# TECHNIQUES FOR SOLVING GENERAL KKT SYSTEMS

GENE H. GOLUB* and CHEN GREIF†

July 10, 2000

**Abstract.** We consider techniques for solving general KKT systems. In particular, we address the situation of a singular (1,1) block, and focus on ways to eliminate the singularity, either by reducing the system size or by employing an augmented Lagrangian technique. The latter is a parameter-dependent approach. We provide some observations regarding the condition number, the spectrum, and a sensible choice of the parameter. The analysis demonstrates how the case of a singular (1,1) block is different from other cases. We also present a few general results regarding inversion of certain KKT matrices, the spectra of their associated Schur complements, and error estimates for regularized systems. Finally, we present a parameter-dependent preconditioning technique, and discuss its spectral properties.

**1. Introduction.** The acronym 'KKT' stands for Karush-Kuhn-Tucker, and generally refers to first order necessary optimality conditions for solving constrained minimization problems. A precise definition in mathematical terms can be found in Nocedal & Wright [39, p. 328]. The conditions were first derived by Karush in 1939, in his Master's thesis at the University of Chicago [30], but it was only in 1950, when the Princeton mathematicians Kuhn and Tucker published their work [36], that the theory of constrained optimization took off and started taking its modern shape. The interested reader can find a detailed account of the history of the derivation of the KKT conditions and the theory of nonlinear programming in [31] and references therein.

In a linear algebra context, KKT linear systems are of the form

$$\mathcal{K}u \equiv \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix} . \tag{1.1}$$

In this work we focus on the case where $A$ is a sparse symmetric $n \times n$ matrix, $B$ is an $n \times p$ sparse matrix of full column rank, with $p \leq n$ (often $p \ll n$), and $\mathcal{K}$ is nonsingular.

When analyzing the properties of the system and appropriate linear solvers, a key observation is that $\mathcal{K}$ is indefinite. Assuming for the moment that $A$ is nonsingular, this can be seen from the block factorization

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^T A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & -B^T A^{-1} B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix} . \tag{1.2}$$

Indeed, since the factorization (1.2) is a congruence transformation, the inertia of $\mathcal{K}$, which is denoted as $i(\mathcal{K})$ and is defined as the ordered triplet that specifies the number of its positive, negative and zero eigenvalues [29, p. 221, Definition 4.5.6], is equal to the sum of inertias of the matrices $A$ and the Schur complement $-B^T A^{-1} B$. If $A$ is positive definite, then $i(\mathcal{K}) = (n, p, 0)$. If $A$ is not positive definite, little can

*SCCM Program, Gates 2B, Stanford University, Stanford CA 94305-9025, USA, golub@sccm.stanford.edu

†SCCM Program, Gates 2B, Stanford University, Stanford CA 94305-9025, USA, greif@sccm.stanford.edu. The work of this author was supported in part by the Natural Sciences and Engineering Research Council of Canada.

be said about the spectrum, and surely one cannot make any definite observations on the inertia. In any event, $\mathcal{K}$ cannot be positive definite since $e_{n+1}^T \mathcal{K} e_{n+1} = 0$.

The purpose of this report is to address the question of how to solve the system efficiently when no particular assumptions are made with regard to $A$ and $B$, other than ones necessary for maintaining nonsingularity of $\mathcal{K}$. In particular, we shall address the situation of an ill-conditioned or singular (1,1) block. Such a situation occurs in numerous applications. Examples are domain decomposition techniques that are based on solving a Neumann problem in each subdomain (see, for example, [35] for application of FETI to linear elasticity problems, and references therein); computation of thin plate splines [45]; geophysical inverse problems [25]; three-dimensional magnetostatic problems [41]; and many other applications.

Throughout our discussion we often assume that $A$ is positive semidefinite. In this case, if $B$ is of full column rank and $\ker(A) \cap \ker(B^T) = \{0\}$, then $\mathcal{K}$ is nonsingular and the linear system (1.1) has a unique solution [26]. Various other sufficient conditions for nonsingularity of the KKT matrix can be found in the literature—see for example [39, Chap. 16].

One obvious difficulty in having a singular or ill-conditioned (1,1) block is that techniques based on computing the Schur complement cannot be straightforwardly applied. Our proposed methodology for attacking this problem is to transform the system into one where the $(1,1)$ block is nonsingular. We suggest two different strategies. If the null space is of low rank, then an inexpensive Lanczos/Stieltjes procedure can be applied to estimate rapidly a vector of unknowns whose size is equal to the dimension of the null space of $A$. Then, for the rest of the unknowns we get a new system with a nonsingular (1,1) block, and we can proceed to solve by traditional means. If necessary, iterative refinement can be applied to improve the accuracy of the solution. On the other hand, if the nullity of the (1,1) block is high or if the matrix is *nearly* singular, then we apply an augmented Lagrangian procedure [10, 19, 21, 27, 42]. This technique has been considered for a positive definite (1,1) block; we show that applying it for the case where the (1,1) block is singular is possible, and in fact forms a unique situation.

A large variety of methods for solving KKT systems, including solvers that can handle a singular (1,1) block, can be found in the literature. Null-space methods can be applied [20, 39]. (See also [1] for a recent report which includes backward error analysis.) Also of interest are splitting schemes such as the one introduced in [13], which was later generalized to nonsymmetric real positive matrices in [24]. A technique of a different flavor altogether is introduced in [35]. Schur complement techniques [37] and the Uzawa algorithm [2] are popular, especially when the (1,1) block is positive definite. Direct sparse solvers based on stable decompositions [11] are examined and reported to be effective in [41]. It is also possible to apply block preconditioning techniques for Krylov subspace solvers, in the spirit of [38, 34, 15, 17], and others. The reader is referred to Nocedal & Wright [39], Gill, Murray & Wright [20], Björck [8], Strang [48, 49] or Benzi [6] for a detailed description of the existing techniques, and pointers to further references.

The rest of the paper is divided as follows. In Section 2 we introduce a technique based on rapid estimation of a small part of the vector of unknowns using the Lanczos/Stieltjes algorithm. In Section 3 we analyze the augmented Lagrangian technique and perform some convergence analysis. In Section 4 we present a parameter-dependent preconditioning technique. In Section 5 we draw some conclusions. Numerical examples are scattered throughout the manuscript.

**2. Low nullity of the (1,1) block.** Suppose $A$ is a singular symmetric $n \times n$ matrix, whose rank is $n-q$, and $q$ is a small integer.

**2.1. Motivation.** Suppose $A$ can be decomposed as follows:

$$A = Z \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} Z^T, \tag{2.1}$$

where $\Lambda$ is a nonsingular $(n-q) \times (n-q)$ diagonal matrix, and $Z$ is an orthogonal $n \times n$ matrix. Below we attach a 'hat' symbol to any matrix or vector that we multiply from the left by $Z^T$. For example, $\hat{B} := Z^T B$. The system (1.1) can be transformed into

$$\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} \hat{x} + \hat{B} y = \hat{c}, \tag{2.2a}$$

$$\hat{B}^T \hat{x} = d. \tag{2.2b}$$

We can separate the first $n$ equations into two groups. Denote $\hat{x} = \begin{pmatrix} \tilde{x} \\ \hat{x}_n \end{pmatrix}$, $\hat{c} = \begin{pmatrix} \tilde{c} \\ \hat{c}_n \end{pmatrix}$, and $\hat{B} = \begin{pmatrix} \tilde{B} \\ \hat{B}_n^T \end{pmatrix}$, where $\hat{x}_n, \hat{c}_n$ are vectors of dimension $q$ and $\hat{B}_n^T$ is the matrix comprised of the last $q$ rows of $\hat{B}$. Then we have

$$\begin{pmatrix} \Lambda & 0 & \tilde{B} \\ 0 & 0 & \hat{B}_n^T \\ \tilde{B}^T & \hat{B}_n & 0 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \hat{x}_n \\ y \end{pmatrix} = \begin{pmatrix} \tilde{c} \\ \hat{c}_n \\ d \end{pmatrix} . \tag{2.3}$$

In (2.3) the dimensions of each of the zero blocks are different yet clear from the dimensions of the other quantities in the matrix. By substituting

$$\tilde{x} = \Lambda^{-1}(\tilde{c} - \tilde{B} y) \tag{2.4}$$

in the third block equation of (2.3) we obtain

$$\tilde{B}^T \Lambda^{-1}(\tilde{c} - \tilde{B} y) + \hat{B}_n \hat{x}_n = d \tag{2.5}$$

and hence

$$y = \left( \tilde{B}^T \Lambda^{-1} \tilde{B} \right)^{-1} \cdot \left( \tilde{B}^T \Lambda^{-1} \tilde{c} + \hat{B}_n \hat{x}_n - d \right) . \tag{2.6}$$

Finally, using the second block equation of (2.3), namely $\hat{B}_n^T y = \hat{c}_n$, we get an expression for $\hat{x}_n$:

$$\hat{x}_n = \left( \hat{B}_n^T (\tilde{B}^T \Lambda^{-1} \tilde{B})^{-1} \hat{B}_n \right)^{-1} \left[ \hat{c}_n + \hat{B}_n^T (\tilde{B}^T \Lambda^{-1} \tilde{B})^{-1} (d - \tilde{B}^T \Lambda^{-1} \tilde{c}) \right]. \tag{2.7}$$

Note that

$$\tilde{B}^T \Lambda^{-1} \tilde{B} = B^T A^\dagger B, \tag{2.8}$$

where $A^\dagger$ is the pseudoinverse of $A$.

At first sight (2.7) might look hard to compute, but in fact all the quantities involving inverting a matrix appear in the form of a bilinear expression; thus the Lanczos/Stieltjes procedure can be used to evaluate them cheaply—see [23].

The important point is that now the linear system for the rest of the unknowns, namely $\tilde{x}$ and $y$, involves a KKT matrix that has a nonsingular (1,1) block, and one can proceed to solve for these unknowns by using methods that rely on inversion of this block. Analysis of this procedure can be done by means similar to the available error analysis. See, for example, [3, 7, 23]. In a sense, the procedure we have just presented is similar to deflation techniques (see for example [12, 46]), however the difference is that the original linear system is nonsingular; it is *part* of the system, namely the (1,1) block, that is singular, and the deflation is performed for this block, without modifying the space of solutions of the original KKT system.

