

MATRIX THEORY AND ITS APPLICATIONS

SELECTED TOPICS

MATRIX THEORY AND ITS APPLICATIONS

SELECTED TOPICS

N.J. Pullman

Department of Mathematics
Queen's University at Kingston
Ontario, Canada

MARCEL DEKKER, INC. New York and Basel

MATRIX THEORY
AND ITS
APPLICATIONS

COPYRIGHT © 1976 MARCEL DEKKER, INC. ALL RIGHTS RESERVED.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage and retrieval system, without permission in writing from the publisher.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 75-40845

ISBN: 0-8247-6420-X

Current printing (last digit):

10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

CHAPTER 1

ALGEBRAIC AND ANALYTIC PRELIMINARIES

1. NOTATION

We shall reserve all the capital Latin Letters (A, B , etc.) for matrices, the letters u, v, w, x, y, z for column vectors and reserve the letters i, j, k, ℓ, m, n for integer indices (except when $i = \sqrt{-1}$). The letters f, g, h are reserved for functions. The entries in a matrix A (or a vector x) are denoted by the same letter with the appropriate subscripts (a_{ij} or x_j). Other scalars are denoted by lowercase Greek letters (α, β , etc.). Unless we specify otherwise, when we say *matrix* we mean complex, $k \times k$ matrix, when we say *scalar* (or *number*) we mean complex number, and usually when we say *vector* we mean a complex $k \times 1$ matrix (a complex *column vector*). Other notation will be explained as we go along. These are collected in the Notation at the end of the book.

2. BACKGROUND

You are expected to recall, or at any rate to review, certain basic ideas of matrix theory: initially you should make sure you know the meaning of *eigenvalue*, *eigenvector* and *characteristic polynomial*. This involves remembering

what the *determinant* of a matrix is, reviewing its properties, and, in particular, recalling its connection with the question of whether or not a given matrix is *singular*. You should also be familiar with the ideas of *vector space* and *subspace*, *linear dependence* and *independence*, *linear combinations of vectors*, the space *spanned by* a set of vectors, the *basis* of a vector space and its *dimension* and *linear mapping* and *linear operator* as given in most standard first courses in linear algebra. When the occasion arises, references to a standard textbook or reference book will be given. The details of the reference (title, publisher, etc.) will be found in the Reference section at the end of the book. Suggestions for collateral reading are also given there.

You are also expected to have had some experience with elementary real analysis (and to a somewhat lesser extent with complex analysis) at least to being able to deal with infinite series and sequences.

3. SIMILARITY AND SIMILARITY INVARIANTS

A matrix A is said to be *similar* to a matrix B (written: $A \sim B$) if and only if there is a matrix P such that $A = PBP^{-1}$. Geometrically speaking, $A \sim B$ means that A and B both represent the same linear operator. You should be able to see why \sim is an equivalence relation (see Exercise 1).

Similarity Invariants

A function f defined on matrices is said to be a *similarity invariant* if $f(A) = f(B)$, whenever $A \sim B$. Such functions cannot distinguish between similar matrices. For example, the determinant, trace, and rank are similarity invariants.

The characteristic polynomial is a similarity invariant (it too is a function with a matrix argument, but it takes on polynomial values). The minimal polynomial is another polynomial-valued similarity invariant.

A set of matrices is said to be *similarity invariant* (or *invariant under similarity*) if all matrices similar to a matrix in the set are also in the set. For instance, given α , β , and γ , the set of all solutions X to $\alpha X^2 + \beta X + \gamma I = 0$ is invariant under similarity (see Exercise 5). A property is said to be *invariant under similarity* (or *similarity invariant*) if the set of matrices having the property is similarity invariant; for example, invertibility is a similarity invariant.

Exercises

1. Prove that similarity is an equivalence relation, i.e., that for all A , B , and C : $A \sim A$; if $A \sim B$, then $B \sim A$; if $A \sim B$ and $B \sim C$, then $A \sim C$.
2. a. If I is the identity matrix, show that λI is the only matrix similar to λI .

