SIAM J. OPTIM.
Vol. 7, No. 1, pp. 162–176, February 1997

NEWTON METHODS FOR LARGE-SCALE LINEAR INEQUALITY-CONSTRAINED MINIMIZATION*

ANDERS FORSGREN† AND WALTER MURRAY‡

Abstract. Newton methods of the linesearch type for large-scale minimization subject to linear inequality constraints are discussed. The purpose of the paper is twofold: (i) to give an active–set-type method with the ability to delete multiple constraints simultaneously and (ii) to give a relatively short general convergence proof for such a method. It is also discussed how multiple constraints can be added simultaneously. The approach is an extension of a previous work by the same authors for equality-constrained problems. It is shown how the search directions can be computed without the need to compute the reduced Hessian of the objective function. The convergence analysis states that every limit point of a sequence of iterates satisfies the *second-order* necessary optimality conditions.

Key words. linear inequality-constrained minimization, negative curvature, modified Newton method, symmetric indefinite factorization, large-scale minimization, linesearch method

AMS subject classifications. 49M37, 65K05, 90C30

PII. S1052623494279122

1. Introduction. We consider a method for finding a local minimizer of the problem

$$\begin{array}{ll} \underset{x\in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & Ax \geq b, \end{array} \tag{1.1}$$

where A is an $m \times n$ matrix and $f \in C^2$. We are interested in the case when n and possibly m are large and when second derivatives of f are available. The method is a Newton method of the linesearch type using an active-set strategy to identify the constraints that are active at the solution, where the active set at each iteration may change significantly. No assumptions are made about the number of constraints active at the solution or in the problem. In the approach advocated, it is not necessary to make any initial transformation of the problem such as transforming it into canonical form. The method proposed builds on a method we proposed recently for the *equality*-constrained problem [11] and requires only a single matrix factorization per iteration.

Linearly constrained optimization has been studied quite extensively over the years; see, e.g., Gill, Murray, and Wright [17, Chapter 5] and Fletcher [10, Chapter 11]. As mentioned above, our interest is in linesearch methods of the active-set type, i.e., methods that solve a sequence of equality-constrained subproblems. Methods of this type, designed to give limit points that satisfy the *first-order* optimality conditions, have been given by, e.g., Rosen [28], Goldfarb [18], Ritter [25, 26, 27], and Byrd and Shultz [6]. Similarly, linesearch methods designed to give limit points that satisfy the

* Received by the editors December 21, 1994; accepted for publication (in revised form) October 18, 1995.
http://www.siam.org/journals/siopt/7-1/27912.html

† Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, S-100 44 Stockholm, Sweden (andersf@math.kth.se). This author was partially supported by the Royal Swedish Academy of Sciences (Magnuson's fund, KVA) and the Swedish Natural Science Research Council (NFR).

‡ Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, CA 94305-4022 (walter@sol-walter.stanford.edu). This author was supported by Department of Energy contract DE-FG03-92ER25117, National Science Foundation grant DDM-9204208, and Office of Naval Research grant N00014-90-J-1242.

second-order necessary optimality conditions have been given by, e.g., McCormick [20] and Gill and Murray [13]. Methods for large-scale linearly constrained problems are given by, e.g., Buckley [1] and Murtagh and Saunders [23]. The motivation for our work is to give a method for large-scale problems together with a concise and comprehensive convergence analysis. The method proposed here gives limit points that satisfy the second-order necessary optimality conditions and it is based on a single matrix factorization per iteration. Although only linesearch methods are considered in this paper, *trust-region* methods with similar convergence properties have been proposed; see, e.g., Gay [12].

2. Notation and assumptions. The method proposed generates a sequence $\{x_k\}_{k=0}^{\infty}$ of iterates of the form

$$x_{k+1} = x_k + \alpha_k p_k,$$

where p_k is a search direction and α_k is determined by a linesearch along p_k. It is assumed that $f_k \equiv f(x_k)$, the gradient $g_k \equiv \nabla f(x_k)$, and the Hessian $H_k \equiv \nabla^2 f(x_k)$ can be evaluated. The definition of p_k is given in section 3 and the conditions on α_k are discussed in section 3.5. We denote by a_i^T the ith row of A and by b_i the ith component of b. At a point x_k, a constraint $a_i^T x \geq b_i$ is said to be *active* if $a_i^T x_k = b_i$, *inactive* if $a_i^T x_k > b_i$, and *violated* if $a_i^T x_k < b_i$. We denote by A_k a matrix comprising a subset of the rows of A that correspond to constraints active at x_k. Similarly, b_k is the vector of the corresponding elements of b. We denote by $\mathcal{W}_k \subseteq \{1, 2, \ldots, m\}$ the indices of the rows of A in A_k and refer to $\mathcal{W}_k$ as the *working set* at iteration k. The notation $\mathcal{W}_{k+1} \backslash \mathcal{W}_k$ is used for the set of indices that belong to $\mathcal{W}_{k+1}$ but not to $\mathcal{W}_k$. (Note that $\mathcal{W}_{k+1} \backslash \mathcal{W}_k$ is defined also when $\mathcal{W}_k \not\subseteq \mathcal{W}_{k+1}$.) The matrix Z_k denotes an orthonormal matrix whose columns form a basis for the null space of A_k. Note that Z_k need not be known; our use of this matrix is for theoretical purposes only. We shall assume that A_0 has full row rank. Then, the rules we give in section 3.6 for updating A_k ensure that A_k has full rank for all k. In section 5.1 it is shown how A_0 may be obtained without making any assumptions about A, and in section 5.6 it is shown how A_k may be updated while maintaining the full row rank. For a symmetric matrix M, we use the notation $\lambda_{\min}(M) \geq 0$ for M positive semidefinite and $\lambda_{\min}(M) > 0$ for M positive definite but this is just for notational purposes, and the eigenvalues are not computed. For a sequence $I \subseteq \{0, 1, \ldots\}$, the abbreviated notation $\lim_{k \in I}$ is used for $\lim_{k \to \infty, k \in I}$.

Throughout, the following assumptions are made:

A1. The objective function f is twice continuously differentiable.

A2. The initial feasible point x_0 is known, and the level set $\{x : Ax \geq b, f(x) \leq f(x_0)\}$ is compact.

A3. The constraint matrix associated with the active constraints has full row rank at all points that satisfy the *second-order* necessary optimality conditions if these constraints are regarded as equalities. Formally, let $\bar{x}$ denote a feasible point of (1.1), let A_A denote the matrix associated with the active constraints at $\bar{x}$, and let Z_A denote a matrix whose columns form an orthonormal basis for the null space of A_A. If it holds that

$$Z_A^T \nabla f(\bar{x}) = 0 \quad \text{and} \quad \lambda_{\min}(Z_A^T \nabla^2 f(\bar{x}) Z_A) \geq 0,$$

then A_A has full row rank.

