INTERNAL REPORTS IN

# SIMULATION, OPTIMIZATION AND CONTROL

No. SOC-284

## NONLINEAR PROGRAMMING USING LAGRANGIAN FUNCTIONS

J.W. Bandler

December 1981

FACULTY OF ENGINEERING
McMASTER UNIVERSITY
HAMILTON, ONTARIO, CANADA

# NONLINEAR PROGRAMMING USING LAGRANGIAN FUNCTIONS

J.W. Bandler

Abstract

A brief review of major features of nonlinear programming methods which employ Lagrangian functions is presented. Following statements and discussion of necessary and sufficient conditions for a solution, the augmented Lagrangian method is described. This method adds a penalty term to permit a sequence unconstrained optimizations to be applied. The motivation behind the formulation and a discussion of Newton and quasi-Newton approaches is given. The Han-Powell algorithm is subsequently presented. This algorithm employs a quadratic approximation to the objective function, describes linearized constraints which lead to a quadratic program to be solved. The results provide the next search direction and appropriate Lagrange multipliers. After a one-dimensional search, the second derivative approximation is updated by a BFGS formula, with steps taken to ensure positive definiteness.

---

This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant A7239.

The author is with the Group on Simulation, Optimization and Control, and the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada L8S 4L7.

Acknowledgement

This text is based principally on lectures given by K. Madsen [1], as well as on papers by Powell [2,3] and a chapter by Fletcher [4].

The Problem

In this review, we address the problem of minimizing w.r.t. $\phi$ the nonlinear differentiable function $U(\phi)$ subject to the nonlinear differentiable constraints

$$\begin{aligned} c_i(\phi) &= 0, \quad i = 1, \ldots, t, \\ c_i(\phi) &\geq 0, \quad i = t+1, \ldots, m. \end{aligned} \tag{1}$$

Necessary Condition for a Solution

Theorem Let $\check{\phi}$ minimize $U(\phi)$ subject to (1) and assume the gradient vectors $a_i(\check{\phi})$ of active constraints $c_i(\check{\phi})$ are linearly independent. Then there exist real multipliers

$$\lambda_i, \quad i = 1, \ldots, m,$$

with

$$\lambda_i \geq 0, \quad i = t+1, \ldots, m, \tag{2}$$

such that the gradient vector $g(\check{\phi})$ of the objective function is given by

$$g(\check{\phi}) = \sum_{i=1}^{m} \lambda_i \, a_i(\check{\phi}), \tag{3}$$

with

$$\lambda_i \, c_i(\check{\phi}) = 0, \quad i = 1, \ldots, m. \tag{4}$$

If the constraints are linear, namely, if

$$c_i(\phi) = a_i^T \phi + b_i, \quad i = 1, \ldots, m, \tag{5}$$

where $a_i$, $i = 1, \ldots, m$, are constant vectors, then the linear independence condition on the constraint functions is unnecessary.

Descent Direction for Linear Equality Constraints

Suppose the necessary condition for a problem with linear equality constraints is not satisfied at some point $\phi$. Then

$$g(\phi) \notin V \overset{\Delta}{=} \{\sum_{i=1}^{t} \lambda_i \, a_i | \lambda_i \in R\}, \tag{6}$$

where R denotes the set of real numbers. Consider a direction $s$ such that

$$a_i^T s = 0, \quad i = 1, \ldots, m. \tag{7}$$

Then

$$\begin{aligned} g^T s &= (v + g - v)^T s \\ &= v^T s + (g - v)^T s \\ &= 0 - s^T s < 0, \end{aligned} \tag{8}$$

where we take

$$v \in V \tag{9}$$

and

$$g = v + s, \tag{10}$$

implying that a descent direction exists.

Implication of the Necessary Condition for Linear Constraints

Suppose the first s constraints are active. Then

$$\begin{aligned} a_i^T \phi + b_i &= 0, \quad i = 1, \ldots, s, \\ a_i^T \phi + b_i &> 0, \quad i = s+1, \ldots, m. \end{aligned} \tag{11}$$

Consider a feasible direction $s$ such that $U(\phi + \alpha s) < U(\phi)$ for small $\alpha >$ 0. Then

$$\begin{aligned} a_i^T(\phi + \alpha s) + b_i &= 0, \quad i = 1, \ldots, t, \\ a_i^T(\phi + \alpha s) + b_i &> 0, \quad i = t+1, \ldots, s. \end{aligned} \tag{12}$$

