Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems – MTNS 2010 • 5–9 July, 2010 • Budapest, Hungary

Linear Matrix Inequalities for Normalizing Matrices

Christian Ebenbauer

***Abstract*— A real square matrix is normal if it can be diagonalized by an unitary matrix. In this paper novel convex conditions are derived to decide if a matrix is similar to a normal matrix and it is shown how to use these conditions to normalize matrices. The conditions are expressed in terms of linear matrix inequalities and hence they are efficiently solvable using semidefinite programming. Since a matrix is similar to a normal matrix if and only if the matrix is diagonalizable, the present results also provide novel convex conditions for deciding if a matrix is diagonalizable or not. Some applications of these results are presented.**

I. Introduction

An $n \times n$ real square matrix A is normal if there exists an unitary matrix U such that UAU^* is diagonal, where $U^* = U^{-1}$ is the conjugate transpose of the $n \times n$ complex square matrix U. Equivalently, A is normal if and only if

$$[A^T, A] = A^T A - AA^T = 0 \tag{1}$$

with the matrix commutator $[A, B] = AB - BA$. Normal matrices have many interesting properties and applications from eigenvalue computation, graph theory, random matrix assembles to control theory [1], [2], [6]–[8], [10], [14], [19].

For example, in [19] properties of normal matrices have been exploit to obtain stability conditions for a network of nonlinear passive systems. A key property of normal matrices in stability analysis is the fact that a normal matrix A is stable (real parts of the eigenvalues are negative) if and only if $A + A^T$ is negative definite. This follows from the fact that the eigenvalues of $A + A^T$ are twice the real parts of the eigenvalues of A, if A is normal. The eigenvalues of $A - A^T$ are twice the imaginary parts of the eigenvalues of A.

If A is parametrized by some parameters $\gamma = (\gamma_1, \ldots, \gamma_m)$ and if $A(\gamma)$ has the property that it be transformed to a normal matrix $B(\gamma) = T(\gamma)^{-1}A(\gamma)T(\gamma)$, then necessary and sufficient stability conditions for $A(\gamma)$ can be derived, if $B(\gamma)$ is know explicitly. This has been exploited for example in [1], [19] for the stability of networks of nonlinear dissipative systems by identifying A as an adjuncacy matrix and by assuming T be to diagonal.

This work mainly focuses on the normalization of constant matrices. The objective is to find (a) necessary and sufficient conditions to decide if a matrix A is similar to a normal matrix and (b) to compute the similarity transformation which makes A normal.

C. Ebenbauer is with the Institute for Systems Theory and Automatic Control, University of Stuttgart, Germany, ce@ist.uni-stuttgart.de

The main ideas and tools utilized to derive the conditions are an observation which allows one to replace a matrix equation by a matrix inequality and convex duality theory. A review of the literature has shown that such convex conditions seem to be unknown. However, there is a considerable amount of literature related to the current work, see e.g. [12], [16] and references therein.

The initial motivation for this work was twofold, namely to explore normality in stability analysis and the question to decide if a matrix is diagonalizable. Notice that a matrix is similar to a normal matrix if and only if it is diagonalizable [8], [13]. However, due to the relatively large literature on normal matrices, there may be other applications where the results presented here are of interest.

The outline of the remainder of the paper is as following. In Section II the main results are presented. In particular, a necessary and sufficient condition as well as a sufficient condition, both expressed in linear matrix inequalities, is derived to decide if a matrix is diagonalizable/normalizable and how to actually normalize a given matrix. In Section III some applications and numerical examples are presented. Finally, Section IV concludes the work with a summary and outlook.

II. Main Results

In the following, the main results are derived. In particular, three results are established. Two novel conditions in terms of linear matrix inequalities to decide if a matrix is normalizable and a semidefinite program which shows how to actually normalize a matrix based on the first condition.

