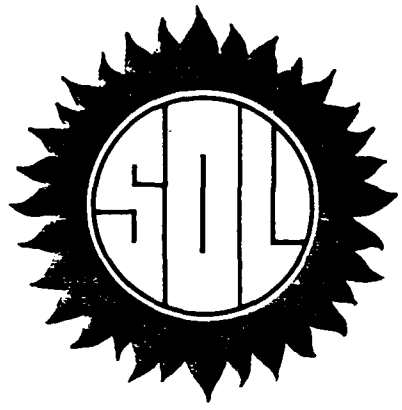

Systems
Optimization
Laboratory

Newton Methods for Large-Scale Linear Equality-Constrained Minimization

by

Anders L. Forsgren* and Walter Murray†

TECHNICAL REPORT SOL 90-6

April 1990

Department of Operations Research
Stanford University
Stanford, CA 94305

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4022

Newton Methods for Large-Scale Linear Equality-Constrained Minimization

by
Anders L. Forsgren* and Walter Murray†

TECHNICAL REPORT SOL 90-6
April 1990

* Research supported by the Göran Gustafsson Foundation.

† Research and reproduction of this report were partially supported by the National Science Foundation Grant DDM-8715153 and the Office of Naval Research Contract N00014-90-J-1242.

NEWTON METHODS FOR LARGE-SCALE LINEAR EQUALITY-CONSTRAINED MINIMIZATION

Anders L. FORSGREN* and Walter MURRAY†

*Optimization and Systems Theory, Department of Mathematics
The Royal Institute of Technology, S-100 44 Stockholm, Sweden

†Systems Optimization Laboratory, Department of Operations Research
Stanford University, Stanford, California 94305-4022, USA

Technical Report SOL 90-6
April 1990

Abstract

Newton methods for large-scale minimization subject to linear equality constraints are discussed. For large-scale problems, it may be prohibitively expensive to reduce the problem to an unconstrained problem in the null space of the constraint matrix. We investigate computational schemes that enable the computation of descent directions and directions of negative curvature without the need to know the null-space matrix. The schemes are based on factorizing a sparse symmetric indefinite matrix. Three different methods are proposed for computing the desired directions. Convergence properties for the different methods are established.

Keywords: Linear equality-constrained minimization, negative curvature, modified Newton method, symmetric indefinite factorization, large-scale minimization, linesearch method.

1. Introduction

We consider methods for finding a local minimizer of the problem

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

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. It is assumed that A is a sparse matrix of full row rank. We also assume that an initial feasible

*Research supported by the Swedish National Board for Technical Development and the Göran Gustafsson Foundation.

†Research supported by the National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

point x_0 is known and that the level set $S(x_0) = \{x : f(x) \leq f(x_0),\ Ax = b\}$ is compact.

If (1.1) is solved by Newton's method, the step p from the current iterate x may be defined as $p = Zp_Z$, where

$$Z^T H Z p_Z = -Z^T g. \tag{1.2}$$

The columns of the matrix Z form a basis for the null space of A, H denotes the Hessian matrix $\nabla^2 f(x)$ and g denotes the gradient $\nabla f(x)$. We shall refer to the matrix $Z^T H Z$ as the *reduced Hessian* and the vector $Z^T g$ as the *reduced gradient.* A mathematically equivalent way to obtain p is to solve the equation

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\mu \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix}. \tag{1.3}$$

We shall refer to the matrix on the left-hand side of (1.3) as the *KKT-matrix* and denote it by K.

Given p, the next iterate is obtained as $x + p$. If Newton's method converges to a solution of (1.1) and the reduced Hessian is positive definite and Lipschitz continuous, convergence occurs at a quadratic rate. However, Newton's method may not converge from every starting point, and in the neighborhood of a saddle-point or a local maximizer, a sequence of iterates generated by Newton's method may converge to such a point. Consequently, if a method that generates a sequence of improving estimates is required, some modification is needed. If $Z^T H Z$ is positive definite, p is a *feasible descent direction*, i.e., a direction p such that $p^T g < 0$ and $Ap = 0$. In this case, the objective function value at $x + p$ may not be reduced but a step along p may be found that yields the next improved iterate. If $Z^T H Z$ is not positive definite, the Newton direction may not be a direction along which the objective function decreases.

Modifications of Newton's method suitable for iterates where the Newton step does not yield a *sufficient* decrease of the objective function exist. The methods can be divided into two types, *trust-region methods* and *linesearch methods.* See Moré and Sorensen [MS84] and Dennis and Schnabel [DS89] for a discussions of these different types of method. We focus on linesearch methods in this paper and in order to simplify the notation, we shall refer to modifications of Newton's method of linesearch type as *modified Newton methods.* Such methods involve the computation of a feasible descent direction s and, if the reduced Hessian has at least one negative eigenvalue, a *feasible direction of negative curvature* d. A direction d is said to be a feasible direction of negative curvature if $d^T H d < 0$ and $Ad = 0$. If s is a direction of *sufficient* descent and d is a direction of *sufficient* negative curvature, the convergence of a modified Newton method for linear equality-constrained problems follows from known results on modified Newton methods, see for example Fiacco and McCormick [FM68], Gill and Murray [GM74], McCormick [McC77], Fletcher and Freeman [FF77], Mukai and Polak [MP78], Kaniel and Dax [KD79], Moré and Sorensen [MS79], Goldfarb [Gol80] and Forsgren *et al.* [FGM89b]. We require the computed directions s and d to be feasible, i.e., $As = 0$ and $Ad = 0$. If the

reduced Hessian is known and has at least one negative eigenvalue, a direction v such that $v^T Z^T H Z v < 0$ may be obtained using techniques for unconstrained problems. Consequently, $d = Zv$ is a feasible direction of negative curvature. Similarly, a positive-definite modification of $Z^T H Z$ enables the computation of a feasible descent direction. Feasibility is therefore not an issue if the reduced Hessian is known, whereas it is not immediately apparent how to satisfy $As = 0$ and $Ad = 0$ when utilizing K.

Although equations (1.2) and (1.3) are mathematically equivalent, they differ in the amount of work required to obtain the search direction p. To obtain p from (1.2), the reduced Hessian is required. A matrix Z may be obtained by forming the LU-factorization of A, see Gill *et al.* [GMSW87a]. Even if Z is sparse, the matrix $Z^T H Z$ may be completely dense, and the amount of work required to form $Z^T H Z$ explicitly may be prohibitive, see Gill *et al.* [GMSW85]. To obtain p from (1.3), the KKT-matrix is required. If H and A are sparse, they yield a sparse K. Consequently, if equations involving K are solved, it is possible to take advantage of the sparsity of the problem.

The goal of the methods described in this paper is to compute search directions directly from equations involving the KKT-matrix K or a modified KKT-matrix. We prefer to use the identical method to compute the Newton direction from (1.3) that we would use if it was known in advance that $Z^T H Z$ was positive definite. Also, a method for which there exists an efficient implementation for a large sparse matrix K is desired. Accordingly, we consider the LBL^T-factorization, described in Section 3. Such a factorization computes $\Pi^T K \Pi = LBL^T$, where Π is a permutation matrix, L is a unit lower-triangular matrix and B is a symmetric block-diagonal matrix whose diagonal blocks are of size 1×1 or 2×2. The permutations are performed in order to obtain a matrix L that is sparse and well-conditioned. For unconstrained problems, the matrices K and H are identical. In this case, Moré and Sorensen [MS79] have shown how to compute a descent direction and a direction of negative curvature, whenever they exist, from L and B. We describe a pivoting strategy in the LBL^T-factorization of K, so that the ability to compute a feasible descent direction and a feasible direction of negative curvature is achieved by a single factorization of K, analogous to the method of Moré and Sorensen. The computed directions are shown to satisfy conditions needed to apply known convergence results for modified Newton methods, which state that limit points of a generated sequence of iterates satisfy the second-order necessary optimality conditions.

The abovementioned pivoting strategy sufficient for computing both a descent direction and a direction of negative curvature is then shown to be necessary in order to guarantee the ability to obtain the directions from a single factorization of K. Two other methods for computing search directions are described that do not require this pivoting strategy. One method always generates a descent direction, whereas the other method generates a search direction that is either a descent direction and/or a direction of negative curvature. Applying known convergence results for modified Newton methods, limit points of a generated sequence of iterates using these methods satisfy the first-order necessary optimality conditions. The main purpose of introducing two additional methods is that the computation of the search

direction is likely to be cheaper than in the method motivated by convergence to a "second-order" point. The intent is that the three methods may be mixed to obtain a more efficient method.

2. Basics

2.1. Terminology

Throughout the paper, the Hessian $\nabla^2 f(x)$ and gradient $\nabla f(x)$ are evaluated at points $x \in S(x_0)$. Given an arbitrary point in $S(x_0)$, H denotes the Hessian and g the gradient at such a point. The set of KKT-matrices for which H is evaluated at some point in $S(x_0)$ is denoted by $\mathcal{K}$. All KKT-matrices considered belong to $\mathcal{K}$. For an iterative sequence $\{x_k\}_{k=0}^{\infty}$ a subscript k is included, so that $H_k = \nabla^2 f(x_k)$ and $g_k = \nabla f(x_k)$. Also, for vectors and matrices, the norm used is always the Euclidean vector norm and the corresponding induced matrix norm. The vector e_i denotes the i-th unit vector of the appropriate dimension. In a number of lemmas matrices of zero dimension may arise. In such circumstances, there is no loss of generality if we make the assumption that matrices of zero dimension have unit eigenvalues and norm zero.

2.2. The inertia of a matrix

Let M be any symmetric matrix. We denote by $i_p(M)$, $i_n(M)$ and $i_z(M)$ respectively the (nonnegative) numbers of positive, negative and zero eigenvalues of M. The *inertia* of M—denoted by $In(M)$—is the associated integer triple (i_p, i_n, i_z). The following lemma states an important relationship between the inertia of the KKT-matrix K and the reduced Hessian $Z^T H Z$.

Lemma 2.1. *If* $\text{rank}(A) = r$, *and* Z *is a matrix whose columns form a basis for the null space of* A, *then* $In(K) = In(Z^T H Z) + (r, r, m - r)$.

Proof. See Gould [Gou85, Lemma 3.4] ∎

This lemma implies that K is singular only if $Z^T H Z$ is singular or A is rank deficient. In this paper, we assume that A has full row rank, so that singularity of K always means singularity of $Z^T H Z$. The following lemma shows that if A has full row rank, there is a uniform relationship between the nonsingularity of K and $Z^T H Z$.

Lemma 2.2. *Assume that* $\text{rank}(A) = m$, *and let* Z *denote a matrix whose columns form a basis for the null space of* A. *For a given positive constant* c_1, *there exists a positive constant* c_2 *such that for any KKT-matrix* $K \in \mathcal{K}$ *having its smallest singular value greater than* c_1, *the smallest singular value of* $Z^T H Z$ *is greater than* c_2.

