SIAM J. MATRIX ANAL. APPL.
Vol. 16, No. 1, pp. 29–39, January 1995

003

EXISTENCE AND UNIQUENESS OF OPTIMAL MATRIX SCALINGS*

V. BALAKRISHNAN† AND S. BOYD‡

Abstract. The problem of finding a diagonal similarity scaling to minimize the scaled singular value of a matrix arises frequently in robustness analysis of control systems. It is shown here that the set of optimal diagonal scalings is nonempty and bounded if and only if the matrix that is being scaled is irreducible. For an irreducible matrix, a sufficient condition is derived for the uniqueness of the optimal scaling.

Key words. diagonal similarity scalings, scaled singular value minimization, irreducible matrices

AMS subject classifications. 65F35, 15A60, 15A12, 47A55

Notation. $\mathbf{R}$ ($\mathbf{C}$) denotes the set of real (complex) numbers. $\mathbf{R}_+$ stands for the set of positive real numbers. For $z \in \mathbf{C}$, $\operatorname{Re} z$ is the real part of z. The set of $m \times n$ matrices with real (complex) entries is denoted $\mathbf{R}^{m\times n}$ ($\mathbf{C}^{m\times n}$). I stands for the identity matrix with size determined from context. For a matrix $P \in \mathbf{C}^{m\times n}$, P^T stands for the transpose and P^* stands for the complex conjugate of P^T. $\|P\|$ is the spectral norm (maximum singular value) of P given by the square root of the maximum eigenvalue of P^*P. (For a vector $v \in \mathbf{C}^n$, $\|v\|$ is just the Euclidean norm.) For $P \in \mathbf{C}^{n\times n}$, $\operatorname{Tr} P$ stands for the trace, that is, the sum of the diagonal entries of P.

1. Introduction.

Given a complex matrix $M \in \mathbf{C}^{n\times n}$ and a nonsingular diagonal matrix $D \in \mathbf{C}^{n\times n}$, the *similarity-scaled singular value* of M corresponding to scaling D is defined as

$$f(M,D) = \left\|DMD^{-1}\right\|.$$

The optimal diagonal scaling problem is to minimize $f(M,D)$ over all diagonal nonsingular matrices D:

$$f_{\min}(M) = \inf\left\{ \left\|DMD^{-1}\right\| \;\middle|\; D \in \mathbf{C}^{n\times n},\ D \text{ is diagonal and nonsingular}\right\}. \tag{1}$$

We refer to $f_{\min}(M)$ as the *optimally scaled singular value* of M.

Problem (1) arises in the robustness analysis of control systems with structured uncertainties. For further details, see [11] and [4]. Much research has focused on the related problem of finding optimal (with various criteria for optimality) diagonal preconditioners for use in iterative algorithms; see, for example, [5] and [7].

Reformulation as a convex optimization problem. We note that $f(M,|D|) = f(M,D)$; we also observe that $f(M,D)$ is homogeneous of degree zero in D, that is, $f(M,\alpha D) = f(M,D)$ for all nonzero $\alpha \in \mathbf{C}$. Therefore, we may rewrite (1) as

$$f_{\min}(M) = \inf\left\{ \left\|e^{D}Me^{-D}\right\| \;\middle|\; D \in \mathbf{R}^{n\times n},\ D \text{ is diagonal}, \operatorname{Tr} D = 0\right\}. \tag{2}$$

* Received by the editors August 10, 1992; accepted for publication (in revised form) by L. Kaufman, September 30, 1993. This research was supported in part by Air Force Office of Scientific Research under contract F49620-92-J-0013.

† School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907 (ragu@ecn.purdue.edu).

‡ Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, California 94305 (boyd@isl.stanford.edu).

The reason for rewriting (1) as (2) is that $\|e^D M e^{-D}\|$ is a *convex* function of D—this fact will prove important in the sequel—while $\|DMD^{-1}\|$ is not [12], [13]. For convenience, we let

$$\mathcal{D} = \{D \mid D \in \mathbf{R}^{n\times n},\ D \text{ is diagonal}, \operatorname{Tr} D = 0\}\ .$$

In this paper, we do not concern ourselves with the solution of (2). We instead investigate the set of *minimizers* for (2), that is, the set of *optimal* scalings $\mathcal{D}_{\text{opt}}$ defined by

$$\mathcal{D}_{\text{opt}} \triangleq \{D \mid D \in \mathcal{D}, \|e^D M e^{-D}\| = f_{\min}(M)\}\ . \tag{3}$$

In the process, we provide a sufficient condition for $\mathcal{D}_{\text{opt}}$ to be nonempty (which means the infimum in (2) is *achieved*) and a sufficient condition for $\mathcal{D}_{\text{opt}}$ to be a singleton (which means that there is a *unique* optimal scaling).

