

MINIMIZING THE EUCLIDEAN CONDITION NUMBER*

RICHARD D. BRAATZ[†] AND MANFRED MORARI[‡]

Abstract. This paper considers the problem of determining the row and/or column scaling of a matrix A that minimizes the condition number of the scaled matrix. This problem has been studied by many authors. For the cases of the ∞ -norm and the 1-norm, the scaling problem was completely solved in the 1960s. It is the Euclidean norm case that has widespread application in robust control analyses. For example, it is used for integral controllability tests based on steady-state information, for the selection of sensors and actuators based on dynamic information, and for studying the sensitivity of stability to uncertainty in control systems.

Minimizing the scaled Euclidean condition number has been an open question—researchers proposed approaches to solving the problem numerically, but none of the proposed numerical approaches guaranteed convergence to the true minimum. This paper provides a convex optimization procedure to determine the scalings that minimize the Euclidean condition number. This optimization can be solved in polynomial-time with off-the-shelf software.

Key words. scaling, conditioning, condition number

AMS subject classifications. 65F35, 93B35, 93D21

1. Introduction. Let $V_1 = \mathbf{C}^n$ be the normed complex vector space with Hölder p -norm $\|\cdot\|_p$, $\|x\|_p = (\sum |x_j|^p)^{1/p}$. For an $n \times n$ matrix $A : V_1 \rightarrow V_1$, the following induced matrix norm is defined:

$$(1) \quad \|A\|_{ip} = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}.$$

If the inverse A^{-1} exists, then the condition number subordinate to the norm $\|\cdot\|_p$ is defined by

$$(2) \quad \kappa_p(A) = \|A\|_{ip} \|A^{-1}\|_{ip}.$$

Define $\mathbf{C}^{n \times n}$ to be the set of complex $n \times n$ matrices. Let $\mathbf{D}^{n \times n}$ be the set of all diagonal invertible matrices in $\mathbf{C}^{n \times n}$. If $A \in \mathbf{C}^{n \times n}$ is the matrix defining a system of linear equations $Ax = b$, scaling the rows of this system is equivalent to premultiplying A by a diagonal matrix $D_1 \in \mathbf{D}^{n \times n}$. Scaling the unknowns is equivalent to postmultiplying A by a diagonal matrix $D_2 \in \mathbf{D}^{n \times n}$. The quality of numerical computations is generally better when the condition number of A is small. Since diagonal scalings of A are trivial modifications, researchers in the 1960s–1970s were led to investigate the following minimizations in order to obtain optimal scalings of a matrix:

$$(3) \quad \begin{aligned} \text{(i)} \quad & \kappa_p^l(A) = \inf_{D_1 \in \mathbf{D}^{n \times n}} \kappa_p(D_1 A), \\ \text{(ii)} \quad & \kappa_p^r(A) = \inf_{D_2 \in \mathbf{D}^{n \times n}} \kappa_p(A D_2), \\ \text{(iii)} \quad & \kappa_p^{lr}(A) = \inf_{D_1, D_2 \in \mathbf{D}^{n \times n}} \kappa_p(D_1 A D_2). \end{aligned}$$

* Received by the editors October 13, 1992; accepted for publication (in revised form) June 22, 1993.

[†] Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125 (rdb@mozart.caltech.edu). This author was supported by the Fannie and John Hertz Foundation.

[‡] Department of Chemical Engineering, California Institute of Technology, Pasadena, California 91125 (mm@imc.caltech.edu).

TABLE 1

Minimized condition numbers. The matrix whose elements are the moduli of the corresponding elements of A is denoted by $|A|$. The spectral radius of A is denoted by $\rho(A)$. The maximum singular value $\bar{\sigma}(A)$ refers to $\|A\|_{i2}$.

	$p = 1, \infty$	$p = 2$
$\inf_{D_1} \kappa_p(D_1 A)$	$\bar{\sigma}(A^{-1} \cdot A)$?
$\inf_{D_2} \kappa_p(A D_2)$	$\bar{\sigma}(A \cdot A^{-1})$?
$\inf_{D_1, D_2} \kappa_p(D_1 A D_2)$	$\rho(A \cdot A^{-1})$?

Problem (3(i)) was present for example in the error analysis of direct methods for the solution of linear equations [34], [2]. Problem (3(ii)) is important for obtaining the best possible bounds for eigenvalue inclusion theorems [3], and is a natural measure of the linear independence of the column vectors that form A [2]. Problem (3(iii)) was used for decreasing the error in calculation of the matrix inverse A^{-1} [14].

