SIAM REV. © 1997 Society for Industrial and Applied Mathematics
Vol. 39, No. 2, pp. 305–309, June 1997 006

A UNIFIED ELEMENTARY APPROACH TO CANONICAL FORMS OF MATRICES*

JOHN KARRO† AND CHI-KWONG LI‡

Abstract. A unified elementary approach is used to obtain various canonical-form theorems for complex and real matrices.

Key words. canonical form, orthogonal matrix, unitary matrix

AMS subject classification. 15A18

PII. S0036144595294862

1. Introduction. One of the most fruitful ideas in the theory of matrices is that of a matrix decomposition or canonical form. In fact, this topic has applications to many pure and applied subjects such as geometry, statistics, differential equations, approximation theory, and control theory (e.g., see [L, Chapter 8], [G, Chapter 9], [HJ1, Chapter 3], [HJ2, Chapter 3], [S], and their references).

While the canonical-form theorems for matrices are beautiful and have many applications, they are usually not introduced to students in the first matrix theory course. This is partly due to the fact that many of the standard proofs of the canonical-form theorems are not simple and require a lot of background such as the theory of eigenvalues, singular values, and eigenvectors. This is especially frustrating for students of other disciplines such as engineering, physics, and statistics, who would not be able to manage the canonical-form theorems needed in their study after a full semester course of linear algebra.

The purpose of this note is to use a unified elementary approach, requiring only some calculus theory and the fact that a continuous function attains its maximum on a compact set, to prove various basic canonical-form theorems for matrices.

We shall illustrate our approach by proving several basic results in the next section. Some additional techniques and results will be discussed in section 3. The standard basis for $\mathbb{R}^n$ or $\mathbb{C}^n$ will be denoted by $\{e_1, \ldots, e_n\}$, and the standard basis for $m \times n$ matrices will be denoted by $\{E_{11}, E_{12}, \ldots, E_{mn}\}$.

2. Basic results. We first present the proof of the canonical-form theorem for real symmetric matrices.

THEOREM 1. *Suppose A is an $n \times n$ real symmetric matrix. Then there is an orthogonal matrix U such that $A = U(\sum_{i=1}^n \lambda_i E_{ii})U^t$ with $\lambda_1 \geq \cdots \geq \lambda_n$ ($\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A).*

Proof. We divide the proof into three steps.

Step 1. For the given symmetric matrix A, let $u \in \mathbb{R}^n$ satisfy $u^t u = 1$ and

$$u^t A u = \max\{x^t A x : x \in \mathbb{R}^n,\ x^t x = 1\}.$$

*Received by the editors May 11, 1995; accepted for publication (in revised form) October 11, 1995.

http://www.siam.org/journals/sirev/39-2/29486.html

†Department of Computer Science, University of Virginia, Charlottesville, VA 22903 (karro@virginia.edu).

‡Department of Mathematics, The College of William and Mary, Williamsburg, VA 23187 (ckli@math.wm.edu).

The existence of such a vector u is guaranteed by the fact that $x \mapsto x^t Ax$ is a continuous function and $\{x \in \mathbb{R}^n : x^t x = 1\}$ is a compact set.

Step 2. Extend u to an orthogonal matrix U with u as the first column. This can be done by applying the Gram–Schmidt process to a basis for $\mathbb{R}^n$ of the form $\{u, e_{j_2}, \ldots, e_{j_n}\}$, where u is not in the linear span of $\{e_{j_2}, \ldots, e_{j_n}\}$. We claim that $B = U^t AU$ is of the form $\left[\begin{smallmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{smallmatrix}\right]$, where $\lambda_1 = u^t Au$ and A_1 is an $(n-1) \times (n-1)$ symmetric matrix. Note that B is symmetric and $B_{11} = \lambda_1$ by our construction of U. For $2 \leq k \leq n$, let $x(\theta) = U(\cos\theta\, e_1 + \sin\theta\, e_k)$ and

$$f(\theta) = x(\theta)^t Ax(\theta) = \cos^2\theta\,\lambda_1 + 2\cos\theta\sin\theta\, B_{1k} + \sin^2\theta\, B_{kk}$$

with $\theta \in \mathbb{R}$. Our assumption on λ_1 implies that the real-valued function f attains a maximum at $\theta = 0$. Hence $0 = f'(0) = 2B_{1k}$, and our claim is proved.

