August 1981 LIDS-P-1124

PROJECTED NEWTON METHODS FOR OPTIMIZATION PROBLEMS WITH SIMPLE CONSTRAINTS*

Dimitri P. Bertsekas
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

We consider the problem min $\{f(x) \mid x \geq 0\}$ and algorithms of the form $x_{k+1} = [x_k - \alpha_k D_k \nabla f(x_k)]^+$ where $[\cdot]^+$ denotes projection on the positive orthant, α_k is a stepsize chosen by an Armijo-like rule, and D_k is a positive definite symmetric matrix which is partly diagonal. We show that D_k can be calculated simply on the basis of second derivatives of f so that the resulting Newton-like algorithm has a typically superlinear rate of convergence. With other choices of D_k convergence at a typically linear rate is obtained. The algorithms are almost as simple as their unconstrained counterparts. They are well suited for problems of large dimension such as those arising in optimal control while being competitive with existing methods for low-dimensional problems.

1. Introduction

We consider the problem

$$\text{minimize} \quad f(x) \qquad (1)$$

$$\text{subject to } x \geq 0$$

where $f : R^n \to R$ is a continuously differentiable function and the vector inequality $x \geq 0$ is meant to be componentwise [i.e. for $x = (x^1, x^2, \ldots, x^n) \in R^n$, we write $x \geq 0$ if $x^i \geq 0$ for all $i = 1, \ldots, n$]. This type of problem arises very often in applications; for example when f is a dual functional relative to an original inequality constrained primal problem, and x represents a vector of nonnegative Lagrange multipliers corresponding to the inequality constraints, and when f represents an Augmented Lagrangian or exact penalty function taking into account other possibly nonlinear equality and inequality constraints. The analysis and algorithms that follow apply also with minor modifications to problems with rectangle constraints such as

$$\text{minimize} \quad f(x) \qquad (2)$$

$$\text{subject to } b_1 \leq x \leq b_2$$

where b_1 and b_2 are given vectors. Problems (1) and (2) are referred to as simply constrained problems, and their algorithmic solution is the primary subject of this paper.

In view of the simplicity of the constraints, one would expect that solution of problem (1) is almost as easy as unconstrained minimization of f. This expectation is partly justified in that the first order necessary condition for a vector $\bar{x} = (\bar{x}^1, \ldots, \bar{x}^n)$ to be a local minimum of problem (1) takes the simple form

$$\frac{\partial f(\bar{x})}{\partial x^i} \geq 0 , \qquad i = 1, \ldots, n \qquad (3a)$$

$$\frac{\partial f(\bar{x})}{\partial x^i} = 0 , \qquad \text{if } \bar{x}^i > 0, \quad i = 1, \ldots, n \qquad (3b)$$

Furthermore the direct analog of the method of steepest descent takes the simple form

$$x_{k+1} = [x_k - \alpha_k \nabla f(x_k)]^+ , \quad k = 0, 1, \ldots \qquad (4)$$

where α_k is a positive scalar stepsize and for any vector $z = (z^1, \ldots, z^n) \in R^n$ we denote

$$[z]^+ = \begin{bmatrix} \max\{0, z^1\} \\ \vdots \\ \max\{0, z^n\} \end{bmatrix}$$

The stepsize α_k may be chosen in a number of ways. In the original proposal of Goldstein [1], and Levitin and Poljak [2], α_k is taken to be a constant $\bar{\alpha}$ (i.e. $\alpha_k \equiv \bar{\alpha}$, for all k) and a convergence result is shown under the assumption that $\bar{\alpha}$ is sufficiently small and ∇f is Lipschitz continuous. In general a proper value for $\bar{\alpha}$ can be found only through experimentation. An alternative suggested by McCormick [3] is to choose α_k by function minimization along the arc of points $x_k(\alpha)$, $\alpha \geq 0$ where

*Work supported by Grant NSF ECS-79-20834 and by Alphatech, Inc. through DOE Contract No. DE-AC01-79ET29243.