This gives

$$\begin{aligned} \underset{\sim}{a}_i^T \underset{\sim}{s} &= 0, \; i = 1, \ldots, t, \\ \underset{\sim}{a}_i^T \underset{\sim}{s} &\geq 0, \; i = t+1, \ldots, s. \end{aligned} \tag{13}$$

A downhill direction satisfies

$$\underset{\sim}{g}^T \underset{\sim}{s} < 0. \tag{14}$$

Let

$$\underset{\sim}{g} = \sum_{i=1}^{s} \lambda_i \underset{\sim}{a}_i, \quad \lambda_i \geq 0 \text{ for } i > t. \tag{15}$$

Suppose $\underset{\sim}{s}$ is feasible. Then

$$\underset{\sim}{g}^T \underset{\sim}{s} = \sum_{i=1}^{s} \lambda_i \underset{\sim}{a}_i^T \underset{\sim}{s} = \sum_{i=t+1}^{s} \lambda_i \underset{\sim}{a}_i^T \underset{\sim}{s} \geq 0. \tag{16}$$

Hence, no downhill direction exists, i.e., (14) cannot be satisfied when (15) is satisfied.

Multiplier Signs for Inequality Constraints

Consider $\bar{\underset{\sim}{\phi}}$ to be a candidate for a minimizing point. Find the active constraints $s \geq t$ such that

$$\begin{aligned} \underset{\sim}{a}_i^T \bar{\underset{\sim}{\phi}} + b_i &= 0, \quad i = 1, \ldots, s, \\ \underset{\sim}{a}_i^T \bar{\underset{\sim}{\phi}} + b_i &> 0, \quad i = s+1, \ldots, m. \end{aligned} \tag{17}$$

Then $\bar{\underset{\sim}{\phi}}$ must be a minimum w.r.t. $\underset{\sim}{a}_i^T \bar{\underset{\sim}{\phi}} + b_i = 0, \; i = 1, \ldots, s$. This implies that

$$\underset{\sim}{g}(\bar{\underset{\sim}{\phi}}) = \sum_{i=1}^{s} \lambda_i \underset{\sim}{a}_i \tag{18}$$

using the theorem for equality constraints. We need to show in addition that $\lambda_i \geq 0$, $i = t+1, \ldots, m$. Suppose, therefore, that $\lambda_k < 0$ for $k > t$. Let

$$V \triangleq \{ \sum_{\substack{i=1 \\ i \neq k}}^{s} \mu_i \, \underset{\sim}{a}_i \mid \mu_i \in R\} \tag{19}$$

with $\underset{\sim}{a}_1, \ldots, \underset{\sim}{a}_s$ linearly independent. Then $\underset{\sim}{a}_k \notin V$. We now show that $\underset{\sim}{s}$ exists which is downhill and feasible. Consider $\underset{\sim}{s}$ such that

$$\underset{\sim}{a}_i^T \underset{\sim}{s} = 0 \quad \text{for} \quad t + 1 \leq i \leq s, \; i \neq k \tag{20}$$

and let

$$\underset{\sim}{a}_k = \underset{\sim}{v} + \underset{\sim}{s}, \tag{21}$$

where

$$\underset{\sim}{v} \in V. \tag{22}$$

Then

$$\underset{\sim}{a}_k^T \underset{\sim}{s} = (\underset{\sim}{v} + \underset{\sim}{s})^T \underset{\sim}{s} = \underset{\sim}{s}^T \underset{\sim}{s} > 0. \tag{23}$$

Hence, $\underset{\sim}{s}$ is feasible since all constraints are satisfied. Now

$$\underset{\sim}{g}^T \underset{\sim}{s} = \sum_{i=1}^{s} \lambda_i \, \underset{\sim}{a}_i^T \underset{\sim}{s} = \lambda_k \, \underset{\sim}{a}_k^T \underset{\sim}{s} < 0 \tag{24}$$

if $\lambda_k < 0$, hence the direction $\underset{\sim}{s}$ is also downhill. Therefore, $\lambda_i \nless 0$, $i = t + 1, \ldots, m$ if $\bar{\underset{\sim}{\phi}}$ is a minimizing point.

## Sufficient Condition for a Solution

Define the Lagrangian function

$$L(\underset{\sim}{\phi}, \underset{\sim}{\lambda}) = U(\underset{\sim}{\phi}) - \sum_{i=1}^{m} \lambda_i \, c_i(\underset{\sim}{\phi}) \tag{25}$$

and consider the first s constraints to be active.

