Determining Subspace Information from the Hessian of a Barrier Function

Margaret H. Wright
AT&T Bell Laboratories
Murray Hill, New Jersey 07974

Abstract

Karmarkar's 1984 paper has produced a resurgence of interest in logarithmic barrier methods for linear and nonlinear programming. Although barrier methods for nonlinear inequality-constrained optimization were widely applied throughout the 1960's, they were largely abandoned in favor of other approaches during the 1970's. This shift was partly caused by the inefficiency of "black-box" methods in solving the unconstrained subproblems. Some of these difficulties were explained by results of Murray and Lootsma showing that barrier Hessian matrices become increasingly ill-conditioned along the trajectory of approach to the solution. This paper discusses several aspects of the Hessian matrix of the logarithmic barrier function. We first show that, except in two special cases, the barrier Hessian is ill-conditioned in an entire region near the solution. At points in a more restricted region (including the barrier trajectory itself), this ill-conditioning displays a special structure connected with subspaces defined by the Jacobian of the active constraints. If a Cholesky factorization with diagonal pivoting detects numerical rank-deficiency in the barrier Hessian, the Cholesky factor can be transformed to reveal information about the underlying subspaces, without making an explicit prediction of the active constraints. Using this subspace information, a close approximation to the Newton direction can be calculated by solving linear systems whose condition reflects that of the original problem.

1. Introduction

1.1. Background

Interior methods for constrained optimization have been the subject of intense activity since the announcement in 1984 of Karmarkar's fast polynomial-time algorithm for linear programming [Kar84]. In the 1960's and early 1970's, nonlinearly constrained optimization problems were widely solved by transformation via penalty and barrier functions to a sequence of *unconstrained* subproblems. Because classical logarithmic barrier techniques are formally related to Karmarkar's method, several polynomial-time algorithms for linear and quadratic programming have been developed since 1984 based on barrier functions (see [Gon92] for a recent survey). Barrier methods are also undergoing a resurgence of interest for nonlinear problems.

In the 1970's, the decline of popularity of barrier methods was attributable in part to inefficiencies experienced by "off-the shelf" unconstrained methods in solving the unconstrained subproblems. An explanation of some of these difficulties was found by Murray [Mur71] and Lootsma [Loot69], who showed that, along the trajectory of approach to the solution, the barrier Hessian matrices become increasingly ill-conditioned, and are singular in the limit. The passage of time has not eliminated this ill-conditioning, which remains as an issue that must be addressed in any practical barrier method.

(Ill-conditioning is less important in linear programming (LP), because of the special nature of an LP solution; see Section 3.)

1.2. Notation and assumptions

Consider the inequality-constrained optimization problem

$$\underset{x\in\mathcal{R}^n}{\text{minimize}}\ f(x) \quad \text{subject to} \quad c_j(x) \geq 0,\ \ j=1,\ldots,m, \tag{1.1}$$

where f and $\{c_j\}$ are smooth. The logarithmic barrier function corresponding to (1.1) is

$$B(x,\mu) = f(x) - \mu \sum_{j=1}^{m} \ln c_j(x), \tag{1.2}$$

where μ is a positive scalar called the *barrier parameter*. This barrier function is defined only at strictly feasible points, and we therefore assume that at least one point $\bar{x}$ exists where $c(\bar{x}) > 0$.

Let x^* denote a local solution of (1.1), and let x_μ denote an unconstrained minimizer of $B(x,\mu)$. Let $g(x)$ denote the gradient of $f(x)$, and $H(x)$ its (symmetric) Hessian. The gradient of $c_j(x)$ is denoted by $a_j(x)$, and its Hessian by $H_j(x)$. The matrix $A(x)$ denotes the $m\times n$ matrix whose jth row is $a_j(x)^T$, i.e. $A(x)$ is the Jacobian of the constraints. The *Lagrangian function* associated with (1.1) is defined as $L(x,\lambda) = f(x) - \lambda^T c(x)$. The Hessian of the Lagrangian with respect to x, $\nabla^2 L(x,\lambda) = H - \sum_{j=1}^m \lambda_j H_j(x)$, will appear in our later discussion.

Let $\hat{m}$ denote the number of constraints active at x^*. The index set $\mathcal{A}$ is defined as $\{k_j\}$, $j=1$, $\ldots$, $\hat{m}$, so that $\mathcal{A}$ contains the indices of the active constraints, $c_j(x^*) = 0$ if $j\in\mathcal{A}$ and $c_j(x^*) > 0$ if $j\notin\mathcal{A}$. We stress that, given x^*, $\mathcal{A}$ is a *fixed* set of indices, and does not vary with x. Since we are concerned only with strictly feasible points, "active" and "inactive" always refer to properties of x^*.

For any point x, $\hat{A}(x)$ means the $\hat{m}\times n$ matrix whose jth row is the gradient of constraint k_j evaluated at x; the matrix $\hat{A}$ is sometimes referred to as the Jacobian of the active constraints. Let $\bar{A}(x)$ denote the $(m-\hat{m})\times n$ matrix whose rows are the gradients of the constraints inactive at x^*, evaluated at x. If v is an m-vector, $\hat{v}$ denotes the $\hat{m}$-dimensional subvector of v with components designated by $\mathcal{A}$.

We use $Z(x)$ to denote a matrix whose columns form an orthonormal basis for the null space of $\hat{A}(x)$, so that $\hat{A}(x)Z(x) = 0$ and $Z(x)^T Z(x) = I$. In general $Z(x)$ is not unique, but may be defined as a smooth function in regions where $\hat{A}(x)$ has full rank (see [CS84]). If a function has Hessian "Hess", the matrix "$Z^T \text{Hess} Z$" is called its reduced Hessian.

We assume throughout that:

(i) $g(x^*) = A^T(x^*)\lambda^*$, where λ^* is called the Lagrange multiplier vector;

(ii) $\lambda_j^* c_j(x^*) = 0$ for $j=1$, $\ldots$, m;

(iii) $\lambda_j^* > 0$ if $j\in\mathcal{A}$, i.e. *strict complementarity* holds at x^*;

(iv) the gradients of the active constraints at x^* are linearly independent, i.e. $\hat{A}(x^*)$ has full row rank;

(v) $Z^{*T} W^* Z^*$ is positive definite, where Z^* denotes $Z(x^*)$ and $W^* = \nabla^2 L(x^*,\lambda^*)$.

It is standard (see, for example, [FM68] or [Fle87]) that, under these conditions, x^* is an isolated local constrained minimizer of (1.1) and λ^* is unique.

The gradient of the barrier function (1.2), $\nabla B(x,\mu)$, is denoted by g_B:

$$g_B(x,\mu) = g(x) - \sum_{j=1}^{m} \frac{\mu}{c_j(x)}\, a_j(x) = g(x) - \mu A^T(x) C^{-1}(x) e. \tag{1.3}$$

The final form in (1.3) introduces the widely used convention that an uppercase version of a letter denoting a given vector means the diagonal matrix whose diagonal elements are those of the vector; thus, $C(x)$ is the $m \times m$ matrix $\mathrm{diag}\big(c_j(x)\big)$, and $C^{-1}(x)$ is the matrix $\mathrm{diag}\big(1/c_j(x)\big)$. The vector e denotes the vector of appropriate dimension whose components are all equal to one, so that $e = (1, 1, \ldots, 1)^T$.

The barrier Hessian $\nabla^2 B(x, \mu)$, denoted by H_B, is given by

$$H_B(x, \mu) = H(x) - \sum_{j=1}^{m} \frac{\mu}{c_j(x)} H_j(x) + \mu A^T(x) C^{-2}(x) A(x). \tag{1.4}$$

Assumptions (i)–(v) imply that, for sufficiently small μ:

(a) an isolated local unconstrained minimizer x_μ of the barrier function (1.2) exists, at which

$$g_B(x_\mu, \mu) = 0, \quad \text{so that} \quad g(x_\mu) = \sum_{j=1}^{m} \frac{\mu}{c_j(x_\mu)} a_j(x_\mu); \tag{1.5}$$

(b) $H_B(x_\mu, \mu)$ is positive definite, and its smallest eigenvalue is bounded away from zero;

(c) if μ is regarded as a continuous parameter, x_μ defines a smooth trajectory converging to x^* as $\mu \to 0$. The points $\{x_\mu\}$ are said to lie on the *barrier trajectory*.

For proofs and additional details, see [FM68], [Fi79] and [Wr92].

We shall use the following standard notation throughout; see [PS82] for details.

Definition 1.1. *(Order notation). Let ϕ be a function of a positive variable h, and let p be fixed.*

- *If there exists a constant $\kappa_u > 0$ such that $|\phi| \le \kappa_u h^p$ for all sufficiently small h, then we write $\phi = O(h^p)$. If $p = 0$, we write $\phi = O(1)$ and say that ϕ is finite.*
- *If there exists a constant $\kappa_l > 0$ such that $|\phi| \ge \kappa_l h^p$ for all sufficiently small h, then we write $\phi = \Omega(h^p)$. If $p = 0$, we write $\phi = \Omega(1)$ and say that ϕ is bounded away from zero.*
- *If there exist constants $\kappa_l > 0$ and $\kappa_u > 0$ such that $\kappa_l h^p \le |\phi| \le \kappa_u h^p$ for all sufficiently small h, then we write $\phi = \Theta(h^p)$. If $p = 0$, we write $\phi = \Theta(1)$ and say that ϕ is both finite and bounded away from zero.*

1.3. Outline of the paper

The remainder of this paper discusses several aspects of Hessian matrices of the logarithmic barrier function. After summarizing results about matrix perturbations (Section 2), we show in Section 3 that, unless $\hat{m} = 0$ or $\hat{m} = n$, the barrier Hessian is ill-conditioned in an entire region near x^*. We then characterize the special structure of invariant subspaces of the Hessian in a more restricted region, including the barrier trajectory itself. Section 4 presents results of Higham [High90] about the Cholesky factorization of a nearly rank-deficient matrix. Section 5 discusses applications of a rank-retaining Cholesky factorization. Finally, Section 6 describes how all of these results may be combined, both to determine subspace information from a barrier Hessian and to compute an approximation to the Newton step.

2. Results on Matrix Perturbations

This section summarizes results from [SS90] that will be applied in Section 3 to analyze the structure of the barrier Hessian in a region near x^*. This background material is important because of the need to focus on *invariant subspaces* rather than eigenvectors.

When a symmetric matrix M has an eigenvalue ω of multiplicity k, there exist k linearly independent eigenvectors corresponding to ω; it is customary to assume that these eigenvectors are orthonormal. Although the eigenvectors associated with ω are not unique, they span a unique invariant subspace of dimension k. (See Chapter V of [SS90] for a general discussion of invariant subspaces.) When a matrix has "close" eigenvalues, individual eigenvectors are ill-conditioned, but the associated invariant subspaces are often well conditioned. Thus, if the eigenvalues of a particular matrix are not known *a priori* to be well separated, it is helpful to focus on invariant subspaces rather than eigenvectors.

The following lemma gives various properties of the matrices $A^T D$ and $A^T D^2 A$, where A has full row rank and D is a positive diagonal matrix. For related material, see, for example, [GV89] or Chapter I of [SS90]. Unless specified otherwise, we always use the two-norm.

Lemma 2.1. *We are given:*

(i) A, an $\ell \times n$ matrix of rank ℓ, with singular values $\sigma_1(A) \geq \cdots \geq \sigma_\ell(A) > 0$; and

(ii) D, an $\ell \times \ell$ diagonal matrix $\mathrm{diag}(d_j)$ with $d_j > 0$, $j = 1, \ldots, \ell$. (The singular values of D, denoted by $\sigma_1(D) \geq \cdots \geq \sigma_\ell(D) > 0$, are the sorted elements of d.)