Assumption A3 states that the problem does not have *primal degenerate* second-order constrained stationary points (dual degeneracy may occur). Any algorithm for general problems we are familiar with, for which primal nondegeneracy does not need to be assumed, requires an iteration that in itself has a subiteration. Our purpose here is to devise algorithms that do not require such subiterations since our primary concern is to solve large problems. Nonetheless, degeneracy (or near degeneracy) is possible and needs to be dealt with in any practical implementation. In practice degeneracy may be dealt with by techniques that allow the standard iteration to be used; see, e.g., Gill et al. [14]. Such a technique is used within the MINOS code, see Murtagh and Saunders [24], which has been used to solve thousands of practical problems. The consequence of using this approach to degeneracy is that the solution obtained may be infeasible. However, the degree of infeasibility may be set at a level similar to that which arises due to finite precision. Indeed, even if degeneracy was not present such techniques are necessary in an endeavor to make the matrix of active constraints well conditioned. Discussions on theoretical aspects of degeneracy are given in Burke and Moré [3, 4], Burke [2], and Burke, Moré, and Toraldo [5].

3. Definition of the algorithm. The search direction p_k is a sum of three directions. More specifically,

$$p_k = s_k + d_k + q_k,$$

where a nonzero s_k is a descent direction of bounded norm in the null space of A_k, a nonzero d_k is a direction of negative curvature with bounded norm in the null space of A_k, and a nonzero q_k is a descent direction of bounded norm such that $A_k q_k \geq 0$ and $a_j^T q_k > 0$ for some $j \in \mathcal{W}_k$. At each iteration, a set of Lagrange multiplier estimates π_k, associated with A_k, is required. In this section, the required properties of s_k, d_k, π_k, and q_k are given, and in section 5 an appropriate way of computing the directions for large-scale problems is discussed.

3.1. Properties of s_k. A nonzero s_k has to have bounded norm and be a descent direction in the null space of A_k, i.e., satisfy $g_k^T s_k < 0$ and $A_k s_k = 0$. We also require that s_k be a *sufficient* descent direction in the following sense:

$$\lim_{k\in I} g_k^T s_k = 0 \quad \Rightarrow \quad \lim_{k\in I} Z_k^T g_k = 0 \quad \text{and} \quad \lim_{k\in I} s_k = 0, \tag{3.1}$$

where I is any subsequence.

3.2. Properties of d_k. We require a nonzero d_k to be a nonascent direction of negative curvature in the null space of the A_k, i.e., $g_k^T d_k \leq 0$, $d_k^T H_k d_k < 0$, and $A_k d_k = 0$. Furthermore, the norm of d_k has to be bounded and the curvature has to be *sufficient* in the sense that

$$\lim_{k\in I} d_k^T H_k d_k = 0 \quad \Rightarrow \quad \liminf_{k\in I} \lambda_{\min}(Z_k^T H_k Z_k) \geq 0 \quad \text{and} \quad \lim_{k\in I} d_k = 0, \tag{3.2}$$

where I is any subsequence.

3.3. Properties of π_k. At each iteration, a vector of Lagrange multiplier estimates, π_k, is required. The vector π_k must satisfy

$$\lim_{k\in I} \|Z_k^T g_k\| = 0 \quad \Rightarrow \quad \lim_{k\in I} \|g_k - A_k^T \pi_k\| = 0, \tag{3.3}$$

where I is any subsequence. We define $\pi_{\min,k} = \min_i(\pi_k)_i$ and use this notation throughout.

3.4. Properties of q_k. If $\pi_{\min,k} \geq 0$ or $\mathcal{W}_k \not\subseteq \mathcal{W}_{k-1}$, we set $q_k = 0$. This is to say that we take at least one step towards minimality for a given A_k before considering deleting constraints. When $q_k \neq 0$ we require it to be a descent direction that moves off at least one constraint in the working set and remains feasible with respect to the others, i.e., $g_k^T q_k < 0$ and $0 \neq A_k q_k \geq 0$. Furthermore, the norm of q_k has to be bounded and it is also required that the q_k's are such that

$$\lim_{k\in I} g_k^T q_k = 0 \Rightarrow \liminf_{k\in I} \pi_{\min,k} \geq 0 \quad \text{and} \quad \lim_{k\in I} q_k = 0, \tag{3.4a}$$

$$a_i^T q_k > 0 \Rightarrow (\pi_k)_i \leq \nu \pi_{\min,k} \quad \text{for} \quad k \in I,\ i \in \mathcal{W}_k, \tag{3.4b}$$

where I is any subsequence such that $\mathcal{W}_k \subseteq \mathcal{W}_{k-1}$ for all $k \in I$ and ν is a preassigned tolerance, $(0 < \nu \leq 1)$.

3.5. Definition of the iterates. We follow Moré and Sorensen [21] and Forsgren and Murray [11] in the linesearch and adapt it to cope with inequality constraints. For the sake of completion, the linesearch is reviewed here, and the properties that are subsequently required for the linear inequality-constrained case are given in Lemmas 4.1, 4.2, and 4.3 below.

Iteration k takes the following form. The search direction is obtained as $p_k = s_k + d_k + q_k$, where s_k, d_k, and q_k satisfy the conditions of sections 3.1–3.4. Define $\phi_k(\alpha) = f(x_k + \alpha p_k)$. Sections 3.1–3.4 give $p_k = 0$ if and only if $\phi_k'(0) = 0$ and $\phi_k''(0) \geq 0$. The linesearch is designed to give $\lim_{k\to\infty} \phi_k'(0) = 0$ and $\liminf_{k\to\infty} \phi_k''(0) \geq 0$. An upper bound on the steplength is computed as

$$\bar{\alpha}_k = \min\left\{\alpha_{\max}, \min_{i:a_i^T p_k < 0} \frac{a_i^T x_k - b_i}{-a_i^T p_k}\right\},$$

where $\alpha_{\max}$, $(\alpha_{\max} \geq 1)$ is a fixed upper bound on the maximum steplength. If $\bar{\alpha}_k = 0$, then $\alpha_k = 0$. Otherwise, the steplength α_k is determined such that $\alpha_k \in (0, \bar{\alpha}_k]$ satisfies

$$\phi_k(\alpha_k) \leq \phi_k(0) + \mu(\phi_k'(0)\alpha_k + \tfrac{1}{2}\min\{\phi_k''(0), 0\}\alpha_k^2) \tag{3.5}$$

and at least one of

$$|\phi_k'(\alpha_k)| \leq \eta|\phi_k'(0) + \min\{\phi_k''(0), 0\}\alpha_k| \quad \text{or} \tag{3.6a}$$

$$\alpha_k = \bar{\alpha}_k, \tag{3.6b}$$

where $0 < \mu < 0.5$ and $\mu \leq \eta < 1$. Finally, $x_{k+1} = x_k + \alpha_k p_k$. The conditions of sections 3.1–3.4 give $\phi_k'(0) \leq 0$ for all k, and $\phi_k'(0) = 0$ if and only if $p_k = d_k$. It follows from Moré and Sorensen [21, Lemma 5.2] that α_k is well defined.

We refer to a step α_k as *restricted* if

$$\alpha_k = \min_{i:a_i^T p_k < 0} \frac{a_i^T x_k - b_i}{-a_i^T p_k},$$

i.e., a constraint is encountered in the linesearch at iteration k. Otherwise, the step is referred to as *unrestricted*. Hence, a restricted step always satisfies (3.6b) whereas an unrestricted step satisfies at least one of $\alpha_k = \alpha_{\max}$ or (3.6a).