Proof. If A has full row rank, Lemma 2.1 implies that K is nonsingular if and only if $Z^T H Z$ is nonsingular. For a symmetric matrix, the singular values are the absolute

values of the eigenvalues. The norm of K is bounded, since $S(x_0)$ is compact. The result follows by observing that the eigenvalues of K and Z^THZ vary continuously with H. ∎

3. The LBL^T-factorization

3.1. The factorization

An efficient method for solving equations involving symmetric indefinite matrices is to use the factorization

$$\Pi^T K \Pi = LBL^T, \tag{3.1}$$

where Π is a permutation matrix that represents column interchanges in K, L is a unit lower-triangular matrix, B is a symmetric block-diagonal matrix whose blocks are of size 1×1 or 2×2. The 2×2 blocks are always nonsingular with one positive and one negative eigenvalue. We shall refer to this factorization as the LBL^T-factorization. Various algorithms for computing the factors have been proposed, see Bunch and Parlett [BP71] and Bunch and Kaufman [BK77].

Forming the LBL^T-factorization may be viewed as a step-wise procedure that repeatedly computes *Schur complements* of decreasing dimension. If K is partitioned as

$$K = \begin{pmatrix} T & N^T \\ N & G \end{pmatrix}, \tag{3.2}$$

and T is nonsingular, the *Schur complement of* T *in* K will be denoted by K/T, and is defined as

$$K/T = G - NT^{-1}N^T.$$

The matrix K/T will be referred to as "the" Schur complement when the matrix T is clear from the context. By convention, we define the Schur complement as K when T has dimension zero. Given the partition from (3.2), the identity

$$\begin{pmatrix} T & N^T \\ N & G \end{pmatrix} = \begin{pmatrix} I & 0 \\ NT^{-1} & I \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & K/T \end{pmatrix} \begin{pmatrix} I & T^{-1}N^T \\ 0 & I \end{pmatrix} \tag{3.3}$$

holds. In the first step of the factorization, the matrix T is a 1×1 or 2×2 principal submatrix of K. Symmetric row and column interchanges may be necessary in order to obtain a T that is suitable as *pivot*. The elimination step yields one or two rows of L^T as $(I \quad T^{-1}N^T)$ and a diagonal block of B of size 1×1 or 2×2 from T. In the next step, the process is repeated for the Schur complement K/T. Eventually the Schur complement has dimension zero, and the algorithm terminates with permutation matrix Π, unit lower-triangular matrix L and block-diagonal matrix B. Algorithms differ in the way the pivot is chosen. In this paper we do not elaborate on which algorithm to use, except to assume the algorithm yields a bounded $\|L\|$. We shall refer to this algorithm as an algorithm that performs a *regular* LBL^T-factorization. Since L is a unit lower-triangular matrix, if the norm of L is bounded then the norm

of L^{-1} is also bounded. Consequently, it follows from (3.1) that K is arbitrarily close to a singular matrix if and only if B is arbitrarily close to a singular matrix. In our applications, the inertia of K is required. This inertia is readily available if B is known, since Sylvester's law of inertia yields $In(K) = In(B)$, see [GV89, page 416].

3.2. Computational considerations

When K is sparse, the LBL^T-factorization of K may be carried out in two steps. First, in the *analyze phase*, a symbolic factorization based on the nonzero elements in K is performed. This symbolic factorization yields an ordering of the rows and columns that attempts to minimize the number of nonzeros in L. Then, in the *numerical phase*, the factors are computed, attempting to maintain the ordering given by the analyze phase. However, for numerical reasons, it may be necessary to perform additional permutations in the numerical phase. Consequently, the permutation matrix Π in (3.1) is the combination of the ordering from the analyze phase and the additional permutations from the numerical phase. In a modified Newton method, where a sequence of K-matrices is factorized, the analyze phase need only be performed once, since the positions of the nonzero elements in the Hessian remain the same. A robust and efficient routine that performs such a two-step factorization is the Harwell routine MA27, see Duff and Reid [DR82] [DR83]. (MA27 may use 2×2 pivots that are not indefinite. To fit the discussion here, such a pivot may be viewed equivalent to two consecutive 1×1 pivots.)

3.3. Definition of pivot types

The matrix B consists of diagonal blocks of size 1×1 or 2×2, where the 2×2-blocks are indefinite. These diagonal blocks are the pivots defined by the factorization algorithm. It is of importance to distinguish between different types of block.

A 1×1-block is defined to be of type H^+ if it is positive, and if the element in the same position in $\Pi^T K \Pi$ is a diagonal element of H. Similarly, 1×1-blocks are defined to be of type H^- and H^0 if they satisfy the same position requirements as the blocks of type H^+, but are negative or zero, respectively. We denote by n_H^+, n_H^- and n_H^0 respectively, the number of such 1×1-blocks.

A 1×1-block is defined to be of type A^+ if it is positive, and if the element in the same position in $\Pi^T K \Pi$ is a diagonal element of the zero-part of K. Similarly, 1×1-blocks are defined to be of type A^- and A^0 if they satisfy the same position requirements as the blocks of type A^+, but are negative or zero, respectively. We denote by n_A^+, n_A^- and n_A^0 respectively, the number of such 1×1-blocks.

A 2×2-block is defined to be of type HH if the elements in $\Pi^T K \Pi$ with the same position are elements of H. We denote by n_{HH} the number of such 2×2-blocks.

A 2×2-block is defined to be of type AA if the elements in $\Pi^T K \Pi$ with the same position are elements of the zero-part of K. We denote by n_{AA} the number of such 2×2-blocks.

A 2×2-block is defined to be of type HA if the elements in $\Pi^T K \Pi$ with the same position consist of one diagonal element from H, one diagonal element from

the zero-part of K and two elements from A.

3.4. Applications in unconstrained minimization

For unconstrained problems, Moré and Sorensen [MS79] have shown how to compute a descent direction s and a direction of negative curvature d, whenever such directions exist, using the LBL^T-factorization of the Hessian H, given by $\Pi^T H \Pi = LBL^T$. The descent direction s satisfies the equation $\Pi L \bar{B} L^T \Pi^T s = -g$, where $\bar{B}$ is a positive-definite modification of B. If H is sufficiently positive definite, then $\bar{B} = B$ and s is the Newton search direction. If H is not positive definite, $\bar{B}$ is obtained as $\bar{B} = B + D$, where D is a block-diagonal matrix with the same block structure as B, and D has rank equal to the number of nonpositive eigenvalues of H. If H has at least one negative eigenvalue, the direction of negative curvature, d, satisfies the equation $L^T \Pi^T d = \pm\sqrt{-\lambda_{\min}(B)}v$, where v is an eigenvector of unit norm corresponding to $\lambda_{\min}(B)$, the smallest eigenvalue of B. The sign is chosen so that $g^T d \leq 0$. For details, see Moré and Sorensen [MS79].

Similarly, in the linear equality-constrained case, a feasible descent direction, (assuming such a direction exists), may be obtained by solving

$$\begin{pmatrix} \bar{H} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} s \\ -\mu \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix}, \tag{3.4}$$

where $\bar{H}$ is a modification of H such that $Z^T \bar{H} Z$ is positive definite. Utilizing the idea of More and Sorensen [MS79] from the unconstrained case, we may try to use the LBL^T-factorization of K to modify blocks of B of type H^-, H^0 and HH, yielding a matrix $\bar{B}$ such that $L\bar{B}L^T$ has n positive eigenvalues and m negative eigenvalues. The following lemma shows that there is a sufficient number of blocks of type H^-, H^0 and HH to create such a matrix $\bar{B}$.

Lemma 3.1. *If* $In(K) = (i_p, i_n, i_z)$, *where* $i_n + i_z - m > 0$, *then* $n_H^- + n_H^0 + n_{HH} \geq i_n + i_z - m$.

Proof. Assume the contrary, that $n_H^- + n_H^0 + n_{HH} < i_n + i_z - m$. The dimensions of H and A imply

$$n_H^+ + n_H^- + n_H^0 + 2n_{HH} + n_{HA} = n \tag{3.5a}$$
$$n_A^+ + n_A^- + n_A^0 + n_{HA} + 2n_{AA} = m. \tag{3.5b}$$

If $In(K) = (i_p, i_n, i_z)$, we get

$$n_H^+ + n_{HH} + n_{HA} + n_A^+ + n_{AA} = i_p \tag{3.6a}$$
$$n_H^- + n_{HH} + n_{HA} + n_A^- + n_{AA} = i_n \tag{3.6b}$$
$$n_H^0 + n_A^0 = i_z. \tag{3.6c}$$

Adding (3.6b) and (3.6c) we obtain $n_{HA} + n_A^- + n_{AA} + n_A^0 > m$, contradicting (3.5b). Thus $n_H^- + n_H^0 + n_{HH} < i_n + i_z - m$ cannot hold. ∎

However, although there are sufficiently many block elements of type H^-, H^0 and HH, if no additional conditions are imposed on Π, the ordering may be insufficient to guarantee that only the H-part of K is altered when a block element of type H^-, H^0 or HH is modified. Consequently, if an equation involving $L\bar{B}L^T$ is solved, there is no straightforward analogous way of obtaining a descent direction s such that $As = 0$. As an example, consider

$$H = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 2 & 1 \end{pmatrix}, \quad \text{so that} \quad K = \begin{pmatrix} 2 & 0 & 2 \\ 0 & -2 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

If Π is the identity matrix, we obtain

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -\frac{1}{2} & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -\frac{3}{2} \end{pmatrix}.$$

The matrix B has $n_H^+ = n_H^- = n_A^- = 1$. Analogous to the unconstrained case, we make the second diagonal element of B positive by adding a positive number d_{22}. However, the matrix $L\bar{B}L^T$ will differ from LBL^T in the A-part and the zero-part, since

$$L(B+D)L^T = \begin{pmatrix} 2 & 0 & 2 \\ 0 & -2 & 1 \\ 2 & 1 & 0 \end{pmatrix} + d_{22} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{4} \end{pmatrix}.$$

Because of the lower triangular structure of L, if a diagonal block of B is altered, the modification may alter elements of the permuted K with greater row and column numbers. Consequently, if there are rows of A with greater row numbers, there is a danger of altering parts of K other than the H-part.

It may appear that, by analogy with the unconstrained case, a feasible direction of negative curvature may be obtained by solving the equation $L^T\Pi^T u = v$, where v is an eigenvector of B corresponding to a negative eigenvalue of a block of type H^- or HH. (Note that such a vector v always exists if the reduced Hessian has at least one negative eigenvalue, since Lemma 2.1 implies that K has more than m negative eigenvalues. Consequently, since there are at most m block elements of type A^-, HA and AA, there must exist a block element of type H^- or HH for K to have more that m negative eigenvalues.) The vector d is then defined as the first n components of u. However, the ordering of the rows and columns of K given by Π may be insufficient to guarantee that $Ad = 0$. In our example matrix, we obtain $d = (0\ 1)^T$ for $v = (0\ \ 1\ \ 0)^T$, and consequently $Ad = 1 \neq 0$.

Conn and Gould [CG84] have proposed an algorithm for finding feasible directions of negative curvature based on the LBL^T-factorization of K. In addition to the original factorization, the algorithm requires the factorization of a "triangular-like" matrix of dimension $m \times m$. In this paper, we adopt a different view and investigate what pivoting strategy is necessary and sufficient to obtain the desired search directions in a single factorization of K, using L and B in an analogous way to the approach of Moré and Sorensen [MS79] for the unconstrained case.

4. Sufficient Pivot Conditions

If K is factorized using a regular LBL^T-factorization, the permutations are performed in order to obtain sparsity of the factors and boundedness of $\|L\|$. In the example given in Section 3.4, it was shown that these permutations may be insufficient for computing a feasible descent direction and a feasible direction of negative curvature. In this section, we investigate what conditions may be imposed on the pivots in order to ensure the ability to compute these directions, using L and B, whenever the directions exist.