Step 3. One can now apply an inductive argument to A_1 to complete the proof. □

It is worth mentioning that the unit vector u in Step 1 of the proof is a unit eigenvector corresponding to the largest eigenvalue λ_1 of the symmetric matrix A. The fact that $\lambda_1 = u^t Au$ is the solution of the maximization problem in Step 1 follows from the Rayleigh principle (e.g., see [HJ1, Theorem 4.2.2]). Of course, this knowledge is not required in our proof.

The theory of complex symmetric matrices is useful in the study of complex function theory and physics (e.g., see [HJ1, section 4.4]). The proof of Theorem 1 can be adapted to obtain the following canonical-form theorem for complex symmetric matrices.

THEOREM 2. *Suppose A is an $n \times n$ complex symmetric matrix. Then there is a unitary matrix U such that $A = U(\sum_{i=1}^n s_i E_{ii})U^t$ with $s_1 \geq \cdots \geq s_n \geq 0$ ($s_1, \ldots, s_n$ are the singular values of A).*

Proof. Again we divide the proof into three steps.

Step 1. For a given complex symmetric matrix A, let $u \in \mathbb{C}^n$ satisfy $u^*u = 1$ and

$$u^t Au = \max\{\mathrm{Re}\,(x^t Ax) : x \in \mathbb{C}^n,\ x^*x = 1\}.$$

The existence of u is guaranteed by the fact that $x \mapsto \mathrm{Re}\,(x^t Ax)$ is a continuous function and $\{x \in \mathbb{C}^n : x^*x = 1\}$ is a compact set.

Step 2. Extend u to a unitary matrix U with u as the first column. We claim that $B = U^t AU$ is of the form $\left[\begin{smallmatrix} s_1 & 0 \\ 0 & A_1 \end{smallmatrix}\right]$, where $s_1 = u^t Au$ and A_1 is an $(n-1) \times (n-1)$ symmetric matrix. Clearly, $s_1 \geq 0$; otherwise one may replace u by μu for a suitable $\mu \in \mathbb{C}$ with $|\mu| = 1$ such that $\mathrm{Re}\,(\mu^2 s_1) = \mathrm{Re}\,((\mu u)^t A(\mu u)) = |s_1| > \mathrm{Re}\,(s_1)$. Note that B is symmetric and $B_{11} = s_1$ by our construction of U. For $2 \leq k \leq n$ and for any $\mu \in \mathbb{C}$ with $|\mu| = 1$, let $x(\theta) = U(\cos\theta\, e_1 + \mu\sin\theta\, e_k)$ and

$$f(\theta) = \mathrm{Re}\,(x(\theta)^t Ax(\theta)) = \cos^2\theta\, s_1 + 2\cos\theta\sin\theta\,\mathrm{Re}\,(\mu\, B_{1k}) + \sin^2\theta\,\mathrm{Re}\,(\mu^2\, B_{kk})$$

with $\theta \in \mathbb{R}$. Our assumption on s_1 implies that the real-valued function f attains a maximum at $\theta = 0$. Hence $0 = f'(0) = 2\mathrm{Re}\,(\mu B_{1k})$. Since this is true for all $\mu \in \mathbb{C}$ with $|\mu| = 1$, we have $B_{1k} = 0$. Thus our claim is proved.

Step 3. One can now apply an inductive argument to A_1 to complete the proof. □

The author in [Ho] gave a simple proof of the singular value decomposition theorem which asserts the following:

Every real $m \times n$ matrix A can be written as $U(\sum_{i=1}^k s_i E_{ii})V^t$, where U and V are orthogonal matrices of appropriate sizes, k is the rank of A, and $s_1 \geq \cdots \geq s_k > 0$ (are the singular values of A).

