ITERATIVE SOLUTION OF AUGMENTED SYSTEMS ARISING IN INTERIOR METHODS

Anders FORSGREN*  Philip E. GILL†  Joshua D. GRIFFIN‡

UCSD Department of Mathematics
Technical Report NA-05-03
August 2005

Abstract

Each step of an interior method for nonlinear optimization requires the solution of a symmetric indefinite linear system of equations known as the KKT system. These equations involve both the primal and dual variables and become increasingly ill-conditioned as the optimization proceeds. As the problem size increases, the solution of the KKT system by direct matrix factorization may become prohibitively expensive, and eventually iterative solvers provide the only viable alternative. A one-parameter family of equivalent linear systems of equations is formulated that includes the KKT system as a special case. The proposed method is based on applying a preconditioned conjugate-gradient method to a particular system in this family that is positive definite with respect to both the primal and dual variables. Constraint preconditioners are proposed that provably eliminate the inherent ill-conditioning in the linear equations. In addition, two active-set constraint preconditioners are proposed that involve only a subset of the constraints. The proposed method is suitable for nonconvex problems, in the sense that it detects if the KKT system has the wrong matrix inertia, indicating that the solution of the KKT system is not useful.

Key words. Large-scale nonlinear programming, nonconvex optimization, interior methods, augmented systems, KKT systems, iterative methods, conjugate-gradient method, preconditioning.

AMS subject classifications. 49J20, 49J15, 49M37, 49D37, 65F05, 65K05, 90C30

*Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (andersf@kth.se). Research supported by the Swedish Research Council (VR).
†Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (pgill@ucsd.edu). Research supported by the National Science Foundation grants DMS-9973276, and CCF-0082100.
‡Sandia National Laboratories, Livermore, CA 94551-9217 (jgriffi@sandia.gov). Research supported by the National Science Foundation grants DMS-9973276 and CCF-0082100, and the Göran Gustafsson Foundation.
§Simultaneously issued as technical reports at: Department of Mathematics, Royal Institute of Technology, Stockholm; Department of Mathematics, University of California, San Diego; and Sandia National Laboratories, Livermore.
1. Introduction

This paper concerns the formulation and analysis of preconditioned iterative methods for the solution of augmented systems of the form $Bx = b$, where $B$ is an $(n + m) \times (n + m)$ real indefinite matrix of the form

$$B = \begin{pmatrix} H & A^T \\ A & -G \end{pmatrix},$$

with $A$ an $m \times n$ matrix, $H$ symmetric and $G$ symmetric positive semidefinite. The ubiquity of these systems in science and engineering has resulted in them being known by a variety of names, including “augmented systems”, “saddle-point systems”, “KKT systems” and “equilibrium systems”. (The bibliography of the survey by Benzi, Golub and Liesen [2] contains 513 related articles.) The main focus of this paper will be on the solution of augmented systems arising in interior methods for general constrained optimization, where $Bx = b$ represents the symmetrized equations associated with Newton’s method for finding values of the primal and dual variables that satisfy the perturbed KKT optimality conditions (see, e.g., Wright [37]; Forsgren, Gill and Wright [10]). In this context $G$ is diagonal and $Bx = b$ is sometimes referred to as the KKT system.

Many of the benefits associated with the methods discussed in this paper derive from formulating the interior method so that the diagonal $G$ is positive definite. (The treatment of systems with singular $G$ are considered in Section 5.) Without loss of generality, we can rewrite the KKT system as $Bx = b$ with $B$ and $b$ defined so that:

$$\begin{pmatrix} H & -A^T \\ -A & -D \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

where $D$ is a positive-definite diagonal matrix. (This format will simplify some of the algebra in later sections.) If $D$ is positive-definite, it is well known that the augmented system is equivalent to the two smaller systems

$$(H + A^TD^{-1}A)x_1 = b_1 + A^TD^{-1}b_2 \quad \text{and} \quad x_2 = D^{-1}(b_2 - Ax_1),$$

where the system for $x_1$ is known as the condensed system. It is less well known that another equivalent system is the doubly augmented system

$$\begin{pmatrix} H + 2A^TD^{-1}A & A^T \\ A & D \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 + 2A^TD^{-1}b_2 \\ b_2 \end{pmatrix},$$

which has been proposed for use with direct factorization methods by Forsgren and Gill [9]. In this paper we investigate the properties of preconditioned iterative methods applied to the system (1.2) directly or to the equivalent systems (1.3) and (1.4).

If the underlying optimization problem is not convex, the matrix $H$ may be indefinite. The KKT matrix $B$ of (1.2) is said to have correct inertia if the matrix $H + A^TD^{-1}A$ is positive definite. This definition is based on the properties of the underlying optimization problem. Broadly speaking, the KKT system has correct
inertia if the problem is locally convex (for further details see, e.g., Forsgren and Gill [9], Forsgren [8] and Griffin [22]). If the KKT matrix has correct inertia, the systems (1.2)–(1.4) have a common unique solution (see Section 2).

1.1. Properties of the KKT system

The main issues associated with using iterative methods to solve KKT systems are (i) termination control, (ii) inertia control, and (iii) inherent ill-conditioning. The first of these issues is common to other applications where the linear system represents a linearization of some underlying nonlinear system of equations. Issues (ii) and (iii), however, are unique to optimization and will be the principal topics of this paper.

In the context of interior methods, the KKT system (1.2) is solved as part of a two-level iterative scheme. At the outer level, nonlinear equations that define the first-order optimality conditions are parameterized by a small positive quantity $\mu$. The idea is that the solution of the parameterized equations should approach the solution of the optimization problem as $\mu \to 0$. At the inner level, equations (1.2) represent the symmetrized Newton equations associated with finding a zero of the perturbed optimality conditions for a given value of $\mu$. Although the systems (1.2)–(1.4) have identical solutions, an iterative method will generally produce a different sequence of iterates in each case (see Section 3 for a discussion of the equivalence of iterative solvers in this context). An iterative method applied to the augmented system (1.2) or the doubly augmented system (1.4) treats $x_1$ and $x_2$ as independent variables, which is appropriate in the optimization context because $x_1$ and $x_2$ are associated with the independent quantities in the perturbed optimality conditions (i.e., the primal and dual variables). In contrast, an iterative solver for the condensed system (1.3) will generate approximations to $x_1$ only, with the variables $x_2$ being defined as $x_2 = D^{-1}(b_2 - Ax_1)$. This becomes an important issue when an approximate solution is obtained by truncating the iterations of the linear solver. During the early outer iterations, it is usually inefficient to solve the KKT system accurately and it is better to accept an inexact solution that gives a residual norm that is less than some factor of the norm of the right-hand side (see, e.g., Dembo, Eisenstat and Steihaug [4]). For the condensed system, the residual for the second block of equations will be zero regardless of the accuracy of $x_1$, which implies that termination must be based on the accuracy of $x_1$ alone. It is particularly important for the solver to place equal weight on $x_1$ and $x_2$ when the system (1.2) is being solved in conjunction with a primal-dual trust-region method (see Gertz and Gill [12], and Griffin [22]). The conjugate-gradient version of this method exploits the property that the norms of the $(x_1, x_2)$ iterates increase monotonically (see Steihaug [34]). This property does not hold for $(x_1, x_2)$ iterates generated for the condensed system.