To be presented at the 20th IEEE Conference on Decision and Control, San Diego, CA. December 1981.

$$x_k(\alpha) = [x_k - \alpha \nabla f(x_k)]^+, \quad \alpha \geq 0 \qquad (5)$$

Thus α_k is chosen so that

$$f[x_k(\alpha_k)] = \min_{\alpha \geq 0} f[x_k(\alpha)]. \qquad (6)$$

Unfortunately the minimization above is very difficult to carry out, particularly for problems of large dimension, since $f[x_k(\alpha)]$ need not be differentiable, convex, or unimodal as a function of α even if f is convex. For most problems we prefer the Armijo-like stepsize rule first proposed in Bertsekas [4] whereby α_k is given by

$$\alpha_k = \beta^{m_k} s \qquad (7a)$$

where m_k is the first nonnegative integer m satisfying

$$f(x_k) - f[x_k(\beta^m s)] \geq \sigma \nabla f(x_k)'[x_k - x_k(\beta^m s)]. \qquad (7b)$$

Here the scalars s, β, and σ are fixed and are chosen so that $s > 0$, $\beta \in (0,1)$, and $\sigma \in (0, \frac{1}{2})$. In addition to being easily implementable and convergent, the algorithm (4), (7), has the advantage that when it converges to a local minimum $\bar{x}$ satisfying the standard second order sufficiency conditions for optimality (including strict complementarity) it identifies the active constraints at $\bar{x}$ in a finite number of iterations in the sense that there exists $\bar{k}$ such that

$$B(\bar{x}) = B(x_k) \quad \forall\, k \geq \bar{k} \qquad (8)$$

where, for every $x \in R^m$, B(x) denotes the set of indices of binding constraints at x

$$B(x) = \{i \mid x^i = 0,\ i = 1,\ldots,n\}. \qquad (9)$$

Minor modifications of the proofs given in [4] show that the results stated above hold also for the algorithm

$$x_{k+1} = [x_k - \alpha_k D_k \nabla f(x_k)]^+ \qquad (10)$$

where D_k is a diagonal positive definite matrix and α_k is chosen by (7) where now $x_k(\alpha)$ is given by

$$x_k(\alpha) = [x_k - \alpha D_k \nabla f(x_k)]^+ \qquad (11)$$

For this it is necessary to assume that the diagonal elements d_k^i, $i = 1,\ldots,n$ of the matrices D_k satisfy

$$\underline{d} \leq d_k^i \leq \bar{d}, \quad i = 1,\ldots,n \quad k = 0,1,\ldots$$

where $\underline{d}$ and $\bar{d}$ are some positive scalars.

While it is often possible to achieve substantial computational savings by proper diagonal scaling of ∇f as in (10), the resulting algorithm is typically characterized by linear convergence rate. Any attempt to construct a superlinearly convergent algorithm must by necessity involve a nondiagonal scaling matrix D_k which is an adequate approximation of the inverse Hessian $\nabla^2 f(x_k)^{-1}$, at least along a suitable subspace. At this point we find that the algorithms available at present are far more complicated than their unconstrained counterparts, particularly when the problem has large dimension. Thus the most straightforward extension of Newton's method is given by

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k) \qquad (12)$$

where $\bar{x}_k$ is a solution of the quadratic program

$$\text{minimize } \nabla f(x_k)'(x - x_k) + \frac{1}{2}(x - x_k)'\nabla^2 f(x_k)(x - x_k)$$
$$\text{subject to } x \geq 0 \qquad (13)$$

and α_k is a stepsize parameter. There are convergence and superlinear rate of convergence results in the literature regarding this type of method (Levitin and Poljak [2], Dunn [5]) and its Quasi-Newton versions (Garcia-Palomares and Mangasarian [6]), however its effectiveness is strongly dependent upon the computational requirements of solving the quadratic program (13). For problems of small dimension, problem (13) can be solved rather quickly by standard pivoting or manifold suboptimization methods but when the problem is of large dimension solution of the quadratic program (13) by standard methods can be very time consuming. Indeed, there are large-scale quadratic programming problems arising in optimal control the solution of which by pivoting methods is unthinkable. In any case the facility or lack thereoff of solving the quadratic program (13) must be accounted for when comparing method (12) against other alternatives.

