SIAM J. Matrix Anal. Appl.
Vol. 23, No. 1, pp. 42–56

ON WEIGHTED LINEAR LEAST-SQUARES PROBLEMS RELATED TO INTERIOR METHODS FOR CONVEX QUADRATIC PROGRAMMING*

ANDERS FORSGREN† AND GÖRAN SPORRE†

Abstract. It is known that the norm of the solution to a weighted linear least-squares problem is uniformly bounded for the set of diagonally dominant symmetric positive definite weight matrices. This result is extended to weight matrices that are nonnegative linear combinations of symmetric positive semidefinite matrices. Further, results are given concerning the strong connection between the boundedness of weighted projection onto a subspace and the projection onto its complementary subspace using the inverse weight matrix. In particular, explicit bounds are given for the Euclidean norm of the projections. These results are applied to the Newton equations arising in a primal-dual interior method for convex quadratic programming and boundedness is shown for the corresponding projection operator.

Key words. unconstrained linear least-squares problem, weighted least-squares problem, quadratic programming, interior method

AMS subject classifications. 65F20, 65F35, 65K05

PII. S0895479800372298

1. Introduction.

In this paper we study certain properties of the weighted linear least-squares problem

$$\underset{\pi \in \mathbb{R}^m}{\text{minimize}} \qquad \|W^{1/2}(A^T\pi - g)\|_2^2, \tag{1.1}$$

where A is an $m \times n$ matrix of full row rank and W is a positive definite symmetric $n \times n$ matrix whose matrix square root is denoted by $W^{1/2}$. (See, e.g., Golub and Van Loan [14, p. 149] for a discussion on matrix square roots.) Linear least-squares problems are fundamental within linear algebra; see, e.g., Lawson and Hanson [20], Golub and Van Loan [14, Chapter 5] and Gill, Murray, and Wright [12, Chapter 6]. An individual problem of the form (1.1) can be converted to an unweighted problem by substituting $\widetilde{A} = AW^{1/2}$ and $\widetilde{g} = W^{1/2}g$. However, our interest is in *sequences* of weighted problems, where the weight matrix W changes and A is constant. The present paper is a continuation of the paper by Forsgren [10], in which W is assumed to be diagonally dominant. Our concern is when the weight matrix is of the form

$$W = (H + D)^{-1}, \tag{1.2}$$

where H is a constant positive semidefinite symmetric matrix and D is an arbitrary positive definite diagonal matrix. Such matrices arise in interior methods for convex quadratic programming. See section 1.1 below for a brief motivation.

The solution of (1.1) is given by the *normal equations*

$$AWA^T\pi = AWg \tag{1.3}$$

*Received by the editors May 18, 2000; accepted for publication (in revised form) by M. Overton November 29, 2000; published electronically April 18, 2001. This research was supported by the Swedish Natural Science Research Council (NFR).

http://www.siam.org/journals/simax/23-1/37229.html

†Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (anders.forsgren@math.kth.se, goran.sporre@math.kth.se).

or alternatively as the solution to the *augmented system* (or *Karush–Kuhn–Tucker* (*KKT*) *system*)

$$\begin{pmatrix} M & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} r \\ \pi \end{pmatrix} = \begin{pmatrix} g \\ 0 \end{pmatrix}, \tag{1.4}$$

where $M = W^{-1}$. In some situations, we will prefer the KKT form (1.4), since we are interested in the case when M is a positive semidefinite symmetric and singular matrix. In this situation, W^{-1} and (1.3) are not defined, but (1.4) is well defined. This would be the case, for example, in an equality-constrained weighted linear least-squares problem; see, e.g., Lawson and Hanson [20, Chapter 22]. For convenience, we will mainly use the form (1.3).

If $M = W^{-1}$, then, mathematically, (1.3) and (1.4) are equivalent. From a computational point of view, this need not be the case. There is a large number of papers giving reasons for solving systems of one type or the other, starting with Bartels, Golub, and Saunders [1], followed by, e.g., Duff et al. [9], Björck [4], Gulliksson and Wedin [17], Wright [29, 31], Björck and Paige [5], Vavasis [26], Forsgren, Gill, and Shinnerl [11], and Gill, Saunders, and Shinnerl [13]. The focus of the present paper is linear algebra, and we will not discuss these important computational aspects.

If A has full row rank and if $\mathcal{W}_+$ is defined as the set of $n \times n$ positive definite symmetric matrices, then for any $W \in \mathcal{W}_+$, the unique solution of (1.1) is given by

$$\pi = (AWA^T)^{-1}AWg. \tag{1.5}$$

In a number of applications, it is of interest to know if the solution remains in a compact set as the weight matrix changes, i.e., the question is whether

$$\sup_{W \in \mathcal{W}} \|(AWA^T)^{-1}AW\|$$

remains bounded for a particular subset $\mathcal{W}$ of $\mathcal{W}_+$. It should be noted that boundedness does not hold for an arbitrary subset $\mathcal{W}$ of $\mathcal{W}_+$. Take for example $A = (0\ 1)$ and let

$$W(\epsilon) = \begin{pmatrix} \frac{2}{\epsilon} & 1 \\ 1 & \epsilon \end{pmatrix}$$

for $\epsilon > 0$. Then $W(\epsilon) \in \mathcal{W}_+$ for $\epsilon > 0$, and

$$(AW(\epsilon)A^T)^{-1}AW(\epsilon) = \begin{pmatrix} \frac{1}{\epsilon} & 1 \end{pmatrix}.$$

This implies that $\|(AWA^T)^{-1}AW\|$ is not bounded when W is allowed to vary in $\mathcal{W}_+$. See Stewart [24] for another example of unboundedness and related discussion. For the case where $\mathcal{W}$ is the set of positive definite diagonal matrices, Dikin [8] gives an explicit formula for the optimal π in (1.1) as a convex combination of the *basic solutions* formed by satisfying m linearly independent equations. From this result, the boundedness is obvious. If A does not have full row rank, it is still possible to show boundedness; see Ben-Israel [2, p. 108]. Later, Wei [28] also studied boundedness in absence of a full row rank assumption on A and has furthermore given some stability results. Bobrovnikova and Vavasis [6] have given boundedness results for complex diagonal weight matrices. The geometry of the set $(AWA^T)^{-1}AWg$ when W varies over the set of positive definite diagonal matrices has been studied by Hanke and Neumann [18]. Based on the formula derived by Dikin [8], Forsgren [10] has given boundedness results when $\mathcal{W}$ is the set of positive definite diagonally dominant matrices.

We show boundedness for the set of weight matrices that are arbitrary nonnegative combinations of a set of fixed positive semidefinite symmetric matrices and the set of inverses of such matrices. As a special case, we then obtain the set of weight matrices of the form (1.2), which was our original interest. The boundedness is shown in the following way. In section 2, we review results for the characterization of π as W varies over the set of symmetric matrices such that AWA^T is nonsingular. Section 3 establishes the boundedness when W is allowed to vary over a set of matrices that are nonnegative linear combinations of a number of fixed positive semidefinite matrices such that AWA^T is positive definite. In section 4, results that are needed to handle the projection using the inverse weight matrix are given. In section 5, we combine results from the previous two sections to show boundedness for the π that solves (1.4) when M is allowed to vary over the nonnegative linear combinations of a set of fixed positive semidefinite symmetric matrices.

