SIAM J. SCI. COMPUT.
Vol. 24, No. 6, pp. 2076–2092

ON SOLVING BLOCK-STRUCTURED INDEFINITE LINEAR SYSTEMS*

GENE H. GOLUB† AND CHEN GREIF‡

Abstract. We consider 2×2 block indefinite linear systems whose $(2,2)$ block is zero. Such systems arise in many applications. We discuss two techniques that are based on modifying the $(1,1)$ block in a way that makes the system easier to solve. The main part of the paper focuses on an augmented Lagrangian approach: a technique that modifies the (1,1) block without changing the system size. The choice of the parameter involved, the spectrum of the linear system, and its condition number are discussed, and some analytical observations are provided. A technique of deflating the (1,1) block is then introduced. Finally, numerical experiments that validate the analysis are presented.

Key words. indefinite linear systems, condition number, Schur complement, singularity, spectrum, augmented Lagrangian

AMS subject classifications. 65F50, 65F35, 65F10

DOI. 10.1137/S1064827500375096

1. Introduction.

Indefinite linear systems arise in many applications. Notable members of this family are linear systems whose associated matrix can be presented as a 2×2 block matrix whose $(2,2)$ block is zero:

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

The matrix A is assumed to be $n \times n$, and B is $n \times p$, where $p \leq n$. (Often $p \ll n$.)

As an example, consider the classical quadratic programming problem:

$$\begin{gathered} \text{Minimize } \tfrac{1}{2} x^T A x - x^T c \\ \text{subject to } B^T x = d. \end{gathered}$$

In optimization terms, A represents the Hessian of the quadratic function to be minimized, and B^T is the Jacobian of the linear constraints.

Solving the problem using Lagrange multipliers amounts to defining a function of the form

$$\phi(x,y) = \frac{1}{2} x^T A x - x^T c + y^T (B^T x - d) \tag{1.2}$$

and computing its stationary points, which satisfy

$$\nabla \phi = 0. \tag{1.3}$$

The result is the linear system (1.1). The components of the vector y are the Lagrange multipliers. See, for example, Nocedal and Wright [40] for more details on this problem and ways to solve it.

*Received by the editors July 10, 2000; accepted for publication (in revised form) November 13, 2002; published electronically July 8, 2003.
http://www.siam.org/journals/sisc/24-6/37509.html

†SCCM Program, Gates 2B, Stanford University, Stanford, CA 94305-9025 (golub@sccm.stanford.edu). This author's work was supported in part by the National Science Foundation.

‡Department of Computer Science, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (greif@cs.ubc.ca). The work of this author was done while he was at Stanford University and was supported in part by the Natural Sciences and Engineering Research Council of Canada.

In the context of constrained optimization, (1.3) is part of what defines *first order optimality conditions* known as the Karush–Kuhn–Tucker (KKT) conditions. A precise definition of the KKT conditions can be found in [40, p. 328]. The conditions were first derived by Karush in 1939, in his master's thesis at the University of Chicago [33], and were later rederived by others, but it was in 1950, when the Princeton mathematicians Kuhn and Tucker published their work [38], that the theory of constrained optimization took off and started taking its modern shape. Throughout the years it has become common practice to call (1.1) *the KKT system* or *the augmented system*. A detailed account of the history of the derivation of the KKT conditions and the theory of nonlinear programming can be found in [34] and references therein.

Problem (1.1) also arises in numerous applications other than optimization, especially in the solution of PDEs, and should therefore be considered in a broad context. Formulations that lead to such linear systems appear in fluid dynamics, electromagnetics, structural analysis, data fitting, linear elasticity, and other areas of applications.

A large variety of methods for solving linear systems of the form (1.1) can be found in the literature. Among them we mention null-space methods [1, 20, 24, 40]; direct solvers [10, 41]; the classical Uzawa algorithm [2] and the inexact Uzawa algorithm [14]; splitting schemes such as the one introduced in [12], which was later generalized to real positive matrices in [27]; preconditioned Krylov subspace solvers based on approximating the Schur complement or other methodologies [13, 15, 16, 17, 35, 37, 39, 50]. See also [5, 7, 24, 40, 47, 48] for surveys of existing methods and further references.

Of particular interest to us are techniques based on the Schur complement $S = -B^T A^{-1} B$. The importance of this matrix is evident by the block factorization

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^T A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & -B^T A^{-1} B \end{pmatrix} \begin{pmatrix} I & A^{-1} B \\ 0 & I \end{pmatrix}. \tag{1.4}$$

Since computing S exactly and inverting it may require practically as much computational work as solving (1.1), many preconditioning techniques (see [16] and references therein) are based on approximating S^{-1} and A^{-1}. Thus the numerical properties of the (1,1) block, such as its condition number and the distribution of its eigenvalues, are important in finding effective approximations.

Conditions for the nonsingularity of $\mathcal{K}$ can be found in various papers. An important result is the following [40, Lem. 16.1]: if Z is a basis for the null space of B^T, and if the reduced Hessian $Z^T A Z$ is positive definite and B has full column rank, then $\mathcal{K}$ is nonsingular.

Since (1.4) is a congruence transformation, the inertia of $\mathcal{K}$, which is denoted $i(\mathcal{K})$ and is defined as the ordered triplet that specifies the number of its positive, negative, and zero eigenvalues [32, p. 221, Def. 4.5.6], is equal to the sum of inertias of the matrices A and the Schur complement. When A is positive definite we have $i(\mathcal{K}) = (n, p, 0)$. A more general result, which applies also if A is singular (in which case decomposition (1.4) is not valid), connects the inertia of $\mathcal{K}$ with the inertia of $Z^T A Z$: $i(\mathcal{K}) = (p + n_+, p + n_-, n_0)$, where $i(Z^T A Z) = (n_+, n_-, n_0)$. See [28] and [23, Thm. 3.1, Cor. 3.1]. Further results on the inertia and the spectrum of $\mathcal{K}$ can be found in [21, 43].

Throughout this paper we consider cases where the (1,1) block is possibly singular or ill-conditioned. Frequently the singularity appears in the form of semidefiniteness. Examples of applications are domain decomposition techniques for solving linear elasticity problems [36], computation of thin plate splines [44], geophysical inverse problems [30], and various optimization problems [40]. In such cases, due to the

singularity, techniques based on computing the Schur complement or approximating its inverse cannot be straightforwardly applied. This motivates in part our discussion: how to modify the linear system so that the singularity no longer poses a difficulty.

Scope of this paper. We refer to the linear system (1.1) in a broad context, without making strict assumptions with regard to the nature of the underlying continuous problem. Our primary goal is to discuss ways of modifying system (1.1) in a way that may make it easier to solve, using existing methods. We thus focus on the "preprocessing" stage that leads to a modified linear system that has the same solution but may be easier to deal with numerically.