Another possible approach for constructing superlinearly convergent algorithms for solving problem (1) stems from the original gradient projection proposal of Rosen [7], and is based on manifold suboptimization and active set strategies as in Gill and Murray [8], Goldfarb [9], Luenberger [10] and other sources, (see Lenard [11] for an up-to-date performance evaluation of various alternatives). Methods of this type are quite efficient for problems of relatively small dimension, but are typically unattractive for large-scale problems with a large number of constraints binding at a solution. The main reason is that typically at most one constraint can be added to the active set at each iteration, so if, for example, 1,000 constraints are binding at the point of convergence and an interior starting point is selected, then the method will require at least 1,000 iterations (and possibly many more) to converge. While several authors [8], [10] have alluded to the possibility of bending the direction of search along the constraint boundary, the only specific proposal known to the author that has been made in the context of the manifold suboptimization approach is the one of McCormick [12] and it does not seem particularly attractive for large-scale problems. (The Quasi-Newton methods proposed by Brayton and Cullum [13] incorporate bending but simultaneously require the solution of quadratic programming subproblems). Manifold suboptimization methods require also additional computation overhead in deciding which constraint to

drop from the currently active set. For the apparently most successful strategies (Lenard [11]) which attempt to drop as many constraints as possible, this overhead can be significant and must be taken into account when comparing the manifold suboptimization approach with other alternatives.

The algorithms proposed in this paper attempt to combine the basic simplicity of the steepest descent iteration (4), (7) with the sophistication and fast convergence of the constrained Newton's method (12), (13). They do not involve solution of a quadratic program thereby avoiding the associated computational overhead, and there is no bound to the number of constraints that can be added to the currently active set thereby bypassing a serious inherent limitation of manifold suboptimization methods. The basic form of the method is

$$x_{k+1} = x_k(\alpha_k) \tag{14}$$

where

$$x_k(\alpha) = [x_k - \alpha D_k \nabla f(x_k)]^+, \quad \alpha \geq 0. \tag{15}$$

D_k is a positive definite symmetric matrix which is partly diagonal, and α_k is a stepsize determined by an Armijo-like rule similar to (7) that will be described later. The convergence and rate of convergence properties of this method are discussed in Section 2. A key property of the method is that under mild assumptions it identifies the manifold of binding constraints at a solution in a finite number of iterations in the sense of (8). This means that eventually the method is reduced to an unconstrained method on this manifold and brings to bear the extensive methodology and analysis relating to unconstrained minimization algorithms.

In reference [17] we discuss how the method (14), (15) can form the basis for constructing algorithms for general linearly constrained problems of the form

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to } b_1 \leq Ax \leq b_2. \end{aligned} \tag{16}$$

The main idea here is to view problem (16) locally as a simply constrained problem via a transformation of variables. For example, if the matrix A is square and invertible, problem (16) is equivalent to the problem

$$\text{minimize} \quad h(y) \triangleq f(A^{-1}y)$$

$$\text{subject to } b_1 \leq y \leq b_2$$

via the transformation

$$y = Ax.$$

A similar approach based on an active set strategy is employed when A is not square and invertible. The ideas are similar to those involved in manifold suboptimization methods where a linear manifold is selected as a "local universe" for the purposes of the current iteration. In our algorithms we take a suitably chosen rectangle (i.e. a set described by upper and lower bounds on the variables) as a "local universe" instead of a manifold.

