A Class of Accelerated Conjugate Direction Methods for Linearly Constrained Minimization Problems*

By Michael J. Best and Klaus Ritter

Abstract. A class of algorithms are described for the minimization of a function of n variables subject to linear inequality constraints. Under weak conditions convergence to a stationary point is demonstrated. The method uses a mixture of conjugate direction constructing and accelerating steps. Any mixture, for example alternation, may be used provided that the subsequence of conjugate direction constructing steps is infinite. The mixture of steps may be specified so that under appropriate assumptions the rate of convergence of the method is two-step superlinear or $(n - p + 1)$-step cubic where p is the number of constraints active at a stationary point. The accelerating step is always superlinearly convergent. A condition is given under which the alternating policy is every step superlinear. Computational results are given for several test problems.

1. **Introduction.** In [3] a conjugate direction method is described for minimizing a nonlinear function subject to linear inequality constraints. An accelerating step is always performed after the construction of $(n - p)$ conjugate directions, where n is the number of variables and p is the number of constraints active at the limit point of the sequence of points constructed by the method. Under appropriate assumptions this results in an $(n - p + 1)$-step cubic rate of convergence.

The idea of accelerating the rate of convergence of methods of conjugate directions for unconstrained optimization has further been pursued in [2] and [9]. In [2] the construction of conjugate directions is based on Zoutendijk's projection method [11], and the accelerating direction is obtained using an approximation to the solution of certain linear equations involving differences of gradients at previous iterations. In [9], conjugate directions are obtained by always choosing the descent direction orthogonal to $n - 1$ differences of gradients; and therefore, a set of n conjugate directions is available at every iteration. This allows an accelerating direction to be used more frequently than every n iterations.

It is the purpose of this paper to extend these methods to minimization problems with linear inequality constraints. The algorithm allows considerable flexibility in the mixture of accelerating and conjugate direction constructing steps. If the algorithm does not terminate in a finite number of steps it is only required that the number of conjugate direction constructing steps be infinite. Under appropriate assumptions then each accelerating step is a superlinear step, and this results in an l-step superlinear rate

of convergence where l and the rate of l-step superlinear convergence depend on the policy used.

As a special case we obtain the rate of convergence characteristic of [3]. Another special case is obtained by alternating accelerating and conjugate direction constructing steps. In most cases this policy gives a one-step superlinear convergence rate and in all cases the rate of convergence is at least two-step superlinear.

2. Notation, Formulation of the Problem. Let $x \in E^n$ and assume that $F(x)$ is a given real-valued function. If $F(x)$ is differentiable at a point x_j, we denote its gradient at x_j by $\nabla F(x_j)$ or g_j. If $F(x)$ is twice differentiable at x_j, we denote the Hessian matrix of $F(x)$ at x_j by $G(x_j)$ or G_j. For any column vector x and any matrix M the transpose is denoted by x' and M', respectively. For any vectors x and y, $L[x, y] \equiv \{w; w = \theta x + (1 - \theta)y, \ 0 \leqslant \theta \leqslant 1\}$ denotes the set of points on the line segment joining x and y.

Let A be an $m \times n$ matrix with rows $a_1', a_2', \ldots, a_m'$ and let b be an m-dimensional column vector with components $b_1, b_2, \ldots, b_m$. Define

$$R = \{x \in E^n / Ax \leqslant b\}.$$

We consider the problem of determining a point $z \in R$ such that $F(z) \leqslant F(x)$ for all $x \in R$.

A point $x \in R$ is said to be stationary if there exist numbers $\lambda_1, \lambda_2, \ldots, \lambda_m$ satisfying

$$\nabla F(x) = \sum_{i=1}^{m} \lambda_i a_i,$$

$$\lambda_i(a_i' x - b_i) = 0, \qquad \lambda_i \leqslant 0 \quad \text{for } i = 1, 2, \ldots, m.$$

The Kuhn-Tucker Theorem states that every local minimizer of $F(x)$ over R is a stationary point; and that if $F(x)$ is a pseudo-convex function, then every stationary point minimizes $F(x)$ over R.

It is the purpose of this paper to describe a class of algorithms which either terminate with a stationary point after a finite number of steps or produce a sequence of points $\{x_j\}$ with the properties:

(1) under a differentiability assumption $\{F(x_j)\}$ is strictly decreasing; every cluster point of $\{x_j\}$ is stationary; and if $\{x_j\}$ has an isolated cluster point, then $\{x_j\}$ converges.

(2) if the sequence $\{x_j\}$ has a cluster point z and in a neighborhood of z, $G(x)$ exists and has certain other properties, then $\{x_j\}$ converges to z. The rate of convergence will be l-step superlinear where l and the rate of superlinear convergence depend on the algorithm specified.

For later reference we formulate:

Assumption I. $F(x)$ is twice continuously differentiable and there are constants $0 < \mu < \eta$ such that $\mu \|x\|^2 \leqslant x' G(y) x \leqslant \eta \|y\|^2$ for all x and y.

For ease of presentation the convergence and convergence rate results of Sections

4 and 5, respectively, are obtained using the rather restrictive Assumption I. In Section 6 a modified class of algorithms is presented for which the convergence and convergence rate results apply under much weaker assumptions (Assumptions II and III). Some computational results are presented in Section 7.

3. **General Description of the Algorithm.** At iteration j the point x_j is known. The next point in the sequence is obtained by constructing a descent direction s_j, a stepsize σ_j and setting $x_{j+1} = x_j - \sigma_j s_j$.

Let $D'_j = [d_{1j}, d_{2j}, \ldots, d_{nj}]$ be an ($n \times n$) nonsingular matrix and let $D_j^{-1} = [c_{1j}, c_{2j}, \ldots, c_{nj}]$. Suppose for simplicity that the first p constraints are active at x_j. If it is appropriate that these same p constraints be active at x_{j+1}, then s_j must be orthogonal to $a'_1, a'_2, \ldots, a'_p$. A simple way to do this is to require that $d_{ij} = a'_i$ for $i = 1, 2, \ldots, p$ and then to take s_j parallel to any of the last $n - p$ columns of D_j^{-1}. By definition of the inverse matrix the required orthogonality property will be satisfied. Choosing s_j in this manner also results in s_j being orthogonal to $n - p - 1$ of the last $n - p$ columns of D'_j. If these columns are required to be normalized differences of gradients at previous points, then $s'_j(g_{j-i+1} - g_{j-i}) = 0$ so that s_j will be approximately conjugate to $n - p - 1$ previous search directions s_{j-i}, thus assuring a rapid rate of convergence.

The columns of D'_j will in general consist of gradients of active constraints a'_i and normalized differences of gradients $(g_{j-i} - g_{j-i+1})/\|\sigma_{j-i}s_{j-i}\|$. To record the origin and type of each column we use the index α_{ij} where

$$\alpha_{ij} = \begin{cases} -k & \text{if } d_{ij} = (g_k - g_{k+1})/\|\sigma_k s_k\|, \\ \\ l & \text{if } d_{ij} = a'_l. \end{cases}$$

An infinite subset of iteration indices J is assumed to have been specified at the beginning of the algorithm. For $j \in J$ we initially consider any column c_{ij} of D_j^{-1} to be a potential search direction and compute the n directional derivatives $(v_j)_i = c'_{ij}g_j/\|c_{ij}\|$. We compute $(v_j)_l$, the largest derivative in a direction which will drop an active constraint and $|(v_j)_k|$ the largest derivative in a direction which will maintain the active constraints. In the first few iterations any direction with a large directional derivative can be used to reduce F. Eventually, however, the search direction should be chosen in such a way as to guarantee convergence to a stationary point and to obtain a rapid rate of convergence. If $(v_j)_l$ is considerably larger than $|(v_j)_k|$, then it is clear that an active constraint should be dropped. To ensure that zigzagging does not occur, we require that if at iteration $j - 1$, a new constraint became active then no constraint may be dropped at iteration j unless there is no other usable direction. The information required to enforce this rule is contained in the indicator β_j which was set equal to one if a new constraint became active at the previous iteration.

If no active constraint is to be dropped, let r be the column number of D'_j corresponding to the oldest difference of gradients. Since the rate of superlinear convergence of the algorithm will be determined by the oldest gradient difference information, the sharpest results will be obtained if we use $s_j = c_{rj}(v_j)_r$ and then replace d_{rj} by more

recent information. In the first few iterations, if $|(v_j)_r|$ is larger than some fixed positive constant β, it is quite reasonable to do this. On the other hand, if $|(v_j)_r| \leqslant \beta$ but $|(v_j)_k|$ is larger than some other fixed positive number γ, it seems more appropriate to search along $c_{kj}(v_j)_k$ to obtain a large local decrease in F.

When x_j is close to a stationary point, then both $|(v_j)_r|$ and $|(v_j)_k|$ will be small. If c_{rj} tends to become orthogonal to g_j so that $|(v_j)_r|$ tends to zero and in addition $|(v_j)_k|$ remains bounded away from zero, then it is clear that convergence to a stationary point cannot be guaranteed using $c_{rj}(v_j)_r$ as a search direction. To avoid this situation we introduce the test $|(v_j)_r| > \beta|(v_j)_k|^2$ and use $c_{rj}(v_j)_r$ only if it passes this test. As j increases, $|(v_j)_k|$ becomes small and the test becomes weaker, making it increasingly more likely that $c_{rj}(v_j)_r$ can be used as the search direction. If the test is not satisfied, then in order to maintain convergence, s_j must have a component parallel to c_{rj}. Furthermore, it may occur infinitely often that the test has failed. In this case if s_j is constructed so that it converges to a direction parallel to c_{rj}, then the strongest convergence rate results will still apply. Choosing $s_j = (\text{sign}(v_j)_r c_{rj} + c_{kj}(v_j)_k)|(v_j)_k|$ when the test fails ensures that both these properties are satisfied. J_1 is the set of iteration indices for which s_j is chosen by this rule.

For $j \notin J$ a special search direction of the form $s_j = \Sigma_{\alpha_{ij} \leqslant 0} w_i c_{ij}$ is constructed to accelerate the rate of convergence of $\{x_j\}$. Assuming that x_j is close to a stationary point z, Taylor's theorem gives $F(x_j - s_j) \simeq F(x_j) - \Sigma w_i g_j' c_{ij} + \frac{1}{2}\Sigma w_i^2 c_{ij}' G c_{ij}$ since it will be shown in Theorem 2 that the c_{ij}'s are approximately conjugate directions. An appropriate way to choose the coefficients w_i is to minimize this quadratic approximation to F. This gives $w_i = g_j' c_{ij}/c_{ij}' G c_{ij}$. In Theorem 2 it will be shown that $G c_{ij} \simeq d_{ij}\|c_{ij}\|$. Thus it is appropriate to set $w_i = g_j' c_{ij}/\|c_{ij}\| \equiv (v_j)_i$.

In Step II of the algorithm the determination of the stepsize is based on the result (Lemma 6) that for $j \notin J_1$ the optimal stepsize converges to unity. If a unit stepsize gives a feasible point, then the Armijo test is applied; and the stepsize is repeatedly reduced until an acceptable stepsize is obtained. It will be shown in Lemma 6 that after a finite number of steps the Armijo test will always be satisfied with a unit stepsize. For $j \in J_1$, the optimal stepsize converges to

$$\frac{|(v_j)_r|}{|(v_j)_k|} + |(v_j)_k| \frac{\|c_{kj}\|'}{\|c_{rj}\|}$$

and the procedure is similar to the above.