**2.2. A practical algorithm.** The technique presented in the previous subsection has the drawback of needing the spectrum of the (1,1) block, which in general is prohibitively expensive to compute. We remark at this point that other types of decompositions can be used in the same manner.

In practice, if the rank of $A$ is known, we may be able to repeat a procedure in the same spirit, while avoiding forming the decomposition.

Given $q$, we split the matrix into blocks, the second block-row of which has $q$ rows:

$$\begin{pmatrix} A_{1,1} & A_{1,2} & B_1 \\ A_{1,2}^T & A_{2,2} & B_2 \\ B_1^T & B_2^T & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ y \end{pmatrix} = \begin{pmatrix} c \\ d \\ e \end{pmatrix}. \tag{2.9}$$

By symmetrically permuting the last two block-rows we obtain

$$\left(\begin{array}{cc|c} A_{1,1} & B_1 & A_{1,2} \\ B_1^T & 0 & B_2^T \\ \hline A_{1,2}^T & B_2 & A_{2,2} \end{array}\right) \left(\begin{array}{c} x_1 \\ y \\ \hline x_2 \end{array}\right) = \left(\begin{array}{c} c \\ e \\ \hline d \end{array}\right) \tag{2.10}$$

and hence

$$\begin{pmatrix} x_1 \\ y \end{pmatrix} = \begin{pmatrix} A_{1,1} & B_1 \\ B_1^T & 0 \end{pmatrix}^{-1} \left[ \begin{pmatrix} c \\ e \end{pmatrix} - \begin{pmatrix} A_{1,2} \\ B_2^T \end{pmatrix} x_2 \right]. \tag{2.11}$$

The third block row now reads

$$A_{2,2}x_2 = d - \begin{pmatrix} A_{1,2}^T & B_2 \end{pmatrix} \begin{pmatrix} A_{1,1} & B_1 \\ B_1^T & 0 \end{pmatrix}^{-1} \left[ \begin{pmatrix} c \\ e \end{pmatrix} - \begin{pmatrix} A_{1,2} \\ B_2^T \end{pmatrix} x_2 \right]. \tag{2.12}$$

Thus we have

$$\begin{aligned} x_2 = &\left[ A_{2,2} - \begin{pmatrix} A_{1,2}^T & B_2 \end{pmatrix} \begin{pmatrix} A_{1,1} & B_1 \\ B_1^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} A_{1,2} \\ B_2^T \end{pmatrix} \right]^{-1} \\ &\cdot \left[ d - \begin{pmatrix} A_{1,2}^T & B_2 \end{pmatrix} \begin{pmatrix} A_{1,1} & B_1 \\ B_1^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} c \\ e \end{pmatrix} \right]. \end{aligned} \tag{2.13}$$

Note that we have tacitly assumed that $\begin{pmatrix} A_{1,1} & B_1 \\ B_1^T & 0 \end{pmatrix}$ is nonsingular. If this is not true, we could recover by eliminating a different set of rows and columns.

In certain applications the choice of the rows and columns to be 'deflated' may be clear from the structure of the matrix. For example, in the formulation of three-dimensional magnetostatic problems that is analyzed in [41], the matrix can be introduced in a KKT form, where the (1,1) block is a $3 \times 3$ block matrix whose third block-row and block-column are identically zero; if they are removed, the resulting matix is a KKT matrix whose new (smaller) (1,1) block is known to be symmetric positive definite.

Again, as in (2.7), at first sight (2.13) looks quite difficult to handle; however the size of this linear system is $q \times q$, i.e., very small. If $\mathrm{rank}(A) = n-1$, for example, this is a scalar equation. It is fairly easy to compute $x_2$, as we can use the Lanczos/Stieltjes procedure for evaluating the bilinear forms. Once $x_2$ is computed, the singularity of the (1,1) block is no longer of concern and now computing $x_1$ and $y$ can be done by solving (2.11), using Schur complement techniques.

If the number of Lanczos iterations is small compared to the system size, the cost of this procedure depends linearly on the latter; iterative refinement can be used to improve the accuracy of the computed solution.

**2.3. Example.** Consider the following geophysical inverse problem, whose full description can be found in [25]. Given observations, $d$, of a field $u$, at some discrete locations, $s$, we wish to recover the *model*, $m$. The quantity $m$ could, for example, stand for conductivity. The mathematical equation that describes the connection between $d$ and $u$ is

$$d = u(s) + \epsilon = Qu + \epsilon, \tag{2.14}$$

where $Q$ projects the field $u$ into the measurement locations $s$ and $\epsilon$ is the noise. The constrained problem formulation in [25] is based on the following:

- Require $\|Qu - d\|^2 \leq Tol$.
- $W$ is a weight operator of our choice and we require that $\|Wm\|^2$ be small.
- The forward problem is $A(m)u = f$, where $A$ is a differential operator depending on the specifics of the problem. This equation needs to be solved exactly.

As a result, we get the following optimization problem:

$$\text{minimize} \quad \phi(u,m) = \tfrac{1}{2}\|Qu - b\|^2 + \tfrac{\beta}{2}\|W(m - m_0)\|^2$$

$$\text{subject to} \quad A(m)u = f.$$

The Lagrangian is given by

$$\mathcal{L}(u,m,\lambda) = \frac{1}{2}\|Qu - b\|^2 + \frac{\beta}{2}\|W(m - m_0)\|^2 + \lambda^T[A(m)u - f], \tag{2.15}$$

where $\lambda$ is a vector of Lagrange multipliers and $\beta$ is a regularization parameter. The Hessian is given by

$$H(m,u,\lambda) = \begin{pmatrix} Q^TQ & K(m,\lambda)^T & A(m)^T \\ K(m,\lambda) & \beta W^TW + R(m,u,\lambda) & G(m,u)^T \\ A(m) & G(m,u) & 0 \end{pmatrix}. \tag{2.16}$$

The matrices $G$ and $K$ are associated with the partial derivatives of the matrix $A$, as follows: $G = \frac{\partial(A(m)u)}{\partial m}$; $K = \frac{\partial(A^T\lambda)}{\partial m}$; $R = \frac{\partial(G^T\lambda)}{\partial m}$. The matrix $Q$ is typically very sparse, and $\begin{pmatrix} Q^TQ & K(m,\lambda)^T \\ K(m,\lambda) & \beta W^TW + R(m,u,\lambda) \end{pmatrix}$ is rank deficient.

We consider a specific set-up that comes from electromagnetic simulation of a magnetotelluric survey. The matrix $A$ corresponds to an operator arising from Maxwell's equations. See [25] for full details.

We present below a particular case in which the nullity is 1 and the (1,1) block is semidefinite. The size of the augmented matrix is $136 \times 136$. The size of the (1,1) block is $100 \times 100$. The small size has been picked merely for reasons of computational convenience; the mathematical and numerical properties of this system well represent the situation for larger linear systems for the same model problem. The largest eigenvalue of the (1,1) block is $1.14 \cdot 10^5$. The smallest positive eigenvalue is approximately 0.998. In order to obtain the desired structure of eigenvalues, the diagonal of the original matrix was modified by adding a matrix of the form $\alpha I$. The condition number of the KKT matrix is $1.64 \cdot 10^5$. Table 2.1 shows the increase in accuracy as the number of Lanczos iterations gets larger.

| $k$ | Relative error |
|---|---|
| 5 | $1.02 \cdot 10^{-1}$ |
| 10 | $5.58 \cdot 10^{-2}$ |
| 20 | $6.05 \cdot 10^{-4}$ |

TABLE 2.1

*Relative error for different Lanczos iterations. k stands for the number of Lanczos iterations. The relative error column refers to the measured accuracy of all components of the solution vector.*

**3. An augmented Lagrangian approach.** The approach presented in Section 2 will work well if the nullity of the (1,1) block is low. However, if it is not, then the block-Lanczos procedure employed to estimate a block of unknowns might be costly. In this case we consider the augmented Lagrangian procedure, which is less sensitive to the nullity, and can also work well if there is near singularity.

Let $W$ be a $p \times p$ matrix. Multiplying the second block-row of system (1.1) by $BW$, and adding the resulting equation to the first block equation of the system, we obtain

$$\left\{ \begin{array}{lcl} \left(A + BWB^T\right) x + By & = & c + BWd \\ B^T x & = & d \end{array} \right. . \tag{3.1}$$

The above can be classified as a generalization of the augmented Lagrangian technique. [19, 10] consider the case where the (1,1) block in the original system is positive definite, along with the choice $W = \gamma I$, where $\gamma$ is a positive scalar. In [19] there is a thorough discussion on ways to select the parameter when the Uzawa algorithm is used.

While in the above mentioned references the focus is on a positive definite (1,1) block, in the present work we extend the interest to a singular matrix $A$. We discuss later the fundamental difference between this case and a case where no singularity exists. One immediate benefit of applying the augmented Lagrangian approach in this case is evident: even if the (1,1) block is originally singular, in the modified KKT system (for an appropriate choice of $W$) it is no longer singular. Notice that this approach cannot be considered as regularization: the exact solution does not change.

Selecting $W(\gamma)$ is difficult. Possible criteria can be:

- Scaling based on physical considerations or norms: pick $W$ so as to have a stabilizing effect, by taking into account the type (and order) of the differential operator with which the matrices $A$ and $B$ are associated. For example, in the Navier-Stokes equation, we can say that in the sense of order of associated differential operators, we have $A \approx BB^T$, and so we may select $W = \gamma I$. Generally, we may try to modify $A$ by giving $BB^T$ an 'appropriate' weight. For example, the choice $W = \gamma I$ with $\gamma = \|A\|/\|B\|^2$ has proven to produce excellent results in a large number of cases we have tested.
- Sparsity considerations: since the sparsity pattern of $B$ could be considerably different from that of $A$, it is desirable to aim at the least possible change of the sparsity pattern of the (1,1) block, that would still accomplish the desired effect of turning it into a nonsingular matrix. Here one of the simplest possibilities for choosing $W$ is to have it as a diagonal matrix with 1s and 0s, meaning in practice that we add $\tilde{B}\tilde{B}^T$, where $\tilde{B}$ is part of $B$: certain columns of it.
- Guarantee positive definiteness of the (1,1) block in the spirit of ideas discussed by Hestenes [28, p. 76].