**Theorem** Let

$$\underset{\sim}{g}(\check{\underset{\sim}{\phi}}) = \sum_{i=1}^{m} \lambda_i \, \underset{\sim}{a}_i(\check{\underset{\sim}{\phi}}), \tag{26}$$

$$\begin{aligned} \lambda_i &> 0, \quad i = t+1, \ldots, s, \\ \lambda_i &= 0, \quad i = s+1, \ldots, m. \end{aligned} \tag{27}$$

Let

$$\underset{\sim}{s}^T[\underset{\sim}{\nabla}_\phi \nabla_\phi^T L(\check{\underset{\sim}{\phi}},\underset{\sim}{\lambda})]\, \underset{\sim}{s} > 0 \tag{28}$$

for all $\underset{\sim}{s} \neq \underset{\sim}{0}$ satisfying

$$\underset{\sim}{s}^T \underset{\sim}{a}_i(\check{\underset{\sim}{\phi}}) = 0, \; i = 1, \ldots, s, \tag{29}$$

and let $\underset{\sim}{a}_i(\check{\underset{\sim}{\phi}})$, $i = 1, \ldots, s$ be linearly independent.

Then $\check{\underset{\sim}{\phi}}$ is a local minimum.

<u>Example</u> [1]

Consider the function of two variables

$$U(\underset{\sim}{\phi}) = 3\phi_2 + \phi_1^2 + \phi_2^2.$$

Then

$$\underset{\sim}{\nabla}_\phi \underset{\sim}{\nabla}_\phi^T U = \underset{\sim}{\nabla}_\phi \, [2\phi_1 \quad 3 + 2\phi_2]$$

$$= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

which is positive definite. Consider the inequality constraint

$$c(\underset{\sim}{\phi}) = \phi_1^2 + (\phi_2 + 1)^2 - 1 \geq 0.$$

Let

$$\underset{\sim}{\phi} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

be a candidate for a solution. Here

$$\underset{\sim}{g} = \underset{\sim}{\nabla}_\phi U = \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \lambda \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Let

$$\underset{\sim}{\phi} = \begin{bmatrix} \sin\theta \\ (\cos\theta) - 1 \end{bmatrix}$$

and consider $\theta = 0$. Then

$$\underset{\sim}{\phi} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

But

$$U(\underset{\sim}{\phi}) = 3((\cos\theta) - 1) + \sin^2\theta + ((\cos\theta) - 1)^2$$
$$= 3\cos\theta - 3 + \sin^2\theta + \cos^2\theta - 2\cos\theta + 1$$
$$= (\cos\theta) - 1.$$

Obviously, $\theta = 0$ does not provide the smallest value of the objective function.

Consider

$$\underset{\sim}{\nabla}_\phi \underset{\sim}{\nabla}_\phi L = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} - \frac{3}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

This is not positive definite. This example shows that the curvature of the Lagrangian needs to be considered.

Obviously, if all constraints are linear only the Hessian of U needs to be examined.

<u>The Augmented Lagrangian Method [3,5]</u>

<u>Algorithm for Equality Constraints</u>

Consider

$$c_i(\underset{\sim}{\phi}) = 0, \quad i = 1, \ldots, t,$$

and let

$$\Phi(\underset{\sim}{\phi}, \underset{\sim}{\lambda}, r) = U(\underset{\sim}{\phi}) - \sum_{i=1}^{t} \lambda_i c_i(\underset{\sim}{\phi}) + r \sum_{i=1}^{t} c_i^2(\underset{\sim}{\phi}). \tag{30}$$

<u>Step 1</u> Choose $\underset{\sim}{\phi}^0$, set $\underset{\sim}{\lambda}^1 \leftarrow \underset{\sim}{0}$, $r_1 \leftarrow 1$, $K_1 \leftarrow \max_i |c_i(\underset{\sim}{\phi}^0)|$, $j \leftarrow 0$.

<u>Step 2</u> Set $j \leftarrow j + 1$. Minimize $\Phi(\underset{\sim}{\phi}, \underset{\sim}{\lambda}^j, r_j)$ w.r.t. $\underset{\sim}{\phi}$ yielding $\underset{\sim}{\phi}^j$.

<u>Step 3</u> If $\max_i |c_i(\underset{\sim}{\phi}^j)| \geq K_j$ go to Step 6.

Step 4  Set $K_{j+1} \leftarrow \max_i |c_i(\underset{\sim}{\phi}^j)|$. If $K_{j+1} \leq \varepsilon$ stop.