3.6. Properties of $\boldsymbol{A_k}$. The initial working-set matrix A_0 is required to have full row rank and contain constraints active at x_0. To give the rule for updating $\mathcal{W}_k$, define

$$\mathcal{W}_k^0 = \{i \in \mathcal{W}_k : a_i^T p_k = 0\}.$$

Let $\mathcal{P}_k^a$ denote the index set of constraints that are encountered in the linesearch at iteration k, i.e.,

$$\mathcal{P}_k^a = \{i \notin \mathcal{W}_k : a_i^T p_k < 0,\ a_i^T x_{k+1} = b_i\}.$$

Note that either of $\mathcal{W}_k^0$ and $\mathcal{P}_k^a$ may be the empty set. We then define $\mathcal{W}_{k+1} = \mathcal{W}_k^0 \cup \mathcal{W}_k^a$, where $\mathcal{W}_k^a \subseteq \mathcal{P}_k^a$ and the associated A_{k+1} are required to satisfy

$$\mathcal{P}_k^a \neq \emptyset \Rightarrow \mathcal{W}_k^a \neq \emptyset \quad \text{and} \tag{3.7a}$$

$$A_{k+1} \text{ has full row rank.} \tag{3.7b}$$

The implication of (3.7a) is that if new constraints are encountered in the linesearch, at least one of them has to be added. If A_k has full row rank, (3.7b) will trivially hold if $\mathcal{W}_k^a = \emptyset$. Otherwise, care has to be taken to ensure that A_{k+1} has full row rank. This is further discussed in section 5.6.

Note that an implication of the above conditions is that a step α_k is restricted if and only if $\mathcal{W}_{k+1} \backslash \mathcal{W}_k \neq \emptyset$.

4. Convergence results for linear inequality constraints. Lemmas 4.1, 4.2, and 4.3 below review results from unconstrained optimization originally proposed by Moré and Sorensen [21]. These give results for unrestricted steps. The remainder of this section then establishes the convergence results for linear inequality-constrained problems.

The following lemma gives some properties of the iterates for a sequence generated by the above linesearch conditions.

LEMMA 4.1. *Given assumptions* A1–A3*, assume that a sequence* $\{x_k\}_{k=0}^\infty$ *is generated as outlined in section* 3*. Then*

(i) $\lim_{k\to\infty} \alpha_k \phi_k'(0) = 0$;
(ii) $\lim_{k\to\infty} \alpha_k^2 \min\{\phi_k''(0), 0\} = 0$;
(iii) $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$.

Proof. Rearrangement of (3.5) gives

$$\phi_k(0) - \phi_k(\alpha_k) \geq -\mu(\phi_k'(0)\alpha_k + \tfrac{1}{2}\min\{\phi_k''(0), 0\}\alpha_k^2).$$

Since $\mu > 0$, $\phi_k'(0) \leq 0$, and the objective function is bounded from below on the feasible region, (i) and (ii) follow.

To show (iii), we write $x_{k+1} - x_k = \alpha_k p_k$ and show that $\lim_{k\to\infty} \|\alpha_k p_k\| = 0$. Since α_k and $\|p_k\|$ are bounded, if $\lim_{k\to\infty} \|\alpha_k p_k\| \neq 0$, there must exist a subsequence I and $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $\alpha_k \geq \epsilon_1$ and $\|p_k\| \geq \epsilon_2$ for $k \in I$. From the existence of ϵ_1, (i) implies $\lim_{k\in I} \phi_k'(0) = 0$ and (ii) implies $\liminf_{k\in I} \phi_k''(0) \geq 0$. Since $\phi_k'(0) = g_k^T p_k = g_k^T(s_k + d_k + q_k)$ and it holds that $g_k^T s_k \leq 0$, $g_k^T d_k \leq 0$, and $g_k^T q_k \leq 0$, (3.1) implies $\lim_{k\in I} s_k = 0$ and (3.4) implies $\lim_{k\in I} q_k = 0$. Hence, since $\phi_k''(0) = p_k^T H_k p_k$ and $\lim_{k\in I} \|p_k - d_k\| = 0$, (3.2) implies $\lim_{k\in I} d_k = 0$. Thus, $\lim_{k\in I} \|p_k\| = 0$. This contradicts the existence of ϵ_2, thus establishing (iii). ∎

The following lemma relates α_k to $\phi_k'(0)$ for an unrestricted step. The implication is that α_k is bounded away from zero if $\phi_k'(0)$ is bounded away from zero.

LEMMA 4.2. *Given assumptions* A1–A3*, assume that a sequence* $\{x_k\}_{k=0}^{\infty}$ *is generated as outlined in section* 3*. If, at iteration k, an unrestricted step is taken, then either* $\alpha_k = \alpha_{\max}$ *or there exists a* θ_k*,* $(0 < \theta_k < \alpha_k)$ *such that*

$$\alpha_k(\phi_k''(\theta_k) + \eta \max\{-\phi_k''(0), 0\}) \geq -(1-\eta)\phi_k'(0). \tag{4.1}$$

Proof. Since $\phi_k'(0) \leq 0$, it follows from (3.6) that if α_k is unrestricted and $\alpha_k < \alpha_{\max}$, it satisfies

$$-\phi_k'(\alpha_k) \leq -\eta\phi_k'(0) + \eta \max\{-\phi_k''(0), 0\}\alpha_k. \tag{4.2}$$

Further, since ϕ_k' is a continuously differentiable univariate function, the mean-value theorem ensures the existence of a $\theta_k \in (0, \alpha_k)$ such that

$$\phi_k'(\alpha_k) = \phi_k'(0) + \alpha_k\phi_k''(\theta_k). \tag{4.3}$$

A combination of (4.2) and (4.3) now gives (4.1), as required. □

Finally, the following lemma gives some properties of subsequences of unrestricted iterates for a sequence generated by the above linesearch conditions.

LEMMA 4.3. *Given assumptions* A1–A3*, assume that a sequence* $\{x_k\}_{k=0}^{\infty}$ *is generated as outlined in section* 3*. Let I denote a subsequence of iterations where unrestricted steps are taken*; *then*

(i) $\lim_{k\in I} \phi_k'(0) = 0$;
(ii) $\liminf_{k\in I} \phi_k''(0) \geq 0$;
(iii) $\lim_{k\in I} Z_k^T g_k = 0$ *and* $\liminf_{k\in I} \lambda_{\min}(Z_k^T H_k Z_k) \geq 0$.

Proof. To show (i), assume by contradiction there is a subsequence $I' \subseteq I$ such that $\phi_k'(0) \leq -\epsilon_1 < 0$ for $k \in I'$. Lemma 4.2 in conjunction with assumptions A1 and A2 then implies that $\limsup_{k\in I'} \alpha_k \neq 0$, contradicting Lemma 4.1. Hence, the assumed existence of I' is false, and we conclude that (i) holds.

Similarly, to show (ii), assume by contradiction that there is a subsequence $I'' \subseteq I$ such that $\phi_k''(0) \leq -\epsilon_2 < 0$ for $k \in I''$. Since $\alpha_k > 0$ and $\phi_k'(0) \leq 0$, Lemma 4.2 implies that for $k \in I''$ there exists $\theta_k \in (0, \alpha_k)$ such that

$$\phi_k''(\theta_k) - \eta\phi_k''(0) \geq 0. \tag{4.4}$$

Lemma 4.1 gives $\lim_{k\in I''} \alpha_k = 0$, and thus (4.4) cannot hold for k sufficiently large. Consequently, the assumed existence of I'' is false, and (ii) holds.