The research was initiated by a paper by Gonzaga and Lara [15]. The link to that paper has subsequently been superseded, but we include a discussion relating our results to the result of Gonzaga and Lara in the appendix.

1.1. Motivation. Our interest in weighted linear least-squares problems is from interior methods for optimization and in particular for convex quadratic programming. There is a vast number of papers on interior methods, and here we give only a brief motivation for the weighted linear least-squares problems that arise. Any convex quadratic programming problem can be transformed to the form

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \frac{1}{2}x^THx + c^Tx \\ \text{subject to} & Ax = b, \\ & x \geq 0, \end{array} \tag{1.6}$$

where H is a positive semidefinite symmetric $n \times n$ matrix and A is an $m \times n$ matrix of full row rank. For $x \in \mathbb{R}^n$, $\pi \in \mathbb{R}^m$, and $s \in \mathbb{R}^n$ such that $x > 0$ and $s > 0$, an iteration of a primal-dual path-following interior method for solving (1.6) typically takes a Newton step towards the solution of the equations

$$Hx + c - A^T\pi - s = 0, \tag{1.7a}$$

$$Ax - b = 0, \tag{1.7b}$$

$$Xs - \mu e = 0, \tag{1.7c}$$

where μ is a positive barrier parameter; see, e.g., Monteiro and Adler [21, p. 46]. Here, $X = \operatorname{diag}(x)$ and similarly below $S = \operatorname{diag}(s)$. Strict positivity of x and s is implicitly required and typically maintained by limiting the step length. If μ is set equal to zero in (1.7) and the implicit requirements $x > 0$ and $s > 0$ are replaced by $x \geq 0$ and $s \geq 0$, the optimality conditions for (1.6) are obtained. Consequently, (1.7) and the implicit positivity of x and s may be viewed as a perturbation of the optimality conditions for (1.6). In a primal-dual path-following interior method, the perturbation is driven to zero to make the method converge to an optimal solution. The equations (1.7) are often referred to as the primal-dual equations. Forming the Newton equations associated with (1.7) for the corrections Δx, $\Delta\pi$, Δs and eliminating Δs gives

$$\begin{pmatrix} H + X^{-1}S & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} -\Delta x \\ \Delta\pi \end{pmatrix} = \begin{pmatrix} Hx + c - \mu X^{-1}e - A^T\pi \\ Ax - b \end{pmatrix}. \tag{1.8}$$

If x and s are strictly feasible, i.e., x and s are strictly positive and x satisfies $Ax = b$, then a comparison of (1.4) and (1.8) shows that the Newton equations (1.8) can be associated with a weighted linear least-squares problem with a positive definite weight matrix $(H + X^{-1}S)^{-1}$. A sequence of strictly feasible iterates $\{x_k\}_{k=0}^{\infty}$ gives rise to a sequence of weighted linear least-squares problems, where the weight matrix changes but A is constant.

In a number of convergence proofs for linear programming, a crucial step is to ensure boundedness of the step $(\Delta x, \Delta s)$; see, e.g., Vavasis and Ye [27, Lemma 4] and Wright [30, Lemmas 7.2 and A.4]. Since linear programming is the special case of convex quadratic programming where $H = 0$, we are interested in extending this boundedness result to convex quadratic programming. Therefore, the boundedness of

$$\|(A(H + X^{-1}S)^{-1}A^T)^{-1}A(H + X^{-1}S)^{-1}\| \tag{1.9}$$

as $X^{-1}S$ varies over the set of diagonal positive definite matrices is of interest. This boundedness property of (1.9) is shown in section 5.

1.2. Notation. When we refer to matrix norms and make no explicit reference to what type of norm is considered, it can be any matrix norm that is induced from a vector norm such that $\|(x^T \ 0)^T\| = \|x\|$ holds for any vector x. To denote the ith eigenvalue and the ith singular value, we use λ_i and σ_i, respectively. For symmetric matrices A and B of equal dimension, $A \succeq B$ means that $A - B$ is positive semidefinite. Similarly, $A \succ B$ means that $A - B$ is positive definite.

The remainder of this section is given in Forsgren [10]. It is restated here for completeness. For an $m \times n$ matrix A of full row rank, we shall denote by $\mathcal{J}(A)$ the collection of sets of column indices associated with the nonsingular $m \times m$ submatrices of A. For $J \in \mathcal{J}(A)$, we denote by A_J the $m \times m$ nonsingular submatrix formed by the columns of A with indices in J. Associated with $J \in \mathcal{J}(A)$, for a diagonal $n \times n$ matrix D, we denote by D_J the $m \times m$ diagonal matrix formed by the elements of D that have row and column indices in J. Similarly, for a vector g of dimension n, we denote by g_J the vector of dimension m with the components of g that have indices in J. The slightly different meanings of A_J, D_J, and g_J are used in order not to make the notation more complicated than necessary. For an example clarifying the concepts, see Forsgren [10, p. 766].

The analogous notation is used for an $m \times n$ matrix A of full row rank and an $n \times r$ matrix U of full row rank in that we associate $\mathcal{J}(AU)$ with the collection of sets of column indices corresponding to nonsingular $m \times m$ submatrices of AU. Associated with $J \in \mathcal{J}(AU)$, for a diagonal $r \times r$ matrix D, we denote by D_J the $m \times m$ diagonal matrix formed by the elements of D that have row and column indices in J. Similarly, for a vector g of dimension r, we denote by g_J the vector of dimension m with the components of g that have indices in J. Since column indices of AU are also column indices of U, for $J \in \mathcal{J}(AU)$, we denote by U_J the $n \times m$ submatrix of full column rank formed by the columns of U with indices in J. Note that each element of $\mathcal{J}(A)$ as well as each element of $\mathcal{J}(AU)$ is a collection of m indices.

2. Background. In this section, we review some fundamental results. The following theorem, which states that the solution of a diagonally weighted linear least-squares problem can be expressed as a certain convex combination, is the basis for our results. As far as we know, it was originally given by Dikin [8] who used it in the convergence analysis of the interior point method for linear programming he proposed [7]. The proof of the theorem is based on the Cauchy–Binet formula and Cramer's rule.

THEOREM 2.1 (Dikin [8]). *Let A be an $m \times n$ matrix of full row rank, let g be a vector of dimension n, and let D be a positive definite diagonal $n \times n$ matrix. Then,*

$$(ADA^T)^{-1}ADg = \sum_{J \in \mathcal{J}(A)} \left(\frac{\det(D_J)\det(A_J)^2}{\sum_{K \in \mathcal{J}(A)} \det(D_K)\det(A_K)^2} \right) A_J^{-T} g_J,$$

where $\mathcal{J}(A)$ is the collection of sets of column indices associated with nonsingular $m \times m$ submatrices of A.