2. Boundedness of $\mathcal{D}_{\text{opt}}$. We start with a few definitions.

DEFINITION 1. A *permutation* matrix P is a real, orthogonal $n \times n$ matrix (i.e., $PP^T = P^T P = I$) with entries that are either one or zero. We let $\mathcal{P}$ denote the set of $n \times n$ permutation matrices.

DEFINITION 2. A complex matrix M is said to be *reducible* if there exists some $P \in \mathcal{P}$ such that PMP^T is block upper triangular, that is,

$$PMP^T = \begin{bmatrix} M_{11} & M_{12} \\ 0 & M_{22} \end{bmatrix},$$

where M_{11}, M_{22} are square matrices of appropriate sizes [6], [1]. A matrix that is not reducible is termed *irreducible*.

Remark. For any permutation matrix P,

$$\|e^D M e^{-D}\| = \|P e^D P^T P M P^T P e^{-D} P^T\|\,.$$

Note that $Pe^D P^T$ is diagonal and corresponds to just a reordering of the diagonal entries of e^D. Therefore, as far as the scaling problem is concerned, if a matrix M is reducible, we may assume without loss of generality that

$$M = \begin{bmatrix} M_{11} & M_{12} \\ 0 & M_{22} \end{bmatrix},$$

bearing in mind that a reordering of the entries of the scaling D might be necessary. In the sequel, the phrase "within a permutation" refers to such a reordering of the entries of D and the corresponding permutation similarity transformation on M.

Let $\mathcal{D}_\gamma$ denote the *sublevel* set

$$\{D \in \mathcal{D} \mid \|e^D M e^{-D}\| < \gamma\}\ .$$

The following theorem relates the irreducibility of M to the boundedness of the sublevel sets.

THEOREM 2.1. *For any $\gamma > f_{\min}(M)$, the sublevel set $\mathcal{D}_\gamma$ is bounded if and only if M is irreducible.*

Proof. We first note the following lemma.

LEMMA 2.2. *It holds that*

$$\left\|\begin{bmatrix} M_{11} & M_{12} \\ 0 & M_{22} \end{bmatrix}\right\| \geq \left\|\begin{bmatrix} M_{11} & 0 \\ 0 & M_{22} \end{bmatrix}\right\|,$$

where M_{11}, M_{12}, and M_{22} are matrices of appropriate sizes.

Proof. The proof is left to the reader.

We first assume that M is reducible. Then to within a permutation,

$$M = \begin{bmatrix} M_{11} & M_{12} \\ 0 & M_{22} \end{bmatrix},$$

where $M_{11} \in \mathbf{C}^{r \times r}$ with $r < n$.

Then given any $\gamma > f_{\min}(M)$ and $D \in \mathcal{D}_\gamma$ (note that $\mathcal{D}_\gamma$ is nonempty), partition D conformally with the block upper triangular structure of M above as

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}.$$

Consider now a sequence of scaling matrices $D^{(i)}$ of the form

$$D^{(i)} = \begin{bmatrix} D_1 - i\,(n-r) & 0 \\ 0 & D_2 + i\,r \end{bmatrix}, \qquad i = 1, 2, \ldots.$$

(Note that $D^{(i)} \in \mathcal{D}$ for $i = 1, 2, \ldots$.)

For such a sequence of scalings, $\left\| e^{D^{(i)}} M e^{-D^{(i)}} \right\|$ converges to

$$\max\left(\left\| e^{D_1} M_{11} e^{-D_1} \right\|, \left\| e^{D_2} M_{22} e^{-D_2} \right\| \right),$$

which is less than or equal to

$$\left\| e^{D^{(i)}} M e^{-D^{(i)}} \right\|$$

for every i, from Lemma 2.2. Thus for every $\gamma > f_{\min}(M)$, the set $\mathcal{D}_\gamma$ is unbounded.

To prove the converse, let us assume that for some $\gamma > f_{\min}(M)$, $\mathcal{D}_\gamma$ is not bounded. Then there is a sequence of scalings $D^{(i)}$ in $\mathcal{D}_\gamma$ with some of the elements of the diagonal scaling matrix $D^{(i)}$ with absolute value tending to infinity. Then, there exists a subsequence $D^{(n_i)}$, which can be partitioned to within a permutation as

$$D^{(n_i)} = \begin{bmatrix} D_{1,n_i} & 0 \\ 0 & D_{2,n_i} \end{bmatrix},$$

where every element of D_{1,n_i} diverges to $-\infty$ with i, while every element of D_{2,n_i} is bounded below. (In fact, at least one of the elements of D_{2,n_i} must diverge to ∞, but we will not use this fact.)