Finally, we show that (i) and (ii) imply (iii). Since $\phi_k'(0) = g_k^T p_k = g_k^T(s_k + d_k + q_k)$ and it holds that $g_k^T s_k \leq 0$, $g_k^T d_k \leq 0$, and $g_k^T q_k \leq 0$, (i) and (3.1) imply $\lim_{k\in I} Z_k^T g_k = 0$ and $\lim_{k\in I} s_k = 0$. Further, (i) and (3.4a) imply $\lim_{k\in I} q_k = 0$. Hence, since $\phi_k''(0) = p_k^T H_k p_k$ and $\lim_{k\in I} \|p_k - d_k\| = 0$, (ii) and (3.2) imply $\liminf_{k\in I} \lambda_{\min}(Z_k^T H_k Z_k) \geq 0$ and, thus, (iii) holds. □

We now extend these results to the case of linear inequality constraints. The first lemma shows that if there exists a subsequence of iterates at which a constraint is deleted with the smallest multiplier negative and bounded away from zero and for which no constraints were deleted at the previous iteration, then eventually a constraint will be added.

LEMMA 4.4. *Given assumptions* A1–A3*, assume that a sequence* $\{x_k\}_{k=0}^{\infty}$ *is generated as outlined in section* 3*. If there is a subsequence I and an* $\epsilon > 0$ *such that* $q_{k-1} = 0$*,* $q_k \neq 0$*, and* $\pi_{\min,k} < -\epsilon$ *for* $k \in I$*, then there is an integer K such that* $\mathcal{W}_{k+1} \backslash \mathcal{W}_k \neq \emptyset$ *for all* $k \in I$ *and* $k \geq K$*.*

Proof. Suppose that there is a subsequence I and an $\epsilon > 0$ such that $q_{k-1} = 0$, $q_k \neq 0$, and $\pi_{\min,k} < -\epsilon$ for $k \in I$. Now assume that there is a subsequence $I' \subseteq I$ such that an unrestricted step is taken for $k \in I'$. Lemma 4.3 implies that $\lim_{k \in I'} \phi_k'(0) = 0$. On the other hand, (3.4a) ensures the existence of a subsequence $I'' \subseteq I'$ and a positive constant ϵ_2 such that $g_k^T q_k \leq -\epsilon_2$ for all $k \in I''$. However, since $g_k^T s_k \leq 0$ and $g_k^T d_k \leq 0$, this implies that $\phi_k'(0) \leq -\epsilon_2$ for all $k \in I'$, which is a contradiction. Hence, the assumed existence of the subsequence I' is false, and there must exist a K such that for $k \in I$ and $k \geq K$ a restricted step is taken, i.e., $\mathcal{W}_{k+1} \backslash \mathcal{W}_k \neq \emptyset$ for all $k \in I$ and $k \geq K$. ◻

Assumption A3 can now be used to show that for a subsequence of iterates where constraints are deleted, but no constraints were deleted at the previous iteration, the smallest multiplier is nonnegative in the limit.

LEMMA 4.5. *Given assumptions* A1–A3*, assume that a sequence* $\{x_k\}_{k=0}^{\infty}$ *is generated as outlined in section* 3*. If there is a subsequence* I *such that* $q_{k-1} = 0$ *and* $q_k \neq 0$ *for* $k \in I$*, then* $\liminf_{k \in I} \pi_{\min,k} \geq 0$*.*

Proof. Assume that there exists a subsequence I and an $\epsilon > 0$ such that $q_{k-1} = 0$, $q_k \neq 0$, and $\pi_{\min,k} < -\epsilon$ for $k \in I$. For each $k \in I$, let l_k denote the following iteration with least index such that $\mathcal{W}_{l_k} = \mathcal{W}_{l_k - 1}$; i.e., an unrestricted step is taken at iteration $l_k - 1$ and $q_{l_k - 1} = 0$. Lemma 4.4 implies that there is an integer K such that $\mathcal{W}_{k+1} \backslash \mathcal{W}_k \neq \emptyset$ for all $k \in I$ and $k \geq K$. The properties of q_k from section 3.4 imply that $q_{k+1} = 0$ for $k \in I$, $k \geq K$. Consequently, for $k \geq K$, l_k is the iteration with least index following k where no constraint is added in the linesearch. Since there can be at most $\min\{m, n\}$ consecutive iterations where a constraint is added, it follows from (iii) of Lemma 4.1 that $\lim_{k \in I} \|x_k - x_{l_k}\| = 0$. Consequently, there must exist a point $\bar{x}$, which is a common limit point to $\{x_k\}_{k \in I}$ and $\{x_{l_k}\}_{k \in I}$. By taking appropriate subsequences, there exists a subsequence $I' \subseteq I$ such that $\lim_{k \in I'} x_k = \bar{x}$ and $\lim_{k \in I'} x_{l_k} = \bar{x}$. Again, by taking appropriate subsequences, there must exist a subsequence $I'' \subseteq I'$ such that $\mathcal{W}_k$ is identical for every $k \in I''$ and $\mathcal{W}_{l_k}$ is identical for every $l_k \in J$, where J denotes the subsequence $\{l_k\}_{k \in I''}$. Define $\mathcal{W}^I \equiv \mathcal{W}_k$ for any $k \in I''$ and $\mathcal{W}^J \equiv \mathcal{W}_{l_k}$ for any $l_k \in J$.

Since all constraints corresponding to $\mathcal{W}^I$ are active at $\bar{x}$ and an infinite number of unrestricted steps are taken where the working set is constant, it follows from assumptions A1 and A2 in conjunction with (iii) of Lemma 4.1 and (iii) of Lemma 4.3 that $\lim_{k \in I''} Z_I^T g_k = 0$ and $\liminf_{k \in I''} \lambda_{\min}(Z_I^T H_k Z_I) \geq 0$, where Z_I denotes a matrix whose columns form an orthonormal basis for the null space of A_I, the constraint matrix associated with $\mathcal{W}^I$. Consequently, (3.3) and the full row rank of A_I imply that $\lim_{k \in I''} \pi_k = \pi^I$, where π^I satisfies

$$\nabla f(\bar{x}) = A_I^T \pi^I = \sum_{i \in \mathcal{W}^I} a_i \pi_i^I. \tag{4.5}$$

By a similar reasoning and notation for Z_J and A_J we have $\lim_{k \in I''} Z_J^T g_{l_k} = 0$, $\liminf_{k \in I''} \lambda_{\min}(Z_J^T H_{l_k} Z_J) \geq 0$, and $\lim_{k \in I''} \pi_{l_k} = \pi^J$, where π^J satisfies

$$\nabla f(\bar{x}) = A_J^T \pi^J = \sum_{i \in \mathcal{W}^J} a_i \pi_i^J. \tag{4.6}$$

Combining (4.5) and (4.6), we obtain

$$\sum_{i \in \mathcal{W}^I \backslash \mathcal{W}^J} a_i \pi_i^I + \sum_{i \in \mathcal{W}^I \cap \mathcal{W}^J} a_i \left(\pi_i^I - \pi_i^J\right) - \sum_{i \in \mathcal{W}^J \backslash \mathcal{W}^I} a_i \pi_i^J = 0. \tag{4.7}$$