Proof. See, e.g., Ben-Tal and Teboulle [3, Corollary 2.1]. ∎

Theorem 2.1 implies that if the weight matrix is diagonal and positive definite, then the solution to the weighted least-squares problem (1.1) lies in the convex hull of the *basic solutions* formed by satisfying m linearly independent equations. Hence, this theorem provides an expression on the supremum of $\|(ADA^T)^{-1}AD\|$ for D diagonal and positive definite, as the following corollary shows.

COROLLARY 2.2. *Let A be an $m \times n$ matrix of full row rank, and let $\mathcal{D}_+$ denote the set of positive definite diagonal $n \times n$ matrices. Then,*

$$\sup_{D \in \mathcal{D}_+} \|(ADA^T)^{-1}AD\| = \max_{J \in \mathcal{J}(A)} \|A_J^{-T}\|,$$

where $\mathcal{J}(A)$ is the collection of sets of column indices associated with nonsingular $m \times m$ submatrices of A.

Proof. See, e.g., Forsgren [10, Corollary 2.2]. ∎

The boundedness has been discussed by a number of authors over the years; see, e.g., Ben-Tal and Teboulle [3], O'Leary [22], Stewart [24], and Todd [25]. Theorem 2.1 can be generalized to the case where the weight matrix is an arbitrary symmetric, not necessarily diagonal, matrix such that AWA^T is nonsingular. The details are given in the following theorem.

THEOREM 2.3 (Forsgren [10]). *Let A be an $m \times n$ matrix of full row rank, and let W be a symmetric $n \times n$ matrix such that AWA^T is nonsingular. Suppose $W = UDU^T$, where D is diagonal. Then,*

$$(AWA^T)^{-1}AW = \sum_{J \in \mathcal{J}(AU)} \left(\frac{\det(D_J)\det(AU_J)^2}{\sum_{K \in \mathcal{J}(AU)} \det(D_K)\det(AU_K)^2} \right) (AU_J)^{-T} U_J^T,$$

where $\mathcal{J}(AU)$ is the collection of sets of column indices associated with nonsingular $m \times m$ submatrices of AU.

Proof. See Forsgren [10, Theorem 3.1]. ∎

3. Nonnegative combinations of positive semidefinite matrices. Let A be an $m \times n$ matrix of full row rank and assume that we are given an $n \times n$ symmetric weight matrix $W(\alpha)$, which depends on a vector $\alpha \in \mathbb{R}^t$ for some t. If $W(\alpha)$ can be decomposed as $W(\alpha) = UD(\alpha)U^T$, where U does *not* depend on α and $D(\alpha)$ is diagonal, Theorem 2.3 can be applied, provided $AW(\alpha)A^T$ is nonsingular and the matrices $(AU_J)^{-T}U_J^T$ involved do not depend on α. If, in addition $D(\alpha) \succeq 0$, then the linear combination of Theorem 2.3 is a convex combination. Consequently, the norm remains bounded as long as the supremum is taken over a set of values of α for which $AW(\alpha)A^T \succ 0$ and $D(\alpha) \succeq 0$. In particular, we are interested in the case where a set of positive semidefinite and symmetric matrices, W_i, $i = 1, \ldots, t$, are given and $W(\alpha)$ is defined as $W(\alpha) = \sum_{i=1}^t \alpha_i W_i$. The following two lemmas and associated corollary concern the decomposition of $W(\alpha)$. The first lemma concerns the set of all possible decompositions of a positive semidefinite matrix W as $W = UU^T$ and the relation between different decompositions of this type.

LEMMA 3.1. *Let W be a symmetric positive semidefinite $n \times n$ matrix of rank r, and let $\bar{\mathcal{U}} = \{U \in \mathbb{R}^{n\times r} : UU^T = W\}$. Then, $\bar{\mathcal{U}}$ is nonempty and compact. Further, if U and $\widetilde{U}$ belong to $\bar{\mathcal{U}}$, then there is an $r \times r$ orthogonal matrix Q such that $U = \widetilde{U}Q$.*

Proof. It is possible to decompose W as $W = UU^T$, where U is an $n \times r$ matrix of full column rank, for example, using a Cholesky factorization with symmetric interchanges; see, e.g., Golub and Van Loan [14, section 4.2.9]. Therefore, $\bar{\mathcal{U}}$ is nonempty. If U and $\widetilde{U}^T$ both belong to $\bar{\mathcal{U}}$, then

$$U^T x = 0 \;\Leftrightarrow\; UU^T x = 0 \;\Leftrightarrow\; \widetilde{U}\widetilde{U}^T x = 0 \;\Leftrightarrow\; \widetilde{U}^T x = 0.$$

Hence, U^T and $\widetilde{U}^T$ have the same null space, which implies that the range spaces of U and $\widetilde{U}$ are the same. Therefore, there is a nonsingular $r \times r$ matrix M such that $U = \widetilde{U}M$, from which it follows that $\widetilde{U}\widetilde{U}^T = \widetilde{U}MM^T\widetilde{U}^T$. Premultiplying this equation by $\widetilde{U}^T$ and postmultiplying it by $\widetilde{U}$ gives

$$\widetilde{U}^T\widetilde{U}\widetilde{U}^T\widetilde{U} = \widetilde{U}^T\widetilde{U}MM^T\widetilde{U}^T\widetilde{U}. \tag{3.1}$$

Since $\widetilde{U}^T\widetilde{U}$ is nonsingular, (3.1) gives $MM^T = I$. Compactness is established by proving boundedness and closedness. Boundedness holds because $\|U^T e_i\|_2^2 = W_{ii}$, $i = 1, \ldots, n$, where e_i is the ith unit vector. Let $\{U^{(i)}\}_{i=1}^{\infty}$ be a sequence converging to U^* such that $U^{(i)} \in \bar{\mathcal{U}}$ for all i. From the continuity of matrix multiplication, U^* belongs to $\bar{\mathcal{U}}$, and the closedness of $\bar{\mathcal{U}}$ follows. □

A consequence of this lemma is that we can decompose each W_i, $i = 1, \ldots, t$, as stated in the following corollary.

COROLLARY 3.2. *For $i = 1, \ldots, t$, let W_i be an $n \times n$ symmetric positive semidefinite matrix of rank r_i. Let $r = \sum_{i=1}^t r_i$. Then*

$$\mathcal{U} = \left\{ U \in \mathbb{R}^{n\times r} : U = \begin{pmatrix} U_1 & U_2 & \cdots & U_t \end{pmatrix}, U_i \in \mathbb{R}^{n\times r_i}, U_i U_i^T = W_i, i = 1, \ldots, t \right\}$$

is a well-defined compact subset of $\mathbb{R}^{n\times r}$. Furthermore, if U and $\widetilde{U}$ belong to $\mathcal{U}$, then, for $i = 1, \ldots, t$, there are orthogonal $r_i \times r_i$ matrices Q_i such that $U_i = \widetilde{U}_i Q_i$.

