TRANSFORM

LINEAR ALGEBRA

FRANK UHLIG

[Web Chapters and Solutions]

Contents

No diagonals

Chapter 14

Nondiagonalizable Matrices, the Jordan Normal Form

We study sparse matrix representations for square matrices A that cannot be diagonalized under similarity.

14.1 Lecture Fourteen (The Jordan Normal Form)

We develop the Jordan normal form for matrices A_{nn} that do not have a complete eigenvector basis.

This chapter extends Chapter 9. Both versions of Lecture 9 in Sections 9.1 and 9.1.D terminated with a subsection on diagonalizable matrices. We recall that a square matrix $A \in \mathbb{R}^{n,n}$ (or $\mathbb{C}^{n,n}$) is diagonalizable by matrix similarity $D = X^{-1}AX = \text{diag}(\lambda_i)$ if and only if there is a basis of eigenvectors in $\mathbb{R}^n$ (or $\mathbb{C}^n$) for A that appears in the columns of X. See specifically equation (7.4) and Theorems 9.4 and 9.2.D in Sections 7.1, 9.1, and 9.1.D, respectively.
If A is diagonalizable and $D = X^{-1}AX$ is diagonal, then A's eigenvalues λ_i occur on the diagonal of D and its eigenvectors x_i are the columns of X.
If A is not diagonalizable, then A cannot have a complete eigenvector basis in $\mathbb{R}^n$ (or $\mathbb{C}^n$). Recalling the notion of eigenspace $E(\lambda_i) = \ker(A - \lambda_i I)$ of A for each distinct eigenvalue $\lambda_i,\ i = 1, ..., k \leq n$, we observe that for a **non diagonalizable matrix** A the join of its eigenspaces cannot encompass the whole space, i.e., with the subspace join notation from Section 4.2, we must have

$$E(\lambda_1) \ + \ ... \ + \ E(\lambda_k) \ \subsetneqq \ \mathbb{R}^n \ \ (\text{or } \mathbb{C}^n).$$

According to Section 9.1, a non diagonalizable matrix A has a minimal polynomial of the form $p_A(x) = \prod_{i=1}^{k} (x - \lambda_i)^{j_i}$ with at least one eigenvalue λ_i of index, or exponent $j_i > 1$. Likewise according to Theorem 9.8, a non diagonalizable matrix A must have at least one eigenvalue λ_i of differing geometric and algebraic multiplicity, namely

$$\dim(E(\lambda_i)) \ = \ \text{geom. mult.}(\lambda_i) \ < \ \text{alg. mult.}(\lambda_i) \ = \ n_i \ ,$$

where $f_A(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{n_i}$ is the characteristic polynomial $\det(\lambda I - A)$ of A and the $\lambda_i \in \mathbb{R}$ or $\mathbb{C}$ are distinct. According to the Corollary that is common to both, the end of Sections 9.1 and 9.1.D, to study non diagonalizable matrices A_{nn} thus means to study matrices with at least one repeated eigenvalue.

If a matrix A is not diagonalizable, and if λ is one eigenvalue of A with insufficiently many eigenvectors, with $\text{index}(\lambda) > 1$, or equivalently with geom. mult.$(\lambda) <$ alg. mult.(λ), then $\ker(A - \lambda I)^2 \supsetneqq \ker(A - \lambda I) = E(\lambda)$ as we shall shortly see. For simplicity, we start this section with matrices A that have only one eigenvalue λ.

(a) *A matrix A with only one eigenvalue λ*

If A is diagonalizable and has only one eigenvalue λ, then $X^{-1}AX = \lambda I$, or $A = \lambda I$.
For general, not necessarily diagonalizable matrices A we study a generalization of matrix eigenspaces, namely the **principal subspaces**.

Definition 1: For an eigenvalue $\lambda \in \mathbb{R}$ or $\mathbb{C}$ of $A_{n,n}$, the subspace $P_k(\lambda) := \ker(A - \lambda I)^k$, $k = 0, 1, 2, \ldots$, of $\mathbb{R}^n$ or $\mathbb{C}^n$ is called the **principal subspace** of order k for A and λ. ◀

Clearly $P_0(\lambda) = \ker(A - \lambda I)^0 = \ker(I) = \{0\}$ while $P_1(\lambda) = \ker(A - \lambda I) = E(\lambda)$ is the eigenspace of λ for A. Note that in general, i.e., for diagonalizable and non diagonalizable matrices alike, each principal subspace $P_k(\lambda)$ is a matrix kernel, i.e., according to Chapter 4 it is a subspace of $\mathbb{R}^n$ (or of $\mathbb{C}^n$ if $\lambda \notin \mathbb{R}$). Moreover $P_1(\lambda) \subset P_2(\lambda) \subset \ldots$ is an ascending chain of subspaces of $\mathbb{R}^n$ (or $\mathbb{C}^n$) for each eigenvalue λ of A. This follows since $y \in P_k(\lambda)$ implies $(A - \lambda I)^k y = 0$ and thus $(A - \lambda I)^{k+\ell} y = (A - \lambda I)^\ell \left((A - \lambda I)^k y\right) = (A - \lambda I)^\ell 0 = 0$, or $y \in P_{k+\ell}(\lambda)$ for each $\ell \geq 0$. More is true about the vectors in a principal subspace chain $P_1(\lambda) \subset P_2(\lambda) \subset \ \ldots$ of A.

Lemma 1: Let A be an arbitrary square matrix with the eigenvalue λ. If for some order k the principal subspaces $P_k(\lambda) = P_{k+1}(\lambda)$ are equal, then $P_{k+\ell}(\lambda) = P_k(\lambda)$ for all $\ell \geq 0$.
In other words, if the sequence of principal subspaces $P_i(\lambda)$ becomes stationary for one order $i = k$, then it remains so for all larger orders $k + \ell$. ◀

Proof: We have already noticed that $P_k(\lambda) \subset P_{k+\ell}(\lambda)$ for all k and $\ell \geq 0$. Hence we only need to show that $P_{k+\ell}(\lambda) \subset P_k(\lambda)$ in order to ascertain equality, given that

$P_k(\lambda) = P_{k+1}(\lambda)$.
If $x \in P_{k+\ell}(\lambda)$ for $\ell > 1$, then $(A-\lambda I)^{k+\ell}x = 0$, or $(A-\lambda I)^{k+1}\left((A-\lambda I)^{\ell-1}x\right) = 0$. Thus $(A-\lambda I)^{\ell-1}x \in P_{k+1}(\lambda) = P_k(\lambda)$ as assumed. Thus $(A-\lambda I)^k\left((A-\lambda I)^{\ell-1}x\right) = (A-\lambda I)^{k+\ell-1}x = 0$ as well, or $x \in P_{k+\ell-1}(\lambda)$. By repeating this order reduction, we eventually obtain that $x \in P_{k+1}(\lambda) = P_k(\lambda)$. I.e., $P_{k+\ell}(\lambda) \subset P_k(\lambda)$ for all $\ell \geq 0$, provided that $P_k(\lambda) = P_{k+1}(\lambda)$. ■

We specialize this result now to a non diagonalizable matrix A with only one eigenvalue λ. Such a matrix has a minimal polynomial of the form $p_A(x) = (x-\lambda)^j$ for an exponent $2 \leq j \leq n$ according to Section 9.1. But the minimality of p_A together with the fact that $p_A(A) = (A-\lambda I)^j = O_n$ make $P_j(\lambda) = \mathbb{R}^n$ (or $\mathbb{C}^n$) while $P_{j-1}(\lambda) \subsetneqq P_j(\lambda)$. Thus we have the following relations among the principal subspaces for an eigenvalue λ of precise index $j > 1$:

$$\{0\} = P_0(\lambda) \subsetneqq E(\lambda) = P_1(\lambda) = \ker(A-\lambda I) \subsetneqq P_2(\lambda) = \ker(A-\lambda I)^2 \subsetneqq \ldots.$$
$$\ldots. \subsetneqq P_{j-1}(\lambda) = \ker(A-\lambda I)^{j-1} \subsetneqq P_j(\lambda) = \ker(A-\lambda I)^j = \mathbb{R}^n \quad (\text{or } \mathbb{C}^n)\ .$$

Of particular interest for us are those vectors that lie in one principal subspace $P_\ell(\lambda)$, but do not lie in $P_{\ell-1}(\lambda)$. Since $P_{\ell-1}(\lambda) \subsetneqq P_\ell(\lambda)$, such vectors always exist as long as $1 \leq \ell \leq j$, where j is the index of λ.

Definition 2: The vectors $x \in P_\ell(\lambda)$ with $x \notin P_{\ell-1}(\lambda)$ are called **principal vectors** of order ℓ for A and λ. ◀

Note that not all vectors in the principal vector subspace $P_k(\lambda)$ of A and λ are principal vectors of order k. Clearly $0 \in P_k(\lambda)$ has order zero and any principal vector for λ of order less than k will be annihilated by $(A-\lambda I)^k$, i.e., any lower order principal vector belongs to $P_k(\lambda)$ by default. Principal vectors of order one are customarily called eigenvectors for λ. Principal vectors of arbitrary order help us find sparse and revealing matrix normal forms for non diagonalizable matrices A. Note that principal vectors exist for each eigenvalue λ of a square matrix A up to and including the order equal to the index of λ.

In Chapter 9 we have seen that eigenvectors corresponding to distinct eigenvectors of one matrix are linearly independent. In this chapter we show that the same is true for carefully chosen **chains of principal vectors**. Moreover we learn how these principal vector chains help us complete the incomplete eigenvector set for a non diagonalizable matrix to a full basis of $\mathbb{R}^n$ or $\mathbb{C}^n$. And thereby we will achieve a non diagonal sparse normal form for such matrices under similarity.

Definition 3: If $x^{(k)}$ is a principal vector of order $k \geq 1$ for an arbitrary square matrix A and its eigenvalue λ, then the k vectors $x^{(k)}$, $x^{(k-1)} := (A-\lambda I)x^{(k)} \in P_{k-1}(\lambda)$, ... , and $x^{(1)} := (A-\lambda I)^{k-1}x^{(k)} = (A-\lambda I)x^{(2)} \in P_1(\lambda) = E(\lambda)$ form a **principal vector chain** for A and λ of length k. ◀

Lemma 2: The vectors $x^{(k)}$, $x^{(k-1)}$, ... $x^{(1)}$ of a principal vector chain are linearly independent. ◀

Proof: Clearly if $x^{(k)}$ is a principal vector of order k, then $(A-\lambda I)^{k-\ell}(A-\lambda I)^{\ell}x^{(k)} = (A-\lambda I)^k x^{(k)} = 0$ and therefore $(A-\lambda I)^{\ell}x^{(k)} \in P_{k-\ell}(\lambda)$ for each $k-1 \geq \ell \geq 0$. To show linear independence we use the linear independence Definition 2 of Section 5.1: Assume that $\alpha x^{(k)} + \alpha_{k-1}x^{(k-1)} + \ \dots \ + \alpha_1 x^{(1)} = 0$. Then

$$-\alpha_k x^{(k)} \ = \ \alpha_{k-1}x^{(k-1)} \ + \ \dots \ + \ \alpha_1 x^{(1)} \ . \tag{14.1}$$

Since $x^{(k)}$ is a principal vector for A and its eigenvalue λ of order k by assumption, the left hand vector $-\alpha_k x^{(k)}$ in (14.1) does not lie in $P_{k-1}(\lambda) = \ker(A-\lambda I)^{k-1}$ unless $\alpha_k = 0$. On the other hand, the right hand side linear combination in (14.1) does lie in $P_{k-1}(\lambda)$. Thus $\alpha_k = 0$ and we are left to analyze the shortened equation $\alpha_{k-1}x^{(k-1)} + \ \dots \ + \alpha_1 x^{(1)} = 0$, or $-\alpha_{k-1}x^{(k-1)} = \alpha_{k-2}x^{(k-2)} + \ \dots \ + \alpha_1 x^{(1)}$. Again, the left hand side vector does not belong to $P_{k-2}(\lambda)$ unless $\alpha_{k-1} = 0$, while the right hand side does, implying that $\alpha_{k-1} = 0$. By repeating this argument we conclude that $\alpha_k = \alpha_{k-1} = \ \dots \ = \alpha_2 = \alpha_1 = 0$, i.e., linear independence of the chain of specially chosen principal vectors $x^{(k)}, \ \dots \ , x^{(1)}$ for A and λ. ■

If we define the integers $d_k := \dim P_k(\lambda) - \dim P_{k-1}(\lambda) \geq 0$ and $p_k := \dim P_k(\lambda)$, then for each k there is a basis $\{x_1, ..., x_{p_{k-1}}\}$ for $P_{k-1}(\lambda)$ and a basis $\{x_1, ..., x_{p_k}\}$ for $P_k(\lambda)$ whose first p_{k-1} members form a basis for the principal subspace $P_{k-1}(\lambda)$. Similar to Lemma 2, the collection of all $d_k \cdot k$ principal vectors that originate from the d_k linearly independent principal vectors $x_{p_{k-1}+1}, x_{p_{k-1}+2}, .., x_{p_k}$ of order k for A and λ are linearly independent. The key argument in proving this is the obvious assertion that any linear combination of the linearly independent principal vectors $x_{p_{k-1}+1}, x_{p_{k-1}+2}, .., x_{p_k} \in P_k(\lambda) \setminus P_{k-1}(\lambda)$ of order k which lies in $P_{k-1}(\lambda)$, or has the order $k-1$ or less as a principal vector, must be the zero vector.

Lemma 3: If $y_1, ..., y_{p_{k-1}}$ is a basis for the principal subspace $P_{k-1}(\lambda)$ of order $k-1$ for A and an eigenvalue λ of order at least k, and if $y_1, ..., y_{p_{k-1}}, ..., y_{p_k}$ is a basis for $P_k(\lambda)$, then the $d_k \cdot k$ vectors $(A-\lambda I)^{\ell} y_i$, $i = p_{k-1}+1, ..., p_k$, with $d_k = \dim P_k(\lambda) - \dim P_{k-1}(\lambda)$ and $\ell = 0, ..., k-1$ are linearly independent. ◀

Proof: Assume that for certain coefficients α_i, β_i, ..., δ_i the linear combination

$$\sum_{i=p_{k-1}+1}^{p_k} \alpha_i y_i + (A-\lambda I) \sum_{i=p_{k-1}+1}^{p_k} \beta_i y_i + \ \dots \ + (A-\lambda I)^{k-1} \sum_{i=p_{k-1}+1}^{p_k} \delta_i y_i \ = \ 0 \ . \tag{14.2}$$

This vector equation contains k sum terms on its left. If we multiply (14.2) on the left by $(A-\lambda I)^{k-1}$, then the last $k-1$ sum terms vanish since by construction each y_i is a principal vector of order k for each $p_{k-1} < i \leq p_k$. Thus we are left with the equation $(A-\lambda I)^{k-1}\sum \alpha_i y_i = 0$, which in turn implies $\sum_i \alpha_i y_i = 0$ since $\sum_i \alpha_i y_i \notin P_{k-1}(\lambda)$. As the y_i were chosen linearly independent, all $\alpha_i = 0$. Thus equation (14.2) simplifies to

$$(A-\lambda I)\sum_{i=p_{k-1}+1}^{p_k}\beta_i y_i + \ \dots\ + (A-\lambda I)^{k-1}\sum_{i=p_{k-1}+1}^{p_k}\delta_i y_i \ = \ 0$$

with $k-1$ sum terms. If this equation is multiplied from the left by $(A-\lambda I)^{k-2}$ it yields the simpler equation $(A-\lambda I)^{k-1}\sum_i \beta_i y_i = 0$ since each of the remaining $k-2$ sum terms vanishes because $(A-\lambda I)^k y_i = 0$. As before, we conclude that all $\beta_i = 0$. Continuing in this fashion proves the linear independence of all $d_k \cdot k$ principal vectors $(A-\lambda I)^\ell y_i$ in the d_k distinct principal vector chains for A and λ of index k where $i = p_{k-1}+1, \dots, p_k$ and $\ell = 0, 1, \dots, k-1$. ■

Let us look at the effect that this linear independence of all principal vector chains has for a matrix A with one eigenvalue λ. In particular we want to represent A with respect to each principal vector chain of order k. If $x^{(k)} \in P_k(\lambda)$ and $x^{(k)} \notin P_{k-1}(\lambda)$ is a principal vector of order k for A and λ, then its associated principal vector chain consists of the vectors $x^{(k)}$, $x^{(k-1)} := (A-\lambda I)x^{(k)}$, $x^{(k-2)} := (A-\lambda I)x^{(k-1)}$, ..., $x^{(1)} := (A-\lambda I)x^{(2)}$, where the upper index (ℓ) indicates the order of the principal vector $x^{(\ell)} \in P_\ell(\lambda)$ with $x^{(\ell)} \notin P_{\ell-1}(\lambda)$ throughout. Now look at the matrix product $A_{nn} \cdot \begin{pmatrix} | & | & & | \\ x^{(k)} & x^{(k-1)} & \dots & x^{(1)} \\ | & | & & | \end{pmatrix}_{nk}$.

Clearly $Ax^{(\ell)} = \lambda x^{(\ell)} + x^{(\ell-1)}$ for each $\ell = k, \dots, 2$, while $Ax^{(1)} = x^{(1)}$. This is due to the construction of the principal vector chain, i.e., to our setting $x^{(\ell-1)} := (A-\lambda I)x^{(\ell)}$. Thus

$$A_{nn}\begin{pmatrix} | & & | \\ x^{(k)} & \dots & x^{(1)} \\ | & & | \end{pmatrix}_{nk} = \begin{pmatrix} | & & | & | \\ \lambda x^{(k)} + x^{(k-1)} & \dots & \lambda x^{(2)} + x^{(1)} & \lambda x^{(1)} \\ | & & | & | \end{pmatrix} =$$

$$= \begin{pmatrix} | & & | \\ x^{(k)} & \dots & x^{(1)} \\ | & & | \end{pmatrix}_{nk} \begin{pmatrix} \lambda & 0 & & & 0 \\ 1 & \ddots & & & \\ & \ddots & \ddots & & \\ & & \ddots & \lambda & 0 \\ 0 & & & 1 & \lambda \end{pmatrix}_{kk}.$$

The k by k matrix $\begin{pmatrix} \lambda & 0 & & & 0 \\ 1 & \ddots & & & \\ & \ddots & \ddots & & \\ & & \ddots & \lambda & 0 \\ 0 & & & 1 & \lambda \end{pmatrix}$ is called a (lower) **Jordan block** $J(\lambda, k)$ of size k for the eigenvalue λ. Note that a Jordan block $J(\lambda, k)$ with $k > 1$ cannot be diagonalized: its single eigenvalue λ appears k–fold on the diagonal, yet $\text{rank}(J(\lambda,k) - \lambda I_k) = k-1$, since the lower co-diagonal contains precisely $k-1$ pivot candidates. Thus $J(\lambda, k)$ has precisely one eigenvector for the repeated eigenvalue λ. If the single repeated eigenvalue

λ of A has the index j and if there are $d_j = \dim P_j(\lambda) - \dim P_{j-1}(\lambda)$ linearly independent principal vectors $x_m^{(j)}$ of order j for $m = 1, ..., d_j$, then we set

$$X := \begin{pmatrix} | & & | & & | & & | \\ x_1^{(j)} & \dots & x_1^{(1)} & \dots & x_{d_j}^{(j)} & \dots & x_{d_j}^{(1)} \\ | & & | & & | & & | \end{pmatrix}_{n, d_j \cdot j}$$

and note that

$$A_{nn} X = X \begin{pmatrix} J(\lambda, j) & & \\ & \ddots & \\ & & J(\lambda, j) \end{pmatrix}_{d_j \cdot j, d_j \cdot j} = X \operatorname{diag}(J(\lambda, j)) \ . \tag{14.3}$$

Here $\operatorname{diag}(J(\lambda, j))$ is a block diagonal matrix with d_j diagonal Jordan blocks and $d_j \cdot j \leq n$. If $d_j \cdot j < n$, then for one first lower level r, less than the maximal index level j for λ, the principal subspace $P_r(\lambda)$ contains more linearly independent principal vectors of order r than those generated from the maximal order principal vectors $x_1^{(j)}, ..., x_{d_j}^{(j)}$ through repeated vector iteration with $(A - \lambda I)$. The extra $s := d_r - d_{r+1}$ linearly independent principal vectors of precise order r and their principal vector chains of length r can be appended to X. And for the augmented principal vector matrix $X = X_{n, d_j \cdot j + s \cdot r}$ we have

$$AX = X \left(\operatorname{diag}(J(\lambda, j)),\ \operatorname{diag}(J(\lambda, r)) \right) \ .$$

Here the first diagonal block $\operatorname{diag}(J(\lambda, j))$ contains d_j Jordan blocks for λ of size j according to (14.3) , and the second diagonal block, denoted by $\operatorname{diag}(J(\lambda, r))$ above, contains $s = d_r - d_{r+1}$ Jordan blocks of size r for λ. If we collect all such lower order maximal principal vectors and their principal vector chains in X, then eventually X will be n by n and $AX = XJ$, where J_{nn} is the **Jordan normal form** of the matrix A_{nn} with just one eigenvalue λ. This Jordan normal form consists of various Jordan blocks for λ on its block diagonal. The sizes of the Jordan blocks derive from the dimension differences between $P_k(\lambda)$ and $P_{k-1}(\lambda)$ for each $k = j,\ j-1,\ ....,\ 2$ if the single repeated eigenvalue λ of A has index j. The matrix X that transforms A to its Jordan normal form by similarity contains the vectors of a **Jordan basis** for A in its columns.