Let the singular values of $A^T D$ be denoted by $\tau_1 \geq \cdots \geq \tau_\ell$, and the eigenvalues of $A^T D^2 A$ by $\omega_1 \geq \cdots \geq \omega_n$, with corresponding eigenvectors $\{u_i\}$, $i = 1, \ldots, n$. Then

1. *$A^T D$ and $A^T D^2 A$ have rank ℓ;*
2. *$\sigma_i(D)\sigma_\ell(A^T) \leq \tau_i \leq \sigma_i(D)\sigma_1(A^T)$, $i = 1, \ldots, \ell$;*
3. *$\sigma_i(A^T)\sigma_\ell(D) \leq \tau_i \leq \sigma_i(A^T)\sigma_1(D)$, $i = 1, \ldots, \ell$;*
4. *For $i = 1, \ldots, \ell$, $\omega_i = \tau_i^2$ and $u_i = A^T v_i$ for some ℓ-vector v_i;*
5. *For $i = \ell + 1, \ldots, n$, $\omega_i = 0$ and $u_i = Z v_i$ for some $(n - \ell)$-vector v_i, where Z denotes an $n \times (n - \ell)$ matrix whose columns form a basis for the null space of A.*

Proof. The first result is straightforward. Results 2 and 3 follow from Theorem I.4.5 of [SS90]. The statements in results 4 and 5 about the eigenvalues of $A^T D^2 A$ follow from the definition of singular values; the statements about eigenvectors follow from properties of the range space of A^T and the null space of A. $\square$

The following result about interlacing of eigenvalues is Corollary IV.4.4 from [SS90].

Lemma 2.2. *Let U be an $n \times \ell$ matrix with orthonormal columns, and M an $n \times n$ symmetric matrix with eigenvalues $\omega_1 \geq \cdots \geq \omega_n$. Denote the eigenvalues of $U^T M U$ by $\beta_1 \geq \cdots \geq \beta_\ell$. Then*

$$\omega_i \geq \beta_i \geq \omega_{i+n-\ell}, \qquad i = 1, 2, \ldots, \ell.$$

The next result about the perturbation of eigenvalues of symmetric matrices is Corollary IV.4.10 from [SS90].

Lemma 2.3. *Let M and $\tilde{M}$ be $n \times n$ symmetric matrices, with eigenvalues $\{\omega_i\}$ and $\{\tilde{\omega}_i\}$ respectively. Then*

$$\max\{|\tilde{\omega}_i - \omega_i|\} \leq \|\tilde{M} - M\|_2.$$

Mirsky's Theorem (Theorem IV.4.11 from [SS90]) involves the singular values of perturbed matrices.

Theorem 2.1. *(Mirsky's Theorem.) Let X and $\tilde{X}$ be matrices of the same dimensions with singular values $\sigma_1 \geq \cdots \geq \sigma_p$ and $\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_p$, respectively. Then, for any unitarily invariant norm,*

$$\|\,\mathrm{diag}(\tilde{\sigma}_i - \sigma_i)\| \leq \|\tilde{X} - X\|.$$

For real symmetric matrices M and N, the function $\mathrm{sep}(M, N)$ measures the smallest distance between their eigenvalues. For general matrices, a much broader definition of sep is given in [SS90]; see Theorem V.3.1 of [SS90] for a proof that the following definition is valid.

Definition 2.1. *(The* sep *function.) Given real symmetric matrices M and N with eigenvalues $\{\omega_i\}$ and $\{\eta_j\}$ respectively,*

$$\mathrm{sep}(M, N) = \min_{i,j} |\omega_i - \eta_j|.$$

An invariant subspace is formally defined as follows.

Definition 2.2. *(Invariant subspace.) Given a matrix M, the subspace $\mathcal{X}$ is an invariant subspace of M if $Mx \in \mathcal{X}$ for any vector $x \in \mathcal{X}$.*

For real symmetric M, the following result (a specialization of Theorem V.1.1 of [SS90]) is useful.

Theorem 2.2. *Let M be an $n \times n$ real symmetric matrix. Assume that the $n \times \ell$ matrix X_1 has linearly independent columns, and that the $n \times (n - \ell)$ matrix X_2 has linearly independent columns that span the null space of X_1^T. Then the range spaces of X_1 and X_2 are* invariant subspaces *of M if and only if*

$$(X_1 \;\; X_2)^T M (X_1 \;\; X_2) = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}, \tag{2.1}$$

where L_1 is $\ell \times \ell$.

Relation (2.1) implies that

$$M = X_1 L_1 X_1^T + X_2 L_2 X_2^T. \tag{2.2}$$

Either form (2.1) or (2.2) is called the *spectral resolution* of M with respect to X_1 and X_2.

Finally, we define a *simple* invariant subspace.

Definition 2.3. *(Simple invariant subspace.) Given an $n \times n$ real symmetric matrix M and an invariant subspace $\mathcal{X}$, let X_1 be an $n \times \ell$ matrix whose columns are orthonormal and span $\mathcal{X}$. Assume that the $n \times (n - \ell)$ matrix X_2 has orthonormal columns and spans the null space of X_1^T, and that (2.1) holds. Then $\mathcal{X}$ is a* simple *invariant subspace of M if L_1 and L_2 have no common eigenvalues.*

It follows from these definitions that, given a symmetric matrix with an eigenvalue of exact multiplicity ℓ, the associated ℓ orthonormal eigenvectors define a simple invariant subspace. We stress, however, that invariant subspaces can have a more general form.

For a given symmetric matrix M and an orthonormal set of vectors that span a simple invariant subspace, the following result (a specialization of Theorem V.2.7 of [SS90]) gives conditions under which the same set of vectors is close to an orthonormal basis for an invariant subspace of a *perturbed* matrix $\tilde{M} = M + E$. When the invariant subspace corresponds to a particular eigenvalue ω, the result holds if the separation of ω from other eigenvalues of M is sufficiently large compared to portions of a transformed version of E.

Theorem 2.3. *Let $(X_1 \;\; X_2)$ be orthogonal and suppose that range(X_1) is a simple invariant subspace of the real symmetric matrix M, with spectral resolution (2.1). Given a symmetric perturbation matrix E, let*

$$\tilde{M} = M + E \quad \text{and} \quad (X_1 \;\; X_2)^T E (X_1 \;\; X_2) = \begin{pmatrix} E_{11} & E_{12} \\ E_{12}^T & E_{22} \end{pmatrix}.$$

We define

$$\tilde{\gamma} = \|E_{12}\|_F, \quad \tilde{\delta} = \mathrm{sep}(L_1, L_2) - \|E_{11}\|_F - \|E_{22}\|_F \quad \text{and} \quad \beta = \tilde{\gamma}/\tilde{\delta}.$$

Assume that $\tilde{\delta} > 0$ and $\beta < \frac{1}{2}$. Then there is a unique matrix P satisfying $\|P\|_F < 2\beta$ for which there are matrices $\tilde{X}_1$ and $\tilde{X}_2$, defined by

$$\tilde{X}_1 = (X_1 + X_2 P)(I + P^T P)^{-\frac{1}{2}} \quad \text{and} \quad \tilde{X}_2 = (X_2 - X_1 P^T)(I + P P^T)^{-\frac{1}{2}}, \tag{2.3}$$

such that the columns of $\tilde{X}_1$ and $\tilde{X}_2$ form orthonormal bases for simple invariant subspaces of $\tilde{M}$.

The following corollary is included because its form is convenient for our later treatment of the barrier Hessian.

Corollary 2.1. *Under the assumptions of Theorem 2.3, there are orthonormal bases* $\tilde{X}_1$ *and* $\tilde{X}_2$ *for simple invariant subspaces of* $\tilde{M}$ *satisfying*

$$\|X_1 - \tilde{X}_1\| = O(\beta) \quad \text{and} \quad \|X_2 - \tilde{X}_2\| = O(\beta). \tag{2.4}$$

The following is Theorem V.3.3 of [SS90], specialized to the case of a real matrix M. It is assumed that the eigenvalues $\{\omega_i\}$ of an $n \times n$ symmetric matrix are ordered so that $\omega_1 \geq \cdots \geq \omega_n$.

Theorem 2.4. *With the same assumptions and definitions as in Theorem 2.3, let* $\{\breve{\omega}_i\}$ *be the eigenvalues of* $X_1^T \tilde{M} X_1$*, where* X_1 *has* ℓ *columns. Then, if* $\tilde{\delta} > 0$ *and* $\beta < \frac{1}{2}$*, there are eigenvalues* $\tilde{\omega}_{j_1}, \ldots, \tilde{\omega}_{j_\ell}$ *of* $\tilde{M}$ *satisfying*

$$\sum_{i=1}^{\ell} (\tilde{\omega}_{j_i} - \breve{\omega}_i)^2 \;\leq\; \Big(2\beta\tilde{\gamma}(1 + 2\,\beta^2)\Big)^2.$$

3. Properties of the Barrier Hessian

If we wish to minimize a barrier function using a Newton-based method, the barrier Hessian H_B (1.4) appears in the Newton equations at the current iterate x:

$$H_B p_{\text{newt}} = -g_B. \tag{3.1}$$

When H_B satisfies certain conditions (e.g., is sufficiently positive definite), the next iterate is given as $x + \alpha p_{\text{newt}}$, where the *Newton step* p_{newt} is the solution of (3.1) and α is a positive step chosen to produce a sufficient decrease in a merit function—most often, the barrier function. For dense problems, the Newton equations are typically solved via the *Cholesky factorization*, in which H_B is decomposed into the form $R^T R$, where R is an upper-triangular matrix; see [DS83], [Fle87] and [GMW81] for discussions of Newton-based methods. When the Hessian is not positive definite, various Cholesky-based techniques can be applied to create modified Newton methods; see [GMW81] and [SE90]. The Cholesky factorization will be examined with specific reference to the barrier Hessian in Section 6.

Near x^*, H_B has a special structure with important algorithmic implications. In analyzing its behavior, we express separately the contributions of active and inactive constraints, and write H_B in the form

$$H_B = V + \mu \hat{A}^T \hat{C}^{-2} \hat{A} + \mu S, \tag{3.2}$$

with

$$V = H - \sum_{j=1}^{m} \frac{\mu}{c_j} H_j \quad \text{and} \quad S = \bar{A}^T \bar{C}^{-2} \bar{A}, \tag{3.3}$$

where all functions are evaluated at x.

Let the m-vector $d(x, \mu)$ and the $m \times m$ positive diagonal matrix D be given by

$$d_j(x, \mu) = \frac{\mu}{c_j(x)} \quad \text{and} \quad D(x, \mu) = \operatorname{diag}(d_j), \quad j = 1, \ldots, m. \tag{3.4}$$

We shall omit the arguments of d and D when they are clear from context. The vector $\hat{d}$ denotes the $\hat{m}$-dimensional subvector of d associated with the active constraints. The $\hat{m} \times \hat{m}$ diagonal matrix $\hat{D}$ is defined as $\operatorname{diag}(\hat{d}_j)$, and satisfies

$$\frac{1}{\mu} \hat{A}^T \hat{D}^2 \hat{A} = \mu \hat{A}^T \hat{C}^{-2} \hat{A}, \quad \text{so that} \quad H_B = \frac{1}{\mu} \hat{A}^T \hat{D}^2 \hat{A} + V + \mu S.$$

3.1. Near the solution

The Hessian matrices of penalty and barrier functions at points on the barrier trajectory were originally analyzed in [Loot69] and [Mur71], and were shown to be ill-conditioned if $0 < \hat{m} < n$. We wish to analyze the barrier Hessian at more general points, to determine whether there is a pattern of ill-conditioning that might affect the performance of Newton's method. This analysis is motivated by the "approximate" nature of modern barrier algorithms, which tend to perform only a few Newton iterations (perhaps just one) for each value of μ, and then to reduce μ to move along the trajectory, closer to x^*. Hence exact minimizers of the barrier function are unlikely to occur often.

It is crucial to distinguish the phenomenon of systematic ill-conditioning of the Hessian from *occasional* (unstructured) ill-conditioning. When (as we shall see) the Hessian is known to display ill-conditioning with a predictable structure in an entire region near the solution, it seems advisable to devise specialized techniques to treat this ill-conditioning; see Section 6.

We assume that, for some positive δ, the Hessian is evaluated at any point x_δ satisfying

$$\|x_\delta - x^*\| \le \delta \quad \text{and} \quad c(x_\delta) > 0. \tag{3.5}$$

As the iterates converge, δ will decrease. Our interest is in small values of δ, so that $\delta \ll 1$.