**3.1. Estimates of the condition number.** We start this part of our discussion by providing a connection between the inverses of the original and the modified KKT matrices. The result applies to a large class of matrices, including the unsymmetric case.

THEOREM 3.1. *Suppose that $A$ is a general $n \times n$ matrix, $B$ and $C$ are full column rank $n \times p$ matrices ($p \leq n$), and $W$ is a $p \times p$ matrix. Define*

$$\mathcal{K}(W) := \begin{pmatrix} A + BWC^T & B \\ C^T & 0 \end{pmatrix}. \tag{3.2}$$

*For simplicity of notation, denote $\mathcal{K}(0)$ simply by $\mathcal{K}$. Suppose $\mathcal{K}$ is nonsingular. Then for any $W \neq 0$ such that $\mathcal{K}(W)$ is nonsingular,*

$$\mathcal{K}^{-1}(W) = \mathcal{K}^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & W \end{pmatrix}. \tag{3.3}$$

*Proof.* Write

$$\begin{pmatrix} A + BWC^T & B \\ C^T & 0 \end{pmatrix} \begin{pmatrix} X_W & Y_W \\ Z_W & V_W \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}. \tag{3.4}$$

Specifically, for $W = 0$,

$$\begin{pmatrix} A & B \\ C^T & 0 \end{pmatrix} \begin{pmatrix} X_0 & Y_0 \\ Z_0 & V_0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}. \tag{3.5}$$

From (3.4), the equations for the $(1,1)$ and the $(1,2)$ block-elements of the inverse are

$$(A + BWC^T)X_W + BZ_W = I, \tag{3.6a}$$

$$C^T X_W = 0. \tag{3.6b}$$

Substituting (3.6b) in (3.6a), we get

$$AX_W + BZ_W = I, \tag{3.7}$$

which means that (3.6a) and (3.6b) are identical to the two block equations of (3.5) that determine $X_0$ and $Z_0$. From the uniqueness of the inverse it follows that $X_W = X_0$ and $Z_W = Z_0$.

The equations for the $(1,2)$ and $(2,2)$ blocks of $\mathcal{K}^{-1}(W)$ are given by:

$$(A + BWC^T)Y_W + BV_W = 0, \tag{3.8a}$$

$$C^T Y_W = I. \tag{3.8b}$$

The same equations, applied to $W = 0$, are

$$AY_0 + BV_0 = 0, \tag{3.9a}$$

$$C^T Y_0 = I. \tag{3.9b}$$

Subtracting (3.9a) and (3.9b) from (3.8a) and (3.8b) respectively, it readily follows that

$$\begin{pmatrix} A & B \\ C^T & 0 \end{pmatrix} \begin{pmatrix} Y_W - Y_0 \\ W + V_W - V_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.10}$$

Eq. (3.10) can be thought of as a multiple right-hand-side linear system, with $p$ vectors of unknowns. Since $\begin{pmatrix} A & B \\ C^T & 0 \end{pmatrix}$ is nonsingular, we must have

$$Y_W - Y_0 = 0, \tag{3.11a}$$

$$W + V_W - V_0 = 0, \tag{3.11b}$$

which leads to the desired result. □

Theorem 3.1 allows us to present the following upper bound on the condition number of $\mathcal{K}(W)$.

PROPOSITION 3.2. *The condition number of $\mathcal{K}(W)$ defined in (3.2) satisfies*

$$\kappa\left(\mathcal{K}(W)\right) \leq \kappa\left(\mathcal{K}\right) + \|W\|_2\|B\|_2\|C\|_2 \cdot \left(\|\mathcal{K}^{-1}\|_2 + \frac{\|\mathcal{K}\|_2}{\|B\|_2\|C\|_2} + \|W\|_2\right). \tag{3.12}$$

*Proof.* Clearly,

$$\|\mathcal{K}(W)\|_2 \leq \|\mathcal{K}\|_2 + \|W\|_2\|B\|_2\|C\|_2. \tag{3.13}$$

From Theorem 3.1 it follows that

$$\|\mathcal{K}^{-1}(W)\|_2 \leq \|\mathcal{K}^{-1}\|_2 + \|W\|_2. \tag{3.14}$$

The last two inequalities lead to (3.12). □

An important outcome of Prop. 3.2 is that one can test several choices of $W$ and get a good estimate of the condition number with practically no computational effort. This can be helpful because picking an effective $W$ might require a few iterations.

More can be said if we refer to the case $W = W(\gamma) = \gamma I$. An immediate consequence of Theorem 3.1 is:

COROLLARY 3.3. *Let*

$$\mathcal{K}(\gamma) = \begin{pmatrix} A + \gamma BB^T & B \\ B^T & 0 \end{pmatrix}. \tag{3.15}$$

*Under the assumptions of Theorem 3.1, with $C = B$, we have*

$$\mathcal{K}^{-1} - \mathcal{K}^{-1}(\gamma) = \begin{pmatrix} 0 & 0 \\ 0 & \gamma I \end{pmatrix}. \tag{3.16}$$

COROLLARY 3.4. *For $W(\gamma) = \gamma I$, as $\gamma \to \infty$, the bound given in Proposition 3.2 behaves asymptotically like $\|B\|^2\gamma^2$.*

If the bound is tight, Cor. 3.4 forms a qualitative result that illustrates the rate at which the condition number of the modified augmented matrix behaves in terms of $\gamma$. Figure 3.1 demonstrates this. Here we use randomly generated block-matrices, manipulated so that the (1,1) block is singular. The specifics of the KKT matrix are given in the caption of the figure. For $\gamma$ sufficiently large, the condition number of the modified augmented matrix becomes increasingly large, and indeed tends to $\|B\|^2\gamma^2$. For $\gamma < 1$ the condition number of the matrix hardly changes.

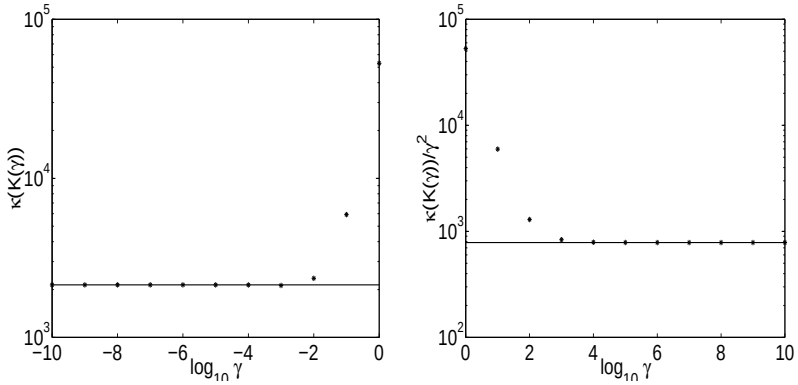



FIG. 3.1. *The condition number of the KKT matrix, for a range of values of $\gamma$. The original matrix in this example, $\mathcal{K} \equiv \mathcal{K}(0)$, is nonsingular and is of size $130 \times 130$. The matrices $A$ and $B$ were randomly generated using the Matlab command 'randn'. Plot (a) shows the condition number without any scaling. The horizontal line is the value of $\kappa(\mathcal{K})$. In plot (b) we depict the condition number divided by $\gamma^2$. The horizontal line for this graph is $\|B\|^2$. As is evident, the condition number divided by $\gamma^2$ for the larger values of $\gamma$ coincides with $\|B\|^2$, which illustrates the tightness and effectiveness of the bound (3.12).*

If the (1,1) block is positive definite, more can be said about the condition number of $\mathcal{K}(\gamma)$, as we discuss below.

PROPOSITION 3.5. *Suppose that $A$ is an $n \times n$ symmetric positive definite matrix and $B$ is an $n \times p$ matrix of full column rank. Let $\Lambda$ be the diagonal matrix whose entries are the generalized eigenvalues of the problem*

$$\lambda A x = B B^T x. \tag{3.17}$$

*Then $\Lambda$ has $p$ nonzeros at most, and there exists an $n \times n$ matrix $G$ such that $A = GG^T$ and $BB^T = G\Lambda G^T$.*

*Proof.* The result can be obtained by employing a similar technique to the fashion in which simultaneous diagonalization of two matrices is obtained: see [22, pp. 461–463] or [47, p. 281]. Let $A = FF^T$. Denote the singular value decomposition of $F^{-1}B$ by

$$F^{-1}B = U\Sigma V^T, \tag{3.18}$$

and define $G := FU$. Note that $U$ is $n \times n$, $\Sigma$ is of size $n \times p$ with its last $n - p$ rows identically zero, and $V$ is $p \times p$. From (3.18) we have that $B = G\Sigma V^T$ and thus $BB^T = G\Sigma\Sigma^T G^T$. Since $U$ is orthogonal, we have

$$A = FF^T = FUU^T F^T = GG^T. \tag{3.19}$$

By (3.18) we have

$$(F^{-1}B)(F^{-1}B)^T = F^{-1}BB^T F^{-T} = U\Sigma\Sigma^T U^T, \tag{3.20}$$

but since the eigenvalues of $F^{-1}BB^T F^{-T}$ are equal to those of $F^{-T}F^{-1}BB^T = A^{-1}BB^T$ (see [29, p. 53] for justification), which are exactly the generalized eigenvalues defined in (3.17), we have $\Lambda = \Sigma\Sigma^T$. Since $\Sigma$ is $n \times p$, the matrix $\Lambda$ cannot have more than $p$ nonzeros. The last statement can also be deduced by observing that $\ker(B^T)$ forms a linear space for the zero generalized eigenvalues—see for example [19, p. 9]. □

Using Prop. 3.5, we have

$$\begin{aligned} \begin{pmatrix} A + \gamma BB^T & B \\ B^T & 0 \end{pmatrix} &= \begin{pmatrix} GG^T + \gamma G\Sigma\Sigma^T G^T & G\Sigma V^T \\ V\Sigma^T G^T & 0 \end{pmatrix} \\ &= \begin{pmatrix} G & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} I + \gamma\Sigma\Sigma^T & \Sigma \\ \Sigma^T & 0 \end{pmatrix} \begin{pmatrix} G^T & 0 \\ 0 & V^T \end{pmatrix}. \end{aligned} \tag{3.21}$$

Since

$$\left\| \begin{pmatrix} G & 0 \\ 0 & V \end{pmatrix} \right\| = \left\| \begin{pmatrix} FU & 0 \\ 0 & V \end{pmatrix} \right\| = \max\{\|FU\|, \|V\|\} = \max\{\|F\|, 1\} \tag{3.22}$$

and similarly for the inverse, it follows that

$$\kappa\left(\mathcal{K}(\gamma)\right) \leq \alpha \cdot \kappa \begin{pmatrix} I + \gamma\Sigma\Sigma^T & \Sigma \\ \Sigma^T & 0 \end{pmatrix}, \tag{3.23}$$

where $\alpha$ depends on $\|F\|$ and $\|F^{-1}\|$, but not on $\gamma$.