A. First Condition

Consider an $n \times n$ real square matrix A, the question whether or not A is similar to a normal matrix is equivalent to search for an $n \times n$ real invertible matrix T such that

$$\begin{aligned}[(TAT^{-1})^T, TAT^{-1}] = (TAT^{-1})^T TAT^{-1} \\ -TAT^{-1}(TAT^{-1})^T = 0,\end{aligned} \tag{2}$$

i.e.

$$T^{-T}A^T T^T TAT^{-1} - TAT^{-1}T^{-T}A^T T^T = 0. \tag{3}$$

Now notice that the above matrix equation arise from a commutator, consequently the trace of the left-hand side is identically zero (for any T!). Indeed, any matrix with trace zero can be written in commutator form, i.e. $\mathrm{tr}(Z) = 0$ if and only if there exists an X, Y such that $Z = [X, Y]$ [15]. Since the trace of the left-hand side is always zero, one can replace the matrix equation (3) by the matrix inequality

$$T^{-T}A^T T^T TAT^{-1} - TAT^{-1}T^{-T}A^T T^T \geq 0, \tag{4}$$

where the inequality sign indicates positive semidefiniteness of the left-hand side. This is possible since a positive semidefinite matrix with zero trace is the zero matrix. Consequently, (2) and (3) are equivalent. Next, by defining the positive definite matrix

$$X = T^T T \tag{5}$$

and by multiplying from left by T^T and from the right T, one obtains

$$A^T X A - X A X^{-1} A^T X \geq 0. \tag{6}$$

Notice that (4) and (6) are equivalent to due the fact that every positive definite matrix can be factorized in the form (5) and due to the fact that $P \geq 0$ if and only if $T^T P T \geq 0$, T nonsingular (more specifically, P and $T^T P T$ have the same inertia - congruence transformation). In order to obtain a linear matrix inequality condition, one has to apply the Schur complement argument, i.e. the following statements are equivalent: (a) $[A\ B^T,\ B\ C] \geq 0$, A, C symmetric and A nonsingular (b) $A - B^T C^{-1} B \geq 0$. Hence, (6) is equivalent to

$$\begin{bmatrix} A^T X A & X A \\ A^T X & X \end{bmatrix} \geq 0. \tag{7}$$

Therefore, a necessary and sufficient condition for deciding whether or not a matrix A is normalizable is the following one:

Theorem 1: An $n \times n$ real square matrix A is similar to a normal matrix if and only if there exists a symmetric (positive definite) matrix X such that

$$X \geq I, \; \begin{bmatrix} A^T X A & X A \\ A^T X & X \end{bmatrix} \geq 0. \tag{8}$$

Proof: Since (2) is equivalent to (7) with X as defined in (5), the result follows from the arguments above. Notice that (7) is homogeneous in X, thus if (7) is satisfied for $X > 0$, then so for $\alpha X > 0$, $\alpha > 0$. Consequently, $X \geq I$ is not restrictive. ■

Remarks. First notice that (8) is also a necessary and sufficient condition to decide diagonalizability of A since a matrix is diagonalizable if and only if it is normalizable [8], [13].

Second, due to (3), (6) will either be infeasible or the left-hand side is zero. This is one of the interesting properties of (8) (see also Section IV). Hence, if (8) is feasible, then the Schur complement $A - B^T C^{-1} B$ is the zero matrix and therefore the rank of the $2n \times 2n$ matrix (7) is at most n [5], i.e. if $A^T X A = X A X^{-1} A^T X$ then

$$\begin{bmatrix} A^T X A & X A \\ A^T X & X \end{bmatrix} = \begin{bmatrix} X A \\ X \end{bmatrix} X^{-1} \begin{bmatrix} A^T X & X \end{bmatrix}. \tag{9}$$

Condition (8) is a linear matrix inequality, i.e. X appears linearly in the above inequalities. Thus, semidefinite programming (convex optimization) can be used to check if X exists or not. An important step in deriving condition (8) was the observation that a matrix equation which arises from a commutator can be expressed as a matrix inequality, i.e. a matrix *equation* can be replaced by a matrix *inequality*. That might be useful in other situations as well.