Theorem 14.1: (**Jordan Normal Form** (for a matrix with one eigenvalue))

Each matrix $A \in \mathbb{R}^{n,n}$ (or $\mathbb{C}^{n,n}$) with a single eigenvalue λ is similar over $\mathbb{R}$ (or $\mathbb{C}$, if $\lambda \notin \mathbb{R}$) to a block diagonal matrix $J = \operatorname{diag}\left(J, \lambda_{n_{s_i}}\right)$ of Jordan blocks $J(\lambda, n_{s_i})$. Here the dimension n_{s_i} are equal to the maximal orders of principal vectors of A for λ that form a basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$), called the **Jordan basis** of A.

In other words, the various dimensions n_{s_i} of Jordan blocks in the Jordan normal form of A are equal to the lengths of all complete length principal vector chains for A and λ.

The matrix J is called the **Jordan normal form** of A. ◀

The Jordan form of any matrix is generally not unique since a rearrangement of the principal vector chains in X rearranges the order of the individual Jordan blocks in the Jordan normal form J of A. However, the list of Jordan block dimensions $\{n_{s_i}\}$ is uniquely determined by A.

Example 1: Assume that a matrix $A_{14,14}$ has the single repeated eigenvalue λ. Assume further that the dimensions of the principal subspaces for A and λ are given as $p_1 = \dim E(\lambda) = \dim P_1(\lambda) = 5$, $p_2 = \dim P_2(\lambda) = 9$, $p_3 = 12$, and $p_4 = p_5 = 14$. Thus λ has index 4. By Theorem 14.1, our matrix A is similar to a block diagonal matrix with diagonal Jordan blocks for λ of sizes 4 by 4 or smaller for λ.
Which Jordan block sizes do occur here, given the above data?
Since $d_4 := p_4 - p_3 = 14 - 12 = 2$, there are two linearly independent principal vector chains of maximal length 4 for A and λ. These result in 2 copies of $J(\lambda, 4)$ occuring in the Jordan normal form J of A. Next, $d_3 := p_3 - p_2 = 12 - 9 = 3$ gives A three linearly independent principal vectors of order 3. Two of these are accounted for inside the two principal vector chains of order 4. Thus the third linearly independent principal vector of order 3 gives rise to one Jordan block $J(\lambda, 3)$ of size 3 in J. Since $d_2 := p_2 - p_1 = 9 - 5 = 4$, there is one additional Jordan block $J(\lambda, 2)$ of size 2 by 2 in J, because only three order 2 principal vectors are associated with the three larger dimensional Jordan blocks $J(\lambda, 4)$ (twice) and $J(\lambda, 3)$. Finally $d_1 := p_1 = 5$ counts the number of linearly independent eigenvectors for λ. Four of these occur in the two principal vector chains of length 4, in the one of length 3, and in the one of length 2. Thus there is one additional 1 by 1 Jordan block in J. And A is similar to its Jordan normal form

$$J_{14,14} \quad = \quad \text{diag}(J(\lambda,4),\ J(\lambda,4),\ J(\lambda,3),\ J(\lambda,2),\ J(\lambda,1)) \ .$$

◄

For one eigenvalue λ the set of dimensions of its asociated Jordan blocks can be illustrated in **diagram** form.

Example 2: We display the linearly independent principal vector chains from the previous example by the following dot diagram:

								no. of dots:
In $P_4(\lambda)$:	2 l. i. principal vectors of order 4 :	•	•					$d_4 = 2$ dots
In $P_3(\lambda)$:	3 l. i. principal vectors of order 3 :	•	•	•				$d_3 = 3$ dots
In $P_2(\lambda)$:	4 l. i. principal vectors of order 2 :	•	•	•	•			$d_2 = 4$ dots
In $P_1(\lambda)$:	5 l. i. eigenvectors :	•	•	•	•	•		$d_1 = 5$ dots
	principal vector chain lengths :	4	4	3	2	1		

Here the number of dots in each row is equal to the number of respective $d_{\dots} =$

$p_{\dots} - p_{\dots-1}$, while the number of dots in one column describes the size of a Jordan block in the Jordan normal form of A. And the total number of dots (14) in the diagram equals the size of $A_{14,14}$
In this example we have $5 = p_1$ principal vector chains of lengths 4, 4, 3, 2, and 1 that give rise to Jordan blocks of the same dimensions as their principal vector chain lengths. Note that the sequence $\{d_i\} = \{5,\ 4,\ 3,\ 2\}$ of numbers $d_i = p_i - p_{i-1}$ for $i = 1, \dots, j$ of linearly independent principal vectors of order i is strictly decreasing in our example, namely $d_4 = 2 < d_3 = 3 < d_2 = 4 < d_1 = 5$.
For general matrices A and one eigenvalue λ of index j, we always have $d_i \leq d_{i-1}$ for each $i \leq j$, the index of the eigenvalue of λ. And equality $d_i = d_{i-1}$ holds for $i \leq j$ if there is no Jordan block of dimension $i - 1$ in the Jordan normal form of A. ◀

Example 3: (a) The sequence $p_1 = 4,\ p_2 = 8,\ p_3 = 9$, and $p_4 = p_5 = 10$ of principal subspace dimensions $p_i = \dim P_i(\lambda)$ does occur for a matrix A with a single repeated eigenvalue λ. The d_k dot diagram of Example 2 looks as follows with $d_1 = p_1 = 4,\ d_2 = p_2 - p_1 = 4,\ d_3 = p_3 - p_2 = 1,\ d_4 = p_4 - p_3 = 1$:

					the d_i:
order 4 :	•				$d_4 = 1$
order 3 :	•				$d_3 = 1$
order 2 :	•	•	•	•	$d_2 = 4$
order 1 :	•	•	•	•	$d_1 = 4$
Jordan block sizes :	4	2	2	2	

This tableau with 10 dots gives rise to the Jordan normal form

$$J = \text{diag}(J(\lambda, 4), J(\lambda, 2), J(\lambda, 2), J(\lambda, 2))$$

of size 10 by 10 for A. Note that the Jordan normal form for A contains no Jordan blocks of size 3 since $d_3 = d_4 = 1$, nor does it contain any of size 1 since $d_1 = d_2 = 4$.

(b) The sequence $p_1 = 3,\ p_2 = 7,\ p_3 = 9$, and $p_4 = p_5 = 10$ is not a legitimate principal subspace dimension sequence for any matrix A with one eigenvalue λ. To see this, look at the difference sequence $d_i = p_i - p_{i-1}$. Namely $d_1 = 3 < d_2 = 4 > d_3 = 2 > d_4 = 1$. Clearly the inequality $d_i \geq d_{i+1}$ is not satisfied for $i = 1$. This contradicts the fact that the number $d_2 = 4$ of linearly independent principal vectors of level 2 must not be less than the number of linearly independent eigenvectors in $P_1(\lambda) = E(\lambda)$. The d_i Jordan diagram does not look appropriate at all:

					the d_i:
order 4 :	•				$d_4 = 1$
order 3 :	•	•			$d_3 = 2$
order 2 :	•	•	•	•	$d_2 = 4$
order 1 :	•	•	•	?	$d_1 = 3$
Jordan block sizes :	4	2	2	?	

◀

Corollary 1: A sequence p_i of principal subspace dimensions of order i is admissible for a matrix A and one eigenvalue λ if and only if the numbers $d_i := p_i - p_{i-1} \geq 1$ satisfy $d_i \leq d_{i-1}$ for all $i = 1, ..., \text{index}(\lambda)$.
Specifically if $d_i < d_{i-1}$ for one $i \leq \text{index}(\lambda)$, then the Jordan normal form of A contains $d_{i-1} - d_i$ Jordan blocks of size $i-1$ by $i-1$ for λ. ◀

Remark 1: According to Chapter 11, every real symmetric matrix $A = A^T$ has a diagonal Jordan normal form for real eigenvalues. In particular, it follows for $A = A^T \in \mathbb{R}^{n,n}$ that $\ker(A - \lambda I) = \ker(A - \lambda I)^k$ for all $k > 1$ and all $\lambda \in \mathbb{R}$. ◀

(b) *A matrix A with several distinct eigenvalues*

The ideas and results of subsection (a) can be applied to matrices with several eigenvalues. If a sparse normal form N can be obtained for A via matrix similarity $X^{-1}AX = N$, then this form can be achieved for every matrix B that is similar to A: In particular, if A and B are two similar matrices, i.e., if $Y^{-1}AY = B$ for a nonsingular matrix Y and if $X^{-1}AX = N$ is a normal form of A, then with $Z := Y^{-1}X$ and $Z^{-1} = X^{-1}Y$ according to Theorem 6.4, we have

$$Z^{-1}BZ \;=\; X^{-1}YBY^{-1}X \;=\; X^{-1}AX \;=\; N \;.$$

Therefore $B = Y^{-1}AY$ is also similar to the normal form N of A.

Thus, when looking for sparse and structure revealing matrix representations of a given matrix A_{nn} under basis change $X^{-1}..X$, we may start with any matrix B that is similar to A. Its normal form will be identical to the one for A. For the multi–eigenvalue and non diagonalizable case, we shall start with the Schur normal form S of A; see Theorem 11.1. S is upper triangular and has some repeated entries on its diagonal since the non diagonalizable matrix A must have repeated eigenvalues. We assume further without loss of generality that the distinct eigenvalues λ_1, ..., λ_k of S are arranged in consecutive diagonal entry clusters as depicted below

$$S \;=\; \begin{pmatrix}
\lambda_1 & * & & & & & & & & & * \\
0 & \ddots & & & & & & & & & \\
& & \lambda_1 & & & & * & & & & \\
& & & \lambda_2 & & & & & & & \\
& & & & \ddots & & & & & & \\
& & & & & \lambda_2 & & & & & \\
& & & & & & \ddots & & & * & \\
& & & & & & & \ddots & & & \\
& & & & & & & & \lambda_k & & \\
& & & & & & & & & \ddots & * \\
0 & & & & & & & & & 0 & \lambda_k
\end{pmatrix}$$

with $\lambda_i \neq \lambda_j$ for $i \neq j$. Note that this Schur form can be partially achieved by choosing $\lambda_2 = \lambda_1$ in the second step of the proof of Theorem 11.1 in case λ_1 is repeated, and it is completely achieved by further repetition. Our main task now is to use matrix similarity to reduce the Schur normal form S of A with its repeated eigenvalues appearing in clusters on its diagonal further, namely to a block diagonal matrix of the form

$$\widehat{S} = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & 0 & 0 \\ & \boxed{\begin{matrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2 \end{matrix}} & 0 \\ & & \ddots \\ 0 & & & \ddots \end{pmatrix}$$

with zeros in all of its upper off-diagonal blocks.

If A_{nn} has only two distinct eigenvalues $\lambda_1 \neq \lambda_2$ of multiplicities $r > 1$ and $s \geq 1$, respectively, then $r + s = n$ and its Schur normal form can be written as

$$S = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & * \\ 0 & \boxed{\begin{matrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2 \end{matrix}} \end{pmatrix},$$

where the first diagonal block is r by r, and the second one is s by s. We shall perform matrix similarities on S using Gaussian elimination matrices $X_{i,j}(\alpha) := (\alpha I_n + \alpha E_{i,j})$ of the form (6.3) for indices $1 \leq i \leq r$ and $r + 1 \leq j \leq r + s = n$ and reduce S to a matrix of type $\widehat{S}$ with zeros in its (1,2) block. Here $E_{i,j}$ is the elementary n by n matrix with zeros in each of its n^2 positions, except for a 1 in position (i, j). As stated in Lemma 2 of Section 6.2, $(X_{i,j}(\alpha))^{-1} = X_{i,j}(-\alpha)$ for all i, j, and α. We start the process with α in position $i = r$ and $j = r + 1$ and form $(X_{r,r+1}(\alpha))^{-1} S X_{r,r+1}(\alpha) =$

$$\left(\begin{array}{c|c} I_r & \begin{matrix} & 0 \\ -\alpha & \end{matrix} \\ \hline O_{s,r} & I_s \end{array}\right) \cdot \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & \begin{matrix} & * \\ s_{r,r+1} & \end{matrix} \\ O_{r,s} & \boxed{\begin{matrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2 \end{matrix}} \end{pmatrix} \cdot \left(\begin{array}{c|c} I_r & \begin{matrix} & 0 \\ \alpha & \end{matrix} \\ \hline O_{s,r} & I_s \end{array}\right) =$$

$$
= \left(\begin{array}{c|cc} & & 0 \\ I_r & -\alpha & \\ \hline O_{s,r} & I_s & \end{array}\right) \cdot \left(\begin{array}{ccc} \boxed{\begin{matrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & \begin{matrix} \dagger \\ \dagger \\ \alpha\lambda_1 + s_{r,r+1} \end{matrix} & * \\ O_{s,r} & \boxed{\begin{matrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2 \end{matrix}} & \end{array}\right) =
$$

$$
= \left(\begin{array}{ccc} \boxed{\begin{matrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & \begin{matrix} \dagger \\ \dagger \\ \alpha\lambda_1 + s_{r,r+1} - \alpha\lambda_2 \end{matrix} & \begin{matrix} * \\ \\ \ddagger \quad \ddagger \end{matrix} \\ O_{r,s} & \boxed{\begin{matrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2 \end{matrix}} & \end{array}\right) = \widetilde{S}\ .
$$

Here we have formed the product of the two matrix factors S and $X_{r,r+1}(\alpha)$ on the right first by updating the first $r-1$ entries in the $(r+1)^{st}$ column, denoted by $\dagger$, and by updating the $(r, r+1)$ entry $s_{r,r+1}$ of S to become $\alpha\lambda_1 + s_{r,r+1}$. Multiplying the resulting product matrix on the left by $X_{r,r+1}(-\alpha)$ updates the $(r, r+1)$ entry further to $\widetilde{s}_{r,r+1} := \alpha\lambda_1 + s_{r,r+1} - \alpha\lambda_2$ and changes the last $s-1$ entries of the r^{th} row. This update is denoted by $\ddagger$. Note that no other entries of the original Schur form S are affected by the similarity.
If we choose $\alpha := \dfrac{s_{r,r+1}}{\lambda_2 - \lambda_1}$, then the $(r, r+1)$ entry in $\widetilde{S} := (X_{r,r+1}(\alpha))^{-1}\, S X_{r,r+1}(\alpha)$ becomes zero.
Next, if $s > 1$ we repeat this process with an eye on zeroing out the $(r, r+2)$ entry of $\widetilde{S}$ via the matrix similarity $X_{r,r+2}(-\alpha)\ \widetilde{S}\ X_{r,r+2}(\alpha)$ for a new α in position $(r, r+2)$ of X:

$$
\left(\begin{array}{c|ccc} & 0 & & 0 \\ I_r & & & \\ & 0 & -\alpha & 0 \\ \hline O_{s,r} & & I_s & \end{array}\right) \cdot \left(\begin{array}{cc} \boxed{\begin{matrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & \begin{matrix} * & * & * \\ * & * & * \\ 0 & \widetilde{s}_{r,r+2} & * \end{matrix} \\ O_{s,r} & \boxed{\begin{matrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2 \end{matrix}} \end{array}\right) \cdot \left(\begin{array}{c|ccc} & 0 & & 0 \\ I_r & & & \\ & 0 & \alpha & 0 \\ \hline O_{s,r} & & I_s & \end{array}\right) =
$$

$$= \left(\begin{array}{c|ccc} & 0 & & 0 \\ I_r & & & \\ & 0 & -\alpha & 0 \\ \hline O_{s,r} & & I_s & \end{array}\right) \cdot \left(\begin{array}{cc} \boxed{\begin{matrix}\lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1\end{matrix}} & \begin{matrix} * & \dagger & * \\ * & \dagger & * \\ 0 & \alpha\lambda_1 + \widetilde{s}_{r,r+2} & * \end{matrix} \\ O_{s,r} & \boxed{\begin{matrix}\lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2\end{matrix}} \end{array}\right) =$$

$$= \left(\begin{array}{cc} \boxed{\begin{matrix}\lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1\end{matrix}} & \begin{matrix} * & \dagger & * \\ * & \dagger & * \\ 0 & \alpha\lambda_1 + \widetilde{s}_{r,r+2} - \alpha\lambda_2 & \ddagger \end{matrix} \\ O_{s,r} & \boxed{\begin{matrix}\lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2\end{matrix}} \end{array}\right) = \widetilde{\widetilde{S}} \ .$$

Here the only entries in $\widetilde{\widetilde{S}}$ that differ from those in $\widetilde{S}$ lie above and to the right of the $(r, r+2)$ position while the $(r, r+2)$ entry changes to $\widetilde{\widetilde{s}}_{r,r+2} = \alpha\lambda_1 + \widetilde{s}_{r,r+2} - \alpha\lambda_2$. If we pick $\alpha := \dfrac{s_{r,r+2}}{\lambda_2 - \lambda_1}$ in this step, then the $(r, r+2)$ entry in $\widetilde{\widetilde{S}} = (X_{r,r+2}(\alpha))^{-1}\, \widetilde{S} X_{r,r+2}(\alpha)$ becomes zero.

Note the pattern here: We have chosen specific constants α and elementary matrices $X_{i,j}(\alpha)$ as similarity transformations for the Schur normal form S in sequence for $i \leq r$ and $r+1 \leq j \leq r+s$. Namely we start at the lower left position $(r, r+1)$ in the $(1,2)$ block of S with indices so that $i + j = 2r+1$ first, followed by those with $i + j = 2r+2$ from $i = r$ to $i = r-1$, and then by those with $i + j = 2r+3$ from $i = r$ back to $i = r-2$, and so forth. This off–diagonal sweep pattern allows us to clear out the lower triangle of the $(1,2)$ block in S, one co-diagonal at a time, all by elementary similarities. Next we go up the last column of S from position $(r-1, r+s)$ to position $(1, r+s)$ and zero out the entries on their respective co-diagonals. All in all, if we sweep S via such repeated similarities starting with the extreme row positions along the bottom and right edge inside the $(1,2)$ block of S in the column sequence $r+1, r+2, ..., r+s$ in row r followed by the row sequence $r+s-1, r+s-2, ..., 2, 1$ in column $r+s = n$, then all of the co–diagonals (from the bottom left corner to the right and on up) of the $(1,2)$ block in S will be zeroed out for appropriate choices of the respective α since $\lambda_1 \neq \lambda_2$. And thus we achieve the form $\widehat{S}$.

This process readily generalizes to Schur normal forms S of matrices A with several distinct eigenvalues. Combining this with Theorem 14.1 proves our main result:

Theorem 14.2: (**The Jordan Normal Form**)

Every matrix $A_{nn} \in \mathbb{R}^{n,n}$ (or $\mathbb{C}^{n,n}$) is similar over the complex numbers $\mathbb{C}$ to a block diagonal matrix consisting of certain Jordan blocks for its eigenvalues.
In other words, for each A_{nn} there exists a Jordan basis or a nested principal vector basis of $\mathbb{C}^n$ collected in the columns of X_{nn}, so that

$$\begin{aligned} X^{-1}AX &= \operatorname{diag}\left(J(\lambda_1, n_{1,1}), ..., J(\lambda_1, n_{1,\ell_1}), J(\lambda_2, n_{2,1}), ..., J(\lambda_2, n_{2,\ell_2}), ..., \right. \\ &\qquad \left. ..., J(\lambda_k, n_{k,1}), ..., J(\lambda_k, n_{k,\ell_k})\right) \ . \end{aligned}$$

Here $n_{j,i} \geq 1$, $\ell_j \geq 1$, and $\sum_{j=1}^{k} \sum_{i=1}^{\ell_j} n_{j,i} = n$. Each sequence of Jordan block dimensions $n_{m,1}, ..., n_{m,\ell_m}$ is uniquely determined for the eigenvalue $\lambda_m \in \mathbb{C}$ of A.

To prove this theorem, we simply invoke subsection (a) on each of the diagonal Schur blocks $\begin{pmatrix} \lambda_i & & * \\ & \ddots & \\ 0 & & \lambda_i \end{pmatrix}$ that has been obtained via elementary similarities from the Schur normal form S of A in $\widehat{S}$.

Remark 2: The Jordan normal form of a matrix is generally not unique since the order of the eigenvalues can be changed at will, as can the order of the sequence of Jordan block dimensions $n_{j,i}$ for each repeated eigenvalue λ_j of A.
The Jordan normal form of a matrix is generally not diagonal. It contains one or several non zero co–diagonal entries of 1 in general.
The Jordan normal form represents A sparsely with respect to a principal vector chain basis, called the Jordan basis of A.
If A is diagonalizable, then the Jordan normal form of A is diagonal and A has an eigenvector basis, and vice versa. ◀

Example 4: Find the Jordan normal form J for $A = \begin{pmatrix} 1 & 0 & -1 & -1 \\ 1 & 2 & 1 & 1 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 3 \end{pmatrix}$.