In fact, our approach uses some of the ideas in [Ho]. One can reorganize the proof of Hoechsmann in terms of our three basic steps as follows.

Step 1. For a given real $m \times n$ matrix A, determine unit vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ such that

$$u^t Av = \max\{x^t Ay : x \in \mathbb{R}^m,\ y \in \mathbb{R}^n, x^t x = y^t y = 1\}.$$

The existence of such u and v is guaranteed by the fact that $(x, y) \mapsto x^t Ay$ is a continuous function and $\{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n : x^t x = y^t y = 1\}$ is a compact set.

Step 2. Extend u (respectively, v) to an orthogonal matrix U (respectively, V) with u (respectively, v) as the first column. Show that $U^t AV$ is of the form $\left[\begin{smallmatrix} s_1 & 0 \\ 0 & A_1 \end{smallmatrix}\right]$ with $A_1 \in \mathbb{R}^{(m-1)\times(n-1)}$.

Step 3. Apply an inductive argument to A_1 to complete the proof.

Horn informed us that the idea of this type of variational proof of the singular value decomposition theorem has also been used by other authors, including Jordan in 1874 (see [HJ2, section 3.0]). It is worth mentioning that a different choice of the function to be optimized in Step 1 may make a significant difference in the computations required in Step 2 (e.g., see [HJ2, Theorem 3.1.1], [Ho, Theorem 1], and [PS]). That is why in each of our proofs we try to indicate clearly how to define the objective function and reduce the problem to a lower dimension.

Hoechsmann [Ho] pointed out that his proof of the singular value decomposition can be adapted to the complex case. Furthermore, after proving the singular value decomposition theorem, he showed that a symmetric matrix has a one-dimensional invariant subspace using elementary arguments and then obtained a proof for our Theorem 1.

3. Additional techniques and results. Recall that a complex matrix is *Hermitian* if $A = A^*$, where A^* denotes the conjugate transpose of A. Complex Hermitian matrices are always regarded as the analog of real symmetric matrices in the study of operator theory and quadratic forms. Also, complex symmetric matrices do not seem to occur in applications nearly as often as complex Hermitian matrices. We have the following canonical-form theorem for complex Hermitian matrices.

THEOREM 3. *Suppose A is an $n \times n$ complex Hermitian matrix. Then there is a unitary matrix U such that $A = U(\sum_{i=1}^n \lambda_i E_{ii})U^*$ with $\lambda_1 \geq \cdots \geq \lambda_n$ ($\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A).*

Proof. Again we divide the proof into three steps.

Step 1. For a given complex Hermitian matrix A, let $u \in \mathbb{C}^n$ satisfy $u^* u = 1$ and

$$u^* Au = \max\{x^* Ax : x \in \mathbb{C}^n,\ x^* x = 1\}.$$

It is routine to check that $\overline{x^* Ax} = (x^* Ax)^* = x^* A^* x = x^* Ax$ and hence $x^* Ax \in \mathbb{R}$ for any $x \in \mathbb{C}^n$, and the existence of u is guaranteed by the fact that $x \mapsto x^* Ax$ is a continuous function and $\{x \in \mathbb{C}^n : x^* x = 1\}$ is a compact set.