Upon completion of the the factorization of K, we have computed a permutation matrix Π, a lower-triangular matrix L and a block-diagonal matrix B such that $LBL^T = \Pi^T K \Pi$. We will consider a specific step in the factorization. Therefore the permutation matrix Π is partitioned as

$$\Pi = \begin{pmatrix} \Pi_1 & \Pi_2 \end{pmatrix}, \tag{4.1}$$

where the matrix $\Pi_1^T K \Pi_1$ is the principal submatrix of K for which the factors have been computed. Consequently, $\Pi_1^T K \Pi_1 = L_{11} B_1 L_{11}^T$, where L_{11} and B_1 are the leading principal submatrices of L and B, respectively. Note that at this step of the factorization, the matrix Π_2 has not yet been determined.

The matrix $\Pi_1^T K \Pi_1$ is a principal submatrix of K, and we introduce a permutation $\bar{\Pi}$ for which

$$\bar{\Pi}^T K \bar{\Pi} = \begin{pmatrix} H_{11} & H_{12} & A_{11}^T & A_{21}^T \\ H_{21} & H_{22} & A_{12}^T & A_{22}^T \\ A_{11} & A_{12} & 0 & 0 \\ A_{21} & A_{22} & 0 & 0 \end{pmatrix}.$$

The matrix $\bar{\Pi}$ and the partition of $\bar{\Pi}^T K \bar{\Pi}$ are such that if we define K_{11} as

$$K_{11} = \begin{pmatrix} H_{11} & A_{11}^T \\ A_{11} & 0 \end{pmatrix}, \tag{4.2}$$

the matrix $\Pi_1^T K \Pi_1$ is a permuted version of K_{11}. The matrix K_{11} is a principal submatrix of K such that H_{11} is an $n_1 \times n_1$ principal submatrix of H, and A_{11} is an $n_1 \times m_1$ submatrix of A. To simplify the notation, we say that K_{11} *contains* n_1 rows of H and m_1 rows of A.

Since $\Pi_1^T K \Pi_1$ denotes the part of K that has been factorized at a particular step, its size increases from step to step. We say that $\Pi_1^T K \Pi_1$ and the associated K_{11} are *expanded* when one step in the factorization is performed. For example, if in one step of the factorization, a pivot of type HA is selected, the size of $\Pi_1^T K \Pi_1$ and K_{11} is increased by two and both n_1 and m_1 are increased by one.

It is shown in this section, that by selecting sufficiently nonsingular pivots of type H^+, A^- and HA, until the corresponding K_{11} contains all rows of A, the ability to compute a feasible descent direction and a feasible direction of negative curvature is

achieved. A matrix K_{11} created in such a way will prove crucial for the computation of descent directions and directions of negative curvature.

Properties of $\Pi_1^T K \Pi_1$ that we want to describe are also properties of K_{11}. In particular, the eigenvalues of K_{11} and $\Pi_1^T K \Pi_1$ are identical. We consider K_{11} since its properties are easier to describe. The following lemma gives an explicit representation of K_{11}^{-1}.

Lemma 4.1. *If K_{11} is nonsingular, then*

$$K_{11}^{-1} = \begin{pmatrix} H_{11} & A_{11}^T \\ A_{11} & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T & \bar{Y} \\ \bar{Y}^T & -\bar{Y}^T H_{11} \bar{Y} \end{pmatrix},$$

where $\bar{Z}$ is an orthonormal matrix whose columns form a basis for the null space of A_{11}, and the matrix $\bar{Y}$ satisfies the matrix equation

$$\begin{pmatrix} H_{11} & A_{11}^T \\ A_{11} & 0 \end{pmatrix} \begin{pmatrix} \bar{Y} \\ -\Lambda \end{pmatrix} = \begin{pmatrix} 0 \\ I \end{pmatrix}, \tag{4.3}$$

where $\Lambda = \bar{Y}^T H_{11} \bar{Y}$. If A_{11} is square, the matrix $\bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T$ is to be interpreted as a zero matrix of dimension $m_1 \times m_1$.

Proof. Direct substitution in (4.3) shows that $\Lambda = \bar{Y}^T H_{11} \bar{Y}$. Lemma 2.1 shows that nonsingularity of K_{11} yields nonsingularity of $\bar{Z}^T H_{11} \bar{Z}$. Thus, it remains to verify that

$$\begin{pmatrix} H_{11} & A_{11}^T \\ A_{11} & 0 \end{pmatrix} \begin{pmatrix} \bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T \\ \bar{Y}^T \end{pmatrix} = \begin{pmatrix} I \\ 0 \end{pmatrix}. \tag{4.4}$$

This holds if and only if $H_{11} \bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T + A_{11}^T \bar{Y}^T = I$. If A_{11} is square, then $\bar{Y} = A_{11}^{-1}$, and (4.4) holds for $\bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T = 0$. Otherwise, (4.3) gives $\bar{Z}^T H_{11} \bar{Y} = 0$ and it follows that

$$\begin{pmatrix} \bar{Z}^T \\ \bar{Y}^T \end{pmatrix} \left(H_{11} \bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T + A_{11}^T \bar{Y}^T \right) = \begin{pmatrix} \bar{Z}^T \\ \bar{Y}^T \end{pmatrix}. \tag{4.5}$$

The proof is now completed by showing that $\begin{pmatrix} \bar{Z} & \bar{Y} \end{pmatrix}$ is nonsingular. Assume that $\bar{Z} v_z + \bar{Y} v_y = 0$. Premultiplication by A_{11} yields $v_y = 0$. Since the columns of $\bar{Z}$ are linearly independent, $v_z = 0$. ∎

The following lemma shows that when $\Pi_1^T K \Pi_1$ has inertia $(n_1, m_1, 0)$, a row in $\Pi^T K \Pi / \Pi_1^T K \Pi_1$ corresponding to a row of A cannot be almost linearly dependent on the other rows of $\Pi^T K \Pi / \Pi_1^T K \Pi_1$ unless A is almost rank-deficient.

Lemma 4.2. *For given positive constants c_1 and M, let K be any KKT-matrix in $\mathcal{K}$ with a principal submatrix K_{11} such that $In(K_{11}) = (n_1, m_1, 0)$, where the*

smallest singular value of K_{11} is greater than c_1. Define a matrix S, which is a permuted version of the Schur complement $\Pi^T K \Pi / \Pi_1^T K \Pi_1$, in partitioned form as

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} H_{22} & A_{22}^T \\ A_{22} & 0 \end{pmatrix} - \begin{pmatrix} H_{21} & A_{12}^T \\ A_{21} & 0 \end{pmatrix} \begin{pmatrix} H_{11} & A_{11}^T \\ A_{11} & 0 \end{pmatrix}^{-1} \begin{pmatrix} H_{12} & A_{21}^T \\ A_{12} & 0 \end{pmatrix}.$$

If $m_1 < m$, there exists a positive constant c_2, such that if

$$-e_i^T S_{22} e_i \le \frac{\epsilon}{2Mn} \quad \text{and} \tag{4.6a}$$

$$\|e_i^T S_{21}\| \le \sqrt{\epsilon}, \tag{4.6b}$$

for some i, then

$$\left\| \begin{pmatrix} -e_i^T A_{21} \bar{Y} & e_i^T \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \right\| \le c_2 \sqrt{\epsilon}, \tag{4.7}$$

where $\bar{Y}$ is defined in (4.3) and ϵ is any nonnegative scalar.

Proof. Assume that $m_1 < m$ and that there is an i such that (4.6a) and (4.6b) hold for some nonnegative ϵ. Utilizing the notation of Lemma 4.1 we obtain

$$\begin{aligned}
S_{11} &= H_{22} - H_{21}\bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T H_{12} - A_{12}^T \bar{Y}^T H_{12} - H_{21} \bar{Y} A_{12} + A_{12}^T \bar{Y}^T H_{11} \bar{Y} A_{12} \\
S_{12} &= A_{22}^T - H_{21}\bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T A_{21}^T - A_{12}^T \bar{Y}^T A_{21}^T \\
S_{21} &= A_{22} - A_{21}\bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T H_{12} - A_{21} \bar{Y} A_{12} \\
S_{22} &= -A_{21}\bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T A_{21}^T.
\end{aligned}$$

The definition of S_{22} yields

$$-e_i^T S_{22} e_i = e_i^T A_{21} \bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T A_{21}^T e_i.$$

If the smallest singular value of K_{11} is greater than a positive constant c_1 and $In(K_{11}) = (n_1, m_1, 0)$, Lemmas 2.1 and 2.2 guarantee that $\bar{Z}^T H_{11} \bar{Z}$ is positive definite with its smallest eigenvalue greater than a positive constant. By the assumed compactness of $S(x_0)$, the elements in H_{11} are bounded in magnitude. Consequently, there exists a positive constant $\bar{c}_1$, such that

$$\frac{\epsilon}{2Mn} \ge -e_i^T S_{22} e_i \ge \bar{c}_1 \|\bar{Z}^T A_{21}^T e_i\|^2, \tag{4.8}$$

where the left-hand inequality follows from the assumption that (4.6a) holds. Using the definition of S_{21} we obtain

$$e_i^T A_{22} - e_i^T A_{21} \bar{Y} A_{12} = e_i^T S_{21} + e_i^T A_{21} \bar{Z}(\bar{Z}^T H_{11} \bar{Z})^{-1} \bar{Z}^T H_{12}.$$

It follows from this equation, (4.6b), (4.8), the nonsingularity of K_{11} and the boundedness of $\|H\|$ that there exists a positive constant $\bar{c}_2$, such that

$$\|e_i^T A_{22} - e_i^T A_{21} \bar{Y} A_{12}\| \le \bar{c}_2 \sqrt{\epsilon}. \tag{4.9}$$

The identities $A_{11}\bar{Y} = I$ and $A_{21}^T e_i = A_{11}^T(A_{11}A_{11}^T)^{-1}A_{11}A_{21}^T e_i + \bar{Z}\bar{Z}^T A_{21}^T e_i$ yield

$$e_i^T A_{21}\bar{Y}A_{11} = e_i^T A_{21} - e_i^T A_{21}\bar{Z}\bar{Z}^T + e_i^T A_{21}\bar{Z}\bar{Z}^T\bar{Y}A_{11}.$$

Utilizing (4.8) we obtain

$$\|e_i^T A_{21} - e_i^T A_{21}\bar{Y}A_{11}\| \leq \bar{c}_3\sqrt{\epsilon}, \tag{4.10}$$

for some positive constant $\bar{c}_3$. Consequently, (4.9) and (4.10) yield the existence of a positive constant c_2, such that

$$\| \begin{pmatrix} -e_i^T A_{21}\bar{Y} & e_i^T \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \| \leq c_2\sqrt{\epsilon}.$$

∎

In the following lemma we use the full-rank assumption on A and Lemma 4.2 to show that if K_{11} has exactly m_1 ($m_1 < m$) negative eigenvalues and has its smallest singular value greater than a positive constant, it is always possible to expand K_{11} so that m_1 is increased by one.