Starting with $x = \begin{pmatrix} -2 \\ 1 \\ -1 \\ 2 \end{pmatrix}$ for example, we compute the vector iteration matrix

$$\begin{pmatrix} | & | & & | \\ x & Ax & \dots & A^4x \\ | & | & & | \end{pmatrix} = \begin{pmatrix} -2 & -3 & -4 & -4 & 0 \\ 1 & 1 & 0 & -4 & -16 \\ -1 & -2 & -4 & -8 & -16 \\ 2 & 3 & 4 & 4 & 0 \end{pmatrix}$$

and its RREF as $\begin{pmatrix} 1 & 0 & -4 & -16 & -48 \\ 0 & 1 & 4 & 12 & 32 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$. Thus the minimal polynomial of A contains the factor $p(x) = x^2 - 4x + 4 = (x-2)^2$ that we can read off the first linearly dependent column, the third column, in the RREF. As the minimal polynomial of A contains a quadratic factor, A is not diagonalizable according to Chapter 9.
To find the Jordan block structure of A, we analyze $\ker(A - 2I_4)^2$ next:

Note that $A - 2I = \begin{pmatrix} -1 & 0 & -1 & -1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}$ and $(A-2I)^2 = O_2$. Thus $p(x) = (x - 2)^2 = p_A(x)$ is A's minimal polynomial and $p_2 = \dim P_2(2) = \dim(\ker(A-2I)^2) = 4$ while $p_1 = \dim P_1(2) = \dim E(2) = \dim(\ker(A-2I)) = 3$. This makes $d_2 = p_2 - p_1 = 4 - 3 = 1$ and $d_1 = p_1 = 3$. Thus the Jordan normal form J of A has $1 = d_2$ Jordan block $J(2,2)$ of size two and $2 = d_2 - d_1$ Jordan blocks $J(2,1)$ of size one, all for the single repeated eigenvalue $\lambda = 2$ of A: The Jordan normal form of A is

$$J \; = \; \mathrm{diag}(J(2,2), \; J(2,1), \; J(2,1)) \; .$$

Incidentally, the Jordan normal form of our matrix $A_{4,4}$ could have been deduced by a counting argument alone: $A_{4,4}$ has the single eigenvalue $\lambda = 2$ with geometric multiplicity 3; just look at $A - 2I$ above. Hence there must be precisely one principal vector for $\lambda = 2$ of index 2 in order that the Jordan dot diagram has 4 dots:

order 2 :	$\underline{\bullet}$!		
order 1 :	$\bullet$	$\bullet$	$\bullet$
Jordan block sizes :	2	1	1

Or alternatively, one could argue differently again: The eigenvalue 2 of A has the index 2 since $p_A(x) = (x-2)^2$. Therefore the matrix A of size 4 could possibly have a Jordan normal form consisting of 2 Jordan blocks of size 2 for its single eigenvalue, or its Jordan form could have one Jordan block of size 2 for $\lambda = 2$ and two more 1–dimensional Jordan blocks, all for $\lambda = 2$. The eigenspace $E(2)$ for A has dimension 3; just look at $A - 2I$. Hence the latter must be the Jordan normal form of A, since in the former case A could only have two linearly independent eigenvectors. ◀

(c) *Practicalities*

For ease with developing the Jordan normal form theory, we have so far introduced principal vector chains $x^{(j)},\ x^{(j-1)} = (A - \lambda I)x^{(j)},\ ...$ from the top order principal vector(s) $x^{(j)}$ of A_{nn} on down. This assumes implicitly that we know the index j of an eigenvalue λ of A and the sequence of principal subspace dimensions $p_i = \dim P_i(\lambda)$ beforehand. There is an easier way than to construct all nested kernels $E(\lambda) = P_1(\lambda) \subset P_2(\lambda) \subset ...$ and to compute their dimensions p_i explicitly. We can start instead with an eigenvector basis for an eigenvalue λ of A. From the Jordan dot diagrams, we know that there are precisely $\dim E(\lambda)$ many Jordan blocks for λ in the Jordan normal form of A. What we have to ascertain is their dimensions, or the length of each principal vector chain for λ from the eigenspace generators of A themselves.

If $x^{(1)}$ is a known eigenvector for A and λ, and if $x^{(2)}$ is a solution of the singular linear system $(A - \lambda I)x = x^{(1)}$, then $x^{(2)}$ is a principal vector of order 2 for A and λ. In general, the principal vector chain that has $x^{(1)}$ as its eigenvector can be constructed from $x^{(1)}$ by solving the linear systems $(A - \lambda I)x = x^{(k)}$ successively for $k = 1, 2, ...$, until the linear system becomes inconsistent and thus has no solution. This allows us to find the index of an eigenvalue λ and the length of its principal vector chains, as well as the sizes of its Jordan blocks by linear means from λ. We have to choose the generating eigenvectors $x^{(1)}$ carefully in this process, though, so that we obtain useful principal vector chains. The selection process will become clear in the next example. It also lets us readily determine the similarity transformation matrix X that sends A to its Jordan normal form.

Example 5: We use the matrix $A = \begin{pmatrix} 1 & 0 & -1 & -1 \\ 1 & 2 & 1 & 1 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 3 \end{pmatrix}$ of Example 4 and try to find its principal vector chains from its eigenvectors.

For the eigenvalue $\lambda = 2$ of A we first compute a basis for the kernel of $A - \lambda I$:

$$\ker(A - 2I_4) \ : \quad \begin{array}{rrrr|rl} -1 & 0 & -1 & -1 & 0 & \\ 1 & 0 & 1 & 1 & 0 & +\ row_1 \\ 0 & 0 & 0 & 0 & 0 & \\ 1 & 0 & 1 & 1 & 0 & +\ row_1 \\ \hline -1 & 0 & -1 & -1 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \end{array}$$

Thus the three vectors $x_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $x_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$, and $x_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$ form a basis for the eigenspace for $\lambda = 2$ and A. Note that none of these three eigenvectors

leads to a solvable linear system of the form $(A-2I)x = x_i$ for $i = 1,2,3$. This should be checked by the students.
Next we check whether there is a linear combination $y := \alpha x_1 + \beta x_2 + \gamma x_3 = \begin{pmatrix} \beta+\gamma \\ \alpha \\ -\beta \\ -\gamma \end{pmatrix} \in E(2) = P_1(2)$ of the three eigenvectors for which $(A-2I)x = y$ is solvable.

$$\begin{array}{rrrr|c}
-1 & 0 & -1 & -1 & \beta+\gamma \\
1 & 0 & 1 & 1 & \alpha \\
0 & 0 & 0 & 0 & -\beta \\
1 & 0 & 1 & 1 & -\gamma \\
\hline
-1 & 0 & -1 & -1 & \beta+\gamma \\
0 & 0 & 0 & 0 & \alpha+\beta+\gamma \\
0 & 0 & 0 & 0 & -\beta \\
0 & 0 & 0 & 0 & \beta
\end{array}
\qquad
\begin{array}{l}
\\ +\,row_1 \\ \\ +\,row_1 \\ \\ \\ \\ \\
\end{array}$$

This system is solvable precisely when $\beta = 0$ and $\alpha = -\gamma$. If we set $\alpha := 1$ then $x^{(1)} := x_1 - x_3 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$ is an eigenvector that belongs to a principal vector chain of order at least 2 for A and $\lambda = 2$. The corresponding principal vector $x^{(2)}$ of order 2 solves $(A-2I)x = x^{(1)}$. We find $x^{(2)}$ from the scheme

$$\begin{array}{rrrr|r}
-1 & 0 & -1 & -1 & -1 \\
1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 \\
\hline
-1 & 0 & -1 & -1 & -1 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{array}
\qquad
\begin{array}{l}
\\ +\,row_1 \\ \\ +\,row_1 \\ \\ \\ \\ \\
\end{array}$$

as $x^{(2)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ for example. Now, the next level linear system $(A-2I)x = x^{(2)} = e_1$ is inconsistent, because its first and second row equations are contradictory. Therefore there are no higher order principal vectors than of order 2 for the eigenvalue $\lambda = 2$.

If we write the principal vector matrix X in the form $X = \begin{pmatrix} | & | & | & | \\ x^{(2)} & x^{(1)} & x_1 & x_2 \\ | & | & | & | \end{pmatrix}$ using the first found eigenvectors x_1 and x_2 in columns 3 and 4 for example, then

$$X^{-1}AX = J = \operatorname{diag}(J(2,2), J(1,1), J(1,1)) \ .$$

And we have found an explicit matrix similarity that transforms A to its Jordan normal form J.

◀

We have generalized the concept of diagonalizability of a matrix and have obtained a sparse normal form representation of matrices that are not diagonalizable.

14.1.P Problems

1. Write down a 5 by 5 matrix that is not diagonalizable.

2. Write down a 7 by 7 matrix that is not diagonal but diagonalizable.

3. Find a 12 by 12 triangular matrix that is diagonalizable, and one that is not.

4. Find the Jordan normal form of $\begin{pmatrix} -4 & 4 \\ -1 & 0 \end{pmatrix}$.

5. Find the Jordan normal form of $A = \begin{pmatrix} 0 & 0 & 1 \\ -3 & -1 & -3 \\ -1 & 0 & -2 \end{pmatrix}$, given that -1 is an eigenvalue of A. (Hint: What is the algebraic multiplicity of -1?)

6. Find the Jordan normal form of $B = \begin{pmatrix} -3 & -7 & -3 \\ 2 & 5 & 2 \\ -1 & -3 & -1 \end{pmatrix}$, given that 1 and 0 are eigenvalues of B.

7. Write out the Jordan dot diagram for A in Example 4 and for Problem 5.

8. Decide whether the matrix $A = \begin{pmatrix} -4 & 9 \\ -1 & 2 \end{pmatrix}$ is diagonalizable or not.
   What is its Jordan normal form?

9. Show that the real matrix $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ is similar to the complex matrix $\begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix}$ where $\lambda = a + bi \in \mathbb{C}$.

10. Determine whether the two matrices $A = \begin{pmatrix} 2 & 5 & 0 & -2 \\ -3 & -6 & 0 & 2 \\ -7 & -12 & -1 & 5 \\ -3 & -5 & 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -4 & -2 & 0 \\ -1 & 1 & 1 & 0 \\ -1 & -11 & -4 & 1 \\ 8 & -3 & -4 & -2 \end{pmatrix}$ are similar.
    (Note that both matrices have the 4 fold eigenvalue $\lambda = -1$.)

11. Find the Jordan normal form of $C = \begin{pmatrix} -6 & 12 & -22 \\ 2 & -2 & 5 \\ 3 & -5 & 10 \end{pmatrix}$, given that -1 and 2 are eigenvalues of C. What is the third eigenvalue of C? (Think 'trace'.)

12. Determine whether $A = \begin{pmatrix} 4 & -2 & -1 \\ -6 & 8 & 3 \\ 16 & -16 & -6 \end{pmatrix}$ has the Jordan normal form $J = \begin{pmatrix} 2 & & \\ & 2 & 0 \\ & 1 & 2 \end{pmatrix}$, $K = \begin{pmatrix} 2 & & \\ 1 & 2 & \\ 0 & 1 & 2 \end{pmatrix}$, or neither.
    What other Jordan normal form could A possibly have, given that it has the three fold eigenvalue $\lambda = 2$?

13. What are all possible Jordan diagrams for a matrix A with a single 4 fold real eigenvalue λ.

14. Repeat the previous problem for a matrix B with the 4 fold eigenvalue $\lambda = 4$ of

    (a) index 3;

(b) index 1;

(c) index 4.

15. Write down all possible Jordan block dimensions for a matrix $A_{n,n}$ with a single 6 fold eigenvalue λ.

16. For each of the possible Jordan forms in the previous 3 problems, determine the algebraic and the geometric multiplicities of the eigenvalue λ.

17. If A has one real eigenvalue with geometric multiplicity 3, algebraic multiplicity 8, and index 4, how many differing Jordan block dimensions can there be for this eigenvalue of A? What Jordan diagrams are possible for this data?

18. Repeat the previous problem for an eigenvalue λ with

(a) geometric multiplicity 4, algebraic multiplicity 8, and index 5;

(b) geometric multiplicity 6 and algebraic multiplicity 8;

(c) equal geometric multiplicity, algebraic multiplicity, and index, if possible.

19. Find the possible range of indices of an eigenvalue λ which has algebraic multiplicity 12 and geometric multiplicity 7.

20. Find a matrix $X_{5,5}$ with

$$X \begin{pmatrix} 2 & 1 & & 0 & \\ 0 & 2 & & & \\ & & 1 & 1 & 0 \\ & & 0 & 1 & 1 \\ 0 & & 0 & 0 & 1 \end{pmatrix} X^{-1} =$$

$$= \begin{pmatrix} 1 & 0 & 0 & & 0 \\ 1 & 1 & 0 & & \\ 0 & 1 & 1 & & \\ & & & 2 & 0 \\ 0 & & & 1 & 2 \end{pmatrix}.$$

21. (a) If the eigenvalue λ of A has an algebraic multiplicity exceeding its geometric one by 1, what Jordan blocks are possible for λ and A?

(b) If the eigenvalue λ of A has an algebraic multiplicity two larger than its geometric one, what Jordan blocks are possible for λ and A?

(c) If the algebraic multiplicity of the eigenvalue λ of A is 6 and if this exceeds its geometric multiplicity by 3, how many different Jordan structures are possible for λ and A, disregarding interchanges of Jordan blocks?

22. Show that for a symmetric matrix $A = A^T \in \mathbb{R}^{n,n}$, we have $\ker(A - \lambda I) = \ker(A - \lambda I)^k$ for all $\lambda \in \mathbb{R}$ and all $k \geq 1$.
(Hint: start with $k = 2$.)

23. (a) Show that the inverse of a nonsingular Jordan block $J(\lambda, k)$ has the form

$$\begin{pmatrix} 1/\lambda & 0 & & \dots & 0 \\ -1/\lambda^2 & 1/\lambda & & & \\ 1/\lambda^3 & -1/\lambda^2 & 1/\lambda & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ (-1)^{k+1}/\lambda^k & \dots & 1/\lambda^3 & -1/\lambda^2 & 1/\lambda \end{pmatrix}.$$

(b) If $J = \operatorname{diag}(J_i)$ is the Jordan normal form of a nonsingular matrix A, what is the Jordan normal form of A^{-1}?

24. True or false:

(a) If J_{kk} is a Jordan block, then J^2 is a Jordan block.

(b) The matrix $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ has the eigenvalue $\lambda = 1$.

(c) The matrix $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ has the eigenvalue $\lambda = 0$.

(d) The matrix A of part (b) satisfies $A^2 - 2I = O_3$.

(e) The matrix A of part (b) satisfies $A^2(A - 2I) = O_3$

(f) The matrix A of part (b) satisfies $A(A - 2I) = O_3$

(g) The matrix B of part (c) satisfies $B(B^2 - 3B + 2I) = O_3$

(h) The eigenvalue -1 of $J = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ has the algebraic multiplicity 2.

(i) The eigenvalue 3 of J in part (h) has the algebraic multiplicity 2.

(j) 0 is an eigenvalue of J in part (h).

(k) The matrix J in part (h) has the eigenvalues $\lambda = 3$ and $\mu = -1$, both of geometric multiplicity equal to 1.

(l) The matrix J in part (h) has the same minimal and characteristic polynomial.

(m) The matrix J in part (h) is diagonalizable over $\mathbb{C}$, but not over $\mathbb{R}$.

25. Show that $A = \begin{pmatrix} -4 & -4 \\ 9 & 8 \end{pmatrix}$ is not diagonalizable. What is the Jordan normal form of A?

26. Show that a matrix of the form $A = \begin{pmatrix} 1 & 0 & 0 \\ * & 2 & 0 \\ 1 & * & 1 \end{pmatrix}$ cannot be diagonalized, no matter what the entries marked by $*$ are.

27. (a) Show that for a fixed λ all matrices of the form $A = \begin{pmatrix} \lambda & \alpha \\ 0 & \lambda \end{pmatrix}$ are similar for every $\alpha \neq 0$. What if $\alpha = 0$?

(b) Show that for a fixed λ all matrices of the form $A = \begin{pmatrix} \lambda & \alpha_1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & \alpha_{k-1} \\ 0 & & & \lambda \end{pmatrix}$ are similar for any choice of α_i as long as $\prod \alpha_i \neq 0$. What if some $\alpha_j = 0$?

(c) Are the matrices $B = \begin{pmatrix} \lambda & 2 \\ 0 & \lambda \end{pmatrix}$ and $C = \begin{pmatrix} \lambda & 0 \\ -201 & \lambda \end{pmatrix}$ similar or not? State your reasons.

28. Find the Jordan normal form of $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

29. Let A be an n by n matrix whose k^{th} column is a multiple ae_k of the k^{th} unit vector e_k.

(a) Show that A has the eigenvalue a. What is the corresponding eigenvector?

(b) Give several examples of 4 by 4 matrices A with second column $\begin{pmatrix} 0 \\ -3 \\ 0 \\ 0 \end{pmatrix}$ that are diagonalizable and likewise examples that are not diagonalizable.

(c) In part (b), which entries of your example matrices A seem to determine whether A is diagonalizable or not?

30. (a) If A has the Jordan normal form $J = \operatorname{diag}(J(\lambda,3), J(\lambda,2), J(\lambda,1), J(\mu,4), J(\mu,4), J(3,2))$ for $\lambda \neq \mu$, what are the dimensions of the eigenspaces $E(\lambda)$ and $E(\mu)$, depending on whether $\lambda = 3$, $\mu = 3$, or neither.

(b) What are the characteristic and minimal polynomials for J in each case of (a)? Repeat the question for A.

31. (a) Can a complex matrix $A \notin \mathbb{R}^{n,n}$ have both real and nonreal eigenvalues?

(b) Can a complex matrix $A \notin \mathbb{R}^{n,n}$ have only real eigenvalues?

(c) Can parts (a) or (b) hold for a 2 by 2 matrix A?

Hint: Try to construct a complex matrix $A_{2,2}$ with real trace and negative determinant, if possible.

Teacher's Problem–Making Exercise

T 14. To construct dense integer matrices A_{nn} with integer principal vector chains and a predetermined Jordan structure is a charm: Start with an integer Jordan normal form matrix J_{nn}, composed of integer Jordan blocks J_ℓ on its block diagonal. The diagonal Jordan blocks J_ℓ of J can have the standard form $J_\ell = J(\lambda_\ell, k) = \begin{pmatrix} \lambda_\ell & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_\ell \end{pmatrix}_{kk}$ for an integer λ_ℓ or the real Jordan normal form (see Section 14.2 below) $J_\ell = J(a_\ell, b_\ell, 2r) = \begin{pmatrix} a_\ell & b_\ell & & & & 0 \\ -b_\ell & a_\ell & & & & \\ 1 & 0 & a_\ell & b_\ell & & \\ 0 & 1 & -b_\ell & a_\ell & & \\ & & \ddots & & \ddots & \\ 0 & & & & \ddots & \ddots \end{pmatrix}_{2r,2r}$ for two integers a_ℓ, b_ℓ with $b_\ell \neq 0$. Next construct an n by n unimodular matrix X as in Teacher Problem T 6 of Section 6.1.P from two unit triangular matrices, one lower and the other one upper triangular, i.e., $X = RL$ or $X = LR$. Then the matrix $A := XJX^{-1}$ has integer entries. Moreover A has the Jordan normal form J for the eigenvalues $\lambda_\ell \in \mathbb{R}$ and $a_\ell + b_\ell i \in \mathbb{C}$ and integer principal vector chains contained in the respective columns of X.

Example : For $n = 5$ and the Jordan normal form $J = \text{diag}\left(\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & & \\ 1 & 1 & \\ & 1 & 1 \end{pmatrix} \right)$ with eigenvalues 1 (3 fold) and $1 \pm 2i$, we use $X = RL =$

$$\begin{pmatrix} -1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & 1 & -1 & 0 & 0 \\ -1 & 1 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & -1 & 1 & 1 \\ -1 & 4 & 0 & 0 & -1 \\ 0 & 2 & -1 & -1 & -1 \\ -1 & 0 & -1 & -2 & -1 \\ -2 & 1 & 0 & -1 & -1 \end{pmatrix}$$

and compute $A = XJX^{-1} = \begin{pmatrix} 0 & 5 & -4 & 4 & -6 \\ 7 & 1 & -8 & 1 & 14 \\ 5 & 1 & -5 & 2 & 8 \\ 3 & -2 & 0 & 0 & 6 \\ 3 & -3 & 0 & -2 & 9 \end{pmatrix}$. The dense integer matrix A has the same eigenvalues 1 (3 fold) and $1 \pm 2i$ as J. The corresponding eigenvectors are $\begin{pmatrix} 3 \\ -1 \\ 0 \\ -1 \\ -2 \end{pmatrix} + \begin{pmatrix} 0 \\ 4 \\ 2 \\ 0 \\ 1 \end{pmatrix} i$ for $\lambda_1 = 1 + 2i$ and $\begin{pmatrix} 3 \\ -1 \\ 0 \\ -1 \\ -2 \end{pmatrix} - \begin{pmatrix} 0 \\ 4 \\ 2 \\ 0 \\ 1 \end{pmatrix} i$ for $\lambda_2 = 1 - 2i$. The respective real and imaginary part vectors of these complex eigenvectors appear in columns 1 and 2 of X that correspond to the first diagonal block $\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$ in J. The eigenvector $\begin{pmatrix} 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}$ for $\lambda_3 = 1$ appears in the last column of X. For $\lambda_3 = 1$ with index equal to three, a complete principal vector chain $x^{(3)}$, $x^{(2)}$, $x^{(1)}$ appears in descending order in the columns 3, 4, and 5 of X.

If we form the matrix $B = XJ^TX^{-1} = \begin{pmatrix} 1 & -1 & -1 & 1 & 0 \\ -8 & 3 & 6 & 2 & -18 \\ -6 & -1 & 8 & -1 & -10 \\ -3 & -2 & 6 & -2 & -2 \\ -3 & 1 & 3 & 0 & -5 \end{pmatrix}$ instead, then this matrix has the same eigenvalues as A and as J and its Jordan normal form has the same Jordan structure. The corresponding eigenvectors and principal vector chains are a bit harder to find from the construction data: We have $X^{-1}BX = J^T$ and $EJ^TE = J$ where $E = \text{diag}(E_i)$ is conformally partitioned as $J = \text{diag}(J_i)$ and each $E_i = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix} = E_i^{-1}$. – [The students might want to check the last assertion on transforming a block matrix of transposed upper Jordan blocks $J^T = \text{diag}(J_i^T)$ to its lower block form $J = \text{diag}(J_i)$.] – Thus

$$\text{diag}(E_i)X^{-1}BX\,\text{diag}(E_i) \;=\; E^{-1}J^TE \;=\; J\ .$$

And the eigenvectors and principal vectors of B appear in the respective columns of X diag(E_i) instead. ◀

14.2 Theory (The Real Jordan Normal Form, the Companion Matrix Normal Form, and Symmetric Matrix Products)

We study normal forms that can be achieved over the reals for non diagonalizable real matrices, as well as symmetric matrix products.