If the KKT matrix does not have the correct inertia, the solution of (1.2) is not useful, and the optimization continues with an alternative technique based on either implicitly or explicitly modifying the matrix $H$ (see, e.g., Toint [30], Steihaug [34], Gould et al. [21], Hager [23] and Griffin [22]). It is therefore important that the iterative solver is able to detect when $B$ does not have correct inertia.
As the perturbation parameter $\mu$ is reduced, the KKT systems become *increasingly ill-conditioned*. The precise form of this ill-conditioning depends on the formulation of the interior method, but a common feature is that some diagonal elements of $D$ are big and some are small. (It is almost always possible to formulate an interior method that requires the solution of an *unsymmetric* system that does not exhibit inevitable ill-conditioning as $\mu \to 0$. This unsymmetric system could be solved using an unsymmetric solver such as GMRES or QMR. Unfortunately, this approach is unsuitable for general KKT systems because an unsymmetric solver is unable to determine if the KKT matrix has correct inertia.) In Section 3 we consider a preconditioned conjugate-gradient (PCG) algorithm that provably removes the inherent ill-conditioning. In particular, we define a one-parameter family of preconditioners related to the class of so-called *constraint preconditioners* proposed by Keller, Gould, and Wathen [24]. Several authors have used constraint preconditioners in conjunction with the conjugate-gradient method to solve the indefinite KKT system (1.2) with $b_2 = 0$ and $D = 0$ (see, e.g., Lukšan and Vlček [26], Gould, Hribar and Nocedal [20], Perugia and Simoncini [29], and Bergamaschi, Gondzio and Zilli [3]).

Recently, Dollar [5], and Dollar et al. [6] have proposed constraint preconditioners for the system (1.2) with no explicit inertial or diagonal condition on $D$, but a full row-rank requirement on $A$ and the assumption that $b_2 = 0$. Methods that require $b_2 = 0$ perform an initial projection step that effectively shifts the right-hand side to zero. The constraint preconditioner then forces the $x_1$ iterates to lie in the null space of $A$. A disadvantage with this approach is that the constraint preconditioner must be exact if subsequent iterates are to lie in the null space. This limits the ability to perform approximate solves with the preconditioner, as is often required when the matrix $A$ has a PDE-like structure that must also be handled using an iterative solver (see, e.g., Saad [31], Notay [27], Simoncini and Szyld [33] and Elman et al. [7]). In Section 3 we consider preconditioners that do not require the assumption that $b_2 = 0$, and hence do not require an accurate solve with the preconditioner.

### 1.2. A PCG method for the KKT system

The goal of this paper is to formulate an iterative method that is able to provide termination control and inertia control as well as provably eliminate the inevitable ill-conditioning associated with interior methods. All these features are present in an algorithm based on applying a preconditioned conjugate-gradient method to the doubly augmented system (1.4). This system is positive definite if the KKT matrix has correct inertia, and gives equal weight to $x_1$ and $x_2$ for early terminations. As preconditioner we use the matrix

$$P = \begin{pmatrix} M + 2A^TD^{-1}A & A^T \\ A & D \end{pmatrix},$$

where $M$ is an approximation of $H$ such that $M + A^TD^{-1}A$ is positive definite.

The equations $Pv = r$ used to apply the preconditioner are solved by exploiting the
2. A Parameterized System of Linear Equations

The following proposition shows how we may embed the indefinite KKT system (1.2) in a family of equivalent linear systems, parameterized by a scalar \( \nu \). This parameterization will allow us to analyze the three systems (1.2)–(1.4) simultaneously.

The equivalence of the systems:

\[
\begin{pmatrix}
M + 2A^T D^{-1} A & A^T \\
A & D
\end{pmatrix}
\begin{pmatrix}
v_1 \\
v_2
\end{pmatrix} =
\begin{pmatrix}
r_1 \\
r_2
\end{pmatrix},
\quad (1.6a)
\]

\[
\begin{pmatrix}
M & -A^T \\
-A & -D
\end{pmatrix}
\begin{pmatrix}
v_1 \\
v_2
\end{pmatrix} =
\begin{pmatrix}
r_1 - 2A^T D^{-1} r_2 \\
r_2
\end{pmatrix},
\quad \text{and} \quad (1.6b)
\]

\[
(M + A^T D^{-1} A)v_1 = r_1 - A^T D^{-1} r_2, \quad v_2 = D^{-1}(r_2 - A v_1), \quad (1.6c)
\]

(see Section 3). This allows us to compute the solution of (1.6a) by solving either (1.6b) or (1.6c) (the particular choice will depend on the relative efficiency of the methods available to solve the condensed and augmented systems.)

We emphasize that the doubly augmented systems are never formed or factored explicitly. The matrix associated with the doubly augmented equations (1.4) is used only as an operator to define products of the form \( v = Bu \). As mentioned above, the equations (1.6a) that apply the preconditioner are solved using either (1.6b) or (1.6c). An important property of the method is that these equations may also be solved using an iterative method. (It is safe to use the augmented or condensed system for the preconditioner equations \( P v = r \) because the inertia of \( P \) is guaranteed by the choice of \( M \) (see Section 3).)

Finally, in Section 4 we formulate and analyze two variants of the preconditioner (1.5) that exploits the asymptotic behavior of the elements of \( D \). The use of these so-called active-set preconditioners may require significantly less work when the underlying optimization problem has more constraints than variables.