Throughout the paper we emphasize Newton-like methods as prototypes for broad classes of superlinearly converging algorithms that fit the framework of the paper. We often make positive definiteness assumptions on the Hessian matrix of f in order to avoid getting bogged down in technical details relating to modifications of Newton's method such as those employed in unconstrained minimization [14]-[16] to account for the possibility that $\nabla^2 f$ is not positive definite. Quasi-Newton versions of the Newton-like methods presented are possible but the discussion of specific implementations is beyond the scope of the paper. More generally it may be said that the nature of the algorithms proposed is such that almost every useful idea from unconstrained minimization can be fruitfully adapted within the constrained minimization framework considered here, however, the precise details of how this should be done may involve considerable further research and experimentation.

The notation employed throughout the paper is as follows. All vectors are considered to be column vectors. A prime denotes transposition. The standard norm in R^n is denoted by $|\cdot|$, i.e. for $x = (x^1,\ldots,x^n)$ we write $|x| = [\sum_{i=1}^{n} (x^i)^2]^{1/2}$. The gradient and Hessian of a function $f: R^n \to R$ are denoted by ∇f and $\nabla^2 f$ respectively. Proofs of all results stated may be found in reference [17].

2. Algorithms for Minimization Subject to Simple Constraints

We consider first the problem $\min \{f(x) \mid x \geq 0\}$ of (1). Any vector $\bar{x}$ satisfying the first order necessary condition (3) will be referred to as a critical point with respect to problem (1). We focus attention at iterations of the form

$$x_{k+1} = [x_k - \alpha_k D_k \nabla f(x_k)]^+$$

where D_k is a positive definite symmetric matrix and α_k is chosen by search along the arc of points

$$x_k(\alpha) = [x_k - \alpha D_k \nabla f(x_k)]^+, \quad \alpha \geq 0.$$

It is easy to construct examples where an arbitrary positive definite choice of the matrix D_k leads to situations where it is impossible to reduce the value of the objective by suitable choice of the stepsize α (i.e. $f[x_k(\alpha)] \geq f(x_k)$, $\forall\, \alpha \geq 0$). The following proposition identifies a class of matrices D_k for which an objective function reduction is possible. Define for all $x \geq 0$

$$I^+(x) = \{i \mid x^i = 0, \ \frac{\partial f(x)}{\partial x^i} > 0\}. \qquad (17)$$

We say that a symmetric nxn matrix D with elements d^{ij} is diagonal with respect to a subset of indices $I \subset \{1,2,\ldots,n\}$ if

$$d^{ij} = 0, \quad \forall\, i \in I,\ j = 1,2,\ldots,n, \quad j \neq i. \qquad (18)$$

Proposition 1: Let $x \geq 0$ and D be a positive definite symmetric matrix which is diagonal with respect to $I^+(x)$ and denote

$$x(\alpha) = [x - \alpha D\nabla f(x)]^+, \ \forall \alpha > 0. \qquad (19)$$

a) The vector x is a critical point with respect to problem (1) if and only if

$$x = x(\alpha), \quad \forall \alpha \geq 0.$$

b) If x is not a critical point with respect to problem (1) there exists a scalar $\bar{\alpha} > 0$ such that

$$f[x(\alpha)] < f(x), \quad \forall \alpha \in (0,\bar{\alpha}]. \qquad (20)$$

Based on Proposition 1 we are led to the conclusion that the matrix D_k in the iteration

$$x_{k+1} = [x_k - \alpha_k D_k \nabla f(x_k)]^+$$

should be chosen diagonal with respect to a subset of indices that contains

$$I^+(x_k) = \{i \mid x_k^i = 0, \ \frac{\partial f(x_k)}{\partial x^i} > 0\}.$$

Unfortunately the set $I^+(x_k)$ exhibits an undesirable discontinuity at the boundary of the constraint set whereby given a sequence $\{x_k\}$ of interior points that converges to a boundary point $\bar{x}$ the set $I^+(x_k)$ may be strictly smaller than the set $I^+(\bar{x})$. This causes difficulties in proving convergence of the algorithm and may have an adverse effect on its rate of convergence. (This phenomenon is quite common in feasible direction algorithms and is referred to zigzagging or jamming). For this reason we will employ certain enlargements of the sets $I^+(x_k)$ with the aim of bypassing these difficulties.