According to Chapters 7 through 12, a matrix normal form $N = X^{-1}AX$ represents the standard matrix $A = A_{\mathcal{E}}$ of a linear transformation with respect to a particularly chosen basis $\mathcal{X}$. Normal forms are designed to reveal certain aspects of the linear transformation $x \mapsto Ax$. We have described several normal forms for matrices, such as the diagonal form under similarity in Sections 9.1, 9.1.D, 11.1, and 11.2, the singular value decomposition of Chapter 12, as well as the complex Jordan normal form of Section 14.1. Among these, the SVD is the only normal form that is real for all real matrices A.

The similarity normal forms, such as a diagonal or a Jordan normal form representation of a given matrix $A \in \mathbb{R}^{n,n}$, generally involve complex matrix representations for real linear mappings $A : \mathbb{R}^n \to \mathbb{R}^n$, and thus they may not be all that revealing and desirable for real problems. The main culprit here is the Fundamental Theorem of Algebra which places the roots of a polynomial, such as the minimal or the characteristic polynomial of a real matrix A in $\mathbb{C}$. Thus the Jordan normal forms of Theorems 14.1 and 14.2 can generally only be achieved for a real matrix by using a complex Jordan basis of $\mathbb{C}^n$. For many problems, such as for finding real solutions of linear differential equations for example, it is, however, important to find sparse real representations of real system matrices A. In Section 9.3

we have partially dealt with this issue. Now we extend those results by using the Jordan normal form of Theorem 14.2.

If $A \in \mathbb{R}^{n,n}$ has the eigenvalue $\lambda = a + bi \notin \mathbb{R}$ for a necessarily complex eigenvector $z = u + iw$ with $u, w \in \mathbb{R}^n$, then $Az = \lambda z$ and $A\overline{z} = \overline{Az} = \overline{\lambda z} = \overline{\lambda}\overline{z}$ since $\overline{A} = A$ for real A. Therefore the complex number $\overline{\lambda} = a - bi \neq \lambda$ is also an eigenvalue of A for the eigenvector $\overline{z} = u - iw$ if λ is for z. The main idea that establishes a real Jordan form representation for real matrices A lies in combining the two complex conjugate eigenvalues λ and $\overline{\lambda}$ of $A \in \mathbb{R}^{n,n}$ and their corresponding eigenvectors and principal vectors chains into one.

Lemma 4: If $A \in \mathbb{R}^{n,n}$ and $Az = \lambda z$ for $\lambda = a + bi \notin \mathbb{R}$ and $z = u + iw \in \mathbb{C}^n$ with $u, w \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$, then the two vectors u and w are linearly independent. Moreover $Au = au - bw$ and $Aw = bu + aw$. ◀

Proof: By linearity and complex arithmetic we see that $Az = A(u + iw) = Au + iAw = \lambda z = (a + bi)(u + iw) = au - bw + i(bu + aw)$. And both equations $Au = au - bw$ and $Aw = bu + aw$ follow by comparing the real and imaginary parts in this identity.
To prove linear independence of u and $w \in \mathbb{R}^n$, we may assume without loss of generality that $u = \alpha w$ for $\alpha \in \mathbb{R}$. Then $u + iw = (\alpha + i)w$ and

$$Az = A(u + iw) = (\alpha + i)Aw = \lambda z = \lambda\alpha w + i\lambda w = \lambda(\alpha + i)w.$$

Thus $Aw = \lambda w$ with $w \in \mathbb{R}^n$. But the left hand side Aw of this equation lies in $\mathbb{R}^n$, while the right hand side λw does not, a contradiction that makes the real and imaginary parts u and w of a complex eigenvector z of $A \in \mathbb{R}^{n,n}$ linearly independent. ■

If we link the two complex conjugate eigenvalues $\lambda \neq \overline{\lambda}$ and the corresponding eigenvectors $z, \overline{z}$ of a real matrix A, we observe that

$$A \begin{pmatrix} | & | \\ z & \overline{z} \\ | & | \end{pmatrix} = \begin{pmatrix} | & | \\ z & \overline{z} \\ | & | \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix},$$

i.e., we see a partial complex diagonalization $\begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix}$ of A with respect to its two complex eigenvectors z and $\overline{z}$. Using the real and imaginary parts vectors u and w of the complex eigenvector $z = u + iw$ instead, we arrive at an analogous partial real representation of A with respect to u and $w \in \mathbb{R}^n$. Namely

$$A \begin{pmatrix} | & | \\ u & w \\ | & | \end{pmatrix} = \begin{pmatrix} | & | \\ u & w \\ | & | \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \tag{14.4}$$

according to Lemma 4. Note that the real matrix $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ is similar to the complex matrix $\begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix}$; see Problem 9 in Section 14.1.P. We can repeat this capture of two complex conjugate eigenvalues λ and $\overline{\lambda}$ of a real matrix A in one 2 by 2 real matrix in (14.4) for principal vectors of order k and establish linear independence of their real and imaginary parts vectors as well.

Lemma 5: If $z^{(k)} = u^{(k)} + iw^{(k)}$ with $x^{(k)}, w^{(k)} \in \mathbb{R}^n$ is a complex principal vector of order k for the eigenvalue $\lambda = a + bi \notin \mathbb{R}$ and $A \in \mathbb{R}^{n,n}$, then

(a) $\overline{z^{(k)}} = u^{(k)} - iw^{(k)}$ is a principal vector of order k for the eigenvalue $\overline{\lambda}$ of A, and

(b) the vectors $u^{(k)}$ and $w^{(k)} \in \mathbb{R}^n$ are linearly independent. ◀

Proof: Part (a) follows readily by complex conjugation of the two defining equations $(A - \lambda I)^k z^{(k)} = 0$ and $(A - \lambda I)^{k-1} z^{(k)} \neq 0$.
In (b), we establish the result for $k = 2$ only. The case of eigenvectors, i.e., $k = 1$, has been settled in Chapter 9. The higher dimensional cases follow along similar lines as our proof for $k = 2$ below.
If the real and imaginary parts $u^{(2)}$ and $w^{(2)} \in \mathbb{R}^n$ of the principal vector $z^{(2)} = u^{(2)} + iw^{(2)} \neq 0$ of order two are linearly dependent for $\lambda = a + bi \notin \mathbb{R}$ and $A \in \mathbb{R}^{n,n}$, then we may assume without loss of generality that $u^{(2)} = \alpha w^{(2)} \neq 0$ for $\alpha \in \mathbb{R}$. In this case, $w^{(2)}$ is a real principal vector of order 2 for A and λ, since $(A - \lambda I)^2 z^{(2)} = (\alpha + i)(A - \lambda I)^2 w^{(2)} = 0$ with $\alpha \in \mathbb{R}$. Thus for $k = 2$ and one complex principal vector of order 2 with linearly dependent real and imaginary parts, the real vector $x := w^{(2)} \neq 0$ is also a principal vector of order 2. If we write out $(A - \lambda I)^2 x = (A^2 - 2\lambda A + \lambda^2 I)x = 0$, we note that $A^2 x = (2\lambda A - \lambda^2 I)x$ is a real vector. Thus $A^2 x = \overline{A^2 x} = (2\overline{\lambda} A - \overline{\lambda}^2 I)x$, since x, A, and I are all real. Consequently

$$0 = A^2 x - \overline{A^2 x} = (2(\lambda - \overline{\lambda})A - (\lambda^2 - \overline{\lambda}^2)I)x = (4biA - 4abiI)x = 4bi(A - aI)x ,$$

or $(A - aI)x = 0$ since $\lambda - \overline{\lambda} = 2bi \neq 0$ and $\lambda^2 - \overline{\lambda}^2 = 4abi$. If $Ax = ax$ for $x \in \mathbb{R}^n$ and $a \in \mathbb{R}$, then the identity $0 = (A - \lambda I)^2 x = (A^2 - 2\lambda A + \lambda^2 I)x = a^2 x - 2(a + bi)ax + (a^2 - b^2 + 2abi)x$ makes its real parts vector equal to zero. Namely $(a^2 - 2a^2 + a^2 - b^2)x = -b^2 x = 0$. This forces $b = 0$, unless $x = 0$. But if $b = 0$, then $\lambda = a \in \mathbb{R}$, a contradiction. And if $x = w^{(2)} = 0$, then $z^{(2)} = 0$ gives another contradiction. Thus the real and imaginary parts of an order 2 principal vector for a complex eigenvalue of a real matrix must be linearly independent. ■

The following follows directly from Lemma 5 (a).

Corollary 2: If A is a real square matrix, then the Jordan blocks for nonreal eigenvalues occur in pairs in the Jordan normal form J of A. Specifically if $\lambda = a + bi \notin \mathbb{R}$ is an eigenvalue of $A \in \mathbb{R}^{n,n}$ and if the Jordan block $J(\lambda, m)$ occurs in J, so does the block $J(\overline{\lambda}, m)$. ◀

If $\{x^{(\ell)} = u^{(\ell)} + iw^{(\ell)}\}$ denotes a chain of principal vectors of order k for $\ell = k, ..., 1, 0$ and real vectors $u^{(\ell)}$ and $w^{(\ell)}$ for the nonreal eigenvalue $\lambda = a + bi$ of $A \in \mathbb{R}^{n,n}$ (with $x^{(0)} := 0$ for convenience), then for any $1 \leq \ell \leq k$ we have

$$\begin{aligned}
A \begin{pmatrix} | & | \\ u^{(\ell)} & w^{(\ell)} \\ | & | \end{pmatrix} &= \begin{pmatrix} | & | \\ u^{(\ell)} & w^{(\ell)} \\ | & | \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} + \begin{pmatrix} | & | \\ u^{(\ell-1)} & w^{(\ell-1)} \\ | & | \end{pmatrix} \\
&= \begin{pmatrix} | & | & | & | \\ u^{(\ell)} & w^{(\ell)} & u^{(\ell-1)} & w^{(\ell-1)} \\ | & | & | & | \end{pmatrix} \left(\begin{array}{cc} a & b \\ -b & a \\ \hline 1 & 0 \\ 0 & 1 \end{array} \right) .
\end{aligned} \tag{14.5}$$

To see this, observe that $x^{(\ell-1)} = u^{(\ell-1)} + iw^{(\ell-1)} = (A - \lambda I)x^{(\ell)}$ by the definition of a principal vectors chain. And equation (14.5) becomes apparent by expansion and by sorting $x^{(\ell-1)}$'s real and imaginary part vectors as follows:

$$\begin{aligned}
x^{(\ell-1)} &= (A - \lambda I)x^{(\ell)} = (A - aI - biI)(u^{(\ell)} + iw^{(\ell)}) \\
&= Au^{(\ell)} - au^{(\ell)} + bw^{(\ell)} + i(Aw^{(\ell)} - aw^{(\ell)} - bu^{(\ell)}) \\
&=: u^{(\ell-1)} + iw^{(\ell-1)} .
\end{aligned}$$

Comparing the real and imaginary part vectors above, gives us the two identities $Au^{(\ell)} - au^{(\ell)} + bw^{(\ell)} = u^{(\ell-1)}$ and $Aw^{(\ell)} - aw^{(\ell)} - bu^{(\ell)} = w^{(\ell-1)}$, or $Au^{(\ell)} = au^{(\ell)} - bu^{(\ell)} + u^{(\ell-1)}$ and $Aw^{(\ell)} = bu^{(\ell)} + aw^{(\ell)} + w^{(\ell-1)}$. This establishes (14.5) . Note that (14.5) is a real matrix equation that generalizes the real and imaginary parts eigenvector equation (14.4) to principal vectors.

For each complex eigenvalue $\lambda = a + bi \notin \mathbb{R}$ of a matrix $A \in \mathbb{R}^{n,n}$ there may be several Jordan blocks $J(\lambda, n_1), ..., J(\lambda, n_j)$ associated with $\lambda \in \mathbb{C}$. Each of these has an equal sized complex conjugate Jordan block $J(\overline{\lambda}, n_1), ..., J(\overline{\lambda}, n_j)$ in the complex Jordan normal form of A. The **real Jordan normal form** of a real matrix A uses the two corresponding Jordan blocks $J(\lambda, m)$ and $J(\overline{\lambda}, m)$ in tandem to represent A with respect to their joint principal real and imaginary parts vector subspaces by the $2m$ by $2m$ real matrix

$$J(a,b,2m) = \left(\begin{array}{cc|cc|cccc|cc}
a & b & 0 & 0 & & & & & 0 & 0 \\
-b & a & 0 & 0 & & & & & 0 & 0 \\
\hline
1 & 0 & a & b & & & & & & \\
0 & 1 & -b & a & & & & & & \\
& & & \ddots & & \ddots & & & & \\
& & & & & \ddots & & \ddots & 0 & 0 \\
& & & & & & & & 0 & 0 \\
\hline
0 & 0 & & & & & 1 & 0 & a & b \\
0 & 0 & & & & & 0 & 1 & -b & a
\end{array}\right)$$

with m diagonal 2 by 2 blocks $\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \in \mathbb{R}^{2,2}$, $m-1$ subdiagonal 2 by 2 identity matrices, and zeros elsewhere. The matrix $J(a,b,2m)$ is called the **real Jordan block** for the complex eigenvalue $\lambda = a + bi$ of A where $b \neq 0$.

Notation: We denote Jordan normal forms and Jordan blocks by the same letter J throughout. A specific Jordan block can carry either two or three variables as in $J(\lambda, 5)$ or in $J(a,b,2k)$. Jordan blocks with 2 variables $J(\lambda, k)$ always denote the standard Jordan block $\begin{pmatrix} \lambda & & & 0 \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ 0 & \dots & 1 & \lambda \end{pmatrix}$ with $\lambda \in \mathbb{C}$ or $\lambda \in \mathbb{R}$ and k a positive integer. Jordan blocks with three attached variables $J(a,b,\ell)$ always denote a real Jordan block as just described. Here a and b are always real numbers and ℓ is a positive even integer. ◀

Theorem 14.3: (**The Real Jordan Normal Form**)

Every real square matrix A is similar over the reals to its real Jordan normal form J. J is a block diagonal matrix comprised of real Jordan blocks. Namely if $\lambda \in \mathbb{R}$ is an eigenvalue of A, the Jordan blocks associated with it have the standard form $J(\lambda, m) = \begin{pmatrix} \lambda & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix} \in \mathbb{R}^{m,m}$ for various sizes m, while for $\lambda = a + bi \notin \mathbb{R}$ and $a, b \in \mathbb{R}$, the associated Jordan blocks have the 2 by 2 lower block triangular form

$$J(a,b,2r) = \begin{pmatrix} a & b & & & & 0 \\ -b & a & & & & \\ 1 & 0 & a & b & & \\ 0 & 1 & -b & a & & \\ & & \ddots & & \ddots & \\ & 0 & & & \ddots & \ddots \end{pmatrix}.$$

The individual eigenvalues and their respective Jordan block dimensions and types are determined uniquely by A. The order of the Jordan blocks that appear in J is not unique.

Each Jordan block $J(a,b,2r) = \begin{pmatrix} a & b & & & & \\ -b & a & & & & \\ 1 & 0 & a & b & & \\ 0 & 1 & -b & a & & \\ & & \ddots & & \ddots & \\ & & & & \ddots & \ddots \end{pmatrix}$ in Theorem 14.3 can be represented in an even sparser form as $\begin{pmatrix} a & b & & & & \\ -b & a & & & & \\ 0 & 1 & a & b & & \\ 0 & 0 & -b & a & & \\ & & \ddots & & \ddots & \\ & & & & \ddots & \ddots \end{pmatrix}$. Here the off diagonal 2 by 2 blocks I_2 of the real Jordan normal form J have been replaced by

$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ while all other entries are left unchanged in J. This sparse form shows that every real square matrix can be reduced to a real tridiagonal matrix via a real similarity.

Corollary 3: Every real square matrix is similar over $\mathbb{R}$ to a real tridiagonal matrix. ◀

Example 6: Find the real Jordan normal form of $A = \begin{pmatrix} 4 & 7 & -1 & -6 \\ -2 & -3 & 2 & 3 \\ 0 & 1 & 2 & -1 \\ 0 & -1 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{4,4}$.

Using the vector iteration method of Section 9.1 with $x = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$ for example gives us the vector iteration matrix $\begin{pmatrix} | & & | \\ x & \dots & A^4x \\ | & & | \end{pmatrix} = \begin{pmatrix} 0 & 5 & 16 & 30 & 32 \\ 0 & -1 & -4 & -8 & -8 \\ 1 & 3 & 6 & 8 & 4 \\ -1 & -1 & 0 & 4 & 12 \end{pmatrix}$ with the RREF $\begin{pmatrix} 1 & 0 & 0 & 0 & -4 \\ 0 & 1 & 0 & 0 & 8 \\ 0 & 0 & 1 & 0 & -8 \\ 0 & 0 & 0 & 1 & 4 \end{pmatrix}$. Thus the minimal polynomial $p_A(\lambda) = x^4 - 4x^3 + 8x^2 - 8x + 4 = (x^2 - 2x + 2)^2$. Since it has degree $4 = n$, it coincides with the characteristic polynomial $f_A(\lambda) = \det(\lambda I - A)$ of A. Moreover since $p_A(\lambda)$ is a square of the irreducible real polynomial $x^2-2x+2 = (x-1-i)(x-1+i)$, A is not diagonalizable. A has the double complex conjugate eigenvalues $1 \pm i$. Thus the real Jordan normal form of A is

$$J = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 \end{pmatrix}$$

according to Theorem 14.3. The real Jordan normal form J can be achieved via the similarity $J = X^{-1}AX$ of A for $X = \begin{pmatrix} 3 & 1 & 3 & -1 \\ 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}$. The first two columns of X contain the real and imaginary parts of the principal vector $x^{(2)} = \begin{pmatrix} 3+i \\ 0 \\ 1+i \\ 1 \end{pmatrix}$ of order 2 for A and $\lambda = 1+i$. Here $x^{(2)} \in \ker(A-(1+i)I)^2$, while $x^{(2)} \notin \ker(A-(1+i)I)$. The last two columns of X contain the real and imaginary part vectors of the eigenvector $x^{(1)} = (A-(1+i)I)x^{(2)} = \begin{pmatrix} 3-i \\ -1 \\ 1 \\ -i \end{pmatrix}$ for λ.

The students should verify all of the assertions by hand computations and row reducing $A-(1+i)I$ and $(A-(1+i)I)^2$. ◀

In the sequel we demonstrate how to represent a rational matrix $A \in \mathbb{Q}^{n,n}$ in a sparse normal form if its entries are rational numbers of the form $\frac{r}{s}$ for integers r and s with $s \neq 0$. (For the definition of rational numbers $\mathbb{Q}$, see Appendix A.) The **companion matrix normal form** of a rational matrix is a sparse rational matrix that is reminiscent of the Jordan normal form, but it uses different types of diagonal blocks. If p is a monic polynomial of degree n, i.e., if $p(x) = x^n + a_{n-1}x^{n-1} + ... + a_1x + a_0$ with coefficients $a_j \in \mathbb{R}$, then the **companion matrix**

$$C(p) = \begin{pmatrix} 0 & & & -a_0 \\ 1 & & & \vdots \\ & \ddots & & \vdots \\ & & 1 & -a_{n-1} \end{pmatrix}_{nn}$$

has the characteristic polynomial $f_{C(p)}(\lambda) = \det(\lambda I - C(p)) = p(\lambda)$; see Problem 9 in Section 9.R or expand $\det(\lambda I - C(p))$ along its last column. The companion matrix $C(p)$ has the same characteristic polynomial as its complex and as its real Jordan normal form, since these three matrices are all similar to each other.
We start by investigating the real Jordan normal form for $C(p)$ in light of the possible factorizations of $p(x)$ over the reals before studying the same over the rational numbers.
The real Jordan normal form J of $C(p)$ is lower block triangular according to Theorem 14.3. Specifically for a real root λ_k of $p(x)$, the real eigenvalue λ_k of $C(p)$ appears on the diagonal of J, repeated according to its algebraic multiplicity. For a complex root $\lambda_k = a_k + b_k i \notin \mathbb{R}$ of p, the two corresponding complex conjugate eigenvalues λ_k and $\overline{\lambda_k}$ of $C(p)$ give rise to a 2 by 2 real diagonal block $\begin{pmatrix} a_k & b_k \\ -b_k & a_k \end{pmatrix} \in \mathbb{R}^{2,2}$, possibly repeated; all according to Theorem 14.3. Clearly for each complex eigenvalue $\lambda_k = a_k + b_k i \notin \mathbb{R}$ we have $\det\left(\lambda I_2 - \begin{pmatrix} a_k & b_k \\ -b_k & a_k \end{pmatrix}\right) = (\lambda - a_k)^2 + b_k^2$. The lower co-diagonal entries of 1 or of I_2 in J for real or complex eigenvalues of $C(p)$ of index $j > 1$, respectively, can be ignored when forming $C(p)$'s characteristic polynomial since the characteristic polynomial of a block triangular matrix is equal to the product of the characteristic polynomials of its individual diagonal blocks. This follows from the determinant Proposition, part 13, of Chapter 8. Therefore

$$\begin{aligned} p(\lambda) &= f_{C(p)}(\lambda) = \det(\lambda I_n - C(p)) = \\ &= \prod_{\lambda_\ell \text{ real}} (\lambda - \lambda_\ell)^{n_\ell} \prod_{\substack{\lambda_j \text{ complex} \\ \lambda_j = a_j + b_j i}} \left((\lambda - a_j)^2 + b_j^2\right)^{n_j} \end{aligned}$$

with $n_\ell,\ n_j \geq 1$. This factors the given polynomial $p(\lambda)$ over the reals into degree one or degree 2 real polynomial factors, the latter of which have no real roots since $b_j \neq 0$. This proves the following well known result about factoring real polynomials p into **irreducible factors**. Here we call a real polynomial q **irreducible** over its field of coefficients $\mathbb{R}$ if it cannot be factored into real polynomials of lower degree, such as $q(\lambda) = (\lambda - a_k)^2 + b_k^2$ for $b_k \neq 0$.