By assumption A3, the vectors a_i, $i \in \mathcal{W}^I \cup \mathcal{W}^J$ are linearly independent. Hence, it follows from (4.7) that

$$\pi_i^I = 0 \quad \text{for } i \in \mathcal{W}^I \backslash \mathcal{W}^J, \tag{4.8a}$$

$$\pi_i^I = \pi_i^J \quad \text{for } i \in \mathcal{W}^I \cap \mathcal{W}^J, \tag{4.8b}$$

$$\pi_i^J = 0 \quad \text{for } i \in \mathcal{W}^J \backslash \mathcal{W}^I. \tag{4.8c}$$

Since Lemma 4.4 implies that there is an integer K such that $\mathcal{W}_{k+1} \backslash \mathcal{W}_k \neq \emptyset$ for all $k \in I$ and $k \geq K$, we conclude that $\mathcal{W}^J \backslash \mathcal{W}^I \neq \emptyset$. Since no constraints have been deleted between iterations k and l_k for $k \in I''$, any constraints whose index is in the set $\mathcal{W}^I \backslash W^J$ must have been deleted in an iteration $k \in I''$. Since $I'' \subseteq I$, it follows that $\pi_{\min,k} \leq -\epsilon$ for $k \in I''$. From the rule for moving off a constraint, (3.4b), we can deduce that $(\pi_k)_i \leq -\nu\epsilon$ for $k \in I''$ and $i \in \mathcal{W}^I \backslash W^J$, where $\nu \in (0,1)$. Since $\lim_{k \in I''} \pi_k = \pi^I$, we conclude that $\pi_i^I \leq -\nu\epsilon$ for $i \in \mathcal{W}^I \backslash W^J$. Hence, (4.8a) implies that $\mathcal{W}^I \backslash \mathcal{W}^J = \emptyset$. Consequently, it must hold that $|\mathcal{W}^J| \geq |\mathcal{W}^I| + 1$ and, by (4.8c), π^J has at least one component zero.

We can conclude from (4.8b) that $\pi_{\min,l_k} < -0.5\epsilon$ for $k \in I''$ and k sufficiently large. The rules for computing q_k, (3.4a), ensure that there is a subsequence $I''' \subseteq I''$ such that $q_{l_k} \neq 0$ for all $k \in I'''$. From the definition of l_k, it holds that $q_{l_k - 1} = 0$ for all $k \in I'''$. Therefore, if $J' = \{l_k : k \in I'''\}$, we may replace I by J' and repeat the argument. Since $|\mathcal{W}^J| \geq |\mathcal{W}^I| + 1$ and $|\mathcal{W}_k| \leq \min\{m, n\}$ for any k, after having repeated the argument at most $\min\{m, n\}$ times we have a contradiction to assumption A3, implying that the assumed existence of a subsequence I such that $q_{k-1} = 0$ and $q_k \neq 0$ and $\pi_{\min,k} < -\epsilon$ for $k \in I$ is false. Thus, the result of the lemma follows. ∎

We are now in the position to give the main convergence result. In addition to the global convergence established here, we also add a well-known rate-of-convergence result from Moré and Sorensen [22].

THEOREM 4.6. *Given assumptions* A1–A3, *assume that a sequence* $\{x_k\}_{k=0}^{\infty}$ *is generated as outlined in section* 3. *Then, any limit point* x^* *satisfies the second-order necessary optimality conditions*; *i.e., if the constraint matrix associated with the active constraints at* x^* *is denoted by* A_A, *there is a vector* π_A *such that*

$$\nabla f(x^*) = A_A^T \pi_A, \quad \pi_A \geq 0,$$

and it holds that

$$\lambda_{\min}(Z_A^T \nabla^2 f(x^*) Z_A) \geq 0,$$

where Z_A *denotes a matrix whose columns form an orthonormal basis for the null space of* A_A.

If, in addition, $\lambda_{\min}(Z_A^T \nabla^2 f(x^*) Z_A) > 0$ *and* $\pi_A > 0$ *hold, then* $\lim_{k \to \infty} x_k = x^*$. *Further, for* k *sufficiently large, it follows that if* $s_k = -Z_A(Z_A^T H_k Z_A)^{-1} Z_A^T g_k$ *then* s_k *is sufficient in the sense of* (3.1), $p_k = s_k$, *and* $\alpha_k = 1$ *satisfies* (3.5) *and* (3.6). *Moreover, for this choice of* s_k *and* α_k, *the rate of convergence is at least q-quadratic, provided the second-derivative matrix is Lipschitz continuous in a neighborhood of* x^*.

Proof. Let x^* denote a limit point of a generated sequence of iterates. By assumption A2, there is a subsequence I such that $\lim_{k \in I} x_k = x^*$. We claim that this implies the existence of a subsequence I' such that $\lim_{k \in I'} x_k = x^*$, $q_{k-1} = 0$ and $A_{k-1} = A_k = \hat{A}$ for each $k \in I'$, where $\hat{A}$ denotes a matrix which is identical for each $k \in I'$. For $k \in I$, an iterate l_k is defined as follows. If $q_k \neq 0$, let l_k be the iteration with largest index that does not exceed k for which $q_{l_k-1} = 0$. Since no constraints are deleted immediately upon adding constraints, we obtain $q_{l_k-1} = 0$, $q_{l_k} \neq 0$, $\mathcal{W}_{l_k-1} = \mathcal{W}_{l_k}$, and $k - m \leq l_k \leq k$. If $q_k = 0$, let l_k denote the following iteration with least index such that $\mathcal{W}_{l_k} = \mathcal{W}_{l_k-1}$. If $q_{l_k-1} \neq 0$, the properties of q_{l_k-1} and the rules for updating the working set give $\mathcal{W}_{l_k} \neq \mathcal{W}_{l_k-1}$. Hence, for this case, we must have $q_{l_k-1} = 0$. Since no constraints are deleted immediately upon adding constraints, it follows that l_k is the following iteration with least index when no constraint is added. For this case, we obtain $q_{l_k-1} = 0$, $\mathcal{W}_{l_k-1} = \mathcal{W}_{l_k}$, and $k+1 \leq l_k \leq k+m$. It follows from (iii) of Lemma 4.1 that $\lim_{k\in I} \|x_k - x_{l_k}\| \to 0$, and hence $\lim_{k\in I} x_{l_k} = x^*$. With $\{l_k\}_{k\in I}$ defined this way, since there is only a finite number of different active-set matrices, the required subsequence I' can be obtained as a subsequence of $\{l_k\}_{k\in I}$.

Since, for each $k \in I'$, an unrestricted step is taken at iteration $k-1$, assumptions A1 and A2 in conjunction with property (iii) of Lemma 4.3 give

$$\hat{Z}^T \nabla f(x^*) = 0 \quad \text{and} \quad \lambda_{\min}(\hat{Z}^T \nabla^2 f(x^*) \hat{Z}) \geq 0, \tag{4.9}$$

where $\hat{Z}$ denotes an orthonormal matrix whose columns form a basis for the null space of $\hat{A}$. Since $\lim_{k\in I'} \hat{Z}^T g_k = 0$ and $\hat{A}$ has full row rank, it follows from (3.3) and (4.9) that

$$\nabla f(x^*) = \hat{A}^T \hat{\pi} \quad \text{for} \quad \hat{\pi} = \lim_{k\in I'} \pi_k. \tag{4.10}$$

It remains to show that $\min_i \hat{\pi}_i \geq 0$. Assume that there is a subsequence $I'' \subseteq I'$ and an $\epsilon > 0$ such that $\pi_{\min,k} < -\epsilon$ for $k \in I''$. Lemma 4.5 shows that there exists a K such that $q_k = 0$ for $k \in I''$ and $k \geq K$. But this contradicts (3.4a), and since $\hat{\pi} = \lim_{k\in I'} \pi_k$, we conclude that

$$\min_i \hat{\pi}_i \geq 0. \tag{4.11}$$

A combination of (4.9), (4.10), and (4.11) now ensures that x^* satisfies the second-order necessary optimality conditions. If there are constraints in A_A that are not in $\hat{A}$, the associated Lagrange multipliers are zero, i.e., π_A equals $\hat{\pi}$ possibly extended by zeros. Also, in this situation, the range space of Z_A is contained in the range space of $\hat{Z}$. Hence, $\lambda_{\min}(\hat{Z}^T \nabla^2 f(x^*) \hat{Z}) \geq 0$ implies $\lambda_{\min}(Z_A^T \nabla^2 f(x^*) Z_A) \geq 0$.