In Step III of the algorithm D_{j+1} is obtained from D_j by replacing one of its columns with a new vector. The column to be replaced is either one which contains the gradient of a constraint which has just become inactive or the column corresponding to the oldest gradient difference information. The new column is either the gradient of a newly active constraint or the normalized gradient difference $d_j = (g_j - g_{j+1})/\|\sigma_j s_j\|$. If $s_j = (\text{sign}(v_j)_r c_{rj} + c_{kj}(v_j)_k)|(v_j)_k|$ and $|c_{rj}' d_j| < |c_{kj}' d_j(v_j)_x|$, then it is not possible to ensure that $\|D_{j+1}^{-1}\|$ will remain bounded. In this case and when $j \notin J$, the update is not performed; and we set $D_{j+1}^{-1} = D_j^{-1}$. It will be shown, however, in Lemma 5 that after a finite number of steps D_j is always updated for $j \in J$.

Since D_{j+1} differs from D_j by exactly one column, it is straightforward to compute D_{j+1}^{-1} from D_j^{-1} using a standard simplex transformation. Finally, it should be emphasized that only D_j^{-1} is required for the computations of the algorithm; and that since D_j' is not required, it need not be stored.

An important property of the algorithm is that under the assumption of strict complementary slackness for all j sufficiently large the constraints active at x_j are precisely those which are active at z. Suppose these constraints are p in number. The algorithm is designed in such a way that the mixture of conjugate direction constructing and accelerating steps is flexible. Performing $n - p$ consecutive regular steps and using an appropriate approximation to the optimal stepsize gives an $n - p$-step superlinear or quadratic convergence rate (Theorem 5). Following these steps with an accelerating step always gives a superlinear rate (Theorem 3) and this rate depends on the oldest gradient difference information in D_j'. An important feature of the algorithm is that regular and accelerating steps can be performed alternately. This policy results in a two-step superlinear rate (Theorem 4); and in addition, if the accelerating step does not converge faster than expected, then both regular and accelerating steps are superlinear steps (Theorem 6).

4. Detailed Description of the Algorithm and Convergence. In the algorithm we use constants $\alpha > 0$, $\beta > 0$, $\gamma > 0$, and $0 < \delta < \frac{1}{2}$. Furthermore, we require an infinite subsequence $J \subset \{0, 1, 2, \ldots \}$. For every $j \notin J$ we choose a special search direction s_j which is designed to accelerate the convergence of the sequence $\{x_j\}$ generated by the algorithm.

Let $x_0 \in R$ be the initial point and suppose that

$$a_i' x_0 = b_i, \quad i = 1, \ldots, q; \qquad a_i' x_0 < b_i, \quad i = q + 1, \ldots, m.$$

Set

$$d_{i0} = a_i, \quad i = 1, \ldots, q,$$

and, if $q < n$, let $d_{q+1,0}, \ldots, d_{n0}$ be any set of vectors such that with $D_0' = (d_{10}, \ldots, d_{n0})$ the matrix $D^{-1} = (c_{10}, \ldots, c_{n0})$ exists. Finally, set $\beta_0 = 1, J_1 = \{0\}$ and, $J(x_0) = \{\alpha_{10}, \ldots, \alpha_{n0}\}$ with $\alpha_{i0} = i$, $i = 1, \ldots, q$, and $\alpha_{i0} = 0$, $i = q + 1, \ldots, n$.

A general cycle of the algorithm consists of three steps which are described below. At the beginning of the jth cycle the following data are available: $x_j \in R$, $g_j = \nabla F(x_j)$, β_j, $J(x_j) = \{\alpha_{1j}, \ldots, \alpha_{nj}\}$ and $D_j^{-1} = (c_{1j}, \ldots, c_{nj})$.

Step I: *Computation of the Direction of Descent s_j.* Let $u_j' = g_j' D_j^{-1}$, $u_j' = ((u_j)_1, \ldots, (u_j)_n)$ and set

$$(v_j)_i = (u_j)_i / \|c_{ij}\|, \quad i = 1, \ldots, n.$$

Define $l = l_j$ and $k = k_j$ such that

$$(v_j)_l \geqslant (v_j)_i \quad \text{for all } i \text{ with } \alpha_{ij} > 0,$$

$$|(v_j)_k| \geqslant |(v_j)_i| \quad \text{for all } i \text{ with } \alpha_{ij} \leqslant 0.$$

If $\alpha_{ij} \leqslant 0$ for all i, set $(v_j)_l = 0$; and if $\alpha_{ij} > 0$ for all i, set $(v_j)_k = 0$.

(i) $j \notin J$, i.e., a special direction s_j is chosen.

Set

$$\bar{s}_j = \sum_{\alpha_{ij} \leqslant 0} c_{ij}(v_j)_i,$$

and

$$s_j = \begin{cases} \bar{s}_j & \text{if } g_j' \bar{s}_j > 0, \\ c_{lj}(v_j)_l & \text{if } g_j' \bar{s}_j \leqslant 0 \quad \text{and} \quad (v_j)_l > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(ii) $j \in J$, i.e., a regular direction s_j is chosen.

Set

$$s_j = c_{lj}(v_j)_l \quad \text{if } ((v_j)_l \geqslant \alpha |(v_j)_k| \quad \text{and} \quad \beta_j = 0) \text{ or}$$

$$((v_j)_k = 0 \quad \text{and} \quad (v_j)_l \geqslant 0);$$

otherwise determine r such that

$$-\alpha_{rj} \leqslant -\alpha_{ij} \quad \text{for all } i \text{ with } \alpha_{ij} \leqslant 0,$$

and set

$$s_j = \begin{cases} c_{rj}(v_j)_r & \text{if } |(v_j)_r| > \beta \text{ or } (|(v_j)_r| > \beta |(v_j)_k|^2 \text{ and } |(v_j)_k| < \gamma), \\ c_{kj}(v_j)_k & \text{if } |(v_j)_r| < \beta \text{ and } |(v_j)_k| > \gamma, \\ (c_{rj} \text{ sign}(v_j)_r + c_{kj}(v_j)_k)|(v_j)_k| & \text{otherwise.} \end{cases}$$

If $s_j = (\text{sign}(v_j)_r c_{rj} + c_{kj}(v_j)_k)|(v_j)_k|$, set $J_1 = J_1 \cup \{j\}$. Go to Step II.

Step II: *Computation of the Stepsize* σ_j. If $g_j' s_j \leqslant 0$, stop, otherwise let

$$\sigma_j^* = \min\left\{ \frac{a_i' x_j - b_i}{a_i' s_j} \text{ for all } i \text{ with } a_i' s_j < 0 \right\}^{**}$$

and

$$h(x_j; \sigma) = \frac{F(x_j) - F(x_j - \sigma s_j)}{\sigma g_j' s_j}, \quad \sigma > 0.$$

Set

$$\tilde{\sigma}_j = \begin{cases} \min\{1, \sigma_j^*\} & \text{if } j \notin J_1, \\ \min\left\{ \left| \frac{(v_j)_r}{(v_j)_k} \right| + |(v_j)_k| \frac{\|c_{kj}\|}{\|c_{rj}\|}, \sigma_j^* \right\} & \text{if } j \in J_1, \end{cases}$$

** If, for $i = 1,\ldots, m$, $a_i' s_j \geqslant 0$, set $\sigma_j^* = \infty$. If the minimal vlaue σ_j^* is attained for more than one index, we apply a small perturbation to b in such a way that for the perturbed constraints, σ_j^* is attained for exactly one index.

and let ν_j be the smallest nonnegative integer for which $h(x_j, \widetilde{\sigma}_j(\frac{1}{2})^{\nu_j}) \geqslant \delta$.

Set $\sigma_j = \widetilde{\sigma}_j(\frac{1}{2})^{\nu_j}$ and $x_{j+1} = x_j - \sigma_j s_j$. Compute g_{j+1} and go to Step III.

Step III: *Computation of D_{j+1}^{-1}, $J(x_{j+1})$, and β_{j+1}.*

Case 1. $\sigma_j < \sigma_j^*$, i.e., no new active constraint occurs at x_{j+1}.

Set $\beta_{j+1} = 0$ and if $j \in J$, $d_j = (g_j - g_{j+1})/\|\sigma_j s_j\|$. Let

$$\nu = \begin{cases} l & \text{if } s_j = c_{lj}(v_j)_l, \\ k & \text{if } s_j = c_{kj}(v_j)_k, \\ r & \text{if } s_j = c_{rj}(v_j)_k \text{ or } j \in J_1 \quad \text{and} \quad |c'_{rj}d_j| \geqslant |c'_{kj}d_j(v_j)_k|. \end{cases}$$

If $j \notin J$ or $j \in J_1$ and $|c'_{rj}d_j| < |c'_{kj}d_j(v_j)_k|$, set $D_{j+1}^{-1} = D_j^{-1}$ and $J(x_{j+1}) = J(x_j)$; otherwise, replace the νth column of D'_j by d_j, denote the new matrix by D'_{j+1} and determine D_{j+1}^{-1} from D_j^{-1}. Set

$$J(x_{j+1}) = \{\alpha_{1,j+1}, \ldots, \alpha_{n,j+1}\},$$

where $\alpha_{i,j+1} = \alpha_{ij}$, $i \neq \nu$, and $\alpha_{\nu,j+1} = -j$.

Case 2. $\sigma_j = \sigma_j^*$, i.e., a new constraint becomes active at x_{j+1}.

Set $\beta_{j+1} = 1$ and

$$\nu = \begin{cases} l & \text{if } s_j = c_{lj}(v_j)_l, \\ k & \text{if } s_j = c_{kj}(v_j)_k \quad \text{or} \quad s_j = \sum_{\alpha_{ij} \leqslant 0} c_{ij}(v_j)_i, \\ r & \text{if } s_j = c_{rj}(v_j)_r \quad \text{or} \quad j \in J_1. \end{cases}$$

Let $\xi = \xi_j$ be the index for which the minimum σ_j^* is attained. Suppose

$$|c'_{\nu j}a_\xi| \geqslant \gamma \|c_{\nu j}\|.\text{***}$$

Replace the νth column of D'_j by a_ξ, denote the new matrix by D'_{j+1} and determine D_{j+1}^{-1} from D_j^{-1}.

Set

$$J(x_{j+1}) = \{\alpha_{1,j+1}, \ldots, \alpha_{n,j+1}\},$$

where $\alpha_{i,j+1} = \alpha_{ij}$, $i \neq \nu$, and $\alpha_{\nu,j+1} = \xi_j$.

PROPOSITION 1. *Let Assumption I be satisfied, $x_j \in R$ and suppose s_j, σ_j and D_j are determined by the algorithm. Then*

(i) *D_j is nonsingular, i.e., D_j^{-1} exists.*

(ii) *$g_j s_j \leqslant 0$ if and only if x_j is a stationary point.*

(iii) *If x_j is not a stationary point, then σ_j is well defined, $x_{j+1} = x_j - \sigma_j s_j \in R$ and $F(x_{j+1}) < F(x_j)$.*

Proof. (i) By assumption D_0 is nonsingular. For every j, D'_j and D'_{j+1} are either identical or differ in exactly the νth column, i.e., $d_{\nu j} \neq d_{\nu,j+1}$. It suffices, therefore, to show that $d'_{\nu,j+1}c_{\nu j} \neq 0$. If $\sigma_j = \sigma_j^*$, this follows immediately from the assumption in Case 2 of Step III of the algorithm. If $\sigma_j < \sigma_j^*$, Assumption I and Lagrange's for-

---

*** The case that this assumption is not satisfied is discussed in Remark 1.

mula [10] imply that there is $\xi_j \in L[x_j, x_{j+1}]$ such that

$$s_j' d_{\nu, j+1} = s_j' d_j = s_j' \frac{g_j - g_{j+1}}{\|\sigma_j s_j\|} = s_j' G(\xi_j) \frac{s_j}{\|s_j\|} \geqslant \mu \|s_j\| > 0.$$

It follows now from the definition of ν in Step III of the algorithm that s_j is parallel to $c_{\nu j}$ and, therefore, $d_{\nu, j+1}' c_{\nu j} \neq 0$.