Assumptions throughout.

- The matrix $\mathcal{K}$ is assumed to be nonsingular.
- With the exception of one case (Proposition 2.1), the matrix associated with the linear system is symmetric, just as it appears in (1.1).
- A is sparse and square, and its singularity or nonsingularity (as well as definiteness) are to be explicitly mentioned in the specific context. B is sparse and of full column rank. As mentioned earlier, A and B are assumed to be $n \times n$ and $n \times p$, respectively.

Part of our analysis and findings applies only to a positive definite or a positive semidefinite (1,1) block, but experimental evidence suggests that the analysis can in some cases predict the behavior in circumstances it does not strictly satisfy. When singularity is present, our methodology attacks it by transforming the system into one where the $(1,1)$ block is nonsingular. The benefit is that methods that rely on the nonsingularity of the (1,1) block can then be applied.

We discuss two different strategies: an augmented Lagrangian approach [22, 31, 42] (section 2) and a procedure that leads to a smaller indefinite system by deflating the (1,1) block (section 3). The augmented Lagrangian approach is parameter-dependent and may be useful both when the (1,1) block is singular and when it is nonsingular. In the main part of the paper we provide some analytical observations regarding the spectrum of the matrix, and discuss ways of selecting the parameter.

In section 4 we present numerical examples. Finally, in section 5 we draw some conclusions.

2. An augmented Lagrangian approach. 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

$$\left\{ \begin{array}{lcl} \left(A + BWB^T\right) x + By & = & c + BWd, \\ B^T x & = & d. \end{array} \right. \tag{2.1}$$

This new linear system has the same solution and may be easier to solve, depending on the method used. For methods that rely on the Schur complement, for example, there may be one immediate benefit: even if the original (1,1) block was singular or ill-conditioned, the (1,1) block of the modified linear system, (2.1), may be nonsingular, and with attractive properties such as positive definiteness or a small condition number.

Performing the step that leads to (2.1) can be considered an augmented Lagrangian technique [31, 42]. In constrained optimization, the augmented Lagrangian technique combines introducing both quadratic penalty terms and Lagrange multipliers [40, Chap. 17]. The technique is also called the method of multipliers [31]. It is useful in restraining the ill-conditioning inherent in quadratic penalty methods, by which it is necessary to drive the penalty parameter to zero so that the constraints

are not violated. The name "augmented Lagrangian" has been used also to describe the underlying linear system (see, for example, [22]). Certain numerical properties of linear systems of the form (2.1) were analyzed in [9, 22] (and in other places) for applications arising from elliptic PDEs with boundary conditions. The (1,1) block in the original system was in many cases assumed to be positive definite, and the choice $W = \gamma I$ (where γ is a scalar) was considered.

Selecting W as a function of the parameter γ may be difficult. Possible choices are the following:

- *Scaling based on norms.* For example, the choice $W = \gamma I$ with $\gamma = \|A\|/\|B\|^2$ may often force the norm of the matrix A to be of the same magnitude as the norm of the added term BWB^T. This in turn may cause a significant difference in the spectrum and the condition number of the matrix $A+BWB^T$ in comparison to A. More on this is to be discussed in section 2.3.
- *Sparsity considerations.* Since the sparsity pattern of B could be considerably different than that of A, it may be desirable to aim at the least possible change of the sparsity pattern of the (1,1) block which would still accomplish turning it into a nonsingular matrix. Here one of the simplest possibilities for choosing W is to have it as a diagonal matrix with 1's and 0's, meaning that we add $\tilde{B}\tilde{B}^T$ to A, where $\tilde{B}$ is a matrix comprised of some but not all of the columns of B.
- *Desired numerical properties.* A possible goal may be to obtain positive definiteness of the (1,1) block. For example, if A is semidefinite, it is relatively easy to achieve this goal. See Hestenes [31, pp. 76–77] for a related discussion. Working with a positive definite (1,1) block may make it easier to solve the system, in particular if methods that rely on inverting the (1,1) block are used.

2.1. Estimates of the condition number. We start this part of our discussion by providing a connection between the inverses of the original matrix (1.1) and the modified one, associated with (2.1). The following result appeared in [18] (and in other places) for symmetric matrices, and holds for nonsymmetric matrices as well.

PROPOSITION 2.1. *Suppose that A is a general $n \times n$ matrix, B and C are full column rank $n \times p$ matrices ($p \le n$), and W is a $p \times p$ matrix. Define*

$$\mathcal{K}(W) = \begin{pmatrix} A + BWC^T & B \\ C^T & 0 \end{pmatrix}. \tag{2.2}$$

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

$$\mathcal{K}^{-1}(W) = \mathcal{K}^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & W \end{pmatrix}. \tag{2.3}$$

There is more than one way to prove this result. It can be verified by constructing the inverse directly, or by using the Sherman–Morrison–Woodbury formula [25, p. 50].

Proposition 2.1 allows us to present the following upper bound on the condition number of the matrix $\mathcal{K}(W)$.

COROLLARY 2.2. *The condition number of $\mathcal{K}(W)$ defined in* (2.2) *satisfies*

$$\kappa_2\left(\mathcal{K}(W)\right) \le \kappa_2\left(\mathcal{K}\right) + \|W\|_2\|B\|_2\|C\|_2 \cdot \left(\|\mathcal{K}^{-1}\|_2 + \|W\|_2\right) + \|\mathcal{K}\|_2\|W\|_2. \tag{2.4}$$

Proof. Clearly,

$$\|\mathcal{K}(W)\|_2 \le \|\mathcal{K}\|_2 + \|W\|_2\|B\|_2\|C\|_2. \tag{2.5}$$

From Proposition 2.1 it follows that

$$\|\mathcal{K}^{-1}(W)\|_2 \le \|\mathcal{K}^{-1}\|_2 + \|W\|_2. \tag{2.6}$$

The last two inequalities lead to (2.4). ∎

2.2. Condition number and spectrum analysis for the case $W = \gamma I$. The specific choice $W = \gamma I$ allows us to make further observations that are relevant to the rate of convergence and accuracy of iterative solvers; see also section 2.3. Two immediate consequences of Proposition 2.1 follow.

COROLLARY 2.3. *Let*

$$\mathcal{K}(\gamma) = \begin{pmatrix} A + \gamma BB^T & B \\ B^T & 0 \end{pmatrix}. \tag{2.7}$$

Under the assumptions of Proposition 2.1*, with* $C = B$*,*

$$\mathcal{K}^{-1} - \mathcal{K}^{-1}(\gamma) = \begin{pmatrix} 0 & 0 \\ 0 & \gamma I \end{pmatrix}. \tag{2.8}$$

COROLLARY 2.4. *For* $W = \gamma I$*, as* $\gamma \to \infty$*,*

$$\frac{\kappa_2\left(\mathcal{K}(\gamma)\right)}{\gamma^2} \to \|B\|^2. \tag{2.9}$$

The analysis that follows is performed for the case of a positive definite (1,1) block.