1.3. Notation

Unless explicitly indicated otherwise, \( \| \cdot \| \) denotes the vector two-norm or its subordinate matrix norm. Given a real symmetric matrix \( A \), the inertia of \( A \)—denoted by \( \text{In} (A) \)—is the associated integer triple \( (a_+, a_-, a_0) \) indicating the number of positive, negative and zero eigenvalues of \( A \). The spectrum of a (possibly unsymmetric) matrix \( A \) will be denoted by \( \text{eig}(A) \). If \( A \) is symmetric, then \( \text{eig}_{\min} (A) \) denotes the least (i.e., “left-most”) eigenvalue of \( A \). Given vectors \( x_1 \) and \( x_2 \), the column vector consisting of the elements of \( x_1 \) augmented by the elements of \( x_2 \) is denoted by \( (x_1, x_2) \). The vector \( e \) denotes the vector of ones with dimension determined by the context. Throughout the paper, we will consider quantities related to matrices that are implicitly dependent on the parameter \( \mu \). The notation \( O(\mu) \) will be used to denote a quantity that converges to zero at least as fast as \( \mu \). Analogously, \( \Theta(1) \) denotes a bounded quantity that is bounded away from zero as \( \mu \to 0 \), and \( \Omega(1/\mu) \) denotes a quantity whose inverse converges to zero at least as fast as \( \mu \).
Proposition 2.1. Let \( \nu \) denote a scalar parameter. If \( \nu \) is nonzero, then the system (1.2) and the parameterized system

\[
\begin{pmatrix}
H + (1 + \nu)A^T D^{-1} A & \nu A \\
\nu A & \nu D
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
=
\begin{pmatrix}
b_1 + (1 + \nu)A^T D^{-1} b_2 \\
\nu b_2
\end{pmatrix}
\] (2.1)

are equivalent, i.e., (1.2) has a solution \((x_1, x_2)\) if and only if \((x_1, x_2)\) is a solution to (2.1). If \( \nu = 0 \), then (1.2) has a solution \((x_1, x_2)\) if and only if \( x_1 \) is a solution to (2.1) and \( x_2 = D^{-1}(b_2 - Ax_1) \).

The parameterized system may be written as \( B(\nu)x = b(\nu) \), where

\[
B(\nu) = \begin{pmatrix}
H + (1 + \nu)A^T D^{-1} A & \nu A \\
\nu A & \nu D
\end{pmatrix}
\quad \text{and} \quad
b(\nu) = \begin{pmatrix}
b_1 + (1 + \nu)A^T D^{-1} b_2 \\
\nu b_2
\end{pmatrix}.
\]

We are particularly interested in the values \( \nu = -1, 0, 1 \). If \( \nu = -1 \) we obtain the symmetric KKT system (1.2). If \( \nu = 0 \) we obtain the condensed system

\[
(H + A^T D^{-1} A)x_1 = b_1 + A^T D^{-1} b_2 \quad \text{and} \quad x_2 \text{ arbitrary.}
\] (2.2)

Note that for \( \nu = 0 \), the \( x_2 \) variables do not appear in the system and are hence arbitrary. Finally, if \( \nu = 1 \), we obtain the doubly-augmented system

\[
\begin{pmatrix}
H + 2A^T D^{-1} A & A \\
A & D
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
=
\begin{pmatrix}
b_1 + 2A^T D^{-1} b_2 \\
b_2
\end{pmatrix}.
\] (2.3)

The next result follows from Lemma A.1 and gives the inertia of \( B(\nu) \) as a function of \( \nu \).

Proposition 2.2. Consider the matrix \( B(\nu) \) such that

\[
B(\nu) = \begin{pmatrix}
H + (1 + \nu)A^T D^{-1} A & \nu A \\
\nu A & \nu D
\end{pmatrix}.
\]

If \( D \) is symmetric positive definite, then

(i) \( \text{In}(B(\nu)) = \text{In}(H + A^T D^{-1} A) + (m, 0, 0) \) if \( \nu > 0 \),

(ii) \( \text{In}(B(\nu)) = \text{In}(H + A^T D^{-1} A) + (0, m, 0) \) if \( \nu = 0 \), and

(iii) \( \text{In}(B(\nu)) = \text{In}(H + A^T D^{-1} A) + (0, 0, m) \) if \( \nu < 0 \).

This proposition implies that the inertia of \( B(\nu) \) may be determined from the inertia of \( H + A^T D^{-1} A \) and the sign of \( \nu \). In particular, the result gives the inertia of the three alternate systems in the situation where the inertia of \( B \) is correct and the solver can be allowed to continue solving the system. If the inertia of \( B \) is correct then: (i) the doubly augmented system is positive definite; (ii) the KKT system has \( n \) positive eigenvalues and \( m \) negative eigenvalues; and (iii) the condensed system is positive definite (when regarded as a system involving only \( x_1 \)).
Proposition 2.2 implies that it is not worth applying a conjugate-gradient method to the general indefinite KKT system (1.2) because this method is unable to estimate the number of negative eigenvalues of an indefinite matrix. In contrast, the conjugate-gradient method is appropriate for both the doubly augmented system and the condensed system because indefiniteness is immediately indicated by the occurrence of a negative value of $p_j^T C p_j$, where $p_i$ is a conjugate direction and $C$ is either the doubly augmented matrix or the matrix for the condensed system. In other words, the occurrence of a negative value of $p_j^T C p_j$ indicates that the inertia of the system is incorrect and the search for a solution of (1.2) should be abandoned.

3. Constraint Preconditioning for the Linear Equations

The rate of convergence of the CG method may be accelerated by choosing an appropriate symmetric positive-definite preconditioner of the form $P = R^T R$, and applying the CG method to the preconditioned system $R^{-T} B R^{-1} Rx = R^{-T} b$. As is well known, the computations may be arranged so that the preconditioner is applied by solving systems of the form $P v = r$. It is then the eigenvalues of the preconditioned matrix $R^{-T} B R^{-1}$ that determine the rate of convergence. As $\text{eig}(R^{-T} B R^{-1}) = \text{eig}(R^{-1} R^{-T} B R^{-1} R) = \text{eig}(P^{-1} B)$, the analysis may be written in terms of $P^{-1} B$ without regard to $R$. However, it must be emphasized that $P$ must be symmetric positive definite for the standard preconditioned conjugate-gradient method to be well-defined.

Several authors have suggested constraint preconditioners for (1.2) and (1.3), in which $H$ is replaced by a “simpler” approximation matrix $M$ such that $M + A^T D^{-1} A$ is positive definite (see, e.g., Keller, Gould, and Wathen [24], and Bergamaschi, Gondzio and Zilli [3]).

Under certain circumstances, PCG can be applied to all three systems and will give identical results in exact arithmetic.

**Proposition 3.1.** Assume that $H + A^T D^{-1} A$ is positive definite. Consider PCG applied to the KKT system (1.2), the condensed system (1.3) and the doubly augmented system (1.4) with preconditioners

$$
\begin{pmatrix}
M & -A^T \\
-A & -D
\end{pmatrix}, \quad M + A^T D^{-1} A \quad \text{and} \quad \begin{pmatrix}
M + 2A^T D^{-1} A & A^T \\
A & D
\end{pmatrix},
$$

(3.1)

respectively. If $b_2 = 0$, then PCG generates the same sequence of iterates for all three systems (with the definition $x_2 = D^{-1}(b_2 - A x_1)$ for the condensed system).

