem 3.2

Let f, g and h be continuously differentiate and:

$$\alpha||x||^2 \le x^T B x \le \beta||x||^2 \quad ,\forall x$$

**for some $\alpha$, $\beta$ > 0**

Then any sequence $\{x_k\}$ generated by the algorithm either terminates at a Kuhn-Tucker point of (1.1) or any accumulation point $\bar{x}$, satisfying the Mangasarian-Fromowitz Constraint Qualification (MFCQ):

$$\left\langle \begin{array}{l} \exists z \in R^n \,|\, \nabla g_k(\bar{x})^T z < 0,\ k \in K \quad \text{where } K = \{k \,|\, g_k(x) = 0\} \\ \nabla h(\bar{x})^T z = 0 \\ \text{and } \nabla h(\bar{x})^T \text{ is linearly independent} \end{array} \right\rangle$$

is a Kuhn-Tucker point of (1.1).

**Proof**

If $d^i=0$, then $(x^1, u^1, v^1)$ satisfies Kuhn-Tucker conditions of (1.1) **and** we are done. Suppose $d^i \wedge 0$ then $x^{i+1} = x^i + Xd^i$ exists for some $X > 0$ such that

$$LV^{\wedge}.USV^1, «) * i/V.uSvSa) - 6X4$$

Let $\bar{x}$ be an accumulation point of $x^i$ satisfying the constraint qualification. If $B^i$ is formed by well-known quasi-Newton updates, then:'

$$\{x^1\} + \bar{x} \text{ and } \{B^1) \to \bar{B}$$

It follows from definition of $B^i$ and MFCQ that $Q(\bar{x},\bar{B})$ has a Kuhn-Tucker point, $\bar{d}$. If $\bar{d} = 0$ then $\bar{x}$ is a Kuhn-Tucker point of (1.1) **and** the theorem is proved. Suppose that $\bar{d} \wedge 0$. By Lemma 3.2 and the solution of $Q(\bar{x},\bar{B})$:

$$d^1 \to \bar{d} \qquad i$$

$$u^i \to \bar{u}$$

**Now** if $\bar{d}$ *4 0* then $\{X^1\}.- 0$ as $\{x^1\} + \bar{x}$.

We know that if $L^*(x,u,v,a)$ is not .at an optimum, 3 **a X** > 0 **that** gives the descent direction and a decrease in $\bar{L}^*$.

**For** some $i \geq I$ we can find some $X^i > 0$

$$L^*(x^i + \lambda^i d^i,u$$

**and** because $\{X_i\}$ ↛ 0, ∃some $y > 1$ such that:

$$L^*(x^i + \gamma\lambda^i d^i, u^i, v^i, \alpha) > L^*(x^i, u^i, v^i, \alpha) - \delta\gamma\lambda\phi(x^i)$$

Writing a Taylor series expansion on the last inequality gives

$$\gamma\lambda^i \nabla_x L^*(x^i, u^i, v^i, \alpha)^T d^i + \gamma\lambda^i O(\lambda^i) > -\delta\gamma\lambda\phi(x^i)$$

or
$$\nabla_x L^*(x^i, u^i, v^i, \alpha)^T d + O(\lambda^i) > -\delta\phi(x^i)$$

From Theorem 3.1 we know that

$$\nabla L^*(x^i, u^i, v^i, \alpha)^T d^i \leq \text{^}(x^i)$$

Since $d^1$ *4* 0, $d^l$ • $\bar{d}$ and $X^1$ • 0, for i $\geq$ I we write:

0 > -0.6rfcx$^1$)& Ty^xSuSv$^1$,*)** + OCX$^1$) > $-\delta\phi(x^1)$

or 0 > -(0.6-6) ^(x$^1$) a ^L^x$^1$, *$^1$, *$^1$.")$^1$*$^1$ + ^(x$^1$) + ©'(X$^1$) > 0

which is a contradiction. Therefore $\bar{d}$ = 0 and the theorem is proved.