Corollary 4: Every real polynomial $p(x)$ can be factored over the reals into first and second degree real polynomial factors. ◀

Example 7: Factor the two real polynomials $f(x) = x^5 + 1$ and $g(x) = x^5 - 1$ into the product of irreducible real polynomials.

We approach this problem by computing the eigenvalues λ_i and μ_j of the companion matrices $C(f) = \begin{pmatrix} 0 & & & -1 \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{pmatrix}_{5,5}$ and $C(g) = \begin{pmatrix} 0 & & & 1 \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{pmatrix}_{5,5}$ in MATLAB. They are $\lambda_1 = -1,\ \lambda_{2,3} = -0.309 \pm 0.951i,\ \lambda_{4,5} = 0.809 \pm 0.568i$ and $\mu_1 = 1,\ \mu_{2,3} = 0.309 \pm 0.951i,\ \mu_{4,5} = -0.809 \pm 0.588i$, respectively.

Thus over $\mathbb{C}$,

$$\begin{aligned} f(x) &= (x+1)(x-\lambda_2)(x-\overline{\lambda_2})(x-\lambda_4)(x-\overline{\lambda_4}) \\ &= (x+1)(x^2+0.618x+1)(x^2-1.618x+1) \\ &= (x+1)(x^2-(0.5-\frac{\sqrt{5}}{2})x+1)(x^2-(0.5+\frac{\sqrt{5}}{2})x+1)\ , \end{aligned}$$

while

$$\begin{aligned} g(x) &= (x-1)(x^2-0.618x+1)(x^2+1.618x+1) \\ &= (x+1)(x^2+(0.5-\frac{\sqrt{5}}{2})x+1)(x^2+(0.5+\frac{\sqrt{5}}{2})x+1)\ . \end{aligned}$$

Here we have expressed both f and g as a product of irreducible real factor polynomials. ◀

If we attempt to factor rational polynomials $p(x)$ with coefficients in $\mathbb{Q}$ into irreducible rational polynomial factors instead, we must realize that there is no equivalent result to Corollary 4 for the maximal factor degree of real factorizations. Namely, arbitrary degree rational polynomials may be irreducible over $\mathbb{Q}$, such as $p(x) = x^n - 2$ is for any $n \geq 1$. Note that the two rational polynomials in Example 7 factor irreducibly over $\mathbb{Q}$ into $f(x) = (x+1)(x^4 - x^3 + x^2 - x + 1)$ and $g(x) = (x-1)(x^4 + x^3 + x^2 + 1)$. This can be verified by long division and by looking back at the the real factors of f and g in Example 7.

Now we turn to rational sparse matrix representations of rational matrices $A_{nn} \in \mathbb{Q}^{n,n}$. We assume that $f_A(\lambda) = \det(\lambda I - A) = \prod_{i=1}^{k} p_i(x)^{n_i}$ and $p_A(\lambda) = \prod_{i=1}^{k} p_i(x)^{m_i}$ with $m_i \leq n_i$ for each i. Furthermore we assume that the polynomials $p_i(x)$ are irreducible over $\mathbb{Q}$. Then there exists a basis of $\mathbb{Q}^n$ that represents the linear transformation $x \mapsto Ax$ of $\mathbb{Q}^n$ as a block diagonal matrix. This matrix contains a companion matrix for each irreducible factor p_i of f_A (or p_A) as its diagonal block, with or without accompanying co-diagonal identity matrices as follows:

Theorem 14.4: (**The Companion Matrix Normal Form**)

Let $A \in \mathbb{Q}^{n,n}$ be a matrix with rational entries. Assume that the characteristic polynomial of A is $f_A(\lambda) = \det(\lambda I - A) = \prod_{i=1}^{k} p_i^{n_i}$ and that the minimum polynomial $p_A(\lambda) = \prod_{i=1}^{k} p_i^{m_i}$ and $f_A(x)$ are both factored irreducibly over $\mathbb{Q}$.

Then $m_i \leq n_i$ for each $i = 1, ..., k$ and A is similar over $\mathbb{Q}$ to a block diagonal matrix $C \in \mathbb{Q}^{n,n}$ comprised of k diagonal blocks C_i. In turn, each diagonal block C_i is in lower block diagonal form

$$C_i = \begin{pmatrix} C(p_i) & & & 0 \\ I_{d_i} & C(p_i) & & \\ & \ddots & \ddots & \\ 0 & & I_{d_i} & C(p_i) \end{pmatrix}_{\ell_i \cdot d_i \times \ell_i \cdot d_i} ,$$

comprised of $\ell_i \leq m_i$ companion matrices $C(p_i)$ for the irreducible factor p_i of f_A (or p_A) on its block diagonal and copies of the identity matrix I_{d_i} of dimension $d_i =$ degree(p_i) on its lower block co–diagonal.

The matrix $C =$ diag(C_i) is called the **companion matrix normal form** of A over $\mathbb{Q}$.

The order of the individual blocks C_i in C is not unique.

Example 8: The companion matrix $A = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$ for the polynomial $f(x) = x^5 + 1$ of Example 7 has the following normal forms, depending on the base field that we consider:

Over $\mathbb{C}$, A is similar to the complex diagonal matrix

$$\begin{pmatrix} -1 & & & & \\ & -0.309+0.951i & & & \\ & & -0.309-0.951i & & \\ & & & 0.809+0.568i & \\ & & & & 0.809-0.568i \end{pmatrix}$$

according to Example 7. This is the complex Jordan normal form of A. Over $\mathbb{C}$, A is diagonalizable because it has five distinct complex eigenvalues.
Over $\mathbb{R}$, A is similar to the block diagonal real matrix

$$\begin{pmatrix} -1 & & & & \\ & -0.309 & 0.951 & & \\ & -0.951 & -0.309 & & \\ & & & 0.809 & 0.568 \\ & & & -0.568 & 0.809 \end{pmatrix},$$

which is A's real Jordan normal form.
Finally over $\mathbb{Q}$, A is similar to its companion matrix normal form

$$\begin{pmatrix} -1 & & & & \\ & 0 & 0 & 0 & -1 \\ & 1 & 0 & 0 & 1 \\ & 0 & 1 & 0 & -1 \\ & 0 & 0 & 1 & 1 \end{pmatrix}$$

since $f(x) = x^5 + 1 = (x+1)(x^4 - x^3 + x^2 - x + 1)$ with both polynomial factors irreducible over $\mathbb{Q}$. ◀

Matrix normal forms allow us to classify all matrices that represent the same underlying linear transformation, except for different bases.

Proposition: Two n by n matrices A and B are similar if and only if the two matrices have the same complex Jordan normal form, the same real Jordan normal form, or the same companion matrix normal form.
Here we call two matrix normal forms identical, if the two forms contain the same collection of normal form blocks, disregarding their specific order. ◀

Finally, the Jordan normal form allows us to prove a classical result, namely that every real or complex matrix can be written as the product of two symmetric matrices, real or complex, respectively.

Corollary 5: Every real or complex matrix A_{nn} can be written as the product of two symmetric (real or complex, respectively) matrix factors $A = S_1 S_2$ with $S_i^T = S_i$ and S_1 nonsingular. ◀

Note that not all complex matrices can be written as the product of two hermitian matrices, see Problems 8 – 12 below. Symmetry is a stronger matrix attribute than being hermitian in the complex matrix case.

Proof: Every Jordan block $J(\lambda,k) = \begin{pmatrix} \lambda & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix}$ can be factored as

$$J(\lambda,k) = \begin{pmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{pmatrix}_{k,k} \begin{pmatrix} 0 & & 1 & \lambda \\ & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \\ 1 & \cdot^{\cdot^{\cdot}} & & \\ \lambda & & & 0 \end{pmatrix}_{kk}$$

into the product of two real (or complex, if $\lambda \notin \mathbb{R}$) symmetric matrices. Likewise every real Jordan block

$$J(a,b,2m) = \begin{pmatrix} a & b & & & & \\ -b & a & & & & \\ 1 & 0 & \ddots & & & \\ 0 & 1 & & \ddots & & \\ & & \ddots & & \ddots & \\ & & & 1 \quad 0 & & a \quad b \\ & & & 0 \quad 1 & & -b \quad a \end{pmatrix}_{2m,2m}$$

for a complex eigenvalue $a + bi \notin \mathbb{R}$, a, $b \in \mathbb{R}$ factors symmetrically as

$$J(a,b,2m) = \begin{pmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{pmatrix}_{2m,2m} \begin{pmatrix} & & & 0 \quad 1 & -b \quad a \\ & & & 1 \quad 0 & a \quad b \\ & & \cdot^{\cdot^{\cdot}} & & \cdot^{\cdot^{\cdot}} \\ 0 \quad 1 & & & \cdot^{\cdot^{\cdot}} & \\ 1 \quad 0 & \cdot^{\cdot^{\cdot}} & & & \\ -b \quad a & & & & \\ a \quad b & & & & \end{pmatrix}.$$

This can be readily checked. Consequently any Jordan normal form matrix J, real or complex, can be written as the product $J = S_1 S_2$ of two block diagonal matrices S_1 and S_2 that carry symmetric diagonal blocks of the form $\begin{pmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{pmatrix}$ in S_1, and $\begin{pmatrix} 0 & & 1 & \lambda \\ & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \\ 1 & \cdot^{\cdot^{\cdot}} & & \\ \lambda & & & 0 \end{pmatrix}$ or $\begin{pmatrix} & & & & 0 & 1 & -b & a \\ & & & & 1 & 0 & a & b \\ & & & \cdot^{\cdot^{\cdot}} & & \cdot^{\cdot^{\cdot}} & & \\ 0 & 1 & & & \cdot^{\cdot^{\cdot}} & & & \\ 1 & 0 & \cdot^{\cdot^{\cdot}} & & & & & \\ -b & a & & & & & & \\ a & b & & & & & & \end{pmatrix}$ in S_2. By construction, both S_1 and S_2 are symmetric while S_1 is nonsingular.

If $X^{-1}AX = J$ is the Jordan normal form of A and if $J = S_1S_2$, then $A = XJX^{-1} = XS_1S_2X^{-1} = XS_1X^T\left(X^T\right)^{-1}S_2X^{-1} = T_1T_2$ for the nonsingular symmetric matrix $T_1 = XS_1X^T = T_1^T$ and the symmetric matrix $T_2 = \left(X^T\right)^{-1}S_2X^{-1} = T_2^T$. ■

The symmetric factorization that exists for every square matrix allows us to classify real diagonalizable matrices via their symmetric factors.

Corollary 6: A real matrix A_{nn} is diagonalizable over $\mathbb{R}$ if and only if A can be factored into the product of one positive definite real symmetric matrix and a real symmetric matrix. ◀

Proof: If A is diagonalizable over $\mathbb{R}$, i.e., if $X^{-1}AX = J = \text{diag}(\lambda_i)$ for $\lambda_i \in \mathbb{R}$ and $X \in \mathbb{R}^{n,n}$, then

$$A = XIX^T \cdot (X^T)^{-1}JX^{-1} = S_1S_2$$

with $S_1 = XIX^T = XX^T = S_1^T$ positive definite according to the definition in Section 11.3 and $S_2 = X^{-T}\text{diag}(\lambda_i)X^{-1}$ real symmetric. Here we have abbreviated $(X^T)^{-1}$ by X^{-T}.

Conversely, if $A = S_1S_2$ with $S_i = S_i^T \in \mathbb{R}^{n,n}$ and S_1 positive definite, then there is a real orthogonal matrix U according to Section 11.1 with $U^TS_1^{-1}U = \text{diag}\left(\frac{1}{\mu_i}\right)$ for the positive eigenvalues $\mu_i > 0$ of S_1 since S_1^{-1} has the eigenvalues $\frac{1}{\mu_i}$. With $M := \text{diag}(\sqrt{\mu_i}) = M^T \in \mathbb{R}^{n,n}$ we then have

$$M^TU^TS_1^{-1}UM = \text{diag}(\sqrt{\mu_i})\text{diag}\left(\frac{1}{\mu_i}\right)\text{diag}(\sqrt{\mu_i}) = I ,$$

or $X^TS_1^{-1}X = I$ for $X := UM \in \mathbb{R}^{n,n}$. This implies that $S_1^{-1} = X^{-T}X^{-1}$, or $S_1 = XX^T$. For this specific matrix X we look at

$$X^{-1}AX = X^{-1}S_1S_2X = X^{-1}(XX^T)X^{-T}X^TS_2X = IX^TS_2X = X^TS_2X .$$

Thus A is similar to the real symmetric matrix X^TS_2X which in turn is orthogonally diagonalizable over $\mathbb{R}$ according to Chapter 11. Therefore A itself is diagonalizable over the reals. ■

This result for diagonalizable matrices relies upon our understanding of nondiagonalizable ones; giving us a circular closing to the subject.

14.2.P Problems

1. Find the real Jordan normal form of $A = \begin{pmatrix} 1 & -6 & -3 & -1 \\ 4 & 7 & 12 & -3 \\ -3 & -3 & -7 & 2 \\ -2 & -5 & -6 & 1 \end{pmatrix}$. (Hint: Decide first whether to use vector iteration as in Chapter 9.1 or determinants as in Chapter 9.1.D to find the eigenvalues of A.)

2. What is the companion matrix for $p(x) = x^7 - x^5 + 4x^2 - 3x + 12$?

3. Write out all possible real Jordan normal forms for a 13 by 13 real matrix A with the eigenvalues $\lambda = 2 + 3i$ (3 fold), $\mu = 1 - 2i$ (double), and $\nu = 7$. What is the algebraic multiplicity of ν? What are the possible geometric multiplicities of λ?

4. Determine the real Jordan normal form of the companion matrix $C = \begin{pmatrix} 0 & & & & 0 \\ 1 & & & & 6 \\ & 1 & & & -5 \\ & & 1 & & -5 \\ 0 & & & 1 & 5 \end{pmatrix}$. Is C diagonalizable over $\mathbb{R}$ or not? What are C's eigenvalues?

5. Let $J = \begin{pmatrix} \lambda & & & 0 \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix}$ be a Jordan block of size k by k.

   (a) Find an eigenvector of J for λ.

   (b) Find the index of the eigenvalue λ of J.

   (c) Find a principal vector of order k for J and λ.

   (d) Find a principal vector of order $1 < j < k$ for J and λ.

6. We call a nonzero row vector $y \in \mathbb{C}^n$ a **left eigenvector** of a matrix A for the eigenvalue λ, if $yA = \lambda y$. Likewise a row vector z with $z(A - \lambda I)^j = 0$ and $z(A - \lambda I)^{j-1} \neq 0$ is called a **left principal vector** of order j for A and λ.

   (a) For the Jordan block $J = \begin{pmatrix} \lambda & & & 0 \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix}$ of size k by k, find a left eigenvector.

   (b) For the Jordan block $J = J(\lambda, k)$ find the left principal vectors for all possible orders.

7. Let $A = \begin{pmatrix} 0 & & & -a_0 \\ 1 & & & \vdots \\ & \ddots & & \vdots \\ & & 1 & -a_{n-1} \end{pmatrix} = C(p)$ be the companion matrix for $p(x) = x^2 + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$.
   Show: If z is a root of the polynomial p, then the row vector $v = \begin{pmatrix} 1 & z & z^2 & \ldots & z^{n-1} \end{pmatrix}$ is a left eigenvector for the eigenvalue z of A.

8. Show that $A = \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{C}^{2,2}$ can be expressed as the product of two symmetric complex matrices $S_i = S_i^T \in \mathbb{C}^{2,2}$. Find S_1 and S_2.

9. Show that $B = \begin{pmatrix} 2i & -2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{C}^{2,2}$ can be expressed as the product of two symmetric complex matrices $S_i = S_i^T \in \mathbb{C}^{2,2}$.

10. Show that $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} 4 & -1 & 1 \\ -1 & -1 & -1 \\ 1 & -1 & 0 \end{pmatrix}$ can be diagonalized over $\mathbb{R}$.

11. Factor $A = \begin{pmatrix} -2 & 1 \\ 4 & -4 \end{pmatrix}$ into the product of two real symmetric matrices, if possible.

12. (a) Show that $A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ is the product of $Y = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$ and $S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

    (b) Which of the matrices S or Y is symmetric? Which is positive definite?

(c) How does this square with Corollary 5?

(d) Can you find a symmetric factorization of A in part (a) into the product of two real symmetric matrices with one positive definite factor?

(e) If the answer to (d) is affirmative, find such a symmetric factorization of A.

13. Show that the two real Jordan blocks $J(a,b,2)$ and $J(a,-b,2)$ are always similar.

14.3 Applications (Differential Equations and MATLAB)

(a) *Linear systems of differential equations*

In Section 9.3 we have solved **systems of differential equations** via linear algebra. Specifically for diagonalizable real system matrices A_{nn} we have described the solutions $x(t) \in \mathbb{R}^n$ of the DE $x'(t) = Ax(t)$ in 9.3.

In this section we assume that $J = U^{-1}AU$ is the Jordan normal form, real or complex, diagonal or not, of $A \in \mathbb{R}^{n,n}$ and that U is nonsingular. As in equation (9.7), we rewrite the DE $x' = Ax = UJU^{-1}x$ as

$$U^{-1}x'(t) \;=\; U^{-1}Ax(t) \;=\; JU^{-1}x(t)\ .$$

By setting $v(t) := U^{-1}x(t)$ we only need to solve the nearly separated system of DEs $v'(t) = Jv(t)$. If such a solution v is known, then clearly $x(t) := Uv(t)$ solves the original system of DEs $x'(t) = Ax(t)$.

If the Jordan normal form $J = \operatorname{diag}(J_1, ..., J_k)$ for A consists of k Jordan blocks J_i, then the problem $v'(t) = Jv(t)$ separates into k subproblems, one for each of the Jordan blocks J_i in J. The diagonalizable case $J_i = (\lambda_i) \in \mathbb{C}^{1,1}$ has been treated in Section 9.3. By Theorems 14.1 and 14.2, we assume that a Jordan block $J_i = \begin{pmatrix} \lambda & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & \lambda \end{pmatrix}$ for A has size ℓ by ℓ for $\ell > 1$ and $\lambda \in \mathbb{C}$ and try to solve $w'(t) = J_i w(t)$ for $w(t) = \begin{pmatrix} w_1(t) \\ \vdots \\ w_\ell(t) \end{pmatrix}$.

Since

$$w'(t) \;=\; \begin{pmatrix} w_1'(t) \\ \vdots \\ w_\ell'(t) \end{pmatrix} \;=\; \begin{pmatrix} \lambda & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & \lambda \end{pmatrix} \begin{pmatrix} w_1(t) \\ \vdots \\ w_\ell(t) \end{pmatrix}\ , \tag{14.6}$$

we immediately note that $w_1'(t) = \lambda w_1(t)$ from row one and that $w_1(t) = ce^{\lambda t}$ according to elementary calculus. Next $w_2'(t) = w_1(t) + \lambda w_2(t)$ in row 2 of (14.6) is solved by $w_2(t) :=$

$cte^{\lambda t} = tw_1(t)$ since $w_2'(t) = ce^{\lambda t} + \lambda cte^{\lambda t} = w_1(t) + \lambda w_2(t)$ according to the product rule of differentiation. For arbitrary $j > 1$ we set $w_j(t) := c\dfrac{t^{j-1}}{(j-1)!}e^{\lambda t} = \dfrac{t}{j-1}w_{j-1}(t)$ and observe that

$$w_j'(t) = c\frac{t^{j-2}}{(j-2)!}e^{\lambda t} + \lambda c\frac{t^{j-1}}{(j-1)!}e^{\lambda t} = w_{j-1}(t) + \lambda w_j(t)$$

satisfies the j^{th} equation of (14.6) . Therefore for one Jordan block J_i of size ℓ by ℓ with $\ell > 1$ and the eigenvalue $\lambda \in \mathbb{C}$, the general solution to $w'(t) = J_i w(t)$ is given in vector form by

$$w(t) = ce^{\lambda t}\begin{pmatrix} 1 \\ t \\ t^2/2 \\ \vdots \\ \dfrac{t^{\ell-1}}{(l-1)!} \end{pmatrix} \in \mathbb{C}^n \ . \tag{14.7}$$

Next, a solution $v(t)$ to the Jordan normal form DE $v'(t) = Jv(t)$ with $J = \text{diag}(J_1, ..., J_k)$ can be concatenated from the k individual solutions to the Jordan block DEs $w'(t) = J_i w(t)$, $i = 1, ..., k$, that have just been described. Finally $x(t) := Uv(t)$ is the solution to the original problem $x'(t) = Ax(t)$ if $U^{-1}AU = J = \text{diag}(J_1, ..., J_k)$.

For real system matrices A with complex eigenvalues, the solution that we have just described involves complex functions for all eigenvalues λ_i of A that do not lie in $\mathbb{R}$. This is unfortunate because a complex solution $x(t)$ gives us little applicable information for the underlying real world problem. In Section 9.3, namely in equations (9.9), (9.10), and in Example 10 there, we have dealt with the case of one complex eigenvalue $\lambda = a + bi$ with $b \neq 0$ of index 1 for $A \in \mathbb{R}^{2,2}$. If we use the Real Jordan Normal Form Theorem 14.3 for complex eigenvalues of $A \in \mathbb{R}^{n,n}$, we can generalize the results from Section 9.3 to obtain real solutions for complex eigenvalues of A of index greater than 1.