The behavior of H_B at x_δ obviously depends in part on the relationship between δ and μ, the barrier parameter that appears in the Hessian. Broadly speaking, we expect the Newton step p_{newt} (see (3.1)) to move from x_δ *toward* the barrier trajectory, namely toward x_μ for the given value μ. It is consequently assumed that the distance from x_δ to x^* is *larger* than the distance from x^* to x_μ, i.e. that

$$\|x_\delta - x^*\| \ge \|x_\mu - x^*\|. \tag{3.6}$$

To quantify these distances, we define κ_μ as

$$\|x_\mu - x^*\|_2 = \kappa_\mu \mu. \tag{3.7}$$

Under assumptions (i)–(v) of Section 1.2, κ_μ is $\Theta(1)$, i.e. for sufficiently small μ there exist constants η_l and η_u such that

$$0 < \eta_l \le \kappa_\mu < \eta_u < \infty. \tag{3.8}$$

The property (3.8) can be derived from the nontangential approach of the barrier trajectory to x^* under the given assumptions; see, for example, [FM68], [Fi79] and [Wr92]. We stress that relation (3.8) need not hold unless *all* of the assumptions (i)–(v) are satisfied. For example, consider minimizing $\frac{1}{2}x^2$ subject to $x \ge 0$. The unique solution is $x^* = 0$, with $\lambda^* = 0$, so that strict complementarity does not hold. The unique minimizer of the barrier function is $x_\mu = \sqrt{\mu}$, and there are no positive constants η_u and η_l such that (3.8) is satisfied for all sufficiently small μ.

Relations (3.6) and (3.7) together imply $\kappa_\mu \mu \le \delta$, so that $\mu/\delta \le 1/\kappa_\mu$. Because $\kappa_\mu \ge \eta_l$ (see (3.8)),

$$\frac{\mu}{\delta} \le \frac{1}{\eta_l}, \quad \text{which means that} \quad \frac{\mu}{\delta} = O(1) \quad \text{or} \quad \mu = O(\delta). \tag{3.9}$$

Relation (3.5) and smoothness of f and $\{c_j\}$ imply that

$$|f(x_\mu) - f(x^*)| = O(\delta), \qquad |c_j(x_\mu) - c_j(x^*)| = O(\delta), \tag{3.10a}$$
$$\|g(x_\mu) - g(x^*)\| = O(\delta), \qquad \|a_j(x_\mu) - a_j(x^*)\| = O(\delta), \tag{3.10b}$$
$$\|H(x_\mu) - H(x^*)\| = O(\delta) \quad \text{and} \quad \|H_j(x_\mu) - H_j(x^*)\| = O(\delta), \tag{3.10c}$$

for $j = 1, \ldots, m$. For sufficiently small δ, the inactive constraints are bounded away from zero at x_δ, which means that there exists $\bar{\kappa}$ such that

$$c_j(x_\delta) \ge \bar{\kappa}, \quad j \notin \mathcal{A}. \tag{3.11}$$

We define

$$d_{\max}(x_\delta,\mu) \;=\; \max_j \hat{d}_j(x_\delta,\mu) = \frac{\mu}{\min\, \hat{c}_j(x_\delta)} \quad \text{and} \tag{3.12a}$$

$$d_{\min}(x_\delta,\mu) \;=\; \min_j \hat{d}_j(x_\delta,\mu) = \frac{\mu}{\max\, \hat{c}_j(x_\delta)}. \tag{3.12b}$$

The following lemma states properties of the matrix $\hat{A}^T\hat{D}^2\hat{A}$ evaluated at (x_δ,μ).

Lemma 3.1. *Assume that $\hat{m} > 0$, and consider a point x_δ satisfying (3.5). Denote the singular values of $\hat{A}(x_\delta)$ by $\alpha_1 \ge \cdots \ge \alpha_{\hat{m}}$. (The singular values of $\hat{D}(x_\delta,\mu)$ are its sorted diagonal elements.) Then, for sufficiently small δ and μ satisfying (3.9),*

1. *$\hat{A}^T\hat{D}$ and $\hat{A}^T\hat{D}^2\hat{A}$ have rank $\hat{m}$;*
2. *There exist positive constants η'_α and η_α such that $\alpha_1 < \eta'_\alpha$ and $\alpha_{\hat{m}} > \eta_\alpha$;*
3. *There exists a positive constant η_d such that $d_{\min} \ge \eta_d$;*
4. *$\hat{A}^T\hat{D}^2\hat{A}$ has $\hat{m}$ positive eigenvalues $\omega_1 \ge \cdots \ge \omega_{\hat{m}}$, where ω_1 and $\omega_{\hat{m}}$ satisfy*

$$\omega_1 \;\ge\; d_{\max}^2\alpha_{\hat{m}}^2 \quad \text{and} \quad \omega_{\hat{m}} \;\ge\; d_{\min}^2\alpha_{\hat{m}}^2; \tag{3.13}$$

5. *The eigenvectors corresponding to the positive eigenvalues of $\hat{A}^T\hat{D}^2\hat{A}$ lie in range$(\hat{A}^T)$;*
6. *$\hat{A}^T\hat{D}^2\hat{A}$ has $n-\hat{m}$ zero eigenvalues, and the associated eigenvectors lie in the null space of $\hat{A}$.*

Proof. Since $\hat{A}(x^*)$ has full rank, the matrix $\hat{A}(x)$ must also have full rank for all x sufficiently close to x^*. Because x_δ is strictly feasible, the diagonal elements of $\hat{D}(x_\delta,\mu)$ are positive. These two properties imply the first result.

By smoothness of $\hat{A}$ and Mirsky's Theorem, every singular value of $\hat{A}(x)$ is $\Theta(1)$ for all x sufficiently close to x^*, which is the second result.

From (3.10a), the *largest* possible value of any active constraint at x_δ is a constant times δ. Since μ/δ is $O(1)$ (see (3.9)), it follows from (3.12b) that

$$d_{\min}(x_\delta,\mu) = \Omega(1), \tag{3.14}$$

which gives the third result. In contrast, the *smallest* active constraint at a general point x_δ can be arbitrarily small, so that there is no upper bound on the largest singular value $d_{\max}$ of $\hat{D}(x_\delta,\mu)$.

The lower bounds on ω_1 and $\omega_{\hat{m}}$ follow from Lemma 2.1.

The final two results are immediate from properties of the range space of $\hat{A}^T$. □

Next, we bound the difference between μH_B and $\hat{A}^T\hat{D}^2\hat{A}$ (Lemma 3.2), and then the difference between their eigenvalues (Lemma 3.3).

Lemma 3.2. *For x_δ satisfying (3.5), and sufficiently small δ and μ satisfying (3.9), it holds that*

$$\|\mu H_B - \hat{A}^T\hat{D}^2\hat{A}\| = d_{\max}O(\mu).$$

(For compactness, we abuse notation slightly by using $O(\mu)$ to denote a function that is $O(\mu)$.)

Proof. Recall that

$$\mu H_B - \hat{A}^T\hat{D}^2\hat{A} = \mu(H - \sum_{j=1}^{m} d_j H_j) + \mu^2\bar{A}^T\bar{C}^{-2}\bar{A} = \mu V + \mu^2 S,$$

where V and S are defined by (3.3). At x_δ for sufficiently small δ, $\|H\| = O(1)$ and $\|H_j\| = O(1)$ for $j = 1, \ldots, m$. At x_δ, we have

$$\|V\| \le \|H\| + \sum_{j=1}^{m} |d_j|\,\|H_j\| \le \|H\| + d_{\max}\sum_{j=1}^{m} \|H_j\|.$$

In the proof of Lemma 3.1, it is shown that $d_{\min}$ is bounded below by a constant (see (3.14)), which implies that $d_{\max}$ is similarly bounded below. Hence we may write

$$\|V\| = (1 + d_{\max})O(1) \quad \text{or} \quad \|V\| = d_{\max}O(1). \tag{3.15}$$

For sufficiently small δ, the inactive constraints at x_δ are bounded away from zero (see (3.11)), so that $\|S\|$ is $O(1)$. The desired result follows from these properties. $\square$

Lemma 3.3. *Let $\omega_1 \geq \cdots \geq \omega_m$ denote the eigenvalues of $\hat{A}^T\hat{D}^2\hat{A}$ evaluated at (x_δ, μ) satisfying (3.5) and (3.9). For sufficiently small μ and δ, the eigenvalues of $\mu H_B(x_\delta, \mu)$, denoted by $\zeta_1 \geq \cdots \geq \zeta_n$, satisfy*

$$|\zeta_j - \omega_j| = d_{\max}O(\mu), \quad j = 1, \ldots, m.$$

Proof. The result is immediate from Lemmas 3.2 and 2.3. $\square$

Putting these results together, we obtain the following theorem, which states that the barrier Hessian is ill-conditioned in a region near x^*.

Theorem 3.1. *Assume that $0 < \hat{m} < n$. For x_δ satisfying (3.5), and sufficiently small δ and μ satisfying (3.9), it holds that*

$$\operatorname{cond}(\mu H_B) = \operatorname{cond}(H_B) = d_{\max}\, \Omega\Big(\frac{1}{\mu}\Big).$$

Proof. We know from Lemma 3.1 that, for sufficiently small μ and δ, $\hat{A}^T\hat{D}^2\hat{A}$ has $\hat{m}$ positive eigenvalues, where the largest positive eigenvalue ω_1 is bounded below by $d_{\max}^2$ times a positive constant. By assumption, $\hat{m} \geq 1$; relation (3.14) shows that $d_{\max}$ is bounded below by a constant. Thus the closeness of ζ_1 and ω_1 asserted in Lemma 3.3 implies that, for sufficiently small μ, the largest eigenvalue ζ_1 of μH_B is positive and bounded below by $d_{\max}^2$ times a constant. Since $\hat{A}^T\hat{D}^2\hat{A}$ has $n - \hat{m}$ zero eigenvalues and $\hat{m} < n$ by assumption, Lemma 3.3 also implies that the magnitude of the *smallest* eigenvalue ζ_n of μH_B is bounded above by $\mu d_{\max}$ times a constant. It follows that

$$\operatorname{cond}(\mu H_B) = \operatorname{cond}(H_B) = \frac{|\zeta_1|}{|\zeta_n|} = d_{\max}\, \Omega\Big(\frac{1}{\mu}\Big),$$

which is the desired result. $\square$

This result generalizes the well known property derived by [Mur71] and [Loot69] that the barrier Hessian matrices at x_μ become increasingly ill-conditioned as μ goes to zero, and are singular in the limit. We see that the barrier Hessian is ill-conditioned not only on (or near) the trajectory, but also in an entire region close to the solution.

The theorem excludes the special cases $\hat{m} = 0$ and $\hat{m} = n$. If $\hat{m} = 0$, so that no constraints are active, H_B simply converges to $H(x^*)$. If $\hat{m} = n$, then *all* eigenvalues of μH_B are within $O(\mu)$ of the positive eigenvalues of $\hat{A}^T\hat{D}^2\hat{A}$, so that $\operatorname{cond}(H_B) \approx \operatorname{cond}(\hat{A}^T\hat{D}^2\hat{A})$. If we consider only points on the barrier trajectory when $\hat{m} = n$ (as in a nondegenerate linear program), $\hat{A}^T\hat{D}^2\hat{A}$ is converging to $\hat{A}^{*T}(\hat{\Lambda}^*)^2\hat{A}^*$, whose condition number is finite; in this case, although the barrier Hessian is becoming unbounded in norm, it is *not* necessarily ill-conditioned.

The following relationship between the barrier Hessian and the reduced Hessian matrix Z^TVZ is sometimes useful, and is stated formally for completeness.

Lemma 3.4. *Given any strictly feasible point x and any barrier parameter $\mu > 0$, let the eigenvalues of Z^TVZ evaluated at (x, μ) be denoted by $\nu_1 \geq \cdots \geq \nu_{n-\hat{m}}$. Then the eigenvalues of $H_B(x_\delta, \mu)$, denoted by $\xi_1 \geq \cdots \geq \xi_n$, satisfy*

$$\xi_k \geq \nu_k \geq \xi_{k+n-\hat{m}}, \quad k = 1, \ldots, n - \hat{m}.$$

Proof. The result follows immediately from the interlacing property given in Lemma 2.2. It is worth stressing that, for *any* x and μ, the smallest eigenvalue of the reduced Hessian matrix Z^TVZ evaluated at (x, μ) is always an upper bound on the smallest eigenvalue of the barrier Hessian. $\square$

3.2. Not close to the boundary

To establish more about H_B than the simple ill-conditioning shown by Theorem 3.1, we need to make further assumptions about x_δ. If $\hat{m} > 0$ and we are prepared to limit attention to points x_δ for which all active constraints are bounded below by a constant times μ, then $d_{\max}$ (which can be arbitrarily large at a general point x_δ) is known to be $\Theta(1)$. If $d_{\max}$ is suitably bounded, the invariant subspaces of H_B can be characterized in terms of the null space of $\hat{A}^T$.

