Mathematical
Inequalities
& Applications
Volume 1, Number 1 (1998), 149–152

A NOTE ON SIMULTANEOUSLY DIAGONALIZABLE MATRICES

ABRAHAM BERMAN[1] AND ROBERT J. PLEMMONS[2]

(*communicated by A. Ben-Israel*)

Abstract. Consider the functional $f(U) = \sum_{i=1}^{n} \max_j \{(U^T M_j U)_{ii}\}$ over orthogonal matrices U, where the collection of n-by-n symmetric matrices M_j are pairwise commutative, and thus simultaneously diagonalizable. Selecting an orthogonal matrix U which maximizes $f(U)$ has applications in adaptive optics. A proof is given here that any orthogonal matrix Q which simultaneously diagonalizes the M_j maximizes the function f.

1. Introduction

Recently, there has been a growing interest in adaptive optics, i.e., methods to overcome the effects of distortion in imaging through a medium, such as the atmosphere [6, 4, 7]. In adaptive optics, the following optimization problem arises [4]. The problem involves maximizing the functional

$$f(U) = \sum_{i=1}^{n} \max_j \{(U^T M_j U)_{ii}\} \tag{1}$$

over all n-by-n orthogonal matrices U. In this note we show that if the matrices M_j are simultaneously diagonalizable, then any orthogonal matrix Q which simultaneously diagonalizes the M_j maximizes f.

2. An Optimization Result

Recall that a set of real symmetric matrices can be simultaneously diagonalized by an orthogonal matrix if and only if they are pairwise commutative, e.g., [5], Theorems 1.3.19 and 2.3.3. For such matrices we state and prove our main result.

Mathematics subject classification (1991): 15A27, 15A45, 15A51, 15A90.
Key words and phrases: Simultaneous diagonalization, commutativity, adaptive optics, trace maximization.

[1] Research supported by the Fund for Promotion of Research at the Technion.

[2] Research sponsored by the Air Force Office of Scientific Research and the National Science Foundation.

Paper MIA-01–13

THEOREM. *Suppose* $\{M_j\}$, $1 \leqslant j \leqslant k$, *is a collection of pairwise commuting matrices. Let* Q *be any orthogonal matrix which simultaneously diagonalizes the* M_j. *Then* Q *maximizes the functional given in* (1).

Proof. Let $B_j = U^T M_j U$, where U is an arbitrary orthogonal matrix. We can rewrite B_j using the orthogonal matrix $V = Q^T U$:

$$B_j = U^T M_j U = (QV)^T M_j (QV) = V^T Q^T M_j Q V = V^T D_j V \, ,$$

where $D_j = Q^T M_j Q$ is the diagonalization of M_j using Q. Observe that

$$f(U) = \sum_{i=1}^{n} \max_j \{(B_j)_{ii}\}$$

and that

$$f(Q) = \sum_{i=1}^{n} \max_j \{(D_j)_{ii}\}.$$

We have to show that $f(Q) \geqslant f(U)$. To do this we define matrices $D = \operatorname{diag}(d_{ss})$ and $B = (b_{st})$ as follows: $d_{ss} = \max_j (D_j)_{ss}$ and $b_{st} = \max_j (B_j)_{st}$. For example, d_{11} is the largest element of $\{(D_1)_{11}, (D_2)_{11}, \ldots, (D_k)_{11}\}$. In other words,

$$D = \max_{1 \leqslant j \leqslant k} \{D_j\} \, , \quad \text{and} \quad B = \max_{1 \leqslant j \leqslant k} \{B_j\} \, ,$$

where the maxima are evaluated *elementwise*. We will use $\tilde{d}$, $\tilde{d}_j$, $\tilde{b}$, and $\tilde{b}_j$ to represent the column vectors of the diagonals of D, D_j, B, and B_j, respectively. For example, $\tilde{b}_1$ is the column vector whose entries are the diagonal elements of B_1. Finally, we let e be a column vector of ones. With this notation, the traces of D and B can be written in the form $e^T \tilde{d}$ and $e^T \tilde{b}$. So, proving the theorem is equivalent to establishing the inequality

$$e^T \tilde{d} \geqslant e^T \tilde{b} \, . \tag{2}$$

Let $S = V \circ V$, the Hadamard product of V with itself; namely,

$$s_{ij} = v_{ij}^2.$$

A matrix which can be expressed as $S = V \circ V$, where V is orthogonal is called orthostochastic. Every orthostochastic matrix is doubly stochastic, i. e., $s_{ij} \geqslant 0$ and $e^T S = e^T S^T = e^T$, e. g., [1].