The matrix $T(\gamma) := \begin{pmatrix} I + \gamma\Sigma\Sigma^T & \Sigma \\ \Sigma^T & 0 \end{pmatrix}$ has a simple structure, depicted in Figure 3.2(a). On the diagonal, the first $p$ elements are $\{1 + \gamma\sigma_i^2\}, i = 1, \ldots, p$, the next $n - p$ are all equal to 1, and the last $p$ elements on the diagonal are equal to zero. On the $n$th superdiagonal and $n$th subdiagonal, we have $\sigma_i, i = 1, \ldots, p$. Here $\{\sigma_i\}$ refer to the components of the diagonal of the matrix $\Sigma$.

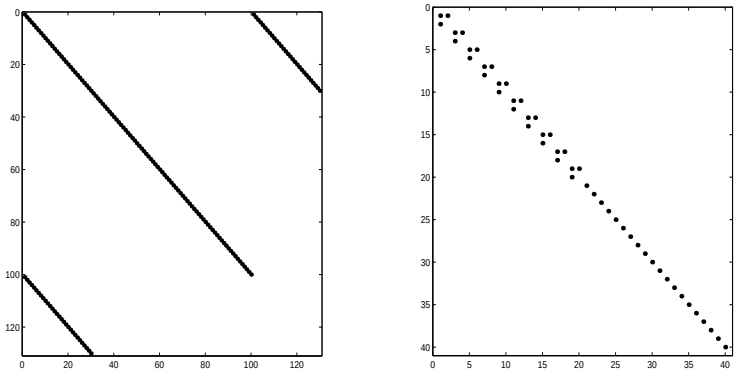
Consider the permutation vector defined by

$$\tilde{p} = [1, n+1, 2, n+2, 3, n+3, 4, n+4, \ldots, m, n+m, m+1 : n]. \tag{3.24}$$

The sparsity structure of the matrix corresponding to the symmetric permutation of $T(\gamma)$ associated with $\tilde{p}$ is depicted in Figure 3.2(b). The permuted matrix is block diagonal with either $2 \times 2$ or $1 \times 1$ blocks. All $1 \times 1$ blocks are equal to 1. The first $p$ $2 \times 2$ blocks are of the form

$$\operatorname{diag} \begin{pmatrix} 1 + \gamma\lambda_i & \sqrt{\lambda_i} \\ \sqrt{\lambda_i} & 0 \end{pmatrix}, \tag{3.25}$$

where $\lambda_i$ are the generalized eigenvalues (3.17).



(a) Original : T (b) Permuted : $\tilde{p}^{T}T\tilde{p}$

FIG. 3.2. *Sparsity patterns of $T(\gamma)$ and its symmetric permutation.*

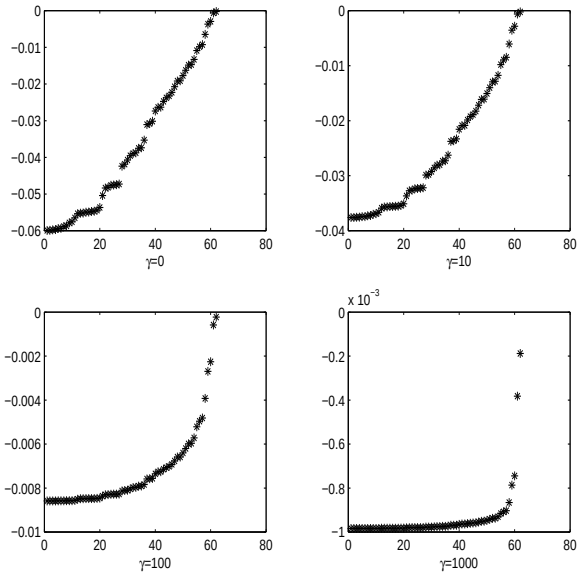



FIG. 3.3. *The negative eigenvalues of the Stokes matrix, for a $16 \times 16$ grid, as $\gamma$ varies. The matrix was created using Elman & Silvester's FEM code [18]. The effect of clustering as $\gamma$ grows larger is evident: the negative eigenvalues tend to $-\frac{1}{\gamma}$.*

The dependence of the condition number of the modified augmented matrix on the parameter $\gamma$ boils down to the mathematical problem of determining the condition number of a block diagonal matrix with $1 \times 1$ or $2 \times 2$ blocks.

The eigenvalues of a $2 \times 2$ symmetric matrix of the form

$$\begin{pmatrix} 1+\gamma\sigma^2 & \sigma \\ \sigma & 0 \end{pmatrix} \tag{3.26}$$

are given by

$$\lambda(\gamma) = \frac{1+\gamma\sigma^2 \pm \sqrt{(1+\gamma\sigma^2)^2+4\sigma^2}}{2}. \tag{3.27}$$

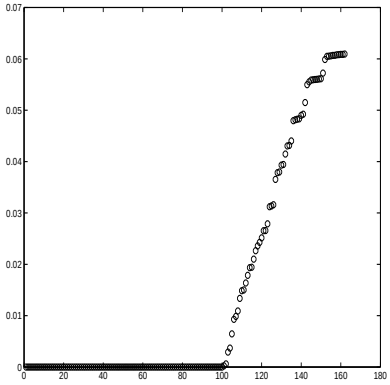



FIG. 3.4. *The generalized eigenvalues $\lambda Ax = BB^T x$ for the Stokes problem, for the same grid as in Fig. 3.3.*

For $\gamma$ sufficiently large, we have

$$\frac{2\sigma}{1+\gamma\sigma^2} \ll 1, \tag{3.28}$$

which allows us to use the first order Taylor expansion $\sqrt{1+x} \approx 1 + \frac{x}{2}$ (valid when $x$ is sufficiently small). Hence

$$\lambda(\gamma) \approx \frac{1+\gamma\sigma^2}{2} \cdot \left(1 \pm \left(1 + \frac{2\sigma^2}{(1+\gamma\sigma^2)^2}\right)\right). \tag{3.29}$$

It should be stressed that the condition number of the block diagonal matrix (3.25) is a fraction of two eigenvalues (the maximal and the minimal in absolute value) which do *not* belong to the same $2 \times 2$ block. Some elementary calculus shows that the expression for the eigenvalues with the positive sign is monotonically increasing as a function of $\sigma$. From this we can deduce that the maximal positive eigenvalue is

$$\lambda_{\max}(\gamma) = \frac{1+\gamma\sigma_M^2 + \sqrt{(1+\gamma\sigma_M^2)^2 + 4\sigma_M^2}}{2}, \tag{3.30}$$

where $\sigma_M$ is the maximal singular value of $F^{-1}B$. Using Taylor expansion, this simplifies to

$$\lambda_{\max}(\gamma) \approx 1 + \gamma\sigma_M^2 + \frac{\sigma_M^2}{1+\gamma\sigma_M^2}. \tag{3.31}$$

A more delicate situation occurs for the minimal eigenvalue. Here we have to determine whether the minimum in absolute value occurs for the smallest *positive* eigenvalue, or for the negative eigenvalue closest to zero. The smallest positive eigenvalue is approximately given by $1 + \gamma\sigma_m^2 + \frac{\sigma_m^2}{1+\gamma\sigma_m^2}$, with $\sigma_m$ denoting the minimal singular value of $F^{-1}B$. On the other hand, for the negative eigenvalues we have, by (3.29), expressions of the form $-\frac{\sigma^2}{1+\gamma\sigma^2}$. Again, by calculus considerations, this is a monotonic function in $\sigma$. Clearly, the magnitude of the negative eigenvalue for $\sigma = \sigma_m$ is smaller than the smallest positive eigenvalue. We thus have

$$\min_i |\lambda_i(\gamma)| \approx \frac{\sigma_m^2}{1+\gamma\sigma_m^2}. \tag{3.32}$$

We can summarize everything in a lemma.

LEMMA 3.6. *For $\gamma$ sufficiently large, the condition number of $T(\gamma)$ satisfies*

$$\kappa\left(T(\gamma)\right) \approx \frac{(1+\gamma\sigma_m^2)\left((1+\gamma\sigma_M^2)^2+\sigma_M^2\right)}{\sigma_m^2(1+\gamma\sigma_M^2)}. \tag{3.33}$$

*As $\gamma \to \infty$ we have*

$$\frac{\kappa\left(T(\gamma)\right)}{\gamma^2} \to C \neq 0. \tag{3.34}$$

*In addition, the $p$ negative eigenvalues of $T(\gamma)$ tend to $-\frac{1}{\gamma}$.*

Hence a Krylov solver will converge typically within $p$ fewer iterations for the modified system, compared to solving the original system, however accuracy might be compromised. The strong connection between $T(\gamma)$ and $\mathcal{K}(\gamma)$, given in (3.21), leads us to believe that similar effects occur for both, as $\gamma$ changes. Figure 3.3 demonstrates the clustering effect of the negative eigenvalues as $\gamma$ gets large, for Elman & Silvester's discretization of the Stokes problem [18]. Figure 3.4 depicts the generalized eigenvalues of this problem.