Formally, we consider x_δ for all sufficiently small μ and δ satisfying (3.5) and (3.9), such that

$$\min_{i \in \mathcal{A}} c_i(x_\delta) = \Theta(\mu), \quad \text{or, equivalently,} \quad d_{\max}(x_\delta, \mu) = \Theta(1). \tag{3.16}$$

This relation excludes points very close to the boundary of the strictly feasible region, i.e. points where at least one of the active constraints is "too small" relative to μ. In a barrier algorithm, the iterates should remain "centered", which means that points with inordinately small constraint values tend not to occur as iterates (except, by chance, as an initial point). Broadly speaking, a point very close to the boundary is unlikely to qualify as a new iterate because of its failure to produce a sufficient decrease in the barrier function.

The following theorem shows that for suitable points near x^*, the barrier Hessian has invariant subspaces close to those spanned by $Z(x_\delta)$ and its orthogonal complement, and close to those spanned by Z^* and its orthogonal complement Y^*.

Theorem 3.2. *(Invariant subspaces of $\hat{A}^T$ and H_B.) Consider the set of points x_δ satisfying (3.5), (3.9) and (3.16) for all sufficiently small δ and μ. Let $(Z \quad Y)$ be an orthogonal matrix, where the columns of Z span the null space of $\hat{A}(x_\delta)$. Then the barrier Hessian H_B evaluated at (x_δ, μ) has simple invariant subspaces that are "close to" those defined by Z and Y, and to those defined by Z^* and Y^*, in the sense that there are matrices $\tilde{Z}$ and $\tilde{Y}$ satisfying*

$$\|Z - \tilde{Z}\| = O(\mu) \quad \text{and} \quad \|Y - \tilde{Y}\| = O(\mu); \tag{3.17a}$$

$$\|Z^* - \tilde{Z}\| = O(\delta) \quad \text{and} \quad \|Y^* - \tilde{Y}\| = O(\delta), \tag{3.17b}$$

such that the columns of $\tilde{Z}$ and $\tilde{Y}$ form orthonormal bases for simple invariant subspaces of H_B.

Proof. To avoid the appearance of factors of $(1/\mu)$, we consider initially the matrix $\hat{A}^T\hat{D}^2\hat{A}$. (Invariant subspaces are unaffected by multiplication by a scalar.) We know from Lemma 3.1 that the matrix Z forms an orthonormal basis for the subspace of eigenvectors associated with zero eigenvalues of $\hat{A}^T\hat{D}^2\hat{A}$. The spectral resolution (2.1) of $\hat{A}^T\hat{D}^2\hat{A}$ is

$$(Z \quad Y)^T \, \hat{A}^T\hat{D}^2\hat{A} \, (Z \quad Y) = \begin{pmatrix} 0 & 0 \\ 0 & Y^T\hat{A}^T\hat{D}^2\hat{A}Y \end{pmatrix},$$

so that

$$L_1 = 0 \quad \text{and} \quad L_2 = Y^T\hat{A}^T\hat{D}^2\hat{A}Y. \tag{3.18}$$

We now consider the matrix L_2 of (3.18). The matrices Y and Z are orthogonal complements, which implies that the columns of Y form an orthonormal basis for range($\hat{A}^T$). We know from part 4 of Lemma 2.1 that the positive eigenvalues of $\hat{A}^T\hat{D}^2\hat{A}$ are equal to the squared singular values of $\hat{A}^T\hat{D}$, with corresponding eigenvectors that lie in range($\hat{A}^T$). Because $d_{\min}$ and $d_{\max}$ are $\Theta(1)$ (see (3.14) and (3.16)), it follows from parts 2 and 3 of Lemma 2.1 that all singular values of $\hat{A}^T\hat{D}$ are $\Theta(1)$. The *smallest* eigenvalue of the $\hat{m} \times \hat{m}$ matrix $Y^T\hat{A}^T\hat{D}^2\hat{A}Y$ is consequently positive and $\Theta(1)$. Because all eigenvalues of L_1 are zero, L_1 and L_2 have no common eigenvalues, and Definition 2.3 shows that range(Z) is a simple invariant subspace of $\hat{A}^T\hat{D}^2\hat{A}$.

Turning to the scaled matrix $(1/\mu)\hat{A}^T\hat{D}^2\hat{A}$, we invoke Theorem 2.3 with

$$X_1 = Z, \quad X_2 = Y, \quad M = \frac{1}{\mu}\,\hat{A}^T\hat{D}^2\hat{A} \quad \text{and} \quad \tilde{M} = H_B.$$

Writing $\tilde{M} = M + E$, the matrix E is $V + \mu S$, where V and S are defined by (3.3). The proof of Lemma 3.2 shows that $\|E\| = d_{\max} O(1)$. Because x_δ satisfies the new restriction (3.16), $d_{\max}$ is bounded above by a constant, so that $\|E\| = O(1)$.

We now examine the values $\tilde{\gamma}$, $\tilde{\delta}$ and β of Theorem 2.3 for the scaled matrix $(1/\mu)\hat{A}^T\hat{D}^2\hat{A}$. Because $\|E\| = O(1)$ and $\tilde{\gamma} = \|E_{12}\|_F$, it follows that $\tilde{\gamma} = O(1)$. The quantity $\tilde{\delta}$ is defined as

$$\tilde{\delta} = \frac{1}{\mu}\,\mathrm{sep}(L_1, L_2) - \|E_{11}\|_F - \|E_{22}\|_F,$$

where L_1 and L_2 are given by (3.18). Because $L_1 = 0$, Definition 2.1 shows that $\mathrm{sep}(L_1, L_2)$ is the smallest eigenvalue of $Y^T\hat{A}^T\hat{D}^2\hat{A}Y$, which was shown previously to be $\Theta(1)$. Since $\|E_{11}\|_F$ and $\|E_{22}\|_F$ are $O(1)$, it follows that, for sufficiently small μ, $\tilde{\delta}$ is both positive and $\Theta(1/\mu)$.

Finally, β is defined as $\tilde{\gamma}/\tilde{\delta}$. Since $\tilde{\gamma} = O(1)$ and $\tilde{\delta}$ is $\Theta(1/\mu)$, β is $O(\mu)$ for sufficiently small μ. Corollary 2.1 implies the existence of $\tilde{Z}$ and $\tilde{Y}$, orthonormal bases for simple invariant subspaces of H_B, satisfying

$$\|Z - \tilde{Z}\| = O(\mu) \quad \text{and} \quad \|Y - \tilde{Y}\| = O(\mu),$$

which is the first desired result.

The matrix $\hat{A}(x)$ has full rank in a sufficiently small neighborhood of x^*. If $Z(x)$ is always chosen as a suitable smooth function, then $\|Z^* - Z(x_\delta)\| = O(\delta)$, and similarly for Y^* and $Y(x_\delta)$. Since $\mu = O(\delta)$ (see (3.9)), the second result (3.17b) follows immediately. □

An interpretation of Theorem 3.2 that will be useful in Section 6 is that the invariant subspace associated with the unbounded ("large") eigenvalues of the barrier Hessian is close to the range space of $\hat{A}^T$, and the invariant subspace associated with the bounded ("small") eigenvalues is close to the null space of $\hat{A}$. For the scaled matrix μH_B, one might say informally that its "small space" (the subspace of vectors v for which $\|\mu H_B v\|$ is small relative to $\|v\|$), is close to the null space of $\hat{A}$. Similarly, the "almost range space" of μH_B is close to the range space of $\hat{A}^T$. We stress that no general bound is given on the closeness of individual eigenvectors, since for nearly-multiple eigenvalues the corresponding eigenvectors become ill-conditioned (see Section 2).

The following theorem shows that H_B has $n - \hat{m}$ eigenvalues close to those of the reduced Hessian matrix Z^TVZ evaluated at (x_δ, μ). If we focus on the barrier minimizers $\{x_\mu\}$, $n - \hat{m}$ eigenvalues of $H_B(x_\mu, \mu)$ are close to those of the reduced Hessian of the Lagrangian function evaluated at (x^*, λ^*); this latter result was given in [Mur71] and [Loot69].

Theorem 3.3. *(Eigenvalues of H_B, Z^TVZ and $Z^{*T}W^*Z^*$.) Consider the set of points x_δ satisfying (3.5), (3.9) and (3.16) for all sufficiently small δ and μ. Let $Z(x)$ denote a matrix whose columns form an orthonormal basis for the null space of $\hat{A}(x)$. Then H_B has $n - \hat{m}$ eigenvalues, denoted by $\xi_{\hat{m}+1}, \ldots, \xi_n$, satisfying*

$$|\xi_{\hat{m}+k} - \nu_k| = O(\mu), \quad k = 1, \ldots, n - \hat{m}, \tag{3.19}$$

where $\{\nu_k\}$ are the eigenvalues of Z^TVZ evaluated at (x_δ, μ).

For the set of points $\{x_\mu\}$ (the minimizers of the barrier function), it holds that

$$|\xi_{\hat{m}+k} - \nu_k^*| = O(\mu), \quad k = 1, \ldots, n - \hat{m}, \tag{3.20}$$

where $\{\nu_k^*\}$ are the eigenvalues of $Z^{*T}W^*Z^*$, the reduced Hessian of the Lagrangian function at x^*.

Proof. First, we follow the proof of Theorem 3.2, with

$$M = \frac{1}{\mu}\hat{A}^T\hat{D}^2\hat{A}, \quad \tilde{M} = H_B, \quad \text{and} \quad \|E\| = O(1).$$

The matrix X_1 is again taken as Z, an orthonormal basis for the null space of $\hat{A}$. The matrix X_2 is Y, an orthonormal basis for the range space of $\hat{A}^T$. Exactly as in the proof of Theorem 3.2, we have that $\tilde{\gamma} = O(1)$, $\tilde{\delta} > 0$, and β is $O(\mu)$.

Let $\{\breve{\nu}_i\}$, $i = 1, \ldots, n - \hat{m}$, denote the eigenvalues of $Z^T H_B Z$. Theorem 2.4 implies that there are $n - \hat{m}$ eigenvalues $\{\xi_{j_k}\}$, $k = 1, \ldots, n - \hat{m}$, of H_B satisfying

$$\sum_{k=1}^{n-\hat{m}} (\xi_{j_k} - \breve{\nu}_k)^2 \le \left(2\beta\tilde{\gamma}(1 + 2\beta^2)\right)^2,$$

which means that

$$|\xi_{j_k} - \breve{\nu}_k| = O(\mu), \quad k = 1, \ldots, n - \hat{m}. \tag{3.21}$$

For each eigenvalue $\breve{\nu}_k$ of $Z^T H_B Z$, there is thus a "close" eigenvalue of H_B.

Because Z has orthonormal columns and $\|S\| = O(1)$, the matrices $Z^T H_B Z$ and $Z^T V Z$ satisfy

$$Z^T H_B Z = Z^T V Z + \mu Z^T S Z = Z^T V Z + O(\mu).$$

Let $\{\nu_k\}$, $k = 1, \ldots, n - \hat{m}$, denote the eigenvalues of $Z^T V Z$. Lemma 3.2 implies that

$$|\breve{\nu}_k - \nu_k| \le O(\mu), \quad k = 1, \ldots, n - \hat{m}, \tag{3.22}$$

which gives the first result of the theorem.

Now we consider the special properties of x_μ. Since $Z(x)$ has been chosen as a smooth function of x and $\hat{A}(x)$ has full rank in a neighborhood of x^*, (3.7) implies that

$$\|Z(x_\mu) - Z^*\| = O(\mu).$$

Because of (3.7), the full rank of $\hat{A}(x^*)$, and relation (1.5), the value of $d_j(x_\mu)$ is within $O(\mu)$ of λ_j^*. We conclude that

$$\|Z^{*T} W^* Z^* - Z^T V Z\| = O(\mu). \tag{3.23}$$

Applying Lemma 3.2 once again, and combining the results (3.21), (3.22) and (3.23), we obtain the second conclusion of the theorem. □

3.3. Numerical examples

To illustrate properties of the barrier Hessian, we consider a four-variable example:

$$\begin{array}{ll} \text{minimize} & x_1 + x_2 + x_3^2 + x_4^2 \\ \text{subject to} & \frac{1}{2}x_1^2 + (x_1 - x_2)^2 + (x_1 - 1)x_4 - \frac{1}{2} \ge 0 \\ & x_1^2 + x_2^2 + x_3 + 2x_4 - \frac{1}{2} \ge 0 \\ & \frac{1}{2}x_2^2 + 2x_3^3 - \frac{1}{2} \ge 0 \\ & x_1 \ge 0 \\ & x_2 \ge 0. \end{array} \tag{3.24}$$

The optimal solution is $x^* = (1, 1, 0, 0)^T$, where the first and third constraints are active, with Lagrange multiplier vector $\lambda^* = (1, 0, 1, 0, 0)^T$. This example has been constructed so that both matrices $\hat{A}^{*T}(\hat{A}^*)^2\hat{A}^*$ and $Z^{*T} W^* Z^*$ have multiple eigenvalues.