The first preconditioner of (3.1) is not positive definite, which implies that it does not fit within the standard PCG framework. However, Proposition 3.1 implies that PCG may be applied safely to the KKT system (1.2) in the special situation where $b_2 = 0$. The resulting algorithm is then a projected PCG method that can be shown to be formally equivalent to the standard method applied to the condensed system, see, e.g, Lukšan and Vlček [26] and Gould, Hribar and Nocedal [20].
The condition $b_2 = 0$ may be achieved by choosing a special initial point $y$. In particular, consider the point $(y_1, y_2)$ such that $Ay_1 + Dy_2 = b_2$, and the appropriate preconditioner is used for each system. Let $x$ denote the generic vector of unknowns (the dimension of $x$ will depend on which of the three systems is to be solved). We may for example solve $Py = b$, where $P$ is one of the three appropriate preconditioners. Or we may set $y_1 = 0$, $y_2 = D^{-1}b_2$. Then use PCG with preconditioner $P$ to solve

$$B\hat{x} = b - By,$$

and set $x = y + \hat{x}$.

In general, if $b_2 \neq 0$, it is not safe to apply PCG to the indefinite system (1.2). Moreover, PCG will usually generate different iterates for the condensed system (1.3) and the doubly augmented system (1.4).

Finally, we note that the condensed system (1.3) and doubly augmented system (1.4) may be viewed as being preconditioned versions of each other, as defined in the following result.

**Proposition 3.2.** Consider the CG method applied to the symmetric system $Bx = b$ with symmetric positive-definite preconditioner $P$ and initial iterate $x_0 = 0$. Let $L$ be a nonsingular matrix with the same dimension as $B$. Then, if the CG method is applied to $LBL^T\hat{x} = Lb$ with preconditioner $LPL^T$ and initial iterate $\hat{x}_0 = 0$, the CG iterates are related by the transformation $x = L^T\hat{x}$.

If we consider the decomposition

$$
\begin{pmatrix}
M + 2A^TD^{-1}A & A^T \\
A & D
\end{pmatrix} =
\begin{pmatrix}
I & A^TD^{-1} \\
A^T & I
\end{pmatrix} \begin{pmatrix}
M + A^TD^{-1}A & D \\
D^{-1}A & I
\end{pmatrix},
$$

then Proposition 3.2 implies that the doubly augmented system may be viewed as a particular preconditioned version of the condensed system augmented by the diagonal $D$ for the $x_2$ variables (or vice versa). This is further illustration that the proposed approach gives equal weight to $x_1$ and $x_2$. We prefer to do the analysis in terms of the doubly augmented system because it provides the parameterization based on the scalar parameter $\nu$.

### 3.1. Properties of the constraint preconditioners

We now embed the preconditioners of (3.1) within a family of preconditioners, parameterized by a scalar parameter $\nu$. This parameterization is analogous to the parameterization of the matrices of Proposition 2.1. The parameterization allows a unified analysis of the three preconditioners given in (3.1).

**Definition 3.1.** Associated with the matrix $B(\nu)$ of Proposition 2.2, we define the preconditioner $P(\nu)$ as

$$P(\nu) = \begin{pmatrix}
M + (1 + \nu)A^TD^{-1}A & \nu A^T \\
\nu A & \nu D
\end{pmatrix},$$

with $M$ a symmetric matrix chosen such that $M + A^TD^{-1}A$ is positive definite.
We will assume throughout Section 3 that the problem is such that the sequence of matrices are well defined in the sense that $\sigma_{\min}(\begin{bmatrix} A & D \end{bmatrix}) = \Theta(1)$. In addition, we assume that the preconditioner $M$ is such that $\sigma_{\min}(P(-1)) = \Theta(1)$, which implies that $P(\nu)$ is sufficiently removed from singularity for $\nu = 1$ and $\nu = -1$. This can be achieved, for example, by using a suitable factorization when solving with $P(-1)$ (see Forsgren and Murray [11], Forsgren and Gill [9], and Forsgren [8]).

Proposition 2.2 gives $\ln(P(\nu)) = \ln(M + AT^{-1}A) + \ln(\nu D)$. It follows that $P(\nu)$ is positive definite for $\nu > 0$ and $\ln(P(\nu)) = (n, m, 0)$ for $\nu < 0$. Hence, $P(\nu)$ is nonsingular for $\nu \neq 0$, and it is straightforward to show that for $\nu \neq 0$, the eigenvalues of $P(\nu)^{-1}B(\nu)$ are real and independent of $\nu$. The first lemma reveals the structure of $P(\nu)^{-1}B(\nu)$.

**Lemma 3.1.** Let $B(\nu)$ and $P(\nu)$ be as defined in Proposition 2.2 and Definition 3.1 respectively. Then, for $\nu \neq 0$,

$$
P(\nu)^{-1}B(\nu) = \begin{pmatrix} I & S^{-1}_M S_H & I \\ -D^{-1}A & I & D^{-1}A \\ \end{pmatrix} = \begin{pmatrix} S^{-1}_M S_H \\ D^{-1}AS^{-1}_M(H - M) \end{pmatrix},
$$

where $S_M = M + AT^{-1}A$ and $S_H = H + AT^{-1}A$. In addition, the eigenvalues of $P(\nu)^{-1}B(\nu)$ are independent of $\nu$ and consist of $m$ unit eigenvalues and the $n$ eigenvalues of $(M + AT^{-1}A)^{-1}(H + AT^{-1}A)$.

**Proof.** The expressions for $P(\nu)^{-1}B(\nu)$ follow from the decomposition given in Lemma A.1. The similarity transform of (3.2a) shows that $P(\nu)^{-1}B(\nu)$ has the eigenvalues of $(M + AT^{-1}A)^{-1}(H + AT^{-1}A)$ and $m$ unit eigenvalues, as required.

Next we show that if $m_S$ elements in $D$ corresponding to linearly independent rows of $A$ are $O(\mu)$, then $P(\nu)^{-1}B(\nu)$ has $m$ unit eigenvalues and $m_S$ eigenvalues that are $1 + O(\mu)$.

**Lemma 3.2.** Let $H$, $A$, and $D$ be defined as in (1.2) and let $M$ be a symmetric matrix such that $M + AT^{-1}A$ is positive definite and $\sigma_{\min}(M + AT^{-1}A) = \Theta(1)$. If there exist $k$ diagonal elements of $D$ of order $O(\mu)$ associated with $k$ linearly independent rows in $A$, then the matrix $(M + AT^{-1}A)^{-1}(H + AT^{-1}A)$ has all eigenvalues of order $O(1)$, and at least $k$ eigenvalues $1 + O(\mu)$.