Step 2. Extend u to a unitary matrix U with u as the first column. We claim that $B = U^* AU$ is of the form $\left[\begin{smallmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{smallmatrix}\right]$, where $\lambda_1 = u^* Au$ and A_1 is an $(n-1) \times (n-1)$ symmetric matrix. Note that B is hermitian and $B_{11} = \lambda_1 \in \mathbb{R}$ by our construction of U. For $2 \leq k \leq n$ and for any $\mu \in \mathbb{C}$ with $|\mu| = 1$, let $x(\theta) = U(\cos\theta\, e_1 + \mu \sin\theta\, e_k)$ and

$$f(\theta) = x(\theta)^* Ax(\theta) = \cos^2\theta\, \lambda_1 + 2\cos\theta \sin\theta\, \mathrm{Re}\,(\mu\, B_{1k}) + \sin^2\theta\, B_{kk}$$

with $\theta \in \mathbb{R}$. Our assumption on λ_1 implies that the real-valued function f attains a maximum at $\theta = 0$. Hence $0 = f'(0) = 2\mathrm{Re}\,(\mu B_{1k})$. Since this is true for all $\mu \in \mathbb{C}$ with $|\mu| = 1$, we have $B_{1k} = 0$. Thus our claim is proved.

Step 3. One can now apply an inductive argument to A_1 to complete the proof. □

As pointed out by Horn, once a "basic" canonical-form theorem is obtained one may get additional canonical-form theorems for other classes of matrices using the special structure of them (e.g., see [A]). We illustrate this idea by proving the canonical-form theorem for normal matrices in Theorem 4. Recall that an $n \times n$ complex matrix A is *normal* if $AA^* = A^*A$. If $H = (A + A^*)/2$ and $G = (A - A^*)/(2i)$, then H and G are Hermitian matrices such that $A = H + iG$. Moreover, A is normal if and only if $HG = GH$.

THEOREM 4. *Suppose A is an $n \times n$ complex normal matrix. Then there is a unitary matrix U such that $A = U(\sum_{i=1}^n \lambda_i E_{ii})U^*$ with $\lambda_i \in \mathbb{C}$ ($\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A).*

Proof. Let $A = H + iG$ with $H = (A + A^*)/2$ and $G = (A - A^*)/(2i)$. By Theorem 3, there is a unitary matrix W such that $W^*HW = \alpha_1 I_{n_1} \oplus \cdots \oplus \alpha_k I_{n_k}$, where $\alpha_1 > \cdots > \alpha_k$ and $n_1 + \cdots + n_k = n$. Since $HG = GH$, we see that $\tilde{H} = W^*HW$ and $\tilde{G} = W^*GW$ also satisfy $\tilde{H}\tilde{G} = \tilde{G}\tilde{H}$. It follows that $\tilde{G} = G_1 \oplus \cdots \oplus G_k$, where G_i is an $n_i \times n_i$ Hermitian matrix for $1 \le i \le k$. By Theorem 3 again, there are unitary matrices V_i such that $V_i^* G_i V_i$ is in diagonal form for all i. Let $U = W(V_1 \oplus \cdots \oplus V_k)$. Then $U^*AU = U^*(H + iG)U$ is in diagonal form as asserted. □

In [Ho] the author remarked that his elementary approach could also be used to prove the spectral theorem of complex normal matrices. It is unclear to us whether the method in their mind is the same as ours.

A less well-known and more involved result is the canonical-form theorem for skew-symmetric matrices, those matrices A satisfying $A = -A^t$ (e.g., see [G, Chapter 9]). The canonical-form theorem for skew-symmetric matrices can also be proved by our variational method, as shown in the following.

THEOREM 5. *Suppose A is an $n \times n$ real skew-symmetric matrix. Then there is an orthogonal matrix U such that $A = U[\sum_{i \le n/2} s_i(E_{2i-1,2i} - E_{2i,2i-1})]U^t$, where $s_1 \ge s_2 \ge \cdots$ are nonnegative real numbers ($s_1, s_2, \ldots$, are the singular values of A).*

Proof. Again we divide the proof into three steps.