Step 5  If $(K_{j+1} \leq K_j/4)$ or $(\underset{\sim}{\lambda}^{j-1} = \underset{\sim}{\lambda}^j$ and $j > 1)$
set $\underset{\sim}{\lambda}^{j+1} \leftarrow \underset{\sim}{\lambda}^j - \underset{\sim}{c}(\underset{\sim}{\phi}^j)/(2r_j)$, $r_{j+1} \leftarrow r_j$ and go to Step 2.

Step 6  Set $K_{j+1} \leftarrow K_j$, $\underset{\sim}{\lambda}^{j+1} \leftarrow \underset{\sim}{\lambda}^j$, $r_{j+1} \leftarrow 10\, r_j$. Go to Step 2.

Comment

The equation

$$\underset{\sim}{\nabla}_\phi \Phi = \underset{\sim}{g} - \sum_{i=1}^{t} \lambda_i \underset{\sim}{a}_i + r \sum_{i=1}^{t} 2c_i \underset{\sim}{a}_i = \underset{\sim}{0} \tag{31}$$

is a necessary condition for a minimum of $\Phi$ w.r.t. $\underset{\sim}{\phi}$ for given values of $\underset{\sim}{\lambda}$ and r. Consider the trajectory of

$$\check{\underset{\sim}{\phi}}(\underset{\sim}{\lambda}) \text{ s.t. } \underset{\sim}{\nabla}_\phi \Phi = \underset{\sim}{0}. \tag{32}$$

Here, we have expressed solution points in terms of $\underset{\sim}{\lambda}$. We aim to create a sequence of points such that $\underset{\sim}{c}(\underset{\sim}{\lambda}) \to \underset{\sim}{0}$. Now, let

$$\underset{\sim}{\nabla}_\lambda \Phi = \frac{\partial \underset{\sim}{\phi}^T}{\partial \underset{\sim}{\lambda}} \nabla_\phi \Phi + \frac{\partial \Phi}{\partial \underset{\sim}{\lambda}}, \tag{33}$$

where $\partial\Phi/\partial\underset{\sim}{\lambda}$ is defined as the gradient vector of $\Phi$ of (30) w.r.t. $\underset{\sim}{\lambda}$ and $\underset{\sim}{\nabla}_\lambda \Phi$ is defined as the gradient vector subject to (31). Since $\underset{\sim}{\nabla}_\phi \Phi = \underset{\sim}{0}$

$$\underset{\sim}{\nabla}_\lambda \Phi = \frac{\partial \Phi}{\partial \underset{\sim}{\lambda}} = -\underset{\sim}{c}(\underset{\sim}{\lambda}). \tag{34}$$

Further, it can be shown that

$$\underset{\sim}{\nabla}_\lambda \underset{\sim}{\nabla}_\lambda^T \Phi = -[\underset{\sim}{A}^T \underset{\sim}{G}_\Phi^{-1} \underset{\sim}{A}]\Big|_{\underset{\sim}{\phi}(\underset{\sim}{\lambda})}, \tag{35}$$

where $\underset{\sim}{A}$ contains the $\underset{\sim}{a}_i$ as appropriately ordered columns and

$$\underset{\sim}{G}_\Phi \triangleq \underset{\sim}{\nabla}_\phi \underset{\sim}{\nabla}_\phi^T \Phi. \tag{36}$$

Furthermore, for large r

$$[\underset{\sim}{A}^T \underset{\sim\Phi}{G}^{-1} \underset{\sim}{A}]^{-1} \simeq 2r \underset{\sim}{1}. \tag{37}$$

The iteration

$$\underset{\sim}{\lambda}^{j+1} \leftarrow \underset{\sim}{\lambda}^{j} - [\underset{\sim}{A}^T \underset{\sim\Phi}{G}^{-1} \underset{\sim}{A}]^{-1} \underset{\sim}{c}\Big|_{\underset{\sim}{\phi}^j} \tag{38}$$

is a Newton step in the $\underset{\sim}{\lambda}$ parameters to a maximum of $\Phi(\underset{\sim}{\lambda})$, whereas the iteration shown in Step 5 of the algorithm is a steepest ascent step w.r.t. $\underset{\sim}{\lambda}$.