If $A$ is singular, we cannot obtain (3.23) for obvious reasons, however we can still obtain a decomposition which is similar to (3.21).

LEMMA 3.7. *Suppose $C = \alpha A + \beta BB^T$ is positive definite. Then there exists a nonsingular matrix $X$ such that both $D_A := X^T A X$ and $D_B := X^T(BB^T)X$ are diagonal.*

*Proof.* Let $C = FF^T$. Then

$$I = \alpha F^{-1} A F^{-T} + \beta F^{-1} BB^T F^{-T}. \tag{3.35}$$

Let $F^{-1}AF^{-T} = QD_AQ^T$, where $D_A$ is diagonal and $Q$ is orthogonal, and define

$$X = F^{-T}Q. \tag{3.36}$$

Clearly, $X^TAX = D_A$. Now,

$$\begin{aligned} \beta X^T(BB^T)X &= \beta Q^T F^{-1} BB^T F^{-T} Q \\ &= Q^T\left(I - \alpha F^{-1} A F^{-T}\right) Q \\ &= I - \alpha D_A = \beta D_B. \end{aligned}$$

□

An important case is when $A$ is positive semidefinite.

PROPOSITION 3.8. *If $A$ is positive semidefinite and $\ker(A) \cap \ker(B^T) = \{0\}$, the matrix $A + \gamma BB^T$ is symmetric positive definite for any $\gamma > 0$.*

*Proof.* The result can be shown by observing that any vector $x$ such that $x^TAx = 0$ cannot satisfy $x^TBB^Tx$ due to the condition that the kernels of $A$ and $B^T$ do not intersect. Alternatively, an SVD argument can be used, as follows. Let $B^T = U\Sigma V^T$ be the SVD decomposition of $B^T$. Since $B^T$ is of rank $p$, we have $B^T\left(v_{p+1}, \ldots, v_n\right) = 0$. On the other hand, $BB^T = V\Sigma^T\Sigma V^T$, thus $BB^T\left(v_{p+1}, \ldots, v_n\right) = 0$. Since $\ker(A)$ and $\ker(B^T)$ do not intersect, the desired result readily follows. □

We can now rely on the nonnegativity of $D_A$. We have $A = GG^T$, as in (3.19), but now $G = X^{-T}D_A^{1/2}$. By repeating the same steps, it is straightforward to obtain a decomposition similar to (3.21), and derive further estimates of the condition number of the matrix.

**3.2. Convergence.** For the augmented Lagrangian approach, in addition to the usual issues that arise in performing convergence analysis, we have to address the question of a choice for the parameter $\gamma$ that will guarantee convergence.

We choose to focus on the Uzawa algorithm [2], which is quite popular and is based on constructing a sequence of approximations to $x$ and $y$ (defined in (1.1)), as follows:

For $k = 0, 1, \ldots$
Solve $Ax_{k+1} = c - By_k$
Compute $y_{k+1} = y_k + \alpha(B^T x_{k+1} - d)$

Convergence analysis shows that when $A$ is positive definite, the optimal Uzawa parameter, $\alpha$, is given by

$$\alpha_{opt} = \frac{2}{\lambda_{min}(S) + \lambda_{max}(S)}, \tag{3.37}$$

where $S$ is the Schur complement $B^T A^{-1} B$ [16]. Another important observation is that in general, there is no need to solve *exactly* the linear system involving $A$ in each iteration. See any of [16, 4, 9] for ways of performing the inner iteration. See also [34] for further references.

If the matrix $A$ is replaced by $A + \gamma BB^T$, we have to determine *two* parameters, namely $\gamma$ and $\alpha$. While the convergence of the outer iteration depends on properties of the Schur complement, in each iteration a linear system involving $A + \gamma BB^T$ must be solved. Thus there is a delicate balance between the spectral properties of the Schur complement and those of the (1,1) block, and our aim is to find a value of $\gamma$ that is 'good' in both aspects.

Fortin & Glowinski performed a comprehensive analysis for the case of $A$ positive definite, and discussed at length the optimal Uzawa parameter. The interested reader is referred to [19], particularly Chapter 1 therein. Here we take a different mathematical approach, and add a few observations for the case of a semidefinite (1,1) block.

Since the spectrum of the Schur complement is fundamental to the convergence behavior of the Uzawa scheme, we start by finding a connection between $S = B^T A^{-1} B$ and the set of matrices

$$S(\gamma) := B^T (A + \gamma BB^T)^{-1} B, \tag{3.38}$$

in the case $A$ is nonsingular.

PROPOSITION 3.9. *Suppose $A$ is nonsingular, and let $S(\gamma)$ be as defined in (3.38). For notational convenience, denote $S(0) \equiv B^T A^{-1} B$ simply by $S$. Then*

$$S(\gamma) = S \cdot (I + \gamma S)^{-1}, \tag{3.39}$$

*and consequently, if the eigenvalues of $S$ are $\{\mu_i\}$, the eigenvalues of $S(\gamma)$ are*

$$\mu_i(\gamma) = \frac{\mu_i}{1 + \gamma \mu_i}. \tag{3.40}$$

*Proof.* The Sherman-Morrison-Woodbury formula (see, e.g. [22, p. 50]) states that for an $n \times n$ nonsingular matrix $A$ and $n \times k$ matrices $U$ and $V$ the following relation holds (provided that $I + V^T A^{-1} U$ is nonsingular):

$$(A + UV^T)^{-1} = A^{-1} - A^{-1} U (I + V^T A^{-1} U)^{-1} V^T A^{-1}. \tag{3.41}$$

For $U = V = \sqrt{\gamma}B$ we have

$$\begin{aligned} B^T(A+\gamma BB^T)^{-1}B &= B^T\left(A^{-1} - \gamma A^{-1}B(I+\gamma B^TA^{-1}B)^{-1}B^TA^{-1}\right)B \\ &= B^TA^{-1}B - \gamma B^TA^{-1}B\left(I+\gamma B^TA^{-1}B\right)^{-1}B^TA^{-1}B \\ &= B^TA^{-1}B\cdot\left[I-\gamma(I+\gamma B^TA^{-1}B)^{-1})(B^TA^{-1}B)\right] \\ &= S\cdot\left(I-\gamma(I+\gamma S)^{-1}S\right). \end{aligned}$$

It is straightforward to verify that $I-\gamma(I+\gamma S)^{-1}S = (I+\gamma S)^{-1}$. ∎

We can now determine the circumstances in which the Uzawa algorithm converges. From [16] we know that in order for the scheme to converge for any initial guess, $\alpha$ must be such that $\rho\left(I-\alpha S(\gamma)\right) < 1$. (As usual, $\rho$ denotes the spectral radius of a matrix [50, p. 9].) From Proposition 3.9 it follows that the matrix $I-\alpha S(\gamma)$ has the eigenvalues $\dfrac{1+(\gamma-\alpha)\mu_i}{1+\gamma\mu_i}$. Suppose that $\gamma > 0$. If $A$ is positive definite, then so is $B^TA^{-1}B$ (provided that $B$ is of full column rank, which we assume throughout this report). We thus have $\mu_i > 0$ for all $i$, hence $1+\gamma\mu_i > 0$, and imposing

$$-1 < \frac{1+(\gamma-\alpha)\mu_i}{1+\gamma\mu_i} < 1 \tag{3.42}$$

is equivalent to $0 < \alpha < 2\gamma + \frac{2}{\mu_i}$. This will hold if

$$0 < \alpha < 2\gamma + \frac{2}{\mu_{max}}, \tag{3.43}$$

and is satisfied for any arbitrary set of eigenvalues $\{\mu_i\}$ if $0 < \alpha \leq 2\gamma$. This result is identical to the result in [19], though obtained by different means.

If $A$ is positive semidefinite, Prop. 3.9 does not apply directly, however we can obtain the eigenvalues of $S(\gamma)$ using a continuity argument, by taking $\delta$ small enough so that (3.43) is satisfied. $A+\delta BB^T$ is positive definite by Prop. 3.8 and hence Prop. 3.9 can be straightforwardly applied to this matrix. It readily follows that the condition for convergence in this case is $0 < \alpha < 2\gamma$.

We remark that a mathematical model that results in a KKT system with a semidefinite (1,1) block might turn out to be a linear system with a numerically indefinite (1,1) block when the computation is carried out. However, it can be shown that a small negative eigenvalue of the matrix $A$ is mapped into an eigenvalue of $B^T(A+\gamma BB^T)^{-1}B$ that is nearly equal to $1/\gamma$, and in general the numerical indefiniteness will manifest itself by the need to replace the condition $0 < \alpha < 2\gamma$ by a condition $0 < \alpha < 2\gamma - \varepsilon$, with $\varepsilon$ being of magnitude equal to the magnitude of the small negative eigenvalues of $A$.

We can now summarize all that has been said above as follows.

THEOREM 3.10. *Consider the Uzawa algorithm applied to the modified KKT system, i.e., to a KKT system in which the (1,1) block is $A+\gamma BB^T$. If $A$ is positive semidefinite, any choice of $\gamma > 0$ with $0 < \alpha < 2\gamma$ converges.*

**3.3. Practical choice of $\gamma$.** A question that has still not been answered is how in practice to pick the value of $\gamma$ or the weight matrix $W$. At least for the case $W(\gamma) = \gamma I$, it turns out that if the (1,1) block is singular, there may be a particular choice of $\gamma$ that leads to faster convergence. This is illustrated in Figure 3.5. Here we have a semidefinite (1,1) block based on an originally randomly generated matrix. In the graph, the value of $\gamma$ ranges from $10^{-10}$ to $10^{10}$. It is evident that there

exists a certain range of values of $\gamma$, around $\gamma = 0.1$, for which all three condition numbers, namely those of the (1,1) block, the Schur complement, and the whole KKT matrix, are relatively small. It is important to note that in this example, the value of $\|A\|/\|B\|^2$ is approximately 0.25; thus this is an excellent choice for the parameter $\gamma$.