To show the second half of the theorem, note that if $\pi_A > 0$, then we must have $\hat{\pi} = \pi_A$, and it follows from (4.10) that there cannot exist a subsequence $\tilde{I}' \subseteq I'$ such that $\pi_{\min,k} < 0$ for $k \in \tilde{I}'$. This implies that there is an iteration $\tilde{K}$ such that $A_k = \hat{A}$ and $q_k = 0$ for $k \geq \tilde{K}$. Then the problem may be written as an equality-constrained problem in the null space of $\hat{A}$, namely

$$\begin{array}{ll} \underset{x\in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & \hat{A}x = \hat{b}, \end{array} \tag{4.12}$$

where $\hat{b}$ denotes the corresponding subvector of b. If $\hat{Z}^T \nabla^2 f(x^*) \hat{Z}$ is now positive definite, then (iii) of Lemma 4.1 and (3.5) ensure that the limit point is unique, i.e., $\lim_{k\to\infty} x_k = x^*$. From the continuity of f, it follows that $\hat{Z}^T H_k \hat{Z}$ is positive definite for k sufficiently large. Hence, it must hold that $d_k = 0$ and $p_k = s_k$ for k sufficiently large. If $s_k = -Z_A(Z_A^T H_k Z_A)^{-1} Z_A^T g_k$, then s_k is sufficient in the sense of

(3.1) provided that $k \geq K$ and x_k is sufficiently close to x^*. Also, this choice of s_k is the Newton step for solving (4.12), and it follows from Moré and Sorensen [22, p. 53] that $\alpha_k = 1$ eventually satisfies (3.5) and (3.6). Moreover, Moré and Sorensen [22, Theorem 2.3] show that under these assumptions $\lim_{k\to\infty} x_k = x^*$ and the rate of convergence is q-quadratic provided the second-derivative matrix is Lipschitz continuous in a neighborhood of x^* [22, Theorem 2.8]. $\square$

5. Computation of the search direction for large-scale problems.

We now show how to compute s_k, d_k, π_k, and q_k that satisfy the properties of sections 3.1, 3.2, 3.3, and 3.4, respectively. A way of updating A_k to satisfy the properties of section 3.6 is also given. Our particular interest is large-scale problems for which no prior assumptions are made about the number of constraints in the problem or the number of constraints active at the solution. This precludes the use of the reduced Hessian.

Forsgren and Murray [11] describe how suitable search directions can be computed for large-scale linear *equality-constrained* problems without the need to form the reduced Hessian. The technique they describe can be utilized also in the current context for computing a suitable descent direction s_k, a suitable Lagrange multiplier vector π_k, and a suitable direction of negative curvature d_k. We briefly review this approach here. The key procedure is an indefinite symmetric factorization of the Karush–Kuhn–Tucker (KKT) matrix K_k, where

$$K_k = \begin{pmatrix} H_k & A_k^T \\ A_k & 0 \end{pmatrix}. \tag{5.1}$$

The factorization is an LBL^T *factorization*, i.e., a factorization of the form

$$\Pi_k^T K_k \Pi_k = L_k B_k L_k^T,$$

where Π_k is a permutation matrix, L_k is a unit lower-triangular matrix, and B_k is a symmetric block-diagonal matrix whose diagonal blocks are of size 1×1 or 2×2. For a general LBL^T factorization, the permutations are performed in order to obtain a matrix L_k that is sparse and well conditioned; see, e.g., Duff and Reid [8], [9]. It is shown in Forsgren and Murray [11] that by potentially requiring additional permutations, suitable s_k and d_k can be computed from one single factorization of K_k. We demonstrate below that the additional quantities π_k and q_k can also be computed from the same factors. In the discussion below, the inertia of $Z_k^T H_k Z_k$ is required. Note that this inertia can be deduced from the inertia of K_k; see Gould [19, Lemma 3.4]. First we show how to choose A_0.

5.1. Finding an A_0 with full row rank.

It is required that A_0 has full row rank. Let $\bar{A}_0$ denote the matrix composed of all the rows of A corresponding to the active set at x_0. A straightforward way to determine A_0 is to form an LU-factorization of $\bar{A}_0^T$. An alternative approach, which fits well with the discussion of section 5, is to form the symmetric factorization of K_0 described in Forsgren and Murray [11], with $A_0 = \bar{A}_0$. In forming the factorization a redundant constraint is identified if its associated pivot is zero. The factorization may then be terminated prematurely when only redundant rows are left.

5.2. Computation of s_k and π_k. The computation of s_k and π_k is identical to the computation of s_k in Forsgren and Murray [11]. We solve

$$\begin{pmatrix} \bar{H}_k & A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} s_k \\ -\pi_k \end{pmatrix} = \begin{pmatrix} -g_k \\ 0 \end{pmatrix}, \tag{5.2}$$

and $\bar{H}_k = H_k$ when $Z_k^T H_k Z_k$ is sufficiently positive definite; otherwise, $\bar{H}_k$ is a modification of H_k such that $Z_k^T \bar{H}_k Z_k$ is sufficiently positive definite and has bounded norm. It is shown in Forsgren and Murray [11] how the factors of $\bar{K}_k$ may be obtained directly from those of K_k, where $\bar{K}_k$ denotes the modified matrix of (5.2) and K_k is given by (5.1). The matrix $\bar{K}_k$ is bounded away from a singular matrix, $\bar{H}_k$ is bounded, and $Z_k^T \bar{H}_k Z_k$ is positive definite with bounded condition number and smallest eigenvalue bounded away from zero. It is straightforward to verify that s_k from (5.2) can be written as

$$s_k = -Z_k (Z_k^T \bar{H}_k Z_k)^{-1} Z_k^T g_k, \tag{5.3}$$

and it follows that s_k is sufficient in the sense of (3.1). Moreover, assumptions A1 and A2 ensure that s_k has bounded norm if evaluated in the set $\{x : Ax \geq b, f(x) \leq f(x_0)\}$.

A combination of (5.2) and (5.3) gives

$$g_k - A_k^T \pi_k = \bar{H}_k Z_k (Z_k^T \bar{H}_k Z_k)^{-1} Z_k^T g_k,$$

and it follows that π_k satisfies (3.3).