(ii) Writing g_j in the form

$$g_j = \sum_{i=1}^{n} \lambda_{ij} d_{ij},$$

we observe that x_j is a stationary point if and only if $\lambda_{ij} \leqslant 0$ for i with $\alpha_{ij} > 0$ and $\lambda_{ij} = 0$ for i with $\alpha_{ij} \leqslant 0$. Since $\lambda_{ij} = c_{ij}' g_j = \|c_{ij}\| (v_j)_i$, x_j is a stationary point if and only if $(v_j)_k = 0$ and $(v_j)_l \leqslant 0$. By Step I of the algorithm this is equivalent to $s_j = 0$.

(iii) Since $a_i' s_j \geqslant 0$ for all i with $\alpha_{ij} > 0$, it follows that $\sigma_j^* > 0$ and $x_j - \sigma s_j \in R$ for $0 \leqslant \sigma \leqslant \sigma_j^*$. By Taylor's theorem there is $\xi_j \in L[x_j, x_j - \sigma s_j]$ such that

$$F(x_j) - F(x_j - \sigma s_j) = \sigma g_j' s_j - \sigma (g_j - \nabla F(\xi_j)) s_j'.$$

This implies that

$$h(x_j; \sigma) = 1 - (g_j - \nabla F(\xi_j))' s_j / g_j' s_j,$$

where $h(x_j; \sigma)$ is defined in Step II of the algorithm. It follows that $\lim_{\sigma \to 0} h(x_j; \sigma) = 1$ so that ν_j is well defined and $F(x_{j+1}) < F(x_j)$.

Remark 1. In the convergence proof we need that the sequence $\{\|D_j^{-1}\|\}$ is bounded. This is shown in Lemma 1 provided $|c_{\nu j}' a_\xi| \geqslant \gamma \|c_{\nu j}\|$ for some $\gamma > 0$. If this assumption is not satisfied in Case 2 of Step III of the algorithm, the matrix D_{j+1}' is reset in the sense that all columns $d_{i,j+1}$ such that $\alpha_{i,j+1} \leqslant 0$ are replaced by unit vectors orthogonal to all other columns of D_{j+1}'. The details of this procedure are discussed in [7].

The proof that the sequence $\{x_j\}$ generated by the algorithm converges to the global minimizer of $F(x)$ over R is rather lengthy. Since the algorithm given in this paper is a generalization of a method described in [8], we shall not give a new convergence proof but show that the algorithm has the properties upon which the convergence proof given in [8] depends. First, we need

LEMMA 1. *Let Assumption I be satisfied. Then the sequences $\{D_j\}, \{D_j^{-1}\}$, and $\{s_j\}$ are bounded.*

Proof. Using the arguments given in the proof of Lemma 1 in [7], we see that $\{D_j\}$ is bounded and that it suffices to show that there is some $\omega > 0$ such that for all $j \in J$

(1) $$|c_{\nu j}' d_{\nu, j+1}| \geqslant \omega \|c_{\nu j}\|,$$

where ν is defined as in Step III of the algorithm.

If $\sigma_j = \sigma_j^*$, it follows from Case 2 in Step III of the algorithm that (1) is satisfied for every $\omega \leqslant \gamma$. If $\sigma_j < \sigma_j^*$, we see as in the proof of part 1 of Proposition 1 that

$s_j' d_{\nu, j+1} \geqslant \mu \|s_j\|$, which in conjunction with the definition of ν and s_j implies (1) if $j \notin J_1$.

For $j \in J_1$ we have $s_j = (c_{rj} \operatorname{sign}(v_j)_r + c_{kj}(v_j)_k)|(v_j)_k|$ and, therefore,

$$s_j' d_{\nu, j+1} = s_j d_j = (d_j' c_{rj} \operatorname{sign}(v_j)_r + d_j' c_{kj}(v_j)_k)|(v_j)_k| \geqslant \mu \|s_j\|,$$

or

$$d_j' c_{rj} \operatorname{sign}(v_j)_r + d_j' c_{kj}(v_j)_k \geqslant \mu \|d_j' c_{rj} \operatorname{sign}(v_j)_r + d_j' c_{kj}(v_j)_h\|.$$

Since $\{D_j\}$ is bounded, the right-hand side of this inequality is bounded away from zero; i.e., there is a constant $\delta > 0$ such that

(2) $$d_j' c_{rj} \operatorname{sign}(v_j)_r + d_j' c_{kj}(v_j)_h \geqslant \delta \quad \text{for all } j \in J_1,$$

which implies (1).

The algorithm given in [8] always uses as search direction either $c_{kj}(v_j)_k$ or $c_{lj}(v_j)_l$. The stepsize σ_j is the optimal stepsize, i.e., the solution to

$$\min\{F(x_j - \sigma s_j)|0 \leqslant \sigma_j \leqslant \sigma_j^*\},$$

where σ_j^* is the maximal stepsize defined in Step II of the algorithm.

For $s_j \neq c_{lj}(v_j)_l$ the property of s_j which is used in the convergence proof given in [8] is the following: (Lemma 5 in [8]).

Let $\{x_j, j \in I\}$ be a subsequence of $\{x_j\}$ which converges to z such that for all $j \in I$,

$$a_i' x_j = b_i, \quad i \in I_m; \qquad a_i x_j < b, \quad i \notin I_m,$$

where I_m is some subset of $\{1, \ldots, m\}$. If the orthogonal projection of $\nabla F(z)$ onto $\{x|a_i' x = 0, i \in I_m\}$ is different from zero, then there is $\epsilon > 0$ such that $g_j' s_j \geqslant \epsilon$ for $j \in I$ sufficiently large.

The critical properties of the stepsize σ_j are established in Proposition 1 and Lemma 3 of [8]. The main property used is the following: For every $\epsilon > 0$ there is $\delta(\epsilon) > 0$ such that $g_j s_j \geqslant \epsilon$ and $\sigma_j^* \geqslant \epsilon$ imply $F(x_{j+1}) \leqslant F(x_j) - \delta(\epsilon)$.

In the following two lemmas we shall show that the search direction s_j and the stepsize σ_j generated by the new algorithm have these properties too.

LEMMA 2. *Let Assumption I be satisfied and let $\{s_j\}$ and $\{x_j\}$ be determined by the algorithm. Let $I_m = \varnothing$ or $I_m \subset \{1, \ldots, m\}$. Suppose $\{x_j, j \in I\}$ is a subsequence of $\{x_j\}$ which converges to some $z \in E^n$. If, for every $j \in I$,*

$$a_i' x_j = b_i, \quad i \in I_m \quad \text{and} \quad a_i' x_j < b_i, \quad i \notin I_m,$$

and the orthogonal projection of $\nabla F(z)$ onto $\{x|a_i' x = 0, i \in I_m\}$ is different from zero, then there is $\epsilon > 0$ such that for $j \in I$ sufficiently large

$$|(v_j)_k| \geqslant \epsilon \quad \text{and} \quad s_j \neq c_{lj}(v_j)_l \quad \text{implies } g_j' s_j \geqslant \epsilon.$$

Proof. Let Px denote the orthogonal projection of x onto $\{x|a_i' x = 0, i \in I\}$. It follows from Lemma 2 in [8] that, for $j \in I$, $\rho_1 \|Pg_j\| \leqslant |(v_j)_k|$. Hence, $P\nabla F(z) \neq 0$

implies that there is $\epsilon > 0$ such that $|(v_j)_k| \geqslant \epsilon$ for $j \in I$ sufficiently large.

Since $\{D_j\}$ is bounded above, there is a constant $\overline{\omega} > 0$ such that $\|c_{ij}\| \geqslant \overline{\omega}$ for all i and j.

Thus, if $s_j = c_{kj}(v_j)_k$, we have

$$g_j' s_j = g_j' c_{kj}(v_j)_k = \|c_{kj}\|((v_j)_k)^2 \geqslant \overline{\omega}((v_j)_k)^2.$$

If $s_j = (c_{rj} \operatorname{sign}(v_j)_r + c_{kj}(v_j)_k)|(v_j)_k|$, we have

(1) $$g_j' s_j = (|g_j' c_{rj}| + \|c_{kj}\|((v_j)_k)^2)|(v_j)_k| \geqslant \overline{\omega}|(v_j)_k|^3.$$

Finally, if $s_j = \Sigma_{\alpha_{ij} < 0} c_{ij}(v_j)_i$, then

$$g_j' s_j = \sum_{\alpha_{ij} \leqslant 0} \|c_{ij}\|((v_j)_i)^2 \geqslant \|c_{kj}\|((v_j)_k)^2 \geqslant \overline{\omega}((v_j)_k)^2,$$

which completes the proof of the lemma.

LEMMA 3. *Let Assumption I be satisfied, and let $\{s_j\}$ and $\{\sigma_j\}$ be determined by the algorithm. For every $\epsilon > 0$ there is $\delta(\epsilon) > 0$ such that for all j,*

$$g_j' s_j \geqslant \epsilon \text{ and } \sigma_j^* \geqslant \epsilon \quad \text{imply } F(x_{j+1}) \leqslant F(x_j) - \delta(\epsilon),$$

$$\|x_{j+1} - x_j\| \geqslant \epsilon \qquad \text{implies } g_j' s_j \geqslant \delta(\epsilon),$$

$$g_j' s_j \geqslant \epsilon \text{ and } \sigma_j < \sigma_j^* \quad \text{imply } \|x_{j+1} - x_j\| \geqslant \delta(\epsilon).$$

Proof. By Taylor's theorem there is $\xi_j \in L[x_j, x_j - \sigma s_j]$ such that for $\sigma > 0$,

(1) $$\frac{F(x_j) - F(x_j - \sigma s_j)}{\sigma g_j' s_j} = 1 + \frac{(\nabla F(\xi_j) - g_j)' s_j}{g_j' s_j}.$$

Because $\nabla F(x)$ is uniformly continuous and $\{s_j\}$ is bounded, there is $\tau > 0$ such that for all j,

(2) $$\|\nabla F(\xi_j) - g_j\| \|s_j\| \leqslant \epsilon/2 \quad \text{for } 0 \leqslant \sigma \leqslant \tau.$$

If $g_j' s_j \geqslant \epsilon$, it follows, therefore, from (1) and (2) that for $0 < \sigma \leqslant \tau$,

(3) $$\frac{F(x_j) - F(x_j - \sigma s_j)}{\sigma g_j' s_j} \geqslant 1 - \frac{\|\nabla F(\xi_j) - g_j\| \|s_j\|}{g_j' s_j} \geqslant \frac{1}{2} > \delta.$$

By inequality (1) in the proof of Lemma 2 we have

(4) $$|(v_j)_k| \geqslant \epsilon_1 > 0 \quad \text{for all } j \in J_1 \text{ with } g_j' s_j \geqslant 0.$$

Hence, $g_j' s_j \geqslant \epsilon$ and $\sigma_j^* \geqslant \epsilon$ imply that $\widetilde{\sigma}_j \geqslant \epsilon_2 > 0$ for some ϵ_2. By (3) and the definition of v_j we have, therefore,

(5) $$\sigma_j \geqslant \tfrac{1}{2}\min\{\tau, \epsilon_2\} > 0$$

and

$$F(x_j) - F(x_{j+1}) \geqslant \sigma_j \delta g_j' s_j \geqslant \tfrac{1}{2}\delta\epsilon \min\{\tau, \epsilon_2\} > 0.$$

If $g_j' s_j \geqslant \epsilon$ and $\sigma_j < \sigma_j^*$, it follows from $\tilde{\sigma}_j \leqslant \sigma_j^*$, (3), and the definition of ν_j and σ_j that $\sigma_j^* \geqslant \tau$. Thus, (5) holds in this case too. Furthermore, $g_j' s_j \geqslant \epsilon$ implies $\|s_j\| \geqslant \epsilon_3$ for some $\epsilon_3 > 0$, which proves the last part of the lemma.