Thus the maximum singular value of

$$M = \begin{bmatrix} e^{D_{1,n_i}} M_{11} e^{-D_{1,n_i}} & e^{D_{1,n_i}} M_{12} e^{-D_{2,n_i}} \\ e^{D_{2,n_i}} M_{21} e^{-D_{1,n_i}} & e^{D_{2,n_i}} M_{22} e^{-D_{2,n_i}} \end{bmatrix},$$

remains bounded with every element of D_{1,n_i} diverging to $-\infty$ while the elements of D_{2,n_i} are bounded below. This immediately means that $M_{21} = 0$, which shows that M must be reducible. □

COROLLARY 2.3. *$\mathcal{D}_{\text{opt}}$ is nonempty and bounded if M is irreducible.*

Proof. If M is irreducible, the sublevel set $\mathcal{D}_\gamma$ is bounded for every $\gamma > f_{\min}(M)$; since $\left\| e^{D} M e^{-D} \right\|$ is a continuous function of D over $\mathcal{D}_\gamma$, the infimum in (2) is achieved.

Thus $\mathcal{D}_{\text{opt}}$ is nonempty if M is irreducible. Boundedness of $\mathcal{D}_{\text{opt}}$ follows from an argument similar to the one in the proof of Theorem 2.1. □

We note that this sufficient condition for the existence of optimal scalings can also be found in [3, Prop. 4].

Remark. Thus, irreducibility of M is a sufficient condition for the existence of optimal matrix scalings. If M is reducible, two cases are possible: $\mathcal{D}_{\text{opt}}$ may be empty or it may be nonempty and unbounded. The following examples illustrate this.

Example 1 ($\mathcal{D}_{\text{opt}}$ empty).

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is shown in the Appendix that $\mathcal{D}_{\text{opt}}$ is empty. The optimally scaled singular value is the limit of the sequence of scaled singular values corresponding to scalings $D(d)$ with $d \downarrow -\infty$:

$$D(d) = \begin{bmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & -2d \end{bmatrix}.$$

Example 2 ($\mathcal{D}_{\text{opt}}$ nonempty and unbounded).

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is shown in the Appendix that

$$\mathcal{D}_{\text{opt}} = \left\{ D \;\middle|\; D = \begin{bmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & -2d \end{bmatrix},\ d \in (-\infty, \log(3/2)/6] \right\}.$$

3. $\mathcal{D}_{\text{opt}}$ for irreducible matrices. We next derive a sufficient condition for $\mathcal{D}_{\text{opt}}$ to be a singleton.

We first state without proof a condition for optimality of a scaling D.

THEOREM 3.1. *Suppose the maximum singular value of $e^D M e^{-D}$ is* isolated, *i.e., of unit multiplicity. Then D is an optimal scaling for Problem* 2 *if and only if there exist vectors u and v, with $\|u\| = \|v\| = 1$, such that*

$$\begin{aligned} e^D M e^{-D}\, v &= f_{\min}(M)\, u \\ e^{-D} M^* e^D\, u &= f_{\min}(M)\, v, \end{aligned} \quad \text{and} \quad \left|u^{(i)}\right| = \left|v^{(i)}\right|,\ i = 1, 2, \ldots, n,$$

where $u^{(i)}$ and $v^{(i)}$, $i = 1, 2, \ldots, n$ are the components of u and v, respectively.

Theorem 3.1, which is a "magnitude-matching" condition on the components of the left and right singular vectors of the scaled matrix, follows immediately from simple gradient calculations (see, for example, [9]).

We also need the following theorem about the analyticity properties of the singular values of a complex matrix that depends on a real parameter (see [2], [10], [8]).

THEOREM 3.2. *Let $A(x)$ be a (complex) $m \times n$ matrix, the entries of which are analytic functions of a real parameter x. There are real analytic functions $f_i : \mathbf{R} \to \mathbf{R}, i = 1, \ldots, \min(m,n)$ such that, for all $x \in \mathbf{R}$,*

$$\{\sigma_i(A(x)), i = 1, \ldots, \min(m,n)\} = \{|f_i(x)|, i = 1, \ldots, \min(m,n)\}, \tag{4}$$

where $\sigma_i(A(x))$ stands for the ith singular value of $A(x)$. (Thus, the f_i's are the unordered and unsigned singular value functions of $A(x)$.)

For convenience, we let $\gamma = f_{\min}(M)$. With D being an optimal scaling, suppose that (i) γ is the isolated maximum singular value of $e^D M e^{-D}$ and (ii) the left and right singular vectors of $e^D M e^{-D}$ (i.e., u and v in Theorem 3.1) belong to the same *coordinate subspace*, i.e., a subspace of the form $\bigcup_{i \in \mathbf{I}} \operatorname{span}\{e_i\}$, where $\mathbf{I}$ is a proper subset of the set of indices $\{1, \ldots, n\}$ and $\{e_i,\ i = 1, \ldots, n\}$ are coordinate vectors (i.e., unit vectors of $\mathbf{R}^n$ in the standard basis). We will show that this means that $\mathcal{D}_{\text{opt}}$ is not a singleton.