Obviously, second derivatives of $\Phi$ w.r.t. $\underset{\sim}{\phi}$ are needed for the true Newton step, which is a disadvantage. However, quasi-Newton methods can provide a good approximation to

$$\underset{\sim\Phi}{G}^{-1} (\underset{\sim}{\phi}(\underset{\sim}{\lambda}^j)).$$

Inequality Constraints

We let

$$\Phi(\underset{\sim}{\phi},\underset{\sim}{\lambda},r) = U(\underset{\sim}{\phi}) - \sum_{i=1}^{t} \lambda_i c_i(\underset{\sim}{\phi}) + r \sum_{i=1}^{t} c_i^2(\underset{\sim}{\phi})$$

$$+ r \sum_{i=t+1}^{m} \left(c_i(\underset{\sim}{\phi}) - \frac{\lambda_i}{2r}\right)_-^2 - \sum_{i=t+1}^{m} \frac{\lambda_i^2}{4r} \quad , \tag{39}$$

where

$$a_- \overset{\Delta}{=} \min\{0,a\} = \begin{cases} a \text{ if } a < 0, \\ 0 \text{ if } a \geq 0. \end{cases} \tag{40}$$

In this case, at $\underset{\sim}{\lambda} = \underset{\sim}{\lambda}^j$, $\underset{\sim}{\phi} = \underset{\sim}{\phi}^j$,

$$\frac{\partial \Phi}{\partial \lambda_i} = \begin{cases} -c_i \text{ if } i \leq t \text{ or } c_i - \dfrac{\lambda_i^j}{2r_i} \leq 0, & (41a) \\ -\dfrac{\lambda_i^j}{2r_i} \text{ otherwise.} & (41b) \end{cases}$$

Then

$$\underset{\sim}{\lambda}^{j+1} \leftarrow \underset{\sim}{\lambda}^{j} - [\underset{\sim}{A}^T \underset{\sim}{G}_{\Phi}^{-1} \underset{\sim}{A}]^{-1} \left.\frac{\partial \Phi}{\partial \underset{\sim}{\lambda}}\right|_{\underset{\sim}{\phi}^j} \tag{42}$$

except that rows/columns in $\underset{\sim}{\nabla}_{\lambda}\ \underset{\sim}{\nabla}_{\lambda}^{T}\Phi$ associated with (41b) have nonzero components only on the diagonal of value $-1/(2r_j)$.

Comment

To derive the form of (39) we have set, taking $\underset{\sim}{\phi}$ as k-dimensional,

$$c_i(\underset{\sim}{\phi}) = 0, \qquad i = 1, \ldots, t,$$

$$c_{i+t}(\underset{\sim}{\phi}) - \phi_{k+i} = 0, \quad i = 1, \ldots, m-t,$$

$$\phi_{k+i} \geq 0, \qquad i = 1, \ldots, m-t,$$

the latter being slack variables. It can be proved that

$$\phi_{k+i} = \max\{0,\ c_{i+t} - \frac{\lambda_{i+t}}{2r}\},$$

resulting in (39).

Discussion

It is well-known that a numerical solution of the necessary conditions for optimality using, for example, the Newton method for solving the resulting nonlinear equations, is not guaranteed to find a minimum. This provides the motivation for searching for alternative approaches.

Consider the penalty function

$$\Phi(\underset{\sim}{\phi},r) = U(\underset{\sim}{\phi}) + r_1 \sum_{i=1}^{t} c_i^2(\underset{\sim}{\phi}) + r_2 \sum_{i=t+1}^{m} (c_i(\underset{\sim}{\phi}))_-^2$$

minimized w.r.t. $\underset{\sim}{\phi}$ for increasing values of $r_1$, $r_2$, $\to \infty$. The features of this approach can be illustrated by considering

$$P(\underset{\sim}{\phi},r) = U(\underset{\sim}{\phi}) + r \sum_{i=1}^{t} c_i^2(\underset{\sim}{\phi}).$$

Now,

$$\underset{\sim}{\nabla}_\phi P(\check{\underset{\sim}{\phi}}) = \underset{\sim}{g}(\check{\underset{\sim}{\phi}}) + 2r \sum_{i=1}^{t} c_i(\check{\underset{\sim}{\phi}}) \underset{\sim}{a}_i(\check{\underset{\sim}{\phi}})$$

$$= \underset{\sim}{g}(\check{\underset{\sim}{\phi}}) \neq \underset{\sim}{0}.$$

This function will usually have an ill-conditioned Hessian. Consider instead

$$\Phi(\underset{\sim}{\phi},r) = U(\underset{\sim}{\phi}) - \sum_{i=1}^{t} \hat{\lambda}_i \, c_i(\underset{\sim}{\phi}) + r \sum_{i=1}^{t} c_i^2(\underset{\sim}{\phi}),$$

where $\hat{\underset{\sim}{\lambda}}$ are the Lagrange multipliers at $\check{\underset{\sim}{\phi}}$. Then

$$\underset{\sim}{\nabla}_\phi \Phi(\check{\underset{\sim}{\phi}},r) = \underset{\sim}{g}(\check{\underset{\sim}{\phi}}) - \sum_{i=1}^{t} \hat{\lambda}_i \, \underset{\sim}{a}_i(\check{\underset{\sim}{\phi}}) + 2r \sum_{i=1}^{t} c_i(\check{\underset{\sim}{\phi}}) \, \underset{\sim}{a}_i(\check{\underset{\sim}{\phi}})$$

$$= \underset{\sim}{0},$$

which is a desirable result for the application of unconstrained minimization techniques.