**Proof.** Without loss of generality it may be assumed that the elements of $D$ are ordered so that the $k$ elements that are $O(\mu)$ appear first. This implies that $A$ and $D$ may be partitioned conformally as

$$
A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix},
$$
where $A_1$ is $k \times m$, and $D_1$ is $k \times k$ with all eigenvalues $O(\mu)$. Then

$$A^T D^{-1} A = A_1^T D_1^{-1} A_1 + A_2^T D_2^{-1} A_2.$$  

First we show that $(M + A^T D^{-1} A)^{-1}$ has at least $k$ eigenvalues that are $O(\mu)$. Let $A_1 = U \Sigma V^T$ be the singular-value decomposition of $A_1$, with $U \in \mathbb{R}^{k \times k}$, $\Sigma \in \mathbb{R}^{k \times k}$ and $V \in \mathbb{R}^{n \times k}$. Let $p$ be a $k$-vector of unit length. Then,

$$p^T V A_1^T D_1^{-1} A_1 V p = p^T \Sigma U^T D_1^{-1} U \Sigma p \geq \sigma_k^2 \text{eig}_{\min}(D_1^{-1}),$$  

(3.3)

using the orthonormality of $U$ and $V$. Since $A_1$ has full row rank, we have $\sigma_k > 0$. In addition, all eigenvalues of $D_1$ are $O(\mu)$. Therefore, with $v = V p$, we conclude from (3.3) that $v^T A_1^T D_1^{-1} A_1 v = \Omega(1/\mu)$. It now follows that

$$v^T (M + A^T D^{-1} A) v = v^T M v + v^T A_1^T D_1^{-1} A_1 v + v^T A_2^T D_2^{-1} A_2 v = \Omega(1/\mu),$$  

(3.4)

since $v^T M v = O(1)$ and $v^T A_2^T D_2^{-1} A_2 v \geq 0$. Since $p$ is an arbitrary vector of unit length in $\mathbb{R}^k$, and $v = V p$, we conclude from (3.4) that there is a $k$-dimensional subspace of vectors $v$ such that $v^T (M + A^T D^{-1} A) v$ is $\Omega(1/\mu)$. The Courant-Fischer minimax theorem then ensures that $M + A^T D^{-1} A$ has at least $k$ eigenvalues that are $\Omega(1/\mu)$, see, e.g., [18, Theorem 8.1.2, p. 394]. Since $M + A^T D^{-1} A$ is positive definite, we conclude that $(M + A^T D^{-1} A)^{-1}$ is also positive definite with at least $k$ eigenvalues that are $O(\mu)$.

Let $u_j$ be an eigenvector of unit length corresponding to an eigenvalue $\lambda_j$ of the matrix $(M + A^T D^{-1} A)^{-1}$. Then,

$$u_j^T (M + A^T D^{-1} A)^{-1/2} (H - M)(M + A^T D^{-1} A)^{-1/2} u_j \leq \lambda_j \|H - M\|.$$  

(3.5)

Since $(M + A^T D^{-1} A)^{-1}$ has at least $k$ eigenvalues that are $O(\mu)$, we may apply (3.5) for each such eigenvalue and corresponding orthonormal eigenvector. Analogous to above, the Courant-Fischer minimax theorem then ensures that the matrix

$$(M + A^T D^{-1} A)^{-1/2} (H - M)(M + A^T D^{-1} A)^{-1/2}$$

has at least $k$ eigenvalues that are $O(\mu)$. Note that, by a similarity transform,

$$\text{eig}((M + A^T D^{-1} A)^{-1} (H - M)) = \text{eig}((M + A^T D^{-1} A)^{-1/2} (H - M)(M + A^T D^{-1} A)^{-1/2}).$$

Consequently, $(M + A^T D^{-1} A)^{-1} (H - M)$ has at least $k$ eigenvalues that are $O(\mu)$. In addition, since $\sigma_{\min}(M + A^T D^{-1} A) = \Theta(1)$ and $\|H - M\| = O(1)$, we conclude that $\text{eig}((M + A^T D^{-1} A)^{-1} (H - M)) = O(1)$. The desired results now follow from the identity

$$(M + A^T D^{-1} A)^{-1} (H + A^T D^{-1} A)^{-1} = I + (M + A^T D^{-1} A)^{-1} (H - M),$$

completing the proof.  

A combination of Lemma 3.1 and Lemma 3.2 gives the following result on the eigenvalues of $P(\nu)^{-1} B(\nu)$.
Proposition 3.3. Let $B(\nu)$ and $P(\nu)$ be as defined in Proposition 2.2 and Definition 3.1 respectively. Let $A_S$ denote the submatrix of rows of $A$ for which the corresponding diagonal elements of $D$ are $O(\mu)$. Let $r$ denote the rank of $A_S$. Then, for all $\nu \neq 0$, the matrix $P(\nu)^{-1}B(\nu)$ has all eigenvalues of order $O(1)$, and at least $m + r$ eigenvalues $1 + O(\mu)$. In addition, at least $m$ of the $m + r$ eigenvalues $1 + O(\mu)$ are exactly 1.

Proof. Lemma 3.1 states that $P(\nu)^{-1}B(\nu)$ has $m$ unit eigenvalues plus the eigenvalues of $(M + A^TD^{-1}A)^{-1}(H + A^TD^{-1}A)$. Lemma 3.2 shows that the matrix $(M + A^TD^{-1}A)^{-1}(H + A^TD^{-1}A)$ has all eigenvalues $O(1)$, and also that it has at least $r$ eigenvalues that are $1 + O(\mu)$, completing the proof.

A consequence of this analysis is that if $m_S$ denotes the number of eigenvalues of $D$ that are $O(\mu)$ and the corresponding $m_S \times n$ submatrix $A_S$ has full row rank, then the preconditioned conjugate-gradient method can be expected to give a solution that is $O(\mu)$ accurate in at most $n - m_S$ iterations.

4. Active-set Preconditioning

An advantage of interior methods is that all inequality constraints are treated in the same way—i.e., the solution path does not depend on an explicit prediction of which constraints are active at the solution. However, this advantage can also be a weakness because all constraint gradients are included in the linear system, even those having little or no influence on the solution. For example, if an interior method is applied to a problem with 100 variables and 100,000 inequality constraints, then a KKT system with 100,100 rows and columns must be solved at each iteration. However, if only 50 (say) of the inequalities are active at the solution, an active-set method would need to solve a KKT system of order 150. In the context of an interior method, the partition of constraints into “active” and “inactive” is determined by the magnitude of the elements of the diagonals of $D$ in the KKT system (1.2). Broadly speaking, the active constraints at the solution are estimated by the “small” elements of $D$, and the inactive constraints are estimated by the “large” elements of $D$. In this section we define an active-set preconditioner based on partitioning the elements of $D$ according to magnitude. The preconditioner can be applied with a cost comparable to that of solving the KKT system in an active-set method and allows considerable flexibility in how the diagonals are partitioned—the choice of partition affects only the rate of convergence of the iterative solver, not the rate of convergence of the interior method. Preconditioners similar to those considered here have been proposed by Gertz and Griffin [13] in the context of support vector machine classifiers for large data sets. Preconditioners for the solution of linear programs in standard form have been considered by Gill et al. [15] and Oliveira and Sorensen [28]. An active-set preconditioner for general nonlinear optimization has been proposed by Lukšan, Matonoha and Vlček [25].