First note that to within a permutation, we have

$$u = \begin{bmatrix} u_1 \\ 0 \end{bmatrix}, \qquad v = \begin{bmatrix} v_1 \\ 0 \end{bmatrix}, \tag{5}$$

where $u_1, v_1 \in \mathbf{C}^r$ with $1 \le r < n$; we then partition $e^D M e^{-D}$ as

$$e^D M e^{-D} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

where $M_{11} \in \mathbf{C}^{r \times r}$. Of course, $\left|u_1^{(i)}\right| = \left|v_1^{(i)}\right|$, $i = 1, \ldots, r$, and γ is the optimally scaled maximum singular value of M_{11}. Now, with

$$D(\lambda) = \begin{bmatrix} \lambda(n-r)I_1 & 0 \\ 0 & -\lambda r I_2 \end{bmatrix} + D,$$

where I_1 is the $r \times r$ identity matrix, consider

$$e^{D(\lambda)} M e^{-D(\lambda)} = \begin{bmatrix} M_{11} & e^{\lambda n} M_{12} \\ e^{-\lambda n} M_{21} & M_{22} \end{bmatrix}.$$

For every $\lambda \in \mathbf{R}$, γ is a singular value of $e^{D(\lambda)} M e^{-D(\lambda)}$, with u and v in (5) being the corresponding left and right singular vectors. Moreover, every entry of $e^{D(\lambda)} M e^{-D(\lambda)}$ is an analytic function of λ. Then, using Theorem 3.2 and the assumption that the maximum singular value of $e^D M e^{-D}$ is isolated, we conclude that the maximum singular value of $e^{D(\lambda)} M e^{-D(\lambda)}$ is isolated, and hence a real analytic function of λ for $\lambda \in [-\epsilon, \epsilon]$, where $\epsilon > 0$ is sufficiently small. It follows immediately that for $\lambda \in [-\epsilon, \epsilon]$, γ is the maximum singular value of $e^{D(\lambda)} M e^{-D(\lambda)}$. In other words, $D(\lambda)$ is also an optimal scaling for M, for $\lambda \in [-\epsilon, \epsilon]$.

Conversely, let us assume that $\mathcal{D}_{\text{opt}}$ is not a singleton, so that there exist $D_1, D_2 \in \mathcal{D}_{\text{opt}}$, with $D_1 \neq D_2$. Then, from the convexity of $\mathcal{D}_{\text{opt}}$, $D(\lambda) = \lambda D_1 + (1-\lambda) D_2 \in \mathcal{D}_{\text{opt}}$ for every $\lambda \in [0,1]$. Moreover, let us assume that γ is the isolated maximum singular value of $e^{D(\lambda)} M e^{-D(\lambda)}$ for $\lambda \in [0,1]$.

Since $D_1 \neq D_2$, to within a permutation,

$$D_1 - D_2 = \begin{bmatrix} d_1 I_1 & 0 & \cdots & 0 \\ 0 & d_2 I_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p I_p \end{bmatrix},$$

where $p > 1$, $d_1 > d_2 > \cdots > d_p$, and $I_1, I_2, \ldots, I_p$ are identity matrices of sizes $n_1, n_2, \ldots, n_p$ respectively. Of course, $\sum_{i=1}^{p} n_i = n$, and $\sum_{i=1}^{p} n_i d_i = 0$. Note that every entry of $e^{D(\lambda)} M e^{-D(\lambda)}$ is an analytic function of λ, more specifically equal to a ratio of polynomials of (the components of) $z = \left[e^{\lambda d_1}\ e^{\lambda d_2} \cdots e^{\lambda d_p}\right]$. Then, using Theorem 3.2, we conclude that since γ is the maximum singular value of $e^{D(\lambda)} M e^{-D(\lambda)}$ for $\lambda \in [0,1]$, it must be a singular value of $e^{D(\lambda)} M e^{-D(\lambda)}$ for *all* $\lambda \in \mathbf{R}$.

Next, let $u(\lambda)$ and $v(\lambda)$ be the left and right singular vectors of $e^{D(\lambda)} M e^{-D(\lambda)}$ corresponding to the singular value γ, so that

$$
(6) \qquad \begin{aligned} e^{D(\lambda)} M\, e^{-D(\lambda)}\, v(\lambda) &= \gamma\, u(\lambda), \\ e^{-D(\lambda)} M^*\, e^{D(\lambda)}\, u(\lambda) &= \gamma\, v(\lambda), \end{aligned}
$$

with $\|u(\lambda)\| = \|v(\lambda)\| = 1$. Then, by a direct calculation, $u(\lambda)$ and $v(\lambda)$ can be chosen as analytic functions of λ whose every entry can be expressed as a ratio of a polynomial of z and the square root of a polynomial of z. Therefore, the limits, as $\lambda \to \pm\infty$, of $u(\lambda)$ and $v(\lambda)$ exist. Next, from Theorem 3.1 we have $\left|u^{(i)}(\lambda)\right| = \left|v^{(i)}(\lambda)\right|$ for $i = 1, 2, \ldots, n$ and $\lambda \in [0,1]$, and therefore $\left|u^{(i)}(\lambda)\right| = \left|v^{(i)}(\lambda)\right|$ for $i = 1, 2, \ldots, n$ and for *all* $\lambda \in \mathbf{R}$.