<u>The Han-Powell Algorithm [1-3,5-7]</u>

At $\underset{\sim}{\phi}^j$ we consider the following quadratic approximation. We minimize w.r.t. $\underset{\sim}{s}$ the function

$$Q^j(\underset{\sim}{s}) \triangleq U(\underset{\sim}{\phi}^j) + \underset{\sim}{g}^T(\underset{\sim}{\phi}^j)\underset{\sim}{s} + 0.5 \, \underset{\sim}{s}^T \underset{\sim}{B}^j \underset{\sim}{s} \qquad (43)$$

subject to

$$\left.\begin{aligned} &\underset{\sim}{a}_i^T(\underset{\sim}{\phi}^j)\underset{\sim}{s} + c_i(\underset{\sim}{\phi}^j) = 0, \quad i = 1, \ldots, t, \\ &\underset{\sim}{a}_i^T(\underset{\sim}{\phi}^j)\underset{\sim}{s} + c_i(\underset{\sim}{\phi}^j) \geq 0, \quad i = t+1, \ldots, m, \end{aligned}\right\} \qquad (44)$$

where $\underset{\sim}{B}^j$ is positive definite.

Main Algorithm

Step 1 Choose $\underset{\sim}{\phi}^0$, set $\underset{\sim}{B}^0 \leftarrow \underset{\sim}{1}$, $j \leftarrow 0$, $\mu_i^{(-1)} \leftarrow 0$, $i \leftarrow 1, \ldots, m$.

Step 2 Solve the quadratic program (43) - (44) w.r.t. $\underset{\sim}{s}$ yielding $\underset{\sim}{s}^j$. Solve for $\underset{\sim}{\lambda}^j$ the system

$$A_s^T A_s \underset{\sim}{\lambda}^j = A_s^T \underset{\sim}{\nabla}_s Q(\underset{\sim}{s}^j), \tag{45}$$

where $A_s$ is the k x s matrix whose columns are the vectors $\underset{\sim}{a}_i(\underset{\sim}{\phi}^j)$ corresponding to the active constraints.

Step 3 Find $\alpha^j$ by the one-dimensional search algorithm which follows.

Step 4 Set $\underset{\sim}{\phi}^{j+1} \leftarrow \underset{\sim}{\phi}^j + \alpha^j \underset{\sim}{s}^j$.

Step 5 If

$$|Q^j(\alpha^j \underset{\sim}{s}^j) - U(\underset{\sim}{\phi}^{j+1})| + \sum_{j \in J} \mu_i^j |c_i(\underset{\sim}{\phi}^{j+1})| + \sum_{\substack{j=1 \\ j \notin J}}^{m} |(c_i(\underset{\sim}{\phi}^{j+1}))_-| < \varepsilon, \tag{46}$$

where J denotes the indices of the constraints found active by the quadratic program, then stop.

Step 6 Update $\underset{\sim}{B}^j$.

Comment See the section on updating $\underset{\sim}{B}^j$.

Step 7 Set $j \leftarrow j+1$ and go to Step 2.

Line Search Algorithm

Step 1 $\alpha^j \leftarrow 1$.

Comment This usually turns out to be the final value for almost all j.

Step 2 If $y(\alpha^j) \leq y(0) + 0.1\ \Delta\ \alpha^j$ stop.

Comment See the definition of $y(\alpha)$ and $\Delta$ following this algorithm.

Step 3 Find the extreme point $\beta$ of a quadratic approximation using $y(0)$, $y'(0) \leftarrow \Delta$, $y(\alpha^j)$.

Step 4 Set $\alpha^j \leftarrow \min\ \{\max\{0.1\ \alpha^j,\ \beta\},\ 0.9\ \alpha^j\}$.

Step 5 Go to Step 2.