PROPOSITION 2.5. *Suppose that* A *is an* $n \times n$ *symmetric positive definite matrix and* B *is an* $n \times p$ *matrix of full column rank. There exists an* $n \times n$ *matrix* G *such that* $A = GG^T$ *and* $BB^T = G(\Sigma\Sigma^T)G^T$*, where* Σ *is an* $n \times p$ *matrix such that the* $n \times n$ *matrix* $\Sigma\Sigma^T$ *is diagonal and consists of the generalized eigenvalues of the problem*

$$\sigma^2 Ax = BB^T x. \tag{2.10}$$

Proof. The result can be obtained by employing a technique of simultaneous diagonalization (see [25, pp. 461–463] or [46, p. 281]). Let $A = FF^T$. Denote the singular value decomposition of $F^{-1}B$ by

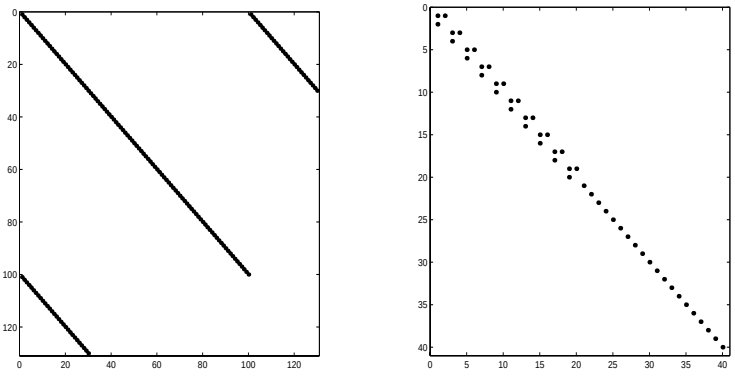
$$F^{-1}B = U\Sigma V^T, \tag{2.11}$$

where U is $n \times n$, Σ is of size $n \times p$ with its last $n - p$ rows identically zero, and V is $p \times p$. Define $G = FU$. From (2.11) we have that $B = G\Sigma V^T$ and thus $BB^T = G\Sigma\Sigma^T G^T$. Since U is orthogonal, we have

$$A = FF^T = FUU^T F^T = GG^T. \tag{2.12}$$

By (2.11) we have

$$(F^{-1}B)(F^{-1}B)^T = F^{-1}BB^T F^{-T} = U\Sigma\Sigma^T U^T, \tag{2.13}$$



(a) Original (b) Permuted

FIG. 2.1. *Sparsity patterns of $T(\gamma)$ and its symmetric permutation.*

and since the eigenvalues of $F^{-1}BB^TF^{-T}$ are equal to those of $F^{-T}F^{-1}BB^T = A^{-1}BB^T$ (see [32, p. 53] for justification), these are the generalized eigenvalues defined in (2.10). Since Σ is $n \times p$, the matrix $\Sigma\Sigma^T$ cannot have more than p nonzeros: this follows since $\ker(B^T)$ forms a linear space for the zero generalized eigenvalues (see [22, p. 9]). □

Using Proposition 2.5, we may rewrite the augmented Lagrangian matrix $\mathcal{K}(\gamma)$ as

$$\begin{pmatrix} A+\gamma BB^T & B \\ B^T & 0 \end{pmatrix} = \begin{pmatrix} GG^T+\gamma G\Sigma\Sigma^TG^T & G\Sigma V^T \\ V\Sigma^TG^T & 0 \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} I+\gamma\Sigma\Sigma^T & \Sigma \\ \Sigma^T & 0 \end{pmatrix} \begin{pmatrix} G^T & 0 \\ 0 & V^T \end{pmatrix}. \tag{2.14}$$

Since

$$\left\| \begin{pmatrix} G & 0 \\ 0 & V \end{pmatrix} \right\| = \left\| \begin{pmatrix} FU & 0 \\ 0 & V \end{pmatrix} \right\| = \max\{\|FU\|, \|V\|\} = \max\{\|F\|, 1\} \tag{2.15}$$

and similarly for the inverse, it follows that

$$\kappa_2\left(\mathcal{K}(\gamma)\right) \le \alpha \cdot \kappa_2 \begin{pmatrix} I+\gamma\Sigma\Sigma^T & \Sigma \\ \Sigma^T & 0 \end{pmatrix}, \tag{2.16}$$

where α depends on $\|F\|$ and $\|F^{-1}\|$ but not on γ.

The matrix

$$T(\gamma) = \begin{pmatrix} I+\gamma\Sigma\Sigma^T & \Sigma \\ \Sigma^T & 0 \end{pmatrix} \tag{2.17}$$

has a simple structure, depicted in Figure 2.1(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 nth superdiagonal and nth subdiagonal, we have $\sigma_i, i = 1, \ldots, p$.

LEMMA 2.6. *Let σ_m^2 and σ_M^2 denote the minimal and maximal generalized eigenvalues, respectively, of* (2.10). *For γ sufficiently large, the condition number of $T(\gamma)$ satisfies*

$$\kappa_2\left(T(\gamma)\right) \approx \frac{(1+\gamma\sigma_m^2)\left((1+\gamma\sigma_M^2)^2+\sigma_M^2\right)}{\sigma_m^2(1+\gamma\sigma_M^2)}. \tag{2.18}$$

In addition, the p negative eigenvalues of $T(\gamma)$ tend to cluster around $-\frac{1}{\gamma}$ as γ becomes large.

Proof. Consider the permutation vector defined, using MATLAB notation, by

$$\tilde{p} = [1, n+1, 2, n+2, 3, n+3, 4, n+4, \ldots, m, n+m, m+1 : n]. \tag{2.19}$$

The sparsity structure of the matrix corresponding to the symmetric permutation of $T(\gamma)$ associated with $\tilde{p}$ is depicted in Figure 2.1(b). The permuted matrix is block diagonal with either 2×2 or 1×1 blocks. All the 1×1 blocks are equal to 1. The 2×2 blocks are of the form

$$\operatorname{diag}\begin{pmatrix} 1+\gamma\sigma^2 & \sigma \\ \sigma & 0 \end{pmatrix}, \tag{2.20}$$

where σ^2 is a generalized eigenvalue of (2.10).