Lemma 1: Suppose the parametrized symmetric matrix $Z(x)$ (not necessarily linear in $x \in \mathbb{R}^m$) can be written in commutator form[1], i.e.

$$Z(x) = [X(x), Y(z)], \tag{10}$$

then the two sets

$$\{x \in \mathbb{R}^m : \; Z(x) = 0\}, \quad \{x \in \mathbb{R}^m : \; Z(x) \geq 0\} \tag{11}$$

are identical.

Proof: The proof follows immediately from the fact that the trace of $Z(x)$ is identical zero and from the fact that a positive semidefinite matrix with zero trace is the zero matrix, i.e. if the sum of all nonnegative eigenvalues is zero, then all eigenvalues are zero. ■

B. Second Condition

Since the first condition for normalization is expressed in terms of convex functions, there is a natural way to derive a second condition based on convex duality. In particular, with the help of theorems of weak alternative from convex analysis, a sufficient condition can be derived when a matrix is not similar to a normal matrix.

Theorem 2: An $n \times n$ real square matrix A is *not* similar to a normal matrix if there exist a nonzero $2n \times 2n$ symmetric (positive semidefinite) matrix Y and a nonzero $n \times n$ symmetric (positive semidefinite) matrix Z such that

$$\begin{bmatrix} Y_1 & Y_2 \\ Y_2^T & Y_3 \end{bmatrix} \geq 0, \; Z \geq 0, \tag{12}$$

$$A Y_1 A^T + Y_2 A^T + A Y_2^T + Y_3 = -Z. \tag{13}$$

[1] Notice that $Z(x)$ is symmetric whenever $X(x)$ is symmetric and $Y(x)$ skewsymmetric. Moreover, every matrix with zero trace can be written in commutator form [15].

Proof: The proof is based on weak duality of semidefinite programming (see [3], p.269f). Consider the optimization problem

$$\begin{array}{ll} min & 0 \\ s.t. & I - X \leq 0 \\ & \begin{bmatrix} -A^T XA & -XA \\ -A^T X & -X \end{bmatrix} \leq 0. \end{array} \tag{14}$$

The Lagrangian of this convex optimization problem is

$$\begin{aligned} L(X,Y,Z) &= \operatorname{tr}(Z(I-X)) \\ &- \operatorname{tr}\left(\begin{bmatrix} Y_1 & Y_2 \\ Y_2^T & Y_3 \end{bmatrix} \begin{bmatrix} A^T XA & XA \\ A^T X & X \end{bmatrix} \right), \end{aligned} \tag{15}$$

where $Y = [Y_1 \ Y_2, Y_2^T \ Y_3], Z$ are positive semidefinite matrices (Lagrange multiplier) of appropriate dimension. The dual function is given by $g(Y,Z) = \inf_X L(X,Y,Z)$. From weak duality follows that if (14) is feasible, then $\sup_{Y \geq 0, Z \geq 0} g(Y,Z) \leq 0$ must hold. By direct calculation can be shown that

$$g(Y,Z) = \begin{cases} \operatorname{tr}(Z) & \text{if } AY_1A^T + Y_2A^T + AY_2 + Y_3 = -Z \\ -\infty & \text{otherwise.} \end{cases} \tag{16}$$

Thus, (14) is infeasible (i.e. A is not similar to a normal matrix) whenever there exist positive semidefinite Y, Z such that $\operatorname{tr}(Z) > 0$ and $AY_1A^T + Y_2A^T + AY_2 + Y_3 = -Z$. ■

Remark. In order to show strong duality, i.e. infeasibility of $\operatorname{tr}(Z) > 0$ and $AY_1A^T + Y_2A^T + AY_2 + Y_3 = -Z$ implies feasiblity of (14), one has to show for example that

$$\begin{array}{ll} min & s \\ s.t. & I - X \leq sI \\ & \begin{bmatrix} -A^T XA & -XA \\ -A^T X & -X \end{bmatrix} \leq \begin{bmatrix} sI & 0 \\ 0 & sI \end{bmatrix}. \end{array} \tag{17}$$

attains its minimum (see [3], p.270, eq. (5.101)). However, strong duality seems to fail in this case. This is reasonable due to the fact that generically any matrix is diagonalizable (the set of diagonalizable matrices is dense).