If the (1,1) block is singular or nearly singular, then in many cases an appropriate choice of $\gamma$ could make the matrix $A + \gamma BB^T$ better-conditioned than $A$. Note that $\gamma$ cannot be too large because then the singular matrix $BB^T$ would play a dominant role and may cause ill-conditioning. Therefore we should seek a value of $\gamma$ that is large enough so as to eliminate the effect of the ill-conditioning of $A$, and is not too large, so as to avoid the effect of the ill-conditioning of $BB^T$.

If $A$ is well-conditioned, the choice of a value of $\gamma$ that is significantly better than other choices may be less obvious or may not exist. This illustrates a difference between the case of a singular (1,1) block and the case of a nonsingular one.

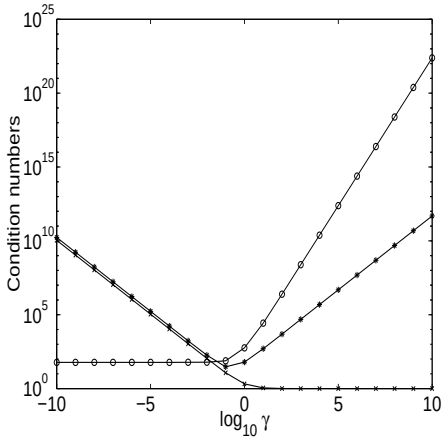



FIG. 3.5. *Condition numbers of the (1,1) block (curves with '*'), the whole KKT matrix (curves with 'o'), and the Schur complement (curves with 'x'), as a function of $\gamma$. In this example, the matrix is positive semidefinite. It was generated randomly, and then modified so as to obtain symmetry and semidefiniteness.*

In Fig. 3.6 we present the error for the nonlinear inverse problem presented in Section 2. The graph is a plot of the actual relative error, when the (exact) Uzawa algorithm was applied. Here we fix the number of iterations, and then check the norm of the relative error. Numerical computations of the condition numbers of the KKT matrix, the (1,1) block and the Schur complement confirm that there is excellent correspondence between the expected 'good' values of $\gamma$, and the actual error. The value of $\frac{\|A\|}{\|B\|^2}$, which we believe should be used as the choice for $\gamma$ when no other information is available, is equal to $3.03 \cdot 10^{-4}$. It is clear from the graph that the performance of the solver for this choice is very good. Fig. 3.7 shows how the structure of the singular values of the matrix changes. Notice the cluster of singular values at $\frac{1}{\gamma}$.

**3.4. Regularization.** Regularizing the system may remedy potential accuracy problems. Common regularization techniques are based on modifying the (2,2) block (which was originally zero) so as to obtain a better conditioned matrix. See for example [39, chap. 17] and [14].

We limit our discussion below to the case of $d \equiv 0$ in the system (1.1). Consider

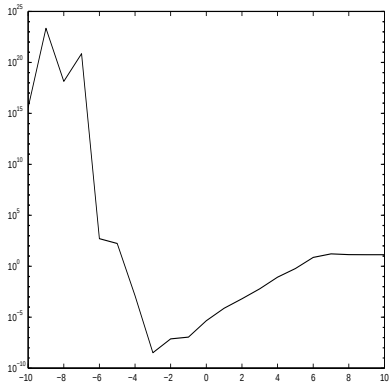


FIG. 3.6. *The relative error for the inverse problem introduced in Section 2. The problem size is* $763 \times 763$*, and we plot the error after 100 Uzawa iterations, for various values of* $\gamma$*: the* $x$*-axis represents* $\log_{10}\gamma$*, and the* $y$*-axis is the 2-norm of the relative error.*

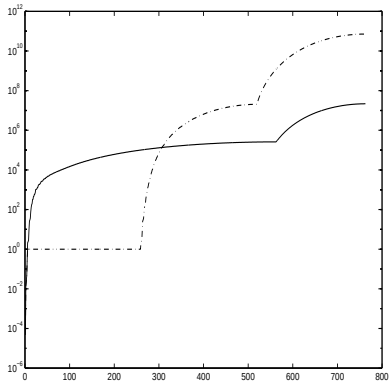


FIG. 3.7. *The singular values for the model problem, for the same set-up as in Fig. 3.6. The dashed curve corresponds to the singular values of the* modified *KKT system, with* $\gamma = 1$*. Here* $p = 254$*, and it is evident that* $p$ *of the singular values tend to* $|\frac{1}{\gamma}|$*. The solid line corresponds to the singular values of the original KKT matrix.*

the linear system

$$\mathcal{K}u \equiv \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ 0 \end{pmatrix}. \tag{3.44}$$

The (1,1) block $A$ is assumed to be positive definite. (Note that even if $A$ is originally singular, $A + \gamma BB^T$ can in certain cases transform it to a positive definite matrix, and the discussion below applies, with $A$ replacing $A + \gamma BB^T$.)

Let us examine the effect of regularizing (3.44) by modifying it to be a parameter-dependent system of the form

$$\mathcal{K}_\mu u_\mu \equiv \begin{pmatrix} A & B \\ B^T & -\mu I \end{pmatrix} \begin{pmatrix} x_\mu \\ y_\mu \end{pmatrix} = \begin{pmatrix} c \\ 0 \end{pmatrix}. \tag{3.45}$$

Below we are mainly interested in the computed error for the vector $x$. Denote the (negative) Schur complement of the regularized matrix by

$$S_\mu = \mu I + B^T A^{-1} B. \tag{3.46}$$

Clearly, $S_0$ is the Schur complement of $\mathcal{K}$. We start by expressing the solution of the unregularized system (3.44) in terms of the matrices involved. Suppose that

$$\mathcal{K}^{-1} = \begin{pmatrix} X & Y \\ Y^T & Z \end{pmatrix}. \tag{3.47}$$

It is straightforward to verify that

$$X = A^{-1} - A^{-1}B(B^TA^{-1}B)^{-1}B^TA^{-1} = A^{-1} - A^{-1}BS_0^{-1}B^TA^{-1}. \tag{3.48}$$

Since $A$ is symmetric and positive definite, by observing that

$$P := A^{-1/2}B(B^TA^{-1}B)^{-1}B^TA^{-1/2} \tag{3.49}$$

is a *projection matrix*, we can rewrite $X$ in a simpler form:

$$X = A^{-1/2}(I - P^2)A^{-1/2}. \tag{3.50}$$

$x$ can be expressed in a form which clarifies how far the unconstrained solution (namely that $A^{-1}c$) is from the constrained one:

$$x = A^{-1}c - A^{-1}BS_0^{-1}B^TA^{-1}c. \tag{3.51}$$

From the last two equations it follows that

$$\|x\| \leq \|A^{-1}\| \cdot \|c\|. \tag{3.52}$$

Also,

$$\|c - A^{-1}BS_0^{-1}B^TA^{-1}c\| \leq \|A\|\|x\|. \tag{3.53}$$

Consider now the regularized linear system (3.45). Let

$$\mathcal{K}_\mu^{-1} = \begin{pmatrix} X_\mu & Y_\mu \\ Y_\mu^T & Z_\mu \end{pmatrix}. \tag{3.54}$$

PROPOSITION 3.11. *The* $(1,1)$ *block of* $\mathcal{K}_\mu^{-1}$ *is given by*

$$X_\mu = A^{-1} - A^{-1}BS_\mu^{-1}B^TA^{-1}, \tag{3.55}$$

*or by the simpler expression*

$$X_\mu = (A + \frac{1}{\mu}BB^T)^{-1}. \tag{3.56}$$

*Proof.* Write

$$\begin{pmatrix} A & B \\ B^T & -\mu I \end{pmatrix}\begin{pmatrix} X_\mu & Y_\mu \\ Y_\mu^T & Z_\mu \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}, \tag{3.57}$$

and write down the equations for the $(1,1)$ and $(1,2)$ blocks:

$$\begin{cases} AX_\mu + BY_\mu^T = I; \\ B^TX_\mu - \mu Y_\mu^T = 0. \end{cases} \tag{3.58}$$

By substituting $X_\mu = A^{-1} - A^{-1}BY_\mu^T$ in the bottom equation of (3.58) we get

$$Y_\mu^T = S_\mu^{-1} B^T A^{-1}, \tag{3.59}$$

from which (3.55) follows.

Alternatively, we can proceed from (3.58) by substituting

$$Y_\mu^T = \frac{1}{\mu} B^T X_\mu \tag{3.60}$$

in the top equation of (3.58), and (3.56) readily follows. □

We note that interestingly, (3.56) is not defined for $\mu = 0$, yet the limit exists. The error in the computed solution is given by

$$u_\mu - u_0 = (X_\mu - X)c. \tag{3.61}$$

From Proposition 3.11 it follows that

$$\begin{aligned} X_\mu - X &= (A^{-1} - A^{-1}BS_\mu^{-1}B^TA^{-1}) - (A^{-1} - A^{-1}BS_0^{-1}B^TA^{-1}) \\ &= A^{-1}B\left(S_0^{-1} - S_\mu^{-1}\right)B^TA^{-1}. \end{aligned}$$

In general, for any two nonsingular matrices $M$ and $N$,

$$M^{-1} - N^{-1} = M^{-1}(N - M)N^{-1}; \tag{3.62}$$

hence

$$S_0^{-1} - S_\mu^{-1} = \mu S_0^{-1} S_\mu^{-1}, \tag{3.63}$$

and

$$X_\mu - X = -\mu A^{-1}BS_\mu^{-1}S_0^{-1}B^TA^{-1}. \tag{3.64}$$

Therefore

$$x_\mu - x = -\mu A^{-1}BS_\mu^{-1}S_0^{-1}B^TA^{-1}c, \tag{3.65}$$

and the norm of the error is bounded by

$$\|x_\mu - x\| = \|\mu A^{-1}BS_\mu^{-1}S_0^{-1}B^TA^{-1}c\| \le \|\mu S_\mu^{-1}\| \cdot \|B^TA^{-1}\|^2 \|S_0^{-1}\| \cdot \|c\|. \tag{3.66}$$

Using (3.53) we can also get an expression for the *relative* error. All the quantities in either the absolute error or the relative error, except $\|\mu S_\mu^{-1}\|$, depend on the given problem, not on the regularization parameter. The latter can be determined so as to make the error as small as desired.