The eigenvalues of the 2×2 blocks of (2.20), for a given $\sigma = \sigma_i$, are

$$\lambda(\gamma) = \frac{1+\gamma\sigma^2 \pm \sqrt{(1+\gamma\sigma^2)^2+4\sigma^2}}{2}. \tag{2.21}$$

For γ sufficiently large, we have

$$\frac{2\sigma}{1+\gamma\sigma^2} \ll 1, \tag{2.22}$$

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

$$\lambda(\gamma) \approx \frac{1+\gamma\sigma^2}{2} \cdot \left(1 \pm \left(1 + \frac{2\sigma^2}{(1+\gamma\sigma^2)^2}\right)\right). \tag{2.23}$$

The condition number of the block diagonal matrix (2.20) is a ratio of two eigenvalues (the maximal and the minimal in absolute value) that do not necessarily belong to the same 2×2 block. The expression for the eigenvalues with the positive sign is monotonically increasing as a function of σ. Thus the maximal positive eigenvalue is

$$\lambda_{\max}(\gamma) = \frac{1+\gamma\sigma_M^2 + \sqrt{(1+\gamma\sigma_M^2)^2+4\sigma_M^2}}{2}. \tag{2.24}$$

Using the Taylor expansion, this simplifies to

$$\lambda_{\max}(\gamma) \approx 1 + \gamma\sigma_M^2 + \frac{\sigma_M^2}{1+\gamma\sigma_M^2}. \tag{2.25}$$

For the block diagonal matrix (2.20) we have to determine whether the minimum in absolute value is attained for the smallest positive eigenvalue or for the negative

eigenvalue closest to zero. The smallest positive eigenvalue is approximately given by $1 + \gamma\sigma_m^2 + \frac{\sigma_m^2}{1+\gamma\sigma_m^2}$. On the other hand, for the negative eigenvalues we have, by (2.23), expressions of the form $-\frac{\sigma^2}{1+\gamma\sigma^2}$. As γ becomes large, these eigenvalues tend to cluster around $-\frac{1}{\gamma}$. In addition, the magnitude of the negative eigenvalue for $\sigma = \sigma_m$ is smaller than the smallest positive eigenvalue, and we have

$$\min_i |\lambda_i(\gamma)| \approx \frac{\sigma_m^2}{1 + \gamma\sigma_m^2}. \tag{2.26}$$

Equations (2.25) and (2.26) lead to (2.18). ∎

2.3. Practical choice of γ. Lemma 2.6 shows that as γ grows larger the condition number of $T(\gamma)$ grows larger. At the same time, the negative eigenvalues of $T(\gamma)$ tend to cluster near $-\frac{1}{\gamma}$, while its positive eigenvalues spread over a larger range of values in comparison with the positive eigenvalues of A. For Krylov solvers, it is known that the condition number and the distribution of the eigenvalues play a role in the speed of convergence [29]. It should be noted, though, that such solvers will typically be applied after preconditioning, and hence further analytical observations will require taking into consideration the particular preconditioner used.

Lemma 2.6 addresses the matrix $T(\gamma)$, but for the matrices we have tested experimentally we have noticed that similar effects occur for $\mathcal{K}(\gamma)$, both with a positive definite or an indefinite (including singular) (1,1) block.

Three matrices that seem relevant for a general discussion (without getting into the specifics of any particular method) are the following:

1. the matrix $\mathcal{K}(\gamma)$,
2. the (1,1) block: $A + \gamma BB^T$,
3. the Schur complement $-B^T(A + \gamma BB^T)^{-1}B$.

If the (1,1) block is singular or nearly singular, we should seek a value of γ that is large enough so as to eliminate the effect of the ill-conditioning of A, but not too large, so as to avoid the effect of the singular matrix BB^T.

If A is well-conditioned to begin with, a choice of a value of γ that is significantly better than other choices may be less obvious or may not exist, and as long as γ is not large, the condition numbers of the (1,1) block and the Schur complement may change slowly as functions of γ. This illustrates an important difference between the case of a singular (1,1) block and the case of a nonsingular one.

The Uzawa algorithm [2] is a good illustration of a method whose performance depends on the above mentioned three matrices in more than one way. The algorithm 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 $y_{k+1} = y_k + \alpha(B^T x_{k+1} - d)$.

The solution of the linear system involving the matrix A, if done iteratively, is called the *inner iteration*, and an inner iteration combined with the computation of y_{k+1} that follows forms the *outer iteration*. It is possible to perform the inner iteration inexactly (see [4, 8, 14]). Convergence analysis shows that when A is positive definite, the optimal Uzawa parameter, α, is given by

$$\alpha_{opt} = \frac{2}{\lambda_{\min}(S) + \lambda_{\max}(S)}, \tag{2.27}$$

where $S = B^T A^{-1} B$ [14]. If the technique discussed in this section with $W = \gamma I$ is applied, the iteration takes the following form:

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

The spectrum of the Schur complement is fundamental to the convergence behavior of the Uzawa scheme. See [22, Chap. 1] for a comprehensive analysis of the case where A is positive definite.

PROPOSITION 2.7. *Let*

$$S(\gamma) = B^T(A + \gamma BB^T)^{-1}B. \tag{2.28}$$

If there exists a scalar ν satisfying $\gamma \geq \nu \geq 0$ such that $A + \nu BB^T$ is positive definite and the eigenvalues of $S(\nu)$ are $\{\mu_i(\nu)\}$, then the eigenvalues of $S(\gamma)$ are

$$\mu_i(\gamma) = \frac{\mu_i(\nu)}{1 + (\gamma - \nu)\mu_i(\nu)}. \tag{2.29}$$

A possible way to prove this result is by applying the Sherman–Morrison–Woodbury formula to the matrix $(\tilde{A} + \beta BB^T)^{-1}$, where $\tilde{A} = A + \nu BB^T$ and $\beta = \gamma - \nu$. The result also follows from [19, p. 327, exercise 12.12].

In order for the Uzawa scheme to converge for any initial guess, α must be such that $\rho(I - \alpha S(\gamma)) < 1$ [14]. (As usual, ρ denotes the spectral radius of a matrix [49, p. 9].) From Proposition 2.7 it follows that the matrix $I - \alpha S(\gamma)$ has the eigenvalues $\frac{1+(\beta-\alpha)\mu_i(\nu)}{1+\beta\mu_i(\nu)}$. If $A + \nu BB^T$ is positive definite, then so is $B^T(A + \nu BB^T)^{-1}B$. (We assume that B has full rank.) We thus have $\mu_i(\nu) > 0$ for all i, and hence $1 + \beta\mu_i(\nu) > 0$, and imposing

$$-1 < \frac{1 + (\beta - \alpha)\mu_i(\nu)}{1 + \beta\mu_i(\nu)} < 1 \tag{2.30}$$

is equivalent to $0 < \alpha < 2\beta + \frac{2}{\mu_i(\nu)}$. This will hold if

$$0 < \alpha < 2\beta + \frac{2}{\mu_{max}(\nu)} \tag{2.31}$$

and is satisfied for any arbitrary set of positive eigenvalues $\{\mu_i(\nu)\}$ if $0 < \alpha \leq 2\beta$. It is desirable to find the minimal value of ν that yields a positive definite matrix $A + \nu BB^T$, so as to obtain as large as possible a range of values of α for which there is convergence. If A itself is positive definite, then $\nu = 0$ and the results in [22] apply. If A is positive semidefinite, then, since $\mathcal{K}$ is nonsingular, for any arbitrarily small $\nu > 0$ the matrix $A + \nu BB^T$ is positive definite.