C. Normalization

A question which often appears is applications is to ask how far away is a certain square matrix from the set of normal matrices. Such nearness problems have received considerable attention in the literature and are in general nontrival problems [12], [14], [16]. Applications lie for example in eigenvalue computation, where finding a normal matrix which is similar to given square matrix might serve as a preconditioner for eigenvalue solvers. Additionally, a measure for nonnormality can serve as an indicator for a highly sensitive (ill-posed) problem [18].

A direct consequence of the derivation of Theorem 1 is the fact that one obtains a matrix T from X such that TAT^{-1} is normal.

Corollary 1: If X satisfy (8) then TAT^{-1} is normal with $X = T^T T$ (Cholesky factorization) or $T = X^{\frac{1}{2}}$.

Moreover, by minimizing some matrix norm of X, i.e. $\|X\|_2, \|X\|_F$, one gets a sort of measure for how far are the eigenvectors from orthogonality and consequently also a measure how ill-conditioned the eigenvalue problem is [18]. For example, the optimal value t^{opt} of the semidefinite program

$$\begin{array}{ll} min & t \\ s.t. & X \geq I, \ tI - X \geq 0 \\ & \begin{bmatrix} A^T XA & XA \\ A^T X & X \end{bmatrix} \geq 0 \end{array} \tag{18}$$

provides such a simple measure (spectral condition number).

Remarks. It is important to point out that semidefinite programming solvers based on interior point methods are probably not the best choice to solve such problems efficiently and accurently, since the solution of (18) has always low rank (see Remarks in Section 2, Section IV). However, the convex condition in Theorem 1 might induce novel ways to (approximately) normalize matrices based on (8) (or on (6) - which has the advantage of not having double dimension). Nevertheless, numerical experiments have shown that interior point methods work well for small to mid-sized matrices ($n \leq 50$, standard laptop with Matlab and Yalmip). Another subtle property of (18) is, that it is feasible for almost every A, since a matrix is normalizable if and only if it is diagonaliable and the set of diagonalizable matrices is dense. It might be also of interst to explore in detail the relationship between semidefinite programs like (18) and nearness normal matrix problems, i.e. does (18), or problems with other matrix norms on X, solve or approximate nearness problems.

Notice that if one diagonalizes the normalized matrix $B = TAT^{-1} = U^*\Lambda U$, $U^*U = I$, $X = T^T T$, which can be done efficiently and reliably - see e.g. [20], then one obtains a diagonal decomposition of A, i.e.

$$A = T^{-1}U^*\Lambda UT = (UT)^{-1}\Lambda UT. \tag{19}$$

A final interesting observation with respect to the SDP (18) is the following. Suppose one replaces A by A^T. Then nothing special will change with respect to feasibility or the optimal value of the SDP. However, one could ask how the new solution Y for A^T relates to the solution X for A. This question is encoded in the following SDP:

$$\begin{array}{ll} min & s + t \\ s.t. & X \geq I, \ tI - X \geq 0 \\ & Y \geq I, \ sI - Y \geq 0 \\ & \begin{bmatrix} A^T XA & XA \\ A^T X & X \end{bmatrix} \geq 0 \\ & \begin{bmatrix} AYA^T & YA^T \\ AY & Y \end{bmatrix} \geq 0. \end{array} \tag{20}$$

It is quite interesting to observe that numerical experiments almost always, when using random matrices, lead to solution pairs (X, Y) with

$$XY = YX = tI = sI. \tag{21}$$

To understand this, notice that in (4), one could also use a $\leq$-sign instead of a $\geq$-sign, i.e.

