SIAM J. OPTIM.
Vol. 15, No. 1, pp. 96–100

UNIFORM BOUNDEDNESS OF A PRECONDITIONED NORMAL MATRIX USED IN INTERIOR-POINT METHODS*

RENATO D. C. MONTEIRO†, JEROME W. O'NEAL†, AND TAKASHI TSUCHIYA‡

Abstract. Solving systems of linear equations with "normal" matrices of the form AD^2A^T is a key ingredient in the computation of search directions for interior-point algorithms. In this article, we establish that a well-known basis preconditioner for such systems of linear equations produces scaled matrices with uniformly bounded condition numbers as D varies over the set of all positive diagonal matrices. In particular, we show that when A is the node–arc incidence matrix of a connected directed graph with one of its rows deleted, then the condition number of the corresponding preconditioned normal matrix is bounded above by $m(n-m+1)$, where m and n are the number of nodes and arcs of the network.

Key words. linear programming, interior-point methods, polynomial bound, network flow problems, condition number, preconditioning, iterative methods for linear equations, normal matrix

AMS subject classifications. 65F35, 90C05, 90C35, 90C51

DOI. 10.1137/S1052623403426398

1. Introduction.

Consider the linear programming (LP) problem $\min\{c^Tx : Ax = b,\ x \geq 0\}$, where $A \in \mathbb{R}^{m\times n}$ has full row rank. Interior-point methods for solving this problem require that systems of linear equations of the form $AD^2A^T\Delta y = r$, where D is a positive diagonal matrix, be solved at every iteration. It often occurs that the "normal" matrix AD^2A^T, while positive definite, becomes increasingly ill-conditioned as one approaches optimality. In fact, it has been proven (e.g., see Kovacevic and Asic [3]) that for degenerate LP problems, the condition number of the normal matrix goes to infinity. Because of the ill-conditioned nature of AD^2A^T, many methods for solving the system $AD^2A^T\Delta y = r$ become increasingly unstable. The problem becomes even more serious when conjugate gradient methods are used to solve this linear system. Hence, the development of suitable preconditioners which keep the condition number of the coefficient matrix of the scaled system under control is of paramount importance. We should note, however, that in practice the ill-conditioning of AD^2A^T generally does not cause difficulty when the system $AD^2A^T\Delta y = r$ is solved using a backward-stable direct solver (see, e.g., [12, 13] and references therein).

In this paper, we analyze a preconditioner for the normal matrix AD^2A^T that has been proposed by Resende and Veiga [7] in the context of the minimum cost network flow problem and subsequently by Oliveira and Sorensen [6] for general LP problems. The preconditioning consists of pre- and postmultiplying AD^2A^T by $D_B^{-1}B^{-1}$ and its transpose, respectively, where B is a suitable basis of A and D_B is the corresponding diagonal submatrix of D. Roughly speaking, B is constructed in such a way that

*Received by the editors May 30, 2003; accepted for publication (in revised form) February 3, 2004; published electronically October 14, 2004.
http://www.siam.org/journals/siopt/15-1/42639.html

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (monteiro@isye.gatech.edu, joneal@isye.gatech.edu). The first author was supported in part by NSF grants CCR-9902010, CCR-0203113, and INT-9910084 and ONR grant N00014-03-1-0401.

‡The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku, Tokyo, 106-8569, Japan (tsuchiya@sun312.ism.ac.jp). This author was supported in part by Japan-U.S. Joint Research Projects of the Japan Society for the Promotion of Science, "Algorithms for Linear Programs over Symmetric Cones."

columns of A corresponding to larger diagonal elements of D have higher priority to be in B. Our main result is that such a preconditioner yields coefficient matrices with a uniformly bounded condition number regardless of the value of the diagonal elements of D. In the context of interior-point methods, this means that the condition number of the preconditioned normal matrix has a bound that does not depend on the current iterate, regardless of whether it is well centered. We also show that when applied to network flow-based LPs, the condition number of the preconditioned normal matrix is bounded by $m(n-m+1)$, where m and n are the number of nodes and arcs of the network.