At x^*, we have

$$\hat{A}^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad \hat{A}^{*T}(\hat{A}^*)^2\hat{A}^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \tag{3.25}$$

$$W^* = \begin{pmatrix} -3 & 2 & 0 & -1 \\ 2 & -3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ -1 & 0 & 0 & 2 \end{pmatrix}.$$

The matrix Z^* is not unique. Since the columns of $\hat{A}^{*T}$ are the first two columns of the 4×4 identity, the most obvious choice for Z^* is the matrix whose columns are the remaining two coordinate vectors:

$$Z^* = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{which gives} \quad Z^{*T} W^* Z^* = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \tag{3.26}$$

The matrix $\hat{A}^{*T}(\hat{A}^*)^2\hat{A}^*$ has two eigenvalues of unity, and two zero eigenvalues. Both eigenvalues of the reduced Hessian of the Lagrangian function are equal to 2.

All calculations described in this paper were performed in double precision (around sixteen decimal digits) on a Silicon Graphics 4D/240S, which has IEEE arithmetic. All numbers are normally displayed to show at least the first "interesting" digit, and are correctly rounded to the number of digits shown.

To illustrate the behavior of the Hessian at points close to x^* but not on the barrier trajectory, we assume that $\mu = 10^{-4}$, and that the current iterate is

$$x_\delta = \begin{pmatrix} 1+2\times 10^{-4} \\ 1+10^{-8} \\ 10^{-4} \\ 10^{-4} \end{pmatrix}, \quad \text{with } c(x_\delta) = \begin{pmatrix} 2.0008\times 10^{-4} \\ 1.5007 \\ 1.0002\times 10^{-8} \\ 1.0002 \\ 1.0000 \end{pmatrix} \quad \text{and } d(x_\delta) = \begin{pmatrix} 0.49980 \\ 6.6636\times 10^{-5} \\ 9.9980\times 10^{3} \\ 9.9980\times 10^{-5} \\ 1.0000\times 10^{-4} \end{pmatrix}.$$

The value of $\|x^* - x_\delta\|$ is 2.45×10^{-4}. The third constraint (the second active constraint) is "inordinately small" relative to μ, and $d_{\max}$ is "large"; hence we could reasonably categorize x_δ as a point that fails to satisfy (3.16). The matrices V, $\hat{A}^T\hat{D}^2\hat{A}$ and μH_B evaluated at (x_δ, μ) are

$$V = \begin{pmatrix} -1.4995 & 0.99960 & 0 & -0.49980 \\ 0.99960 & -9.9990\times 10^{3} & 0 & 0 \\ 0 & 0 & -9.9976 & 0 \\ -0.49980 & 0 & 0 & 2.0000 \end{pmatrix},$$

$$\hat{A}^T\hat{D}^2\hat{A} = \begin{pmatrix} 0.25015 & -9.9967\times 10^{-5} & 8.8824\times 10^{-9} & 5.0013\times 10^{-5} \\ -9.9967\times 10^{-5} & 9.9960\times 10^{7} & 5.99760 & -2.2218\times 10^{-9} \\ 8.8824\times 10^{-9} & 5.9976 & 3.6430\times 10^{-7} & 8.8806\times 10^{-9} \\ 5.0013\times 10^{-5} & -2.2218\times 10^{-9} & 8.8806\times 10^{-9} & 2.7753\times 10^{-8} \end{pmatrix} \quad \text{and}$$

$$\mu H_B = \begin{pmatrix} 0.25000 & -7.2118\times 10^{-9} & 8.8824\times 10^{-9} & 3.2753\times 10^{-8} \\ -7.2118\times 10^{-9} & 9.9960\times 10^{7} & 5.9976 & -2.2218\times 10^{-9} \\ 8.8824\times 10^{-9} & 5.9976 & -9.9940\times 10^{-4} & 8.8806\times 10^{-9} \\ 3.2753\times 10^{-8} & -2.2218\times 10^{-9} & 8.8806\times 10^{-9} & 2.0003\times 10^{-4} \end{pmatrix}.$$

Because $d_{\max}$ is large and the elements of d differ from λ^*, V and $\hat{A}^T\hat{D}^2\hat{A}$ are not close to W^* and $\hat{A}^{*T}(\hat{A}^*)^2\hat{A}^*$ (see (3.25)).

The positive eigenvalues of $\hat{A}^T\hat{D}^2\hat{A}$ are 9.9960×10^7 and 0.25015. The eigenvalues of μH_B are 9.9960×10^7, 0.25000, 2.0003×10^{-4} and -9.9976×10^{-4}. As predicted by Theorem 3.1, the barrier Hessian is ill-conditioned, with $\text{cond}(\mu H_B) = 4.997\times 10^{11}$; this value is (substantially) larger than $d_{\max}/\mu$ because the smallest eigenvalue of μH_B is much smaller than its upper bound $d_{\max}O(\mu)$ (see the proof of Theorem 3.1). Notice also that the barrier Hessian is not positive definite, as it is at points near the trajectory close to x^*.

To see the effect of moving away from the boundary so that condition (3.16) holds, consider the case when μ is again 10^{-4}, and

$$x_\delta = \begin{pmatrix} 1+4\times 10^{-4} \\ 1+4\times 10^{-4} \\ 2\times 10^{-4} \\ 10^{-4} \end{pmatrix}, \text{ with } c(x_\delta) = \begin{pmatrix} 4.001\times 10^{-4} \\ 1.502 \\ 4.001\times 10^{-4} \\ 1.0004 \\ 1.0004 \end{pmatrix} \text{ and } d(x_\delta) = \begin{pmatrix} 0.2499 \\ 6.657\times 10^{-5} \\ 0.2450 \\ 9.996\times 10^{-5} \\ 9.996\times 10^{-5} \end{pmatrix}.$$

(We have deliberately chosen x_δ to display the ill-conditioning of eigenvectors associated with nearly-equal eigenvalues.) Because $d_{\max}$ is not "large", this point can safely be considered to satisfy (3.16). The corresponding matrices V, $\hat{A}^T\hat{D}^2\hat{A}$ and μH_B are

$$V = \begin{pmatrix} -0.74991 & 0.49985 & 0 & -0.24993 \\ 0.49985 & -0.74993 & 0 & 0 \\ 0 & 0 & 1.9994 & 0 \\ -0.24993 & 0 & 0 & 2 \end{pmatrix},$$

$$\hat{A}^T\hat{D}^2\hat{A} = \begin{pmatrix} 6.2525\times 10^{-2} & 1.7745\times 10^{-8} & 8.8688\times 10^{-9} & 2.5015\times 10^{-5} \\ 1.7745\times 10^{-8} & 6.2525\times 10^{-2} & 2.3869\times 10^{-8} & 1.7738\times 10^{-8} \\ 8.8688\times 10^{-9} & 2.3869\times 10^{-8} & 4.43269\times 10^{-9} & 8.8652\times 10^{-9} \\ 2.5015\times 10^{-5} & 1.7738\times 10^{-8} & 8.8652\times 10^{-9} & 2.7725\times 10^{-8} \end{pmatrix} \quad \text{and}$$

$$\mu H_B = \begin{pmatrix} 6.2450\times 10^{-2} & 5.0003\times 10^{-5} & 8.8688\times 10^{-9} & 2.2735\times 10^{-8} \\ 5.0003\times 10^{-5} & 6.2450\times 10^{-2} & 2.3869\times 10^{-8} & 1.7738\times 10^{-8} \\ 8.8688\times 10^{-9} & 2.3869\times 10^{-8} & 1.9994\times 10^{-4} & 8.8652\times 10^{-9} \\ 2.2735\times 10^{-8} & 1.7738\times 10^{-8} & 8.8652\times 10^{-9} & 2.0003\times 10^{-4} \end{pmatrix}.$$

Because d is *not* close to λ^*, the matrix V does not resemble W^*; nonetheless, $\hat{A}^T\hat{D}^2\hat{A}$ and μH_B are "close".

The (nearly equal) positive eigenvalues of $\hat{A}^T\hat{D}^2\hat{A}$ are 6.252501×10^{-2} and 6.252505×10^{-2}. The eigenvalues of the scaled barrier Hessian μH_B are 6.2400027×10^{-2}, 6.2500032×10^{-2}, 2.00029×10^{-4} and 1.99944×10^{-4}, so that $\operatorname{cond}(\mu H_B) = 3.12\times 10^4$. The matrix of eigenvectors of $\hat{A}^T\hat{D}^2\hat{A}$ is

$$Q_A = \begin{pmatrix} 0.57778 & -0.81619 & -1.7885\times 10^{-4} & -3.5788\times 10^{-4} \\ -0.81619 & -0.57778 & 2.1456\times 10^{-7} & -4.2439\times 10^{-7} \\ -2.2959\times 10^{-7} & -3.3639\times 10^{-7} & -0.89436 & 0.44736 \\ 2.3093\times 10^{-4} & -3.2671\times 10^{-4} & 0.44736 & 0.89436 \end{pmatrix},$$

where the first two columns correspond to the positive eigenvalues. The matrix of eigenvectors of μH_B is

$$Q_H = \begin{pmatrix} 0.70709 & -0.70712 & 1.0317\times 10^{-7} & -3.7786\times 10^{-7} \\ -0.70712 & -0.70709 & 3.5141\times 10^{-7} & -3.2321\times 10^{-7} \\ -1.7053\times 10^{-7} & -3.7157\times 10^{-7} & -0.99451 & 0.10469 \\ 5.6796\times 10^{-8} & -4.5936\times 10^{-7} & 0.10469 & 0.99451 \end{pmatrix},$$

where the columns are ordered by the size of the eigenvalues. We stress that, although the *individual columns* of Q_H are not close to those of Q_A, the subspace defined by the first two columns of Q_H *is* close to that defined by the first two columns of Q_A, as predicted by Theorem 3.2. Similarly, the subspace defined by the last two columns of Q_H is close to that defined by the last two columns of Q_A. The subspace defined by the last two columns of Q_H is also close to null($\hat{A}^*$) (see (3.25)). The fact that μH_B contains information about range($\hat{A}^T$) and null($\hat{A}$) will be exploited in Section 6.

Finally, we consider a minimizer of the barrier function. For $\mu = 10^{-4}$,

$$x_\mu = \begin{pmatrix} 1.00010 \\ 1.00010 \\ 3.33219 \times 10^{-5} \\ 1.16629 \times 10^{-4} \end{pmatrix}, \quad c(x_\mu) = \begin{pmatrix} 1.0005 \times 10^{-4} \\ 1.5007 \\ 1.0003 \times 10^{-4} \\ 1.0001 \\ 1.0001 \end{pmatrix}, \quad d(x_\mu) = \begin{pmatrix} 0.99955 \\ 6.6637 \times 10^{-5} \\ 0.99967 \\ 9.9990 \times 10^{-5} \\ 9.9990 \times 10^{-5} \end{pmatrix},$$

with $\|x^* - x_\mu\| = 1.86 \times 10^{-4}$. The matrices $V(x_\mu, \mu)$, $\hat{A}^T\hat{D}^2\hat{A}$ and H_B are

$$V(x_\mu, \mu) = \begin{pmatrix} -2.9988 & 1.9991 & 0 & -0.99955 \\ 1.9991 & -2.9989 & 0 & 0 \\ 0 & 0 & 1.9996 & 0 \\ -0.99955 & 0 & 0 & 2.0000 \end{pmatrix}, \tag{3.27}$$

$$\hat{A}^T\hat{D}^2\hat{A} = \begin{pmatrix} 0.99953 & 1.7767 \times 10^{-8} & 8.8819 \times 10^{-9} & 9.9978 \times 10^{-5} \\ 1.7767 \times 10^{-8} & 0.99953 & 1.5540 \times 10^{-8} & 1.7764 \times 10^{-8} \\ 8.8819 \times 10^{-9} & 1.5540 \times 10^{-8} & 4.4405 \times 10^{-9} & 8.8810 \times 10^{-9} \\ 9.9978 \times 10^{-5} & 1.7764 \times 10^{-8} & 8.8810 \times 10^{-9} & 2.7759 \times 10^{-8} \end{pmatrix}, \quad \text{and}$$

$$\mu H_B = \begin{pmatrix} 0.99923 & 1.9993 \times 10^{-4} & 8.8819 \times 10^{-9} & 2.2762 \times 10^{-8} \\ 1.9993 \times 10^{-4} & 0.99923 & 1.5540 \times 10^{-8} & 1.7764 \times 10^{-8} \\ 8.8819 \times 10^{-9} & 1.5540 \times 10^{-8} & 1.9996 \times 10^{-4} & 8.8810 \times 10^{-9} \\ 2.2762 \times 10^{-8} & 1.7764 \times 10^{-8} & 8.8810 \times 10^{-9} & 2.0003 \times 10^{-4} \end{pmatrix}.$$

The similarities between $V(x_\mu)$ and W^*, and between $\hat{A}^T\hat{D}^2\hat{A}$ and $\hat{A}^{*T}(\hat{A}^*)^2\hat{A}^*$, are apparent (see (3.25)). As predicted, these matrices differ by $O(\mu)$.