Since $\{\sigma_j\}$ is bounded, it suffices now to show that $\|s_j\| \geqslant \epsilon$ implies $g_j' s_j \geqslant \delta(\epsilon)$. By the definition of s_j and the boundedness of $\{D_j^{-1}\}$ we have, for some $\epsilon_4 > 0$,

$$|(v_j)_k| \geqslant \epsilon_4 \quad \text{if } s_j \neq c_{lj}(v_j)_l \quad \text{and} \quad \|s_j\| \geqslant \epsilon,$$

and

$$(v_j)_l \geqslant \epsilon_4 \quad \text{if } s_j = c_{lj}(v_j)_l \quad \text{and} \quad \|s_j\| \geqslant \epsilon.$$

As in the proof of Lemma 2, it follows then that $g_j' s_j \geqslant \omega \epsilon^2$ where ω is a lower bound for $\{\|c_{lj}\|\}$.

Based on the results of Lemmas 1–3 we can now use the convergence proof given in [8] to derive the following.

THEOREM 1. *Suppose Assumption I is satisfied. Then, there is a unique $z \in R$ such that $F(z) < F(x)$ for every $x \in R$, $x \neq z$. The algorithm either terminates after a finite number of iterations with z or generates an infinite sequence $\{x_j\}$ which converges to z.*

Assume that $a_i' z = b_i$, $i = 1, \ldots, p$, and $a_i' z < b_i$, $i = p + 1, \ldots, m$. By the Kuhn-Tucker Theorem there are $\lambda_1, \ldots, \lambda_p$ such that

$$\nabla F(z) = \sum_{i=1}^{p} \lambda_i a_i, \qquad \lambda_i \leqslant 0, \ i = 1, \ldots, p.$$

If $\lambda_i < 0$, $i = 1, \ldots, p$, the strict complementary slackness condition is said to be satisfied at z. In this case, it can be shown that after a finite number of iterations, the set of constraints active at the elements of $\{x_j\}$ does not change.

PROPOSITION 2. *Suppose Assumption I and the strict complementary slackness condition is satisfied. Then, there is j_0 such that, for $j \geqslant j_0$,*

$$a_i' x_j = b_i, \qquad i = 1, \ldots, p \quad \text{and} \quad a_i' x_j < b_i, \qquad i = p + 1, \ldots, m.$$

The proof of this proposition is identical with the proof of Proposition 2 in [8] and, therefore, omitted.

5. Superlinear Rate of Convergence. In this section we demonstrate the rate of convergence associated with several important specifications of the index set J of regular or conjugate direction constructing steps. To obtain these results it is reasonable to require that for all j sufficiently large the set of constraints which are active at x_j are precisely those which are active at z. Therefore, we assume throughout this section that Assumption I and the strict complementary slackness assumption are satisfied so that Proposition 2 applies.

For notational convenience we assume without loss of generality that for all $j \geqslant j_0$ the columns of D_j' have been ordered so that the last p columns contain the gradients $a_1', a_2', \ldots, a_p'$ of active constraints and the first $q \equiv n - p$ columns contain normalized differences of gradients.

The convergence rate results will be based on upper bounds on errors associated

with various approximations. The following lemma demonstrates the relationship among several of these error estimates. The matrix E_j of part (c) occurs when Taylor's theorem is used to express g_{j+1} in terms of g_j and $G \equiv G(z)$.

LEMMA 4. (a) *For any integer* $l \geqslant 0$, $\|x_{j+1} - z\| = O(\|x_j - z\|)$.

(b) $\|x_j - z\| = O(\|g_j - \nabla F(z)\|)$, $\|g_j - \nabla F(z)\| = O(\|x_j - z\|)$.

(c) *Define* $E_j = \int_0^1 G(x_j + t(x_{j+1} - x_j)) dt - G$; *then* $\|E_j\| \to 0$ *as* $j \to \infty$. *If* $G(x)$ *satisfies a Lipschitz condition in a neighborhood of* z, *then* $\|E_j\| = O(\|x_j - z\|)$.

Proof. (a) Let $j \geqslant j_0$. Since by Proposition 2, $(x_j - z)$ is orthogonal to $\nabla F(z)$, it follows from Taylor's theorem that

$$F(x_j) - F(z) = \tfrac{1}{2}(x_j - z)'G(\xi_j)(x_j - z) \quad \text{where } \xi_j \in L[x_j, z].$$

Thus, from Assumption I,

$$\frac{\mu}{2}\|x_j - z\|^2 \leqslant F(x_j) - F(z) \leqslant \frac{\eta}{2}\|x_j - z\|^2.$$

Since $l \geqslant 0$, by Proposition 1, $F(x_{j+l}) < F(x_j)$. Therefore,

$$\frac{\mu}{2}\|x_{j+l} - z\|^2 \leqslant F(x_{j+l}) - F(z) \leqslant \frac{\eta}{2}\|x_j - z\|^2$$

and

$$\|x_{j+l} - z\| \leqslant \sqrt{\eta/\mu}\|x_j - z\|.$$

(b) From the Lagrange formula [10] and Assumption I,

$$\|g_j - \nabla F(z)\| \|x_j - z\| \geqslant (g_j - \nabla F(z))'(x_j - z) = (x_j - z)'G(\xi_j)(x_j - z)$$

$$\geqslant \mu\|x_j - z\|^2,$$

where $\xi_j \in L[x_j, z]$. Therefore, $\|x_j - z\| \leqslant (1/\mu)\|g_j - \nabla F(z)\|$. Again from the Lagrange formula and Assumption I,

$$\|g_j - \nabla F(z)\|^2 = (g_j - \nabla F(z))'(g_j - \nabla F(z)) = (x_j - z)'G(\xi_j)(g_j - \nabla F(z))$$

$$\leqslant \eta\|x_j - z\| \|g_j - \nabla F(z)\|,$$

where $\xi_j \in L[x_j, z]$, so that $\|g_j - \nabla F(z)\| \leqslant \eta\|x_j - z\|$.

(c) Since $F(x)$ has continuous second derivatives and since by Theorem 1 $x_j \to z$ as $j \to \infty$, it follows that $\|E_j\| \to 0$ as $j \to \infty$. If $G(x)$ satisfies a Lipschitz condition in a neighborhood of z, then there is a number $L > 0$ such that for all j sufficiently large

$$\|E_j\| \leqslant \sup_{0 \leqslant t \leqslant 1} \|G(x_j + t(x_{j+1} - x_j)) - G(z)\|$$

$$\leqslant L \sup_{0 \leqslant t \leqslant 1} \|t(x_{j+1} - z) + (1 - t)(x_j - z)\|$$

$$\leqslant L\{\|x_{j+1} - z\| + \|x_j - z\|\}.$$

It now follows from part (a) that $\|E_j\| = O(\|x_j - z\|)$.

LEMMA 5. (a) *For* $i = 1, 2, \ldots, q$, $|(v_j)_i| = O(\|x_j - z\|)$.

(b) *For any integer $l \geqslant 0$,*

$$\|x_{j+l} - z\| = O(\max\{|g'_{j+l}c_{ij}|, \ i = 1, 2, \ldots, q\}).$$

(c) *For all $j \in J$ sufficiently large, $D_{j+1}^{-1} \neq D_j^{-1}$.*

Proof. (a) For $j \geqslant j_0$ and $i = 1, 2, \ldots, q$, by definition of the inverse matrix c_{ij} is orthogonal to $a'_1, a'_2, \ldots, a'_p$ and is thus also orthogonal to $\nabla F(z)$. Therefore,

$$(v_j)_i = \frac{g'_j c_{ij}}{\|c_{ij}\|} = (g_j - \nabla F(z))' \frac{c_{ij}}{\|c_{ij}\|}$$

and by Lemma 4(b), $|(v_j)_i| \leqslant \|g_j - \nabla F(z)\| = O(\|x_j - z\|)$.

(b) Let $T = \{t \in E^n, \ t = \sum_{i=1}^{p} \xi_i a_i\}$. First, we show that

(1)   for any $t \in T$, $\|x_j - z\| = O(\|g_j - t\|)$.

To verify (1) choose any $t \in T$ and let $j \geqslant j_0$. By Proposition 2, $(x_j - z)$ is orthogonal to both t and $\nabla F(z)$. Thus from the Lagrange formula and Assumption I,

$$\|g_j - t\|\|x_j - z\| \geqslant (g_j - t)'(x_j - z) = (g_j - \nabla F(z))'(x_j - z)$$

$$= (x_j - z)'G(\xi_j)(x_j - z) \geqslant \mu\|x_j - z\|^2,$$

where $\xi_j \in L[x_j, z]$. Therefore, $\|x_j - z\| \leqslant \|g_j - t\|/\mu$ which proves (1).

Now let l be any fixed nonnegative integer, $j \geqslant j_0$, and define $w'_j = g'_{j+l}D_j^{-1}$. It follows that

$$(w_j)_i = g'_{j+l}c_{ij} \quad \text{for } i = 1, 2, \ldots, q, \quad \text{and}$$

$$g_{j+l} = D'_j w_j = \sum_{i=1}^{q} (w_j)_i d_{ij} + t_j,$$

where by Proposition 2, $t_j \in T$. From Lemma 1, D_j is bounded so that by (1),

$$\|x_{j+l} - z\| = O(\max\{|(w_j)_i|, \ i = 1, 2, \ldots, q\})$$

$$= O(\max\{|g'_{j+l}c_{ij}|, \ i = 1, 2, \ldots, q\}).$$

(c) From part (a) and Lemma 1 it follows that $|d'_j c_{kj}(v_j)_k| \longrightarrow 0$ as $j \rightarrow \infty$. Therefore, from Eq. (2) of Lemma 1 and for all $j \in J$, sufficiently large $|d'_j c_{rj}| \geqslant \delta/2 > 0$ so that from Step III of the algorithm for all $j \in J$ sufficiently large, $D_{j+1}^{-1} \neq D_j^{-1}$.

The following theorem relates column i of D_j^{-1} to the search direction used when column i of D'_j was last replaced. In addition, it shows that the first q columns of D_j^{-1} are conjugate directions.