- b. Show that $A = \lambda I$ if A is the only matrix similar to A . Try to do this without using Sec. 6. (*Hint:* If A is the only matrix similar to A , then $PA = AP$ for all invertible matrices P . Thus, for example,

$$\text{if } A = \begin{bmatrix} \lambda & \beta \\ \gamma & \delta \end{bmatrix}, \text{ then } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} A = A \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$\begin{bmatrix} \gamma & \delta \\ \lambda & \beta \end{bmatrix} = \begin{bmatrix} \beta & \lambda \\ \delta & \gamma \end{bmatrix}, \text{ and } \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} A = A \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}; \text{ so}$$

$$\begin{bmatrix} \gamma & \delta \\ -\lambda & -\beta \end{bmatrix} = \begin{bmatrix} -\beta & \lambda \\ -\delta & \gamma \end{bmatrix}. \text{ Therefore } \beta = \gamma = -\beta \text{ and}$$

$\lambda = \delta$; hence $\beta = \gamma = 0$ and $A = \lambda I$. Now extend this method so it will work for matrices of arbitrary order.)

3. a. If g is a matrix polynomial (that is, $g(T) = \alpha_0 I + \alpha_1 T + \alpha_2 T^2 + \dots + \alpha_n T^n$ for all square matrices T) and $A = PBP^{-1}$, show that $g(A) = Pg(B)P^{-1}$. [*Hint:* $\alpha_2 A^2 = \alpha_2 (PBP^{-1})^2 = \alpha_2 PB(P^{-1}P)BP^{-1} = \alpha_2 PB^2P^{-1}$.]
- b. If D is a diagonal matrix (that is, $d_{ij} = 0$ if $i \neq j$), show that $g(D)$ is a diagonal matrix whose ii^{th} entry is $g(d_{ii})$.
4. Prove that the determinant and trace of a matrix are similarity invariants.
5. a. Given α , β , and γ , suppose that $\Lambda = \{X : \alpha X^2 + \beta X + \gamma I = 0\}$. Show that Λ is similarity invariant.
- b. Are there α , β , and γ such that $\Lambda = \emptyset$?
6. A matrix A is said to be nilpotent iff $A^\ell = 0$ for some ℓ .
- a. Show that nilpotency is a similarity invariant.

- b. The first $\ell > 0$ such that $A^\ell = 0$ is called the *index of nilpotency of A*. Show that this index is also a similarity invariant.
7. A is *idempotent* iff $A^2 = A$. Show that idempotence is a similarity invariant.
8. Show that the minimal polynomial is a similarity invariant. (Suggestion: The minimal polynomial of A divides every polynomial which has the matrix A for a root.) (See Sec. 17 for a review of minimal polynomial.)
4. HOW IS SIMILARITY USED IN SOLVING PROBLEMS?

The Similarity Method

There is a strategy for solving problems that is used frequently enough to deserve some special attention. We'll call this technique "the similarity method." Here is an outline of the method:

- Step 1. Choose a matrix B similar to A for which the problem is easier to solve.
- Step 2. Solve the problem using the matrix B instead of A (the B-problem).
- Step 3. Interpret the solution to the B-problem in terms of the matrix A.

Example: Given $A = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$, find (each entry in) A^{1010} .

- Step 1. (Choose easier $B \sim A$)

Choose $B = \begin{bmatrix} 5 & 0 \\ 0 & -2 \end{bmatrix}$, where $P = \begin{bmatrix} 1 & -4 \\ 1 & 3 \end{bmatrix}$; then $A = PBP^{-1}$.

Step 2. (Solve B-problem)

$$B^{1010} = \begin{bmatrix} 5^{1010} & 0 \\ 0 & (-2)^{1010} \end{bmatrix}$$

Step 3. (Interpret B-solution)

$$A^{1010} = (BP^{-1})^{1010}$$

$$= PB^{1010}P^{-1} \quad (\text{see Exercise 3})$$

$$\text{Thus } A^{1010} = \frac{1}{7} \begin{bmatrix} 3(5^{1010}) + 2^{1012} & 4(5^{1010}) - 2^{1012} \\ 3(5^{1010}) - 3(2^{1010}) & 4(5^{1010}) + 3(2^{1010}) \end{bmatrix}$$

As you have probably observed already, the first obstacle to using the similarity method is that it may not be clear which B to choose. Indeed, there may be no B similar to A which makes the problem easier. Nevertheless, there are many problems which can be attacked successfully by this method. This is particularly true when A is "diagonalable."