Lemma 4.3. *For a given positive constant c_1, let K be any KKT-matrix in $\mathcal{K}$ with a principal submatrix K_{11} such that $In(K_{11}) = (n_1, m_1, 0)$, where the smallest singular value of K_{11} is greater than c_1. Assume that the factorization of $\Pi_1^T K \Pi_1$ is known, and that the next pivot is to be chosen from the Schur complement $\Pi^T K \Pi / \Pi_1^T K \Pi_1$. If $m_1 < m$, there exists a positive constant c_3 such that there is a pivot in $\Pi^T K \Pi / \Pi_1^T K \Pi_1$ of type A^- or HA whose determinant is less than $-c_3$.*

Proof. Let c_1 be a given positive constant, and K any KKT-matrix in $\mathcal{K}$ with a principal submatrix K_{11} such that $In(K_{11}) = (n_1, m_1, 0)$, where the smallest singular value of K_{11} is greater than c_1. Using the terminology of Lemma 4.2, if K_{11} has inertia $(n_1, m_1, 0)$ and the smallest singular value greater than c_1, Lemma 4.1 and the definition of S guarantee the existence of a constant M such that $\|S\| \leq M$. Let $e_i^T(S_{21} \quad S_{22})$ be a row of S corresponding to a row of A and let l denote the corresponding row number of A. Since A has full row rank, there exists a positive constant $\bar{c}$ such that

$$\|e_l^T A - \sum_{j \neq l} \alpha_j e_j^T A\| \geq \bar{c} \tag{4.11}$$

for all scalars α_j. Given c_1 and M, let c_2 denote the constant from Lemma 4.2. Let $\bar{\epsilon}$ denote the positive constant defined by $c_2\sqrt{\bar{\epsilon}} = \bar{c}/2$. For this choice of $\bar{\epsilon}$, at least one of

$$-e_i^T S_{22} e_i > \frac{\bar{\epsilon}}{2Mn} \quad \text{and} \tag{4.12a}$$

$$\|e_i^T S_{21}\| > \sqrt{\bar{\epsilon}}, \tag{4.12b}$$

must hold, since if (4.12a) and (4.12b) do not hold, Lemma 4.2 implies that there exist α_j-s such that

$$\|e_l^T A - \sum_{j \neq l} \alpha_j e_j^T A\| \leq c_2\sqrt{\bar{\epsilon}}, \tag{4.13}$$

which contradicts (4.11). It is now shown that if (4.12a) holds, there exists a pivot of type A^- and if (4.12a) does not hold but (4.12b) holds there exists a suitable pivot of type HA. Assume that (4.12a) holds. Then $e_i^T S_{22} e_i$ is a pivot of type A^- whose determinant is less than $-\bar{\epsilon}/(2Mn)$. Assume that (4.12b) holds but (4.12a) does not hold. If (4.12b) holds, it must hold that $n - n_1 > 0$, since the length of the row vector $e_i^T S_{21}$ is $n - n_1$. Moreover, there must exist a j such that

$$|e_i^T S_{21} e_j| \geq \sqrt{\frac{\bar{\epsilon}}{n - n_1}}. \tag{4.14}$$

For this j, let P denote the 2×2 matrix defined as

$$P = \begin{pmatrix} e_j^T S_{11} e_j & e_j^T S_{12} e_i \\ e_i^T S_{21} e_j & e_i^T S_{22} e_i \end{pmatrix}.$$

By assumption, (4.12a) does not hold and consequently, since $e_i^T S_{22} e_i \leq 0$, it holds that $|e_i^T S_{22} e_i| \leq \bar{\epsilon}/(2Mn)$. This assumption and the inequality $|e_j^T S_{11} e_j| \leq M$ yield

$$\det(P) \leq \frac{\bar{\epsilon}}{2n} - |e_i^T S_{21} e_j|^2.$$

It follows from (4.14) that $|e_i^T S_{21} e_j| \geq \sqrt{\bar{\epsilon}/n}$, and consequently

$$\det(P) \leq -\frac{\bar{\epsilon}}{2n}.$$

The determinant of P is negative, and therefore P must have one positive and one negative eigenvalue. Consequently, P is a pivot of type HA whose determinant is less than a negative constant. Therefore, there exists a positive constant c_3 for which there is a pivot of type A^- or HA whose determinant is less than $-c_3$. ∎

Based on this lemma, the factorization algorithm is stated in Algorithm 4.1. Initially, the matrix K_{11} has dimension zero, and for such a matrix the assumptions of Lemma 4.3 hold. It follows from Lemma 4.3 that there exists a constant pivot tolerance *tol* so that there is a pivot P of type H^+, A^- or HA for which $|\det(P)| \geq$ *tol*. K_{11} is expanded using this pivot. This process is repeated until K_{11} contains all rows of A. When K_{11} contains all rows of A it is considered the "final" K_{11}-matrix, and the remaining Schur complement is factorized using a regular LBL^T-factorization.

In order to show that this algorithm is well defined, it is essential to show that the required properties of K_{11} are maintained when K_{11} is expanded. The following lemma shows that this is true.

Lemma 4.4. *For a given positive constant c_1, let K be any KKT-matrix in $\mathcal{K}$ with a principal submatrix K_{11} such that $In(K_{11}) = (n_1, m_1, 0)$ and the smallest singular value of K_{11} is greater than c_1. Assume that the factorization of $\Pi_1^T K \Pi_1$ is known, and that the next pivot P is of type H^+, A^- or HA and has $|\det(P)| \geq c_3$, where c_3 is a positive constant. Then, there exists a positive constant c_4 such that if K_{11} is expanded using P as pivot, the expanded K_{11} has inertia $(n_1, m_1, 0)$ and its smallest singular value is greater than c_4.*

Algorithm 4.1. *A factorization of K.*

tol ← pivot tolerance;
repeat
  ex ← pivot P of type H^+, A^- or HA exists with $|\det(P)| \geq tol$;
  if *ex* **then**
    pivot ← P;
  else
    Set *tol* to the largest value for which a pivot P of type H^+, A^- or HA exists with $|\det(P)| = tol$;
    pivot ← P;
  end if
  Expand K_{11} using P as pivot;
  Update m_1 and n_1;
until $m_1 = m$
Factorize the remainder of K using a regular LBL^T-algorithm;

Proof. If K_{11} is expanded using a pivot P of type H^+, A^- or HA, the expanded matrix still has inertia $(n_1, m_1, 0)$. The existence of c_4 remains to be established. Let $\bar{K}_{11}$ denote the expanded K_{11}-matrix. Since the smallest singular value of K_{11} is greater than c_1, there must exist a positive constant $\bar{c}_1$ such that $|\det(K_{11})| \geq \bar{c}_1$. It follows from (3.3) that $\det(\bar{K}_{11}) = \det(K_{11})\det(P)$. If $|\det(P)| \geq c_3$, it follows that $|\det(\bar{K}_{11})| \geq \bar{c}_1 c_3$. Consequently, since the norm of K is bounded, there exists a positive constant c_4 such that the smallest singular value of $\bar{K}_{11}$ is greater than c_4. ∎

This lemma shows that, since A has full row rank, a K_{11}-matrix with inertia $(n_1, m_1, 0)$ and the smallest singular value greater than a positive constant, may be expanded until it contains all rows of A and has m negative eigenvalues. Moreover, the smallest singular value of the final K_{11}-matrix is bounded away from zero. Note also that the pivot tolerance *tol* is bounded away from zero by a constant, since it is greater than a positive constant for each step in the expansion of K_{11}.

If it is required that only sufficiently nonsingular pivots of type H^+, A^- and HA are utilized, until a matrix $\Pi_1^T K \Pi_1$ containing all rows of A is created, this will impose other permutations than the ones provided in a regular algorithm for the LBL^T-factorization. Consequently, it is essential that the boundedness of the norm of the L-matrix is maintained. Since the pivots of the associated B_1-matrix all have singular values bounded away from zero by a constant, the following lemma shows that the boundedness of the norm of L is maintained.

Lemma 4.5. *Assume that $\Pi_1^T K \Pi_1 = L_{11} B_1 L_{11}^T$, with its smallest singular value greater than a positive constant. Furthermore, assume that $\|L_{11}\|$ and $\|L_{21}\|$ are bounded by a constant. If K_{11} is expanded by a pivot that has the absolute value of*

its determinant bounded away from zero by a positive constant, then the expanded L_{11} and L_{21} still have norm bounded by a constant.

Proof. Let P denote the pivot. Since K and $(\Pi_1^T K \Pi_1)^{-1}$ have bounded norm, it follows that the norm of $\Pi^T K \Pi / \Pi_1^T K \Pi_1$ is bounded. Consequently, the norm of P is bounded by a constant. If in addition, the determinant of P is bounded away from zero by a constant in absolute value, the norm of P^{-1} is bounded by a constant. Therefore, the expanded column or columns of L_{21} are bounded in norm, since it follows from (3.3) that they are formed by postmultiplying one or two columns of $\Pi^T K \Pi / \Pi_1^T K \Pi_1$ by P^{-1}. ∎

When a matrix K_{11} that contains all rows of A has been created by accepting pivots of type H^+, A^- and HA, as described in Algorithm 4.1, it is considered the final K_{11}-matrix. For the rest of this section, the matrix K_{11} is assumed to be such a final K_{11}-matrix. The factors from the factorization described in Algorithm 4.1 are also assumed to be known. For this final K_{11}-matrix, the partition of Π from (4.1) is used. This partition induces a partition of L, B and $\Pi^T K \Pi$ as

$$\begin{pmatrix} \Pi_1^T K \Pi_1 & \Pi_1^T K \Pi_2 \\ \Pi_2^T K \Pi_1 & \Pi_2^T K \Pi_2 \end{pmatrix} = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix} \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix}. \quad (4.15)$$

If K_{11} contains all rows of A, all blocks in B_2 correspond to pivots of type H^+, H^- or HH. This means that modifications of B_2 only alter the H-part of K. If the factorization of K given in Algorithm 4.1 is known, the following theorem shows how to obtain a feasible descent direction.