5.3. Computation of d_k. The computation of d_k is identical to the computation of d_k in Forsgren and Murray [11]. If $Z_k^T H_k Z_k$ is positive definite then $d_k = 0$; otherwise, we may define a suitable d_k as the solution of a system of the form

$$\begin{pmatrix} H_k & A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} d_k \\ -\mu_k \end{pmatrix} = \begin{pmatrix} u_k \\ 0 \end{pmatrix}$$

for some suitable vector u_k. Forsgren and Murray [11] show how to compute d_k by a single solve with the triangular factor L_k without the need to form u_k explicitly. They also show that d_k is sufficient in the sense of (3.2) and that it has bounded norm.

5.4. Computation of q_k. We may compute a suitable q_k using the matrix $\bar{K}_k$ and the vector π_k from (5.2). As was mentioned when describing the computation of s_k and π_k, the factors of $\bar{K}_k$ may be obtained directly from those of K_k. For a positive tolerance ν, $(0 < \nu \leq 1)$, we first compute a vector v_k such that $(v_k)_i = -(\pi_k)_i$ if $(\pi_k)_i \leq \nu \pi_{\min,k}$ and $(v_k)_i = 0$ if $(\pi_k)_i > \nu \pi_{\min,k}$. The direction q_k is then obtained from the system

$$\begin{pmatrix} \bar{H}_k & A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} q_k \\ -\eta_k \end{pmatrix} = \begin{pmatrix} 0 \\ v_k \end{pmatrix}. \tag{5.4}$$

The following lemma shows that a nonzero q_k is a descent direction such that $A_k q_k \geq 0$.

LEMMA 5.1. *Let s_k and π_k be defined from* (5.2). *If $\pi_{\min,k} < 0$ and q_k and η_k are defined from* (5.4), *then $q_k^T g_k = \pi_k^T v_k \leq -\pi_{\min,k}^2$ and $A_k q_k = v_k \geq 0$.*

Proof. Premultiplication of both sides of (5.4) by the vector $(s_k^T \ -\pi_k^T)$ from (5.2) yields

$$\begin{pmatrix} s_k^T & -\pi_k^T \end{pmatrix} \begin{pmatrix} \bar{H}_k & A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} q_k \\ -\eta_k \end{pmatrix} = \begin{pmatrix} s_k^T & -\pi_k^T \end{pmatrix} \begin{pmatrix} 0 \\ v_k \end{pmatrix}. \tag{5.5}$$

Utilization of (5.2) and the symmetry of $\bar{H}_k$ in the left-hand side of (5.5) yields

$$\begin{pmatrix} -g_k^T & 0 \end{pmatrix} \begin{pmatrix} q_k \\ -\eta_k \end{pmatrix} = \begin{pmatrix} s_k^T & -\pi_k^T \end{pmatrix} \begin{pmatrix} 0 \\ v_k \end{pmatrix}. \tag{5.6}$$

Simplification of (5.6) gives $q_k^T g_k = \pi_k^T v_k$. The definition of v_k yields

$$\pi_k^T v_k = - \sum_{i:(\pi_k)_i \le \nu \pi_{\min,k}} (\pi_k)_i^2 \le -\pi_{\min,k}^2.$$

Moreover, it follows from the definition of v_k that $v_k \ge 0$, and (5.4) implies $A_k q_k = v_k \ge 0$, as required. □

The norm of π_k is bounded because of the properties of $\bar{K}_k$ and assumptions A1 and A2. Hence, since $Z_k^T \bar{H}_k Z_k$ is positive definite and has bounded norm, we conclude that q_k computed from (5.4) has bounded norm. It follows from (5.4) that $a_i^T q_k = 0$ if $(\pi_k)_i > \nu \pi_{\min,k}$ for $i \in \mathcal{W}_k$, and hence (3.4b) holds. Lemma 5.1 implies that

$$\lim_{k \in I} g_k^T q_k = 0 \quad \Rightarrow \quad \liminf_{k \in I} \pi_{\min,k} \ge 0, \tag{5.7}$$

where I is any subsequence such that q_k is computed from (5.4) for $k \in I$ and hence (3.4a) holds.

5.5. Combination of the search direction. It is not specified in sections 3.1–3.4 exactly how to choose s_k, d_k, π_k, and q_k. Sections 5.2–5.4 give suitable ways of computing these quantities. In certain situations these components are necessarily zero; if $Z_k^T g_k = 0$ then $s_k = 0$, if $Z_k^T H_k Z_k$ is positive semidefinite then $d_k = 0$, and if $\pi_{\min,k} \ge 0$ or $\mathcal{W}_k \not\subseteq \mathcal{W}_{k-1}$ then $q_k = 0$. However, it may be desirable occasionally to let some components be zero even when it is not necessary. For example, having a nonzero q_k whenever possible may not be the most efficient strategy. If the current reduced Hessian has many negative eigenvalues this suggests more constraints should be active rather than less. It is possible to impose a rule that only considers deleting constraints when to do so significantly impacts p_k. The property (3.4a) required of q_k suggests having an additional condition saying that $q_k = 0$ if

$$\pi_{\min,k} \ge \beta \left(g_k^T s_k + d_k^T H_k d_k \right),$$

where β is a positive constant. Since Lemma 4.3 implies that $\lim_{k\to\infty} g_k^T s_k = 0$ and $\lim_{k\to\infty} d_k^T H_k d_k = 0$ for unrestricted steps, such a condition does not impact on (3.4a), and hence it does not alter the convergence analysis. Similar conditions can be imposed to set $s_k = 0$ or $d_k = 0$ at certain iterations.

5.6. The update of A_k. The working-set matrix A_k is required to have full row rank. A straightforward way to ensure this property is to add at most one constraint at every iteration, as the following lemma shows.

LEMMA 5.2. *Given assumptions* A1–A3*, assume that a sequence* $\{x_k\}_{k=0}^{\infty}$ *is generated as outlined in section* 3*. If* A_0 *has full row rank,* $|\mathcal{W}_{k+1}| \le |\mathcal{W}_k^0| + 1$*,* $a_i^T p_k < 0$ *for all* $k \ge 0$*, and* $i \in \mathcal{W}_{k+1} \backslash \mathcal{W}_k$*, then each* A_k *has full row rank.*

Proof. See, e.g., Gill et al. [16, Lemma 2.1]. ∎

Although the computed search directions described in sections 5.2–5.4 are not designed specifically to add more than one constraint per iteration, the convergence analysis presented gives room for defining algorithms that add any number of active constraints, as long as the working-set matrix has full row rank. The issue would be twofold: (i) to modify the definitions of the search directions, so as to make more than one new constraint become active in the linesearch, while still maintaining the required properties of these directions, and (ii) to maintain the full rank of the working-set matrix. This approach may be advantageous for certain problems, e.g., problems where all constraints are simple bounds. In this situation, it is known a priori that any working-set matrix will have full row rank. Techniques similar to gradient projection, see, e.g., Calamai and Moré [7], might prove useful for altering the search direction.