Proof. The result follows by applying Lemma 3.1 to each W_i. □

It should be noted that $\mathcal{U}$ depends on the matrices W_i. This dependence will be suppressed in order to not make the notation more complicated than necessary. From Corollary 3.2, we get a decomposition result for matrices that are nonnegative linear combinations of symmetric positive semidefinite matrices, as is stated in the following lemma. It shows that if we are given a set of positive semidefinite and symmetric matrices, W_i, $i = 1, \ldots, t$, and $W(\alpha)$ is defined as $W(\alpha) = \sum_{i=1}^t \alpha_i W_i$, then we can decompose $W(\alpha)$ into the form $W(\alpha) = UD(\alpha)U^T$, where U does not depend on α and $D(\alpha)$ is diagonal.

LEMMA 3.3. *For $\alpha \in \mathbb{R}^t$, let $W(\alpha) = \sum_{i=1}^t \alpha_i W_i$, where W_i, $i = 1, \ldots, t$, are symmetric positive semidefinite $n \times n$ matrices. Further, let $\mathcal{U}$ be associated with $W_i, i = 1, \ldots, t$, according to Corollary 3.2, and for each i, let r_i denote $\operatorname{rank}(W_i)$ and let I_i be an identity matrix of dimension r_i. Then $W(\alpha)$ may be decomposed as*

$$W(\alpha) = UD(\alpha)U^T,$$

where U is any matrix in $\mathcal{U}$ and $D(\alpha) = \operatorname{diag}(\alpha_1 I_1, \alpha_2 I_2, \ldots, \alpha_t I_t)$.

Proof. Corollary 3.2 shows that we may write

$$W(\alpha) = \sum_{i=1}^{t} \alpha_i W_i = \sum_{i=1}^{t} \alpha_i U_i U_i^T = U D(\alpha) U^T,$$

where U is an arbitrary matrix in $\mathcal{U}$ and $D(\alpha) = \mathrm{diag}(\alpha_1 I_1, \alpha_2 I_2, \ldots, \alpha_t I_t)$. ∎

Note that $D(\alpha)$ is positive semidefinite if $\alpha \geq 0$. An application of Theorem 2.3 to the decomposition of Lemma 3.3 now gives the boundedness result for nonnegative combinations of positive semidefinite matrices, as stated in the following proposition.

PROPOSITION 3.4. *Let A be an $m \times n$ matrix of full row rank. For $\alpha \in \mathbb{R}^t$, $\alpha \geq 0$, let $W(\alpha) = \sum_{i=1}^t \alpha_i W_i$, where W_i, $i = 1, \ldots, t$, are symmetric positive semidefinite $n \times n$ matrices. If $W(\alpha)$ is decomposed as $W(\alpha) = U D(\alpha) U^T$, according to Lemma* 3.3, *then for $\alpha \geq 0$ and $A W(\alpha) A^T \succ 0$,*

$$(AW(\alpha)A^T)^{-1} AW(\alpha) = \sum_{J \in \mathcal{J}(AU)} \left(\frac{\det(D_J(\alpha)) \det(AU_J)^2}{\sum_{K \in \mathcal{J}(AU)} \det(D_K(\alpha)) \det(AU_K)^2} \right) (AU_J)^{-T} U_J^T.$$

Furthermore,

$$\sup_{\substack{\alpha \geq 0: \\ AW(\alpha)A^T \succ 0}} \|(AW(\alpha)A^T)^{-1} AW(\alpha)\| \leq \min_{U \in \mathcal{U}} \max_{J \in \mathcal{J}(AU)} \|(AU_J)^{-T} U_J^T\|, \tag{3.2}$$

where $\mathcal{J}(AU)$ is the collection of sets of column indices associated with nonsingular $m \times m$ submatrices of AU, and $\mathcal{U}$ is associated with W_i, $i = 1, \ldots, t$, according to Corollary 3.2.

Proof. If $AW(\alpha)A^T \succ 0$, Theorem 2.3 immediately gives

$$(AW(\alpha)A^T)^{-1} AW(\alpha) = \sum_{J \in \mathcal{J}(AU)} \left(\frac{\det(D_J(\alpha)) \det(AU_J)^2}{\sum_{K \in \mathcal{J}(AU)} \det(D_K(\alpha)) \det(AU_K)^2} \right) (AU_J)^{-T} U_J^T.$$

Since $\alpha \geq 0$, it follows that $D(\alpha) \succeq 0$. Consequently, $\det(D_J(\alpha)) \geq 0$ for all $J \in \mathcal{J}(AU)$. Thus, the above expression gives

$$\sup_{\alpha \geq 0: AW(\alpha)A^T \succ 0} \|(AW(\alpha)A^T)^{-1} AW(\alpha)\| \leq \max_{J \in \mathcal{J}(AU)} \|(AU_J)^{-T} U_J^T\|.$$

Since this result holds for all $U \in \mathcal{U}$, it holds when taking the infimum over $U \in \mathcal{U}$. To show that the infimum is attained, let

$$f_J(U) = \begin{cases} \|(AU_J)^{-T} U_J^T\| & \text{if } \det(AU_J) \neq 0, \\ 0 & \text{otherwise} \end{cases}$$

for every J that is a subset of $\{1, \ldots, n\}$ such that $|J| = m$. For a fixed J, f_J is continuous at every $\widetilde{U}$ such that $\det(A\widetilde{U}_J) \neq 0$. Further, at $\widetilde{U}$ such that $A\widetilde{U}_J$ is singular, f_J is a lower semicontinuous function; see, e.g., Royden [23, p. 51]. Hence, f_J is lower semicontinuous everywhere. Due to the construction of $f_J(U)$,

$$\max_{J \in \mathcal{J}(AU)} \|(AU_J)^{-T} U_J^T\| = \max_{J : |J| = m} f_J(U).$$

The maximum of a finite collection of lower semicontinuous functions is lower semicontinuous; see, e.g., Royden [23, p. 51], and the set $\mathcal{U}$ is compact by Corollary 3.2. Therefore, the infimum is attained (see, e.g., Royden [23, p. 195]) and the proof is complete. ∎

Note that Proposition 3.4 as special cases includes two known cases: (i) the diagonal matrices, where $W(\alpha) = \sum_{i=1}^n \alpha_i e_i e_i^T$; and (ii) the diagonally dominant matrices, where

$$W(\alpha) = \sum_{i=1}^{n} \alpha_i e_i e_i^T + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left(\alpha_{ij}^{+} (e_i + e_j)(e_i + e_j)^T + \alpha_{ij}^{-} (e_i - e_j)(e_i - e_j)^T \right).$$

In both these cases, the supremum bound of (3.2) is sharp. This is because all the matrices whose nonnegative linear combinations form the weight matrices are of rank one. In that case, the minimum over U in (3.2) is not necessary since it follows from Corollary 3.2 that the columns of U are unique up to multiplication by ± 1. Hence, $D(\alpha)$ may be adjusted so as to give weight one to the submatrix AU_J for which the maximum of the right-hand side of (3.2) is achieved and to give negligible weight to the other submatrices. In general, when not all matrices whose nonnegative linear combinations form the weight matrix have rank one, it is an open question if the supremum bound is sharp.