3. A method of eliminating the nullity of the (1,1) block by reducing the system size. Suppose A is a singular symmetric $n \times n$ matrix, whose rank is $n - q$, and q is a small integer. (The reasons for assuming that q is small have to do with computational cost considerations and will be clarified below.)

3.1. Derivation of the proposed algorithm. Suppose that there exists an orthogonal $n \times n$ matrix Q, partitioned into

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}, \tag{3.1}$$

where Q_1 has $n - q$ columns and Q_2 has q columns, such that $Q_1^T A Q_1$ is nonsingular.

For example, given the spectral decomposition of A,

$$A = Q^T \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} Q^T, \tag{3.2}$$

where Λ is the nonsingular $(n-q)\times(n-q)$ diagonal matrix whose entries are the nonzero eigenvalues of A and Q is the matrix whose columns are the eigenvectors of A, we can define Q_2 as the columns of Q that span the null-space of A.

Let

$$V = \begin{pmatrix} Q & 0 \\ 0 & I \end{pmatrix} \tag{3.3}$$

be an $(n+p)\times(n+p)$ matrix. We then have

(3.4)

$$V^T \mathcal{K} V = \begin{pmatrix} Q^T & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} Q_1^T A Q_1 & Q_1^T A Q_2 & Q_1^T B \\ Q_2^T A Q_1 & Q_2^T A Q_2 & Q_2^T B \\ B^T Q_1 & B^T Q_2 & 0 \end{pmatrix}.$$

We now perform the following two steps.

1. Multiply the original system (1.1) by V^T from the left on both sides and use (3.4) to write the resulting system as a 3×3 block system.
2. Symmetrically permute block rows and columns 2 and 3 of the system obtained in the previous step.

The resulting linear system is

$$\left(\begin{array}{cc|c} Q_1^T A Q_1 & Q_1^T B & Q_1^T A Q_2 \\ B^T Q_1 & 0 & B^T Q_2 \\ \hline Q_2^T A Q_1 & Q_2^T B & Q_2^T A Q_2 \end{array}\right) \left(\begin{array}{c} Q_1^T x \\ y \\ \hline Q_2^T y \end{array}\right) = \left(\begin{array}{c} Q_1^T c \\ d \\ \hline Q_2^T c \end{array}\right). \tag{3.5}$$

We can rewrite it as

$$\begin{pmatrix} \hat{\mathcal{K}} & U \\ U^T & \hat{A}_2 \end{pmatrix} \begin{pmatrix} z \\ x_2 \end{pmatrix} = \begin{pmatrix} e \\ c_2 \end{pmatrix}, \tag{3.6}$$

where

$$\hat{\mathcal{K}} = \begin{pmatrix} Q_1^T A Q_1 & Q_1^T B \\ B^T Q_1 & 0 \end{pmatrix} \tag{3.7}$$

and

$$U = \begin{pmatrix} Q_1^T A Q_2 \\ B^T Q_2 \end{pmatrix}; \quad z = \begin{pmatrix} Q_1^T x \\ y \end{pmatrix}; \quad e = \begin{pmatrix} Q_1^T c \\ d \end{pmatrix}; \tag{3.8}$$

$$\hat{A}_2 = Q_2^T A Q_2; \quad x_2 = Q_2^T y; \quad c_2 = Q_2^T c.$$

Note that $\hat{\mathcal{K}}$ is a block-structured indefinite matrix of a smaller size, $(n+p-q)\times(n+p-q)$, whose (1,1) block is now nonsingular by our initial assumption.

We can now eliminate x_2 by solving for the Schur complement as follows:

$$\left(\hat{A}_2 - U^T \hat{\mathcal{K}}^{-1} U\right) x_2 = c_2 - U^T \hat{\mathcal{K}}^{-1} e. \tag{3.9}$$

Once x_2 has been computed, we can solve for z by solving the system

$$\hat{\mathcal{K}} z = e - U x_2. \tag{3.10}$$

To summarize, the algorithm we propose is the following.

1. Find $Q = [Q_1\ Q_2]$ such that $QQ^T = I$ and $Q_1^T A Q_1$ is nonsingular.
2. Solve (3.9), where the quantities in that system are defined in (3.7)–(3.8).
3. Solve (3.10).
4. Compute x and y by multiplying the required quantities by Q or its components and by using the orthogonality of Q.

3.2. Remarks and discussion of computational aspects.

1. *A decomposition-free algorithm.* The above formulation works for any valid choice of Q, under the conditions specified. There are several ways to reduce the cost of the step involving forming the matrix Q. For example, it is possible to set Q to be a diagonal matrix with either 1's or 0's along its diagonal. Constructing a nonsingular matrix $Q_1^T A Q_1$ amounts to eliminating rows and columns of A such that the resulting smaller matrix is nonsingular.
2. *Deflation.* The algorithm relates to deflation techniques [11, 45], in that the $(1,1)$ block is deflated. The original linear system, however, is nonsingular and its solution space is not changed.

The singularity of the $(1,1)$ block is eliminated by reducing its size by exactly its nullity. The advantage is that methods that rely on the nonsingularity may now be applied to the reduced system. On the other hand, the computational cost of this procedure, which involves solving two separate linear systems, needs to be addressed.

The following factors determine the computational cost of solving (3.9).

- *The rank of A.* The smaller the nullity q is, the smaller the cost would be. For example, if $\text{rank}(A) = n - 1$ (which occurs often, for example, for certain discretized PDEs with Neumann conditions), then (3.9) is merely a scalar equation.
- *Application of the Lanczos/Stieltjes procedure.* The expression $U^T \hat{\mathcal{K}}^{-1} U$ is a quadratic form. It is well known [26] that the Lanczos/Stieltjes algorithm can be applied to rapidly evaluate this quantity. Details and a careful error analysis of the Lanczos/Stieltjes procedure can be found in [3, 6, 25, 26].
- *Inexact solve and iterative refinement.* x_2 does not always need to be computed exactly—for example, in cases where the original problem is nonlinear and solving the linear system is a step in an inexact Newton-type solve. The Lanczos/Stieltjes procedure can then be terminated after a small number of steps. This reduces the overall cost of the algorithm. In the above mentioned references the error analysis allows for having an upper and a lower bound on the error in the computation of a quadratic form. In the event that the resulting residual is still larger than required, iterative refinement can be performed: solve (3.6) using the proposed algorithm; compute the residual and pose it as the right-hand-side vector; repeat those two steps until satisfied.