Local Convergence

This subsection relies heavily on recent convergence results of Boggs et al. (1982). Here we assume that in a neighborhood around the solution:

a) | |3C*–āc| | $\leq$ c for some e > 0

b) the active constraint set (g,h) remains fixed in this neighborhood

c) $\nabla_{xx} L(\bar{x}, \bar{u}, \bar{v})$ is positive definite.

In addition we assume that the well-known quasi-Newton updates, such as BFGS, DFP, etc. have the property (see Boggs et al.):

$$\text{(3.6)} \qquad \lim_{i\to\infty} \frac{\|(B^i - \nabla_{xx}L(\bar{x},\bar{u},\bar{v}))d^i\|}{\|d^i\|} = 0$$

are used for $B$.

It is only necessary to show that $x = \bar{x}$ in the neighborhood of $\bar{x}$ when the local results of Boggs et al. (1982) can be applied.

Lemma 3.3

$$\|h(x^{i+1})\| = O(\|d^i\|)^2$$
$$\|g(x^{i+1})\| = O(\|d^i\|)^2$$

(where $g(x)$ are active inequality constraints)

if $\nabla^2 h(x)$ and $\nabla^2 g(x)$ is finite, $\forall x$.

Proof

from $Q(x^i,B)$:

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

and

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

Thus $\|h(x^{i+1})\| = O(\|d^i\|^2)$

By the same analysis $\|g(x^{i+1})\| = O(\|d^i\|^2)$.

**Theorem 3.3**

As $d^i \to 0$ in some neighborhood about $\bar{x}$, we have

$$L^*(x^{i+1},v^i,\alpha) - L^*(x^i,v^i,\alpha) \le -\delta\phi(x^i)$$

$$\text{where} \qquad x^{i+1} = x^i + d^i$$

for some $i \in I$.

**Proof**

We can write

$$\begin{aligned}
L^*(x^{i+1},u^i,v^i,\alpha) - L^*(x^i,u^i,v^i,\alpha) = \\
L(x^{i+1},u^i,v^i) - L(x^i,u^i,v^i) \\
+ \frac{\alpha}{2} \left(\|g(x^{i+1})_+, h(x^{i+1})\|^2 - \|g(x^i)_+, h(x^i)\|^2\right) \\
= \nabla_x L(x^i,u^i,v^i)^T d^i + \frac{d^{iT} \nabla_{xx} L(x^i,u^i,v^i) d^i}{2} + o(\|d^i\|^3) \\
- \frac{\alpha}{2} \|g(x^i)_+, h(x^i)\|^2
\end{aligned}$$

by Taylor series expansion and Lemma 3.3. From the QP we know:

$$\nabla_x L(x^i,u^i,v^i)^T d^i = -d^{iT} B^i d^i$$

So

$$\begin{aligned}
L^*(x^{i+1},u^i,v^i,\alpha) - L^*(x^i,u^i,v^i,\alpha) = \\
- \frac{d^{iT} B^i d^i}{2} - \frac{d^{iT}(B^i - \nabla_{xx} L(x^i,u^i,v^i)) d^i}{2} \\
- \frac{\alpha}{2} \|g(x^i)_+, h(x^i)\|^2 + o(\|d^i\|^3) \\
= - \left( \frac{d^{iT} B^i d^i}{2} + \frac{\alpha}{2} \|g(x^i)_+, h(x^i)\|^2 \right) \\
+ o(\|d^i\|^2)
\end{aligned}$$

by (3.6).

Thus, for some $i \in I$

$$-f(x^i) - -0.1\,(d^{iT} B^i d^i + \alpha\|h(x^i),\ g(x^i)_+\|^2)$$
$$\geq -\frac{1}{4}\,(d^{iT} B^i d^i + \alpha\|h(x^i),\ g(x^i)_+\|^2)$$
$$\geq L^*(x^{i+1}, u^i, v^i, \alpha) - L^*(x^i, u^i, v^i, \alpha)$$

and the theorem is proved.