Partitioning $e^{D_2} M e^{-D_2}$, $u(\lambda)$ and $v(\lambda)$ as

$$
e^{D_2} M e^{-D_2} = \begin{bmatrix} M_{11} & M_{12} & \ldots & M_{1p} \\ M_{21} & M_{22} & \ldots & M_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ M_{p1} & M_{p2} & \ldots & M_{pp} \end{bmatrix}, \quad u = \begin{bmatrix} u_1(\lambda) \\ u_2(\lambda) \\ \vdots \\ u_p(\lambda) \end{bmatrix}, \quad v = \begin{bmatrix} v_1(\lambda) \\ v_2(\lambda) \\ \vdots \\ v_p(\lambda) \end{bmatrix},
$$

where $M_{ii} \in \mathbf{C}^{n_i \times n_i}$, and $u_i(\lambda)$ and $v_i(\lambda) \in \mathbf{C}^{n_i}$ for $i = 1, 2, \ldots, p$, we now show that γ is the optimally scaled maximum singular value of M_{11} or M_{22} or $\ldots$ M_{pp}.

Consider the following equation, taken from (6).

$$
e^{-\lambda d_1} M_{11} v_1(\lambda) + e^{-\lambda d_2} M_{12} v_2(\lambda) + \cdots + e^{-\lambda d_p} M_{1p} v_p(\lambda) = \gamma e^{-\lambda d_1} u_1(\lambda).
$$

Letting $\lambda \to -\infty$ in the above equation, we get

$$
M_{11} v_1(-\infty) = \gamma u_1(-\infty).
$$

Since $v_1(-\infty)^* v_1(-\infty) = u_1(-\infty)^* u_1(-\infty)$ (this follows from $\left|u^{(i)}(\lambda)\right| = \left|v^{(i)}(\lambda)\right|$ for $i = 1, 2, \ldots, n$ and for $\lambda \in \mathbf{R}$), we conclude that either γ is the optimally scaled maximum singular value of M_{11} or $u_1(-\infty) = v_1(-\infty) = 0$. Continuing similarly, it follows that γ is the optimally scaled maximum singular value of M_{ii}, for some $i = 1, \ldots, p$. (Recall our assumption that γ is the isolated maximum singular value of $e^{D(\lambda)} M e^{-D(\lambda)}$ for $\lambda \in [0,1]$, so that only one of $M_{11}, \ldots, M_{pp}$ can have a maximum singular value of γ.)

Remark. Suppose $u_1(-\infty) \neq 0 \neq v_1(-\infty)$. Then, by replacing λ by $\lambda + \eta$ (where $\eta \in \mathbf{R}$ is fixed) in the preceding argument, we may show that $[u_1(-\infty)^*\ \ 0 \cdots 0]^*$ and $[v_1(-\infty)^*\ \ 0 \cdots 0]^*$ are left and right singular vectors of $e^{D(\eta)} M e^{-D(\eta)}$ corresponding to a singular value γ for *every* $\eta \in \mathbf{R}$, where $D(\eta) = \eta D_1 + (1 - \eta) D_2$.

Remark. If the entries of $D_1 - D_2$ are distinct, then there exist left and right singular vectors of $e^{D_2} M e^{-D_2}$ corresponding to the maximum singular value that both equal the same coordinate vector.[1]

---

[1] We thank Reviewer 1 for drawing our attention to this remark.

In summary, we have shown that there exist two different optimal scalings D_1 and D_2, with the optimally scaled maximum singular value being isolated for all $D(\lambda) = D_2 + \lambda(D_1 - D_2)$, $\lambda \in [0,1]$, if and only if there exist left and right singular vectors of $e^{D_2}Me^{-D_2}$ (indeed, of $e^{D(\lambda)}Me^{-D(\lambda)}$, $\lambda \in [0,1]$) corresponding to the isolated maximum singular value, belonging to the same coordinate subspace.

We thus arrive at the following sufficient condition for the optimal scaling to be unique.

THEOREM 3.3. *For an irreducible matrix M, let D be an optimal scaling, and let the maximum singular value of e^DMe^{-D} be isolated. Then D is the unique optimal scaling if and only if there exists no pair of vectors u and v, with $\|u\| = \|v\| = 1$ satisfying*

$$\begin{aligned} e^D M e^{-D}\, v &= \gamma\, u, \\ e^{-D} M^* e^D\, u &= \gamma\, v \end{aligned} \tag{7}$$

that belong to the same coordinate subspace.