Suppose that the elements of $D$ can be partitioned into two disjoint sets $B$ and $S$ that specify the “big” and “small” elements. In particular, assume that $d_{ii} = O(\mu)$ for $i \in S$, and $1/d_{ii} = O(\mu)$ for $i \in B$. For the purposes of this discussion, the precise
order of the big and small elements of $D$ is not important. The only assumption is that over a sequence of KKT systems, the big diagonals go to $+\infty$ and the small diagonals go to $0$ as $\mu \to 0$. Given the partition induced by $B$ and $S$, the matrix $P(\nu)$ can be partitioned and factored as

$$P(\nu) = \begin{pmatrix} M + (1 + \nu)A^T D^{-1} A & \nu A_S^T & \nu A_B^T \\ \nu A_S & \nu D_S & \nu D_B \\ \nu A_B & \nu D_B \\ \nu A_S & \nu D_S & \nu D_B \end{pmatrix} = \begin{pmatrix} I & A_B^T D_B^{-1} \\ I & \nu A_S \\ I & \nu D_S \\ \nu A_S & \nu D_S \end{pmatrix} \begin{pmatrix} I \\ X \nu A_S \\ \nu D_S \\ \nu D_B \end{pmatrix}, \quad (4.1)$$

where $X = M + A_B^T D_B^{-1} A_B + (1 + \nu)A_S^T D_S^{-1} A_S$. With this partition, the nontrivial step when applying $P(\nu)$ is solving with the matrix

$$\begin{pmatrix} M + A_B^T D_B^{-1} A_B + (1 + \nu)A_S^T D_S^{-1} A_S & \nu A_S^T \\ \nu A_S & \nu D_S \end{pmatrix}. \quad (4.2)$$

The matrix (4.2) corresponds to having eliminated the block $\nu D_S$ in $P(\nu)$. This is however not a viable strategy in general, due to the fill-in caused by the term $A_B^T D_B^{-1} A_B$. Lukšan, Matonoha and Vlček [25] suggest an active-set preconditioner based on forming an incomplete factorization of the $(1,1)$ block that avoids unnecessary fill-in from the term $A_B^T D_B^{-1} A_B$. We suggest an alternative strategy. Upon observing that $\|A_B^T D_B^{-1} A_B\| = O(\mu)$, we may ignore the term $A_B^T D_B^{-1} A_B$ in the $(1,1)$ block of (4.2) without significantly altering the preconditioner. This means replacing (4.2) by the matrix $P_S(\nu)$, given by

$$P_S(\nu) = \begin{pmatrix} M + (1 + \nu)A_S^T D_S^{-1} A_S & \nu A_S^T \\ \nu A_S & \nu D_S \end{pmatrix}. \quad (4.3)$$

In active-set constraint preconditioning, $P_S(\nu)$ plays the role of $P(\nu)$ in the standard case. Analogously with the assumptions earlier in the paper, we will therefore throughout this section assume that the problem is such that $\sigma_{\min}\left((A_S \quad D_S)\right) = \Theta(1)$, and that in addition the preconditioner $M$ is such that $M + A_S^T D_S^{-1} A_S$ is positive definite and $\sigma_{\min}(P_S(-1)) = \Theta(1)$. This implies that $P_S(\nu)$ is sufficiently removed from singularity for $\nu = 1$ and $\nu = -1$.

The use of (4.3) instead of (4.2) in (4.1) leads to the active-set preconditioner $P^1_S(\nu)$, where $P^1_S(\nu) = L_S P^2_S L_S^T$ with

$$L_S = \begin{pmatrix} I & A_B^T D_B^{-1} \\ I & \nu A_S \\ I & \nu D_S \end{pmatrix}, \quad P^2_S(\nu) = \begin{pmatrix} M + (1 + \nu)A_S^T D_S^{-1} A_S & \nu A_S^T \\ \nu A_S & \nu D_S \end{pmatrix}. \quad (4.4)$$

Here $P^2(\nu)$ is an alternative preconditioner, where in addition to replacing (4.2) by (4.3), the off-diagonal term $A_B^T D_B^{-1} A_B$ in $L_S$ has been ignored, so that $L_S$ is approximated by $I$. Multiplying together, we obtain

$$P(\nu) = P^1_S(\nu) + \begin{pmatrix} A_B^T D_B^{-1} A_B & 0 \\ 0 & 0 \end{pmatrix} = P^2_S(\nu) + \begin{pmatrix} A_B^T D_B^{-1} A_B & 0 & \nu A_B^T \\ 0 & \nu A_B \end{pmatrix}. \quad (4.4)$$
It remains to establish the properties of the eigenvalues of $P^1_S(\nu)$ and $P^2_S(\nu)$. The next result indicates that $P^1_S(\nu)$ has the same asymptotic behavior as $P(\nu)$.

**Proposition 4.1.** Let $B(\nu)$ and $P^1_S(\nu)$ be as defined in Proposition 2.2 and (4.4) respectively. Let $r$ denote the rank of $A_S$. Then, $P^1_S(\nu)$ is positive definite for all $\nu > 0$. Moreover, for $\nu \neq 0$, the matrix $P^2_S(\nu)^{-1}B(\nu)$ has at least $m + r$ eigenvalues that are $1 + O(\mu)$, of which at least $m$ are $1$.

**Proof.** Proposition 2.1 implies that the preconditioner $P^1_S(\nu)$ is positive definite for all $\nu > 0$. It follows that $\text{eig}(P^1_S(\nu)^{-1}B(\nu)) = \text{eig}(P^2_S(\nu)^{-1}L^{-1}_SB(\nu)L^{-T}_S)$, where $P^2_S(\nu)^{-1}L^{-1}_SB(\nu)L^{-T}_S$ is $2 \times 2$ block-diagonal, with $(1, 1)$ block
\[
\begin{pmatrix}
M + (1 + \nu)A^T_SD^{-1}_SA_S & \nu A^T_SD^{-1}_S \\
\nu A^T_SD^{-1}_S & \nu D^{-1}_S
\end{pmatrix}^{-1}
\begin{pmatrix}
H + A^T_DB^{-1}DA_B + (1 + \nu)A^T_SD^{-1}_SA_S & \nu A^T_SD^{-1}_S \\
\nu A^T_SD^{-1}_S & \nu D^{-1}_S
\end{pmatrix},
\]
and $(2, 2)$ block consisting of the identity matrix of dimension $m_B$. By successively replacing $H$ by $H + A^T_SB^{-1}DB^{-1}A_B$, $A$ by $A_S$, and $D$ by $D_S$ in Lemma 3.1 and Lemma 3.2, we conclude as in Proposition 3.3 that the $(1, 1)$ block has at least $m_S + r$ eigenvalues that are $1 + O(\mu)$. Moreover, at least $m_S$ of these eigenvalues are exactly $1$. The result now follows by appending the $m_B$ unit eigenvalues from the $(2, 2)$ block. 