Theorem 4.1. *Let K be any KKT-matrix in $\mathcal{K}$. Assume that the LBL^T- factorization of K given in Algorithm 4.1 is known. Assume that the eigenvalue decomposition of B_2 is known, as $B_2 = U \Lambda U^T$. For a positive tolerance c_{TOL}, let the i-th diagonal element of the matrix $\bar{\Lambda}$, denoted by $\bar{\lambda}_i$, be defined using the i-th diagonal element of Λ, denoted by λ_i, as $\bar{\lambda}_i = \max\{|\lambda_i|, c_{TOL}\}$, and let $\bar{B}_2 = U \bar{\Lambda} U^T$. Let $\bar{B}_1 = B_1$ and let $\bar{B}$ denote the matrix consisting of the diagonal blocks $\bar{B}_1$ and $\bar{B}_2$. Let $z = (s^T \ \mu^T)^T$, where s is an n-vector and μ is an m-vector, be defined as $z = \Pi \tilde{z}$, where $\tilde{z}$ satisfies the equation*

$$L \bar{B} L^T \tilde{z} = \Pi^T \begin{pmatrix} -g \\ 0 \end{pmatrix}.$$

Then, $s = Zu$ for some u, and there exist positive constants c_1 and c_2, such that

$$-g^T Z u \geq c_1 \|Z^T g\|^2 \quad \text{and} \quad \|Z^T g\| \geq c_2 \|u\|.$$

Proof. Since the matrix K_{11} has its smallest singular value greater than a positive constant, the compactness of $S(x_0)$ and the boundedness of $\|L^{-1}\|$ guarantee that $\|B_2\|$ is bounded by a constant. Consequently, $\bar{B}_2$ is a bounded block-diagonal matrix whose smallest eigenvalue is greater than c_{TOL}. Consequently, since

$In(B_1) = (n_1, m, 0)$ and the norm of L is bounded, the matrix $L\bar{B}L^T$ has inertia $(n, m, 0)$ and all singular values bounded away from zero by a constant. Since all rows of A are contained in K_{11}, the modification of B_2 only alters H. Consequently,

$$L\bar{B}L^T = \Pi^T \begin{pmatrix} \bar{H} & A^T \\ A & 0 \end{pmatrix} \Pi,$$

where $\bar{H}$ is a modification of H, and z satisfies the equation

$$\begin{pmatrix} \bar{H} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} s \\ \mu \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix},$$

from which it follows that $s = Zu$ for some u satisfying

$$u = -(Z^T\bar{H}Z)^{-1}Z^Tg. \tag{4.16}$$

Utilizing the boundedness of $\|\bar{H}\|$, premultiplication by g^TZ in (4.16) yields the existence of a positive constant c_1, such that $-g^TZu \geq c_1\|Z^Tg\|^2$. Lemmas 2.1 and 2.2 imply that $Z^T\bar{H}Z$ has its smallest eigenvalue greater than a positive constant. Consequently, utilizing norm properties and the boundedness of $\|(Z^T\bar{H}Z)^{-1}\|$, (4.16) guarantees the existence of a positive constant c_2, such that $c_2\|u\| \leq \|Z^Tg\|$. ∎

Note that the direction s computed in Theorem 4.1 is the Newton direction whenever B_2 is sufficiently positive definite, i.e., when the reduced Hessian is sufficiently positive definite. Only if the reduced Hessian is not sufficiently positive definite is the matrix $\bar{B}$ different from B.

If K_{11} is nonsingular, contains all rows of A and has m negative eigenvalues, but K has more than m negative eigenvalues, there exist feasible directions of negative curvature. Moreover, since $In(K) = In(\Pi_1^TK\Pi_1) + In(\Pi^TK\Pi/\Pi_1^TK\Pi_1)$, it must hold that $\Pi^TK\Pi/\Pi_1^TK\Pi_1$ has at least one negative eigenvalue, and consequently B_2 must have at least one negative eigenvalue. Because of the structure of B_2, the smallest eigenvalue of B_2, and its corresponding eigenvector are readily available. The following theorem shows that the factors from the LBL^T-factorization of K given in Algorithm 4.1 enable the computation of a feasible direction of negative curvature.

Theorem 4.2. *Let K be any KKT-matrix in $\mathcal{K}$ with more than m negative eigenvalues. Assume that the LBL^T-factorization of K given in Algorithm 4.1 is known. Let $w = (d^T \;\; \pi^T)^T$, where d is an n-vector and π is an m-vector, be defined as $w = \Pi\tilde{w}$, where $\tilde{w}$ satisfies the equation*

$$\begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix} \begin{pmatrix} \tilde{w}_1 \\ \tilde{w}_2 \end{pmatrix} = \pm\sqrt{-\lambda_{\min}(B_2)} \begin{pmatrix} 0 \\ u_\lambda \end{pmatrix}, \tag{4.17}$$

where u_λ is an eigenvector of unit norm corresponding to $\lambda_{\min}(B_2)$, the smallest eigenvalue of B_2, and the sign is chosen so that $g^Td \leq 0$. Then, d satisfies $d = Zv$ for some vector v which is bounded in norm, and there exists a positive constant c such that

$$v^TZ^THZv \leq -c\,\lambda_{\min}^2(Z^THZ).$$

Proof. From (4.17) and the definition of w we obtain

$$Kw = \pm\sqrt{-\lambda_{\min}(B_2)}\Pi \begin{pmatrix} 0 \\ L_{22}B_2u_\lambda \end{pmatrix}. \tag{4.18}$$

Since K_{11} contains all rows of A, (4.18) yields $Ad = 0$. Consequently, $d = Zv$ for some v and it follows from (4.17) that

$$v^TZ^THZv = w^TKw = -\lambda^2_{\min}(B_2). \tag{4.19}$$

The boundedness of $\|L\|$ and the identity $L_{22}B_2L_{22}^T = \Pi^TK\Pi/\Pi_1^TK\Pi_1$ guarantee the existence of a positive constant c_1, such that

$$\lambda_{\min}(B_2) \leq c_1\lambda_{\min}(\Pi^TK\Pi/\Pi_1^TK\Pi_1) < 0. \tag{4.20}$$

Since K_{11} has inertia $(n_1, m, 0)$, it contains a nonsingular $m \times m$ submatrix A_B of A. Without loss of generality, we may assume that the partition of A is such that $A = \begin{pmatrix} A_B & A_N \end{pmatrix}$, where A_N is the remaining submatrix of A. Let $\bar{Z}$ be a reduced-gradient-type null-space matrix generated by A_B defined as

$$\bar{Z} = \begin{pmatrix} -A_B^{-1}A_N \\ I \end{pmatrix}.$$

Properties of $\Pi^TK\Pi/\Pi_1^TK\Pi_1$ are independent of the pivoting order in which K_{11} is created. Without loss of generality, assume that K_{11} was created by first factorizing the $2m \times 2m$ principal submatrix of K containing A_B. Then, with appropriate substitution in the definition of S in Lemma 4.2, it follows that $\bar{Z}^TH\bar{Z}$ is a permuted version of the Schur complement of this $2m \times 2m$ principal submatrix in $\Pi^TK\Pi$. Moreover, the $2m \times 2m$ principal submatrix has inertia $(m, m, 0)$, and since the inertia of $(\Pi^TK\Pi/\Pi_1^TK\Pi_1)$ is $(n_1, m, 0)$, it must hold that when the $2m \times 2m$ principal submatrix has been processed, the remaining $n_1 - m$ pivots until K_{11} has been processed are all positive. Hence, [FGM89b, Lemma 3.3] implies that

$$\lambda_{\min}(\Pi^TK\Pi/\Pi_1^TK\Pi_1) \leq \lambda_{\min}(\bar{Z}^TH\bar{Z}) < 0. \tag{4.21}$$

Since there exist only a finite number of nonsingular $m \times m$ submatrices of A, it may be asserted that there exists a positive constant c_2, such that

$$\lambda_{\min}(\bar{Z}^TH\bar{Z}) \leq c_2\lambda_{\min}(Z^THZ) < 0. \tag{4.22}$$

Combining (4.19), (4.20), (4.21) and (4.22), there exists a positive constant c, such that

$$v^TZ^THZv \leq -c\,\lambda_{\min}(Z^THZ)^2.$$

Since L is unit lower-triangular, boundedness of $\|L\|$ implies boundedness of $\|L^{-1}\|$. Since all block-diagonal elements in B_1 have norm greater than a positive constant, $\|B_2\|$ is bounded. Consequently, (4.17) and the definition of v guarantee that the norm of v is bounded. ∎

The strategy for choosing pivots given in this section allows the construction of a nonsingular principal submatrix of K that contains all rows of A and has exactly m negative eigenvalues. This strategy only permits pivots of type H^+, A^- and HA until all rows of A have been processed. However, as the following lemma shows, if the *full* Hessian is sufficiently positive definite, the pivot strategy described in this section is of no impact at all, since *all* pivots in such circumstances are of type H^+, A^- and HA.

Lemma 4.6. *If H is positive definite and A has full row rank, then the only pivots that occur in an LBL^T-factorization of K are of type H^+, A^- or HA.*

Proof. Since K is nonsingular, no pivots of type H^0 or A^0 are used. The positive definiteness of H implies that any principal submatrix of H is positive definite. Consequently, *any* nonsingular principal submatrix K_{11} has inertia $(n_1, m_1, 0)$. This property cannot hold if pivots of type H^-, A^+, HH or AA are used. ∎

The conclusion of the lemma does not hold if the Hessian is not positive definite. Consequently, the strategy described in this section may require permutations in addition to those required for a regular LBL^T-algorithm. For example, a pivot of type H^- would not be accepted unless all rows of A have been processed. From the point of view of sparsity, these additional permutations are undesirable since we wish to minimize the change to the ordering determined by the analyze phase.

One possible scheme for choosing the pivots of type H^+, A^- and HA is to use m pivots of type HA. Such a scheme has been proposed by Gill *et al.* [GMSW87b] in the context of large-scale quadratic programming. They name the HA pivot a *tile* and refer to this scheme as tiling.

Finally, if the problem is unconstrained, the matrix K equals H, and the pivot strategy described in this section equals the strategy of a regular LBL^T-factorization, since if A does not exist, all rows of A are processed at the very first step of the factorization. The computed directions are then equivalent to those computed by Moré and Sorensen [MS79].

5. Necessary Pivot Conditions

In the previous section it was shown that a sufficient condition for the computation of a descent direction and a direction of negative curvature is to use pivots of type H^+, A^- and HA until all rows of A have been processed. This scheme creates a principal submatrix $\Pi_1^T K \Pi_1$ which contains all rows of A, has inertia $(n_1, m, 0)$ and its smallest singular value is greater than a positive constant. However, it may also be necessary to obtain such a principal submatrix in order to compute a feasible descent direction and a feasible direction of negative curvature as in the scheme of Moré and Sorensen [MS79]. In the example of Section 3.4, since $\Pi = I$, the only principal submatrix obtained in the factorization that contains all rows of A is K itself. If a block of type A^-, AA or A^0 is modified then the zero-part of K is altered and the subsequent search direction is not feasible. Hence, in the example of Section 3.4, the only part of B that may be modified is the second element.

However, it was shown that such a modification altered both the A-part and the zero-part of K. Similarly, when computing the direction of negative curvature using an eigenvector of B as a right-hand side vector, only eigenvectors corresponding to blocks of type H^- and HH may be used. Again, the example of Section 3.4 shows that solving for the eigenvector corresponding to the second element of B does not yield a feasible descent direction.

It was shown in Section 4 that a sufficient condition for obtaining a nonsingular principal submatrix containing all rows of A with m negative eigenvalues was to use pivots of type H^+, A^- and HA until all rows of A had been processed. Such a scheme may appear unduly restrictive. However, in this section it is shown that these conditions are also necessary in order to guarantee the ability to obtain such a principal submatrix in a single factorization. We utilize the notation of Section 4 and let $\Pi_1^T K \Pi_1$ denote a leading principal submatrix of $\Pi^T K \Pi$, for which the factors L_{11} and B_1 are known. Also, to simplify the notation, the related matrix K_{11} from (4.2) is used.

The following two lemmas show that unless the inertia of $\Pi_1^T K \Pi_1$ is kept equal to $(n_1, m_1, 0)$ until all rows of A have been processed, it may happen that no nonsingular principal submatrix with m negative eigenvalues, containing K_{11} and all rows of A, exists. Consequently, the results from Section 4 are not applicable and there is no guarantee that we are able to compute descent directions and directions of negative curvature as described in Section 4.

Lemma 5.1. *Let M denote an $m \times n$ matrix of full row rank, where $m < n$. Assume that a partition of M is given such that*

$$M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix},$$

where M_1 has full row rank. Given a nonzero vector x such that $M_1 x = 0$, M_2 may be such that $Mx = 0$.