The bound on the condition number of the preconditioned normal matrix developed in this paper is given in terms of a well-known quantity, commonly denoted by $\bar{\chi}_A$, and defined as

$$\bar{\chi}_A \equiv \sup\{\|A^T(ADA^T)^{-1}AD\| : D \in \mathcal{D}_{++}\},$$

where $\mathcal{D}_{++}$ is the set of all positive diagonal matrices. Its finiteness was first established by Dikin [1]. Subsequently, $\bar{\chi}_A$ has been systematically studied by several authors including Stewart [8], Todd [9], and Todd, Tunçel, and Ye [10], and has also played a fundamental role in the analysis of interior-point algorithms (see, for example, Vavasis and Ye [11] and Monteiro and Tsuchiya [4, 5]).

1.1. Notation and terminology. The following notation is used throughout the paper. The superscript ^{T} denotes transpose. $\mathbb{R}^p$ denotes the p-dimensional Euclidean space. The set of all $p \times q$ matrices with real entries is denoted by $\mathbb{R}^{p\times q}$. The jth component of a vector w is denoted by w_j. The jth column and (i,j)th entry of a matrix F is denoted by F_j and F_{ij}, respectively. Given an ordered index set $\alpha \subseteq \{1,\ldots,p\}$, a vector $w \in \mathbb{R}^p$, and a matrix F with p columns, we denote the subvector $(w_j : j \in \alpha)$ by w_α and the submatrix $[F_j : j \in \alpha]$ by F_α. For a matrix Q, we denote its largest and smallest eigenvalues by $\lambda_{\max}(Q)$ and $\lambda_{\min}(Q)$, respectively. The notation $\|\cdot\|$ denotes the vector Euclidean norm or matrix operator norm, respectively, depending on the context. The Frobenius norm of $Q \in \mathbb{R}^{p\times r}$ is given by $\|Q\|_F \equiv (\sum_{i=1}^p \sum_{j=1}^r Q_{ij}^2)^{1/2}$. For a vector d, Diag (d) is the diagonal matrix having the elements of d on its diagonal.

2. The preconditioner and main results. In this section, we describe the preconditioner studied in this paper and establish its main properties.

We first describe a procedure which, given an $n \times n$ diagonal matrix $D \in \mathcal{D}_{++}$, finds a suitable basis of a full row rank matrix $A \in \mathbb{R}^{m\times n}$ obtained by giving higher priority to the columns of A with larger corresponding diagonal elements of D. The use of this basis as a way to obtain preconditioners of matrices of the form AD^2A^T was originally proposed by Resende and Veiga [7] in the context of minimum cost network flow problems and was subsequently extended by Oliveira and Sorensen [6] to the context of general LP problems. Note that when this procedure is used in the context of minimum cost network flow interior-point methods, it produces a maximum spanning tree for the network, whose arc weights are given by the diagonal elements of D.

Algorithm for determining basis B. Let a pair $(A,d) \in \mathbb{R}^{m\times n} \times \mathbb{R}^n_{++}$ be given such that $\text{rank}(A) = m$. Then:

1. Order the elements of d such that $d_1 \geq \cdots \geq d_n$; order the columns of A accordingly.

2. Let $\mathcal{B} = \emptyset$ and set $l = 1$.

3. While $|\mathcal{B}| < m$, do
(i) If A_l is linearly independent of $\{A_j : j \in \mathcal{B}\}$, set $\mathcal{B} \leftarrow \mathcal{B} \cup \{l\}$.
(ii) $l \leftarrow l + 1$.
4. Let $\mathcal{N} = \{1, \ldots, n\} \setminus \mathcal{B}$, $B = A_{\mathcal{B}}$, and $N = A_{\mathcal{N}}$.

We will refer to a basis B produced by the above scheme as a *maximum weight basis* associated with the pair $(A, d) \in \mathbb{R}^{m\times n} \times \mathbb{R}^n_{++}$. We begin with a technical, but important, lemma.