The algorithm that we describe utilizes a scalar $\varepsilon > 0$ (typically small), a fixed† diagonal positive definite matrix M (for example the identity), and two parameters $\beta \in (0,1)$ and $\sigma \in (0,\frac{1}{2})$ that will be used in connection with an Armijo-like stepsize rule. An initial vector $x_0 \geq 0$ is chosen and at the kth iteration of the algorithm we have a vector $x_k \geq 0$. Denote

$$w_k = |x_k - [x_k - M\nabla f(x_k)]^+|$$

$$\varepsilon_k = \min\{\varepsilon, w_k\}.$$

(k+1)st Iteration of the Algorithm

We select a positive definite symmetric matrix D_k which is diagonal with respect to the set I_k^+ given by

$$I_k^+ = \{i \mid 0 \leq x_k^i \leq \varepsilon_k, \ \frac{\partial f(x_k)}{\partial x^i} > 0\}. \qquad (21)$$

Denote

$$p_k = D_k \nabla f(x_k) \qquad (22)$$

$$x_k(\alpha) = [x_k - \alpha p_k]^+, \ \forall \alpha \geq 0. \qquad (23)$$

Then x_{k+1} is given by

$$x_{k+1} = x_k(\alpha_k) \qquad (24)$$

where

$$\alpha_k = \beta^{m_k} \qquad (25)$$

and m_k is the first nonnegative integer m such that

$$f(x_k) - f[x_k(\beta^m)] \geq \sigma\{\beta^m \sum_{i \notin I_k^+} \frac{\partial f(x_k)}{\partial x^i} p_k^i + \sum_{i \in I_k^+} \frac{\partial f(x_k)}{\partial x^i} [x_k^i - x_k^i(\beta^m)]\}. \qquad (26)$$

The stepsize rule (25), (26) may be viewed as a combination of the Armijo-like rule (7) and the Armijo rule usually employed in unconstrained minimization (see e.g. Polak [18]). When I_k^+ is empty the right-hand side of (26) becomes $\sigma\beta^m \nabla f(x_k)' p_k$ and is identical to the corresponding expression of the Armijo rule in unconstrained optimization, while if $I_k^+ = \{1,2,\ldots,n\}$ then inequality (26) is identical with (7). Note that, for all k, we have

$$I_k^+ \supset I^+(x_k) \qquad (27)$$

so the matrix D_k is diagonal with respect to $I^+(x_k)$. It is possible to show that, for all $m \geq 0$, the right-hand side of (26) is nonnegative, and it is positive if and only if x_k is not a critical point.

In conclusion the algorithm is well defined, decreases the value of the objective function at each iteration k for which x_k is not a critical point, and essentially terminates if x_k is critical. We proceed to analyze its convergence and rate of convergence properties. To this end we will make use of the following two assumptions:

†Actually the results that follow can be shown also for the case where M is changed from one iteration to the next in a way that its diagonal elements are bounded above and away from zero.

(A) The gradient ∇f is Lipschitz continuous on each bounded set of R^n, i.e. given any bounded set $S \subset R^n$ there exists a scalar L (depending on S) such that

$$|\nabla f(x) - \nabla f(y)| \leq L|x - y| \quad , \quad \forall\, x, y \in S. \tag{28}$$

(B) There exist positive scalars λ_1, λ_2 and nonnegative scalars q_1, q_2 such that

$$\lambda_1 w_k^{q_1} |z|^2 \leq z' D_k z \leq \lambda_2 w_k^{q_2} |z|^2 \quad , \quad \forall\, z \in R^n, \ k = 0,1,\ldots \tag{29}$$

where

$$w_k = |x_k - [x_k - M\nabla f(x_k)]^+| \quad . \tag{30}$$

Note that when $q_1 = q_2 = 0$, relation (30) takes the form

$$\lambda_1 |z|^2 \leq z' D_k z \leq \lambda_2 |z|^2 \quad , \quad \forall\, z \in R^n, \ k = 0,1\ldots \tag{31}$$

and simply says that the eigenvalues of D_k are uniformly bounded above and away from zero.