6. Primal degeneracy. Assumptions A1 and A2 ensure that the objective function is sufficiently smooth and the iterates remain in a feasible region. Assumption A3 implies that no primal degenerate second-order constrained stationary points exist. Although for nonlinear problems degeneracy is not as common in practice as it is for linear programming problems, there are problems for which A3 does not hold. Consequently, in a practical implementation of our algorithm some technique to handle degeneracy is necessary. The nature of degeneracy is different for nonlinear problems. In linear programming the main concern is degenerate vertices. In effect the iterate is at the degenerate stationary point. In a nonlinear problem we may never be at the stationary point. Moreover it is likely not to be a vertex. What we are likely to encounter is rank deficient active-set matrices for which the number of rows is less than n, and we are not at a constrained stationary point. We need only be concerned if we plan to delete constraints. In exact arithmetic we could define a subiteration to search for a suitable active set. A method of implementing this strategy that makes use of the known factorization of the KKT matrix is described in Gill et al. [15]. Such an approach is an improvement over algorithms based on sequential quadratic programming where a subiteration may be necessary at each iteration of the quadratic programming subproblem. The difficulty with this strategy is the need to define the active set. In inexact arithmetic precisely what is the active set is not clear. We prefer therefore to rely on the approach adopted by Gill et al. [14]. This technique allows infeasibility tolerances on the constraints that are altered at each iteration. The impact on the algorithm is that a zero step is never taken. The consequences of allowing infeasibility tolerances is that the solution obtained may be infeasible. However, the maximum degree of infeasibility may be specified. In practice the maximum infeasibility allowed when solving nonlinear problems is unlikely to be attained and is in any event consistent with the infeasibility that results from the impact of finite precision operations. An advantage of this approach is that it is equally useful for handling near degeneracy. This is likely to be common on problems where the linearly constrained problem being solved is an approximation to a nonlinearly constrained problem whose Jacobian is rank deficient at the solution. The use of a procedure similar to that in [14] is in any event essential in practice for the purpose of trying to introduce a choice in the definition of A_k in an attempt to ensure that the condition number of A_k is not too large. For example, if infeasibilities are allowed then nearly dependent active constraints need not be included in the working set. The search direction will not be exactly orthogonal to the constraint normals of the constraints ignored but it will be close, hence the next iterate will not be too infeasible.

7. Discussion. A convergence analysis for an algorithm to solve linear inequality-constrained optimization problems has been presented. The algorithm is described in broad terms by assuming the availability at each iteration of three directions with certain properties. It has also been shown how to compute all the required search directions from a single symmetric indefinite factorization of the KKT matrix. Such an algorithm is well suited to solving large-scale problems. Unlike some alternatives the efficiency of the method is not dependent on either the active set or the null space of the active set being small.

For convenience of notation, the problem is stated in all-inequality form (1.1), but we emphasize that the analysis can be modified in a straightforward manner to cover the case with a mixture of inequality and equality constraints. A particularly attractive feature of the algorithm described is that the problem does not have to be transformed into a specific form.

Acknowledgments. We thank the referees for their helpful comments and criticism, which significantly improved the presentation.

REFERENCES

[1] A. Buckley, *An alternative implementation of Goldfarb's minimization algorithm*, Math. Prog., 8 (1975), pp. 207–231.
[2] J. Burke, *On the identification of active constraints* II: *The nonconvex case*, SIAM J. Numer. Anal., 27 (1990), pp. 1081–1102.
[3] J. V. Burke and J. J. Moré, *On the identification of active constraints*, SIAM J. Numer. Anal., 25 (1988), pp. 1197–1211.
[4] J. V. Burke and J. J. Moré, *Exposing constraints*, SIAM J. Optim., 4 (1994), pp. 573–595.
[5] J. V. Burke, J. J. Moré, and G. Toraldo, *Convergence properties of trust region methods for linear and convex constraints*, Math. Prog., 47 (1990), pp. 305–336.
[6] R. H. Byrd and G. A. Shultz, *A Practical Class of Globally Convergent Active Set Strategies for Linearly Constrained Optimization*, Tech. rep. CU-CS-238-82, Department of Computer Science, University of Colorado at Boulder, Boulder, CO, 1982.
[7] P. H. Calamai and J. J. Moré, *Projected gradient methods for linearly constrained problems*, Math. Prog., 39 (1987), pp. 93–116.
[8] I. S. Duff and J. K. Reid, *MA*27*: A Set of Fortran Subroutines for Solving Sparse Symmetric Sets of Linear Equations*, Tech. rep. R-10533, Computer Science and Systems Division, AERE Harwell, Oxford, England, 1982.
[9] I. S. Duff and J. K. Reid, *The multifrontal solution of indefinite sparse symmetric linear equations*, ACM Trans. Math. Software, 9 (1983), pp. 302–325.
[10] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., John Wiley Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1987.
[11] A. Forsgren and W. Murray, *Newton methods for large-scale linear equality-constrained minimization*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 560–587.
[12] D. M. Gay, *A trust-region approach to linearly constrained optimization*, in Numerical Analysis, D. F. Griffiths, ed., Lecture Notes in Mathematics 1066, Springer-Verlag, Berlin, Heidelberg, New York, 1984, pp. 72–105.
[13] P. E. Gill and W. Murray, *Newton-type methods for unconstrained and linearly constrained optimization*, Math. Prog., 7 (1974), pp. 311–350.
[14] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *A practical anti-cycling procedure for linearly constrained optimization*, Math. Prog., 45 (1989), pp. 437–474.
[15] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *A Schur-complement method for sparse quadratic programming*, in Reliable Numerical Computation, M. G. Cox and S. J. Hammarling, eds., Oxford University Press, London, England, 1990, pp. 113–138.
[16] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *Inertia-controlling methods for general quadratic programming*, SIAM Rev., 33 (1991), pp. 1–36.

[17] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, New York, 1981.
[18] D. Goldfarb, *Extensions of Davidon's variable metric method to maximization under linear inequality and equality constraints*, SIAM J. Appl. Math., 17 (1969), pp. 739–764.
[19] N. I. M. Gould, *On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem*, Math. Prog., 32 (1985), pp. 90–99.
[20] G. P. McCormick, *A second order method for the linearly constrained nonlinear programming problem*, in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian, and K. Ritter, eds., Academic Press, New York, London, 1970, pp. 207–243.
[21] J. J. Moré and D. C. Sorensen, *On the use of directions of negative curvature in a modified Newton method*, Math. Prog., 16 (1979), pp. 1–20.
[22] J. J. Moré and D. C. Sorensen, *Newton's method*, in Studies in Numerical Analysis, Studies in Mathematics, Vol. 24, G. H. Golub, ed., The Mathematical Association of America, 1984, pp. 29–82.
[23] B. A. Murtagh and M. A. Saunders, *Large-scale linearly constrained optimization*, Math. Prog., 14 (1978), pp. 41–72.
[24] B. A. Murtagh and M. A. Saunders, *MINOS* 5.4 *User's Guide*, Tech. rep. SOL 83-20R, Department of Operations Research, Stanford University, Stanford, CA, 1993.
[25] K. Ritter, *A superlinearly convergent method for minimization problems with linear inequality constraints*, Math. Prog., 4 (1973), pp. 44–71.
[26] K. Ritter, *A method of conjugate direction for linearly constrained nonlinear programming problems*, SIAM J. Numer. Anal., 12 (1975), pp. 274–303.
[27] K. Ritter, *Convergence and superlinear convergence of algorithms for linearly constrained minimization problems*, in Nonlinear Optimization: Theory and Algorithms, Part II, L. C. W. Dixon, E. Spedicato, and G. P. Szegö, eds., Birkhäuser, Boston, MA, 1980, pp. 221–251.
[28] J. B. Rosen, *The gradient projection method for nonlinear programming. Part* I. *Linear constraints*, SIAM J. Appl. Math., 8 (1960), pp. 181–217.