Finally we state two theorems by Boggs et al. (1982) that prove Q-superlinear local convergence. Here we use the additional assumption that $(B^i - \nabla_{xx}L(\bar{x},\bar{u},\bar{v}))$ projected into the null space of $Z^T = \begin{bmatrix} \nabla g_P(\bar{x}) \\ \nabla h(\bar{x}) \end{bmatrix}^T$ is bounded, or

$$(3.7) \qquad \|P(\bar{X})(B^i - \nabla_{xx}L(\bar{x},\bar{u},\bar{v}))\| \leq \xi \qquad \text{(xl)}$$

for some $\xi > 0$

where $P(\bar{x}) = I - Z(Z^T Z)^{-1} Z^T$, the projection matrix.

<u>Theorem 3.4</u>

Let $B^i$ satisfy $\|B^i\| \leq \eta$ for some $\eta > 0$. If

$\|x^1 - \bar{x}\| < c$ for some $c > 0$
and if (3.7) holds for $\forall\, i \in I$,

then the sequence $\{x^i\}$ generated by successive quadratic programming and $\lambda=1$ is well defined and converges linearly to $\bar{x}$.

<u>Proof</u>

(See Theorem 3.3 in Boggs et al. (1982).)

Theorem 3«5

If $\nabla_{xx}L(\bar{x},\bar{u},\bar{v})$ is positive definite and the $B^i$ are obtained by updating formulas, such as DFP or BFGS, that possess the property (3.6) and if $\{x^i\}$ converges linearly to $\bar{x}$, then the convergence is Q-superlinear.

Proof

(See Theorem 3.2 in Boggs et al. (1982).)

## 4. EXAMPLE PROBLEMS

In this section we demonstrate the superiority of an augmented Lagrangian line search on three example problems. The first problem was given by Chamberlin et al (1979) to show the slow convergence of Han's exact penalty function. In that example, using the watchdog technique leads to faster convergence.

In the last two examples, due to Chamberlin (1979), cycling is observed when Powell's line search function is used. The cycling problems are caused by violations of nonlinear constraints that the quadratic programming step determined to be inactive. Cycling in the last example also causes the Hessian matrix and quadratic programming shadow prices to become unbounded.

On all three examples, using the augmented Lagrangian line search function eliminates the problems encountered with other line search functions.

a. Slow convergence with exact penalty function line search (see Chamberlin et al. (1979)).

Consider the problem

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

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

Let $x^k = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ and assume

no errors are made in approximating the matrix

$\nabla_{xx}L(x^*,v^*) = I$ . $x^{k+1}$ is therefore $\begin{bmatrix} \cos\theta + \sin^2\theta \\ \sin\theta[1-\cos\theta] \end{bmatrix} = x^k + d^k$

As mentioned by Chamberlin et ai. (1979)$_f$ it follows that for any small positive number c,

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

can be made less than e if 6 is chosen sufficiently small*.

From the definition of P(x)

$$p(x^k) \text{ « } -\cos e$$

$$P(x^{V4-1}) - -\cos e - \sin^2 e + (T + s) \sin^2 6$$

For reduction to occur in $P(x^{k+1})$,

(4.1) $$T - I - 8 < 1$$

Likewise the Lagrangian and augmented Lagrangian functions are given by:

$$L(x^k{}_t v) \bullet -\cos e$$

$$L(x^{k+1}{}_f v) * -\cos e - \sin^2 e + \left(\frac{\cos e \quad \sin^2 e}{2}\right)$$

$$L^*(x^k{}_f v_f or) \text{ » } -\cos e$$

$$L^*(x^{k+1},v) - -\cos e - \sin^2 e + \left(\frac{\cos 9 \quad 8\ln 2e}{2}\right) + I \text{ »}\sin^4 e$$