Later, it was realized that the appropriate scalings depend on the error in the matrix, not the elements of the matrix itself [10], [31]. This implied, for example, that the scalings solving problem (3(iii)) are not necessarily the best scalings of A to decrease the error in the calculation of A^{-1} . However, problems (3(i))–(3(iii)) still have widespread application in robust control analyses. For example, the minimized condition number (3(iii)) is used for integral controllability tests based on steady-state information [13], [18], and for the selection of sensors and actuators using dynamic information [24], [19], [20]. The sensitivity of stability to uncertainty in control systems is given in terms of the minimized condition number in [29], [30].

Without loss of generality, for each of these problems we need only consider the infimum over the set of real positive diagonal invertible matrices $\mathbf{D}_+^{n \times n}$. This is because any matrix in $\mathbf{D}^{n \times n}$ can be decomposed into a matrix in $\mathbf{D}_+^{n \times n}$ and a unitary diagonal matrix. The unitary diagonal matrix does not affect the value of the condition number in (2) (see [2] for a simple proof). Conditions for the existence of scaling matrices that achieve the infimum are given by Businger [6].

The minimizations were solved for $p = 1$ and $p = \infty$ by Bauer [2] (the results are in Table 1). Many researchers consider the 2-norm as most important for applications [2], [14], [17]. Solving (3(i))–(3(iii)) for the 2-norm has been an open question [28], [35]. In this paper we solve the minimizations for the 2-norm by transforming the minimizations (3(i))–(3(iii)) so that they can be solved via convex programming.

Nonsquare A [33], block diagonal scalings [12], [27], [9], [11], [35], and cross-condition numbers (with B replacing A^{-1} in (2), see [8], [16], [15]) have also received attention. For ease of notation, the results are derived for square matrices with fully diagonal scalings. The results (and proofs) hold for these other cases with the modifications given after the lemmas.

2. Results. The induced matrix norm for the vector 2-norm is commonly referred to as the maximum singular value, $\bar{\sigma}(A) = \|A\|_{i2}$. To simplify notation, drop the subscript on κ_2 , i.e., $\kappa_2 = \kappa$. Let \mathbf{R}_+ be the set of real positive scalars. Let I be the $n \times n$ identity matrix.

LEMMA 2.1. *The following equality holds:*

$$(4) \quad \kappa(A) = \inf_{d_1, d_2 \in \mathbf{R}_+} \bar{\sigma}^2 \left(\begin{bmatrix} d_1 I & 0 \\ 0 & d_2 I \end{bmatrix} \begin{bmatrix} 0 & A^{-1} \\ A & 0 \end{bmatrix} \begin{bmatrix} (d_1)^{-1} I & 0 \\ 0 & (d_2)^{-1} I \end{bmatrix} \right).$$

Proof.

$$\begin{aligned}
 (5) \quad & \inf_{d_1, d_2 \in \mathbf{R}_+} \bar{\sigma}^2 \left(\begin{bmatrix} d_1 I & 0 \\ 0 & d_2 I \end{bmatrix} \begin{bmatrix} 0 & A^{-1} \\ A & 0 \end{bmatrix} \begin{bmatrix} (d_1)^{-1} I & 0 \\ 0 & (d_2)^{-1} I \end{bmatrix} \right) \\
 (6) \quad & = \inf_{d_1, d_2 \in \mathbf{R}_+} \bar{\sigma}^2 \left(\begin{bmatrix} 0 & \frac{d_1}{d_2} A^{-1} \\ \frac{d_2}{d_1} A & 0 \end{bmatrix} \right) \\
 (7) \quad & = \inf_{d_1, d_2 \in \mathbf{R}_+} \max \left\{ \bar{\sigma}^2 \left(\frac{d_1}{d_2} A^{-1} \right), \bar{\sigma}^2 \left(\frac{d_2}{d_1} A \right) \right\} \\
 (8) \quad & = \inf_{d_1, d_2 \in \mathbf{R}_+} \max \left\{ \frac{d_1^2 \bar{\sigma}(A^{-1})}{d_2^2 \bar{\sigma}(A)} \frac{d_2^2 \bar{\sigma}(A)}{d_1^2 \bar{\sigma}(A^{-1})} \right\} \cdot \bar{\sigma}(A) \bar{\sigma}(A^{-1}) \\
 (9) \quad & = \inf_{x \in \mathbf{R}_+} \max \{x, x^{-1}\} \cdot \bar{\sigma}(A) \bar{\sigma}(A^{-1}) \\
 (10) \quad & = \bar{\sigma}(A) \bar{\sigma}(A^{-1}) = \kappa(A).
 \end{aligned}$$

Note that this proof is similar to a proof in [21]. \square

The following lemma gives similar expressions as in (4) for $\kappa^l(A)$, $\kappa^r(A)$, and $\kappa^{lr}(A)$.