Proof. If $x \neq 0$ and $M_1 x = 0$, it follows that the rows of M_1 are orthogonal to x. The null space of x^T has dimension $n - 1$. Consequently, if $m < n$, the rows of M_2 may be orthogonal to x. ∎

Lemma 5.2. *Assume that $n > m$. If $In(K_{11}) = (n_1 - t, m_1 + t, 0)$, where $t > 0$, the matrix A may be such that any nonsingular principal submatrix of K, containing K_{11} and all rows of A, has more than m negative eigenvalues, independently of H_{12}, H_{21} and H_{22}.*

Proof. If $t > 0$, Lemma 2.1 implies that there exists a nonzero vector p_1 such that $p_1^T H_{11} p_1 < 0$ and $A_{11} p_1 = 0$. If $n > m$, Lemma 5.1 shows that A_{12}, A_{21} and A_{22} may be such that $A_{21} p_1 = 0$ in which case

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} p_1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and}$$

$$\begin{pmatrix} p_1^T & 0 \end{pmatrix} \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} p_1 \\ 0 \end{pmatrix} = p_1^T H_{11} p_1 < 0.$$

Consequently, Lemma 2.1 implies that K has more than m negative eigenvalues. If *any* nonsingular principal submatrix of K containing K_{11} and all rows of A is considered, by extending the vector p_1 with the appropriate number of zero elements, it follows that such a principal submatrix has more than m negative eigenvalues. This property depends only on the matrices A and K_{11}, and therefore it is independent of the matrices H_{12}, H_{21} and H_{22}. ∎

The following example illustrates that even when the reduced Hessian is positive semidefinite, if it is singular, then it may be necessary to maintain a K_{11} with the number of negative eigenvalues equal to m_1 for Theorem 4.1 to apply. Let

$$H = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} I & 0 \end{pmatrix}, \tag{5.1}$$

for which the reduced Hessian is zero. If the corresponding K-matrix is factorized, the first pivot must be either of type HA or HH. If an element of type HH is chosen as first pivot, it is not possible to obtain a nonsingular principal submatrix of K, containing this pivot and having as many negative eigenvalues as it contains rows of A. Consequently, Theorem 4.1 does not apply. For this situation, when the reduced Hessian is positive semidefinite and singular, Conn and Gould [CG84] have shown how to compute a descent direction from a single factorization of K. That scheme is not considered here, since it has not been shown that the descent direction given by that scheme satisfies the conditions of Theorem 4.1. In practice, however, the scheme of Conn and Gould [CG84] may be a viable alternative.

On the other hand, it is of interest to detect as early as possible, if K has inertia different from $(n, m, 0)$. However, the next lemma shows that given *any* nonsingular principal submatrix K_{11} with no more than m negative eigenvalues, it may hold that K is nonsingular and has m negative eigenvalues.

Lemma 5.3. *If* $In(K_{11}) = (n_1 - t, m_1 + t, 0)$, *where* $m_1 + t \leq m$, *it may hold that* $In(K) = (n, m, 0)$.

Proof. If $In(K_{11}) = (n_1 - t, m_1 + t, 0)$, where $m_1 + t \leq m$, it may hold that K_{11} can be expanded with all rows of H so that $In(K_{11}) = (n - \bar{t}, m_1 + \bar{t}, 0)$, where $m_1 + t \leq m_1 + \bar{t} \leq m$. Consequently, if we let $\bar{Z}$ denote a matrix whose columns form an orthonormal basis for the null space of the rows of A contained in the expanded K_{11}, Lemma 2.1 yields $In(\bar{Z}^T H \bar{Z}) = (n - m_1 - \bar{t}, \bar{t}, 0)$. Let v_i, $i = 1, \ldots, n - m_1$, denote orthonormal eigenvectors corresponding to eigenvalues λ_i of $\bar{Z}^T H \bar{Z}$, where the eigenvectors are sorted so that $\lambda_i \leq \lambda_{i+1}$. If the remaining rows of A equal $v_i^T \bar{Z}^T$, $i = 1, \ldots, m - m_1$, then A has full row rank and $p^T H p > 0$ for all $p \neq 0$ such that $Ap = 0$. Consequently, Lemma 2.1 yields $In(K) = (n, m, 0)$. ∎

6. A Sufficient Pivoting Method

If the sufficient pivot conditions described in Section 4 are applied when factorizing K, descent directions and directions of negative curvature may be computed as described in Theorems 4.1 and 4.2. In particular, the computed directions satisfy sufficiency requirements for applying known results for unconstrained minimization. In this section we review two different methods and apply them to our linear equality-constrained problem. The first method is a curvilinear linesearch method and the second is a regular linesearch method. In the discussion, at a particular iterate k, the methods use a feasible descent direction s_k and a feasible direction of negative curvature d_k. The direction d_k always satisfies $d_k^T g_k \leq 0$. If no feasible descent direction exists, s_k is to be interpreted as a zero vector. Similarly, if no feasible direction of negative curvature exists, d_k is to be interpreted as a zero vector.

6.1. A curvilinear linesearch method

In the method for unconstrained minimization suggested by Moré and Sorensen [MS79], the next iterate x_{k+1} is determined by searching along an arc emanating from x_k, defined by the univariate function

$$\phi_k(\alpha) = f(x_k + \alpha^2 s_k + \alpha d_k), \tag{6.1}$$

where s_k is a descent direction and d_k is a direction of negative curvature. In this subsection, it is assumed that s_k is computed as in Theorem 4.1. It is also assumed that d_k is computed as in Theorem 4.2 whenever $Z^T H_k Z$ has at least one negative eigenvalue. If $Z^T H_k Z$ has no negative eigenvalues, then $d_k = 0$.

Let $\mu \in (0,1)$, $\eta \in [\mu, 1)$ and $\beta > 0$ denote preassigned constants. The scalar $\alpha_k \in (0, \beta]$ is determined in such a way as to attempt to satisfy

$$\phi_k(\alpha_k) \leq \phi_k(0) + \tfrac{1}{2}\mu\phi_k''(0)\alpha_k^2 \quad \text{and} \tag{6.2a}$$

$$\phi_k'(\alpha_k) \geq \eta(\phi_k'(0) + \phi_k''(0)\alpha_k). \tag{6.2b}$$

In practice, condition (6.2b) is not always satisfied by α_k. For a complete discussion on how to generate α_k we refer to the original paper of Moré and Sorensen [MS79].

The next iterate x_{k+1} is defined as

$$x_{k+1} = x_k + \alpha_k^2 s_k + \alpha_k d_k. \tag{6.3}$$

The vectors s_k and d_k are both feasible, and we may write $s_k = Zu_k$ and $d_k = Zv_k$ for some vectors u_k and v_k. Consequently, utilizing u_k and v_k, the problem is unconstrained. The following lemma shows that u_k and v_k satisfy the requirements of Moré and Sorensen [MS79].

Lemma 6.1. *Let $\{x_k\}_{k=0}^{\infty}$ denote an infinite sequence of points in $S(x_0)$. For each such point x_k, assume that s_k is computed as in Theorem 4.1 and assume that d_k is computed as in Theorem 4.2 whenever $Z^T H_k Z$ has at least one negative eigenvalue,*

and that $d_k = 0$ otherwise. Then, it holds that $s_k = Zu_k$ for some u_k and $d_k = Zv_k$ for some v_k such that

$$g_k^T Z u_k \to 0 \quad \textit{implies} \quad Z^T g_k \to 0 \quad \textit{and} \quad u_k \to 0$$

and

$$v_k^T Z^T H_k Z v_k \to 0 \quad \textit{implies} \quad \lambda_k \to 0 \quad \textit{and} \quad d_k \to 0,$$

where λ_k is the minimum of zero and $\lambda_{\min}(Z^T H_k Z)$.

Proof. Theorem 4.1 implies that $s_k = Zu_k$ for some u_k. Assume that $g_k^T Z u_k \to 0$. Theorem 4.1 implies that $Z^T g_k \to 0$ and $u_k \to 0$.

Theorem 4.2 implies that $d_k = Zv_k$ for some v_k. If $\lambda_{\min}(Z^T H_k Z) \geq 0$ for some k, then $v_k = 0$. Consequently, without loss of generality, it may be assumed that $\lambda_{\min}(Z^T H_k Z) < 0$ for all k. Assume that $v_k^T Z^T H_k Z v_k \to 0$. Theorem 4.2 implies that $\lambda_{\min}(Z^T H_k Z) \to 0$. Using the assumed boundedness of $\|L\|$, it follows from (4.17) and (4.19) that there exists a positive constant $\bar{c}$ such that

$$\|Zv_k\| \leq \bar{c}(-v_k^T Z^T H_k Z v_k)^{1/4}.$$

Hence, since the columns of Z are linearly independent, $v_k^T Z^T H_k Z v_k \to 0$ implies that $v_k \to 0$, as required. ∎

Consequently, using Lemma 6.1, the results of Moré and Sorensen [MS79] may be applied directly.

Theorem 6.1. (Moré and Sorensen [MS79]) *If an infinite sequence $\{x_k\}_{k=0}^{\infty}$ is generated as defined in (6.3), any limit point $\bar{x}$ satisfies*

$$Z^T \nabla f(\bar{x}) = 0 \quad \textit{and} \quad \lambda_{\min}(Z^T \nabla^2 f(\bar{x}) Z) \geq 0.$$

Proof. See Moré and Sorensen [MS79, Theorem 6.2]. ∎

6.2. A regular linesearch method

The method proposed by Forsgren *et al.* [FGM89b] is a method that uses regular linesearch at all iterates. In order to modify this unconstrained method to a method for linear equality-constrained problems utilizing the LBL^T-factorization, a positive constant ϵ is given and the direction of negative curvature d_k is computed as in Theorem 4.2, but $d_k = 0$ if $d_k^T H_k d_k \geq -\epsilon$. Note that $d_k^T H_k d_k$ is available without needing to compute d_k, since Theorem 4.2 yields $d_k^T H_k d_k = -\lambda_{\min}^2(B_2)$. The direction s_k is assumed to be computed as in Theorem 4.1.

At iteration k, the search direction p_k is defined to be a linear combination of the descent direction s_k and the direction of negative curvature d_k, as

$$p_k = s_k + \beta_k d_k,$$

where the scalar β_k is zero if $d_k = 0$, and it is chosen so that $p_k^T H_k p_k \leq d_k^T H_k d_k$ otherwise. We refer to Forsgren *et al.* [FGM89b] for details.

Let μ and γ denote preassigned constants such that $\mu \in (0, \frac{1}{2})$ and $\gamma \in (0,1)$. Given x_k and $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$, where $\alpha_{\min}$ and $\alpha_{\max}$ are positive constants, the number i_k is defined to be the smallest nonnegative integer i such that

$$f(x_k + \gamma^i \alpha_k p_k) \leq f(x_k) + \mu \gamma^i \alpha_k g_k^T p_k \qquad \text{if} \quad d_k = 0; \quad (6.4a)$$

$$f(x_k + \gamma^i \alpha_k p_k) \leq f(x_k) + \mu \gamma^i \alpha_k g_k^T p_k + \frac{\mu^2 \gamma^{2i} \alpha_k^2}{2} p_k^T H_k p_k \quad \text{if} \quad d_k \neq 0. \quad (6.4b)$$

The next iterate x_{k+1} is defined as

$$x_{k+1} = x_k + \gamma^{i_k} \alpha_k p_k. \qquad (6.5)$$

With s_k and d_k computed from Theorems 4.1 and 4.2, p_k satisfies $Ap_k = 0$ and we may write $p_k = Zv_k$ for some v_k. The analogy with the unconstrained case is clearer if this representation is used. The convergence of this linesearch follows from the convergence analysis of Kaniel and Dax [KD79]. The reduced gradient is zero the smallest eigenvalue of the reduced Hessian is greater than a small negative number at all limit points.