**and** the multiplier on the constraint $c_1$ is:

$$v = \frac{\cos e}{2} \text{ " } \bar{T}$$

Since s £ |v| £ 0 for either the Powell or Han line search functions, $P(x^{k**})$ does not decrease at $x^{k**}$ if $t \geq 1$. It can easily be seen that the Lagrangian decreases at $x^{k+1}$ and the augmented Lagrangian also decreases as long as

$$cr< \frac{2(1 - \frac{\cos\theta}{2})}{(1 - \cos^2 6)}$$

(Note that this upper bound is <u>always</u> greater than 1. Also as 8 • 0 ($x^k$ is arbitrarily close to $x^*$) then for a reduction in $L^*$ to occur at $x^{k+1}$ o may take any positive value*)

Moreover, if x is too large the penalty function may not even be minimized at $x=x^*$• If we set $x^h$ at $\begin{bmatrix} h \\ 0 \end{bmatrix}$ with $|h| \leq 1$ but arbitrarily close to 1.

$$P(x^{h}) \ll . _h + (s - \tau)(1 - h^-)$$

and a derivative discontinuity occurs at h=l. Hence,

$P(x^h)$ is minimized at h « 1 when

- 1 - 2h(s - T) £ 0

or

(4.2) 0 fc T - 1/2

Note that the two bounds on s [(4.1) and (4.2)] are inconsistent if x > 3/4. Since the optimum is independent of the parameter T we also need a line search that is unaffected by x.

Both the Lagrangian and augmented Lagrangian functions are unaffected by T. Starting at $x^h = \begin{bmatrix} h \\ 0 \end{bmatrix}$ and using

B « I, we have $$d = \frac{(1-h^2)}{2h}$$

and $$v = -\frac{\wedge}{4h^2}[h^2 + 2h - 1] - T$$

**The functions are therefore:**

$$L(h,v) = -h + \frac{K}{4h^2}[h^2 + 2h - 1]\ (h^2 - 1)$$

$$L^*(h,v,a) = -h + \frac{1}{4h^2}[h^2 + 2h - 1]\ (h^2 - 1) + \frac{f}{}\ (h^2 - 1)^2$$

**Thus, both functions are minimized at h=l ($x^h = x^*$) for <u>any</u> a in $L^*$.**

**This problem illustrates how the Lagrangian and augmented Lagrangian line searches are independent of T and converge to $x^*$. Hence the convergence problems of exact penalty functions are avoided. The next two problems illustrate how the augmented Lagrangian line search (for appropriate values of a) prevents cycling due to changing active sets of nonlinear constraints.**

**b. The Problem Min $x_2$**

$$\text{s.t.}\quad c_1(x) = a(x_1) - x_2 \geq 0$$

$$c_2(x) = a(1-x_1) - x_2 \geq 0$$

$$\text{where}\quad a(x) = 2x^2 - x^3$$

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

$$c_1^0 = 0 \quad \ell 0 . a$$

$$c_2^0 = 1$$

The solution to the quadratic program gives $u^0 = \begin{bmatrix} 0 \\ \pm \end{bmatrix}$ $d = x^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $r^0 = \begin{bmatrix} 0 \\ \cdot \end{bmatrix}$ $\bar{1}$

$$| \quad | \; \bullet \; | \; \bullet \begin{bmatrix} \cdot \\ \cdot \end{bmatrix} \| \; \|$$

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

$$L(x^0,u^0) = 1 \qquad L'(x^0,u^0) = -1$$

$$L^*(x^0,u^0,\alpha) = 1 + \frac{\alpha}{2} \qquad L^{*\prime}(x^0,u^0,\alpha) = -(\alpha + 1)$$

where $\bullet'$ is the directional derivative $\nabla_x \cdot^T d$.