LEMMA 2.2. *The following equalities hold:*

$$(11) \quad \kappa^l(A) = \inf_{D_1 \in \mathbf{D}_+^{n \times n}} \bar{\sigma}^2 \left(\begin{bmatrix} I & 0 \\ 0 & D_1 \end{bmatrix} \begin{bmatrix} 0 & A^{-1} \\ A & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & D_1^{-1} \end{bmatrix} \right).$$

$$(12) \quad \kappa^r(A) = \inf_{D_2 \in \mathbf{D}_+^{n \times n}} \bar{\sigma}^2 \left(\begin{bmatrix} D_2^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 & A^{-1} \\ A & 0 \end{bmatrix} \begin{bmatrix} D_2 & 0 \\ 0 & I \end{bmatrix} \right).$$

$$(13) \quad \kappa^{lr}(A) = \inf_{D \in \mathbf{D}_+^{2n \times 2n}} \bar{\sigma}^2 \left(D \begin{bmatrix} 0 & A^{-1} \\ A & 0 \end{bmatrix} D^{-1} \right).$$

Proof. Substituting $D_1 A D_2$ for A in Lemma 2.1 and rearranging gives

$$(14) \quad \kappa(D_1 A D_2) = \inf_{d_1, d_2 \in \mathbf{R}_+} \bar{\sigma}^2 \left(\begin{bmatrix} d_1 D_2^{-1} & 0 \\ 0 & d_2 D_1 \end{bmatrix} \begin{bmatrix} 0 & A^{-1} \\ A & 0 \end{bmatrix} \begin{bmatrix} (d_1 D_2^{-1})^{-1} & 0 \\ 0 & (d_2 D_1)^{-1} \end{bmatrix} \right),$$

where d_1 and d_2 are real positive scalars.

Take the infimum over D_1 and D_2 on both sides to give

$$(15) \quad \kappa^{lr}(A) = \inf_{D_1, D_2 \in \mathbf{D}_+^{n \times n}} \inf_{d_1, d_2 \in \mathbf{R}_+} \bar{\sigma}^2 \left(\begin{bmatrix} d_1 D_2^{-1} & 0 \\ 0 & d_2 D_1 \end{bmatrix} \begin{bmatrix} 0 & A^{-1} \\ A & 0 \end{bmatrix} \begin{bmatrix} (d_1 D_2^{-1})^{-1} & 0 \\ 0 & (d_2 D_1)^{-1} \end{bmatrix} \right)$$

$$(16) \quad = \inf_{D_1, D_2 \in \mathbf{D}_+^{n \times n}} \bar{\sigma}^2 \left(\begin{bmatrix} D_2^{-1} & 0 \\ 0 & D_1 \end{bmatrix} \begin{bmatrix} 0 & A^{-1} \\ A & 0 \end{bmatrix} \begin{bmatrix} D_2 & 0 \\ 0 & D_1^{-1} \end{bmatrix} \right).$$

Letting $D = \text{diag} \{D_2^{-1}, D_1\}$ gives (13). Expressions (11) and (12) are proved similarly. \square

Let I_r be the $r \times r$ identity matrix. Let $\mathcal{D}^{2n \times 2n} = \text{diag} \{[d_1 I_{r_1}, \dots, d_m I_{r_m}] : d_j \in \mathbf{R}, r_1 + \dots + r_m = 2n\}$, and $M \in \mathcal{C}^{2n \times 2n}$. Consider the following lemma.

LEMMA 2.3. *The following optimization is convex:*

$$(17) \quad \inf_{D \in \mathcal{D}^{2n \times 2n}} \bar{\sigma}^2(e^D M e^{-D}).$$

Proof. See [25]. \square

Because $\{e^D : D \in \mathcal{D}\} = \{D : D \in \mathcal{D}_+\}$, the optimizations in Lemmas 2.1 and 2.2 are equivalent to the optimization in Lemma 2.3. This means that the condition number κ and minimized condition numbers $\kappa^l, \kappa^r, \kappa^{lr}$ can all be calculated through convex programming. Since the optimization (17) is convex, it can have only one minimum.

The optimization (17) has been studied extensively [22], [32], [23], [25], and off-the-shelf software is available for solving these polynomial-time problems (for example, see the program *mu* in [1]). The calculation of the minimized condition numbers is slow, however, since the minimization (17) requires repeated maximum singular value calculations.