$$T^{-T}A^T T^T TAT^{-1} - TAT^{-1}T^{-T}A^T T^T \leq 0. \tag{22}$$

Now, by multiplying from left and right with T^{-1}, T^{-T} and by defining

$$Y = T^{-1}T^{-T} \tag{23}$$

one arrives at (6), but A is now replaced by its transposed

$$AYA^T - YA^TY^{-1}AY \geq 0, \tag{24}$$

i.e. Y is a solution of (20). From (5) and (23) follows now $XY = I$. However, since T is not unique, $XY = I$ is not always satisfied. For example, if

$$A = \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 4 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \tag{25}$$

then XY is not a multiple of the identity. Nevertheless, the solution pair is highly structured and the sparsity pattern of X (Y) corresponds exactly to $Z = S^T S$, where the columns of S are the eigenvectors of A. More research effort around these solution properties is necessary, for example, when is T unique modulo scaling and unitary transformations, i.e. uniqueness of $\tilde{T}$ with $T = \alpha U\tilde{T}$, $U^*U = I$; and what happens when using a different objective function, e.g. minimizing the trace of X and Y (Frobenius-norm) instead of the maximal eigenvalues of X and Y (2-norm)?

III. Examples and Applications

A. Measure of Nonnormality

The companion matrix

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ e & 0 & 0 & 0 & 0 \end{bmatrix} \tag{26}$$

is normal precisely if $e = \pm 1$ and not diagonalizable/normalizable precisely if $e = 0$. The solution of the semidefinite program (18) for $A = \Theta^T C \Theta$, where $\Theta^T \Theta = I$ is a random orthognal matrix, with various values of e are shown Figure 1. Figure 1 confirms the facts shown in the previous section. In all feasible cases ($e \neq 0$), the block matrix $[A^T XA \ XA, A^T X \ X]$ had rank 5.

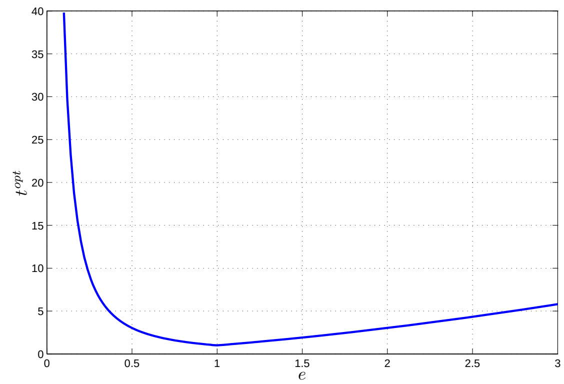

Fig. 1. t^{opt} as a function of $e \in [0.1, 3]$.

B. Matrix Stability

Suppose $B = TAT^{-1}$ is the normalized matrix with $X = T^T T$. Then B has all its eigenvalues in the left half plane (all eigenvalues of B have negative real part) if and only if

$$B + B^T = TAT^{-1} + T^{-T}A^T T^T \leq 0. \tag{27}$$

Multiplying from left with T^T and from right with T yields

$$XA + A^T X \leq 0, \tag{28}$$

which is the well-known Lyapunov inequality. Consequently, a matrix X which satisfies (8) automatically delivers a Lyapunov function $V(x) = x^T Xx$ for a stable linear system $\dot{x} = Ax$.

IV. Summary and Outlook

In this work novel necessary and sufficient conditions for normalizing matrices have been derived. The conditions are expressed in terms of linear matrix inequalities. The conditions also provide a convex formulation of deciding whether or not a matrix is diagonalizable as well as a simple computational scheme to define nonnormality measures for matrices.

Due to the many interesting characterizations of normal matrices, it is hoped that the results in this paper find further interesting applications. For example, counting the number of real eigenvalues of a diagonalizable matrix can be easily done by computing the rank of the skewsymmetric part of the normalized matrix. Another interesting application is the study of the transient behavior of the matrix exponential function $\|e^{At}\|$, e.g. find M, β such that $\|e^{At}\| \leq Me^{\beta t}$. It is easy to see that (18) provide a simple procedure to obtain an M (spectral condition number), [17], p.138.