LEMMA 2.1. *Let $(A, d) \in \mathbb{R}^{m\times n} \times \mathbb{R}^n_{++}$ be given such that* $\operatorname{rank}(A) = m$. *Suppose that B is a maximum weight basis associated with the pair (A, d) and define $D \equiv \operatorname{Diag}(d)$ and $D_{\mathcal{B}} \equiv \operatorname{Diag}(d_{\mathcal{B}})$. Then, for every $j = 1, \ldots, n$,*

$$d_j \, \|D_{\mathcal{B}}^{-1} B^{-1} A_j\| \;\le\; \|B^{-1} A_j\|. \tag{1}$$

As a consequence, we have

$$\|D_{\mathcal{B}}^{-1} B^{-1} AD\| \;\le\; \|D_{\mathcal{B}}^{-1} B^{-1} AD\|_F \;\le\; \|B^{-1} A\|_F \;\le\; \sqrt{m}\, \|B^{-1} A\|. \tag{2}$$

Proof. We first prove (1). For every $j \in \mathcal{B}$, both sides of (1) are the same and hence (1) holds as an equality in this case. Assume now that $j \in \mathcal{N}$. We consider the following two distinct cases: (i) A_j was not considered to enter the basis B in step 3 of the above scheme, and (ii) A_j was a candidate to enter the basis but failed to make it. Consider first case (i). In this case, we have $d_j \le \min(d_{\mathcal{B}})$ since the d_k's are arranged in nonincreasing order at step 1 of the above scheme. Thus, we have

$$d_j \, \|D_{\mathcal{B}}^{-1} B^{-1} A_j\| \;\le\; \frac{d_j}{\min(d_{\mathcal{B}})} \|B^{-1} A_j\| \;\le\; \|B^{-1} A_j\|.$$

Consider now case (ii). Suppose that A_j was a candidate to become the rth column of B. Since it failed to enter B, it must be linearly dependent on the first $r-1$ columns of B. Hence,

$$B^{-1} A_j \;=\; \begin{pmatrix} u \\ 0 \end{pmatrix}$$

for some $u \in \mathbb{R}^{r-1}$. Hence, using the fact that $d_{B_i} \ge d_j$ for every $i = 1, \ldots, r-1$, we conclude that

$$d_j \, \|D_{\mathcal{B}}^{-1} B^{-1} A_j\| \;\le\; d_j \left(\sum_{i=1}^{r-1} \frac{u_i^2}{d_{B_i}^2} \right)^{1/2} \;\le\; \left(\sum_{i=1}^{r-1} u_i^2 \right)^{1/2} \;=\; \|u\| \;=\; \|B^{-1} A_j\|.$$

We next prove (2). The first and third inequalities of (2) are well known. The second inequality follows from (1), the identity $\|R\|_F^2 = \sum_{j=1}^n \|R_j\|^2$ for every $R \in \mathbb{R}^{m\times n}$, and the fact that the jth column of $D_{\mathcal{B}}^{-1} B^{-1} AD$ is $d_j D_{\mathcal{B}}^{-1} B^{-1} A_j$. ∎

Given a pair $(A, d) \in \mathbb{R}^{m\times n} \times \mathbb{R}^n_{++}$ such that $\operatorname{rank}(A) = m$, we next consider a preconditioner for a system of equations of the form $AD^2 A^T p = r$, where $D \equiv \operatorname{Diag}(d)$. Pre- and postmultiplying its coefficient matrix by an invertible matrix $R \in \mathbb{R}^{m\times m}$, we obtain the equivalent system

$$R(AD^2 A^T) R^T \tilde{p} \;=\; Rr,$$

where $\tilde{p} = R^{-T} p$. The following results give a suitable choice of R for which the condition number of the coefficient matrix of the above system is uniformly bounded as d varies over $\mathbb{R}^n_{++}$.