Theorem 6.2. *If an infinite sequence $\{x_k\}_{k=0}^{\infty}$ is generated as defined in (6.5), any limit point $\bar{x}$ satisfies*

$$Z^T \nabla f(\bar{x}) = 0 \quad \textit{and} \quad \lambda_{\min}(Z^T \nabla^2 f(\bar{x}) Z) \geq -c\sqrt{\epsilon},$$

where c is a positive constant.

Proof. Since $p_k^T H_k p_k \leq d_k^T H_k d_k < -\epsilon$ whenever $d_k \neq 0$, it follows from the convergence analysis of Kaniel and Dax [KD79] that there must exist a finite I such that $d_k = 0$ for all $k \geq I$. Moreover, it follows from the convergence analysis of Kaniel and Dax [KD79] that Theorem 4.1 gives a $p_k = Zu_k$ such that $\lim_{k \to \infty} Z^T g_k = 0$. Theorem 4.2 guarantees the existence of a positive constant c such that $\lambda_{\min}(Z^T H_k Z) > -c\sqrt{\epsilon}$ for all $k \geq I$. The result now follows from the continuity of $\nabla^2 f$. ∎

7. Artificial Constraints

In Section 4 we discussed what conditions may be imposed upon the pivots in order to ensure the ability to compute descent directions and directions of negative curvature from a single factorization of K. An alternative strategy for yielding a descent direction or a direction of negative curvature would be to factorize K using a regular LBL^T-factorization, and from the inertia of B deduce the inertia of the reduced Hessian. If the reduced Hessian has at least one negative eigenvalue, an *artificial constraint* may be added to A, so that the number of positive eigenvalues of K is increased by one, and consequently the number of negative eigenvalues of the

reduced Hessian is reduced by one. An artificial constraint is a temporary additional constraint, that is not specified in the original problem. The only requirement for an artificial constraint is linear independence from the original constraints and other artificial constraints. Artificial constraints do not restrict the feasible region, since they are only introduced at a particular iterate, and may be removed from the problem at any iterate. If a scheme for adding artificial constraints is known, a positive definite reduced Hessian could be obtained by adding a sufficient number of such constraints. Unless the dimension of the reduced Hessian has been reduced to zero, a descent direction can be obtained. A method for computing a direction of negative curvature for a positive-definite reduced Hessian in the presence of artificial constraints has been proposed by Forsgren *et al.* [FGM89a].

However, as the following lemma shows, to find an artificial constraint that reduces the number of negative eigenvalues of the reduced Hessian by one is equivalent to finding a direction of negative curvature in the null space of A. Consider the case when a nonsingular KKT-matrix

$$K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}$$

is given, where K has more than m negative eigenvalues. The question of finding an additional artificial constraint a such that

$$\bar{K} = \begin{pmatrix} H & A^T & a \\ A & 0 & 0 \\ a^T & 0 & 0 \end{pmatrix}$$

has one more positive eigenvalue than K is equivalent to finding a direction of negative curvature for the reduced Hessian corresponding to K. The precise statement is given in the following lemma, and uses the solution of the equation

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ \mu \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix}. \tag{7.1}$$

Lemma 7.1. *If $In(\bar{K}) = In(K) + (1,0,0)$, then p from (7.1) is a direction such that $Ap = 0$ and $p^THp < 0$. Conversely, if q is a direction such that $Aq = 0$ and $q^THq < 0$, then $In(\bar{K}) = In(K) + (1,0,0)$ for $a = Hq$.*

Proof. Assume that $In(\bar{K}) = In(K) + (1,0,0)$. Let $w^T = (a^T \;\; 0)$ and let $u^T = (p^T \; \mu^T)$. It follows that

$$\bar{K} = \begin{pmatrix} K & w \\ w^T & 0 \end{pmatrix}$$

and that u solves the equation $Ku = w$. Sylvester's law of inertia implies that $w^TK^{-1}w < 0$. Using the identity $Ku = w$, it follows that $u^TKu < 0$. Consequently, (7.1) yields $p^THp = u^TKu < 0$ and $Ap = 0$.

Assume that q is a vector such that $Aq = 0$ and $q^THq < 0$. Let $u^T = (\ q^T\ 0)$ and $w = Ku$. We get $w^TK^{-1}w = q^THq < 0$. If $a = Hq$, it follows that $w^T = (a^T\ 0)$. Sylvester's law of inertia implies that $In(\bar{K}) = In(K) + (1,0,0)$. ∎

Consequently, the ability to add an artificial constraint is linked with the ability to compute a direction of negative curvature in the null space of A. As an example, consider the case when no constraints exist, and $H = I - ee^T$, where e is an n-vector with all components one. This matrix has one negative eigenvalue and $n-1$ positive eigenvalues. If less than n artificial bounds are added to H, the corresponding K-matrix will have a reduced Hessian that is not positive definite. However, if the single artificial constraint e^T is added, the corresponding reduced Hessian is positive definite. Consequently, although there exist artificial constraints to add, we do not know how to compute them directly. Conn and Gould [CG84] have given a computational scheme for obtaining a direction of negative curvature from the LBL^T-factors of K. However, this scheme requires the solution of a system of equations with an $m \times m$ "triangular-like" matrix.

8. A Descent Method

If no attempt is made to avoid altering the KKT-matrix when the reduced Hessian is positive definite, we may consider an algorithm that yields a descent direction in a single factorization. When forming the factors, pivots corresponding to elements of H are modified, if necessary, so that the principal submatrix factorized at each step is nonsingular and has as many positive eigenvalues as it contains rows of H. This modification corresponds to adding to the diagonal of H, and may be expressed as a positive semidefinite diagonal matrix D of the same dimension as H. If the pivots are modified so that the factorized principal submatrix has its smallest singular value bounded away from zero by a constant, this yields a correction matrix D with bounded norm such that $Z^T(H + D)Z$ is positive definite and has its smallest eigenvalue bounded away from zero by a constant. Such a correction can be computed in a single factorization and requires only the permutations of a regular LBL^T-algorithm. However, in this method, the correction matrix D may be substantial even if the reduced Hessian is positive definite. Therefore, there is no guarantee that this method has the same rate of convergence as Newton's method. For example, if the original ordering of K is used in the factorization, this method will modify H so that $H + D$ is positive definite. Consequently, if Z^THZ is positive definite, but not H, the KKT-matrix is modified unnecessarily. On the other hand, if initially m pivots of type HA are chosen, H will be modified only if the reduced Hessian is not sufficiently positive definite. However, the ordering of the latter case is such that the second-order method described in Section 6 would not require additional permutations.

Given the modified K-matrix, a descent direction p may be obtained by solving the equation

$$\begin{pmatrix} H + D & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\mu \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix}. \tag{8.1}$$

Upon termination of the factorization algorithm described in this section, the LBL^T-factors of the left-hand side matrix of (8.1) are known, and the search direction p may be obtained from these factors. Since $Z^T(H + D)Z$ is positive definite with bounded norm and smallest eigenvalue greater than a positive constant, the search direction p from (8.1) is a sufficient descent direction. Utilizing some inexact linesearch procedure, for example the ones described in Section 6 for the case $d = 0$, it follows that the reduced gradient is zero at all limit points of a generated sequence.

9. A Delta-Method

Assume that for a fix positive constant δ, a search direction p is computed at each iteration from the equation

$$\begin{pmatrix} H + \delta I + D & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\mu \end{pmatrix} = \pm \begin{pmatrix} g \\ 0 \end{pmatrix}, \tag{9.1}$$

where D is a positive semidefinite matrix with bounded norm and the sign of the right-hand side vector is chosen so that $g^T p \leq 0$. The matrix D has all elements zero when the reduced Hessian is sufficiently positive definite. Nonzero elements in D arise for two reasons. Firstly, to ensure that the smallest singular value of the left-hand side matrix of (9.1) is bounded away from zero. Secondly, if having formed an initial factorization there are more than m negative eigenvalues, we may alter B, where possible, to reduce the number of negative eigenvalues. The constant δ may be chosen numerically small, for example in the order of the square-root of the machine precision. Ideally, the ordering in the factorization of the modified K-matrix, where δI is added to H, may then be kept to whatever is satisfactory for preserving sparsity. However, as was pointed out by the example (5.1), the Schur complement may be singular, and there may be no rows of H left, so equation (9.1) cannot be guaranteed to be solved using a single factorization. However, if it is determined that a single factorization is impossible, by refactorizing, utilizing the factorization method described in Section 8 , we may *always* guarantee a positive semidefinite correction matrix D such that the left-hand side matrix of (9.1) is sufficiently removed from a singular matrix. In practice, we may hope that a small value of δ would not impact the rate of convergence compared with Newtons method in the neighborhood of a local minimizer where the reduced Hessian is positive definite. Moreover, it may be expected that the event that a single factorization is impossible, due to singularity arising when factorizing the left-hand side matrix of (9.1), is rare.

The following lemma shows that unless $Z^T g = 0$, the search direction from (9.1) is a nontrivial descent direction or a nontrivial direction of negative curvature.

Lemma 9.1. *The vector p from (9.1) satisfies $p = Zv$ for some v. It holds that $v^T Z^T g \leq 0$ and at least one of the following conditions is satisfied:*

$$v^T Z^T g \leq -\frac{\delta}{2} v^T Z^T Z v \tag{9.2a}$$

$$v^T Z^T H Z v \leq -\frac{\delta}{2} v^T Z^T Z v. \tag{9.2b}$$

Proof. Since $Ap = 0$ it follows that $p = Zv$ for some v. The sign of p is always chosen so that $v^T Z^T g \leq 0$. Utilizing (9.1) we obtain

$$v^T Z^T H Z v \leq -v^T Z^T g - \delta v^T Z^T Z v - v^T Z^T D Z v. \tag{9.3}$$

Assume that (9.2a) does not hold. Since D is positive semidefinite, (9.3) yields

$$v^T Z^T H Z v \leq -\frac{\delta}{2} v^T Z^T Z v,$$

and consequently (9.2b) holds. ∎

If the same linesearch as for the method of Section 6.2 is applied, we can show that the reduced gradient is zero at all limit points. Let μ and γ denote preassigned constants such that $\mu \in (0, \frac{1}{2})$ and $\gamma \in (0,1)$. Given x_k and $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$, where $\alpha_{\min}$ and $\alpha_{\max}$ are positive constants, the number i_k is defined to be the smallest nonnegative integer i such that

$$f(x_k + \gamma^i \alpha_k p_k) \leq f(x_k) + \mu \gamma^i \alpha_k g_k^T p_k \qquad \text{if(9.2a)holds;} \tag{9.4a}$$

$$f(x_k + \gamma^i \alpha_k p_k) \leq f(x_k) + \mu \gamma^i \alpha_k g_k^T p_k + \frac{\mu^2 \gamma^{2i} \alpha_k^2}{2} p_k^T H_k p_k \text{otherwise.} \tag{9.4b}$$

The next iterate x_{k+1} is defined as

$$x_{k+1} = x_k + \gamma^{i_k} \alpha_k p_k. \tag{9.5}$$

Since p_k satisfies $Ap_k = 0$, we may write $p_k = Zv_k$ for some v_k. The analogy with the unconstrained case is clearer if this representation is used.