There are several interesting points for future research, some of them are already mention before. Moreover, as motivated in the introduction, one could consider parametrized matrices $A = A(\gamma)$ and try to find parametrized normalizing transformations $X = X(\gamma)$, possible with the restriction that X is diagonal, as explored in [19]. Another interesting research question is the computational complexity for solving the linear matrix inequality (8)[2]. Due to the fact that a feasible matrix is always low rank, there is no interior point in the feasible set of (8). In general, it is unknown whether or not any linear matrix inequality is solvable in polynomial time since most complexity results of semidefinite programming are based on the assumption that the interior of the feasible set is nonempty (or similar regularity assumptions are imposed) [9]. Interesting in this respect is also the fact that in [11] (see also [4]) it was shown that finding a well-conditioned similarity transformation T to block-diagonalize a nonsymmetric matrix is in general NP-hard.

Finally, the question arise if the novel condition can be used as a design tool, in the sense that if A depends affinely on decision variables z, can the problem be recast (convexified) such that one is able to search for a z^* such that $A(z^*)$ is close to normal. Under what structural conditions on $A(z)$ is this possible?

REFERENCES

[1] M. Arcak and E.D. Sontag. Diagonal stability of a class of cyclic systems and its connection with the secant criterion. *Automatica*, 42:1531–1537, 2006.

[2] R. Bhatia. *Matrix Analysis*. Springer, 1997.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[4] V. Brimkov, B. Codenotti, M. Leoncini, and G. Resta. Strong NP-completeness of a matrix similarity problem. *Theoretical Computer Science*, 165:483–490, 1996.

[5] Y. Chabrillac and J.P. Crouzeix. Definiteness and semidefiniteness of quadratic forms revisited. *Linear Algebra and its Applications*, pages 283–292, 1984.

[6] C. Ebenbauer. A dynamical system that computes eigenvalues and diagonalizes matrices with a real spectrum. In *Proc. of the 46th IEEE Conference on Decision and Control (CDC), New Orleans, USA*, pages 1704–1709, 2007.

[7] C. Ebenbauer and A. Arsie. On an eigenflow equation and its Lie algebraic generalization. *Communications in Information and Systems*, 8:147–170, 2008.

[8] L. Elsner and Kh.D. Ikramov. Normal matrices: an update. *Linear Algebra and its Applications*, 285:291–303, 1998.

[9] M. X. Goemans. Semidefinite programming and combinatorial optimization. *Documenta Mathematica*, Vol III, Extra Volue ICM 1998:657–666, 1996.

[10] R. Grone, C. R. Johnson, E.M. Sa, and H. Wokowicz. Normal matrices. *Linear Algebra and its Applications*, 87:213–225, 1987.

[11] M. Gu. Finding well-conditioned similarities to block-diagonalize nonsymmetric matrices is NP-hard. *Journal of Complexity*, 11:377–391, 1995.

[12] N.J. Higham. Matrix nearness problems and applications. In S. Barnett and J.J.C. Gover, editors, *Applications of matrix theory*, pages 1–27. Oxford University Press, 1989.

[13] B.E. Mitchell. Normal and diagonalizable matrices. *The American Mathematical Monthly*, 60:94–96, 1953.

[14] S. Noschese, L. Pasquini, and L. Reichel. The structured distance to normality of an irreducible real tridiagonal matrix. *Electronic Transactions on Numerical Analysis*, 28:65–77, 2007.

[15] V. Prasolov. *Problems and theorems in linear algebra*. AMS, 1995.

[16] A. Ruhe. Closest normal matrix finally found! *BIT*, 27:585–598, 1987.

[17] L.N. Trefethen and M. Embree. *Spectra and Pseudospectra: The behavior of nonnormal matrices and operators*. Princeton University Press, 2005.

[18] L.N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, 1997.

[19] H.K. Wimmer. Diagonal stability of matrices with cyclic structure and the secant condition. *Systems and Control Letters*, 58:309–313, 2008.

[20] B.B. Zhou and R.P. Brent. An efficient method for computing eigenvalues of a real normal matrix. *Journal of Parallel and Distributed Computing*, 63:638–648, 2003.

[2]The author would like to thank Pablo Parrilo for pointing out this question.