The (nearly-multiple) positive eigenvalues of $\hat{A}^T\hat{D}^2\hat{A}$ are 0.99953353 and 0.99953356. The eigenvalues of μH_B are 0.999436, 0.999034, 2.00029×10^{-4} and 1.99963×10^{-4}, and its condition number is 4.998×10^3. As predicted by our analysis, the two larger eigenvalues of μH_B are close to the positive eigenvalues of both $\hat{A}^T\hat{D}^2\hat{A}$ and $\hat{A}^*(\hat{A}^*)^2\hat{A}^*$. The "small" eigenvalues of μH_B are close to μ times the eigenvalues of $Z^{*T}W^*Z^*$. The eigenvectors of μH_B are

$$Q_H(x_\mu) = \begin{pmatrix} -0.70712 & -0.70710 & -2.3780 \times 10^{-8} & 5.6976 \times 10^{-9} \\ -0.70710 & 0.70712 & -1.9731 \times 10^{-8} & 1.2984 \times 10^{-8} \\ -1.7282 \times 10^{-8} & 4.7139 \times 10^{-9} & 0.13637 & -0.99069 \\ -2.8678 \times 10^{-8} & -3.5381 \times 10^{-9} & 0.99066 & 0.13637 \end{pmatrix}.$$

As expected, the subspace defined by the first two columns of Q_H is close to range($\hat{A}^{*T}$), and the subspace defined by the last two columns is close to null($\hat{A}^*$) (see (3.25)).

All of our results have involved the *exact* barrier Hessian H_B. Since this matrix will be computed in finite precision, it is worthwhile to mention the effects of rounding error. The numerical version of H_B will typically be dominated by the *unbounded* matrix $(1/\mu)\hat{A}^T\hat{D}^2\hat{A}$, and the information contained in V and μS will tend to be "lost" (see (3.2)). The loss of accuracy in the computed H_B tends to *accentuate* its nearness to rank-deficiency, and helps the numerical procedures to be described in Section 6.

4. Cholesky Factors of a Rank-Deficient Matrix

The material is this section is based on the discussion by Higham [High90] of the Cholesky factorization of a positive semidefinite matrix. The major result is Theorem 4.1, an adaptation of a theorem in [High90], which indicates that the Cholesky factorization with diagonal pivoting can successfully reveal the numerical rank of a matrix suitably "close to" a positive semidefinite rank-deficient matrix. For details and many interesting comments about the Cholesky factorization, see [High90].

Lemma 9.1 of [High90] quotes the following result about the existence and uniqueness of a Cholesky decomposition for a positive semidefinite matrix M.

Lemma 4.1. *Given an $n \times n$ symmetric positive semidefinite matrix M of rank r, there is a permutation Π such that $\Pi^T M \Pi$ has a unique Cholesky decomposition of the form*

$$\Pi^T M \Pi = \hat{R}^T \hat{R} = R^T R, \quad \textit{with} \quad \hat{R} = \begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ 0 & 0 \end{pmatrix}, \tag{4.1}$$

where $\hat{R}$ is $n \times n$, R is $r \times n$, and R_{11} is an $r \times r$ upper triangular matrix with positive diagonal elements.

The appearance of the permutation matrix Π in (4.1) is *critical* if M is rank-deficient. When M has full rank and is "sufficiently" positive definite (see [Wil65] and assumption (i) of Theorem 4.1, below), the factorization $M = R^T R$ can be computed in a numerically stable fashion without interchanges. When M is singular, however, interchanges must be allowed to prevent zero pivots.

If $\operatorname{rank}(M) = r$ and $r < n$, a numerically satisfactory factorization (4.1) can be calculated using an interchange strategy that prefers "large" pivots. The most popular approach is to construct Π with *diagonal* pivoting, in which each pivot element is taken as the maximum diagonal element of the remaining (partially factorized) matrix. Once the pivot element is identified, symmetric (row and column) interchanges are performed if necessary. With exact arithmetic, the factorization will terminate with a zero pivot after r steps.

To avoid the inclusion of permutation matrices in our notation, we assume throughout the remainder of this subsection that the matrix called "M" has been pre-permuted using diagonal pivoting, so that $M = R^T R$. Explicit permutations will reappear in later subsections.

Let M be partitioned as

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix}, \tag{4.2}$$

where M_{11} is $r \times r$. Higham's error analysis reveals the importance of the matrix

$$S = M_{11}^{-1} M_{12}, \tag{4.3}$$

which satisfies an *a priori* bound

$$\|S\|^2 = \|M_{11}^{-1} M_{12}\|^2 \le \tfrac{1}{3}(n-r)(4^r - 1). \tag{4.4}$$

Even with diagonal pivoting, this bound can be achieved for a certain parameterized family of matrices analogous to those associated with the worst-case bound for growth in Gaussian elimination with partial pivoting.

Consider an $n \times n$ symmetric matrix $\tilde{M}$ that is "close to" a positive semidefinite matrix, in the sense that

$$\tilde{M} = M + \Delta M, \tag{4.5}$$

where M is symmetric, positive semidefinite and of rank $r < n$, and ΔM is assumed to be "small".

Let R_r denote the $r \times n$ *computed* (partial) Cholesky factor of $\tilde{M}$ after r steps of reduction have been successfully completed. The analysis in [High90] shows that, for some "small" matrix E,

$$\tilde{M} - E = R_r^T R_r + \hat{M}^{(r+1)}, \quad \text{where} \quad \hat{M}^{(r+1)} = \begin{pmatrix} 0 & 0 \\ 0 & \hat{M}_{r+1} \end{pmatrix}. \tag{4.6}$$

Let $\mathbf{u}$ denote the unit roundoff of the floating-point system used in performing the calculations. We define the quantities

$$\epsilon_{r+1} = \frac{(r+1)\mathbf{u}}{1-(r+1)\mathbf{u}} \quad \text{and} \quad \xi = \frac{\epsilon_{r+1}}{1-r\epsilon_{r+1}}. \tag{4.7}$$

Given the relation

$$\tilde{M} - E = M + F, \quad \text{so that} \quad F = \Delta M - E, \tag{4.8}$$

Higham derives the following bound on the difference between $\tilde{M}$ and its rank-r Cholesky factorization:

$$\|\tilde{M} - R_r^T R_r\| \le \tau\,(1+n\xi)\,(\|S\|+1)^2\,(2r\xi\|\tilde{M}\| + \|\Delta M\|) \;+\; O(\|F\|^2), \tag{4.9}$$

where S is defined by (4.3) and

$$\tau = \frac{1}{1-n\xi(\|S\|+1)^2}. \tag{4.10}$$

The following result is a slight generalization of Theorem 9.1 from [High90]. Here, the bounds on ΔM and ΔM_{11} involve a generic small number δ rather than $\mathbf{u}$.

Theorem 4.1. *Let $\tilde{M}$ be a symmetric $n \times n$ matrix of floating-point numbers, with $\tilde{M} = M + \Delta M$, where M is symmetric, positive semidefinite and of rank $r < n$. We partition M as in (4.2), with conformal partitions for $\tilde{M}$ and ΔM. Assume that*

(i) *$\tilde{M}_{11}$ is positive definite, and $20r^{3/2}\,\mathrm{cond}(\tilde{M}_{11})\,\mathbf{u} < 1$;*

(ii) $\max\left\{ \dfrac{\|\Delta M_{11}\|}{\|\tilde{M}_{11}\|}, \dfrac{\|\Delta M\|}{\|\tilde{M}\|} \right\} = \kappa_\delta \delta$, *where κ_δ is a constant, δ is "small" and $\delta > \mathbf{u}$;*

(iii) $(\kappa_\delta\delta + r\xi)\;\mathrm{cond}(\tilde{M}_{11}) < \frac{1}{2}$, *where ξ is defined by (4.7);*

(iv) $n\xi\;(\|S\|+1)^2 < \frac{1}{2}$.

Then in floating-point arithmetic with unit roundoff $\mathbf{u}$, the Cholesky algorithm applied to $\tilde{M}$ successfully completes r stages, and the computed $r \times n$ Cholesky factor R_r satisfies

$$\|\tilde{M} - R_r^T R_r\| \le 2\,(\|S\|+2)^2\,(2r\xi + \kappa_\delta\delta)\,\|\tilde{M}\| \;+ O(\delta^2). \tag{4.11}$$

Proof. We merely indicate the general directions of the proof; complete details are given in [High90]. Condition (i) arises from a classical error analysis of Wilkinson [Wil65], and ensures that r steps of the Cholesky factorization of $\tilde{M}_{11}$ succeed without a numerical breakdown. Conditions (ii) and (iii) can be applied to guarantee that M_{11} is positive definite and $\|M_{11}^{-1}F_{11}\| < 1$, where F is the "small" matrix of (4.8). These properties lead to a bound on $\|\hat{M}^{(r+1)}\|$ of (4.6). Condition (iv) implies that the quantity τ of (4.10) satisfies $\tau \le 2$, and also allows us to bound the quantity $(1+n\xi)\,(\|S\|+1)^2$ by $(\|S\|+2)^2$. Combining all these results and applying them to (4.9), we obtain the desired bound. $\square$

We now examine the bound (4.11) to determine its significance for our application. As with most analyses of floating-point computations, the (possibly large) factor involving $r\xi$ is unlikely to be realistic, since it assumes the maximum possible error in every floating-point operation. If $n\mathbf{u} \ll 1$, the quantity ϵ_{r+1} of (4.7) is bounded above by a quantity very close to $n\mathbf{u}$; similarly, $r\xi$ is bounded above by a quantity very slightly larger than $n^2\mathbf{u}$. Our concern is with *dense* matrices, for which it is reasonable to assume that (say) $n \le 500$. Further, we make the assumption that $\mathbf{u} \le 10^{-14}$, which is quite generous since $\mathbf{u} \approx 10^{-16}$ on IEEE-standard machines using double precision. With these assumptions about n and $\mathbf{u}$, the maximum possible value of $2r\xi$ in the bound (4.11) is approximately 5×10^{-9}. In general, we shall be concerned with relative perturbations ΔM that are substantially larger than this value. If δ dominates $2r\xi$ and $\|S\|$ is not too large, then $R_r^T R_r$ will differ from $\tilde{M}$ by at most a multiple of δ times $\|\tilde{M}\|$.

This analysis has assumed throughout that the rank of M is known *a priori*. In practice, the rank must be estimated during the factorization (4.1). [High90] suggests several rank-estimation strategies, including the simple and effective approach of terminating with estimated rank k when the next candidate for the pivot element (the maximum diagonal element of the remaining matrix $\hat{M}^{(k+1)}$; see (4.6)), is less than some small multiple of the first pivot element, namely

$$\max_{k+1\le i\le n} |\hat{m}_{ii}^{(k+1)}| \le \mathrm{tol}\ \tilde{m}_{11}, \tag{4.12}$$

where tol is an appropriate tolerance. This criterion imposes a lower bound of $\mathrm{tol}^{-1/2}$ on the condition number of the Cholesky factor R_k. The value $\mathrm{tol} = n\mathbf{u}$ is suggested in [High90]. For our purposes, however, a much larger value is appropriate, as we shall discuss in Section 6.

5. A Rank-Retaining Factorization; the Spectral Resolution

In this section, we consider how the Cholesky factor of a positive semidefinite rank-deficient matrix can be transformed to reveal information about the associated range and null spaces, thereby allowing the calculation of the minimum-length least-squares solution. We also mention an application of the spectral resolution.