The Armijo inequality holds for $P(x^0,r^0)$ and $L(x^0,u^0)$ for the step-size, $\lambda = 1$

$$P(x^1,r^0) \le P(x^0,r^0) + \delta\, P'(x^0,r^0)$$

$$L(x^1,u^0) \le L(x^0,u^0) + \delta\, L'(x^0,u^0)$$

both become $\quad 0 \le 1 - \delta$

where $\delta$ is set to 0.1 by Powell (1977). The augmented Lagrangian line search does not satisfy the inequality

$$L^*(x^1,u^0,\alpha) \le L^*(x^0,u^0,\alpha) + \lambda\, L^{*\prime}(x^0,u^0,\alpha)$$

(which is $\left| \wedge 1 + \right| - \delta(\alpha + 1) \Big)$ ,

as long as $\quad a > 1/\lambda - 1$ .

Similarly, during the next iteration with

. • - [ ; ] . . • ⟩⟩ lins unchanged at I

$x^2$ - m -ra Ha na

(same as starting pt.)

| | | |
|---|---|---|
| P^.r$^1$) - 1 | L<xV) - 1 | LV . U$^1$) -1 + f |
| 2 1 . | 2 1 | * 2 1 |
| P(x ,r ) - 1/2 | L(x ,u ) - 0 | L (x .« ) - a/2 |
| P'O^.r$^1$) - - 1 | L'Cx$^1$,!!$^1$) - - 1 | L*'^$^1$,*$^1$) - - (1 + a) |

The Armijo inequality holds for P(x,r) and L(x,r) at *X* – 1,

$$P(x^2,r^1) \le P(x^1,r^1) + \delta P'(x^1,r^1)$$

(1/2·« 1 - 6 for 6 « [0,1/2])

L($x^2$.«V* UX$^1$,!$^1$) + 6 L'(X$^1$,^)

( O i l - ! for 6 a [0,l])

but does not hold for $L^*(x,u,a)$:

$$L^*(x^2,u^1,\alpha) \le L^*(x^1,u^1,\alpha) + \delta L^{*'}(x^1,u^1,\alpha)$$

$\left(\frac{\alpha}{2} < 1 + f - 8(1 + or)\right)$

if a > 1/6-1. Consequently, for an appropriate choice of a, the augmented Lagrangian will prevent cycling between the above two infeasible points.

**Setting** $B^0=1$, $x^0=0$, 6=0.1, the penalty parameter, a, at 10, **and** minimizing a quadratic function along the search direction, this optimization problem converges to the optimum [x «( * r^0 .5 .375. L0. ) in four iterations, if augmented Lagrangian line searches are used. Similar results **are** obtained with a=100. Table 1 lists the iterations for both values of a. As seen from the table, performance in both cases is quite efficient.

c. In this problem, cycling between two infeasible points also causes the Hessian and the multiplier to become unbounded.

**Kin** y **(constant)**

s.t. $c_1(x)$ • -b(x) £ 0

**$c_2(x)$ - -b(l-x) £ 0**

**vhere b(x) « $-1+x+zx-zx^3$**

**for z 2 4**

Beginning at $x^0 = 0$ and $B^0 = 0$, we have

•Hi) •'•[:] •-•[:] **d = 1**

L*<x°,u°,or) - & + a/2 P(x°,r°) - 0

L'V . U°, * ) - a/2 P( ,f°) = 0

L*'(x°,u°,a) - - 0 + of) $P^f$(x°,r°) = - β

L(x* μ° ) - P

L(x¹ ^°) - 0

L'(x⁰,«°) - - P

For the penalty and Lagrangian functions, the line search is satisfied at d = 1, (X = 1) *

$$L(x^1, u^\circ) * L(x^\circ,u^\circ) + 6\, L'(x^0,u^\circ)$$
$$P(x^1, r^\circ) * P(x^\circ,r^\circ) + 6\, P'(x^0,u^\circ)$$
$$(0 * P(1-6)) \quad \text{for} \quad 6 \in [0,1]$$