LEMMA 2.2. *Let* $(A, d) \in \mathbb{R}^{m\times n} \times \mathbb{R}^n_{++}$ *be given such that* $\operatorname{rank}(A) = m$. *Suppose that* B *is a maximum weight basis associated with the pair* (A, d) *and define* $D \equiv \operatorname{Diag}(d)$, $D_{\mathcal{B}} \equiv \operatorname{Diag}(d_{\mathcal{B}})$, *and* $R \equiv D_{\mathcal{B}}^{-1}B^{-1}$. *Then*

$$\operatorname{cond}(RAD^2A^TR^T) \;\le\; \|B^{-1}A\|_F^2 \;\le\; m\,\|B^{-1}A\|^2.$$

Proof. By Lemma 2.1, we have $\|RAD\| \le \|B^{-1}A\|_F$. Hence,

$$\lambda_{\max}(RAD^2A^TR^T) \;=\; \|RAD\|^2 \;\le\; \|B^{-1}A\|_F^2 \;\le\; m\,\|B^{-1}A\|^2.$$

Moreover, since

$$R(ADA^T)R^T \;=\; D_{\mathcal{B}}^{-1}B^{-1}(BD_{\mathcal{B}}^2B^T + ND_{\mathcal{N}}^2N^T)B^{-T}D_{\mathcal{B}}^{-1} \;=\; I + WW^T,$$

where $W \equiv D_{\mathcal{B}}^{-1}B^{-1}ND_{\mathcal{N}}$ and $D_{\mathcal{N}} \equiv \operatorname{Diag}(d_{\mathcal{N}})$, we conclude that

$$\lambda_{\min}(RAD^2A^TR^T) \;\ge\; 1.$$

Hence, the lemma follows. ∎

We will refer to the preconditioner R described in Lemma 2.2 as a *maximum weight basis* preconditioner. Finally, as shown by Todd, Tunçel, and Ye [10] and Vavasis and Ye [11], we have

$$\bar{\chi}_A \;=\; \max\{\,\|B^{-1}A\| : B \text{ is a basis of } A\,\}.$$

Thus we arrive at our main result.

THEOREM 2.3. *Let* $(A, d) \in \mathbb{R}^{m\times n} \times \mathbb{R}^n_{++}$ *be given such that* $\operatorname{rank}(A) = m$. *Suppose that* B *is a maximum weight basis associated with the pair* (A, d) *and define* $D \equiv \operatorname{Diag}(d)$, $D_{\mathcal{B}} \equiv \operatorname{Diag}(d_{\mathcal{B}})$, *and* $R \equiv D_{\mathcal{B}}^{-1}B^{-1}$. *Then*

$$\operatorname{cond}(RAD^2A^TR^T) \;\le\; m\,\bar{\chi}_A^2.$$

Another important consequence of Lemma 2.2 is the following result.

THEOREM 2.4. *Let* $A \in \mathbb{R}^{m\times n}$ *denote the node–arc incidence matrix of a connected directed graph with one of its rows deleted. Suppose that* B *is a maximum weight basis associated with the pair* (A, d) *for some* $d \in \mathbb{R}^n_{++}$. *Letting* $D \equiv \operatorname{Diag}(d)$, $D_{\mathcal{B}} \equiv \operatorname{Diag}(d_{\mathcal{B}})$, *and* $R \equiv D_{\mathcal{B}}^{-1}B^{-1}$, *we have*

$$\operatorname{cond}(RAD^2A^TR^T) \;\le\; m(n-m+1).$$

Proof. Using the structure of A, it is easy to see that $\|B^{-1}A\|_F^2 \le m+(n-m)m = m(n-m+1)$. The result now follows directly from Lemma 2.2. ∎

3. Concluding remarks. As mentioned earlier, using a maximum weight basis preconditioner in the context of network flow problems yields a maximum spanning tree preconditioner first proposed by Resende and Veiga [7]. In such a case, Judice et al. [2] have attempted to show that the condition number of the preconditioned matrix is bounded. However, their proof is incomplete, in that it deals only with three out of four possible cases, the neglected case being the most difficult and interesting one. Our proof does not attempt to correct theirs; rather, it is based on an entirely different approach. Moreover, our approach also holds for general LP problems in standard form, and shows that the derived bounds on the condition number of the preconditioned normal matrices $R(AD^2A^T)R^T$ hold for any $D \in \mathcal{D}_{++}$. In the context

of interior-point methods, this means that the condition number of the preconditioned normal matrix has a bound that does not depend on the current iterate, regardless of whether it is well centered.