5. DIAGONALABLE MATRICES

When a matrix is *diagonalable* (i.e., when it is similar to a diagonal matrix), then the similarity method frequently works, since choosing a diagonal matrix B in Step 1 will probably make Step 2 easy because diagonal matrices are so simple. Step 3 may sometimes turn out to be difficult if not impossible (see Sec. 7 for an illustration), but the previous example and the following example are instances where Step 3 can be carried out easily.

Example: In certain problems in economics the state of a system is described by a matrix $S_n = I + A + A^2 + \dots + A^n$ at time n where A is a given matrix. (We'll go into the background of this problem in more detail later on.) We are interested in the system "in the long run," i.e., when n is large. Suppose $A = \begin{bmatrix} 0.1 & 0.7 \\ 0.3 & 0.6 \end{bmatrix}$. Fortunately, as in the first example, A is diagonalizable because it is a 2×2 matrix having two distinct eigenvalues.[†] Calling these values λ and μ we know that $A = P \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix} P^{-1}$ for some matrix $P^{\dagger\dagger}$. (You should be able to calculate λ , μ , and P . [see e.g., Lipschutz (1968), pp. 198-200].)

Step 1. (Choose $B \sim A$)

So we choose $B = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$.

Step 2. (Solve B -problem)

Let $T_n = I + B + B^2 + \dots + B^n$, then

$$T_n = \begin{bmatrix} 1 + \lambda + \lambda^2 + \dots + \lambda^n & 0 \\ 0 & 1 + \mu + \mu^2 + \dots + \mu^n \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1 + \lambda^{n+1}}{1 - \lambda} & 0 \\ 0 & \frac{1 - \mu^{n+1}}{1 - \mu} \end{bmatrix} \quad \text{for all } n \geq 0$$

Since $|\lambda| < 1$ and $|\mu| < 1$, it follows that T_n

[†] You may recall that a $k \times k$ matrix is diagonalizable if it has k distinct eigenvalues. The converse isn't true.

^{††} Take P to be the matrix whose j^{th} column is an eigenvector for the j^{th} eigenvalue.

approximates $\begin{bmatrix} 1 & 0 \\ 1-\lambda & 1 \\ 0 & 1-\mu \end{bmatrix}$ when n is large. Call

this latter matrix T .

Step 3. (Interpret B -solution)

Since $S_n = PT_nP^{-1}$, it follows that S_n is approximately PTP^{-1} for large enough n (how good an approximation it is depends on how large n is) and so even without computing P we can predict that in the long run the system will "reach a steady state," i.e., S_n isn't appreciably different from S_m when n and m are sufficiently large. If we want a quantitative statement, we calculate P and can then say how large n has to be for S_n to approximate PTP^{-1} (the steady-state matrix) within, say, two-decimal place accuracy or whatever accuracy is required.

Exercises

9. a. Find λ , μ , and P of the previous example.
- b. How large must n be for each entry in S_n to approximate the corresponding entry in PTP^{-1} to two decimal places (S_n , P , T as in previous example)?
10. Prove that every diagonalizable matrix is similar to its transpose. (Later on we'll be able to drop the diagonalizability requirement.) (Hint: Use the similarity method with B the diagonal matrix similar to A .)

11. If A is any diagonalizable matrix and $g(\tau) = \alpha_0 + \alpha_1\tau + \dots + \alpha_{k-1}\tau^{k-1} + \tau^k$ is the characteristic polynomial of A , show that $g(A) = 0$ (i.e., $\alpha_0 I + \alpha_1 A + \dots + \alpha_{k-1} A^{k-1} + A^k = 0$). (This is the Cayley-Hamilton theorem. It's true even if A isn't diagonalizable; we'll see why later on.) [Hint: If B is a diagonal matrix similar to A , use Exercise 3(b) and the fact that the diagonal entries of B are eigenvalues of A to show that $g(B) = 0$.