Proposition 2: Under Assumptions (A) and (B) above, every limit point of a sequence $\{x_k\}$ generated by iteration (24) is a critical point with respect to problem (1).

We now focus attention at a local minimum x^* satisfying the following second order sufficiency conditions. For all $x \geq 0$ we denote by $B(x)$ the set of indices of binding constraints at x, i.e.

$$B(x) = \{i | x^i = 0\}, \quad \forall\, x \geq 0. \tag{32}$$

(C) The local minimum x^* of problem (1) is such that for some $\delta > 0$, f is twice continously differentiable in the open sphere $\{x \,||x-x^*| < \delta\}$ and there exist positive scalars m_1, m_2 such that

$$m_1 |z|^2 \leq z' \nabla^2 f(x) z \leq m_2 |z|^2 \quad , \quad \forall\, x \text{ such that } |x - x^*| < \delta \text{ and } z \neq 0 \text{ such that } z^i = 0, \ i \in B(x^*) \tag{33}$$

Furthermore

$$\frac{\partial f(x^*)}{\partial x^i} > 0, \quad \forall\, i \in B(x^*). \tag{34}$$

The following proposition demonstrates an important property of the algorithm namely that under mild conditions it is attracted by a local minimum x^* satisfying Assumption (C) and identifies the set of active constraints at x^* in a finite number of iterations. Thus if the algorithm converges to x^* then after a finite number of iterations it is equivalent to an unconstrained optimization method restricted on the subspace of binding constraints at x^*. This property is instrumental in proving superlinear convergence of the algorithm when the portion of D_k corresponding to the indices $i \notin I_k^+$ is chosen in a way that approximates the inverse of the portion of the Hessian of f corresponding to these same indices.

Proposition 3: Let x^* be a local minimum of problem (1) satisfying Assumption (C). Assume also that (B) holds in the stronger form whereby, in addition to (41), the diagonal elements d_{ii}^k of the matrices D_k satisfy for some scalar $\bar{\lambda}_1 > 0$

$$\bar{\lambda}_1 \leq d_{ii}^k \quad , \quad \forall\, k = 0,1,\ldots, \ i \in I_k^+. \tag{35}$$

Then there exists a scalar $\bar{\delta} > 0$ such that if $\{x_k\}$ is a sequence generated by iteration (24) and for some index $\bar{k}$ we have

$$|x_{\bar{k}} - x^*| \leq \bar{\delta} \ , \tag{36}$$

then $\{x_k\}$ converges to x^* and we have

$$I_k^+ = B(x_k) = B(x^*), \ \forall\, k \geq \bar{k} + 1. \tag{37}$$

Under the assumptions of Proposition 3 we see that if the algorithm converges to a local minimum x^* satisfying Assumption (C) then it reduces eventually to an unconstrained minimization method restricted to the subspace

$$T = \{x | x^i = 0, \quad \forall\, i \in B(x^*)\} \ . \tag{38}$$

Furthermore as shown in Proposition 3 for some index $\bar{k}$ we will have

$$I_k^+ = B(x^*), \ \forall\, k \geq \bar{k} \ . \tag{39}$$

This shows that if the portion of the matrix D_k corresponding to the indices $i \notin I_k^+$ is chosen to be the inverse of the Hessian of f with respect to these indices then the algorithm eventually reduces to Newton's method restricted on the subspace T.