Theorem 9.1. *If an infinite sequence $\{x_k\}_{k=0}^{\infty}$ is generated as defined in (9.5), any limit point $\bar{x}$ satisfies $Z^T \nabla f(\bar{x}) = 0$.*

Proof. Consider a sequence $\{x_k\}_{k=0}^{\infty}$. Assume that there for some positive constant ϵ exists an infinite subsequence $\{x_k\}_{k \in J}$ such that $\|Z^T g_k\| \geq \epsilon$ for all $k \in J$. The matrix on the left-hand side of (9.1) has its norm bounded by a constant, and consequently there exists a positive constant c_J such that $\|v_k\| \geq c_J$ for all $k \in J$. It follows from Lemma 9.1 and the convergence analysis of Kaniel and Dax [KD79] that there exists a finite I such that (9.2a) holds for all $k \in J$ such that $k \geq I$. Since (9.2a) holds for all $k \in J$ such that $k \geq I$, it follows from the convergence analysis of Kaniel and Dax [KD79] that there exists a finite $\bar{I}$ such that $\|Z^T g_k\| < \epsilon$ for all $k \in J$ such that $k \geq \bar{I}$. Consequently, $\lim_{k \to \infty} Z^T g_k = 0$, as required. ∎

However, since the search direction is zero whenever $Z^T g = 0$, it follows that no stronger result than convergence to a first-order point is possible with this method. A slightly modified method may be obtained by letting δ be variable, i.e., define a value δ_k at the k-th iteration. If, at the k-th iteration, the initial LBL^T-factorization has more than m negative eigenvalues, then $\delta_{k+1} > \delta_k$, in order to reduce the amount of negative curvature in the null space of A. Otherwise, $\delta_{k+1} < \delta_k$. If $\{x_k\}_{k=0}^{\infty}$ converges to a solution where $Z^T H Z$ is positive definite, then δ_k may be reduced so that $\lim_{k \to \infty} \delta_k = 0$, ensuring that the asymptotic rate of convergence is identical to that of Newton's method.

10. Discussion

The pivot strategy discussed in Section 4 is more restrictive than a regular LBL^T-factorization, since it only allows pivots of type H^+, A^- and HA until all rows of A have been processed. Consequently, if an initial ordering is given by the analyze phase, this pivot strategy is likely to change the ordering more than a regular LBL^T-factorization. To attempt to maintain sparsity of the factors, it is desirable to reduce the number of additional pivots required in the numerical phase. In some circumstances it may be possible to accept pivots of other type than H^+, A^- and HA. It was shown in Section 3.4 that modifying a diagonal element of B of type H^- or HH may alter the A-part and the zero-part of K. However, the only altered elements correspond to nonzeros in the outer product created by the corresponding column or columns of L. It is a simple matter to check if these nonzeros of L correspond to rows of A. If they do not, the particular H^- or HH pivot can be accepted even if the number of negative eigenvalues of K_{11} exceeds m_1 by doing so. When K is sparse, it may be a common event that the nonzeros do not correspond to rows of A once a significant portion of the rows of A have been processed. If the nonzeros do not correspond to rows of A, the reduced Hessian has at least one negative eigenvalue, and we can still obtain a feasible direction of negative curvature.

Another scheme is to accept any pivot a regular LBL^T-factorization would accept, keeping track of restart points where K_{11} has inertia $(n_1, m_1, 0)$ and is sufficiently nonsingular. If it turns out that K has inertia $(n, m, 0)$, the reduced Hessian is positive definite, and the Newton direction is a descent direction. Otherwise, when forming the factorization, let K_{11} denote the part of K that is factorized when it is discovered from the Schur-complement that the inertia of K is different from $(n, m, 0)$. If K_{11} contains all rows of A, has inertia $(n_1, m, 0)$ and is sufficiently nonsingular, the results of Section 4 apply. If not all rows of A have yet been processed, an attempt may be made to find pivots of type H^+ and A^+, in order to form a K_{11} that is sufficiently nonsingular, contains all rows of A and has m negative eigenvalues. If this attempt is successful, Theorems 4.1 and 4.2 apply, and the desired directions may be computed. If the attempt is not successful, part of K_{11}, from the latest restart point, may be refactorized, imposing the pivoting strategy of Section 4.

If the number of positive eigenvalues in the reduced Hessian is large compared to the number of negative eigenvalues, we expect the number of pivots of type H^- and HH to be low. Consequently, if the rows of A are processed early in the factorization, there is a high likelihood that these pivots will occur only after all rows of A have been processed. A new version of MA27 (MA47) allows 2×2 pivots in the analyze phase, see Duff *et al.* [DGR*89]. Moreover, these pivots, which in our case would be HA pivots, are preferred over 1×1 pivots. We expect this scheme to make the difference between the additional permutation requirements of the scheme of Section 4 and the additional permutations required by a regular LBL^T-factorization smaller, since in many instances the conditions of Section 4 will be fulfilled automatically.

It may also be observed, that the ability to compute a direction of negative

curvature is only required if the reduced gradient is small in norm. The methods described in Sections 8 and 9 may be applied whenever the norm of the reduced gradient is sufficiently positive. Only at points where the reduced gradient has small norm and the reduced Hessian is not positive definite is it necessary to apply the strategy of Section 4.

As was discussed in Section 7, to add a suitable artificial constraint is equivalent to generating a feasible direction of negative curvature. It can be shown that an artificial constraint that is linearly independent of the constraints in A may be found by one solve with K utilizing a suitable right-hand side. If this artificial constraint increases the number of positive eigenvalues of K by one, a direction of negative curvature may be computed as described in Lemma 7.1. Although this is not guaranteed to be the case, in practice, it may be a viable strategy.

Acknowledgements

We thank Philip Gill for many helpful discussions and suggestions clarifying the results presented. We also thank Per Olov Lindberg for his careful reading of the manuscript and helpful comments.

References

[BK77] J. R. Bunch and L. Kaufman. Some stable methods for calculating inertia and solving symmetric linear systems. *Mathematics of Computation*, 31, 163–179, 1977.

[BP71] J. R. Bunch and B. N. Parlett. Direct methods for solving symmetric indefinite systems of linear equations. *SIAM J. on Numerical Analysis*, 8, 639–655, 1971.

[CG84] A. R. Conn and N. I. M. Gould. On the location of directions of infinite descent for nonlinear programming algorithms. *SIAM J. on Numerical Analysis*, 21, 1162–1179, 1984.

[DGR*89] I. S. Duff, N. I. M. Gould, J. K. Reid, J. A. Scott, and K. Turner. *The factorization of sparse symmetric indefinite matrices.* Technical Report CSS 236, Computer Science and Systems Division, AERE Harwell, Oxford, England, 1989.

[DR82] I. S. Duff and J. K. Reid. *MA27 – a set of Fortran subroutines for solving sparse symmetric sets of linear equations.* Technical Report R-10533, Computer Science and Systems Division, AERE Harwell, Oxford, England, 1982.

[DR83] I. S. Duff and J. K. Reid. The multifrontal solution of indefinite sparse symmetric linear equations. *ACM Transactions on Mathematical Software*, 9, 302–325, 1983.

[DS89] J. E. Dennis, Jr. and R. B. Schnabel. A view of unconstrained optimization. In G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, editors, *Handbooks in Operations Research and Management Science, volume 1. Optimization*, chapter 1, pages 1–72, North Holland, Amsterdam, New York, Oxford and Tokyo, 1989.

[FF77] R. Fletcher and T. L. Freeman. A modified Newton method for minimization. *J. Optimization Theory and Applications*, 23, 357–372, 1977.

[FGM89a] A. L. Forsgren, P. E. Gill, and W. Murray. *On the identification of local minimizers in inertia-controlling methods for quadratic programming.* Report SOL 89-11, Department of Operations Research, Stanford University, 1989. (To appear in SIAM J. on Matrix Analysis and Applications).

[FGM89b] A. L. Forsgren, P. E. Gill, and W. Murray. *A modified Newton method for unconstrained minimization.* Report SOL 89-12, Department of Operations Research, Stanford University, 1989.

[FM68] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* John Wiley and Sons, Inc., New York, London, Sydney and Toronto, 1968.

[GM74] P. E. Gill and W. Murray. Newton-type methods for unconstrained and linearly constrained optimization. *Mathematical Programming*, 7, 311–350, 1974.

[GMSW85] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Software and its relationship to methods. In P. T. Boggs, R. H. Byrd, and R. B. Schnabel, editors, *Numerical Optimization 1984*, pages 139–159, SIAM, Philadelphia, 1985.

[GMSW87a] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Maintaining LU-factors of a general sparse matrix. *Linear Algebra and its Applications*, 88/89, 239–270, 1987.

[GMSW87b] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. *A Schur-complement method for sparse quadratic programming.* Report SOL 87-12, Department of Operations Research, Stanford University, 1987.

[Gol80] D. Goldfarb. Curvilinear path steplength algorithms for minimization which use directions of negative curvature. *Mathematical Programming*, 18, 31–40, 1980.

[Gou85] N. I. M. Gould. On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem. *Mathematical Programming*, 32, 90–99, 1985.

[GV89] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, Maryland, second edition, 1989.

[KD79] S. Kaniel and A. Dax. A modified Newton's method for unconstrained minimization. *SIAM J. on Numerical Analysis*, 16, 324–331, 1979.

[McC77] G. P. McCormick. A modification of Armijo's step-size rule for negative curvature. *Mathematical Programming*, 13, 111–115, 1977.

[MP78] H. Mukai and E. Polak. A second-order method for unconstrained optimization. *J. Optimization Theory and Applications*, 26, 501–513, 1978.

[MS79] J. J. Moré and D. C. Sorensen. On the use of directions of negative curvature in a modified Newton method. *Mathematical Programming*, 16, 1–20, 1979.

[MS84] J. J. Moré and D. C. Sorensen. Newton's method. In G. H. Golub, editor, *Studies in Mathematics, volume 24. Studies in Numerical Analysis*, pages 29–82, The Mathematical Association of America, 1984.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SOL 90-6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Newton Methods for Large-Scale Linear Equality-Constrained Minimization		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Anders L. Forsgren and Walter Murray		8. CONTRACT OR GRANT NUMBER(s) N00014-90-J-1242
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research - SOL Stanford University Stanford, CA 94305-4022		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 1111MA
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research - Dept. of the Navy 800 N. Quincy Street Arlington, VA 22217		12. REPORT DATE April 1990 13. NUMBER OF PAGES 29 pp.
		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Linear equality-Constrained minimization; negative curvature; modified Newton method; symmetric indefinite factorization; large-scale minimization; linesearch method.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(please see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SOL 90-6 Newton Methods for Large-Scale Linear Equality-Constrained Minimization, Anders L. Forsgren and Walter Murray (April 1990, 29 pp.).

Newton methods for large-scale minimization subject to linear equality constraints are discussed. For large-scale problems, it may be prohibitively expensive to reduce the problem to an unconstrained problem in the null space of the constraint matrix. We investigate computational schemes that enable the computation of descent directions and directions of negative curvature without the need to know the null-space matrix. The schemes are based on factorizing a sparse symmetric indefinite matrix. Three different methods are proposed for computing the desired directions. Convergence properties for the different methods are suited.