Step 1. For the given skew-symmetric matrix A, let $u, v \in \mathbb{R}^n$ satisfy $u^t u = v^t v = 1$, $u^t v = 0$, and

$$u^t A v = \max\{x^t A y : x, y \in \mathbb{R}^n,\ x^t x = y^t y = 1, x^t y = 0\}.$$

The existence of the orthonormal pair (u, v) is guaranteed by the fact that $(x, y) \mapsto x^t A y$ is a continuous function and $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : x^t x = y^t y = 1, x^t y = 0\}$ is a compact set. If $A = 0$, we can set $s_i = 0$ for all $i \le n/2$ and we are done. If $A \ne 0$, then

$$s_1 = u^t A v \ge \max\{A_{ij} : 1 \le i, j \le n\} > 0.$$

Step 2. Extend u, v to an orthogonal matrix U with u and v as the first two columns. We claim that $B = U^t A U$ is of the form $\left[\begin{smallmatrix} 0 & -s_1 \\ s_1 & 0 \end{smallmatrix}\right] \oplus A_1$, where A_1 is an $(n-2) \times (n-2)$ skew-symmetric matrix. Since B is also skew-symmetric, $B_{ii} = 0$ for all i. By our construction of U, we have $B_{12} = s_1$, and hence $B_{21} = -s_1$. For $2 < k \le n$, let $x(\theta) = U(\cos\theta\, e_2 + \sin\theta\, e_k)$ and

$$f(\theta) = u^t A x(\theta) = e_1^t B(\cos\theta\, e_2 + \sin\theta\, e_k) = \cos\theta\, s_1 + \sin\theta\, B_{1k}$$

with $\theta \in \mathbb{R}$. Our assumption on s_1 implies that the real-valued function f attains a maximum at $\theta = 0$. Hence $0 = f'(0) = \mu B_{1k}$. Since this is true for all $\mu \in \mathbb{C}$ with $|\mu| = 1$, we have $B_{1k} = 0$. Similarly, for $2 < k \leq n$, one may consider $y(\theta) = U(-\cos\theta\, e_1 + \sin\theta\, e_k)$ and

$$g(\theta) = v^t A y(\theta) = e_2^t B(-\cos\theta\, e_1 + \sin\theta\, e_k) = \cos\theta\, s_1 + \sin\theta\, B_{2k}$$

with $\theta \in \mathbb{R}$ and conclude that $B_{2k} = 0$. Thus our claim is proved.

Step 3. Apply an inductive argument to A_1 until we get a zero block or an empty block to complete the proof. □

The same proof can be applied to the complex case if one modifies Step 1 to the following.

Step 1. For a given complex skew-symmetric matrix A, determine $u, v \in \mathbb{C}^n$ that satisfy $u^* u = v^* v = 1$, $u^* v = 0$, and

$$u^t A v = \max\{\mathrm{Re}\,(x^t A y) : x, y \in \mathbb{C}^n,\ x^* x = y^* y = 1,\ x^* y = 0\}$$

and consider complex unit vectors $x(\theta)$ and $y(\theta)$ in the proof of Step 2.

Note that Theorem 5 (and its complex version) implies that the singular values of a skew-symmetric matrix always occur in pairs and the rank of a skew-symmetric matrix must be even.

Acknowledgments. Thanks are due to Dr. R. Horn and Dr. B. S. Tam for their helpful comments and suggestions.

REFERENCES

[A] L. AUTONNE, *Sur les matrices hypohermitiennes et sur les matrices unitaires*, Ann. Univ. Lyon, Nouvelle Série I, Fasc., 38 (1915), pp. 1–77.

[G] F. R. GANTMACHER, *The Theory of Matrices*, 2 vols., Chelsea, New York, 1959.

[HJ1] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, New York, 1985.

[HJ2] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, New York, 1991.

[Ho] K. HOECHSMANN, *Singular values and the spectral theorem*, AMS Monthly, 97 (1990), pp. 413–414.

[L] D. C. LAY, *Linear Algebra and Its Applications*, Addison–Wesley, New York, 1993.

[PS] C.-T. PAN AND K. SIGMON, *A bottom-up inductive proof of the singular value decomposition*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 59–61.

[S] G. W. STEWART, *On the early history of the singular value decomposition*, SIAM Rev., 35 (1993), pp. 551–566.