4. Inversion of the weight matrix. For a constant positive semidefinite matrix H, our goal is to obtain a bound on $\|(A(H+D)^{-1}A^T)^{-1}A(H+D)^{-1}\|$ when D is an arbitrary positive definite diagonal matrix. One major obstacle in applying Theorem 2.3 is the inverse in the weight matrix $(H+D)^{-1}$. The following proposition and its subsequent corollary and lemma provide a solution to this problem.

PROPOSITION 4.1. *Suppose that an $n \times n$ orthogonal matrix Q is partitioned as $Q = (Z\ Y)$, where Z is an $n \times s$ matrix and $2s \le n$. Further, let W be a symmetric nonsingular $n \times n$ matrix such that $Z^T W^{-1} Z$ and $Y^T W Y$ are nonsingular. Then*

$$(Y^T W Y)^{-1} Y^T W Z = -((Z^T W^{-1} Z)^{-1} Z^T W^{-1} Y)^T$$

and

$$\begin{aligned}
\sigma_i^2((Y^T W Y)^{-1} Y^T W) &= \sigma_i^2((Z^T W^{-1} Z)^{-1} Z^T W^{-1}) \\
&= 1 + \sigma_i^2((Z^T W^{-1} Z)^{-1} Z^T W^{-1} Y) \\
&= 1 + \sigma_i^2((Y^T W Y)^{-1} Y^T W Z), \quad i = 1, \dots, s, \\
\sigma_i((Y^T W Y)^{-1} Y^T W) &= 1, \quad i = s+1, \dots, n-s.
\end{aligned}$$

Proof. The orthogonality of Q ensures that $Y^T Z = 0$ and $ZZ^T + YY^T = I$. This gives

$$0 = Y^T Z = Y^T W (ZZ^T + YY^T) W^{-1} Z = Y^T W Z Z^T W^{-1} Z + Y^T W Y Y^T W^{-1} Z,$$

and hence

$$(Y^T W Y)^{-1} Y^T W Z = -((Z^T W^{-1} Z)^{-1} Z^T W^{-1} Y)^T, \tag{4.1}$$

proving the first part of the proposition.

Since $Z^T W^{-1} Z$ and $Y^T W Y$ are nonsingular, we may write

$$(Z^T W^{-1} Z)^{-1} Z^T W^{-1} \begin{pmatrix} Z & Y \end{pmatrix} = \begin{pmatrix} I & (Z^T W^{-1} Z)^{-1} Z^T W^{-1} Y \end{pmatrix}, \tag{4.2a}$$

$$(Y^T W Y)^{-1} Y^T W \begin{pmatrix} Z & Y \end{pmatrix} = \begin{pmatrix} (Y^T W Y)^{-1} Y^T W Z & I \end{pmatrix}. \tag{4.2b}$$

The orthogonality of Q ensures that

$$\sigma_i((Z^TW^{-1}Z)^{-1}Z^TW^{-1}Q) = \sigma_i((Z^TW^{-1}Z)^{-1}Z^TW^{-1}), \quad i = 1, \ldots, s. \tag{4.3}$$

We also have

$$\sigma_i^2 \begin{pmatrix} I & (Z^TW^{-1}Z)^{-1}Z^TW^{-1}Y \end{pmatrix} = 1 + \sigma_i^2 \begin{pmatrix} (Z^TW^{-1}Z)^{-1}Z^TW^{-1}Y \end{pmatrix}, \tag{4.4}$$

$i = 1, \ldots, s$. A combination of (4.2a), (4.3), and (4.4) gives

$$\sigma_i^2((Z^TW^{-1}Z)^{-1}Z^TW^{-1}) = 1 + \sigma_i^2((Z^TW^{-1}Z)^{-1}Z^TW^{-1}Y), \quad i = 1, \ldots, s. \tag{4.5}$$

An analogous argument applied to (4.2b), taking into account that $2s \le n$, gives

$$\sigma_i^2((Y^TWY)^{-1}Y^TW) = 1 + \sigma_i^2((Y^TWY)^{-1}Y^TWZ), \quad i = 1, \ldots, s, \tag{4.6a}$$
$$\sigma_i^2((Y^TWY)^{-1}Y^TW) = 1, \quad i = s+1, \ldots, n-s. \tag{4.6b}$$

The second part of the proposition follows by a combination of (4.1), (4.5), and (4.6). $\square$

In particular, Proposition 4.1 gives the equivalence between the Euclidean norms of a projection and the projection onto the complementary space using the inverse weight matrix, given that the matrices used to represent the spaces are orthogonal. This is shown in the following corollary.

Corollary 4.2. *Suppose that an $n \times n$ orthogonal matrix Q is partitioned as $Q = (Z\ Y)$, where Y is an $n \times m$ matrix. Further, let W be a symmetric nonsingular $n \times n$ matrix such that $Z^TW^{-1}Z$ and Y^TWY are nonsingular. Then*

$$\|(Y^TWY)^{-1}Y^TW\|_2 = \|(Z^TW^{-1}Z)^{-1}Z^TW^{-1}\|_2.$$

Further, let $\mathcal{W}_+$ denote the set of $n \times n$ positive definite symmetric matrices, and let $\mathcal{W} \subseteq \mathcal{W}_+$. Then,

$$\sup_{W \in \mathcal{W}} \|(Y^TWY)^{-1}Y^TW\|_2 = \sup_{W \in \mathcal{W}} \|(Z^TW^{-1}Z)^{-1}Z^TW^{-1}\|_2.$$

Proof. If $m \ge n/2$, the first statement follows by letting $i = 1$ in Proposition 4.1. The second statement is a direct consequence of the first one. If $m < n/2$, we may similarly apply Proposition 4.1 after interchanging the roles of Y and Z, and W and W^{-1}. $\square$

As noted above, Corollary 4.2 states the equality between the Euclidean norms of two projections, given that the matrices describing the spaces onto which we project are orthogonal. The following lemma relates the Euclidean norms of the projections when the matrices are not orthogonal.