PROPOSITION 3.12. *$\|\mu S_\mu^{-1}\|$ tends to 0 for $\mu \ll 1$, is monotonically increasing as $\mu$ grows larger, and is bounded by 1.*

*Proof.* Recall that $S_\mu = \mu I + B^TA^{-1}B$. Therefore,

$$\mu S_\mu^{-1} = \left(I + \frac{1}{\mu}S_0\right)^{-1}. \tag{3.67}$$

Thus $\mu S_\mu^{-1}$ is positive definite, and

$$\|\mu S_\mu^{-1}\|_2 = \frac{1}{\lambda_{min}(I + \frac{1}{\mu}S_0)} = \frac{\mu}{\mu + \lambda_{min}(S_0)}. \tag{3.68}$$

The latter is a monotonic function in $\mu$, which tends to 0 when $\mu \to 0$, and tends to 1 as $\mu \to \infty$. □

It is worthwhile noting that the function (3.68) is very steep near $\mu = 0$, which is important when determining the range of values of $\mu$ for which the regularization leads to a sufficiently accurate solution. Indeed, if the bound is tight, unless $\mu$ is extremely small there would be a noticeable perturbation in $x$. A possible way to overcome this difficulty is by performing iterative refinement on the system (3.44) by repeatedly solving the better conditioned system (3.45).

**4. A class of parameter-dependent preconditioners.** So far we have not considered techniques related to solving for the whole KKT matrix directly, without separating $x$ from $y$. An appropriate Krylov solver would be, for example, the MINRES algorithm [40]. For indefinite systems, it is widely agreed that a good preconditioner is a necessity [44, 51]. Important results concerning the spectrum of KKT matrices, which may be helpful in understanding the convergence properties of Krylov solvers, can be found in Wathen, Fischer and Silvester [51].

Murphy, Golub and Wathen showed in [38] that if $A$ is nonsingular, then the preconditioner

$$\mathcal{M} = \begin{pmatrix} A & 0 \\ 0 & B^T A^{-1} B \end{pmatrix} \tag{4.1}$$

has the attractive property that the associated preconditioned matrix $\mathcal{M}^{-1}\mathcal{K}$ has at most four distinct eigenvalues: $0, 1, \frac{1}{2} \pm \frac{\sqrt{5}}{2}$. Hence a minimum residual Krylov solver will terminate within four iterations. However, knowing the Schur complement exactly amounts to computing the solution of the linear system and in practice, an effort is made to approximate the Schur complement. See for example [37]. Block preconditioners have also been considered—see for example Klawonn [33],[32], Battermann & Heinkenschloss [5], Elman & Silvester [17].

We are interested in generalizing the preconditioner (4.1) in a manner that allows for application to KKT systems with a singular (1,1) block. We note that one obvious way to do it is to change the (1,1) block in both the matrix and the preconditioner from $A$ to $A + \gamma BB^T$, and use the preconditioner of (4.1), along with replacing the Schur complement by an approximation that allows for reduced computational costs.

Here we consider preconditioning matrices of the form

$$\mathcal{M} = \begin{pmatrix} M(\gamma) & 0 \\ 0 & B^T N(\gamma) B \end{pmatrix}, \tag{4.2}$$

where $M$ and $N$ are to be determined, and $\gamma$ is a positive real scalar.

Let $\lambda$ be an eigenvalue of the preconditioned matrix $\mathcal{M}^{-1}\mathcal{K}$, whose associated eigenvector is $\begin{pmatrix} u \\ v \end{pmatrix}$. Then

$$\begin{cases} Au + Bv &= \lambda M u \\ B^T u &= \lambda B^T N B v \end{cases}. \tag{4.3}$$

Substituting

$$v = \frac{1}{\lambda}(B^T N B)^{-1} B^T u, \tag{4.4}$$

we have a quadratic expression (in $\lambda$) for $u$:

$$\left[\lambda^2 M - \lambda A - B(B^T N B)^{-1} B^T\right] u = 0. \tag{4.5}$$

In the sequel we consider the specific case $M = N^{-1}$. Denote $w = M^{\frac{1}{2}} u$. Then

$$\left[\lambda^2 I - \lambda M^{-\frac{1}{2}} A M^{-\frac{1}{2}} - M^{-\frac{1}{2}} B (B^T M^{-1} B)^{-1} B^T M^{-\frac{1}{2}}\right] w = 0. \tag{4.6}$$

The matrix $P = M^{-\frac{1}{2}} B (B^T M^{-1} B)^{-1} B^T M^{-\frac{1}{2}}$ is an orthogonal projector onto range$(M^{-1/2} B)$. This can be seen by denoting $C = M^{-\frac{1}{2}} B$ and observing that $P = C(C^T C)^{-1} C^T$. Denoting $K = M^{-1/2} A M^{-1/2}$, we now have a quadratic eigenvalue problem of the form

$$(\lambda^2 I - \lambda K - P) z = 0. \tag{4.7}$$

Consider

$$\mathcal{M} = \begin{pmatrix} A + \gamma B B^T & 0 \\ 0 & B^T (A + \gamma B B^T)^{-1} B \end{pmatrix}. \tag{4.8}$$

LEMMA 4.1. *For the matrix $M$ implied by (4.8), namely $M = A + \gamma B B^T$, the matrices $P$ and $K$ share the same eigenvectors.*

*Proof.* Suppose $\mu_i$ is an eigenvalue of $K$, associated with the eigenvector $\mathbf{x}_i$. Then $\mu_i$ is also an eigenvalue of $M^{-1} A$, associated with the eigenvector $\mathbf{z}_i = M^{-1/2} \mathbf{x}_i$, and we have

$$A \mathbf{z}_i = \mu_i (A + \gamma B B^T) \mathbf{z}_i. \tag{4.9}$$

If $A$ is singular, then the matrix $M^{-1} A$ has rank(ker$(A)$) zero eigenvalues, and the corresponding eigenspace is span(ker$(A)$). On the other hand, for $\mu_i \neq 0$ we have

$$\frac{1 - \mu_i}{\gamma \mu_i} A \mathbf{z}_i = B B^T \mathbf{z}_i, \tag{4.10}$$

from which it follows that $\mathbf{z}_i$ are exactly the eigenvectors of the generalized eigenvalue problem

$$s_i A \mathbf{z}_i = B B^T \mathbf{z}_i. \tag{4.11}$$

We now turn to examine the projector $P$. We know that $P$ projects onto range$(M^{-1/2} B)$. Consider $R := (M^{-1/2} B)(M^{-1/2} B)^T = M^{-1/2} B B^T M^{-1/2}$. Using an SVD argument identical to the one used in the proof of Proposition 3.8, range$(R)$ spans the subspace that $P$ projects onto.
$R$ has the same eigenvalues that $M^{-1} B B^T$ has, and if $\tilde{x}_i$ are the eigenvectors of the former, then $\tilde{z}_i = M^{-1/2} \tilde{x}_i$ are the eigenvectors of the latter. If $\beta_i$ are the associated eigenvalues, then

$$\beta_i (A + \gamma B B^T) \tilde{z}_i = B B^T \tilde{z}_i, \tag{4.12}$$

from which it follows that

$$\frac{\beta_i}{1 - \beta_i \gamma} A \tilde{z}_i = B B^T \tilde{z}_i. \tag{4.13}$$

It is now evident that in this case too, as for the matrix $K$, the eigenspace is exactly the space spanned by the generalized eigenvectors $\mathbf{z}_i$. That is, $\tilde{z}_i = \mathbf{z}_i$ and hence $\tilde{x}_i = \mathbf{x}_i$. □

The result stated in Lemma 4.1 simplifies our problem a great deal. The $n \times n$ matrix $P$ has a zero eigenvalue of algebraic multiplicity $n-p$, and an eigenvalue equal to 1, of algebraic multiplicity $p$. The space spanned by the eigenvectors associated with the multiple zero eigenvalue of $P$ is $\ker(B^T)$.

On the other hand, we can say more about the eigenvalues of $K$. From (4.10) and (4.11) it follows that

$$\frac{1-\mu_i}{\gamma\mu_i} = s_i. \tag{4.14}$$

From [19, p. 9] we know the following.

PROPOSITION 4.2. *The space spanned by the eigenvectors associated with the zero eigenvalues of the generalized eigenvalue problem (4.11) is ker$(B^T)$, and its dimension is $n-p$.*

Using Prop. 4.2, from (4.14) we deduce that $K$ has an eigenvalue $\mu_i = 1$ of multiplicity $n-p$, and its associated space of eigenvectors is $\ker(B^T)$.

Since $\ker(B^T)$ is the space of eigenvectors associated with the zero eigenvalues of $P$, we get that $n-p$ eigenvalues of the preconditioned matrix $\mathcal{M}^{-1}\mathcal{K}$ satisfy

$$\lambda^2 - \lambda = 0. \tag{4.15}$$

Since we are interested in a nonsingular preconditioned matrix, we have $\lambda = 1$.

For the rest of the eigenvalues, we have the connection

$$\lambda^2 - \lambda\mu = 1, \tag{4.16}$$

which leads to

$$\lambda = \frac{1}{2} \cdot (\mu_i \pm \sqrt{\mu_i^2 + 4}). \tag{4.17}$$

The zero eigenvalues of $A$ are mapped into $\lambda = \pm 1$. On the other hand, from the relationship (4.14) we can deduce that

$$1 \leq \frac{1}{2}(\mu + \sqrt{\mu^2+4}) \leq \frac{1}{2}(1+\sqrt{5}) \approx 1.618, \tag{4.18}$$

and

$$-0.618 \approx \frac{1}{2}(1-\sqrt{5}) \geq \frac{1}{2}(\mu - \sqrt{\mu^2+4}) \geq -1. \tag{4.19}$$

We conclude, then, that all the eigenvalues of the preconditioned matrix are within two narrow intervals whose ends do not depend on any spectral properties of the particular blocks of the KKT matrix. This is in contrast to the eigenvalues of the latter - see Rusten and Winther [43, Lemma 2.1]. Finally, the connection between the eigenvalues of $K$ and the generalized eigenvalues $s_i$, given in (4.14), gives us the eigenvalues of the preconditioned matrix in terms of the latter.