For the augmented Lagrangian, the inequality

$$L^*(x^1, u^0, \alpha) * L^*(x^0,u^0,\alpha) + 6\, L^{*\prime}(x^0,u^0,\alpha)$$
$$(\alpha/2\ a: p + \alpha/2 - 6(P + a))$$

is not satisfied, if a is chosen > (1-6)6/6. Similarly for the second iteration, at x=l

$$c^1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad B^1 = \beta z \qquad u^1 = \begin{bmatrix} 0 \\ \beta z \end{bmatrix} \qquad r^1 = \begin{bmatrix} \beta/2 \\ \beta z \end{bmatrix}$$

$$d = -1 \quad , \quad x^2 - x^\circ - 0$$

$$P(x^1,r^1) - pz \qquad\qquad M(x^1,u^1) - pz$$
$$P(x^2,r^1) - P/2 \qquad\qquad L(x^2,u^1) - 0$$
$$P'(x^1,r^1) - - Pz \qquad\qquad L'(x^1,u^1) - - pz$$
$$L^*(x^1,u^1,\alpha) = \beta z + \alpha/2$$
$$L^*(x^2,u^1,\alpha) - a/2$$
$$L^{*\prime}(x^1,u^1,\alpha) - - (pz + \alpha)$$

and the line search using a penalty or Lagrange function is satisfied at the initial point of the cycle. Again, the augmented Lagrangian line search breaks this cycle if the penalty parameter

$$a > (1 - 6)\, pz/6 \; .$$

Setting $\beta=1$, $z=4$, $y=0$, $x^0=0$, $\delta=0.1$ and the penalty parameter, $\alpha$, to 10, and minimizing a quadratic along the search direction, the problem converges to the optimum ($x=0.5$) in five iterations, if augmented Lagrangian line searches are used. With $\alpha=100$ only four iterations are required for convergence. The results of both cases are given in Table 2 and show efficient performance.

## 5. CONCLUSIONS

With the augmented Lagrangian line search the difficulties of previously implemented line search functions appear to be eliminated. As shown in the example problems, this line search function does not retard the superlinear convergence of SQP methods, as Han's (and Powell's) penalty function can, and the augmented Lagrangian accurately penalizes infeasible nonlinear constraints that are left inactive by the quadratic program.

Using the augmented Lagrangian line search with SQP has been shown in Section 2 to be an efficient implementation of classical multiplier methods. Here the variables and multipliers are updated simultaneously. It should be mentioned that although multiplier methods may require many unconstrained minimizations, they do have parametrically superlinear convergence properties.

In addition the augmented Lagrangian has fewer derivative discontinuities than the Han or Powell penalty functions. Especially in the equality constrained case, this means that better quadratic approximations are possible during a line search and more*"exact[91] line search minimizations will be performed.

In Section 3 we presented local and global convergence results that do not depend on o as long as a £ 0. Also, since the QP solution $d^{\ell}$ is independent of a, a large value of a does not affect the direction of the algorithm. The example problems in Section 4, however, require some bound on a which is problem dependent but generally quite reasonable. A heuristic that may be employed in choosing a would be to set

$$a^{1+1} - a\hat{\ }lO^{0}) \quad \text{for some } \upsilon \geq 1$$

$$\text{if} \quad ||g(x^{1+1}), h(x^{1+1})|| * H.C^{*1})^{*}. hOc^{1})11 + C .$$

Here, $C \geq 0$ is chosen so that SQP converges superlinearly when the constraint violations are small, but prevents much larger constraint violations from occurring. Controlling this parameter provides.a tradeoff between achieving good convergence close to the solution and approaching the optimum efficiently when large initial constraint violations are present.

Acknowledgments

The author wishes to thank Professors J.M. Borwein and I.E. Grossmann for their discussion* encouragement and helpful suggestions regarding this paper. Financial support from the Robotics Institute of Carnegie-Mellon University and the National Science Foundation under Grant CPE-8204366 is gratefully acknowledged.