THEOREM 2. (a) *For every $j \geqslant j_0$ and $i = 1, 2, \ldots, q$,*

$$c_{ij}/\|c_{ij}\| = s_{-\alpha_{ij}}/\|s_{-\alpha_{ij}}\| + w_{ij}^1,$$

$$Gc_{ij} = \|c_{ij}\|d_{ij} + w_{ij}^2, \quad \text{and}$$

$$|c'_{kj}Gc_{ij}| = O(\|w_{ij}^2\|) \quad \text{for } i < k \leqslant q,$$

where $\|w_{ij}^1\|, \|w_{ij}^2\| \rightarrow 0$ as $j \rightarrow \infty$.

(b) *In addition, if $G(x)$ satisfies a Lipschitz condition in a neighborhood of z,* then

$$\|w_{ij}^1\|, \|w_{ij}^2\| = O(\|x_{-\alpha_{ij}} - z\|).$$

Proof. Let $j \geqslant j_0$ and $1 \leqslant i \leqslant q$ be arbitrary but fixed. Let $\nu = -\alpha_{ij}$, $\epsilon_\nu = \epsilon_\nu(i) = \text{sign}((v_\nu)_i)$, and let k be such that $|(v_\nu)_k| \geqslant |(v_\nu)_l|$ for $l = 1, 2, \ldots, q$. We first show that when column i of D_ν^{-1} is updated it is true that $c_{i,\nu+1}/\|c_{i,\nu+1}\| \simeq s_\nu/\|s_\nu\|$ and that the following updates do not substantially change this ith column.

From Step I of the algorithm, Lemma 5(a), and Proposition 2,

(1)
$$s_\nu = \begin{cases} c_{i\nu}(v_\nu)_i & \text{if } \nu \notin J_1, \\ c_{i\nu}\epsilon_\nu|(v_\nu)_k| + O(|(v_\nu)_k|^2) & \text{if } \nu \in J_1. \end{cases}$$

Column i of $D_{\nu+1}^{-1}$ is obtained from the update formula

$$c_{i,\nu+1} = \frac{c_{i\nu}}{c_{i\nu}'d_\nu} \quad \text{where } d_\nu = \frac{g_\nu - g_{\nu+1}}{\|\sigma_\nu s_\nu\|}.$$

Therefore,

(2)
$$c_{i,\nu+1} = \begin{cases} \dfrac{s_\nu}{\|s_\nu\|}\left(\dfrac{\|s_\nu\|}{(v_\nu)_i c_{i\nu}'d_\nu}\right), & \nu \notin J_1, \\[4mm] \dfrac{s_\nu}{\|s_\nu\|}\left(\dfrac{\|s_\nu\|}{\epsilon_\nu|(v_\nu)_k|c_{i\nu}'d_\nu}\right) + O(|(v_\nu)_k|), & \nu \in J_1. \end{cases}$$

For $\nu \notin J_1$ it follows from (1), the Lagrange formula and Assumption I that

(3)
$$\frac{(v_\nu)_i c_{i\nu}'d_\nu}{\|s_\nu\|} = \frac{s_\nu'G(\xi_\nu)s_\nu}{\|s_\nu\|^2} \geqslant \mu > 0.$$

Similarly, for $\nu \in J_1$,

(4)
$$\frac{\epsilon_\nu|(v_\nu)_k|c_{i\nu}'d_\nu}{\|s_\nu\|} = \frac{s_\nu'G(\xi_\nu)s_\nu}{\|s_\nu\|^2} + O(|(v_\nu)_k|) \geqslant \mu > 0.$$

From (2), (3), (4) and Lemma 5(a) it follows that

(5)
$$\frac{c_{i,\nu+1}}{\|c_{i,\nu+1}\|} = \frac{s_\nu}{\|s_\nu\|} + O(\|x_\nu - z\|).$$

In general D_ν^{-1} may not be updated for several iterations so that

$$D_{\nu+1}^{-1} = D_{\nu+2}^{-1} = \cdots = D_\xi^{-1} \neq D_{\xi+1}^{-1},$$

where ξ is the iteration at which the next update occurs. It follows that $\xi = -\alpha_{rj}$ where

$$-\alpha_{rj} = \min\{-\alpha_{lj}; -\alpha_{lj} > \nu, l = 1, 2, \ldots, q\},$$

and $D'_{\xi+1}$ is obtained from D'_ξ by replacing column r by $d_\xi = (g_\xi - g_{\xi+1})/\|\sigma_\xi s_\xi\|$. Column i of $D^{-1}_{\xi+1}$ is obtained from the update formula

$$(6) \qquad c_{i,\xi+1} = c_{i\xi} - \frac{c'_{i\xi} d_\xi}{c_{r\xi} d_\xi} c_{r\xi}.$$

We next show that $|c'_{i\xi} d_\xi|$ is appropriately small so that column i is left approximately unchanged after one update.

From Taylor's theorem,

$$(7) \qquad c'_{i\xi} d_\xi = c'_{i\xi} \left(G \frac{s_\xi}{\|s_\xi\|} + E_\xi \frac{s_\xi}{\|s_\xi\|} \right),$$

where E_j is defined in Lemma 4(c). By definition of ξ, $c_{i\xi} = c_{i,\nu+1}$ so that from (5),

$$Gc_{i\xi} = \|c_{i,\nu+1}\| G \frac{s_\nu}{\|s_\nu\|} + O(\|x_\nu - z\|).$$

Since $d_{i,\nu+1} = d_{i,\xi+1} = d_\nu$, we have from Taylor's theorem,

$$G \frac{s_\nu}{\|s_\nu\|} = d_{i,\xi+1} - E_\nu \frac{s_\nu}{\|s_\nu\|}.$$

Thus,

$$(8) \qquad c'_{i\xi} G = \|c_{i,\nu+1}\| \left(d_{i,\xi+1} - \frac{s'_\nu}{\|s_\nu\|} E'_\nu \right) + O(\|x_\nu - z\|).$$

Furthermore, from (5),

$$(9) \qquad \frac{s_\xi}{\|s_\xi\|} = \frac{c_{r,\xi+1}}{\|c_{r,\xi+1}\|} + O(\|x_\xi - z\|).$$

Substituting (8) and (9) into (7), applying Lemma 1 and Lemma 4(a) and observing that since $i \neq r$, $d'_{i,\xi+1} c_{r,\xi+1} = 0$ gives

$$(10) \qquad |c'_{i\xi} d_\xi| = O(\max\{\|x_\nu - z\|, \|E_\nu\|, \|E_\xi\|\}).$$

From (1), (3) and (4) it follows that $|c_{r\xi} d_\xi| \geqslant \mu$. Finally, since $c_{i\xi} = c_{i,\nu+1}$ it follows from Lemma 1, (5), (6) and (10) that

$$(11) \qquad \frac{c_{i,\xi+1}}{\|c_{i,\xi+1}\|} = \frac{s_\nu}{\|s_\nu\|} + O(\max\{\|x_\nu - z\|, \|E_\nu\|, \|E_\xi\|\}).$$

From iteration $\nu = -\alpha_{ij}$ to iteration j each column of D'_ν will have been replaced at most once. These updates will occur at iterations $j_1 < j_2 < j_3 < \cdots < j_i < j$, where $j_1 = -\alpha_{ij} = \nu$, $j_2 = -\alpha_{rj} = \xi, \ldots$, etc. Repeating the arguments which yielded Eqs. (6) to (11) at most $q - 1$ times, it follows that

$$(12) \qquad \frac{c_{ij}}{\|c_{ij}\|} = \frac{s_{-\alpha_{ij}}}{\|s_{-\alpha_{ij}}\|} + w^1_{ij},$$

where

$$\|w_{ij}^1\| = O(\max\{\|x_{j_1} - z\|, \|E_{j_1}\|, \|E_{j_2}\|, \ldots, \|E_{j_i}\|\}).$$

Since J is infinite, $-\alpha_{ij} \rightarrow \infty$ as $j \rightarrow \infty$. Therefore, from Lemma 4(c), for $i = 1, 2, \ldots, q$, $\|w_{ij}^1\| \rightarrow 0$ as $j \rightarrow \infty$.

Furthermore, from (12) and Taylor's theorem,

$$(13) \qquad\qquad Gc_{ij} = \|c_{ij}\|d_{ij} + w_{ij}^2,$$

where

$$w_{ij}^2 = \|c_{ij}\|\left(w_{ij}^1 + E_\nu \frac{s_\nu}{\|s_\nu\|}\right);$$

and thus, $\|w_{ij}^2\| \rightarrow 0$ as $j \rightarrow \infty$.

Finally, from (13), since for $k > i$, $c_{kj}'d_{ij} = 0$, it follows that $|c_{kj}'Gc_{ij}| = O(\|w_{ij}^2\|)$, which completes the proof of part (a).

The proof of part (b) is immediate from (12), (13) and Lemma 4 parts (a) and (c).

In Step II of the algorithm if $j \notin J$, the stepsize procedure first attempts to use a unit stepsize. If necessary, this stepsize is reduced by a factor of ½ until an Armijo-Goldstein test is satisfied. The importance of beginning with a unit step is that the optimal stepsize converges to unity for $j \notin J_1$, and that as a consequence after a certain number of iterations the unit stepsize will always be accepted. For $j \in J$, similar remarks apply but with the unit stepsize replaced by

$$\left|\frac{(v_j)_r}{(v_j)_k}\right| + |(v_j)_k| \frac{\|c_{kj}\|}{\|c_{rj}\|}.$$

The precise results are contained in the following lemma.

LEMMA 6. *There is a j_1 such that for all $j \geqslant j_1$, $v_j = 0$ and $\widetilde{\sigma}_j < \sigma_j^*$. In particular, this means that for $j \geqslant j_1$, $\sigma_j = 1$ for $j \notin J_1$ and*

$$\sigma_j = \left|\frac{(v_j)_r}{(v_j)_k}\right| + |(v_j)_k| \frac{\|c_{kj}\|}{\|c_{rj}\|} \quad \text{for } j \in J_1.$$

Furthermore, if $\hat{\sigma}_j$ is such that $F(x_j - \hat{\sigma}_j s_j) \leqslant F(x_j - \sigma s_j)$ for all $0 \leqslant \sigma \leqslant \sigma_j^*$, then $\hat{\sigma} \rightarrow 1$ as $j \rightarrow \infty$, $j \notin J_1$ and $\widetilde{\sigma}_j/\hat{\sigma}_j \rightarrow 1$ as $j \rightarrow \infty$, $j \in J_1$.

Proof. From Lemma 5(a), the definition of s_j and Proposition 2 it follows that $\sigma_j^* \rightarrow \infty$ as $j \rightarrow \infty$. Thus, from the definition of $\widetilde{\sigma}_j$ in Step II of the algorithm it follows that for all j sufficiently large, $\widetilde{\sigma}_j < \sigma_j^*$.