5.1. Subspace information in rank-retaining factors

Suppose that a symmetric $n \times n$ matrix M of rank r can be written as

$$\Pi^T M \Pi = R^T R, \quad \text{with} \quad R = (R_{11} \quad R_{12}), \tag{5.1}$$

where Π is a permutation, R is $r \times n$, and R_{11} is an $r \times r$ nonsingular upper triangle. Because all the matrices R, R^T and M have rank r, the representation (5.1) is an instance of *rank-retaining form*. (See [GMW91] for a general discussion of rank-retaining forms.) We know from the general properties of rank-retaining forms that the columns of R^T form a basis for the range space of $\Pi^T M \Pi$, and that R and $\Pi^T M \Pi$ have the same null space.

Orthonormal bases for these subspaces can be obtained if we know an explicit $n \times n$ orthogonal matrix $\tilde{Q}$ such that

$$R\tilde{Q} = (R_{11} \quad R_{12})\tilde{Q} = (\bar{R} \quad 0), \tag{5.2}$$

where $\bar{R}$ is an $r \times r$ nonsingular upper triangle. The first r columns of $\tilde{Q}$ form an orthonormal basis for the range space of R^T, and the last $n - r$ columns of $\tilde{Q}$ form an orthonormal basis for the null space of R. The most obvious way to obtain $\tilde{Q}$ is to apply r Householder transformations on the right of R to annihilate the columns of R_{12} while retaining the triangular structure of the first r columns, and then to multiply the transformations together to form $\tilde{Q}$. The reduction involves $r^2(n-r)$ flops, and the formation of $\tilde{Q}$ takes $O(r(n-r)\max(r, n-r))$ flops; see, for example, [GV89] or [GMW91] for details.

We define Q as $\Pi\tilde{Q}$, and partition Q into submatrices Y (the first r columns) and Z (the last $n - r$ columns), so that

$$Q = \Pi\tilde{Q} = (Y \quad Z). \tag{5.3}$$

Using the symmetry of M and the orthogonality of Π, it is straightforward to show that Y is an orthonormal basis for the range space of M and Z is an orthonormal basis for the null space of M.

The matrix M can be written as

$$M = Y\bar{R}^T\bar{R}Y^T, \quad \text{which means that} \quad Y^T M Y = \bar{R}^T\bar{R}. \tag{5.4}$$

5.2. Calculating the minimum-length least-squares solution

We now show how the matrices $\bar{R}$ and Y can be used to find the minimum-length least-squares solution of

$$\min_x \|Mx - b\|^2. \tag{5.5}$$

Any n-vector can be written as the sum of two orthogonal vectors, lying respectively in the range of Y and the range of Z. Assume that the right-hand side b of (5.5) is given by

$$b = Yb_Y + Zb_Z \tag{5.6}$$

for some vectors b_Y and b_Z. It is well known that any least-squares solution x of (5.5) must satisfy the linear system

$$Mx = Yb_Y, \tag{5.7}$$

which is necessarily compatible.

The *minimum-length* least-squares solution x_+ is the unique vector satisfying (5.7) that also lies entirely in the range of Y, i.e. $x_+ = Yx_Y$. Because $Y^TY = I$ and $Y^TZ = 0$, application of Y^T annihilates the part of b in the null space of M, and gives $Y^Tb = b_Y$. The result is

$$MYx_Y = Yb_Y, \quad \text{so that } x_Y \text{ satisfies} \quad Y^TMYx_Y = Y^Tb.$$

Because $Y^TMY = \bar{R}^T\bar{R}$, x_+ can be calculated by solving $\bar{R}^T\bar{R}v = Y^Tb$ and setting $x_+ = Yv$.

5.3. Solving linear equations with the spectral resolution

Suppose that $\tilde{M}$ is a symmetric $n \times n$ *nonsingular* matrix, and that we wish to solve $\tilde{M}x = b$. Let $(\tilde{Y} \quad \tilde{Z})$ be an orthogonal matrix, where $\tilde{Y}$ has r columns, such that range$(\tilde{Z})$ is a simple invariant subspace of $\tilde{M}$. Then $\tilde{M}$ can be expressed as its spectral resolution with respect to $\tilde{Y}$ and $\tilde{Z}$ (see (2.2)):

$$\tilde{M} = \tilde{Y}\tilde{M}_Y\tilde{Y}^T + \tilde{Z}\tilde{M}_Z\tilde{Z}^T, \quad \text{where} \quad \tilde{M}_Y = \tilde{Y}^T\tilde{M}\tilde{Y} \quad \text{and} \quad \tilde{M}_Z = \tilde{Z}^T\tilde{M}\tilde{Z}. \tag{5.8}$$

Writing the unknown vector x as $x = \tilde{Y}x_Y + \tilde{Z}x_Z$, we apply $\tilde{Y}^T$ to the equation $\tilde{M}x = b$, substitute for x, and obtain an $r \times r$ linear system satisfied by x_Y:

$$\tilde{Y}^T\tilde{M}(\tilde{Y}x_Y + \tilde{Z}x_Z) = \tilde{M}_Yx_Y = \tilde{Y}^Tb. \tag{5.9}$$

An approximate version of this relation will be applied to the Newton equations in Section 6.

6. Subspace Information; an Approximate Newton Step

The results of Section 3 show that, for x_δ close to x^* and sufficiently small μ and δ satisfying (3.9), the barrier Hessian $H_B(x_\delta, \mu)$ is ill-conditioned. Hence we may expect numerical difficulties in solving the Newton equations (3.1) as the iterates in a barrier method approach the solution. Since a problem is ill-conditioned when the *exact* solution can experience large changes if the problem data undergoes small perturbations, it might also be feared that application of a barrier method is inherently unreliable.

Such a view is overly pessimistic. It is important to remember that solving the Newton equations (3.1) is only one means to achieve the desired end, namely convergence to x^*. Other formulations of a "good" step toward x^* from x_δ can be stated in which the Hessian of the barrier function does not appear. These alternatives rely on properties that hold near x^*, and calculate the search direction in two orthogonal pieces, lying respectively in the range space of $\hat{A}^T$ and the null space of $\hat{A}$ (see, e.g., [Wr76], [MW78] and [NS91]). The equations to be solved involve matrices whose conditioning reflects the "natural" condition of the problem—namely, $\hat{A}^*$, $\hat{\lambda}^*$ and $Z^{*T}W^*Z^*$.

These properties may appear unhelpful to a barrier method because they depend on the set of active constraints, which is *unknown*. In fact, it is a desirable (some might say essential) feature of barrier methods that they do *not* make an explicit prediction of the active set. The question therefore arises as to whether, knowing only H_B, information about the subspaces defined by $\hat{A}$ can be gained *without knowing the active set*.

6.1. Subspace information

Although ill-conditioning in H_B is inevitable, its structure is known in advance to be closely related to $\hat{A}$. In particular, we have seen in Section 3 that, when x_δ satisfies condition (3.16), the small (close to zero) eigenvalues of the scaled barrier Hessian μH_B define an invariant subspace that is close to the null space of $\hat{A}$, and the $\Theta(1)$ eigenvalues define an invariant subspace that is close to the range space of $\hat{A}^T$. It is precisely when consistent ill-conditioning occurs that the iterates are sufficiently close to x^* so that these subspaces can be approximated using the following procedure.

A Cholesky-based algorithm *with diagonal pivoting* is applied at every iteration to the barrier Hessian. The tolerance in the rank test (4.12) can safely be taken as (say) $100n\sqrt{\mathbf{u}}$, where $\mathbf{u}$ is the machine precision. (This value is chosen because the maximum achievable relative accuracy in the solution is typically related to $\sqrt{\mathbf{u}}$; see [GMW81], Section 8.2.2.) When the barrier Hessian is determined to have full rank, the ordinary Newton (or modified Newton) direction can be obtained using the Cholesky factors. As μ decreases, H_B will eventually become so ill-conditioned that its numerical rank is less than n. Once this has happened for a suitably "small" μ, the Hessian is considered as too ill-conditioned for direct use.

We know from Section 3 that, when $0 < \hat{m} < n$, the scaled barrier Hessian matrix μH_B is dominated by the rank-deficient matrix $\hat{A}^T\hat{D}^2\hat{A}$, and hence has the form of $\tilde{M}$ in (4.5) of Section 4, where $\hat{A}^T\hat{D}^2\hat{A}$ plays the role of the positive semidefinite matrix M. Under the conditions of Lemma 3.2, the "perturbation" ΔM is $O(\mu)$. When μ is (say) $100\sqrt{\mathbf{u}}$, the Cholesky algorithm with diagonal pivoting should produce an estimated numerical rank $r = \hat{m}$, and a permutation matrix Π such that

$$\Pi^T H_B \Pi \approx \frac{1}{\mu}\,\Pi^T\hat{A}^T\hat{D}^2\hat{A}\Pi \approx R^T R, \tag{6.1}$$

where R is $r \times n$, with $r < n$, and has the form $(R_{11} \quad R_{12})$, with R_{11} an $r \times r$ nonsingular upper triangle.

We combine the results of Theorem 3.2 and of Section 4: the barrier Hessian has simple invariant subspaces close to those of $\hat{A}$, and the computed Cholesky factor R of H_B is close to the exact Cholesky factor of $(1/\mu)\hat{A}^T\hat{D}^2\hat{A}$. Explicit orthonormal matrices Y and Z spanning the range of R^T and the null space of R can be found by reducing the Cholesky factor R from the right, as in (5.2). Although these orthonormal bases correspond to subspaces close to range($\hat{A}^T$) and null($\hat{A}$), *no identification of the active set occurs.*

In practice, a consistency check for Y and Z is provided by the fact that the gradient of the objective function is known to lie increasingly close to a vector entirely in range($\hat{A}^T$), i.e. in range(Y), as the iterates approach x^*. When μ is small, the vector Z^Tg can be computed and checked to see it is sufficiently "small". If not, this is a signal that Z may not be the correct basis for the null space of $\hat{A}$, since $Z^{*T}g(x^*) = 0$. Observe that we do not need to know $\hat{A}$ to make this test.

We illustrate the calculation of Y and Z for a three-variable problem:

$$\begin{array}{ll} \text{minimize} & 5x_1x_2x_3 - \tfrac{1}{2}x_1^2 - 2x_2x_3 - x_3 - \tfrac{3}{2}x_2^2 - x_3^2 \\ \text{subject to} & -x_1^2 - x_3^2 + 2 \ge 0 \\ & x_1 + \tfrac{3}{4} \ge 0 \\ & (x_1 - x_3)^2 + x_2^3 + 1 \ge 0. \end{array} \tag{6.2}$$

(We do not use problem (3.24) because its barrier Hessian is nearly diagonal near x^*, so that estimating the rank of the Cholesky factor is too easy to be interesting.) An optimal solution of problem (6.2) is $x^* = (1, -1, 1)^T$, where the first and third constraints are active, with $\lambda^* = (3, 0, 2)^T$. At x^*,

$$\hat{A}^* = \begin{pmatrix} -2 & 0 & -2 \\ 0 & 3 & 0 \end{pmatrix} \quad \text{and} \quad W^* = \begin{pmatrix} 1 & 5 & -1 \\ 5 & 9 & 3 \\ -1 & 3 & 0 \end{pmatrix}.$$

The first two columns of the following orthogonal matrix Q_A form a basis for range$(\hat{A}^{*T})$, and the last column is a basis for the (one-dimensional) null space of $\hat{A}^*$:

$$Q_A = \begin{pmatrix} 0 & 0.707107 & 0.707107 \\ 1 & 0 & 0 \\ 0 & 0.707107 & -0.707107 \end{pmatrix}. \tag{6.3}$$

The value of μ is taken as 10^{-5}, which is small enough so that the asymptotic ill-conditioning has made an appearance, but the exact Newton direction can still be calculated accurately with the available precision (see Section 6.2). At the point

$$x = \begin{pmatrix} 1+10^{-5} \\ -1+10^{-5} \\ 1-2\times 10^{-5} \end{pmatrix}, \quad \text{the constraint values are} \quad c = \begin{pmatrix} 1.99995\times 10^{-5} \\ 1.75001 \\ 3.00006\times 10^{-5} \end{pmatrix}, \tag{6.4}$$

and $\|x - x_\mu\| = 2.22\times 10^{-5}$. The barrier gradient g_B and scaled barrier Hessian μH_B are

$$g_B = \begin{pmatrix} -4.99985 \\ 5.0 \\ -4.99996 \end{pmatrix} \quad \text{and} \quad \mu H_B = \begin{pmatrix} 1.00006 & 6.9978\times 10^{-5} & 0.999997 \\ 6.9978\times 10^{-5} & 0.999910 & 1.00017\times 10^{-5} \\ 0.999997 & 1.00017\times 10^{-5} & 0.999993 \end{pmatrix},$$

and the condition of H_B is 6.32×10^4.