4. Numerical examples.

4.1. A randomly generated linear system. Extensive tests with random matrices have been performed and have produced qualitatively similar results.

Our first example is a linear system of the form (1.1) that was constructed in the following manner: A is a 2500×2500 block diagonal matrix of 50 pentadiagonal blocks, consisting of normally distributed random numbers. Each pentadiagonal block $A_i, i = 1, \ldots, 50$, has nullity of at least 1, generated by setting $A_i \leftarrow A_i - \lambda_{min}(A_i) I$ after the construction. Thus A is semidefinite and its rank should be at most 2450. (In this example, the rank was verified to be exactly 2450.) The random matrix B

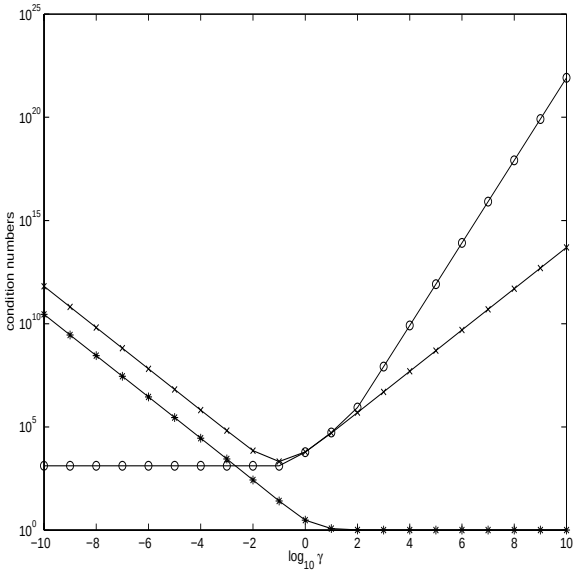


FIG. 4.1. *Condition numbers of $A + \gamma BB^T$ ("-x"), the matrix $\mathcal{K}(\gamma)$ ("-o"), and $B^T(A + \gamma BB^T)^{-1}B$ ("-*"), as a function of γ, for numerical example* 4.1.

is of size 2500×500, and is comprised of five 500×500 tridiagonal blocks assembled together. The matrix $\mathcal{K}$ is thus of size 3000×3000; the right-hand side is a 3000-element randomly generated vector.

The matrix $\mathcal{K}$ is well-conditioned: $\kappa_2(\mathcal{K}) = 1303.2$. Figure 4.1 depicts the condition numbers of the three matrices of interest, as γ changes: the 2×2 block matrix, the (1,1) block, and the Schur complement.

The results validate some of our previously stated observations in section 2. The condition number of $\mathcal{K}(\gamma)$ does not vary significantly as long as γ is small, but as γ gets larger it grows larger and behaves like γ^2; compare with Corollary 2.4.

$A + \gamma BB^T$ is singular for $\gamma = 0$, and when $\gamma = 10^{-10}$ there is only a slight improvement: $\kappa_2(\mathcal{K}(10^{-10})) = 6.53 \times 10^{11}$. As γ grows larger, the condition number becomes dramatically smaller: $\kappa_2(A + 0.1BB^T) = 2100.0$. But as γ grows yet larger, the condition number starts increasing again due to the rank deficiency of the $n \times n$ rank p matrix BB^T.

Finally, the condition number of $B^T(A + \gamma BB^T)^{-1}B$ improves as γ gets larger, as expected.

Even though we do not have an analytical way of determining the optimal value of γ, we note that the choice pointed out as a possibility in section 2, namely $\gamma = \|A\|/\|B\|^2$, works well in this example: it is equal to 0.1748 and is well within the range of values for which the condition numbers of all three matrices are fairly close to their minimum.

In Table 4.1 we present results of running the classical Uzawa algorithm. The Uzawa parameter α was chosen to be equal to γ. As follows from the discussion in section 2.3, the scheme should theoretically converge in this case for any initial guess.

The following observations can be made.

1. For γ very small, no convergence to the desired threshold is reached, within the imposed maximum number of iterations. $A+\gamma BB^T$ is still nearly singular

TABLE 4.1

Performance of the classical Uzawa algorithm for the 3000×3000 *randomly generated sparse linear system of Example* 4.1. *The second, third, and fourth columns correspond, respectively, to the norm of the relative residual, the norm of the error, and the number of iterations. The stopping criterion was the norm of the relative residual smaller than* 10^{-6}. *If this convergence criterion was not satisfied after* 2000 *iterations, the program execution was aborted and the norms presented in these cases are for* 2000 *iterations.*

$\text{Log}_{10}\gamma$	$\|\text{Rel. res.}\|_2$	$\|\text{Error}\|_2$	Iter
-3	6.2322e-03	3.2085e-01	> 2000
-2	9.8987e-07	6.2448e-05	955
-1	9.6296e-07	6.1730e-05	100
0	4.9195e-07	5.9695e-05	14
1	1.2786e-07	1.1557e-04	4
2	4.0451e-08	3.0454e-04	2
3	1.2541e-07	1.2514e-02	1

and as a result, both the inner iteration and the outer iteration (which are affected, respectively, by the conditioning of A and of the Schur complement) are prone to numerical inaccuracies.

2. As γ becomes larger, both $A+\gamma BB^T$ and $B^T(A+\gamma BB^T)^{-1}B$ become better conditioned. As a result the inner iteration produces accurate solutions, and the outer iteration converges increasingly rapidly.
3. As γ grows larger yet, convergence continues to be rapid due to the smaller condition number of the Schur complement. This follows since $\rho(I-\alpha S) = (\kappa-1)/(\kappa+1)$, where κ is the condition number of S [14]. But $A+\gamma BB^T$ and $\mathcal{K}(\gamma)$ are now ill-conditioned; as a result the residual continues to go down, but the norm of the error does not significantly change. We note (details omitted) that the spectra of the matrices involved also change in a way that can explain the difference in speed of convergence.

4.2. A geophysical inverse problem. Consider the following problem, whose full description can be found in [30]. Given observations, d, of a field u at some discrete locations, s, the *model*, m, is to be recovered. The connection between d and u is

$$d = u(s) + \epsilon = Qu + \epsilon, \tag{4.1}$$

where Q projects the field u into the measurement locations s and ϵ is the noise. The constrained problem formulation in [30] is based on the following strategy.