Lemma 4.3. *Let A be an $m \times n$ matrix of full row rank, and let N be a matrix whose columns form a basis for the null space of A. Further, let W be a symmetric nonsingular $n \times n$ matrix such that $N^TW^{-1}N$ and A^TWA are nonsingular. Then*

$$\frac{\sigma_{n-m}(N)}{\sigma_1(A)} \le \frac{\|(AWA^T)^{-1}AW\|_2}{\|(N^TW^{-1}N)^{-1}N^TW^{-1}\|_2} \le \frac{\sigma_1(N)}{\sigma_m(A)}.$$

Proof. Let $Q = (Z\ Y)$ be an orthogonal matrix such that the columns of Z form a basis for the null space of A. Then, there are nonsingular matrices R_Z and R_Y such that $N = ZR_Z$ and $A^T = YR_Y$. Since a matrix norm which is induced from a vector norm is submultiplicative (see, e.g., Horn and Johnson [19, Theorem 5.6.2]) this gives

$$\frac{1}{\|R_Z\|} \le \frac{\|(N^TW^{-1}N)^{-1}N^TW^{-1}\|}{\|(Z^TW^{-1}Z)^{-1}Z^TW^{-1}\|} \le \|R_Z^{-1}\|, \tag{4.7a}$$

$$\frac{1}{\|R_Y\|} \le \frac{\|(AWA^T)^{-1}AW\|}{\|(Y^TWY)^{-1}Y^TW\|} \le \|R_Y^{-1}\|. \tag{4.7b}$$

If the Euclidean norm is used, the bounds in (4.7) can be expressed in terms of singular values of A and N since Y and Z are orthogonal matrices, i.e.,

$$\|R_Z\|_2 = \sigma_1(N), \quad \|R_Z^{-1}\|_2 = 1/\sigma_{n-m}(N), \tag{4.8a}$$

$$\|R_Y\|_2 = \sigma_1(A), \quad \|R_Y^{-1}\|_2 = 1/\sigma_m(A). \tag{4.8b}$$

A combination of Corollary 4.2, (4.7), and (4.8) gives the stated result. □

If the weight matrix is allowed to vary over some subset of the positive definite symmetric matrices, it follows from Lemma 4.3 that the norm of the projection onto a subspace is bounded if and only if the norm of the projection onto the orthogonal complement is bounded when using inverses of the weight matrices. This is made precise in the following corollary.

COROLLARY 4.4. *Let $\mathcal{W}_+$ denote the set of $n \times n$ positive definite symmetric matrices, and let $\mathcal{W} \subseteq \mathcal{W}_+$. Let A be an $m \times n$ matrix of full row rank, and let N be a matrix whose columns form a basis for the null space of A. Then*

$$\sup_{W\in\mathcal{W}} \|(AWA^T)^{-1}AW\| < \infty \quad \text{if and only if} \quad \sup_{W\in\mathcal{W}} \|(N^TW^{-1}N)^{-1}N^TW^{-1}\| < \infty.$$

In particular,

$$\frac{\sigma_{n-m}(N)}{\sigma_1(A)} \sup_{W\in\mathcal{W}} \|(N^TW^{-1}N)^{-1}N^TW^{-1}\|_2 \le \sup_{W\in\mathcal{W}} \|(AWA^T)^{-1}AW\|_2,$$

$$\sup_{W\in\mathcal{W}} \|(AWA^T)^{-1}AW\|_2 \le \frac{\sigma_1(N)}{\sigma_m(A)} \sup_{W\in\mathcal{W}} \|(N^TW^{-1}N)^{-1}N^TW^{-1}\|_2.$$

Proof. The second statement follows by multiplying the inequalities in Lemma 4.3 by $\|(N^TW^{-1}N)^{-1}N^TW^{-1}\|_2$ and then taking the supremum of the three expressions. The first statement of the corollary then follows from the equivalence of matrix norms that are induced from vector norms; see, e.g., Horn and Johnson [19, Theorem 5.6.18]. □

5. Inversion and nonnegative combination. Let A be an $m \times n$ matrix of full row rank, and let Z be a matrix whose columns form an orthonormal basis for the null space of A. Further, let $M(\alpha) = \sum_{i=1}^{t} \alpha_i M_i$, where M_i, $i = 1, \dots, t$, are given symmetric positive semidefinite $n \times n$ matrices. In section 3 the weight matrix was assumed to be the nonnegative combination of symmetric positive semidefinite matrices. This section concerns weight matrices that are the inverse of such combinations, i.e., where the weight matrix is the inverse of $M(\alpha)$. Further, if the problem is originally posed as the KKT system, cf. (1.4),

$$\begin{pmatrix} M(\alpha) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} r(\alpha) \\ \pi(\alpha) \end{pmatrix} = \begin{pmatrix} g \\ 0 \end{pmatrix}, \tag{5.1}$$

it makes sense to study the problem under the assumption that $Z^T M(\alpha) Z \succ 0$ since in our situation, $Z^T M(\alpha) Z \succ 0$ if and only if the matrix of (5.1) is nonsingular; see Gould [16, Lemma 3.4]. Note that $Z^T M(\alpha) Z \succ 0$ is a weaker assumption than $M(\alpha) \succ 0$, which is necessary if the least-squares formulation is to be valid. A combination of Proposition 3.4 and Lemma 4.3 shows that $\pi(\alpha)$ remains bounded under the above-mentioned assumptions. This is stated in the following theorem, which is the main result of this paper.

THEOREM 5.1. *Let A be an $m \times n$ matrix of full row rank and let g be an n-vector. Further, let Z be a matrix whose columns form an orthonormal basis for the null space of A. For $\alpha \in \mathbb{R}^t$, $\alpha \geq 0$, let $M(\alpha) = \sum_{i=1}^t \alpha_i M_i$, where M_i, $i = 1, \ldots, t$, are symmetric positive semidefinite $n \times n$ matrices. Further, let $r(\alpha)$ and $\pi(\alpha)$ satisfy*

$$\begin{pmatrix} M(\alpha) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} r(\alpha) \\ \pi(\alpha) \end{pmatrix} = \begin{pmatrix} g \\ 0 \end{pmatrix}.$$

Then,

$$\sup_{\substack{\alpha \geq 0: \\ Z^T M(\alpha) Z \succ 0}} \|\pi(\alpha)\| < \infty. \tag{5.2}$$

In particular, if $Z^T M(\alpha) Z \succ 0$, then

$$\|\pi(\alpha)\|_2 \leq \frac{1}{\sigma_m(A)} \|(Z^T M(\alpha) Z)^{-1} Z^T M(\alpha)\|_2 \|g\|_2. \tag{5.3}$$

Finally, if $M(\alpha)$ is decomposed according to Lemma 3.3*, then*

$$\sup_{\substack{\alpha \geq 0: \\ Z^T M(\alpha) Z \succ 0}} \|\pi(\alpha)\|_2 \leq \frac{1}{\sigma_m(A)} \min_{U \in \mathcal{U}} \max_{J \in \mathcal{J}(Z^T U)} \|(Z^T U_J)^{-T} U_J^T\|_2 \|g\|_2, \tag{5.4}$$

where $\mathcal{J}(Z^T U)$ is the collection of sets of column indices associated with nonsingular $m \times m$ submatrices of $Z^T U$, and $\mathcal{U}$ is associated with M_i, $i = 1, \ldots, t$, according to Corollary 3.2*.*