We next show that

$$(1) \qquad\qquad \frac{s_j'Gs_j}{g_j's_j} = 1 + \epsilon_j \quad \text{where } |\epsilon_j| \rightarrow 0 \text{ as } j \rightarrow \infty, j \notin J_1.$$

Let $j \in J - J_1$ and $j \geqslant j_0$. Then $s_j = c_{rj}(v_j)_r$ for some r, $1 \leqslant r \leqslant q$. Therefore, $g_j's_j = \|c_{rj}\|((v_j)_r)^2$. From Theorem 2,

$$s'_j G s_j = ((v_j)_r)^2 c'_{rj} G c_{rj} = ((v_j)_r)^2 (\|c_{rj}\| + O(\|w^2_{rj}\|)).$$

Therefore,

$$(2) \qquad\qquad\qquad\qquad s'_j G s_j / g'_j s_j = 1 + O(\|w^2_{rj}\|).$$

Next let $j \notin J$ and $j \geq j_0$. From Step I of the algorithm $s_j = \Sigma^q_{i=1} c_{ij}(v_j)_i$ and $g'_j s_j = \Sigma^q_{i=1} \|c_{ij}\|((v_j)_i)^2$. Let k be such that $|(v_j)_k| \geq |(v_j)_i|$, $i = 1, 2, \ldots, q$. Then, $g'_j s_j \geq \|c_{kj}\|((v_j)_k)^2$. Furthermore, from Theorem 2(a),

$$s'_j G s_j = \left(\sum_{i=1}^q (v_j)_i c_{ij} \right)' \left(\sum_{i=1}^q (v_j)_i \|c_{ij}\| d_{ij} + (v_j)_i w^2_{ij} \right)$$

$$= \sum_{i=1}^q ((v_j)_i)^2 \|c_{ij}\| + O(((v_j)_k)^2 \max\{\|w^2_{ij}\|, \|w^2_{rj}\|, \ldots, \|w^2_{qj}\|\}).$$

Therefore,

$$(3) \qquad\qquad\qquad \frac{s'_j G s_j}{g'_j s_j} = 1 + O(\max\{\|w^2_{1j}\|, \|w^2_{2j}\|, \ldots, \|w^2_{qj}\|\}).$$

Equation (1) now follows from (2), (3) and Theorem 2(a).

From Taylor's theorem,

$$F(x_j - s_j) = F(x_j) - g'_j s_j + \tfrac{1}{2} s'_j G(\xi_j) s_j,$$

where $\xi_j \in L[x_j, x_j - s_j]$. Thus, for $j \notin J_1$,

$$h(x_j, \widetilde{\sigma}_j) = h(x_j, 1) = \frac{F(x_j) - F(x_j - s_j)}{g'_j s_j} = 1 - \frac{1}{2} \frac{s'_j G(\xi_j) s_j}{g'_j s_j}.$$

We have from (1) and Assumption I,

$$\frac{s'_j G(\xi_j) s_j}{g'_j s_j} = \frac{s'_j G(\xi_j) s_j}{s'_j G s_j} (1 + \epsilon_j) = (1 + \epsilon_j) \left(1 + \frac{s'_j (G(\xi_j) - G) s_j}{s'_j G s_j} \right)$$

$$= 1 + O(\max\{|\epsilon_j|, \|G(\xi_j) - G\|\}).$$

Since $s_j \to 0$, $\xi_j \to z$ as $j \to \infty$, $j \notin J_1$, and since $F(x)$ has continuous second derivatives it follows that $h(x_j, \widetilde{\sigma}_j) = h(x_j, 1) \to \tfrac{1}{2}$ as $j \to \infty$, $j \notin J_1$. Since $\delta < \tfrac{1}{2}$, we have for all $j \notin J_1$ sufficiently large that $h(x_j, \widetilde{\sigma}_j) = h(x_j, 1) > \delta$ and thus, $\nu_j = 0$ and $\sigma_j = 1$.

Let $\hat{\sigma}_j$ be defined as in the statement of the lemma. It follows from the Lagrange formula that for all j sufficiently large

$$(4) \qquad\qquad\qquad \hat{\sigma}_j = \frac{g'_j s_j}{s'_j G(\xi_j) s_j} \quad \text{where } \xi_j \in L[x_j, x_j - \hat{\sigma}_j s_j].$$

As in the previous argument,

$$\hat{\sigma}_j = 1 + O(\max\{|\epsilon_j|, \|G(\xi_j) - G\|\}).$$

From (1) and Assumption I, $g_j's_j \leqslant 2\eta\|s_j\|^2$ so that $|\hat{\sigma}_j| < 4\eta/\mu$ for all $j \notin J_1$ sufficiently large. Since $s_j \longrightarrow 0$, it then follows that $\xi_j \longrightarrow z$ as $j \longrightarrow \infty$, $j \notin J_1$. Therefore, $\hat{\sigma}_j \longrightarrow 1$ as $j \longrightarrow \infty$, $j \notin J_1$.

Now let $j \in J_1$ and $j \geqslant j_0$. Let $k = k(j)$ and $r = r(j)$ be such that $|(v_j)_k| \geqslant |(v_j)_i|$ and $-\alpha_{rj} \leqslant -\alpha_{ij}$ for $i = 1, 2, \ldots, q$. We have $s_j = (\text{sign}(k_j)_r)c_{rj} + (c_{kj}(v_j)_k) \cdot |(v_j)_k|$ so that $g_j's_j = (|(u_j)_r| + \|c_{kj}\|((v_j)_k)^2)|(v_j)_k|$. From Theorem 2,

$$Gs_j = (\text{sign}((u_j)_r)\|c_{rj}\|d_{rj} + (v_j)_k\|c_{kj}\|d_{kj} + O(\max\{\|w_{rj}^2\|, |(v_j)_k| \|w_{kj}^2\|\}))|(v_j)_k|.$$

Therefore,

$$s_j'Gs_j = ((v_j)_k)^2(\|c_{rj}\| + O(\max\{\|w_{rj}^2\|, |(v_j)_k| \|w_{kj}^2\|, ((v_j)_k)^2\}))$$

and

$$(5) \qquad \frac{g_j's_j}{s_j'Gs_j} = \left(\frac{|(v_j)_r|}{|(v_j)_k|} + |(v_j)_k| \frac{\|c_{kj}\|}{\|c_{rj}\|}\right)(1 + \epsilon_j)$$

$$= \tilde{\sigma}_j(1 + \epsilon_j), \quad \text{where by Lemma 5(a) and Theorem 2(a),}$$

$$\epsilon_j \longrightarrow 0 \text{ as } j \longrightarrow \infty, j \in J_1.$$

From Taylor's theorem,

$$F(x_j - \tilde{\sigma}_j s_j) = F(x_j) = \tilde{\sigma}_j g_j's_j + \frac{1}{2}\tilde{\sigma}_j^2 s_j'G(\xi_j)s_j,$$

where $\xi_j \in L[x_j, x_j - \tilde{\sigma}_j s_j]$. Therefore,

$$h(x_j, \tilde{\sigma}_j) = 1 - \frac{1}{2}\tilde{\sigma}_j s_j'G(\xi_j)s_j/g_j's_j.$$

It follows from (5) that

$$\tilde{\sigma}_j \frac{s_j'G(\xi_j)s_j}{g_j's_j} = \frac{s_j'G(\xi_j)s_j}{s_j'Gs_j(1 + \epsilon_j)} = \frac{1}{1 + \epsilon_j}\left(1 + \frac{s_j'(G(\xi_j) - G)s_j}{s_j'Gs_j}\right)$$

$$= 1 + O(\max\{|\epsilon_j|, \|G(\xi_j) - G\|\}),$$

where the last equality follows from Assumption I. Since $\tilde{\sigma}_j$ is bounded and $s_j \longrightarrow 0$, it follows that $\xi_j \longrightarrow z$; and thus, $\|G(\xi_j) - G\| \longrightarrow 0$ as $j \longrightarrow \infty$, $j \in J_1$. Therefore, $h(x_j, \tilde{\sigma}_j) \longrightarrow \frac{1}{2}$; and since $\delta < \frac{1}{2}$, we have for all $j \in J_1$ sufficiently large that $h(x_j, \tilde{\sigma}_j) > \delta$; and thus, $v_j = 0$ and $\sigma_j = \tilde{\sigma}_j$.

From (4) and (5), as in the above arguments

$$\tilde{\sigma}_j/\hat{\sigma}_j = 1 + O(\max\{|\epsilon_j|, \|G(\xi_j) - G\|\}),$$

where $\xi_j \in L[x_j, x_j - \hat{\sigma}_j s_j]$. Since $\|\hat{\sigma}_j s_j\| \longrightarrow 0$, it follows that $\xi_j \longrightarrow z$ as $j \longrightarrow \infty$, $j \in$

J_1. Therefore $\widetilde{\sigma}_j/\widehat{\sigma} \rightarrow 1$ as $j \rightarrow \infty$, $j \in J_1$, which completes the proof of the lemma.

The following theorem shows that the rate of convergence of the subsequence of points constructed by the accelerating step is superlinear. The rate of superlinear convergence depends on the oldest gradient difference information contained in D_j'.

THEOREM 3.

(a)
$$\frac{\|x_{j+1} - z\|}{\|x_j - z\|} \rightarrow 0 \quad \text{as } j \rightarrow \infty, j \notin J.$$

(b) *In addition, if $G(x)$ satisfies a Lipschitz condition in a neighborhood of z,* then

$$\frac{\|x_{j+1} - z\|}{\|x_j - z\|} = O(\|x_{r_j} - z\|), \quad j \notin J,$$

where $r_j = \min\{-\alpha_{ij}; 1 \leqslant i \leqslant q\}$.

Proof. (a) We first show that for all $j \notin J$ and $i = 1, 2, \ldots, q$,

(1)
$$|c_{ij}' g_{j+1}| = O(\|x_j - z\| \max\{\|w_{ij}^2\|, \|E_j\|\}),$$

where w_{ij}^2 and E_j are as defined in the statements of Theorem 2(a) and Lemma 4(c), respectively. Let $1 \leqslant i \leqslant q$ and $j \notin J, j \geqslant j_1$. From Lemma 6, $\sigma_j = 1$ so that $x_{j+1} - x_j = -s_j$. From Taylor's theorem,

(2)
$$g_{j+1}' c_{ij} = g_j' c_{ij} - s_j' G c_{ij} - s_j' E_j c_{ij}.$$

From Theorem 2(a), $G c_{ij} = \|c_{ij}\| d_{ij} + w_{ij}^2$; and since $s_j = \Sigma_{l=1}^q (c_{lj}/\|c_{lj}\|) c_{lj}' g_j$ and furthermore since d_{ij} is orthogonal to $c_{1j}, c_{2j}, \ldots, c_{i-1,j}, c_{i+1,j}, \ldots, c_{qj}$, it follows that

$$s_j' G c_{ij} = g_j' c_{ij} + s_j' w_{ij}^2.$$

Substituting this expression into (2) gives

$$g_{j+1}' c_{ij} = -s_j' w_{ij}^2 - s_j' E_j c_{ij}.$$

Since $\|x_{j+1} - x_j\| = \|s_j\| \leqslant \|x_{j+1} - z\| + \|x_j - z\|$, it follows from Lemma 4(a) that $\|s_j\| = O(\|x_j - z\|)$. By Lemma 1, $\|c_{ij}\|$ is bounded. Therefore,

$$|g_{j+1}' c_{ij}| = O(\|x_j - z\| \max\{\|E_j\|, \|w_j^2\|\}),$$

which verifies (1).

Finally, from (1) and Lemma 5(b),

(3)
$$\frac{\|x_{j+1} - z\|}{\|x_j - z\|} = O(\max\{\|E_j\|, \|w_{1j}^2\|, \|w_{2j}^2\|, \ldots, \|w_{qj}^2\|\})$$

and part (a) follows from Lemma 4(c) and Theorem 2(a).

(b) Part (b) follows immediately from (3), the definition of r_j, Theorem 2(b) and Lemma 4, parts (a) and (c).

A particular member of the class of algorithms described in this paper is obtained

by specifying the index set J of regular steps. The convergence rate properties of any particular algorithm will depend to a large extent on the choice of J. The following theorems give specific convergence rate results for a number of important special cases.

One way of mixing regular and accelerating steps is by alternating them. This results in a two-step superlinear convergence rate; and the rate of superlinear convergence depends on the information of the previous $2q$ iterations, this being the number of iterations to update all of the first q columns of D_j^{-1}.