The parallelism between expressions (4), (11), (12), and (13) for $\kappa, \kappa^l, \kappa^r$, and κ^{lr} is interesting. The same optimization can be used for the condition number calculations—the optimizations are just over different “scaling matrices.” This is nice theoretically, since κ^l, κ^r , and κ^{lr} are just the *scaled* condition numbers.

Remark 2.4. Conditions for the existence of scaling matrices that achieve the infimum are given by Businger [6]. When the infimum is achieved, any algorithm that solves (17) provides the minimizing scaling matrices for the condition number. When the infimum is not achieved, the algorithm provides scaling matrices such that the infimum is approached with arbitrary closeness.

Remark 2.5. To generalize to nonsquare A , replace every occurrence of A^{-1} with the respective right or left inverse. More specifically, if $A \in \mathbf{C}^{m \times n}$ and has full row rank with $m < n$, then replace A^{-1} with $A^T(AA^T)^{-1}$ in all proofs and lemmas. For $m > n$ with A having full column rank, replace A^{-1} with $(A^T A)^{-1}A^T$.

Remark 2.6. The Euclidean cross-condition number is defined by

$$(18) \quad \hat{\kappa}(A, B) := \bar{\sigma}(A) \bar{\sigma}(B).$$

Minimized cross-condition numbers can be defined similarly as in (3), for example,

$$(19) \quad \hat{\kappa}^{lr}(A, B) := \inf_{D_1, D_2 \in \mathcal{D}^{n \times n}} \hat{\kappa}(D_1 A D_2, D_2^{-1} B D_1^{-1}).$$

Lemmas 2.1 and 2.2 follow with B , replacing A^{-1} . This problem is important for testing stability of systems with element-by-element uncertainty [7], [8], [16], [15].

Remark 2.7. For block-diagonal scaling matrices, without loss of generality we can take each block to be positive definite Hermitian. This is because any nonsingular complex matrix can be decomposed into a positive definite Hermitian matrix and a unitary matrix [4], and the unitary matrix does not affect the value of the Euclidean condition number. The proofs of Lemmas 2.1 and 2.2 follow exactly as for the fully diagonal case. Lemma 2.3 does not hold for block-diagonal scalings. For block-diagonal scalings it is better to convert the singular value minimizations in Lemma 2.2

into generalized eigenvalue minimizations, as follows:

$$(20) \quad \inf_{D \in \mathbf{D}_+} \bar{\sigma}(DMD^{-1}) = \inf_{D^2 \in \mathbf{D}_+} \{\beta \mid M^* D^2 M - \beta D^2 < 0\}.$$

The condition $M^* D^2 M - \beta D^2 < 0$ is convex in D^2 , so any local minimum is global, and off-the-shelf software is available [1]. Many researchers are working to develop improved computational approaches for these polynomial-time problems (for example, see [5] and the literature cited therein).

3. Conclusions. We have completed Table 1 in the sense that all values in the table can now be calculated with arbitrary precision.

All entries in the table, including the now-filled entries, require the inverse of A to calculate the minimizing scalings and the minimized condition numbers. There are algorithms for numerically determining the minimized condition numbers without predetermining the matrix inverse [26], [35], but these methods are not guaranteed to converge to the true minima.