12. If $x_0 = 1$, $x_1 = 2$, and $x_n = x_{n-1} + 2x_{n-2}$ for all $n \geq 2$, find x_n as a function of n . (Suggestion: Let

$$A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}, \text{ then } A \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 + 2x_0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

$$A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_2 + 2x_1 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}; \text{ so } A^2 \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}, \text{ and in}$$

$$\text{general, } A^n \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} x_n \\ x_{n+1} \end{bmatrix} \text{ for all } n.) \text{ (See Exercise$$

14 for a shortcut to avoid computing P .)

13. If $B_n = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$, find $\det(B_n)$.

[Suggestion: Expand $\det(B_n)$ by minors of the first column to obtain a relation between $\det(B_n)$, $\det(B_{n-1})$, and $\det(B_{n-2})$, then apply the technique of Exercise 12.]

14. If A is diagonalizable and $A^n[x_0, \dots, x_{k-1}]^{\text{tr}} = [x_n, \dots, x_{n+k-1}]^{\text{tr}}$, show that there are scalars $\beta_1, \beta_2, \dots, \beta_\ell$ (independent of n) such that $x_n = \sum_{j=1}^{\ell} \beta_j \lambda_j^n$, where $\lambda_1, \lambda_2, \dots, \lambda_\ell$ are the distinct eigenvalues of A . (This observation enables you to solve problems such as 12 and 13 without finding a matrix P which diagonalizes A as you need only determine the β_i by using the first ℓ values of x_n .)
15. If A is diagonalizable, find necessary and sufficient conditions on the eigenvalues of A for A to be idempotent (see Exercise 7).

6. JORDAN MATRICES

Unfortunately, not all matrices are diagonalizable. [For example, the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ isn't diagonalizable, for if it were, then $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \sim \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$ for some λ, μ and λ, μ must both be 1. (Why?) Therefore, $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$; hence $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (Why?) which would imply that $1 = 0$.]

However, there are many questions about arbitrary matrices A which can be treated by the similarity method if in Step 1 we choose a Jordan matrix B which is similar to A . These matrices are almost as easy to deal with as diagonal matrices.

Some Preliminary Definitions

A *Jordan block* $J_n(\lambda)$ is an $n \times n$ matrix:

1. Each of whose diagonal entries is λ
2. Each of whose superdiagonal entries is 1
3. Each of whose other entries is zero

Thus $J_4(2) = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

and $J_1(2) = 2$ (or $[2]$ if you are a purist).

The next idea we need is the direct sum of matrices:

If A is an $n \times n$ matrix and B is an $m \times m$ matrix then the *direct sum* $A \oplus B$ is the $(n + m) \times (n + m)$ matrix C for which

$$c_{ij} = a_{ij} \quad \text{for all } 1 \leq i, j \leq n$$

$$c_{ij} = b_{ij} \quad \text{for all } n + 1 \leq i, j \leq n + m$$

$$c_{ij} = 0 \quad \text{for all other } i, j$$

Thus $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \oplus \begin{bmatrix} 1 & 0 & 2 \\ 1 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ and

$$2 \oplus \begin{bmatrix} 1 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 4 \end{bmatrix}$$

We extend this definition to cover three summands by defining $A \oplus B \oplus C = (A \oplus B) \oplus C$. In general, we extend it inductively to cover n summands by defining

$$A_1 \oplus A_2 \oplus \cdots \oplus A_{n+1} = (A_1 \oplus A_2 \oplus \cdots \oplus A_n) \oplus A_{n+1}$$

To save space we'll write $\bigoplus_{i=1}^n A_i$ for $A_1 \oplus A_2 \oplus \cdots \oplus A_n$.