Remark. With D being an optimal scaling, if the maximum singular value of e^DMe^{-D} is not isolated, then there always exist v and u with $\|u\| = \|v\| = 1$, satisfying (7) and belonging to the same coordinate subspace. In this case, the optimal scaling may or may not be unique as the following two examples illustrate.

Example 3 ($\mathcal{D}_{\text{opt}}$ is a singleton).

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix}.$$

It is shown in the Appendix that the unique optimal scaling is zero, i.e., the "identity" scaling, though $[1/\sqrt{2}\ \ 1/\sqrt{2}\ \ 0]^T$ is both a left and right singular vector corresponding to the maximum singular value of two. Note that the maximum singular value at the optimal scaling is not isolated.

Example 4 ($\mathcal{D}_{\text{opt}}$ is not a singleton).

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & \dfrac{1}{\sqrt{2}} \end{bmatrix}.$$

It is shown in the Appendix that $\mathcal{D}_{\text{opt}}$ is given by

$$\mathcal{D}_{\text{opt}} = \left\{ D \;\middle|\; D = \begin{bmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & -2d \end{bmatrix},\ d \in [-d_*\,, d_*] \right\},$$

where

$$d_* = \left(\frac{1}{6}\right) \log \frac{9+\sqrt{17}}{8}.$$

For every $D \in \mathcal{D}_{\text{opt}}$, $[1/\sqrt{2}\ \ 1/\sqrt{2}\ \ 0]^T$ is both a left and right singular vector of e^DMe^{-D} corresponding to the maximum singular value of two. Note that the maximum singular value at the optimal scaling

$$\begin{bmatrix} d_* & 0 & 0 \\ 0 & d_* & 0 \\ 0 & 0 & -2d_* \end{bmatrix}$$

is not isolated, as with Example 3.

4. Conclusion. We have derived sufficient conditions for existence and uniqueness of optimal diagonal similarity scalings for scaled singular value minimization. These conditions can be extended to the other structured scaling problems such as block diagonal similarity scaling.

Appendix. More on the examples. *Example* 1. Let d_1, d_2, and d_3 be the diagonal entries of D, with $d_1 + d_2 + d_3 = 0$. Then,

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad e^D M e^{-D} = \begin{bmatrix} 1 & e^{d_1-d_2} & e^{d_1-d_3} \\ e^{d_2-d_1} & 1 & e^{d_2-d_3} \\ 0 & 0 & 1 \end{bmatrix}.$$

We observe that if $d_1 \neq d_2$, then $\|e^D M e^{-D}\| > 2$, since the maximum singular value of the principal 2×2 block exceeds 2. With $d_1 = d_2 = d$, $\|e^D M e^{-D}\| > 2$ once again, since

$$\left(e^D M e^{-D}\right)^* e^D M e^{-D} = \begin{bmatrix} 2 & 2 & 2e^{3d} \\ 2 & 2 & 2e^{3d} \\ 2e^{3d} & 2e^{3d} & 1 + 2e^{6d} \end{bmatrix}$$

is a matrix with positive entries, and therefore its spectral radius (the maximum magnitude of its eigenvalues) is strictly greater than four, which is the spectral radius of its principal 2×2 block (see, for example, [1]). Therefore, it follows that $\|e^D M e^{-D}\| > 2$ for every scaling D.

Finally, we note that with $d_1 = d_2 = d$, as $d \to -\infty$, $\|e^D M e^{-D}\| \to 2$.

A plot of the singular values of of $e^D M e^{-D}$ as a function of d is shown in Fig. 1.

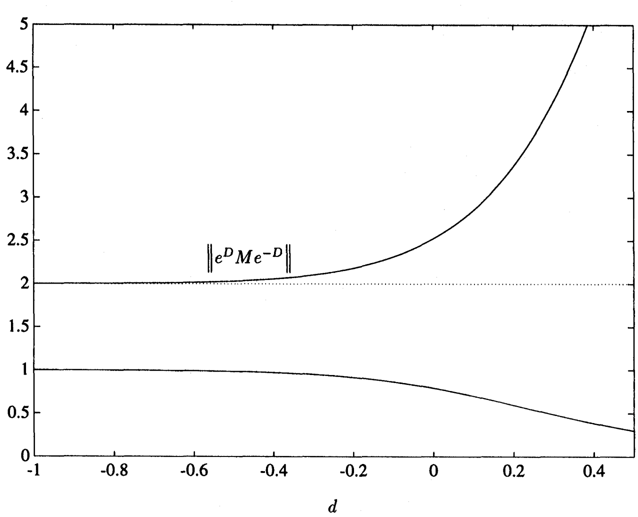

FIG. 1. *Example* 1.