If $J = \text{diag}(J_i)$ is the complex Jordan normal form of a real matrix A and if λ is a complex eigenvalue $\lambda = a + bi \notin \mathbb{R}$ for $a, b \in \mathbb{R}$ with index m exceeding 1, then for each associated Jordan block $J(\lambda)$ in J there is a corresponding equal sized Jordan block $J(\overline{\lambda})$ among the diagonal blocks J_i in J according to Section 14.2. Two such equal sized Jordan blocks $J(\lambda)$ and $J(\overline{\lambda})$ were mated into one real block in the real Jordan normal form of Theorem 14.3. Namely, the two blocks were fused into the real Jordan block $J_j =$

$$J(a,b,2m) = \begin{pmatrix} a & b & & & & \\ -b & a & & & & \\ 1 & 0 & \ddots & & & \\ 0 & 1 & & \ddots & & \\ & & \ddots & & \ddots & \\ & & & 1 \quad 0 & a & b \\ & & & 0 \quad 1 & -b & a \end{pmatrix}_{2m,2m}$$

. The case of index$(\lambda) = m = 1$ was treated in Section 9.3. In this section we try to find real solutions to $x' = Ax$ for eigenvalues $\lambda = a + bi$ of A whose indices exceed 1. In light of our earlier thoughts, the problem of solving $v'(t) = Jv(t)$ with $J = \text{diag}(J_k(\lambda_i),\ J_p(a_j, b_j, 2m_j))$ with real eigenvalues λ_i and complex eigenvalues $\lambda_j = a_j + b_j i$, some of whose indices exceed one, again separates into individual Jordan block problems. Having just solved $w' = J(\lambda)w$ for one real or complex eigenvalue λ and one Jordan block, we need to solve $w'(t) = J(a,b,2m)w(t)$ for w with one 'real Jordan block' of size $2m$ by $2m$ with $m > 1$ and $b \neq 0$ now. To do so we look at the real Jordan block analogue to (14.6) :

$$w'(t) = \begin{pmatrix} w_1'(t) \\ \vdots \\ w_{2m}'(t) \end{pmatrix} = \begin{pmatrix} a & b & & & & \\ -b & a & & & & \\ 1 & 0 & \ddots & & & \\ 0 & 1 & & \ddots & & \\ & & \ddots & & \ddots & \\ & & & 1 \quad 0 & a & b \\ & & & 0 \quad 1 & -b & a \end{pmatrix} \begin{pmatrix} w_1(t) \\ \vdots \\ w_{2m}(t) \end{pmatrix} . \tag{14.8}$$

According to (9.10), the top 2 by 2 system $\begin{pmatrix} w_1'(t) \\ w_2'(t) \end{pmatrix} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix}$ in (14.8) has the real solutions $w_1(t) = ce^{at}\sin(bt)$ and $w_2(t) = ce^{at}\cos(bt)$ for $t \in \mathbb{R}$. This can be quickly verified by applying the differentiation rules for the exponential, the sine, and the cosine.

Let us try to solve the system of DEs (14.8) with $2m = 4$ first in order to find the general pattern for $w_1'(t)$, ..., $w_{2m}'(t)$ that will solve (14.8) . We consider

$$\begin{pmatrix} w_1'(t) \\ w_2'(t) \\ w_3'(t) \\ w_4'(t) \end{pmatrix} = \begin{pmatrix} a & b & 0 & 0 \\ -b & a & 0 & 0 \\ 1 & 0 & a & b \\ 0 & 1 & -b & a \end{pmatrix} \begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \\ w_4(t) \end{pmatrix} . \tag{14.9}$$

Here the first two component functions w_1 and w_2 are known as $w_1(t) = ce^{at}\sin(bt)$ and $w_2(t) = ce^{at}\cos(bt)$. Consequently w_3 and w_4 must satisfy the DEs

$$w_3' = w_1 + aw_3 + bw_4 \quad \text{and} \quad w_4' = w_2 - bw_3 + aw_4 .$$

Taking our lead from the complex case, we multiply w_1 and w_2 by t and set $w_3(t) := cte^{at}\sin(bt) = tw_1(t)$ and $w_4(t) := cte^{at}\cos(bt) = tw_2(t)$. Using the product rule of differentiation twice on w_3 we obtain

$$w_3'(t) \;=\; c[e^{at}\sin(bt) + ate^{at}\sin(bt) + bte^{at}\cos(bt)] \;=\; w_1 + aw_3 + bw_4$$

precisely as desired. Likewise for w_4'. This extends to real Jordan blocks $J(a, b, 2m)$ of arbitrary dimensions $2m$: Simply set the two components $w_{2\ell-1}$ and $w_{2\ell}$ of the solution $w(t) = \begin{pmatrix} w_1(t) \\ \vdots \\ w_{2m}(t) \end{pmatrix}$ of (14.8) equal to

$$\begin{aligned} w_{2\ell-1} &:= c\frac{t^{\ell-1}}{(\ell-1)!}e^{at}\sin(bt) = \frac{t}{\ell-1}w_{2\ell-3} \qquad \text{and} \\ w_{2\ell} &:= c\frac{t^{\ell-1}}{(\ell-1)!}e^{at}\cos(bt) = \frac{t}{\ell-1}w_{2\ell-2} \end{aligned} \tag{14.10}$$

for any $1 < \ell \le \text{index}(\lambda) = m$.

Next we can synthesize the real solution $v(t)$ of $v'(t) = \text{diag}(J_k(\lambda_i),\ J_p(a_j, b_j, 2m_j))v(t)$ from its Jordan block DE parts and ultimately we find the real solution $x(t) := Uv(t)$ via the real principal vector matrix U that transforms $A \in \mathbb{R}^{n,n}$ to its real Jordan normal form $J = \text{diag}(J_k(\lambda_i),\ J_p(a_j, b_j, 2m_j)) = U^{-1}AU$ as before.

Example 9: Let us find a real solution to the linear system of differential equations $x'(t) = \begin{pmatrix} 4 & 7 & -1 & -6 \\ -2 & -3 & 2 & 3 \\ 0 & 1 & 2 & -1 \\ 0 & -1 & 0 & 1 \end{pmatrix} x(t) = Ax(t)$. Here the system matrix A that has been explored in Example 6 of Section 14.2.

A has the real Jordan normal form $J = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 \end{pmatrix}$ for its complex conjugate eigenvalues $\lambda = 1 \pm i$, both of index 2. This form is achieved by the real matrix similarity $J = X^{-1}AX$ with $X = \begin{pmatrix} 3 & 1 & 3 & -1 \\ 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}$, see Example 6. Thus we have to deal with a special case of equation (14.9) for $\lambda = a + bi = 1 + i$ here:

$$\begin{pmatrix} w_1'(t) \\ w_2'(t) \\ w_3'(t) \\ w_4'(t) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \\ w_4(t) \end{pmatrix}.$$

This equation is solved by $w(t) = c\begin{pmatrix} e^t \sin t \\ e^t \cos t \\ te^t \sin t \\ te^t \cos t \end{pmatrix}$ according to (14.10).

To obtain a real solution to $x'(t) = Ax(t)$, we transform the real Jordan nor-

mal form solution w to $x(t) = Xw(t) = c \begin{pmatrix} 3 & 1 & 3 & -1 \\ 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} e^t \sin t \\ e^t \cos t \\ te^t \sin t \\ te^t \cos t \end{pmatrix} =$
$ce^t \begin{pmatrix} 3\sin t + \cos t + t(3\sin t - \cos t) \\ -t\sin t \\ \sin t + \cos t + t\sin t \\ \sin t - t\cos t \end{pmatrix}$.
It should be verified that this actually solves $x' = Ax$ by using calculus differentiation rules on the left and matrix times vector multiplication on the right, see Problem 1 below. ◀

Remark 3: Throughout this chapter we have chosen Jordan blocks J that are lower triangular (block) matrices. Some prefer to express Jordan normal forms via the upper triangular Jordan blocks $\begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} = J_k^T = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix} J_k \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}$.
For such Jordan blocks a partial solution vector $w(t)$ to the linear DE $w' = J_k w$ should be replaced by its 'flipped brother' $\begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix} w(t)$, concatenated with other flipped partial solutions, and then transformed to obtain the solution $x(t)$ of $x'(t) = Ax(t)$ via the different order principal vector matrix Y that transforms A to its upper triangular matrix representation J^T. This variation becomes obvious from the proof of Corollary 5 in Section 14.2. See also our example in Teacher Problem T 14 in Section 14.1.P. ◀

(b) $\mathbf{m^{th}}$ *order linear differential equations with constant coefficients*

Systems $x' = Ax$ of n linear differential equations involve the first derivative of the component functions in $x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} : \mathbb{R} \to \mathbb{R}^n$. Now we study one differential equation for one function $x(t) : \mathbb{R} \to \mathbb{R}$ that involves, however, several order derivatives of x.

To solve the $\mathbf{m^{th}}$ **order differential equation** with constant coefficients

$$x^{(m)}(t) + a_{m-1}x^{(m-1)}(t) + \ldots + a_1 x'(t) + a_0 x(t) = 0 \qquad (14.11)$$

for $x(t) : \mathbb{R} \to \mathbb{R}$ we introduce m auxiliary functions

$$y_1(t) := x(t),\ y_2(t) = x'(t),\ \dots,\ \text{ and }\ y_m(t) = x^{(m-1)}(t)\ .$$

These functions allow us to rewrite (14.11) in terms of the vector valued auxiliary function $y(t) := \begin{pmatrix} y_1(t) \\ \vdots \\ y_m(t) \end{pmatrix}$ as follows:

$$\begin{aligned} y'(t) \; = \; \begin{pmatrix} y_1'(t) \\ \vdots \\ y_m'(t) \end{pmatrix} \; &= \; \begin{pmatrix} y_2(t) \\ \vdots \\ y_m(t) \\ -a_{m-1}y_m(t) - \dots - a_0 y_1(t) \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 & & 0 \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \\ -a_0 & & \dots & & -a_{m-1} \end{pmatrix} \begin{pmatrix} y_1(t) \\ \vdots \\ y_m(t) \end{pmatrix} \\ &= \; C(p)^T y(t)\ . \end{aligned}$$

This is a linear DE system with m differential equations. It has the transposed companion matrix $C(p)^T$ as its system matrix for the polynomial $p(r) = r^m + a_{m-1}r^{m-1} + \cdots + a_1 r + a_0$. Thus solving the m^{th} order differential equation (14.11) with constant coefficients reverts to solving the m by m system of linear DEs $y'(t) = C(p)^T y(t)$. This system can be solved by using the Jordan normal form approach, real if desired, on the system matrix $C(p)^T$ with the coefficients of p read off the m^{th} order equation (14.11) .

Example 10: To solve the third order differential equation $x'''(t) - x''(t) - 5x'(t) - 3x(t) = 0$ for $x(t) : \mathbb{R} \to \mathbb{R}$, we solve the following associated 3 by 3 system of linear DEs instead:

$$\begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 3 & 5 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} .$$

The system matrix is the transposed companion matrix for $p(r) = r^3 - r^2 - 5r - 3$ read off directly from the given third order DE. $p(r)$ has the roots $\lambda_1 = -1$ (double) and $\lambda_2 = 3$ for the corresponding eigenvectors $x_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ $(\lambda_1 = -1)$ and $x_3 = \begin{pmatrix} 1 \\ 3 \\ 9 \end{pmatrix}$ $(\lambda_2 = 3)$. We note that $C(p)^T$ is not diagonalizable, having but one eigenvector x_2 for its repeated eigenvalue -1. To find a principal vector of order 2 for $\lambda_1 = -1$, we look at $\ker\big((A+I)^2\big)$. The matrix $(A+I)^2 = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 6 & 3 \\ 9 & 18 & 9 \end{pmatrix}$ has $x_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ in its kernel. Next, $(A+I)x_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = x_2$ is the eigenvector of $C(p)^T$ found earlier for $\lambda_1 = -1$. Thus taking the three vectors $x_1,\ x_2$, and x_3 into a basis of $\mathbb{R}^3$ and forming their column matrix X, we know that $J = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix} = X^{-1}C(p)^T X$ is the real Jordan normal form of $C(p)^T$. The solution of the Jordan normal form DE is $w(t) = \begin{pmatrix} ce^{-t} \\ cte^{-t} \\ ke^{3t} \end{pmatrix}$ for arbitrary independent constants c and k according to subsection (a). Since $X = \begin{pmatrix} | & & | \\ x_1 & \dots & x_3 \\ | & & | \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 3 \\ -1 & 1 & 9 \end{pmatrix}$ transforms $C(p)^T$ to J, the solution to the equivalent companion matrix form DE $y' = C(p)^T y$ is given by $y(t) = Xw(t) = \begin{pmatrix} ce^{-t}(1+t) + ke^{3t} \\ -cte^{-t} + 3ke^{3t} \\ ce^{-t}(t-1) + 9ke^{3t} \end{pmatrix}$.

And y's first component $x(t) = y_1(t) = ce^{-t}(1+t) + ke^{3t}$ solves the given third order differential equation.

Students should verify this result directly for the original third order differential equation using the rules of calculus. ◀

(c) *Numerical computations of Jordan forms and MATLAB*

MATLAB – and for that matter, all mathematical software – is nearly *incapable* of computing the Jordan normal form, real or complex, of any non diagonalizable matrix A_{nn}, unless n is very small ($n \leq 4$ or 5) if A is dense, or unless A of larger size does break up into several such small dense diagonal blocks and unless A has integer or rational coefficients. This is due to two separate problems with finding a non diagonal Jordan normal form from A.

For one, non diagonalizable matrices must have repeated eigenvalues. All but the smallest sized, integer or rational entry matrices require numerical approximation to find their eigenvalues. There are no "algebraic formulas" such as the 'quadratic formula' for second degree polynomials in Appendices A and B for polynomials of degrees exceeding 4. Therefore, when the eigenvalues of a matrix A_{nn} have been computed approximately, who would be willing or able to decide correctly that the two computed eigenvalues $\lambda_1 = 1.000000000123$ and $\lambda_2 = 0.9999999976$, say, are in reality distinct eigenvalues of A, or that they approximate the 'true' double eigenvalue $\lambda = 1.0000000000438756$ of A?

Even if the first problem of eigenvalue multiplicity could be settled satisfactorily for a given matrix A, the next problem is to reliably determine the dimensions of the principal subspaces $P_j(\lambda) = \ker(A - \lambda I)^j$ for a repeated eigenvalue λ of A. Even the most simple question in this realm, namely whether a repeated eigenvalue λ has index 1 or larger, is generally impossible to decide since the rank of $A - \lambda I$ cannot be computed precisely. For sizable matrices A_{nn}, not even the SVD of Chapter 12 helps much, since it requires a *personal decision* whether the next to last computed singular value $\sigma_{n-1} = 1.23 \cdot 10^{-7}$ of $A - \lambda I$ for example should be taken to be zero, or whether only $\sigma_{n-1} = 2.0237 \cdot 10^{-15}$ should be, giving λ an index greater than 1. Of course most often, not even the smallest singular value σ_n of $A - \lambda I$ will be computed exactly as zero for any approximately computed eigenvalue λ of A, see the problem section.

For properly constructed low dimensional integer matrices $A_{n,n}$ with integer Jordan normal forms and bases (see the Teacher's Problem T 14 in Section 14.1.P), we can, however, use MATLAB successfully to simplify some of the tedious tasks of finding eigenvalues and Jordan bases for A:
First, we can compute the eigenvalues λ_i of A in MATLAB via `eig(A)`. If there are no repeats, the Jordan normal form of A is $\text{diag}(\lambda_i) \in \mathbb{C}^{n,n}$. And the real Jordan normal form of $A \in \mathbb{R}^{n,n}$ is $\text{diag}(\lambda_\ell, J(Re(\lambda_k), Im(\lambda_k), 2))$ where the λ_ℓ denote the real eigenvalues of A and the λ_k are selected from each pair of complex conjugate eigenvalues $\lambda_j = \overline{\lambda_j}$ of A. An eigenvector basis for A can best be found from `rref(A-λ*eye(n))`. For a real eigenvalue λ we then solve the homogeneous system $A - \lambda I = 0$ according to Chapter 3 by hand. Invoking `null(A-λ*eye(n))` instead in MATLAB gives us an orthonormal basis for the kernel and is generally useless for our purpose to work within the integers. For a complex eigenvalue we also solve $A - \lambda I = 0$ from the complex RREF that was computed by MATLAB as before. The complex eigenvalue and its conjugate serve in the eigenvector basis for A. If we look for the real Jordan normal form instead for a real matrix A, then we take the real and the imaginary parts vectors into the Jordan basis for A.
If A has a repeated eigenvalue λ with algebraic multiplicity $k > 1$ according to the result of `eig(A)`, we need to compute the RREFs of $(A - \lambda I)^\ell$ for $\ell = 1, 2, 3, ..., k$ via MATLAB until the number of free variables in these does not increase any longer. This gives us the index m of λ. Next we pick a vector $x^{(m)} \in P_m(\lambda) - P_{m-1}(\lambda)$ and iterate down to the eigenvector $x^{(1)} = (A - \lambda I)^{m-1} x^{(m)}$. This defines a maximal length principal vector chain for λ. We might have to repeat this process for other principal vector chains for λ or A, but remember that the size of A is rather small in our examples that are fit for hand computations. The computed principal vector chain(s) form part of the possibly complex Jordan basis of A. If $A \in \mathbb{R}^{n,n}$ and $\lambda \in \mathbb{C}$, then the real and imaginary parts vectors of each $x^{(j)}$ above should be chosen instead to obtain the desired real Jordan basis of A. These vectors generate the real Jordan block $J(Re(\lambda), Im(\lambda), 2m)$ associated with the pair λ and $\overline{\lambda}$ of index m.
This process leads to the standard or the real Jordan normal form of any properly constructed $A \in \mathbb{R}^{n,n}$ over the integers and to its respective Jordan basis $\mathcal{X}$.

If the Jordan form J of A has been computed as a first step in solving the linear differential equation $x'(t) = Ax(t)$, we solve the Jordan normal form DE $w'(t) = Jw(t)$ using our formulas (14.7) for real eigenvalues and (14.10) for complex ones. And from X, the column vector matrix of the Jordan basis that transforms A to $J = X^{-1}AX$, we then find the solution $x'(t) = Xw(t)$ for $x'(t) = Ax(t)$. These latter steps are best carried out by hand. This process is relatively easy if A is properly constructed according to Teacher's problem T 14; see our exercises below and the solution set.

However in general, unless A has well separated eigenvalues and thus is diagonalizable, the Jordan structure of a general matrix A cannot be derived reliably from numerical computations. The main purpose of this chapter is to give us a theoretical tool that lets us glimpse the similarity invariants of matrices. The Jordan normal form helps us classify matrices and lets us study low dimensional linear DE systems and find their theoretical solutions.
There is, however, some justice. Probabilisticly speaking, every random matrix A_{nn} must have n distinct eigenvalues and therefore every random square matrix is diagonalizable with probability 1. The non diagonalizable matrices form a set of measure zero in matrix space $\mathbb{R}^{n,n}$ and our picture of matrix space in Figure 11.1 of Section 11.2 surely gives too much territory to the non diagonalizable matrices such as J on the right side of the Figure. For random matrices $A \in \mathbb{R}^{n,n}$ or $\mathbb{C}^{n,n}$, the assignment

$$A \longrightarrow \Lambda(A) := \{\lambda_i \mid \lambda_i \text{ is an eigenvalue of } A\} \subset \mathbb{C}$$

of A to its eigenvalue set $\Lambda(A)$ must create random (except for real axis symmetry, if A is real) n points as images, potentially with repetitions. Repetition of an eigenvalue, however, is unlikely since when making n marks randomly in the complex plane, it is unlikely to mark the same spot twice. Incidentally, for the same reason it is highly unlikely that a random matrix will be singular, or have the precise eigenvalue zero.

There is one drawback to these heuristic probability considerations in what amounts to our willful 'human touch'. Through our studies we have become acquainted with many matrices with repeated eigenvalues such as I_n or O_n. We can write down non diagonalizable Jordan normal forms with ease. And dense matrices whose eigenvalues do not repeat seemingly take an effort to construct. We simply are not good random matrix generators. But nature is.

14.3.P Problems

1. (a) Verify that the solution $x(t)$ computed in Example 9 satisfies the DE $x'(t) = Ax(t)$.
   (b) Repeat for Example 10.
2. For a number of randomly generated 5 by 5 matrices, test whether the smallest singular value $\sigma_n(\lambda_i)$ of $A - \lambda_i I$ is actually zero in MATLAB for any of the eigenvalues $\lambda_1, ..., \lambda_n$ of A.
3. Generate ten random 10 by 10 and 100 by 100 matrices each and determine whether the matrix is singular or not. Do so by using (a) the `rref`, (b) the `eig`, and (c) the `svd` functions of MATLAB. Do you obtain consistent results?
4. Repeat the previous problem for ten random 10 by 10 and 100 by 100 matrices each, that you doctor so that their last row is the sum of the first three rows of the given random matrix. Do you obtain consistent results across the three MATLAB functions in Problem 3?

5. Repeat Problem 3 for the matrices A and B of the Example in Teacher Problem T 14 in Section 14.1.P. Use the MATLAB computed eigenvalues of A and B, as well as the theoretically accurate ones and compare.

6. Let $A_c := \begin{pmatrix} c & -1 & 0 \\ 1 & c & 1 \\ 0 & 1 & c \end{pmatrix}$.
   (a) For which value of c does MATLAB compute the eigenvalues of A_c as repeated and for which values of c as distinct?
   (b) For which value of c are the eigenvalues of A_c repeated or distinct?

7. Find the real solution of the DE $x'(t) = Ax(t)$ with $A = \begin{pmatrix} 1 & -6 & -3 & -1 \\ 4 & 7 & 12 & -3 \\ -3 & -3 & -7 & 2 \\ -2 & -5 & -6 & 1 \end{pmatrix}$ from Problem 1 of Section 14.2.P.

8. Find the real Jordan normal form of $B = \begin{pmatrix} 3 & -3 & -1 & 8 \\ 1 & -2 & -7 & 25 \\ -3 & -1 & -8 & 29 \\ 0 & -1 & -3 & 11 \end{pmatrix}$.
   (Hint: B has the eigenvalue $\lambda = 1 - 2i$.)