THEOREM 4. *Let* $J = \{j; j = 2i, i = 0, 1, 2, \ldots \}$. *Then*

(a)
$$\frac{\|x_{2j+2} - z\|}{\|x_{2j} - z\|} \to 0 \quad as\ j \to \infty;$$

(b) *if* $G(x)$ *satisfies a Lipschitz condition in a neighborhood of* z, *then*
$$\|x_{2j+2} - z\|/\|x_{2j} - z\| = O(\|x_{2j-2q+2} - z\|).$$

Proof. Part (a) follows from Theorem 3(a) and Lemma 4(a). To demonstrate part (b) we observe that from Lemma 5(c) and the definition of J,

$$\{-\alpha_{i,2j+1}; i = 1, 2, \ldots, q\} = \{2j, 2j - 2, \ldots, 2j - 2q + 2\},$$

so that in the statement of Theorem 3(b), $r_{2j+1} = 2j - 2q + 2$. The present theorem now follows directly from Theorem 3(b).

Each regular step generates a new conjugate direction. Performing q such steps consecutively in a space of dimension q leads us to expect a q-step superlinear or quadratic rate of convergence. These steps may then be followed by a superlinearly convergent accelerating step to achieve a cubic rate over the entire $q + 1$ iterations. The precise result which is similar to that obtained in [2] and [9] for the unconstrained case and [3] for the linearly constrained case is contained in the following theorem and its corollary.

THEOREM 5. *Let* $J_0 = \{j; j - 1, j - 2, \ldots, j - q \in J\}$ *and suppose* J_0 *is infinite. In addition, if* $\hat{\sigma}_j$ *denotes the optimal stepsize at iteration* j *and for* $j \in J_0$, $\sigma_{j-1} = \hat{\sigma}_{j-1}$, $\sigma_{j-2} = \hat{\sigma}_{j-2}$, $\ldots$, $\sigma_{j-q} = \hat{\sigma}_{j-q}$, *then*

(a)
$$\|x_j - z\|/\|x_{j-q} - z\| \to 0 \quad as\ j \to \infty,\ j \in J_0;$$

(b) *in addition, if* $G(x)$ *satisfies a Lipschitz condition in a neighborhood of* z, *then* $\|x_j - z\| = O(\|x_{j-q} - z\|^2)$.

Proof. Let j and l be such that $j \in J_0$, $j \geqslant j_0$ and $1 \leqslant l \leqslant q$. To simplify notation we assume without loss of generality that the columns of D_j' have been ordered so that

$$-\alpha_{1j} = j - 1, -\alpha_{2j} = j - 2, \ldots, -\alpha_{qj} = j - q.$$

From the identity

$$g_j = -\sum_{i=1}^{l-1} d_{ij} \|x_{j-i} - x_{j-i-1}\| + g_{j-l+1}$$

and the definition of an inverse matrix, it follows that $g_j' c_{lj} = g_{j-l+1}' c_{lj}$. Since $\sigma_{j-l} = \hat{\sigma}_{j-l}$, we have $g_{j-l+1}' s_{j-l} = 0$. Also, from Proposition 2, both c_{lj} and s_{j-l} are orthogonal to $\nabla F(z)$ so that

$$g_j' c_{lj} = (g_{j-l+1} - \nabla F(z))' \left(c_{lj} - \|c_{lj}\| \frac{s_{j-l}}{\|s_{j-l}\|} \right).$$

From Lemma 4 parts (a) and (b) and Theorem 2(a), $|g_j' c_{lj}| = O(\|x_{j-q} - z\| \, \|w_{lj}^1\|)$; and therefore, from Lemma 5(b),

$$\frac{\|x_j - z\|}{\|x_{j-q} - z\|} = O(\max\{\|w_{1j}^1\|, \|w_{2j}^1\|, \dots, \|w_{qj}^1\|\}) \longrightarrow 0 \quad \text{as } j \to \infty, \, j \in J_0.$$

This completes the proof of part (a). Part (b) follows from Theorem 2(b) and Lemma 4(a).

COROLLARY. *Suppose for $j \in J_0$ as defined above, $j \notin J$ and for $j \in J, \sigma_j = \hat{\sigma}_j$. In addition, suppose that $G(x)$ satisfies a Lipschitz condition in a neighborhood of z. Then for $j \in J_0$, $\|x_{j+1} - z\| = O(\|x_{j-q} - z\|^3)$.*

Proof. From part (b) of Theorem 3 for $j \in J_0$,

$$\|x_{j+1} - z\| = O(\|x_j - z\| \, \|x_{j-q} - z\|) \quad \text{where } r_j = j - q.$$

Remark. Theorem 5 and its corollary also apply if for $j \in J$ instead of using $\sigma_j = \hat{\sigma}_j$ the optimal stepsize, σ_j is obtained using a quadratic interpolation procedure described in [8]. Although σ_j as computed in Step II of the algorithm converges to unity, the rate of convergence is not sufficiently rapid to be compatible with the superlinear, quadratic or cubic rates of Theorem 5.

For the policy of alternating regular and accelerating steps if s_j is an accelerating direction, then Theorem 4(b) shows that $\|x_{j+1} - z\|/\|x_j - z\| = O(\|x_{j-2q+1} - z\|)$. If the subsequence of points determined by the accelerating step does not converge faster than this upper bound predicts, then the following theorem shows that $\{x_j\}$ is every step superlinearly convergent. The rate of superlinear convergence is sharper for the accelerating step than for the regular step.

THEOREM 6. *Let J be as in Theorem 4. Assume that $G(x)$ satisfies a Lipschitz condition in a neighborhood of z and there is an $\epsilon > 0$ such that*

$$\frac{\|x_{j+1} - z\|}{\|x_j - z\|} \geqslant \epsilon \|x_{j-2q+1} - z\| \quad \text{for all } j \notin J.$$

Then

$$\frac{\|x_{j+1} - z\|}{\|x_j - z\|} = O(\|x_{j-2q+1} - z\|) \quad \text{for all } j \notin J,$$

and

$$\frac{\|x_{j+1} - z\|}{\|x_j - z\|} = O(\|x_{j-4q+2} - z\|) \quad \text{for all } j \in J.$$

Proof. Let $j \notin J$ such that $j \geqslant j_1$; and thus by Lemma 6, $\sigma_j = 1$. We assume without loss of generality that the columns of D'_j have been ordered so that

$$-\alpha_{1j} = j - 1, -\alpha_{2j} = j - 3, \ldots, -\alpha_{qj} = j - 2q + 1.$$

Let $1 \leqslant \nu \leqslant q$. Since $s_j = \Sigma^q_{i=1} c_{ij}(v_j)_i$, it follows from Taylor's theorem and Theorem 2 that

$$c'_{\nu j} g_{j+1} = c'_{\nu j} g_j - c'_{\nu j} G \left(\sum_{i=1}^{q} c_{ij}(v_j)_i \right) - c'_j E_j s_j$$

$$= c'_{\nu j} g_j - \|c_{\nu j}\| d'_{\nu j} \sum_{i=1}^{q} c_{ij}(v_j)_i - w^{2'}_{\nu j} s_j - c'_{\nu j} E_j s_j$$

$$= - w^{2'}_{\nu j} s_j - c'_{\nu j} E_j s_j.$$

Since $\|s_j\| = O(\|x_j - z\|)$, it follows from Theorem 2 and Lemma 4 that

(1) $$|c'_{\nu j} g_{j+1}| = O(\|x_j - z\| \, \|w^2_{\nu j}\|);$$

and therefore,

$$\max\{|c'_{ij} g_{j+1}|; i = 1, 2, \ldots, q\} = O(\|x_j - z\| \, \|x_{j-2q+1} - z\|).$$

From Lemma 5(b), $\|x_{j+1} - z\|/\|x_j - z\| = O(\|x_{j-2q+1} - z\|)$.

From (1) and the assumption in the statement of the theorem,

(2) $$\frac{|c'_{\nu j} g_{j+1}|}{\|x_{j+1} - z\|} = O \left(\frac{\|w^2_{\nu j}\|}{\|x_{j-2q+1} - z\|} \right) = O \left(\frac{\|x_{j-2\nu+1} - z\|}{\|x_{j-2q+1} - z\|} \right).$$

Thus from Theorem 4, for $\nu = 1, 2, \ldots, q - 1$, $|c'_{\nu j} g_{j+1}|/\|x_{j+1} - z\| \to 0$ as $j \to \infty$, $j \notin J$. However, by Lemma 5(b),

$$1 = O \left(\max \left\{ \frac{|c'_{ij} g_{j+1}|}{\|x_{j+1} - z\|}; i = 1, 2, \ldots, q \right\} \right)$$

so that there is a constant $\theta > 0$ which, for all $j \notin J$ sufficiently large, satisfies $|c'_{qj} g_{j+1}|/\|x_{j+1} - z\| \geqslant \theta$. Since $j \notin J$, $c_{ij} = c_{i,j+1}$. Therefore, $|(v_{j+1})_q| \geqslant |(v_{j+1})_i|$, $i = 1, 2, \ldots, q$; and from Step I of the algorithm, $s_{j+1} = c_{qj}(v_{j+1})_q$.

From Taylor's theorem, Lemma 6 and Theorem 2,

$$c'_{\nu j} g_{j+2} = c'_{\nu j} g_{j+1} - c'_{\nu j} G c_{qj}(v_{j+1})_q - c'_{\nu j} E_{j+1} s_{j+1}$$

$$= c'_{\nu j} g_{j+1} = \|c_{\nu j}\| d'_j c_{qj}(v_{j+1})_q - \|c_{\nu j}\| w^2_{\nu j} s_{j+1} - c'_{\nu j} E_{j+1} s_{j+1}.$$

Therefore, from (2), Lemma 4, parts (a) and (c), and Theorem 4(b)

$$|c'_{\nu j} g_{j+2}| = \begin{cases} O(\|x_{j+1} - z\| \, \|x_{j-4q+3} - z\|), & \nu = 1, 2, \ldots, q - 1, \\ O(\|x_{j+1} - z\| \, \|x_{j-2q+1} - z\|), & \nu = q. \end{cases}$$

Finally, from Lemma 5(b),

$$\|x_{j+2} - z\|/\|x_{j+1} - z\| = O(\|x_{j-4q+3} - z\|);$$

and since $j \notin J$ implies $j + 1 \in J$, the theorem follows.

6. A Modified Algorithm. All convergence results in the previous sections are based on Assumption I. Since it is in general difficult, if not impossible, to verify this assumption it is desirable to try to obtain convergence results under weaker assumptions. It is the purpose of this section to show that for a simple modification of the algorithm, convergence can be proven without assumptions on second order derivatives. Furthermore, it will be shown that, if the sequence generated by the modified algorithm has a cluster point in a neighborhood of which $G(x)$ exists and has certain properties, the results of Section 5 on the rate of convergence are applicable.

We shall use the following

Assumption II. Let $x_0 \in R$ be the starting point of the algorithm. Then there exists a compact convex S such that $\{x \in R | F(x) \leqslant F(x_0)\} \subset S$, and $F(x)$ is continuously differentiable on some open set containing S.

It is clear that under this assumption an optimal solution z exists. By the Kuhn-Tucker Theorem, z is a stationary point. We cannot expect that the sequence $\{x_j\}$ generated by the modified algorithm converges to z if we only have Assumption II. However, we would like to show that every cluster point of $\{x_j\}$ is a stationary point, i.e. satisfies the necessary conditions for an optimal solution.