Example 2. We have

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad e^D M e^{-D} = \begin{bmatrix} 1 & e^{d_1-d_2} & e^{d_1-d_3} \\ e^{d_2-d_1} & 1 & -e^{d_2-d_3} \\ 0 & 0 & 1 \end{bmatrix},$$

with $d_1 + d_2 + d_3 = 0$.

Once again, if $d_1 \neq d_2$, then $\|e^D M e^{-D}\| > 2$. However, in contrast with Example 1, with $d_1 = d_2 = d$, $\|e^D M e^{-D}\|$ is only greater than *or equal to* two. Since

$$\left(e^D M e^{-D}\right)^* e^D M e^{-D} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1+2e^{6d} \end{bmatrix},$$

the singular values of $e^D M e^{-D}$ are $\sqrt{1+2e^{6d}}$, 2, and 0. Therefore if $d \leq d_* = \log(3/2)/6$, $\|e^D M e^{-D}\| = 2$.

A plot of the singular values of of $e^D M e^{-D}$ as a function of d is shown in Fig. 2.

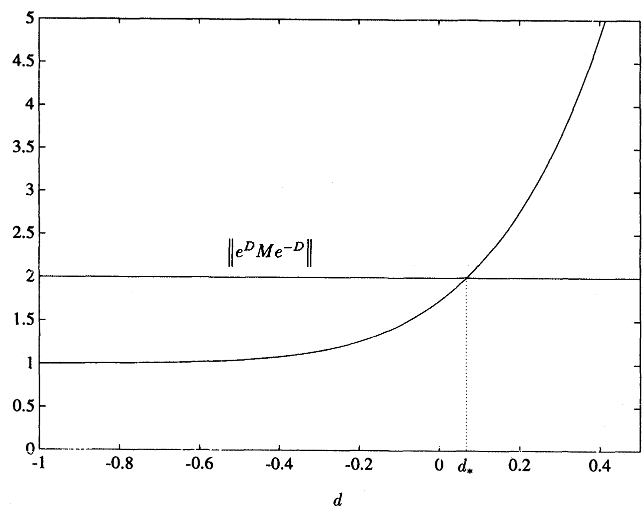

FIG. 2. *Example* 2.

Example 3. We have

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad e^D M e^{-D} = \begin{bmatrix} 1 & e^{d_1-d_2} & -e^{d_1-d_3} \\ e^{d_2-d_1} & 1 & e^{d_2-d_3} \\ -e^{d_3-d_1} & e^{d_3-d_2} & 1 \end{bmatrix},$$

with $d_1 + d_2 + d_3 = 0$.

Once again, if $d_1 \neq d_2$, then $\|e^D M e^{-D}\| > 2$. With $d_1 = d_2 = d$, consider

$$\left(e^D M e^{-D}\right)^* e^D M e^{-D} = \begin{bmatrix} 2+e^{-6d} & 2-e^{-6d} & -e^{-3d} \\ 2-e^{-6d} & 2+e^{-6d} & e^{-3d} \\ -e^{-3d} & e^{-3d} & 1+2e^{6d} \end{bmatrix}.$$

The eigenvalues of this matrix are

$$4, \quad \frac{1}{2}\left((1+2e^{6d}+2e^{-6d}) \pm \sqrt{(1+2e^{6d}+2e^{-6d})^2 - 16}\right).$$

Therefore the maximum singular value of $e^D M e^{-D}$ exceeds two if $d \neq 0$, and equals two if $d = 0$. In other words, the unique optimal scaling is zero, i.e., the "identity" scaling. Note that the maximum singular value at the optimal scaling is not isolated.

A plot of the singular values of of $e^D M e^{-D}$ as a function of d is shown in Fig. 3.

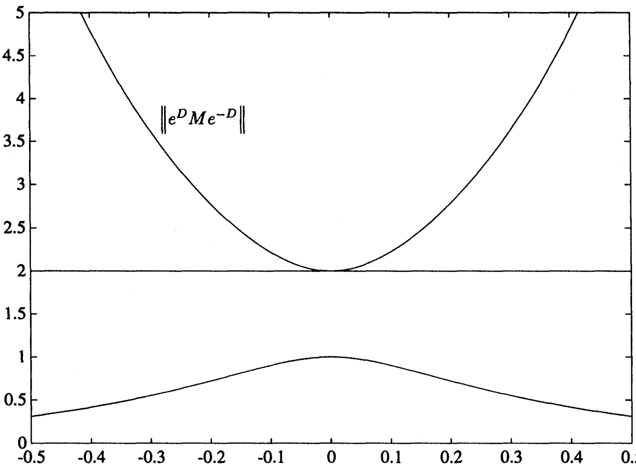

FIG. 3. *Example* 3.