Thus $\bigoplus_{i=1}^3 [i] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, $\bigoplus_{i=1}^n [1]$ is the $n \times n$ identity

matrix, and

$$\bigoplus_{i=1}^3 \begin{bmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{bmatrix} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 1 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_3 & 1 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{bmatrix}$$

Exercises

16. Suppose $g(T) = \sum_{i=0}^n \alpha_i T^i$, for all $k \times k$ matrices T and

for all $k \geq 1$. Show that $g\left(\bigoplus_{j=1}^m T_j\right) = \bigoplus_{j=1}^m g(T_j)$.

17. Show that $\bigoplus_{j=1}^m T_j$ is nilpotent iff each T_j is nilpotent.

(See Exercise 6.)

18. Show that $\bigoplus_{j=1}^m T_j$ is idempotent iff each T_j is

idempotent. (See Exercise 7.)

We say that J is a *Jordan matrix* iff J is a direct sum of Jordan blocks. Thus

$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \oplus 2 \oplus \begin{bmatrix} 4 & 1 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

$$= J_2(2) \oplus J_1(2) \oplus J_3(4)$$

is a Jordan matrix. Another way to write this one is

$\bigoplus_{i=1}^3 J_{n_i}(\lambda_i)$, where $n_1 = 2$, $n_2 = 1$, $n_3 = 3$ and $\lambda_1 = 2$,

$\lambda_2 = 2$, $\lambda_3 = 4$. In general, J is a Jordan matrix iff

$J = \bigoplus_{i=1}^m J_{n_i}(\lambda_i)$. If J is $k \times k$ then $\sum_{i=1}^m n_i = k$. It's also

easy to see that the characteristic polynomial of J is

$(\tau - \lambda_1)^{n_1} (\tau - \lambda_2)^{n_2} \cdots (\tau - \lambda_m)^{n_m}$. [In our example, the

polynomial is $(\tau - 2)^2(\tau - 2)^1(\tau - 4)^3$.] Thus the

multiplicity of the eigenvalue (i.e., the maximum n such that $(\tau - \lambda)^n$ is a factor of the characteristic polynomial) is the sum of those n_i for which $\lambda_i = \lambda$. (In our example, the multiplicity of the eigenvalue 2 is three and $n_1 + n_2 = 2 + 1 = 3$.)

7. USING THE SIMILARITY METHOD ON ARBITRARY MATRICES: THE CAYLEY-HAMILTON THEOREM, EVALUATION OF MATRIX POLYNOMIALS

We said at the outset of Sec. 6 that Jordan matrices are almost as easy to use as diagonal matrices and that many problems concerning matrices can be attacked by the similarity method if we choose a Jordan matrix similar to A at Step 1.

Having seen that not all matrices are diagonalizable:

1. How can we tell whether there is a Jordan matrix similar to A to choose? There are tests for diagonalizability (e.g., if all eigenvalues of A have multiplicity one, then A is diagonalizable).
2. What are the tests for "Jordanability"?

The answer to question 1 was given by Camille Jordan; it neatly disposes of question (2) also.

Jordan's Theorem.[†] Every matrix is similar to some Jordan matrix.

So we see that we can always choose a Jordan matrix (it will be diagonal if A were diagonalizable) at Step 1, and often the "B-problem" can be solved for these matrices. But, as we mentioned in Sec. 4, we may still encounter an insurmountable obstacle at Step 3. For example: Is it true that $\max_{1 \leq i \leq k} \left(\sum_{j=1}^k |a_{ij}| \right) \leq |\lambda|$ for all eigenvalues λ when A is an arbitrary $k \times k$ matrix?

If we choose B to be a Jordan matrix, then the question is answered easily for the matrix B (the answer is yes). But you will find that it is not so easy to perform the third step of the similarity method in this case. This is mainly because the function $f(A) =$

$$\max_{1 \leq i \leq k} \sum_{j=1}^k |a_{ij}|$$

is not similarity invariant. Although

the answer to the original question turns out to be "yes," we need to use a technique other than the similarity method to prove it because of this difficulty with Step 3.

To summarize: Jordan's Theorem enables us to get as far as Step 1 of the similarity method using a Jordan matrix for B no matter what matrix A we begin with. Usually Step 2 is easy (because Jordan matrices have so

[†] For a proof see e.g., Hoffman and Kunze (1961, Secs. 7.2 and 7.3).