One-Dimensional Penalty Function

The function $y(\alpha)$ of Step 2 is given by

$$y(\alpha) \triangleq y^j(\underset{\sim}{\phi}^j + \alpha \underset{\sim}{s}^j), \tag{47}$$

where

$$y^j(\underset{\sim}{\phi}) = U(\underset{\sim}{\phi}) + \sum_{i=1}^{t} \mu_i^j |c_i(\underset{\sim}{\phi})| + \sum_{i=t+1}^{m} \mu_i^j |(c_i(\underset{\sim}{\phi}))_-| \tag{48}$$

and

$$\mu_i^j = \max\ \{|\lambda_i^j|,\ 0.5\ (\mu_i^{j-1} + |\lambda_i^j|)\}. \tag{49}$$

The expression (49) allows positive contributions to the objective function of some inequality constraints that are inactive at the solution to the quadratic program. The approximation to the gradient $\Delta$ is given by the difference

$$\Delta = \ell(1) - \ell(0), \tag{50}$$

where $\ell(\alpha)$ is the penalty function (48) when all functions are taken as linear at $\underset{\sim}{\phi}^j$. It can be shown that $\ell(1) < \ell(0)$. This is because

$$\alpha\ \underset{\sim}{a}_i^T(\underset{\sim}{\phi}^j)\ \underset{\sim}{s} + c_i(\underset{\sim}{\phi}^j) = 0, \qquad i = 1, \ldots, t, \tag{51}$$

$$\min\{0,\ \alpha\ \underset{\sim}{a}_i^T(\underset{\sim}{\phi}^j)\ \underset{\sim}{s} + c_i(\underset{\sim}{\phi}^j)\} = 0, \qquad i = t+1, \ldots, m. \tag{52}$$

for $\alpha = 1$.

## Update for $\underset{\sim}{B}^j$

The BFGS (Broyden-Fletcher-Goldfarb-Shanno) formula

$$\underset{\sim}{B}^{j+1} = \underset{\sim}{B}^j - \frac{\underset{\sim}{B}^j \underset{\sim}{s}^j \underset{\sim}{s}^{j^T} \underset{\sim}{B}^j}{\underset{\sim}{s}^{j^T} \underset{\sim}{B}^j \underset{\sim}{s}^j} + \frac{\underset{\sim}{z}^j \underset{\sim}{z}^{j^T}}{\alpha^j \underset{\sim}{z}^{j^T} \underset{\sim}{s}^j} \tag{53}$$

is recommended, where

$$\underset{\sim}{z}^j = \theta \, \underset{\sim}{\gamma}^j + (1-\theta) \, \underset{\sim}{B}^j \, \Delta\underset{\sim}{\phi}^j, \quad 0 \leq \theta \leq 1, \tag{54}$$

$$\Delta\underset{\sim}{\phi}^j \triangleq \alpha^j \, \underset{\sim}{s}^j, \tag{55}$$

$$\gamma^j \triangleq \underset{\sim}{\nabla}_\phi L \, (\underset{\sim}{\phi}^{j+1}, \underset{\sim}{\lambda}^j) - \underset{\sim}{\nabla}_\phi L \, (\underset{\sim}{\phi}^j, \underset{\sim}{\lambda}^j),$$

$$\theta = \begin{cases} 1 & \text{if } \underset{\sim}{\gamma}^{j^T} \Delta\underset{\sim}{\phi}^j \geq 0.2 \, p^j, \\ \dfrac{0.8 \, p^j}{p^j - \underset{\sim}{\gamma}^{j^T} \Delta\underset{\sim}{\phi}^j} & \text{if } \underset{\sim}{\gamma}^{j^T} \Delta\underset{\sim}{\phi}^j < 0.2 \, p^j, \end{cases} \tag{56}$$

and

$$p^j \triangleq \Delta\underset{\sim}{\phi}^{j^T} \underset{\sim}{B}^j \, \Delta\underset{\sim}{\phi}^j. \tag{57}$$

## Discussion of $\theta=1$

If $\underset{\sim}{B}^j$ is positive definite, $\underset{\sim}{B}^{j+1}$ may not be positive definite for $\theta = 1$ because we are not certain that $\underset{\sim}{\gamma}^{j^T} \underset{\sim}{s}^j > 0$. Hence, we allow $\theta$ to be $< 1$ when $\underset{\sim}{\gamma}^{j^T} \Delta\underset{\sim}{\phi}^j$ is not sufficiently large.