Example 4. We have

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & \dfrac{1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad e^D M e^{-D} = \begin{bmatrix} 1 & e^{d_1-d_2} & -e^{d_1-d_3} \\ e^{d_2-d_1} & 1 & e^{d_2-d_3} \\ -e^{d_3-d_1} & e^{d_3-d_2} & \dfrac{1}{\sqrt{2}} \end{bmatrix},$$

with $d_1 + d_2 + d_3 = 0$.

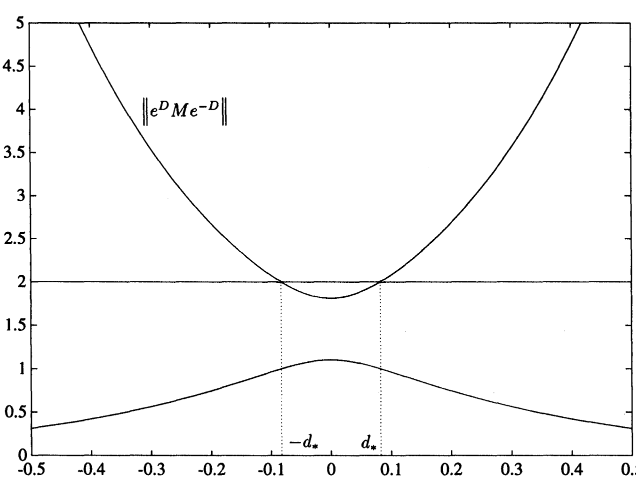

FIG. 4. *Example* 4.

Once again, if $d_1 \neq d_2$, then $\left\|e^D M e^{-D}\right\| > 2$. With $d_1 = d_2 = d$, consider

$$\left(e^D M e^{-D}\right)^* e^D M e^{-D} = \begin{bmatrix} 2 + e^{-6d} & 2 - e^{-6d} & -(1/\sqrt{2})e^{-3d} \\ 2 - e^{-6d} & 2 + e^{-6d} & (1/\sqrt{2})e^{-3d} \\ -(1/\sqrt{2})e^{-3d} & (1/\sqrt{2})e^{-3d} & 1/2 + 2e^{6d} \end{bmatrix}.$$

The eigenvalues of this matrix are

$$4, \quad \frac{1}{2}\left((1/2+2e^{6d}+2e^{-6d}) \pm \sqrt{(1/2+2e^{6d}+2e^{-6d})^2-16}\right).$$

From this, it follows that the maximum singular value of $e^D M e^{-D}$ equals two if $d \in [-d_*, d_*]$, where

$$d_* = (1/6)\log\frac{9+\sqrt{17}}{8}.$$

Note that the maximum singular value of $e^D M e^{-D}$ is isolated for $d \in (-d_*, d_*)$.

A plot of the singular values of of $e^D M e^{-D}$ as a function of d is shown in Fig. 4.

REFERENCES

[1] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Comput. Sci. Appl. Math., Academic Press, New York, 1979.

[2] S. Boyd and V. Balakrishnan, *A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its* $\mathbf{L}_\infty$*-norm*, Systems Control Let., 15 (1990), pp. 1–7.

[3] M. K. Fan and A. L. Tits, *m-form numerical range and the computation of the structured singular value*, IEEE Trans. Automat. Control, 33 (1988), pp. 284–289.

[4] M. K. H. Fan, A. L. Tits, and J. C. Doyle, *Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics*, IEEE Trans. Automat. Control, 36 (1991), pp. 25–38.

[5] G. E. Forsythe and E. G. Straus, *On best conditioned matrices*, Proc. Amer. Math. Soc., 6 (1955), pp. 340–345.

[6] F. R. Gantmacher, *The Theory of Matrices*, Vol. 2, Chelsea, New York, 1959.

[7] A. Greenbaum and G. H. Rodrigue, *Optimal preconditioners of a given sparsity pattern*, BIT, 29 (1989), pp. 610–634.

[8] T. Kato, *A Short Introduction to Perturbation Theory for Linear Operators*, Springer-Verlag, New York, Berlin, 1982.

[9] N. M. Khraishi and A. Emami-Naeini, *A characterization of optimal scaling for structured singular value computation*, Systems Control Lett., 15 (1990), pp. 105–109.

[10] B. D. Moor and S. Boyd, *Analytic properties of singular values and vectors*, Tech. Report ESAT-SISTA Report 1989-28, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, December 1989.

[11] M. G. Safonov, *Stability margins of diagonally perturbed multivariable feedback systems*, IEE Proc., 129-D (1982), pp. 251–256.

[12] R. Sezginer and M. Overton, *The largest singular value of* $e^X A_0 e^{-X}$ *is convex on convex sets of commuting matrices*, IEEE Trans. Automat. Control, 35 (1990), pp. 229–230.

[13] N.-K. Tsing, *Convexity of the largest singular value of* $e^D M e^{-D}$*: a convexity lemma*, IEEE Trans. Automat. Control, 35 (1990), pp. 748–749.