Proof. For $\alpha \geq 0$ and $\epsilon > 0$, $M(\alpha) + \epsilon I \succ 0$. Therefore,

$$\pi(\alpha, \epsilon) = (A(M(\alpha) + \epsilon I)^{-1} A^T)^{-1} A(M(\alpha) + \epsilon I)^{-1} g$$

is well defined. By Lemma 4.3 it follows that

$$\begin{aligned} \|\pi(\alpha, \epsilon)\|_2 &\leq \|(A(M(\alpha) + \epsilon I)^{-1} A^T)^{-1} A(M(\alpha) + \epsilon I)^{-1}\|_2 \|g\|_2 \\ &\leq \frac{1}{\sigma_m(A)} \|(Z^T(M(\alpha) + \epsilon I) Z)^{-1} Z^T (M(\alpha) + \epsilon I)\|_2 \|g\|_2. \end{aligned} \tag{5.5}$$

For α such that $Z^T M(\alpha) Z \succ 0$, the matrix in the system of equations defining $\pi(\alpha)$ and $r(\alpha)$ is nonsingular; see Gould [16, Lemma 3.4]. Then, the implicit function theorem implies that $\lim_{\epsilon \to 0^+} \pi(\alpha, \epsilon) = \pi(\alpha)$. Therefore, letting $\epsilon \to 0^+$ in (5.5) gives (5.3). Taking the supremum over α such that $\alpha \geq 0$ and $Z^T M(\alpha) Z \succ 0$ and using Proposition 3.4 gives (5.4), from which (5.2) follows upon observing that all norms on a real finite-dimensional vector space are equivalent; see, e.g., Horn and Johnson [19, Corollary 5.4.5]. □

As a consequence of Theorem 5.1, we are now able to prove the boundedness of the projection operator for the application of primal-dual interior methods to convex quadratic programming described in section 1.1.

COROLLARY 5.2. *Let H be a positive semidefinite symmetric $n \times n$ matrix, let A be an $m \times n$ matrix of full row rank, and let $\mathcal{D}_+$ denote the space of positive definite diagonal $n \times n$ matrices. Then,*

$$\sup_{D \in \mathcal{D}_+} \|(A(H+D)^{-1}A^T)^{-1}A(H+D)^{-1}\| < \infty.$$

Proof. If $M(\alpha) \succ 0$, then $\pi(\alpha)$ of Theorem 5.1 satisfies

$$\pi(\alpha) = (AM(\alpha)^{-1}A^T)^{-1}AM(\alpha)^{-1}g.$$

Since $\{\alpha \geq 0 : M(\alpha) \succ 0\} \subseteq \{\alpha \geq 0 : Z^T M(\alpha) Z \succ 0\}$, Theorem 5.1 implies that $\pi(\alpha)$ is bounded. This holds for any vector g, and hence

$$\sup_{\alpha \geq 0 : M(\alpha) \succ 0} \|(AM(\alpha)^{-1}A^T)^{-1}AM(\alpha)^{-1}\| < \infty. \tag{5.6}$$

The stated result follows by applying (5.6) with $M_i = e_i e_i^T$, $i = 1, \ldots, m$, $M_{m+1} = H$, and letting $\alpha_{m+1} = 1$. ∎

For convenience in notation, it has been assumed that all variables of the convex quadratic program are subject to bounds. It can be observed that the analogous results hold when some variables are not subject to bounds. In this situation, M of (1.4) may be partitioned as

$$M = \begin{pmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22} \end{pmatrix} + \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

where H is symmetric and positive semidefinite and D_{11} is diagonal and positive definite. Let A be partitioned conformally with M as $A = (A_1 \ \ A_2)$. Then, (1.4) has a unique solution as long as there is no nonzero p_2 such that $A_2 p_2 = 0$ and $p_2^T H_{22} p_2 = 0$; see Gould [16, Lemma 3.4]. Hence, under this additional assumption, Theorem 5.1 can be applied to bound $\|\pi(\alpha)\|$ as D_{11} varies over the set of positive definite diagonal matrices.

6. Summary. It has been shown that results concerning the boundedness of $(AWA^T)^{-1}AW$ for A of full row rank and W diagonal, or diagonally dominant, and symmetric positive definite can be extended to a more general case where W is a nonnegative linear combination of a set of symmetric positive semidefinite matrices such that $AWA^T \succ 0$. Further, boundedness has been shown for the projection onto the null space of A using as the weight matrix the inverse of a nonnegative linear combination of a number of symmetric positive semidefinite matrices. This result has been used to show boundedness of a projection operator arising in a primal-dual interior method for convex quadratic programming.

The main tools for deriving these results have been the explicit formula for the solution of a weighted linear least-squares problem given by Dikin [8] and the relation between a projection onto a subspace with a certain weight matrix and the projection onto the orthogonal complement using the inverse weight matrix.

An interesting question that is left open is whether or not the explicit bounds that are given are sharp. In the case where all the matrices whose nonnegative linear combination form the weight matrix are of rank one, the bounds are sharp. In the general case, this is an open question. On a higher level, an interesting question is whether the results of this paper can be utilized to give new complexity bounds for quadratic programming, analogous to the case of linear programming; see, e.g., Vavasis and Ye [27, section 9].

Appendix. Relationship to partitioned orthogonal matrices. In this appendix we review a result by Gonzaga and Lara [15] concerning diagonally weighted projections onto orthogonally complementary subspaces and combine this result with a result concerning singular values of submatrices of orthogonal matrices. It was these results in fact which lead to the more general results relating weighted projection onto a subspace and the projection onto its complementary subspace using the inverse weight matrix, as described in section 4.

Gonzaga and Lara [15] state that if Y is an $n \times m$ orthogonal matrix and Z is a matrix whose columns form an orthonormal basis for the null space of Y^T, then

$$\sup_{D \in \mathcal{D}_+} \|(Y^T D Y)^{-1} Y^T D\| = \sup_{D \in \mathcal{D}_+} \|(Z^T D Z)^{-1} Z^T D\|,$$

where $\mathcal{D}_+$ is the set of positive definite diagonal $n \times n$ matrices. They use a geometric approach to prove this result. We note that Corollary 4.2, specialized to the case of diagonal positive definite weight matrices, allows us to state the same result. Furthermore, we obtain an explicit expression for the supremum by Corollary 2.2. The following corollary summarizes this result.

Corollary A.1. *Suppose that an $n \times n$ orthogonal matrix Q is partitioned as $Q = (Z\ Y)$, where Y is an $n \times m$ matrix. Let $\mathcal{D}_+$ denote the set of diagonal positive definite $n \times n$ matrices. Then,*

$$\begin{aligned} &\sup_{D \in \mathcal{D}_+} \|(Z^T D Z)^{-1} Z^T D\|_2 = \max_{J \in \mathcal{J}(Z^T)} \frac{1}{\sigma_{\min}(Z_J)} \\ &= \sup_{D \in \mathcal{D}_+} \|(Y^T D Y)^{-1} Y^T D\|_2 = \max_{\tilde{J} \in \mathcal{J}(Y^T)} \frac{1}{\sigma_{\min}(Y_{\tilde{J}})}, \end{aligned}$$

where $\mathcal{J}(Z^T)$ is the collection of sets of column indices associated with nonsingular $(n-m) \times (n-m)$ submatrices of Z^T and $\mathcal{J}(Y^T)$ is the collection of sets of column indices associated with nonsingular $m \times m$ submatrices of Y^T.