The proof of this result requires that the sequences $\{D_j\}$ and $\{D_j^{-1}\}$ be bounded. This has been shown in Lemma 1. The proof is based on the inequalities

$$(1) \qquad\qquad d_j' s_j \geqslant \mu \|s_j\| \quad \text{and} \quad \|d_j\| \geqslant \eta,$$

which follow from Assumption I. Without this assumption, (1) need not be true. Therefore, we have to modify the algorithm in such a way that we test whether (1) is satisfied and update D_j^{-1} only if the answer is affirmative.

We describe now a general iteration of the modified algorithm. The quantities D_j^{-1}, $J(x_j)$ and β_j are defined as before. In addition, we use two constants $0 < \gamma_1 < \gamma_2$.

Step I: *Computation of the Direction of Descent s_j.* Same as in previous algorithm.

Step II: *Computation of the Stepsize.* Same as in previous algorithm.

Step III: *Computation of D_{j+1}^{-1}, $J(x_{j+1})$, and β_{j+1}.*

Case 1. $\sigma_j < \sigma_j^*$, i.e., no new active constraint occurs at x_{j+1}.

Set $\beta_{j+1} = 0$, $d_j = (g_j - g_{j+1})/\|\sigma_j s_j\|$ if $j \in J$ and let ν be defined as in the previous algorithm. If $j \notin J$ or ($j \in J_1$ and $|c_{rj}' d_j| < |c_{kj}' d_j(v_j)_k|$) or

$$|d_j' s_j| < \gamma_1 \|s_j\| \quad \text{or} \quad \|d_j\| > \gamma_2,$$

set $D_{j+1}^{-1} = D_j^{-1}$ and $J(x_{j+1}) = J(x_j)$; otherwise, proceed as in the previous algorithm.

Case 2. $\sigma_j = \sigma_j^*$; i.e., a new constraint becomes active at x_{j+1}.

Same as in previous algorithm.

It is easy to verify that Lemma 1 holds. Since the first two steps of the algorithm are unchanged, Lemmas 2 and 3 remain valid. Therefore, we can again use the

convergence results of [8] to prove the following.

THEOREM 7. *Let Assumption* II *be satisfied. The modified algorithm either terminates after a finite number of iterations with a stationary point or generates an infinite sequence* $\{x_j\}$ *with the following properties*:

(i) $\|x_{j+1} - x_j\| \longrightarrow 0$ *as* $j \longrightarrow \infty$.

(ii) *Every cluster point of* $\{x_j\}$ *is a stationary point.*

(iii) *If* $\{x_j\}$ *has an isolated cluster point* z, *then* $\{x_j\}$ *converges to* z.

Proof. Because of Lemmas 1–3 the second statement of the theorem follows from Theorem 1 in [6]. If there is $\epsilon > 0$ and a subsequence $\{x_j, j \in J\}$ of J such that $\|x_{j+1} - x_j\| \geqslant \epsilon$ for $j \in J$, then it follows from Lemma 3 that

$$g_j' s_j \geqslant \delta(\epsilon) \quad \text{for } j \in J \text{ and some } \delta(\epsilon) > 0.$$

By the definition of s_j, this implies that any cluster point of this subsequence is not a stationary point in contradiction to part (ii) of the theorem. The last statement of the theorem follows from the fact that $\|x_{j+1} - x_j\| \longrightarrow 0$ as $j \longrightarrow \infty$.

In order to derive results concerning the rate of convergence we need the following.

Assumption III. Let z be a cluster point of the sequence $\{x_j\}$ and let $a_i' z = b_i$, $i = 1, \ldots, p$, and $a_i' z < b_i$, $i = p + 1, \ldots, m$. Then

(1) There are numbers λ_i such that

$$\nabla F(z) = \sum_{i=1}^{p} \lambda_i a_i \quad \text{and} \quad \lambda_i < 0, \quad i = 1, \ldots, p.$$

(2) $F(x)$ is twice continuously differentiable in some neighborhood of z.

(3) There are numbers $0 < \mu < \eta$ such that

$$\mu \|x\|^2 \leqslant x' G(z) x \leqslant \eta \|x\|^2$$

for all $x \in E^n$ with $a_i' x = 0, i = 1, \ldots, p$.

Since by Theorem 7 every cluster point of $\{x_j\}$ is a stationary point, the first part of Assumption III states that the strict complementary slackness condition is satisfied at z. Furthermore, it is easy to see that Assumption I implies Assumption II and the last two parts of Assumption III.

As a first consequence of Assumption III we have

PROPOSITION 3. *Let Assumptions* II *and* III *be satisfied. Then* $x_j \longrightarrow z$ *and there is* j_0' *such that for* $j \geqslant j_0'$,

$$a_i' x_j = b_i, \quad i = 1, \ldots, p \quad \text{and} \quad a_i' x_j < b_i, \quad i = p + 1, \ldots, m.$$

Proof. It follows from Assumption III that z is an isolated stationary point. By Theorem 7 this implies that z is an isolated cluster point of $\{x_j\}$. Thus, again by Theorem 7, $x_j \longrightarrow z$. The last statement follows then from Proposition 2.

By Assumption III and Proposition 3, there is a convex neighborhood $u(z)$ of z with the following properties:

(1) $x_j \in u(z)$ for j sufficiently large.

(2) $\mu \|x\|^2 \leqslant x' G(y) x \leqslant \eta \|x\|^2$ for all $y \in u(z)$ and all $x \in E^n$ with $a_i' x = 0, i = 1, \ldots, p$.

Because $a_i' s_j = 0$, $i = 1, \ldots, p$, for j sufficiently large, it follows then from Taylor's theorem that, for j sufficiently large,

$$|d_j' s_j| \geqslant \mu \|s_j\| \quad \text{and} \quad \|d_j\| \leqslant \eta,$$

i.e., the test in Step III of the modified algorithm is satisfied for j sufficiently large provided $\gamma_1 \leqslant \mu$ and $\gamma_2 \geqslant \eta$. But then the two algorithms are identical for j sufficiently large and it follows from (1) and (2) that all the results of Section 5 are applicable. We formulate the result as

THEOREM 8. *Let Assumptions* II *and* III *be satisfied and suppose that* $\gamma_1 \leqslant \mu$ *and* $\gamma_2 \geqslant \eta$. *Then the convergence results of Section 5 also apply to the sequence* $\{x_j\}$ *generated by the modified algorithm.*

7. **Computational Results.** In this section we give the results of some computational tests in which the method presented in this paper was used to solve four test problems. Two of the problems are taken from the Colville study [5] (Nos. 1 and 7) and another two, the Chemical Equilibrium and Weapons Assignment problems, are taken from [4]. Table 1 gives the timing results in standardized units (see [5]). For the Colville problems, timing results are also given for the revised reduced gradient method (RRG). This method was chosen for comparison because it gave the fastest time among the methods considered by Colville. The times reported for RRG are taken from [5].

The computations were performed on a Honeywell 6050 computer at the University of Waterloo. Colville's standard timing programme executed in an average time of 53.8466 seconds on this system, and the standard times in Table 1 were computed using that figure.

The mixture of accelerating and conjugate direction constructing steps was determined as follows. The first n iterates were obtained using only conjugate direction constructing steps since the accelerating step is based on the availability of a complete set of conjugate directions. After n iterations, an alternating policy was used since, by Theorem 6, this gives the sharpest convergence rate.

Step II of the algorithm uses a unit stepsize provided that it is feasible and passes certain tests required for convergence. Lemma 6 shows that after a certain number of iterations the unit stepsize will always be used. Examination of the intermediate output showed that the unit stepsize was used for 76% of the iterations for the four test problems.

The Weapons Assignment problem is of particular interest because of its size of $n = 100$ variables. Although this problem has considerable structure no account was taken of this structure when solving it with the method described in this paper. It is interesting to note that an accurate solution was obtained after approximately $1.7n$ iterations.

Theorem 6 predicts that the rate of convergence of the accelerating steps will be faster than the rate for the conjugate direction constructing steps. This prediction is verified computationally in Table 2 which shows some of the intermediate results for the Weapons Assignment problem.

TABLE 1

Numerical Results for Four Test Problems

	Colville No. 1	Colville No. 7	Chemical Equilibrium	Weapons Assignment
Variables	5	16	10	100
Constraints	15	40	13	112
Standardized {new method	0.0042	0.0220	0.0219	5.3828
Exc. Time {R.R.G.	0.0061	0.0290	——	——
Iterations	11	14	36	168
Function Evaluations	12	17	65	208
Gradient Evaluations	12	15	37	169
Final Objective Value	−32.34867897	−244.8996975	−47.706109086	−1735.569579

TABLE 2

Intermediate Results for Weapons Assignment Problem

Iteration	Objective	Step	Iteration	Objective	Step
0	− 624	—	140	−1735.5684	C
10	−1153.	C	141	−1735.568519	A
20	−1306.	C	142	−1735.568523	C
30	−1424.	C	143	−1735.568572	A
40	−1539.	C	144	−1735.568590	C
50	−1610	C	145	−1735.568992	A
60	−1661	C	146	−1735.568992	C
70	−1693	C	147	−1735.569228	A
80	−1713	C	148	−1735.569230	C
90	−1719	C	149	−1735.569313	A
100	−1726	C	150	−1735.569383	C
110	−1734	C	151	−1735.569447	A
120	−1735.17	C	152	−1735.569450	C
130	−1735.56	C	168	−1735.56957933	A

"A" = accelerated step, "C" = conjugate direction constructing step

Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Ontario, Canada

Mathematisches Institut A
Universität Stuttgart
7 Stuttgart, Pfaffenwaldring 57, West Germany

1. L. ARMIJO, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific J. Math.*, v. 16, 1966, pp. 1–3. MR 32 #8480.
2. M. J. BEST, "A method to accelerate the rate of convergence of a class of optimization algorithms," *Math. Programming*, v. 8, 1975, pp. 139–160.

3. M. J. BEST & K. RITTER, "An accelerated conjugate direction method to solve linearly constrained minimization problems," *J. Comput. System Sci.*, v. 11, 1975, pp. 295–322.

4. J. BRACKEN & G. P. McCORMICK, *Selected Applications of Nonlinear Programming,* Wiley, New York, 1968. MR 38 #1911.

5. A. R. COLVILLE, *A Comparative Study on Nonlinear Programming Codes,* IBM, New York Scientific Center, Report No. 320–2949, 1968.

6. A. A. GOLDSTEIN, *Constructive Real Analysis,* Harper & Row, New York, 1967. MR 36 #705.

7. K. RITTER, "A superlinearly convergent method for minimization problems with linear inequality constraints," *Math. Programming,* v. 4, 1973, pp. 44–71. MR 48 #1711.

8. K. RITTER, "A method of conjugate directions for linearly constrained nonlinear programming problems," *SIAM J. Numer. Anal.,* v. 12, 1975, pp. 273–303.

9. K. RITTER, "Accelerating procedures for methods of conjugate directions," *Computing,* v. 14, 1975, pp. 79–105.

10. M. M. VAĬNBERG, *Variational Methods for the Study of Nonlinear Operators,* GITTL, Moscow, 1956; English transl., Holden-Day, San Francisco, Calif., 1964. MR 19, 567; 31 #638.

11. G. ZOUTENDIJK, *Methods of Feasible Directions, A Study in Linear and Nonlinear Programming,* American Elsevier, New York, 1960. MR 23 #B2156.