Next we show that $P^2_S(\nu)$ has almost the same asymptotic behavior as $P(\nu)$ and $P^1_S(\nu)$. For $P^2_S(\nu)$ we assume $\nu > 0$, which ensures that the eigenvalues of $P^2_S(\nu)^{-1}B(\nu)$ are real. The preconditioner $P^2_S(\nu)$ is less expensive to apply than $P^1_S(\nu)$, but the number of unit eigenvalues of the preconditioned matrix decreases from $m$ to $m_S$ because $A_B$ does not appear in $P^2_S(\nu)$. However, as we show in the next proposition, the total number of $1 + O(\mu)$ eigenvalues for $P^2_S(\nu)^{-1}B(\nu)$ is the same as for $P^1_S(\nu)^{-1}B(\nu)$ and $P(\nu)^{-1}B(\nu)$.

**Proposition 4.2.** Let $B(\nu)$ and $P^2_S(\nu)$ be defined as in Proposition 2.2 and (4.4) respectively. Let $r$ denote the rank of $A_S$. Then, $P^2_S(\nu)$ is positive definite for $\nu > 0$. Moreover, for $\nu > 0$, the matrix $P^2_S(\nu)^{-1}B(\nu)$ has at least $m + r$ eigenvalues that are $1 + O(\mu)$. Moreover, at least $m_S$ of these $m + r$ eigenvalues are $1$.

**Proof.** The positive definiteness of $P^2_S(\nu)$ for $\nu > 0$ follows from Proposition 2.1. Using Proposition 2.2 and Lemma A.1 while replacing $A$ by $A_S$ and $D$ by $D_S$ in the definition of $P(\nu)$, it is straightforward to verify that
\[
P^2_S(\nu)^{-1}B(\nu) = \begin{pmatrix} S & T \\ U & I & V \\ W & I \end{pmatrix},
\]
where the matrices $S$, $T$, $U$, $V$ and $W$ are such that
\[
S = (M + A^T_SD^{-1}_SA_S)^{-1}(H + A^T_DB^{-1}A_B),
T = \nu(M + A^T_DB^{-1}A_S)^{-1}A^T_B,
U = D^{-1}_SA_S(M + A^T_SD^{-1}_SA_S)^{-1}(H + A^T_DB^{-1}A_B - M),
V = -\nu D^{-1}_SA_S(M + A^T_DB^{-1}A_S)^{-1}A^T_B, \quad \text{and}
W = D^{-1}_BA_S.
Using the result that $P_S^2(\nu)^{-1}B(\nu)$ has $m_S$ columns from the identity matrix, we conclude that $P_S^2(\nu)^{-1}B(\nu)$ has $m_S$ unit eigenvalues together with the eigenvalues of the matrix

$$
\begin{pmatrix}
S & T \\
W & I
\end{pmatrix}.
$$

Further, note that $\|W\| = O(\mu)$. In addition, because of the assumption that $\sigma_{\min}(P_S(-1)) = \Theta(1)$, it follows from Lemma A.1 that $\|S\| = O(1)$ and $\|T\| = O(1)$. Hence, since eigenvalues are continuous functions of the matrix elements, the remaining $n + m_B$ eigenvalues of $P_S^2(\nu)^{-1}B(\nu)$ must differ by a term $O(\mu)$ from the eigenvalues of the matrix

$$
\begin{pmatrix}
S & T \\
I & I
\end{pmatrix},
$$

i.e., the eigenvalues of $S$ and the $m_B$ unit eigenvalues of $I$. We may apply Lemma 3.2 to $S$, replacing $H$ by $H + A_T^T b^{-1} A_B$, upon observing that $\|A_T^T b^{-1} A_B\| = O(\mu)$, to conclude that $S$ has at least $r$ eigenvalues that are $1 + O(\mu)$. In summary, by considering the $O(\mu)$ perturbation of the eigenvalues incurred by $W$, we conclude that $P_S^2(\nu)^{-1}B(\nu)$ has at least $m + r$ eigenvalues that are $1 + O(\mu)$, of which at least $m_S$ are 1, as required.

We conclude that it is possible to construct appropriate constraint preconditioners based on solving the smaller system (4.3). Moreover, the matrix $P_S(\nu)$ of (4.3) has exactly the same structure as $P(\nu)$. The difference is that the number of rows and columns in the preconditioner has been reduced from $n + m$ to $n + m_S$. Hence, all the previous analysis applies. For our example with 100 variables and 100,000 inequality constraints, a matrix of dimension 150 would need to be factored instead of a matrix of dimension 100,100.

We stress that the partition in $S$ and $B$ is not critical in the sense that the splitting effects only the efficiency of the preconditioner and not the definition of the linear equations that need to be solved. It is not crucial that the correct choice of $S$ is made when forming the preconditioner. If $S$ is too large, the asymptotic performance of the inner iterates is unchanged.

5. On Semidefinite Diagonal Matrices

Up to this point we have assumed that the matrix $D$ in the (2,2) block of the KKT system is positive-definite. In the general case, the last block of equations in the KKT system has the form

$$
Ax_1 + Gx_2 = b_2,
$$

where $G$ is a diagonal matrix with positive and zero entries. If all the constraints of the optimization problem are nonlinear, it is always possible to formulate the interior method so that $G$ is positive definite. For inequality constraints, standard formulations give positive elements in $G$ that are of the order of the perturbation parameter $\mu$ (see, e.g., Vanderbei and Carpenter [35] and Forsgren and Gill [9]). Typically, zero elements of $G$ are associated with linearized equality constraints,
where the corresponding subset of the equations (5.1) are the Newton equations for a zero of the constraint residual. An alternative to direct constraint linearization is to impose equality constraints approximately via a quadratic penalty function. It can be shown that this approach gives a positive element in $G$ of the order of $\bar{\mu}$, where $\bar{\mu}$ is the inverse of the penalty parameter (see, e.g., Gould [19], and Forsgren and Gill [9]). The parameter $\bar{\mu}$ may be allowed to vary with $\mu$, or may be fixed at some small value (see, e.g., Gill et al. [16], and Saunders and Tomlin [32]). Fixing $\bar{\mu}$ defines a regularization of the problem, which allows the formulation of methods that do not require an assumption on the rank of the equality constraint Jacobian. (For more details on the use of regularization in interior methods, see Gill et al. [14], Vanderbei and Shanno [36], and Altman and Gondzio [1]).