Proof. Since $D \in \mathcal{D}_+$ if and only if $D^{-1} \in \mathcal{D}_+$, Corollary 4.2 shows that

$$\sup_{D \in \mathcal{D}_+} \|(Z^T D Z)^{-1} Z^T D\|_2 = \sup_{D \in \mathcal{D}_+} \|(Y^T D Y)^{-1} Y^T D\|_2.$$

The explicit expressions for the two suprema follow from Corollary 2.2. ∎

Hence, in our setting, we would rather state the result of Gonzaga and Lara [15] in the equivalent form

$$\sup_{D \in \mathcal{D}_+} \|(Y^T D Y)^{-1} Y^T D\| = \sup_{D \in \mathcal{D}_+} \|(Z^T D^{-1} Z)^{-1} Z^T D^{-1}\|$$

with the expressions for the suprema stated in Corollary A.1.

Note that an implication of Corollary A.1 is that if an $n \times n$ orthogonal matrix Q is partitioned as $Q = (Z\ \ Y)$ where Y has m columns, there is a certain relationship between the smallest singular value of all nonsingular $(n-m) \times (n-m)$ submatrices of Z and the smallest singular value of all nonsingular $m \times m$ submatrices of Y. This is in fact a consequence of a more general result, namely, that if Q is partitioned further as

$$Q = \begin{pmatrix} Z_1 & Y_1 \\ Z_2 & Y_2 \end{pmatrix}, \tag{A.1}$$

where Z_1 is $(n-m) \times (n-m)$, then all singular values of Z_1 and Y_2 that are less than one are identical. This in turn is a consequence of properties of singular values of submatrices of orthogonal matrices that can be obtained by the CS-decomposition of an orthogonal matrix; see, e.g., Golub and Van Loan [14, section 2.6.4].

This result relating the singular values of Z_1 and Y_2 of (A.1) implies the existence of J and $\tilde{J}$, which are complementary subsets of $\{1, \dots, n\}$, for which the maxima in Corollary A.1 are achieved. This observation lead us to the result that

$$\|(Y^T D Y)^{-1} Y^T D\|_2 = \|(Z^T D^{-1} Z)^{-1} Z^T D^{-1}\|_2$$

for any positive definite diagonal D. Subsequently, this result was superseded by the more general analysis presented in section 4.

Acknowledgments. We thank the two anonymous referees for their constructive and insightful comments, which significantly improved the presentation.

REFERENCES

[1] R. H. Bartels, G. H. Golub, and M. A. Saunders, *Numerical techniques in mathematical programming*, in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian, and K. Ritter, eds., Academic Press, New York, 1970, pp. 123–176.

[2] A. Ben-Israel, *A volume associated with $m \times n$ matrices*, Linear Algebra Appl., 167 (1992), pp. 87–111.

[3] A. Ben-Tal and M. Teboulle, *A geometric property of the least squares solution of linear equations*, Linear Algebra Appl., 139 (1990), pp. 165–170.

[4] Å. Björck, *Pivoting and stability in the augmented system method*, in Numerical Analysis 1991, Proceedings of the 14th Dundee Conference, University of Dundee, UK, D. F. Griffiths and G. A. Watson, eds., Longman, Wiley, Harlow, UK, New York, 1991, pp. 1–16.

[5] Å. Björck and C. C. Paige, *Solution of augmented linear systems using orthogonal factorizations*, BIT, 34 (1994), pp. 1–24.

[6] E. Y. Bobrovnikova and S. A. Vavasis, *A norm bound for projections with complex weights*, Linear Algebra Appl., 307 (2000), pp. 69–75.

[7] I. I. Dikin, *Iterative solution of problems of linear and quadratic programming*, Soviet Math. Dokl., 8 (1967), pp. 674–675.

[8] I. I. Dikin, *On the speed of an iterative process*, Upravlyaemye Sistemi, 12 (1974), pp. 54–60.

[9] I. S. Duff, N. I. M. Gould, J. K. Reid, J. A. Scott, and K. Turner, *The Factorization of Sparse Symmetric Indefinite Matrices*, IMA J. Numer. Anal., 11 (1991), pp. 181–204.

[10] A. Forsgren, *On linear least-squares problems with diagonally dominant weight matrices*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 763–788.

[11] A. Forsgren, P. E. Gill, and J. R. Shinnerl, *Stability of symmetric ill-conditioned systems arising in interior methods for constrained optimization*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 187–211.

[12] P. E. Gill, W. Murray, and M. H. Wright, *Numerical Linear Algebra and Optimization*, Vol. 1, Addison-Wesley, Redwood City, CA, 1991.

[13] P. E. Gill, M. A. Saunders, and J. R. Shinnerl, *On the stability of the Cholesky factorization for symmetric quasidefinite systems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 35–46.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[15] C. C. Gonzaga and H. L. Lara, *A note on properties of condition numbers*, Linear Algebra Appl., 261 (1997), pp. 269–273.

[16] N. I. M. Gould, *On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem*, Math. Program., 32 (1985), pp. 90–99.

[17] M. Gulliksson and P.-Å. Wedin, *Modifying the QR-decomposition to constrained and weighted linear least squares*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1298–1313.
[18] M. Hanke and M. Neumann, *The geometry of the set of scaled projections*, Linear Algebra Appl., 190 (1993), pp. 137–148.
[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[20] C. L. Lawson and R. J. Hanson, *Solving Least-Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
[21] R. D. C. Monteiro and I. Adler, *Interior path-following primal-dual algorithms. Part* II: *Convex quadratic programming*, Math. Program., 44 (1989), pp. 43–66.
[22] D. P. O'Leary, *On bounds for scaled projections and pseudoinverses*, Linear Algebra Appl., 132 (1990), pp. 115–117.
[23] H. L. Royden, *Real Analysis*, 3rd ed., Macmillan, New York, 1988.
[24] G. W. Stewart, *On scaled projections and pseudoinverses*, Linear Algebra Appl., 112 (1989), pp. 189–193.
[25] M. J. Todd, *A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm*, Oper. Res., 38 (1990), pp. 1006–1018.
[26] S. A. Vavasis, *Stable numerical algorithms for equilibrium systems*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 1108–1131.
[27] S. A. Vavasis and Y. Ye, *A primal-dual interior point method whose running time depends only on the constraint matrix*, Math. Program., 74 (1996), pp. 79–120.
[28] M. Wei, *Upper bound and stability of scaled pseudoinverses*, Numer. Math., 72 (1995), pp. 285–293.
[29] S. J. Wright, *Stability of linear equations solvers in interior-point methods*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1287–1307.
[30] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, PA, 1997.
[31] S. J. Wright, *Stability of augmented system factorizations in interior-point methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 191–222.