## Discussion of $\theta<1$

When $\theta = 0$, we have

$$\underset{\sim}{z}^j = \underset{\sim}{B}^j \, \Delta\underset{\sim}{\phi}^j, \tag{58}$$

which gives

$$\underset{\sim}{z}^{j^T} \Delta\underset{\sim}{\phi}^j = \Delta\underset{\sim}{\phi}^{j^T} \underset{\sim}{B}^j \Delta\underset{\sim}{\phi}^j \geq 0.2\ \Delta\underset{\sim}{\phi}^{j^T} \underset{\sim}{B}^j \Delta\underset{\sim}{\phi}^j. \qquad (59)$$

Consider

$$\underset{\sim}{z}^{j^T} \Delta\underset{\sim}{\phi}^j = 0.2\ \Delta\underset{\sim}{\phi}^{j^T} \underset{\sim}{B}^j \Delta\underset{\sim}{\phi}^j. \qquad (60)$$

Then

$$(\theta\underset{\sim}{\gamma}^j + (1-\theta)\ \underset{\sim}{B}^j \Delta\underset{\sim}{\phi}^j)^T \Delta\underset{\sim}{\phi}^j = 0.2\ p^j \qquad (61)$$

and

$$\theta\underset{\sim}{\gamma}^{j^T} \Delta\underset{\sim}{\phi}^j + \Delta\phi^{j^T} \underset{\sim}{B}^j \Delta\underset{\sim}{\phi}^j(1-\theta) = 0.2\ p^j$$

or

$$\theta\underset{\sim}{\gamma}^{j^T} \Delta\underset{\sim}{\phi}^j - \theta p^j = -0.8\ p^j,$$

from which, finally,

$$\theta = \frac{0.8\ p^j}{p^j - \theta\underset{\sim}{\gamma}^{j^T}\Delta\underset{\sim}{\phi}^j} > 0, \qquad (62)$$

since $p^j > 0$ and the denominator is positive.

## Convergence

Superlinear convergence for the algorithm can be proved, namely,

$$\|\underset{\sim}{\phi}^{j+1} - \check{\underset{\sim}{\phi}}\| \leq k^j \|\underset{\sim}{\phi}^j - \check{\underset{\sim}{\phi}}\|, \text{ where } k^j \to 0 \text{ as } j \to \infty.$$

Han proved global convergence if

$$\mu_i \geq |\lambda_i^j|, \quad i = 1, \ldots, m. \qquad (63)$$

Powell relaxed this requirement in his implementation, because he regarded this as inefficient in that the $\mu_i$ may become too large. To much effort then goes into satisfying the constraints.

References

[1] K. Madsen, "Nonlinear programming", McMaster University, Hamilton, Canada, Lectures given July 3 and 4, 1978.

[2] M.J.D. Powell, "A fast algorithm for nonlinearly constrained optimization calculations", Dundee Conf. Numerical Analysis (Dundee, Scotland, 1977).

[3] M.J.D. Powell, "Algorithms for nonlinear constraints that use Lagrangian functions", Mathematical Programming, vol. 14, 1978, pp. 224-248.

[4] R. Fletcher, "Methods related to Lagrangian functions", in Numerical Methods for Constrained Optimization, P.E. Gill and W. Murray, Eds. New York: Academic Press, 1974.

[5] R. Fletcher, "An ideal penalty function for constrained optimization", J. Inst. Mathematics and its Applications, vol. 15, 1975, pp. 319-342.

[6] S-P. Han, "A globally convergent method for nonlinear programming", J. Optimization Theory and Applications, vol. 22, 1977, pp. 297-309.

[7] S-P. Han, "Superlinearly convergent variable metric algorithms for general nonlinear programming problems", Mathematical Programming, vol. 11, 1976, pp. 263-282.

SOC-284

NONLINEAR PROGRAMMING USING LAGRANGIAN FUNCTIONS

J.W. Bandler

December 1981, No. of Pages: 16

Revised:

Key Words: Nonlinear programming, Lagrangian functions, quasi-Newton methods, Han-Powell algorithm, optimality conditions

Abstract: A brief review of major features of nonlinear programming methods which employ Lagrangian functions is presented. Following statements and discussion of necessary and sufficient conditions for a solution, the augmented Lagrangian method is described. This method adds a penalty term to permit a sequence unconstrained optimizations to be applied. The motivation behind the formulation and a discussion of Newton and quasi-Newton approaches is given. The Han-Powell algorithm is subsequently presented. This algorithm employs a quadratic approximation to the objective function, describes linearized constraints which leads to a quadratic program to be solved. The results provide the next search direction and appropriate Lagrange multipliers. After a one-dimensional search, the second derivative approximation is updated by a BFGS formula, with steps taken to ensure positive definiteness.

Description: Review.

Related Work: SOC-158, SOC-183.

Price: $ 6.00.