REFERENCES

- [1] G. J. BALAS, A. K. PACKARD, J. C. DOYLE, K. GLOVER, AND R. S. R. SMITH, *Development of advanced control design software for researchers and engineers*, in Proc. 1991 American Control Conference, 1991, pp. 996–1001.
- [2] F. L. BAUER, *Optimally scaled matrices*, Numer. Math., 5 (1963), pp. 73–87.
- [3] F. L. BAUER AND A. S. HOUSEHOLDER, *Absolute norms and characteristic roots*, Numer. Math., 3 (1961), pp. 241–246.
- [4] R. BELLMAN, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1960.
- [5] S. BOYD AND L. E. GHAOUI, *Method of centers for minimizing generalized eigenvalues*, Linear Algebra Appl., 188 (1993), pp. 63–111.
- [6] P. A. BUSINGER, *Matrices which can be optimally scaled*, Numer. Math., 12 (1968), pp. 346–348.
- [7] J. CHEN, M. K. H. FAN, AND C. N. NETT, *Robustness analysis with non-diagonally structured uncertainty*, IEEE Trans. Automat. Control, (1994), preprint.
- [8] R. W. DANIEL, B. KOUVARITAKIS, AND H. LATCHMAN, *Principal direction alignment: A geometric framework for the complete solution to the μ -problem*, IEE Proceedings Part D, 133 (1986), pp. 45–56.
- [9] J. W. DEMMEL, *The condition number of equivalence transformations that block diagonalize matrix pencils*, SIAM J. Numer. Anal., 20 (1983), pp. 599–610.
- [10] J. J. DONGARRA, C. B. MOLER, J. R. BUNCH, AND G. W. STEWART, *LINPACK User's Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1979.
- [11] P. V. DOOREN AND P. DEWILDE, *Minimal cascade factorization of real complex rational transfer matrices*, IEEE Trans. Circuits and Systems, 28 (1981), pp. 390–400.
- [12] S. C. EISENSTAT, J. W. LEWIS, AND M. H. SCHULTZ, *Optimal block diagonal scaling of block 2-cyclic matrices*, Linear Algebra Appl., 44 (1982), pp. 181–186.
- [13] P. GROSDIDIER, M. MORARI, AND B. R. HOLT, *Closed-loop properties from steady-state gain information*, Ind. Eng. Chem. Fund., 24 (1985), pp. 221–235.
- [14] A. S. HOUSEHOLDER, *The Theory of Matrices in Numerical Analysis*, Blaisdell, New York, 1964, p. 123.
- [15] B. KOUVARITAKIS AND H. LATCHMAN, *Necessary and sufficient stability condition for systems with structured uncertainties: The major principal direction alignment principle*, Internat. J. Control, 42 (1985), pp. 575–598.
- [16] ———, *Singular-value and eigenvalue techniques in the analysis of systems with structured perturbations*, Internat. J. Control, 41 (1985), pp. 1381–1412.
- [17] C. MCCARTHY AND G. STRANG, *Optimal conditioning of matrices*, SIAM J. Numer. Anal., 10 (1973), pp. 370–388.
- [18] M. MORARI AND E. ZAFIRIOU, *Robust Process Control*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [19] C. N. NETT, *Decentralized control system design for a variable-cycle gas turbine engine*, in Proc. 1990 American Control Conference, 1990.

- [20] C. N. NETT AND K. D. MINTO, *A quantitative approach to the selection and partitioning of measurements and manipulations for the control of complex systems*, in Proc. 1989 American Control Conference, 1989.
- [21] C. N. NETT AND J. A. UTHGENANNT, *An explicit formula and an optimal weight for the 2-block structured singular value interaction measure*, in Proc. 1987 American Control Conference, 1987, pp. 506–511.
- [22] E. E. OSBORNE, *On pre-conditioning of matrices*, J. Assoc. Comp. Mach., 7 (1960), pp. 338–345.
- [23] A. PACKARD AND J. C. DOYLE, *The complex structured singular value*, Automatica, 29 (1993), pp. 71–109.
- [24] D. E. REEVES, C. N. NETT, AND Y. ARKUN, *Control configuration design for complex systems: A practical theory*, Tech. report, Georgia Inst. Tech., Athens, GA, 1992.
- [25] R. SEZGINER AND M. OVERTON, *The largest singular value of $e^X A e^{-X}$ is convex on convex sets of commuting matrices*, IEEE Trans. Automat. Control, 35 (1990), pp. 229–230.
- [26] A. SHAPIRO, *Weighted minimum trace factor analysis*, Psychometrika, 47 (1982), pp. 243–264.
- [27] ———, *Optimal block diagonal l_2 scaling of matrices*, SIAM J. Numer. Anal., 22 (1985), pp. 81–94.
- [28] ———, *Upper bounds for nearly optimal diagonal scaling of matrices*, Linear and Multilinear Algebra, 29 (1991), pp. 145–147.
- [29] S. SKOGESTAD AND M. MORARI, *Design of resilient processing plants—IX. Effect of model uncertainty on dynamic resilience*, Chem. Eng. Sci., 42 (1987), pp. 1765–1780.
- [30] ———, *Implications of large RGA elements on control performance*, Ind. Eng. Chem. Res., 26 (1987), pp. 2323–2330.
- [31] G. W. STEWART AND J. G. SUN, *Matrix Perturbation Theory*, Academic Press, Boston, MA, 1990.
- [32] J. STOER AND C. WITZGALL, *Transformations by diagonal matrices in a normed space*, Numer. Math., 4 (1962), pp. 158–171.
- [33] A. VAN DER SLUIS, *Condition numbers and equilibration of matrices*, Numer. Math., 14 (1969), pp. 14–23.
- [34] ———, *Condition, equilibration and pivoting in linear algebraic systems*, Numer. Math., 15 (1970), pp. 74–86.
- [35] G. A. WATSON, *An algorithm for optimal l_2 scaling of matrices*, IMA J. Numer. Anal., 11 (1991), pp. 481–492.