## 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. Tolie and P. Wang, "On the Local Convergence of Quasi-Newton Methods for Constrained Optimization," SIAM J. Control & Opt., £0, 2, p. 161 (1982).

Chamberlin, R.M., "Some Examples of Cycling in Variable Metric Methods for Constrained Minimization," Math. Prog., 1£, p. 378 (1979).

Chamberlin, R.M., C. Lemarechal, H.C. Pedersen, and M.J.D. Powell, "The Watchdog Technique for Forcing Convergence in Algorithms for Unconstrained Optimization," presented at 10th International Symposium on Math. Prog., Montreal (1979).

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

**Dem<sup>v</sup>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. Mangasarian, "Superlinearly Convergent Quasi-Newton Algorithms for Nonlinearly Constrained Optimization, **Math.** Prog., U., p. 1 (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).

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

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

TABLE 1

Problem b Solved vith Augmented Lagrangian Line Searches

$\alpha = 10$

| $i$ | $x_1^i$ | $x_2^i$ | $c_1(x^i)$ | $c_2(x^i)$ | $L^*(x^i,u^i,\alpha)$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1.0 | 6.0 |
| 2 | 0.55 | 0 | 0.4386 | 0.3139 | 1.6775 |
| 3 | 0.4998 | 0.3737 | $1.0086\cdot10^{-3}$ | $1.5116\cdot10^{-3}$ | 0.37496 |
| 4* | 0.5000 | 0.3750 | $1.49\cdot10^{-8}$ | $1.49\cdot10^{-8}$ | 0.375 |

$\alpha = 100$

| $i$ | $x_1^i$ | $x_2^i$ | $c_1(x^i)$ | $c_2(x^i)$ | $L^*(x^i,u^i,\alpha)$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1.0 | 51.0 |
| 2 | 0.5050 | 0 | 0.38126 | 0.36876 | 14.583 |
| 3 | 0.500 | 0.37499 | $1.225\cdot10^{-5}$ | $1.275\cdot10^{-5}$ | 0.375002 |
| 4 | 1/2 | 3/8 | 0 | 0 | 3/8 |

*Within specified KKT tolerance ($10^{-6}$)

(Results obtained using double precision on a DEC-20 computer. The "O's" in the iterations denote no significant digits remaining.)

**TABU 2**

<u>**problem** c Solved vlth Augmented Lagrangian Line Searches</u>

Of - 10

| i | $x^i$ | $c_1(x^i)$ | V*1) | $L^*(x^i,u^i,\alpha)$ |
|---|---|---|---|---|
| 1 | 0 | 1.0 | 0 . | 6.0 |
| 2 | 0.55 | -.0945 | 0.1045 | 0.55065 |
| 3 | 0.5018 | .00367 | 0.00369 | 6.834-10"5 |
| 4 | * 0.5000 | -6.72-Hf6 | 6.72'UT6 | 2.259 - lO"10 |
| 5 | 1/2 | 0 | 0 | 0 |

*a* - 100

| 1 | $x^1$ | cx(xS | $c_2(x^i)$ | $L^*(x\backslash u_>$ |
|---|---|---|---|---|
| 1 | 0 | 1.0 | 0 | 51.0 |
| 2 | 0.505 | $9.949 \cdot 10^{-3}$ | $1.0049 \cdot 10^{-2}$ | $5.0546 \cdot 10^{-3}$ |
| 3 | 0.5000 | $-4.85 \cdot 10^{-5}$ | $4.85 \cdot 10^{-5}$ | $1.1764 \cdot 10^{-7}$ |
| 4 | 1/2 | 0 | 0 | 0 |

(**Results** obtained using double precision on **a** DEC-20 **computer. The last Iteration shows no significant** digits remaining for 0.)