9. Find the real solution of the DE $x'(t) = Bx(t)$ with B from the previous problem.

10. Convert the 4^{th} order DE $x^{(4)}(t) - 6x^{(3)} + 3x'(t) = 4x(t)$ into a system of linear DEs.

11. Solve the 5^{th} order DE $y^{(5)}(t) - 5y^{(4)}(t) + 5y^{(3)}(t) + 5y^{(2)}(t) - 6y'(t) = 0$.

12. Solve the 5^{th} order DE $x^{(5)}(t) + x^{(4)}(t) - 9x^{(3)}(t) - 5x^{(2)}(t) + 16x'(t) + 12x(t) = 0$.

14.R Review Problems

1. For the complex or real Jordan normal form J of a matrix A prove:
   (a) $\text{rank}(J) = \text{rank}(J^T)$;
   (b) $\det(J) = \det(A)$.
   (c) $f_J = f_{J^T}$ for the characteristic polynomial $f_{..}$.
   (d) $p_J = p_{J^T}$ for the minimum polynomial $p_{..}$.
   (e) $f_A = f_{A^T}$.
   (f) $p_A = p_{A^T}$.
   (g) $f_A = f_J$.
   (h) $p_A = p_J$.

2. Show that the number of Jordan blocks for one eigenvalue λ of A is equal to the geometric multiplicity of the eigenvalue λ.
   How is the geometric multiplicity of an eigenvalue defined?

3. (a) Show that the index of an eigenvalue λ of A is equal to the size of the largest Jordan block for λ in the Jordan normal form of A.
   (b) Should the above result be modified when talking about the real Jordan normal form of A instead? How?

4. Rephrase and reprove the previous problem for the <u>real</u> Jordan normal form of $A \in \mathbb{R}^{n,n}$.

5. If $x \in \mathbb{R}^n$ is a principal vector of order $k > 2$ for $\lambda \in \mathbb{R}$ and A, show that $y := (A - \lambda I)^2 x$ is a principal vector of order $k - 2$ for λ and A.

6. For which real square matrices A do the Jordan normal form and the real Jordan normal form coincide?

7. Which matrices have a diagonal companion matrix normal form?

8. (a) Construct a matrix A with the characteristic polynomial $f_A(\lambda) = \lambda(\lambda + 2)^2(\lambda - 3)^5$. What is A's size?
   (b) Construct a matrix A with the minimum polynomial $p_A(x) = (x + 2)^2(x - 3)^5$. What are A's possible sizes?

9. Construct 7 by 7 and 12 by 12 matrices A and B, respectively, that have the minimum polynomial $x^3(x+2)^4$.

10. Find the minimum and the characteristic polynomials $p_A(x)$ and $f_A(x)$ for a matrix A with the Jordan normal form $J = \operatorname{diag}(J(2,4),\ J(2,3),\ J(2,1),\ J(-1,3),\ J(-1,5),\ J(0,1))$.

11. What is the minimal size of a matrix B whose minimum polynomial is that of the matrix J in the previous problem? What is the possible maximal size of such a B?

12. Assume that $A \in \mathbb{R}^{n,n}$ has the Jordan normal form $J = \operatorname{diag}(J_i) \in \mathbb{C}^{n,n}$.
    (a) Show that A^2 is similar to J^2.
    (b) For a Jordan block $J_i = J(\lambda, k) \in \mathbb{C}^{n,n}$ compute J_i^2.
    (c) What are the eigenvalues and eigenvectors of $J_i(\lambda, k)^2$?
    (Careful: there are two different answers for $\lambda = 0$ and $\lambda \neq 0$.)

13. Find the Jordan normal form of A^2 if A has the Jordan normal form $J = \operatorname{diag}(J_i)$.
    (Hint: Make use of the previous problem.)

Standard Questions and Tasks:

1. What is a principal subspace and a principal vector for a square matrix?
2. What does the order of a principal vector indicate?
3. What is the maximal order that is possible for a principal vector and an eigenvalue of index j? Must such a maximal order principal vector exist?
4. Find a principal vector chain of maximal order for a given eigenvalue of a matrix A.
5. When is a matrix A_{nn} diagonalizable, when is it not?
6. Find the Jordan normal form of a given matrix.
7. Write out the Jordan diagram for a given sequence of principal subspace dimensions for one eigenvalue λ of a given matrix A_{nn} and determine the Jordan blocks associated with λ .

Subheadings of Lecture Fourteen :

Basic Equations :

p. W-2 $P_k(\lambda) = \ker(A - \lambda I)^k$ (Principal subspace)

p. W-3 $x^{(k)} \in P_k(\lambda)$ with $x^{(k)} \notin P_{k-1}(\lambda)$ (Principal vector of order k)

p. W-3 $x^{(k)},\ x^{(k-1)} = (A - \lambda I)x^{(k)},\ \ldots$
$\ldots,\ x^{(1)} = (A - \lambda I)^{k-1}x^{(k)}$ (Principal vector chain)

Inner city space

Appendix D

Inner Product Spaces

The inner product, taken of any two vectors in an arbitrary vector space, generalizes the dot product of two vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$.

For two column vectors x and $y \in \mathbb{R}^n$ we can form two different vector products, namely

- the **outer product**

$$xy^T \;:=\; \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} y_1, ..., y_n \end{pmatrix} \;=\; \begin{pmatrix} x_1y_1 & \dots & x_1y_n \\ \vdots & & \vdots \\ x_ny_1 & \dots & x_ny_n \end{pmatrix} \;\in\; \mathbb{R}^{n,n}$$

and

- the standard **inner product**

$$x^Ty \;:=\; \begin{pmatrix} x_1, ..., x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \;=\; x_1y_1 + ... + x_ny_n \;\in\; \mathbb{R}\;.$$

Here we interpret the two respective vectors as 1 by n or n by 1 matrices and multiply according to the rules of matrix multiplication. The outer product is a **dyadic product** since it creates an n by n **dyad** from two vectors, of which the first appears in column and the second in row form. It allows us to express **matrix multiplication** as a sum of rank 1 dyadic generators; see Sections 6.2, 10.2, and the proof of Lemma 3 in Section 12.2. A matrix product can be written as the sum of dyads of the columns and rows of the two matrix factors as follows:

$$
\begin{aligned}
A_{mn}B_{nk} &= \begin{pmatrix} | & & | \\ a_1 & \dots & a_n \\ | & & | \end{pmatrix} \begin{pmatrix} - & b_1 & - \\ & \vdots & \\ - & b_k & - \end{pmatrix} \\
&= \begin{pmatrix} | \\ a_1 \\ | \end{pmatrix} \begin{pmatrix} - & b_1 & - \end{pmatrix} + \ldots + \begin{pmatrix} | \\ a_n \\ | \end{pmatrix} \begin{pmatrix} - & b_k & - \end{pmatrix} \in \mathbb{R}^{m,k} ,
\end{aligned}
$$

where the vectors a_i denote the columns of the first factor A and the b_j denote the rows of the second factor B.

The standard inner product $x^T y$ of two vectors in $\mathbb{R}^n$ is the same as the **dot product** $x \cdot y \in \mathbb{R}$ of the two vectors. It was introduced in Chapter 1 and interprets the first factor as a row and the second one as a column vector. The inner or dot product is also handy to express matrix multiplication, namely

$$
A_{mn}B_{nk} = \begin{pmatrix} - & \widetilde{a_1} & - \\ & \vdots & \\ - & \widetilde{a_m} & - \end{pmatrix} \begin{pmatrix} | & & | \\ \widetilde{b_1} & \dots & \widetilde{b_k} \\ | & & | \end{pmatrix} = \begin{pmatrix} \widetilde{a_1} \cdot \widetilde{b_1} & \dots & \widetilde{a_1} \cdot \widetilde{b_k} \\ \vdots & & \vdots \\ \widetilde{a_m} \cdot \widetilde{b_1} & \dots & \widetilde{a_m} \cdot \widetilde{b_k} \end{pmatrix} \in \mathbb{R}^{m,k} ,
$$

where the $\tilde{a}_i$ now denote the rows of A and the $\tilde{b}_j$ the columns of the second factor B. In addition, the inner or dot product helps define angles and orthogonality of two vectors in $\mathbb{R}^n$, see Chapters 10 through 12.

We start by listing four fundamental properties of the standard inner product of two vectors $.. \cdot .. : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

Proposition 1: (**Real Inner Product**)

The standard inner product of two vectors x and $y \in \mathbb{R}^n$ is defined as

$$
x \cdot y \ := \ x^T y \ = \ \begin{pmatrix} x_1, \ldots, x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i \ \in \ \mathbb{R} .
$$

It satisfies the following four properties:

(a) $x \cdot y \ = \ y \cdot x$ for all $x, y \in \mathbb{R}^n$;

(b) $x \cdot (y + z) \ = \ x \cdot y + x \cdot y$ for all $x, y, z \in \mathbb{R}^n$;

(c) $(\alpha x) \cdot y \ = \ x \cdot (\alpha y) \ = \ \alpha(x \cdot y)$ for all $x, y \in \mathbb{R}^n$ and all $\alpha \in \mathbb{R}$.

(d) $x \cdot x \ \geq \ 0$ for all $x \in \mathbb{R}^n$, and $x \cdot x = 0 \in \mathbb{R}$ if and only if $x = 0 \in \mathbb{R}^n$. ◀

For two **complex vectors** $x, y \in \mathbb{C}^n$, several modifications are in order in the definition and the properties of an inner product due to the effects of complex conjugation, see Appendix A.

Proposition 2: (**Complex Inner Product**)

The standard inner product of two vectors x and $y \in \mathbb{C}^n$ is defined as

$$x \cdot y \; := \; x^* y \; = \; \left(\; \overline{x_1}, ..., \overline{x_n} \; \right) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \; = \; \sum_{i=1}^{n} \overline{x_i} y_i \; \in \; \mathbb{C} \; .$$

It satisfies the following four properties:

(a) $x \cdot y \; = \; \overline{y \cdot x}$ for all $x, y \in \mathbb{C}^n$;

(b) $x \cdot (y + z) \; = \; x \cdot y + x \cdot y$ for all $x, y, z \in \mathbb{C}^n$;

(c) $(\alpha x) \cdot y \; = \; x \cdot (\overline{\alpha} y) \; = \; \alpha(x \cdot y)$ for all $x, y \in \mathbb{C}^n$ and all $\alpha \in \mathbb{C}$.

(d) $x \cdot x \; \geq \; 0$ for all $x \in \mathbb{C}^n$, and $x \cdot x = 0 \in \mathbb{C}$ if and only if $x = 0 \in \mathbb{C}^n$. ◀

Proof: We deduce the four properties for the complex inner product only.

The properties of the real inner product in Proposition 1 follow immediately by dropping all complex conjugation bars in this proof.

(a) $x \cdot y \; = \; \sum_i \overline{x_i} y_i \; = \; \overline{\sum_i x_i \overline{y_i}} \; = \; \overline{\sum_i \overline{y_i} x_i} \; = \; \overline{y \cdot x}$ since double conjugation $\overline{\overline{c}}$ gives c back for any c in $\mathbb{C}$.

(b) $x \cdot (y + z) \; = \; \sum_i \overline{x_i}(y_i + z_i) \; = \; \sum_i \overline{x_i} y_i + \sum \overline{x_i} z_i \; = \; x \cdot y + x \cdot z$.

(c) $(\alpha x) \cdot y \; = \; \sum_i \overline{(\alpha x_i)} y_i \; = \; \overline{\alpha} \sum_i \overline{x_i} y_i \; = \; \overline{\alpha} x \cdot y$ and $\sum_i \overline{(\alpha x_i)} y_i \; = \; \sum_i \overline{x_i}(\overline{\alpha} y_i) \; = \; x \cdot (\overline{\alpha} y)$.

(d) $x \cdot x = \sum_i |x_i|^2 \geq 0$ as the sum of real squares. And equality holds precisely when $|x_i| = 0$ for each $i = 1, ..., n$, or when $x = 0 \in \mathbb{C}^n$. ■

The standard dot or inner product of $\mathbb{R}^n$ (or $\mathbb{C}^n$) serves very well in many aspects of linear algebra, such as when defining angles, orthogonality, and the length of vectors. More generally, an inner product can be defined in an arbitrary vector space V by requiring the four properties of a dot product; see Appendix C for abstract vector spaces.

Definition 1: Let V be an arbitrary vector space over a field of scalars $\mathbb{F}$.

(1) A function $\langle .., .. \rangle : V \times V \to \mathbb{F}$ that maps any two vectors f and $g \in V$ to the scalar $\langle f, g \rangle$ in $\mathbb{F}$ is **bilinear** if $\langle .., .. \rangle$ is linear in each of its arguments, i.e., if $\langle \alpha f + \beta g, h \rangle = \langle \alpha f, h \rangle + \langle \beta g, h \rangle$ and $\langle x, \delta u + \epsilon v \rangle = \langle x, \delta u \rangle + \langle x, \epsilon v \rangle$ for all scalars $\alpha, \beta, \delta, \epsilon \in \mathbb{F}$ and all vectors $x, u, v, f, g, h \in V$.

(2) A bilinear function $\langle .., .. \rangle : V \times V \to \mathbb{C}$ (or $\mathbb{R}$), operating on a complex (or real) vector space V, is an **inner product** on V if it satisfies the following four properties.

(a) $\langle x, y \rangle \; = \; \overline{\langle y, x \rangle}$ for all $x, y \in V$;

(b) $\langle x, (y + z) \rangle \; = \; \langle x, y \rangle + \langle x, y \rangle$ for all $x, y, z \in V$;

(c) $\langle (\alpha x), y \rangle \; = \; \langle x, (\overline{\alpha} y) \rangle \; = \; \alpha \langle x, y \rangle$ for all $x, y \in V$ and all $\alpha \in \mathbb{C}$.

(d) $\langle x, x\rangle \geq 0$ for all $x \in V$, and $\langle x, x\rangle = 0 \in \mathbb{C}$ if and only if $x = 0 \in V$. ◀

Note that if V is a vector space with the scalar field $\mathbb{R}$, then the complex conjugation in parts 2(a) and 2(c) above should simply be dropped.

For $V = \mathbb{R}^n$ and two vectors $x, y \in \mathbb{R}^n$, the standard dot product $\langle x, y\rangle := x \cdot y$ clearly defines an inner product on $\mathbb{R}^n$. We can express the dot product as $x \cdot y = x^T I y$ via the n by n identity matrix I. The matrix I_n is symmetric with n positive eigenvalues equal to 1 on its diagonal. Positive definite matrices generalize the properties of I; see Section 11.3. For example, every positive definite matrix $P = P^T \in \mathbb{R}^{n,n}$ can be expressed as a matrix product $P = A^T A$ with a nonsingular real square matrix A. For any $P = A^T A$ that is positive definite, we may set $\langle x, y\rangle_P := x^T P y = x^T A^T A y = (Ax)^T \cdot (Ay)$ and thereby obtain an inner product $\langle x, y\rangle_P := \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that differs from the ordinary dot product $x \cdot y = \langle x, y\rangle_I$. All one needs to do to verify this statement, is to show that $\langle .., ..\rangle_P$ is bilinear and satisfies the four standard properties of an inner product of Definition 1, see Problem 2 below.

Proposition 3: (a) Every bilinear form $f(x, y) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ can be expressed as $f(x, y) = x^T S y$ for a real symmetric n by n matrix S.

(b) If $P = P^T \in \mathbb{R}^{n,n}$ is a positive definite real matrix, then $\langle x, y\rangle_P := x^T P y \in \mathbb{R}$ defines an inner product on $\mathbb{R}^n$.

(c) If $P = P^* \in \mathbb{C}^{n,n}$ is a positive definite complex matrix, then $\langle x, y\rangle_P := x^* P y \in \mathbb{C}$ defines an inner product on $\mathbb{C}^n$. ◀

Example 1: (a) To write $f(x, y) = 3x_1y_1 - 2x_1y_2 + 3x_3y_2 - x_2y_2 + 4x_4y_4 \ : \ \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$ in the form $x^T S y$ with $S = S^T \in \mathbb{R}^{4,4}$, we define the diagonal entries s_{ii} of S as the coefficients of x_iy_i in f, or as $s_{11} = 3,\ s_{22} = -1,\ s_{33} = 0$, and $s_{44} = 4$. For $i > j$ we set the off-diagonal entries s_{ij} in S equal to half the coefficient of x_iy_j and then define $s_{ji} = s_{ij}$. Thus $s_{12} = -1 = s_{21}$ and $s_{32} = 1.5 = s_{23}$. Thus

$$S = \begin{pmatrix} 1 & -1 & & \\ -1 & 0 & 1.5 & \\ & 1.5 & -1 & \\ & & & 4 \end{pmatrix} = S^T.$$

(b) Determine whether $\langle x, y\rangle := x^T \begin{pmatrix} 10 & 3 \\ 3 & 1 \end{pmatrix} y$ is an inner product on $\mathbb{R}^2$.

Clearly $\langle x.y\rangle$ is bilinear in both x and $y \in \mathbb{R}^2$ since it is matrix generated. The generating matrix $S := \begin{pmatrix} 10 & 3 \\ 3 & 1 \end{pmatrix}$ with $\langle x.y\rangle = \langle x.y\rangle_S = x^T S y$ is symmetric, i.e., $S = S^T$, and hence its eigenvalues are real according to Section 11.1. Using the trace and determinant conditions of Theorem 9.5, we observe that the two eigenvalues of S add to 11 and multiply to $10 - 9 = 1$. Thus both eigenvalues of S must be positive real, making $S = S^T$ positive definite and $\langle x.y\rangle = \langle x.y\rangle_S$ an inner product on $\mathbb{R}^2$ according to Proposition 3.

(c) Determine whether $\langle u, v\rangle := u^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} v$ is an inner product on $\mathbb{C}^2$.

Clearly the function $\langle u.v\rangle$ maps any two complex 2-vectors to a complex number and it is bilinear. For $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = X^*$ we observe that $X \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $X \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = - \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Therefore X has the eigenvalues 1 and –1 and X is not positive definite. Thus $\langle u, v\rangle = \langle u, v\rangle_X$ is not necessarily an inner product on $\mathbb{C}^n$ since Proposition 3 does not apply. In fact, $\langle u, v\rangle_X$ violates the fourth property (d) of inner products for $u = v = e_1 \neq 0 \in \mathbb{C}^2$:

$$\langle e_1, e_1\rangle_X \;=\; \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \;=\; \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \;=\; 0 \;\in\; \mathbb{C}\;.$$

Therefore $\langle u, v\rangle_X$ is not an inner product on $\mathbb{C}^2$. ◀

Inner products can help us measure and navigate in abstract vector spaces, such as in spaces of functions. As an example, we now consider the space of continuous functions $\mathcal{F}_{[0,1]} := \{f \;:\; [0, 1] \to \mathbb{R} \mid f \text{ continuous}\}$ defined on the interval $[0, 1] \subset \mathbb{R}$. This space is infinite dimensional, see Section 7.2(b). By setting $\langle f, g\rangle := \int_0^1 f(x)g(x)\, dx$ for all functions $f, g \in \mathcal{F}_{[0,1]}$, we have made $\mathcal{F}_{[0,1]}$ into an inner product space. Clearly $\langle f, g\rangle$ is linear in both of its function variables f and g since integration is linear in the sense that $\int u + v\; dx = \int u\; dx + \int v\; dx$. Identities such as

$$\langle \alpha u + \beta v, w\rangle = \int (\alpha u + \beta v)w\; dx = \alpha \int uw\; dx + \beta \int vw\; dx = \alpha\langle u, w\rangle + \beta\langle v, w\rangle$$

prove the properties (b) and (c) of an inner product. Next we observe that property (a) holds since $\langle f, g\rangle = \int fg\; dx = \int gf\; dx = \langle g, f\rangle$. And the first part of property (d) $\langle f, f\rangle = \int_0^1 f^2(x)\; dx \geq 0$ holds for any integrable function f since $f^2(x) \geq 0$. To show that $\langle f, f\rangle = 0$ for $f \in \mathcal{F}_{[0,1]}$ implies that $f = 0$ on $[0, 1]$ requires more thought: Every function $f \in \mathcal{F}_{[0,1]}$ is continuous. If f is not the zero function on $[0, 1]$, then $f(x_0) \neq 0$ for some $x_0 \in [0, 1]$. By continuity there is an interval $[a, b]$, $0 \leq a < b \leq 1$, with $x_0 \in [a, b]$ and $f(x) > \epsilon > 0$ for some given $\epsilon > 0$ and all $x \in [a, b]$. Using the additivity of the integral over its domain of integration, we observe that

$$\begin{aligned} \langle f, f\rangle \;&=\; \int_0^1 f^2(x)\; dx \;=\; \int_0^a f^2\; dx + \int_a^b f^2\; dx + \int_b^1 f^2\; dx \\ &\geq\; \int_a^b f^2\; dx \;\geq (b-a)\epsilon^2 \;>\; 0\;. \end{aligned}$$

Consequently if f is continuous and $\langle f, f\rangle = 0$, then $f = 0$ on $[0, 1]$. Thus

$$\langle f, g\rangle \ := \ \int_0^1 f(x)g(x)\ dx$$

is an inner product on $\mathcal{F}_{[0,1]} = \{f : [0,1] \to \mathbb{R} \mid f \text{ continuous }\}$. Different inner product can be defined on $\mathcal{F}_{[0,1]}$ by setting

$$\langle f, g\rangle_w \ := \ \int_0^1 f(x)g(x)w(x)\ dx$$

for an arbitrary continuous **weight function** $w \in \mathcal{F}_{[0,1]}$ that is positive on $[0,1]$, see Problem 5.