Using a tolerance of 10^{-4} in the rank estimation test (4.12), the barrier Hessian is considered to have rank two. (We use a relatively large tolerance for the rank test to indicate how the closeness to rank-deficiency is revealed.) The rank-two approximate Cholesky factor of μH_B is

$$R = \begin{pmatrix} 1.00003 & 6.99956\times 10^{-5} & 0.999965 \\ & 0.999955 & -5.99941\times 10^{-5} \end{pmatrix}.$$

The matrices $\bar{R}$ and Q produced by Householder reduction of R from the right are

$$\begin{aligned} \bar{R} &= \begin{pmatrix} -1.4142 & -1.0001\times 10^{-5} \\ 0 & -0.99996 \end{pmatrix} \quad \text{and} \\ Q &= \begin{pmatrix} -0.70713 & 0 & -0.70708 \\ -4.2423\times 10^{-5} & -1.0000 & 4.2426\times 10^{-5} \\ -0.70708 & 5.9997\times 10^{-5} & 0.70713 \end{pmatrix}, \end{aligned} \tag{6.5}$$

where the first two columns of Q are Y, and Z is the last column of Q. Note the similarity between the subspaces defined by Y and Z, and the analogous subspaces defined by Q_A of (6.3). The vector $Q^T g$ is $(8.4849,\ -6.0003,\ -1.1312\times 10^{-4})^T$, where the (small) last component is $Z^T g$.

6.2. An approximate Newton step

Once Y and Z have been determined, an approximate solution $\tilde{p}$ of the Newton equations can be computed in two orthogonal pieces, $Y\tilde{p}_Y$ and $Z\tilde{p}_Z$, by applying a procedure analogous to that mentioned in Section 5.2. Because H_B is dominated by $(1/\mu)\hat{A}^T\hat{D}^2\hat{A}$, we know that

$$YH_BY^T \approx \frac{1}{\mu} Y\hat{A}^T\hat{D}^2\hat{A}Y^T \approx \bar{R}^T\bar{R}.$$

Treating this relation as an equality in applying (5.9) to solve the Newton equations, we obtain

$$\bar{R}^T\bar{R}\tilde{p}_Y = -Y^T g_B, \tag{6.6}$$

so that $\tilde{p}_Y$ is obtained by solving two triangular systems of equations: $\bar{R}^T s = -Y^T g_B$ and $\bar{R}\tilde{p}_Y = s$. As the iterates approach the solution, the matrix $\bar{R}$ is *not* inherently ill-conditioned; rather, it reflects the ratio of the largest to the smallest positive singular values of $\hat{A}^T\hat{D}$, which approaches $\hat{A}^{*T}\hat{A}^*$ and thereby reflects the natural condition of the problem.

An alternative way to derive (6.6) is to observe that the gradient of the barrier function is given by

$$g_B = g - \mu A^T C^{-1} e = g - \mu\hat{A}^T(\hat{C})^{-1}e - \mu\bar{A}^T(\bar{C})^{-1}e.$$

At x^*, $g(x^*) = \hat{A}^{*T}\hat{\lambda}^*$. Close to x^*, because of continuity of g and A, g is close to a linear combination of the columns of $\hat{A}^T$, which means that the barrier gradient is dominated by a linear combination of the columns of $\hat{A}^T$. Since H_B is approaching a rank-deficient matrix whose range space is defined by $\hat{A}^T$, the Newton equations $H_B p_{\text{newt}} = -g_B$ are nearly compatible. The vector $Y\tilde{p}_Y$, an approximation to the range-space part of the Newton step, can be viewed as the minimum-length solution of the least-squares problem min $\|H_B\tilde{p} + g_B\|_2^2$; see (5.5).

Assuming that the range-space part $Y\tilde{p}_Y$ is known, the null-space part of $\tilde{p}$ is obtained as follows. The Newton equations are

$$H_B(Y\tilde{p}_Y + Z\tilde{p}_Z) = -g_B. \tag{6.7}$$

Multiplying through by Z^T and rearranging, we have

$$Z^T H_B Z\tilde{p}_Z = -Z^T g_B - Z^T H_B Y\tilde{p}_Y. \tag{6.8}$$

The term $Z^T H_B Y\tilde{p}_Y$ appearing on the right-hand side would be exactly zero if Z were an exact basis for an invariant subspace of H_B. However, because $\|Z^T H_B Y\tilde{p}_Y\|$ is in general *not* negligible with respect to $\|Z^T g_B\|$, the vector $Z^T H_B Y\tilde{p}_Y$ is included in the right-hand side.

Although the reduced barrier Hessian $Z^T H_B Z$ appears in the formal relation (6.8), it is inadvisable to obtain this matrix from the computed version of H_B. If infinite precision were available, application of Z^T to H_B would exactly annihilate the "large" matrix $(1/\mu)\hat{A}^T\hat{D}^2\hat{A}$, so that

$$Z^T H_B Z = Z^T V Z + \mu Z^T S Z.$$

With finite precision, however, the computed matrix H_B is likely to have "lost" most of the information represented in V and μS. Since the active constraints are unknown, we cannot calculate S separately from $(1/\mu)\hat{A}^T\hat{D}^2\hat{A}$. Fortunately, the matrix V does *not* depend on the active set, and $Z^T V Z$ can be calculated accurately if a separate copy of V is retained. Because $\|\mu S\|$ is $O(\mu)$ near x^*, $Z^T V Z$ differs by $O(\mu)$ from $Z^T H_B Z$ for sufficiently small μ. The vector $\tilde{p}_Z$ is accordingly obtained by solving

$$Z^T V Z\tilde{p}_Z = -Z^T g_B - Z^T H_B Y\tilde{p}_Y. \tag{6.9}$$

Unless $\|Z^T V Z\|$ is very small, neglecting μS has only a minor effect on the relative accuracy of $\tilde{p}_Z$, in contrast with the need to include $Z^T H_B Y\tilde{p}_Y$ in (6.9). The vector $\tilde{p}$ is then defined as

$$\tilde{p} = Y\tilde{p}_Y + Z\tilde{p}_Z. \tag{6.10}$$

For points on the barrier trajectory, V is close to the Lagrangian function, which means that the matrix $Z^T V Z$ in (6.9) resembles the reduced Hessian of the Lagrangian function. Equation (6.9) is therefore strongly reminiscent of the equation defining the null-space part of the search direction in sequential quadratic programming (SQP) methods, which are arguably the most popular general technique for solving nonlinearly constrained problems (see, e.g., [GMW81] and [Fle87]). Suppose that the SQP subproblem is

$$\begin{array}{ll} \text{minimize} & \tfrac{1}{2}p^T W p + p^T g \\ \text{subject to} & Ap \geq b, \end{array} \tag{6.11}$$

where the vector b depends on the particular SQP formulation, and g and W represent the gradient and Hessian of the Lagrangian function. Under the same assumptions made in Section 1, it is known that, close to x^*, the standard formulation of the SQP subproblem (6.11) correctly identifies the active set of the original nonlinear problem. Let $\hat{A}$ denote the Jacobian of the active constraints of the quadratic program (6.11), Y a basis for the range space of $\hat{A}^T$ and Z a basis for the null space of $\hat{A}$. The solution $\bar{p}$ of (6.11) can be expressed in two orthogonal pieces as $\bar{p} = Y\bar{p}_Y + Z\bar{p}_Z$, where

$$\hat{A}Y\bar{p}_Y = \hat{b} \quad \text{and} \quad Z^TWZ\bar{p}_Z = -Z^Tg - Z^TWY\bar{p}_Y,$$

where $\hat{b}$ denotes the components of b corresponding to active constraints. See [Wr76] and [MW78] for further discussion of the connections between SQP methods and methods based on following the barrier trajectory.

To illustrate the calculation of $\tilde{p}$, we return to example (6.2), with x from (6.4) and $\mu = 10^{-5}$. With the matrices $\bar{R}$, Y and Z from (6.5), relations (6.6) and (6.9) give

$$Y\tilde{p}_Y = \begin{pmatrix} 2.5000 \times 10^{-5} \\ -5.0006 \times 10^{-5} \\ 2.5001 \times 10^{-5} \end{pmatrix}, \quad Z\tilde{p}_Z = \begin{pmatrix} 3.2736 \times 10^{-6} \\ -1.9642 \times 10^{-10} \\ -3.2738 \times 10^{-6} \end{pmatrix} \quad \text{and} \quad \tilde{p} = \begin{pmatrix} 2.8274 \times 10^{-5} \\ -5.0007 \times 10^{-5} \\ 2.1728 \times 10^{-5} \end{pmatrix}.$$

The Newton direction p_{newt} is

$$p_{\text{newt}} = \begin{pmatrix} 2.8273 \times 10^{-5} \\ -5.0007 \times 10^{-5} \\ 2.1728 \times 10^{-5} \end{pmatrix}, \quad \text{with} \quad p_{\text{newt}} - \tilde{p} = \begin{pmatrix} -7.7978 \times 10^{-10} \\ 9.50 \times 10^{-14} \\ 9.1845 \times 10^{-11} \end{pmatrix},$$

and $\|p_{\text{newt}} - \tilde{p}\| = 7.85 \times 10^{-10}$.

References

[CS84] T. F. Coleman and D. C. Sorensen (1984), A note on the computation of an orthogonal basis for the null space of a matrix, *Mathematical Programming* 29, 234–242.

[DS83] J. E. Dennis, Jr. and R. B. Schnabel (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey.

[Fi79] A. V. Fiacco (1979), Barrier methods for nonlinear programming, in *Operations Research Support Methodology* (A. G. Holzman, ed.), Marcel Dekker (New York), 377–440.

[FM68] A. V. Fiacco and G. P. McCormick (1968), *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons (New York). Republished by SIAM (Philadelphia), 1990.

[Fle87] R. Fletcher (1987), *Practical Methods of Optimization* (second edition), John Wiley and Sons (Chichester).

[GMW81] P. E. Gill, W. Murray and M. H. Wright (1981), *Practical Optimization*, Academic Press, London and New York.

[GMW91] P. E. Gill, W. Murray and M. H. Wright (1991), *Numerical Linear Algebra and Optimization, Volume 1*, Addison-Wesley, Redwood City, California.

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

[Gon92] C. C. Gonzaga (1992), Path following methods for linear programming, *SIAM Review* 34, to appear.

[High90] N. J. Higham (1990), Analysis of the Cholesky decomposition of a semi-definite matrix, in *Reliable Numerical Computation* (M. G. Cox and S. Hammarling, eds.), Clarendon Press (Oxford), 161–185.

[Kar84] N. K. Karmarkar (1984), A new polynomial time algorithm for linear programming, *Combinatorica* 4, 373–395.

[Loot69] F. A. Lootsma (1969). Hessian matrices of penalty functions for solving constrained optimization problems, *Philips Res. Repts* **24**, 322–331.

[Mur71] W. Murray (1971), Analytical expressions for the eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions, *J. Optim. Theory Appl.* 7, 189–196.

[MW78] W. Murray and M. H. Wright (1978), Projected Lagrangian methods based on the trajectories of penalty and barrier functions, Report SOL 78-23, Department of Operations Research, Stanford University, Stanford, California.

[NS91] S. G. Nash and A. Sofer, A barrier method for large-scale constrained optimization, Technical Report 91-10, Department of Operations Research and Applied Statistics, George Mason University, Fairfax, Virginia, 1991.

[PS82] C. R. Papadimitriou and K. Steiglitz (1982), *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, New Jersey.

[SE90] R. B. Schnabel and E. Eskow (1990), A new modified Cholesky factorization, *SIAM J. Sci. Stat. Comp.* 11, 1136–1158.

[SS90] G. W. Stewart and J. Sun (1990), *Matrix Perturbation Theory*, Academic Press, Boston.

[Wil65] J. H. Wilkinson (1965), *The Algebraic Eigenvalue Problem*, The Clarendon Press, Oxford.

[Wr76] M. H. Wright (1976), *Numerical Methods for Nonlinearly Constrained Optimization*, Ph. D. thesis, Stanford University, California.

[Wr92] M. H. Wright (1992), Interior methods for constrained optimization, *Acta Numerica 1992*, 341–407.