- The quantity to be minimized is $\|Qu - d\|$.
- A forward problem (typically a second order PDE) $A(m)u = f$ needs to be solved exactly and forms a constraint.

Since the problem is ill-posed, it is regularized and the result is the following constrained minimization problem:

$$\begin{aligned} \text{minimize} \quad & \phi(u,m) = \frac{1}{2}\|Qu-d\|^2 + \frac{\beta}{2}\|W(m-m_0)\|^2 \\ \text{subject to} \quad & A(m)u = f, \end{aligned}$$

where m_0 is a reference model and W could be, for example, a discretized second order differential operator.

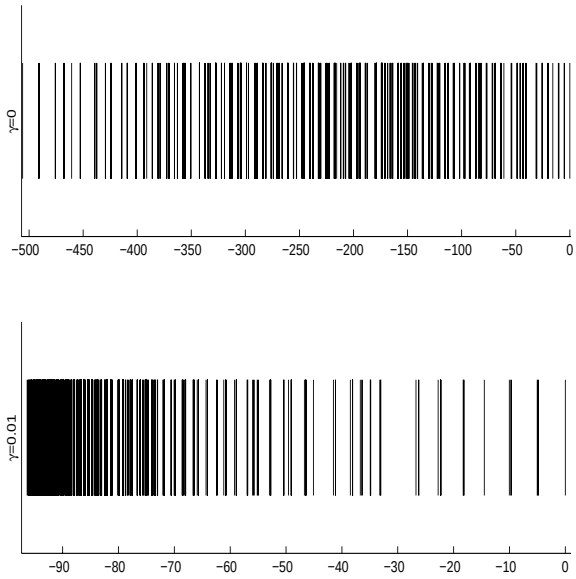


FIG. 4.2. *The* 729 *negative eigenvalues of the original matrix of numerical example* 4.2*, i.e., for* $\gamma = 0$*, vs. those of a modified matrix with* $\gamma = 0.01$*. The eigenvalues closest to zero are approximately* -0.0019 *in both cases.*

A Gauss–Newton solver is used, and in each step an indefinite linear system is solved, where the (1,1) block is typically rank deficient, due to under-sampling. See [30] for full details.

We present results of running a three-dimensional problem using finite volume discretization. The regularization parameter β is equal to 10^{-4} and the noise level is approximately 0.01. The projection matrix Q was chosen to be a discretized gradient operator in one set of tests, and a diagonal matrix in another set of tests, corresponding to sampling of half of the points in the model. Both sets of tests produce results that are qualitatively similar, and the graphs show the results for the first set of tests.

Figures 4.2 and 4.3 were generated for a 1970×1970 linear system, where A is of size 1241×1241 and is of rank 876 and B is of size 1241×729.

Figure 4.2 depicts the 729 negative eigenvalues of $\mathcal{K}(\gamma)$. Our analysis (section 2) does not apply to this case; nevertheless, the behavior is similar to our analytical observations. For $\gamma = 0.01$ there is some clustering of negative eigenvalues near -100. On the other hand, the largest positive eigenvalues of the linear system corresponding to $\gamma = 0.01$ are significantly larger than those of the original matrix. The maximal eigenvalue and the condition number of the matrix corresponding to $\gamma = 0.01$ are 2659.0 and 1.44×10^8, respectively, whereas for A (i.e., for $\gamma = 0$) they are 506.42 and 2.32×10^6, respectively.

Figure 4.3 examines the condition number and shows similar behavior to Figure 4.1. There is a range of values of γ, roughly between 10^{-6} to 10^{-3}, for which the condition numbers of the (1,1) block and the system's indefinite matrix are close to their minimum. ($\|A\|/\|B\|^2 = 3.90 \times 10^{-6}$; thus this value is within this range.) The condition number of the Schur complement is approximately 10^7 for this range of values and is still far from its minimal value. Nevertheless, for this range of values all three condition numbers are either close to or much smaller than their corresponding

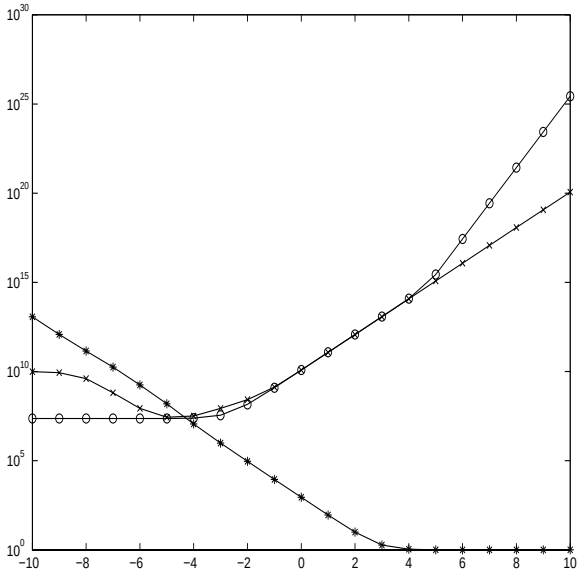


FIG. 4.3. *Condition numbers of the* $(1,1)$ *block ("-x"), the whole KKT matrix ("-o"), and the Schur complement ("-*"), as a function of* γ*, for numerical example* 4.2.

values in the original problem.

5. Concluding remarks. We have discussed indefinite linear systems with a 2×2 block structure, possibly having a singular (1,1) block. Our focus has been on how to modify the linear systems in a way that may make it easier to solve them.

In the main part of this paper we have examined some aspects of the augmented Lagrangian technique. In particular, we have looked at the case where the weight matrix W is equal to γI, and we have made some observations regarding the spectrum of the associated matrix.

There is no obvious way to choose the parameter γ, and our analysis and experiments indicate that the choice is a delicate issue. The condition numbers of the (1,1) block, the indefinite matrix, and the Schur complement depend on γ in different ways. For the (1,1) block (in the event that it is singular) there is typically a decrease in the condition number as γ gets far from zero, but the condition number starts rapidly growing again for large values of γ. The indefinite matrix has a condition number that is monotonically increasing with γ, but the increase is rapid only for large values of γ. Finally, the condition number of the Schur complement monotonically decreases as γ grows larger.

The important point is that there seems to be a range of values of γ where at least two of those condition numbers, and possibly all three, are close to their minimal value. We have experimentally noticed that $\gamma = \|A\|/\|B\|^2$ may be a good choice.

For low nullity of the (1,1) block we have introduced a two-step procedure, whereby in the first step a small subset of the vector of unknowns is computed. The size of this vector is equal to the rank of the null space of the (1,1) block. It is possible to use the Lanczos/Stieltjes procedure to evaluate the quadratic forms involved, and most of the computational work is devoted to solving the linear system associated with the rest of the unknowns. This is now a block-structured indefinite system with a nonsingular (1,1) block, and techniques that rely on the nonsingularity

can be straightforwardly applied.