More specifically, by rearranging indices if necessary, assume without loss of generality that

$$I_k^+ = \{r_{k+1}, \ldots, n\}, \tag{40}$$

where r_k is some integer. Then D_k has the form

$$D_k = \begin{bmatrix} \bar{D}_k & \vdots & 0 \\ \cdots & r_{k+1} & \cdots \\ & d_k & 0 \\ 0 & \vdots & \ddots \\ & 0 & \ddots \ d_k^n \end{bmatrix} \tag{41}$$

where $d_k^i > 0$, $i = r_{k+1}, \ldots, n$ and $\bar{D}_k$ can be an arbitrary positive definite matrix. Suppose we choose $\bar{D}_k$ to be the inverse of the Hessian of f with respect to the indices $i = 1, \ldots, r_k$, i.e. the elements $[\bar{D}_k^{-1}]_{ij}$ of $\bar{D}_k^{-1}$ are

$$[D_k^{-1}]_{ij} = \frac{\partial^2 f(x_k)}{\partial x^i \partial x^j}, \quad \forall i,j \notin I_k^+. \qquad (42)$$

By Assumption (C), $\nabla^2 f(x^*)$ is positive definite on T so it follows from (39) that this choice is well defined and satisfies the assumption of Proposition 3 for k sufficiently large. Since the conclusion of this proposition asserts that the method eventually reduces to Newton's method restricted on the subspace T a superlinear convergence rate result follows. This type of argument can be used to construct a number of Newton-like and Quasi-Newton methods and prove corresponding convergence and rate of convergence results.

References

[1] A.A. Goldstein, "Convex Programming in Hilbert Space", Bull. Amer. Math. Soc., Vol. 70, 1964, pp. 709-710.

[2] E.S. Levitin and B.T. Poljak, "Constrained Minimization Problems", U.S.S.R. Comput. Math. Phys., Vol. 6, 1966, pp. 1-50.

[3] G.P. McCormick, "Anti-Zig-Zagging by Bending", Management Science, Vol. 15, 1969, pp. 315-319.

[4] D.P. Bertsekas, "On the Goldstein-Levitin-Poljak Gradient Projection Method", IEEE Transactions on Aut. Control, Vol. AC-21, 1976, pp. 174-184.

[5] J.C. Dunn, "Newton's Method and the Goldstein Step Length Rule for Constrained Minimization Problems", SIAM J. on Control and Optimization, Vol. 6, 1980, pp. 659-674.

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

[7] J.B. Rosen, "The Gradient Projection Method for Nonlinear Programming, Part I: Linear Constraints", J. SIAM, Vol. 8, 1960, pp. 181-217.

[8] P.E. Gill and M. Murray (eds.), Numerical Methods for Constrained Optimization, Academic Press, N.Y., 1974, Ch. 2 and 3.

[9] D. Goldfarb, "Extension of Davidon's Variable Metric Algorithm to Maximization Under Linear Inequality and Equality Constraints", SIAM J. of Applied Math., Vol. 17, 1969, pp. 739-

[10] D.G. Luenberger, Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Mass. 1973, Ch. 11.

[11] M.L. Lenard, "A Computational Study of Active Set Strategies in Nonlinear Programming with Linear Constraints", Mathematical Programming, Vol. 16, 1979, pp. 81-97.

[12] G.P. McCormick, "The Variable Reduction Method for Nonlinear Programming", Management Science, Vol. 17, 1970, pp. 146-160.

[13] R.K. Brayton and J. Cullum, "An Algorithm for Minimizing a Differentiable Function Subject to Box Constraints and Errors", Journal of Optimization Theory and Applications, Vol. 29, 1979, pp. 521-558.

[14] W. Murray, "Second Derivative Methods", in Numerical Methods for Unconstrained Optimization, W. Murray (ed.), Academic Press, N.Y., 1972, pp. 57-71.

[15] R. Fletcher, and T.L. Freeman, "A Modified Newton Method for Minimization", Journal of Optimization Theory and Applications, Vol. 23, 1977, pp. 357-372.

[16] J.J. More and D.C. Sorensen, "On the Use of Directions of Negative Curvature in a Modified Newton Method", Mathematical Programming, Vol. 16, 1979, pp. 1-20.

[17] D.P. Bertsekas, "Projected Newton Methods for Optimization Problems with Simple Constraints", SIAM J. on Control and Optimization, (to appear).

[18] E. Polak, Computational Methods in Optimization: A Unified Approach, Academic Press, 1971.