However, it may not always be beneficial to regularize linear constraints. Regularization in this context is less crucial because reliable techniques exist for discarding dependent equality constraints. Moreover, interior methods can be defined so that every iterate satisfies the linear equality constraints (see below). With an appropriate choice of constraints, this feature can be used to guarantee that the nonlinear functions and their derivatives are well defined at all points generated by the interior method.

In order to consider KKT systems with a semidefinite $(2,2)$ block, we assume that the variables and equations are preordered to give a system $Bx = b$ such that
\[
\begin{pmatrix}
H & -A^T & -F^T \\
-A & -D & -F \\
-F & & \\
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{pmatrix}
= 
\begin{pmatrix}
b_1 \\
b_2 \\
b_3 \\
\end{pmatrix},
\]
where $D$ is positive definite. Note that we cannot compute the condensed or doubly augmented system for these equation because of the zero block. In this case, the $B$ has correct inertia if $N^T(H + A^TD^{-1}A)N$ is positive definite, where the columns of $N$ form a basis for the null space of $F$ (see Forsgren [8]).

The KKT system (5.2) can be solved using a projection technique similar to that described in Section 3. First, an initial point $y$ is found with first $n$ components $y_1$ such that $Fy_1 = 0$. This vector may be computed in various ways—e.g., by computing an LU factorization of $F^T$ (see, e.g., Gill, Murray and Saunders [17]), or by solving a system for the preconditioning matrix associated with (5.2), where $H$ is replaced by a suitable approximation $M$ (see Gould, Hribar and Nocedal [20]). Once $y$ is known, we may use PCG to solve the system
\[
\begin{pmatrix}
H & -A^T & -F^T \\
-A & -D & -F \\
-F & & \\
\end{pmatrix}
\begin{pmatrix}
\hat{x}_1 \\
\hat{x}_2 \\
\hat{x}_3 \\
\end{pmatrix}
= 
\begin{pmatrix}
b_1 - Hy_1 + A^Ty_2 + Fy_3 \\
b_2 - Ay_1 + Dy_2 \\
0 \\
\end{pmatrix},
\]
and obtain $x = y + \hat{x}$. If the zero elements of $G$ are associated with linear constraints, and the system (5.3) is solved exactly, it suffices to compute the special step $y$ only once, when solving the first system. Then the right-hand side of (5.3) will remain zero for all subsequent iterations.

If the preconditioner can be applied only approximately, then an alternative method can be used based on computing products of the form $N^Tv$ and $Nu$. (Gill,
Murray and Saunders [17] describe how these products may be computed in a numerically stable way without needing to store the columns of $N$.) The requirement that $F\hat{x}_1 = 0$, implies that $\hat{x}_1$ can be written as $\hat{x}_1 = N\hat{p}_1$. Substituting this expression in (5.3) gives the reduced KKT system

$$\begin{pmatrix} N^T H N & -(AN)^T \\ -AN & -D \end{pmatrix} \begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix} = \begin{pmatrix} N^T(b_1 - H y_1 + A^T y_2) \\ -b_2 + A y_1 + D y_2 \end{pmatrix},$$

from which we can define $\hat{x}_1 = N\hat{p}_1$ and $\hat{x}_2 = \hat{p}_2$. This system has a nonsingular (2,2) block and has correct inertia if (5.2) has correct inertia. Moreover, the iterates define exact projections regardless of the accuracy of the solves with the preconditioner. Hence, all the conditions needed for the application of the PCG method proposed in Section 3 apply.

6. Summary and Further Research

A framework has been proposed for applying the preconditioned conjugate-gradient method to KKT systems of the form (1.1) that arise in interior methods for general nonconvex optimization. The proposed methods are based on applying the conjugate-gradient method to the doubly augmented system (1.4), which is positive definite if the underlying optimization problem satisfies the second-order sufficient conditions for optimality. An advantage of the doubly augmented system is that it is positive definite with respect to all the variables.

We have also proposed a class of constraint preconditioners for the doubly augmented system. In particular, we have analyzed two ways of using an estimate of the active set to reduce the cost of applying the preconditioner when there are many inequality constraints. As the solution of the optimization problem is approached, these active-set preconditioners have theoretical performance comparable to constraint preconditioners that include all the constraints.

A useful feature of the proposed method is that the linear equations used to apply the preconditioner need not be solved exactly. Future work will consider the analysis associated with these approximate preconditioners. Future research will also consider preconditioners that utilize better approximations to the matrix $H$ in the (1,1) block. For example, one approach is to generate an approximate preconditioner based on an incomplete inertia-controlling factorization of the KKT system (1.2). For more details on the inertia-controlling factorization for augmented systems in interior methods, see Forsgren and Gill [9], and Forsgren [8].

This paper describes the theoretical basis of our work on the solution of KKT systems using preconditioned conjugate-gradient methods. The practical implications of the analysis in the context of solving very large nonconvex optimization problems will be the subject of a follow-up paper.

Acknowledgments

Part of this work was carried out during the spring of 2003 while Joshua Griffin was visiting KTH with financial support from the Göran Gustafsson Foundation. The
authors would also like to thank Michael Gertz for many stimulating discussions on the theoretical and practical aspects of interior methods.

A. Linear Algebra

In this appendix we review a linear algebra result that is used extensively in the analysis. The following lemma may be verified by direct multiplication.

Lemma A.1. Given a nonsingular matrix $D$, consider the matrix

$$B(\nu) = \begin{pmatrix} H + (1 + \nu)A^T D^{-1} A & \nu A^T \\ \nu A & \nu D \end{pmatrix},$$

where $H$ and $A$ are matrices and $\nu$ is a scalar. Then $B(\nu)$ can be factored in the form

$$B(\nu) = \begin{pmatrix} I & A^T D^{-1} \\ I & D^{-1} A \end{pmatrix} \begin{pmatrix} H + A^T D^{-1} A & \nu D \\ \nu D & \nu I \end{pmatrix} \begin{pmatrix} I & A^T D^{-1} \\ D^{-1} A & I \end{pmatrix}.$$

Moreover, if $H + A^T D^{-1} A$ is nonsingular and $\nu \neq 0$, then $B(\nu)$ is nonsingular, with inverse

$$B(\nu)^{-1} = \begin{pmatrix} (H + A^T D^{-1} A)^{-1} & -\frac{1}{\nu} D^{-1} \\ -\frac{1}{\nu} D^{-1} A (H + A^T D^{-1} A)^{-1} & I \end{pmatrix}.$$

References