Acknowledgments. We are grateful to Jim Varah for his insightful observations, and to Eldad Haber for many fruitful discussions and for providing us with his code for example 4.2. Many thanks to Michael Saunders for his careful reading of the original manuscript, and to Michele Benzi for providing some useful references. We have benefited from the referees' thorough reading of the manuscript and their helpful suggestions. In particular, we thank one of the referees for his suggested reformulation of the algorithm proposed in section 3.1, which improved the quality of the exposition.

REFERENCES

[1] M. Arioli and L. Baldini, *A backward error analysis of a null space algorithm in sparse quadratic programming*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 425–442.
[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 (1997), pp. 29–38.
[4] R. E. Bank, B. D. Welfert, and H. Yserentant, *A class of iterative methods for solving saddle point problems*, Numer. Math., 55 (1990), pp. 645–666.
[5] M. Benzi, *Solution of equality-constrained quadratic programming problems by a projection iterative method*, Rend. Mat. Appl. (7), 13 (1993), pp. 275–296.
[6] M. Benzi and G. H. Golub, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT, 39 (1999), pp. 417–438.
[7] Å. Bjorck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[8] J. H. Bramble, J. E. Pasciak, and A. T. Vassilev, *Analysis of the inexact Uzawa algorithm for saddle point problems*, SIAM J. Numer. Anal., 34 (1997), pp. 1072–1092.
[9] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York, 1991.
[10] J. R. Bunch and B. N. Parlett, *Direct methods for solving symmetric indefinite systems of linear equations*, SIAM J. Numer. Anal., 8 (1971), pp. 639–655.
[11] T. F. Chan, *Deflated decomposition of solutions of nearly singular systems*, SIAM J. Numer. Anal., 21 (1984), pp. 738–754.
[12] N. Dyn and W. E. Ferguson, *The numerical solution of equality-constrainted quadratic programming problems*, Math. Comp., 41 (1983), pp. 165–170.
[13] H. C. Elman, *Preconditioning for the steady-state Navier–Stokes equations with low viscosity*, SIAM J. Sci. Comput., 20 (1999), pp. 1299–1316.
[14] H. C. Elman and G. H. Golub, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1645–1661.
[15] H. Elman and D. Silvester, *Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations*, SIAM J. Sci. Comput., 17 (1996), pp. 33–46.
[16] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Performance and analysis of saddle point preconditioners for the discrete steady-state Navier-Stokes equations*, Numer. Math., 90 (2002), pp. 641–664.
[17] B. Fischer, A. Ramage, D. J. Silvester, and A. J. Wathen, *Minimum residual methods for augmented systems*, BIT, 38 (1998), pp. 527–543.
[18] R. Fletcher, *An ideal penalty function for constrained optimization*, in Nonlinear Programming, 2, Academic Press, New York, 1974, pp. 121–163.
[19] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., Wiley, Chichester, UK, 1987.
[20] R. Fletcher and T. Johnson, *On the Stability of Null-Space Methods for KKT Systems*, Numerical Analysis report NA/167, University of Dundee, Department of Mathematics and Computer Science, Dundee, UK, 1995.
[21] A. Forsgren and W. Murray, *Newton methods for large-scale linear equality-constrained minimization*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 560–587.
[22] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, North-Holland, Amsterdam, 1983.
[23] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *Inertia-controlling methods for general quadratic programming*, SIAM Rev., 33 (1991), pp. 1–36.
[24] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, New York, 1981.

[25] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[26] G. H. Golub and G. Meurant, *Matrices, moments and quadrature*, in Numerical Analysis 1993, Pitman Res. Notes Math. Ser. 3, Longman Sci. Tech., Harlow, UK, 1993, pp. 105–156.
[27] 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 (1998), pp. 530–539.
[28] N. I. M. Gould, *On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem*, Math. Programming, 32 (1985), pp. 90–99.
[29] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, Frontiers Appl. Math., 17, SIAM, Philadelphia, 1997.
[30] E. Haber, U. M. Ascher, and D. Oldenburg, *On optimization techniques for solving nonlinear inverse problems*, Inverse Problems, 16 (2000), pp. 1263–1280.
[31] M. Hestenes, *Conjugate Direction Methods in Optimization*, Springer-Verlag, New York, 1980.
[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[33] W. Karush, *Minima of Functions of Several Variables with Inequalities as Side Conditions*, Dissertation, Department of Mathematics, University of Chicago, Chicago, IL, 1939.
[34] T. H. Kjeldsen, *A contextualized historical analysis of the Kuhn-Tucker theorem in nonlinear programming: The impact of World War* II, Historia Math., 27 (2000), pp. 331–361.
[35] A. Klawonn, *An optimal preconditioner for a class of saddle point problems with a penalty term*, SIAM J. Sci. Comput., 19 (1998), pp. 540–552.
[36] A. Klawonn and O. B. Widlund, *A domain decomposition method with Lagrange multipliers and inexact solvers for linear elasticity*, SIAM J. Sci. Comput., 22 (2000), pp. 1199–1219.
[37] P. Krzyzanowski, *On block preconditioners for nonsymmetric saddle point problems*, SIAM J. Sci. Comput., 23 (2001), pp. 157–169.
[38] H. W. Kuhn and A. W. Tucker, *Nonlinear programming*, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1950, J. Neyman, ed., University of California Press, Berkeley, Los Angeles, 1951, pp. 481–492.
[39] M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (2000), pp. 1969–1972.
[40] J. Nocedal and S. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
[41] I. Perugia, V. Simoncini, and M. Arioli, *Linear algebra methods in a mixed approximation of magnetostatic problems*, SIAM J. Sci. Comput., 21 (1999), pp. 1085–1101.
[42] M. J. D. Powell, *A Method for Nonlinear Constraints in Minimization Problems*, Academic Press, London, 1969.
[43] T. Rusten and R. Winther, *A preconditioned iterative method for saddlepoint problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 887–904.
[44] R. Sibson and G. Stone, *Computation of thin plate splines*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 1304–1313.
[45] G. W. Stewart, *On the implicit deflation of nearly singular systems of linear equations*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 136–140.
[46] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
[47] G. Strang, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, Wellesley, MA, 1986.
[48] G. Strang, *A framework for equilibrium equations*, SIAM Rev., 30 (1988), pp. 283–297.
[49] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.
[50] A. J. Wathen, B. Fischer, and D. J. Silvester, *The convergence of iterative solution methods for symmetric and indefinite linear systems*, in Numerical Analysis 1997, Pitman Res. Notes Math. Ser. 380, Addison Wesley Longman Sci. Tech., Harlow, UK, 1997, pp. 230–243.