Certain computational issues arise from our analysis. When determining the maximum weight basis B, one must determine whether a set of columns in A is linearly independent. This process tends to be sensitive to roundoff errors. Once the set $\mathcal{B}$ is determined, the matrix R can be computed in a stable fashion, since the condition numbers for all possible bases B are uniformly bounded and multiplication by the diagonal matrix $D_{\mathcal{B}}^{-1}$ can be done in a componentwise manner.

The authors believe that the preconditioner studied in this paper will be computationally effective only for some special types of LP problems. For example, the papers [2, 7] developed effective iterative interior-point methods for solving the minimum cost network flow problem based on maximum spanning tree preconditioners. Another class of LP problems for which iterative interior-point methods based on maximum weight basis preconditioners might be useful are those for which bases of A are sparse but the normal matrices AD^2A^T are dense; this situation generally arises in the context of LP problems for which n is much larger than m. Investigation of the many classes of LP problems which would benefit from the application of such methods is certainly an important area for future research.

Acknowledgments. The authors would like to thank the two anonymous referees and the associate editor, Margaret Wright, for their insightful comments, which have helped to substantially improve the presentation of this paper.

REFERENCES

[1] I. I. Dikin and V. I. Zorkalcev, *Iterative Solution of Mathematical Programming Problems: Algorithms for the Method of Interior Points*, Nauka, Novosibirsk, USSR, 1980 (in Russian).

[2] J. J. Judice, J. Patricio, L. F. Portugal, M. G. C. Resende, and G. Veiga, *A study of preconditioners for network interior point methods*, Comput. Optim. Appl., 24 (2003), pp. 5–35.

[3] V. V. Kovacevic-Vujcic and M. D. Asic, *Stabilization of interior-point methods for linear programming*, Comput. Optim. Appl., 14 (1999), pp. 331–346.

[4] R. D. C. Monteiro and T. Tsuchiya, *A variant of the Vavasis–Ye layered-step interior-point algorithm for linear programming*, SIAM J. Optim., 13 (2003), pp. 1054–1079.

[5] R. D. C. Monteiro and T. Tsuchiya, *A new iteration-complexity bound for the MTY predictor-corrector algorithm*, SIAM J. Optim., to appear.

[6] A. R. L. Oliveira and D. C. Sorensen, *Computational Experience with a Preconditioner for Interior Point Methods for Linear Programming*, Technical report, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1997.

[7] M. G. C. Resende and G. Veiga, *An implementation of the dual affine scaling algorithm for minimum-cost flow on bipartite uncapacitated networks*, SIAM J. Optim., 3 (1993), pp. 516–537.

[8] G. W. Stewart, *On scaled projections and pseudoinverses*, Linear Algebra Appl., 112 (1989), pp. 189–193.

[9] M. J. Todd, *A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm*, Oper. Res., 38 (1990), pp. 1006–1018.

[10] M. J. Todd, L. Tunçel, and Y. Ye, *Probabilistic analysis of two complexity measures for linear programming problems*, Math. Program., 90 (2001), pp. 59–69.

[11] 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.

[12] M. H. Wright, *Ill-conditioning and computational error in interior methods for nonlinear programming*, SIAM J. Optim., 9 (1998), pp. 84–111.

[13] S. J. Wright, *Effects of finite-precision arithmetic on interior-point methods for nonlinear programming*, SIAM J. Optim., 12 (2001), pp. 36–78.