The doubly stochastic matrix S^T can be used to relate $\tilde{b}_j$ and $\tilde{d}_j$:

$$\tilde{b}_j = (V \circ V)^T \tilde{d}_j = S^T \tilde{d}_j, \quad 1 \leqslant j \leqslant k \, . \tag{3}$$

By definition, $\tilde{d} \geqslant \tilde{d}_j$ for all j. Since S^T is nonnegative, we can write

$$S^T \tilde{d} \geqslant S^T \tilde{d}_j = \tilde{b}_j, \quad 1 \leqslant j \leqslant k \, . \tag{4}$$

This inequality holds elementwise; thus, we have

$$\left[S^T\tilde{d}\right]_i \geqslant \left[\tilde{b}_j\right]_i$$

for *all* values of i. Therefore,

$$S^T\tilde{d} \geqslant \max_j \tilde{b}_j = \tilde{b}\,. \tag{5}$$

Here again, the maximum is taken elementwise. After multiplying both sides of inequality (5) by e^T and noting that $e^T S^T = e^T$, we obtain inequality (2), thereby proving the theorem. □

3. Remarks

The note is concluded with some remarks on the theorem and on the computational aspects of the problem.

• The converse of the theorem is not true; i.e., an orthogonal matrix Q can maximize $f(U)$ without simultaneously diagonalizing the pairwise commuting M_j. In fact, let $M_1, \cdots, M_k$ be n-by-n diagonal matrices such that $M_1 \geqslant M_j$ for all j. Then for *all* n-by-n orthogonal matrices U, $f(U) = \text{trace}(M_1)$. But, in the case where one of the matrices has distinct diagonal entries, the only matrices that simultaneously diagonalize the (diagonal matrices) M_j are products of permutation and signature matrices.

• Although the theorem establishes the existence of an orthogonal matrix Q that maximizes the functional $f(U)$ where the M_j are pairwise commutative, the process of computing the simultaneous diagonalizer Q can be quite nontrivial. A comprehensive study of numerical methods for simultaneous diagonalization of a single pair of commuting matrices is given in [2].

• The work on adaptive optics that motivated our discussion concerns real matrices. A similar theorem can be stated for Hermitian matrices M_j and a unitary matrix Q. Other recent applications of simultaneous diagonalization and related algorithms can be found in multivariate statistics [3].

REFERENCES

[1] A. BERMAN AND R. J. PLEMMONS, *Nonnegative Matrices in the Mathematical Sciences*, 2nd Edition, SIAM Press Classics Series no. 9, Philadelphia, 1994.
[2] A. BUNSE-GERSTNER, R. BYERS, AND V. MEHRMANN, *Numerical methods for simultaneous diagonalization*, SIAM. J. Matrix Anal. and Applic., **14** (1993), 927–949.
[3] B. D. FLURY AND B. E. NEUENSCHWANDER, *Simultaneous diagonalization algorithms with applications in multivariate statistics*, in Approximation and Computation (R. Zahar, ed.), pp. 179–205, Basel: Birkhauser, 1995.
[4] J. W. HARDY, *Adaptive optics*, Scientific American, **270, no. 6**, June (1994), 60–65.
[5] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, England, 1985.
[6] J. NELSON, *Reinventing the telescope*, Popular Science, 85, January (1995), 57–59.

Mathematical Inequalities & Applications
www.ele-math.com
mia@ele-math.com

[7] R. A. SMITH, *Seeing the universe through new eyes*, National Geographic Magazine, **185, no. 1**, January (1994), 2–41.

(Received July 7, 1997)

Abraham Berman
Department of Mathematics
Technion–Israel Institute of Technology
Haifa 32000
Israel
e-mail: berman@tx.technion.ac.il

Robert J. Plemmons
Department of Computer Science
Wake Forest University
Winston-Salem, NC 27109
e-mail: plemmons@mthcsc.wfu.edu
URL: http://www.mthcsc.wfu.edu/ ~ plemmons/