It should be noted that the construction of this preconditioner in its current form is costly for large problems. It is therefore necessary to find efficient approximations

to the shifted Schur complement in the (2,2) block of (4.2). The theoretical foundation presented above is nevertheless useful for illustrating the spectral properties of preconditioners of this form.

In Fig. 4.1 we illustrate the performance of the proposed preconditioner. For reasons of convenience, MATLAB's QMR (which is usually applied to nonsymmetric systems) was used. The high effectiveness of the preconditioner is evident.

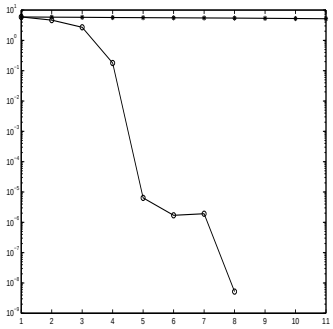


FIG. 4.1. *Convergence after preconditioning, for the inverse problem. The problem size is* $763 \times 763$. *The upper curve is the relative residual without preconditioning. The lower curve corresponds to the residual when our preconditioner was applied: the relative residual is smaller than* $10^{-8}$ *after 8 iterations.*

**5. Conclusions.** We have presented a few approaches for solving general KKT systems. In particular we have addressed the case of a singular (1,1) block. For low nullity we have introduced an approach based on fast computation of a small subset of the solution vector, whose size is equal to the rank of the null space of the (1,1) block. The Lanczos/Stieltjes procedure is used, and iterative refinement can be applied to improve the accuracy. For the case of high rank deficiency, or near singularity of the (1,1) block, we have analyzed the augmented Lagrangian approach and have shown how it is different than the classical case of a positive definite (1,1) block. In particular, we have pointed out that when the weight matrix $W$ is equal to $\gamma I$, if the (1,1) block is singular the performance of the solver can be significantly improved by a good choice of $\gamma$. If there is no obvious way to choose $\gamma$, we suggest to use $\gamma = \|A\|/\|B\|^2$.

An observation on how the inverse of the KKT matrix changes when the (1,1) block is changed has been provided. The connection between the Schur complement of the original KKT matrix and that of the modified one has been introduced, and we have discussed how it can be used to pick values of the Uzawa parameter and the augmented Lagrangian parameter to guarantee convergence. Finally, a class of preconditioners has been introduced and we have analyzed the spectral properties of the associated preconditioned matrix.

**Acknowledgments.** We are grateful to Jim Varah, for numerous fruitful discussions, and for his insightful observations and suggestions. In particular, we thank Jim for pointing out Lemma 3.7. We are indebted to Michael Saunders for his careful reading of the manuscript, and for his invaluable comments and corrections. Many thanks to Eldad Haber, for his input and for providing code, references and explanation for the inverse problem used as an example, and to Rasmus Larsen, for his Lanczos routines. Finally, we thank Dan Boley for his helpful comments and Michele Benzi for providing some useful references.

[1] M. Arioli. The use of QR factorizarion in sparse quadratic programming and backward error issues. *SIAM J. Matrix Anal. Appl.*, 21(3):825–839, 2000.

[2] K. Arrow, L. Hurwicz, and H. Uzawa. *Studies in nonlinear programming.* Stanford University Press, Stanford, CA, 1958.

[3] Z. Bai and G. H. Golub. Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices. *Ann. Numer. Math.*, 4:29–38, 1997.

[4] R. E. Bank, B. D. Welfert, and H. Yserentant. A class of iterative methods for solving saddle point problems. *Numer. Math.*, 55:645–666, 1990.

[5] A. Battermann and M. Heinkenschloss. Preconditioners for Karush-Kuhn-Tucker Matrices arising in the optimal control of distributed systems. *In W. Desch, F. Kappel, K. Kunisch, eds., Optimal Control of Partial Differential Equations*, pages 15–32, 1998.

[6] M. Benzi. Solution of equality-constrained quadratic programming problems by a projection iterative method. *Rendiconti di Matematica*, 13:227–296, 1993.

[7] M. Benzi and G. H. Golub. Bounds for the entries of matrix functions with applications to preconditioning. *BIT*, 39(3):417–438, 1999.

[8] Bjorck. *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, 1996.

[9] J. H. Bramble, J. E. Pasciak, and A. E. Vassilev. Analysis of the inexact Uzawa algorithm for saddle point problems. *SIAM J. Numer. Anal.*, 34(3):1072–1092, 1997.

[10] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag, New York, 1991.

[11] J. R. Bunch and B. N. Parlett. Direct methods for solving symmetric indefinite systems of linear equations. *SIAM J. Num. Anal.*, 8:639–655, 1971.

[12] T. F. Chan. Deflated decompositions of solutions of nearly singular systems. *SIAM J. Sci. Stat. Comput.*, 21(4):738–754, 1984.

[13] N. Dyn and W. E. Ferguson. The numerical solution of equality-constrainted quadratic programming problems. *Math. Comput.*, 41:165–170, 1983.

[14] H. C. Elman. Multigrid and Krylov subspace methods for the discrete Stokes equations. *International Journal for Numerical Methods in Fluids*, 22:755–770, 1996.

[15] H. C. Elman. Preconditioning for the steady-state Navier-Stokes equations at low viscosity. *Technical Report CS-TR-3712, University of Maryland*, 1996.

[16] H. C. Elman and G. H. Golub. Inexact and preconditioned Uzawa algorithms for saddle point problems. *SIAM J. Numer. Anal.*, 31(6):1645–1661, 1994.

[17] H. C. Elman and D. Silvester. Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations. *SIAM J. Sci. Comput.*, 17:33–46, 1996.

[18] H. C. Elman, D. J. Silvester, and A. J. Wathen. Iterative methods for problems in computational fluid dynamics. *Report CS-TR-3675, UMIACS-TR-96-58*, 1996.

[19] M. Fortin and R. Glowinski. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems.* North-Holland, 1983.

[20] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization.* Academic Press, New York, 1981.

[21] R. Glowinski and P. Le Tallec. Numerical solution of problems in incompressible finite elasticity by augmented Lagrangian methods II. three-dimensional problems. *SIAM J. Appl. Math.*, 44(4):710–733, 1984.

[22] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Third Edition, Johns Hopkins University Press, Baltimore, MD, 1996.

[23] G. H. Golub and G. Meurant. Matrices, moments and quadrature. *Numerical Analysis (Dundee)*, pages 105–156, 1993.

[24] G. H. Golub and A. J. Wathen. An iteration for indefinite systems and its application to the Navier-Stokes equations. *SIAM J. Sci. Comput.*, 19:530–539, 1998.

[25] E. Haber, U. M. Ascher, and D. Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse Problems. To appear*, 2000.

[26] G. Hadley. *Nonlinear and Dynamic Programming.* Addison-Wesley, Reading, Mass., 1964.

[27] M. Hestenes. Multiplier and gradient methods. *J. Optimization Theory and Applications*, 4:303–320, 1969.

[28] M. Hestenes. *Conjugate Direction Methods in Optimization.* Springer-Verlag, 1980.

[29] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[30] W. Karush. *Minima of Functions of Several Variables with Inequalities as Side Conditions.* Dissertation, Department of Mathematics, University of Chicago, 1939.

[31] T. H. Kjeldsen. A contextualized historical analysis of the Kuhn-Tucker theorem in nonlinear programming: the impact of World-War II. *Historia Mathematica. To appear.*

[32] A. Klawonn. Preconditioners for indefinite problems. *Doctoral Dissertation, Westfalische Wilhelms-Universitat Munster*, 1995.
[33] A. Klawonn. Block-triangular preconditioners for saddle point problems with a penalty term. *SIAM J. Sci. Comput.*, 19(1):172–184, 1998.
[34] A. Klawonn. An optimal preconditioner for a class of saddle point problems with a penalty term. *SIAM J. Sci. Comput.*, 19(2):540–552, 1998.
[35] A. Klawonn and O. B. Widlund. A domain decomposition method with Lagrange multipliers for linear elasticity. *Technical Report 780, Courant Institute*, 1999.
[36] H. W. Kuhn and A. W. Tucker. Nonlinear programming. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, ed. J. Neyman*, pages 481–492, 1950.
[37] L. Little and Y. Saad. Block preconditioners for symmetric and nonsymmetric saddle point problems. *UMSI 99/104*, 1999.
[38] M. F. Murphy, G. H. Golub, and A. J. Wathen. A note on preconditioning for indefinite linear systems. *SCCM Technical report, number 99-03*, 1999.
[39] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 1999.
[40] C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12:617–629, 1975.
[41] I. Perugia, V. Simoncini, and M. Arioli. Linear algebra methods in a mixed approximation of magnetostatic problems. *SIAM J. Sci. Comput.*, 21(3):1085–1101, 1999.
[42] M. J. D. Powell. A method for nonlinear constraints in minimization problems in optimization. *J. Optimization Theory and Applications*, 4:303–320, 1969.
[43] T. Rusten and R. Winther. A preconditioned iterative method for saddlepoint problems. *SIAM J. SIAM J. Matrix Anal. Appl.*, 13(3):887–904, 1992.
[44] Y. Saad. ILUT: a dual threshold incomplete LU factorization. *Numerical Linear Algebra with Applications*, 1:387–402, 1994.
[45] R. Sibson and G. Stone. Computation of thin plate splines. *SIAM J. Sci. Stat. Comput.*, 12(6):1304–1313, 1991.
[46] G. W. Stewart. On the implicit deflation of nearly singular systems of linear equations. *SIAM J. Sci. Stat. Comput.*, 2(2):136–140, 1981.
[47] G. W. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.
[48] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Massachusetts, 1986.
[49] G. Strang. A framework for equilibrium equations. *SIAM Review*, pages 283–297, 1988.
[50] R. S. Varga. *Matrix Iterative Analysis*. Prentice-Hall, 1962.
[51] A. J. Wathen, B. Fischer, and D. J. Silvester. The convergence of iterative solution methods for symmetric and indefinite linear systems. *Report 97/16, Oxford University*, 1996.