Inner products define angles and orthogonality in abstract vector spaces V just as the standard dot products do in $\mathbb{R}^n$ and $\mathbb{C}^n$; recall Section 10.1. If both f and $g \neq 0 \in V$ and V is an inner product space with the inner product $\langle .., ..\rangle$, then the **angle** between f and g is defined with respect to a given inner product $\langle .., ..\rangle$ by the formula

$$\cos\angle(f,g) \ := \ \frac{\langle f, g\rangle}{\langle f, f\rangle^{\frac{1}{2}}\langle g, g\rangle^{\frac{1}{2}}}\ .$$

And $f \in V$ is **orthogonal** to $g \in V$, or $f \perp g \in V$, if $\langle f, g\rangle = 0$ for the inner product $\langle .., ..\rangle$ of V. Here $\langle f, g\rangle$ may be complex for two functions f and g when V is a complex vector space. In this case the complex valued cosine function $\cos(z)$ is used.

Example 2: (a) In $V = \mathcal{F}_{[0,1]}$ with the inner product $\langle f, g\rangle := \int_0^1 2f(x)g(x)\ dx$, find the angle between the two functions $f(x) = 1$ and $g(x) = x \in V$.
To find the angle we have to evaluate three different inner products: $\langle f, g\rangle = \int_0^1 2x\ dx = x^2\ \big|_0^1 = 1$, $\langle f, f\rangle = \int_0^1 2\ dx = 2x\ \big|_0^1 = 2$, and $\langle g, g\rangle = \int_0^1 2x^2\ dx = \frac{2x^3}{3}\ \big|_0^1 = \frac{2}{3}$. Therefore $\cos\angle(f,g) = \dfrac{\langle f, g\rangle}{\langle f, f\rangle^{\frac{1}{2}}\langle g, g\rangle^{\frac{1}{2}}} = \dfrac{1}{\sqrt{2}\sqrt{\frac{2}{3}}} = \dfrac{\sqrt{3}}{2}$. And the angle between f and g has the radian measure of $\arccos\left(\frac{\sqrt{3}}{2}\right)$. Note that the inner product contains the **weight function** $w(x) = 2$.

(b) Show that the two functions $h(x) = 1$ and $k(x) = x - \frac{1}{2}$ are orthogonal in V of part (a) with its given inner product.
We evaluate $\langle h, k\rangle = \int_0^1 2(x - \frac{1}{2})\ dx = \int_0^1 2x\ dx - \int_0^1\ dx = x^2\ \big|_0^1\ - x\ \big|_0^1 = 1 - 1 = 0$.

(c) Find an orthogonal basis for the subspace $\text{span}\{x, x^2\} \subset \mathcal{F}_{[0,1]}$ with respect to the inner product $\langle f, g\rangle := \int_0^1 f(x)g(x)\ dx$.

Here we use the modified Gram–Schmidt process of Chapter 10 for the functions $u_1 = x$ and $u_2 = x^2$.

$$\begin{aligned} v_1 &:= u_1 = x \; ; \\ v_2 &:= \langle v_1, v_1\rangle u_2 - \langle u_2, v_1\rangle v_1 \; . \end{aligned}$$

We have $\langle v_1, v_1\rangle = \int_0^1 x^2\ dx = \frac{x^3}{3}\ \Big|_0^1 = \frac{1}{3}$ and $\langle u_2, v_1\rangle = \int_0^1 x^3\ dx = \frac{x^4}{4}\ \Big|_0^1 = \frac{1}{4}$. Thus $v_1 = x$ and $v_2 = \frac{1}{3}u_2 - \frac{1}{4}v_1 = \frac{1}{3}x^2 - \frac{1}{4}x$ are orthogonal in $\mathcal{F}_{[0,1]}$ with respect to the particular inner product. Normalizing the v_i with respect to the given inner product $\langle .., ..\rangle$ makes $w_1 = \frac{1}{\sqrt{\langle v_1, v_1\rangle}}\, v_1 = \sqrt{3}\, x$ and $w_2 = \frac{1}{\sqrt{\langle v_2, v_2\rangle}}\, v_2 = 4\sqrt{5}\ x^2 - 3\sqrt{5}\ x$ since $\langle v_1, v_1\rangle = 1/3$ and $\langle v_2, v_2\rangle = 1/(12^2 \cdot 5)$. The students should check this assertion by evaluating $\langle v_1, v_2\rangle = \int_0^1 v_1(x)v_2(x)\ dx$, $\langle v_1, v_1\rangle$, and $\langle v_2, v_2\rangle$. ◀

Inner products define vector norms in a natural way.

Proposition 4: If $\langle .., ..\rangle$ is an inner product on a real vector space V, then

$$\|x\|_{\langle ..,..\rangle} := \langle x, x\rangle^{1/2} \; : \; V \to \mathbb{R}$$

defines a **vector norm** for every $x \in V$ with the following properties:

(1) $\|x\|_{\langle ..,..\rangle} \geq 0$ for all $x \in V$, and $\|x\|_{\langle ..,..\rangle} = 0 \in \mathbb{R}$ if and only if $x = 0 \in V$.

(2) $\|\alpha x\|_{\langle ..,..\rangle} = |\alpha| \|x\|_{\langle ..,..\rangle}$ for all vectors $x \in V$ and all scalars $\alpha \in \mathbb{R}$.

(3) $|\langle x, y\rangle| \leq \|x\|_{\langle ..,..\rangle}\|y\|_{\langle ..,..\rangle}$ for all vectors $x, y \in V$. (**Cauchy–Schwarz inequality**)

(4) $\|x + y\|_{\langle ..,..\rangle} \leq \|x\|_{\langle ..,..\rangle} + \|y\|_{\langle ..,..\rangle}$ for all vectors $x, y \in V$. (**triangle inequality**)

◀

A vector norm $\|..\|$ is called **induced** by the inner product $\langle .., ..\rangle$ if $\|..\| = \langle .., ..\rangle^{1/2}$ as it is in Proposition 4. General vector norms without an underlying inner product can, however, be solely defined by the three norm defining properties (1), (2), and (4) of Proposition 4. Proofs of both the Cauchy–Schwarz and the triangle inequality for the standard **euclidean norm** $\|x\| = \sqrt{x^*x}$ of $\mathbb{R}^n$ or $\mathbb{C}^n$ are outlined in Section 10.1 and Problems 23 and 26 in Section 10.1.P.

Definition 2: A function $g(x) : V \to \mathbb{R}$ is a **vector norm** on a real vector space V if it satisfies the properties (1), (2), and (4) of Proposition 4.

Example 3: The function $g(x) := \max_{i=1}^{n}\{|x_i|\} : \mathbb{R}^n \to \mathbb{R}$ is a vector norm on $\mathbb{R}^n$.

To see this we check the three conditions of a vector norm: Clearly $g(x) \geq 0$ for all $x \in \mathbb{R}^n$, and $g(x) = 0$ if and only if $\max |x_i| = 0 \in \mathbb{R}$, or if and only if $x = 0 \in \mathbb{R}^n$, establishing property (1). Next $g(x)$ satisfies property (2) since $g(\alpha x) = \max\{|\alpha x_i|\} = \max\{|\alpha||x_i|\} = |\alpha| \max\{|x_i|\} = |\alpha| g(x)$. Finally, the triangle inequality $|\alpha + \beta| \leq |\alpha| + |\beta|$ for scalars $\alpha, \beta \in \mathbb{R}$ helps us prove property (4):

$$g(x+y) = \max\{|x_i+y_i|\} \leq \max\{|x_i|+|y_i|\} \leq \max\{|x_i|\}+\max\{|y_i|\} = g(x)+g(y) \ .$$

Thus $g(x)$ is a vector norm.

In Example 5 we learn that g is not induced by any inner product of $\mathbb{R}^n$. ◀

A vector norm $\|..\|$ measures the length of vectors in V, just as the standard euclidean norm $\|u\| := \sqrt{u^T u}$ measures the length of vectors in $\mathbb{R}^n$ via the standard dot product.

Example 4: (a) Find the length of the standard unit vector $e_1 \in \mathbb{R}^2$ in terms of the vector norm that is induced by the inner product $\langle x, y\rangle = x^T \begin{pmatrix} 10 & 3 \\ 3 & 1 \end{pmatrix} y$ of Example 1(a).

We compute

$$\|e_1\|^2_{\langle ..,..\rangle} = \langle e_1, e_1\rangle_{\langle ..,..\rangle} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 10 & 3 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 3 \end{pmatrix} = 10 \ .$$

Therefore $\|e_1\|_{\langle ..,..\rangle} = \sqrt{10}$.

(b) Find the norm of the function $f(x) = x^2$ in the function space V of Example 2(a).

We have $\langle f, f\rangle = \int_0^1 2x^4 \, dx = \frac{2x^5}{5} \big|_0^1 = \frac{2}{5}$, giving f the length, or norm $\|f\|_{\langle ..,..\rangle} = \langle f, f\rangle^{1/2} = \sqrt{\frac{2}{5}}$.

(c) Find the length of the vector $w = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ for the norm that is induced by the inner product $\langle x, y\rangle := x^T \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} y$ on $\mathbb{R}^3$.

Clearly the matrix $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ is positive definite as a positive diagonal matrix. Therefore $\langle .., ..\rangle$ is an inner product according to Proposition 3. Next we compute the induced vector norm of w:

$$\begin{aligned} \|w\|^2_{\langle ..,..\rangle} &= \langle w, w\rangle = \begin{pmatrix} 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 6 \end{pmatrix} = 2+1+12 = 15 \,, \end{aligned}$$

or $\|w\|_{\langle ..,..\rangle} = \sqrt{15}$. Note that in the euclidean norm $\|w\|_2 := \sqrt{w^T I w} = \sqrt{w^T w} = \sqrt{6}$.

(d) Show that for $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$, the function $f(x) = \sqrt{3x_1^2 + 4x_1x_2 + 3x_2^2} : \mathbb{R}^2 \to \mathbb{R}$ is an induced vector norm.
We have $f^2(x) = 3x_1^2 + 4x_1x_2 + 3x_2^2 = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x^T \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} x$; see Example 6 in Section 11.3 for more on quadratic forms such as f. The matrix $A := \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} = A^T$ is real symmetric with eigenvalues that sum to its trace 6 and that multiply to its determinant $9 - 4 = 5$, according to Theorem 9.5. Thus the eigenvalues of A are 5 and 1, making A positive definite and $\langle x, y\rangle := x^T \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} y$ an inner product on $\mathbb{R}^2$ due to Proposition 3. This inner product induces $f(x)$ as a norm on $\mathbb{R}^2$. ◀

All inner products $\|..\|_{\langle ..,..\rangle}$ that are induced by an inner product $\langle .., ..\rangle$ satisfy the **parallelogram identity**.

Proposition 5: If $\|..\|_{\langle ..,..\rangle}$ is the induced vector norm for the inner product $\langle .., ..\rangle$ of an arbitrary real vector space V, then the **parallelogram identity**

$$\frac{1}{2}\left(\|x+y\|^2_{\langle ..,..\rangle} + \|x-y\|^2_{\langle ..,..\rangle}\right) = \|x\|^2_{\langle ..,..\rangle} + \|y\|^2_{\langle ..,..\rangle}$$

holds for all $x, y \in V$. ◀

The parallelogram identity states that the sum of the squared lengths of the two sides x and y of any parallelogram equals the average of the lengths of its two diagonals $x+y$ and $x-y$ squared for any induced vector norm.

Proof: To prove the parallelogram identity in a real vector space we expand its left hand side by using the properties of the norm inducing inner product.

$$\begin{aligned}
\frac{1}{2}\left(\|x+y\|^2_{\langle ..,..\rangle} + \|x-y\|^2_{\langle ..,..\rangle}\right) &= \frac{1}{2}\left(\langle x+y, x+y\rangle + \langle x-y, x-y\rangle\right) \\
&= \frac{1}{2}\langle x,x\rangle + \langle x,y\rangle + \frac{1}{2}\langle y,y\rangle + \frac{1}{2}\langle x,x\rangle - \langle x,y\rangle + \frac{1}{2}\langle y,y\rangle \\
&= \langle x,x\rangle + \langle y,y\rangle \;=\; \|x\|^2_{\langle ..,..\rangle} + \|y\|^2_{\langle ..,..\rangle} \;.
\end{aligned}$$

■

Example 5: The **maximum vector norm** $\|x\|_\infty := \max_{i=1}^{n}\{|x_i|\}$ of $\mathbb{R}^n$ from Example 3 is not induced by any inner product of $\mathbb{R}^n$ since it does not satisfy the parallelogram identity.

For example, in $\mathbb{R}^2$ we have for $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ that $\|x\|_\infty = \|y\|_\infty = 1 = \|x-y\|_\infty$ and $\|x+y\|_\infty = 2$ since $x+y = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $x-y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Thus

$$\frac{1}{2}\left(\|x+y\|^2_\infty + \|x-y\|^2_\infty\right) = \frac{1}{2}(4+1) = \frac{5}{2} \;\neq\; 2 = 1+1 = \|x\|^2_\infty + \|y\|^2_\infty \;.$$

◀

Proposition 4 makes every inner product space a normed vector space. However, in Example 5 and more generally in *functional analysis*, it has been shown that not all vector norms derive from inner products. To complete our elementary explorations of inner product spaces and normed vector spaces, we mention without proof that all normed vector spaces whose norm $\|..\|$ satisfies the parallelogram identity of Proposition 5 can be made into an inner product space by setting

$$\langle x, y\rangle \;:=\; \frac{\|x+y\|^2 - \|x\|^2 - \|y\|^2}{2} \;.$$

Moreover, the parallelogram identity can be generalized to complex vector spaces, but this is beyond the scope of this appendix and elementary linear algebra.

A.D.P Problems

1. Show that the function $f(x,y) = 2x_1y_1 - x_2y_2 + 4x_3y_3 + x_2y_3 : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ is a bilinear function for $x, y \in \mathbb{R}^3$. Is this function an inner product on $\mathbb{R}^3$? Does f induce a norm on $\mathbb{R}^3$?

2. (a) Show that $\langle x,y\rangle_A := x^T A y$ is a bilinear function for every real n by n matrix A.

   (b) Show that if $P = P^T \in \mathbb{R}^{n,n}$ is positive definite, then P can be expressed as $P = A^T A$ for some nonsingular real matrix A.
   (Hint: Use Chapter 11: Diagonalize P orthogonally as $U^T P U = D = \sqrt{D}\sqrt{D}$ for a positive diagonal matrix D and extract P from this matrix equation.)

   (c) Show: If $P = P^T \in \mathbb{R}^{n,n}$ is positive definite, then $\langle x,y\rangle_P := x^T P y$ is an inner product on $\mathbb{R}^n$.
   (Hint: Use part (b).)

3. Test whether the following functions are (a) bilinear and (b) inner products on their respective spaces:
(1) $h(x,y) = x^T \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 1 & 1 \end{pmatrix} y$ for $x, y \in \mathbb{R}^3$.
(2) $k(x,y) = x^T \begin{pmatrix} 4 & 2 \\ 1 & 1 \end{pmatrix} y$ for $x, y \in \mathbb{R}^2$.
(3) $\ell(x,y) = x^T \begin{pmatrix} 4 & 2 \\ 1 & -1 \end{pmatrix} y$ for $x, y \in \mathbb{R}^2$.
(4) $m(x,y) = x^T \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix} y$ for $x, y \in \mathbb{R}^2$.
(5) $n(x,y) = 2x_1^2 - 3x_1y_1 + 4y_1^2$ on $\mathbb{R}^2$.
(6) $p(x,y) = -x_1y_2$ on $\mathbb{R}^2$.

4. (a) Find the length of the two vectors $x = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}$ and $y = \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \end{pmatrix} \in \mathbb{R}^n$ for both the standard euclidean vector norm and for the maximum vector norm.

(b) Find the cosine of the angle between x and $y \in \mathbb{R}^n$ in part (a) for
(1) the standard inner product and the
(2) inner product $x^T \begin{pmatrix} 3 & 1 & & 0 \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & 1 & 3 \end{pmatrix} y$
of $\mathbb{R}^n$.

5. If $w(x) > 0$ is continuous on the interval $[0,1]$, prove that $\langle f,g\rangle_w := \int_0^1 f(x)g(x)w(x)\ dx$ is an inner product on the space of continuous functions $\mathcal{F}_{[0,1]}$.

6. Construct a positive and continuous weight function $w(x)$ so that the two functions $f(x) = 1$ and $g(x) = x \in \mathcal{F}_{[0,1]}$ become orthogonal with respect to the inner product $\langle f,g\rangle_w := \int_0^1 f(x)g(x)w(x)\ dx$, if possible.

7. Construct a positive and continuous weight function $w(x)$ so that the two functions $f(x) = 1$ and $g(x) = x - \frac{1}{2} \in \mathcal{F}_{[0,1]}$ are not orthogonal with respect to the inner product $\langle f,g\rangle_w := \int_0^1 f(x)g(x)w(x)\ dx$, if possible.

8. Orthogonalize the two functions $f(x) = 1$ and $g(x) = x \in \mathcal{F}_{[0,1]}$ with respect to the weighted inner product $\langle f,g\rangle_w := \int_0^1 f(x)g(x)x^2\ dx$.

9. Orthogonalize the three functions $f(x) = 1$, $g(x) = x$, and $h(x) = x^2$ in $\mathcal{F}_{[0,1]}$ with respect to the inner product $\langle f,g\rangle := \int_0^1 f(x)g(x)\ dx$.

10. Show that the standard euclidean vector norm $\|x\| := \sqrt{x^Tx}$ for $x \in \mathbb{R}^n$ satisfies the parallelogram identity.

11. Show that $\|x\|_1 := \sum_{i=1}^n |x_i|$ is a vector norm on $\mathbb{R}^n$. Is it induced by an inner product or not?

12. Examine whether the vector norm $\|x\|_1$ of $\mathbb{R}^n$ in the previous problem is an induced vector norm.

13. (a) Assume that the norm $\|..\|$ is an induced norm on V. If $\|u\| = 4$, $\|u+v\| = 6$, and $\|u-v\| = 5$, what is the length of v? What is the distance between u and v?

(b) Repeat part (a) for $\|u\| = 7$.

14. Let V be a real inner product space. Let $\|..\|$ be the vector norm that is induced by the inner product $\langle ..,..\rangle$ on V. Show that

$$\langle x,y\rangle = \frac{\|x+y\|^2 - \|x-y\|^2}{4}$$

holds for all $x, y \in V$.

15. Let $\{u_1, ..., u_k\}$ be an orthonormal set of vectors in an inner product space of finite dimension $n \geq k$. Prove that $x \in \operatorname{span}\{u_1, ..., u_k\}$ if and only if $\|x\|^2 = |\langle x,u_1\rangle|^2 + \ \dots\ + |\langle x,u_k\rangle|^2$ for the induced vector norm.

16. (a) Does $\langle x,y\rangle := x^* \begin{pmatrix} 2 & -i \\ i & 1 \end{pmatrix} y$ define an inner product on $\mathbb{C}^2$?

(b) Repeat part (a) for $\langle x,y\rangle := x^* \begin{pmatrix} i & i \\ i & 1 \end{pmatrix} y$.

(c) Repeat part (a) for $\langle x,y\rangle := \overline{x_1}y_1 - 2\overline{x_2}y_2$.

(d) Are any of the functions in parts (a) through (c) bilinear?

17. In a normed vector space V we define the **distance function** between any two vectors as $d(x, y) := \|x - y\|$. Show that

(a) $d(x, y) = d(y, x)$,

(b) $d(x, y) \leq d(x, z) + d(z, y)$ for all x, y, and $z \in V$.

(c) Evaluate the distance between $x = \begin{pmatrix} 1 & -2 & 4 & -3 \end{pmatrix}$ and $y = \begin{pmatrix} 0 & 2 & -1 & -3 \end{pmatrix} \in \mathbb{R}^4$ for the euclidean norm and the maximum norm.

18. (a) Assume that $\langle x, y \rangle_H := x^T H x$ and $\langle x, y \rangle_K := x^T K x$ are two bilinear forms on $\mathbb{R}^n$ with H and K both n by n and positive definite.
If $\langle x, y \rangle_H = \langle x, y \rangle_K$ for all $x, y \in \mathbb{R}^n$, show that $H = K$ as matrices.

(b) Show: If $x^T A y = 0$ for a real symmetric matrix A and all vectors $x, y \in \mathbb{R}^n$, then $A = O_n$.

19. Let $x = \begin{pmatrix} 2-i & 1+i & 3 \end{pmatrix}$ and $y = \begin{pmatrix} i & 2 & -i \end{pmatrix} \in \mathbb{C}^3$. For the standard euclidean vector norm of $\mathbb{C}^3$, compute

(a) $\|x\|$ and $\|y\|$, (b) the distance $d(x, y)$,
(c) $\|y - x\|$, (d) $\|x + y\|^2$, and
(e) the angle between x and $y \in \mathbb{C}^3$.

20. In $\mathcal{F}_{[0,1]}$ with the standard inner product $\langle f, g \rangle := \int_0^1 f(x)g(x)\, dx$, determine the 'length' of the two functions $f(x) = 3x - 2$ and $g(x) = x^2 + x \in \mathcal{F}_{[0,1]}$, as well as the cosine of the angle between f and g.

21. Consider the function $f(p, q) : \mathcal{P}_n \times \mathcal{P}_n \to \mathbb{R}$ defined by $f(p, q) = p(1)q(1) + 2p(2)q(2) + 3p(3)q(3)$, where $\mathcal{P}_n$ denotes the real variable polynomials of degree not exceeding n.

(a) Show that f is bilinear on $\mathcal{P}_n \times \mathcal{P}_n$.

(b) Show that f induces a norm $\|..\|$ on $\mathcal{P}_m$ for all $m \leq 2$.

(c) Show that f does not induce a norm on any $\mathcal{P}_m$ if $m > 2$.

(d) Find $\|2x^2 - 3x\|$ and $\|1 - x^2 - x^3 + 4x\|$ in the vector norm that is induced by f.

(e) Write down a formula for the induced norm $\|p\|$ and $p \in \mathcal{P}_2$.

Solutions to Selected Problems of the Web Chapters

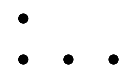
Chapter 14

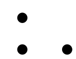
14.1.P **p. W-17**
$\downarrow$

1. The two 5 by 5 matrices comprised of zeros everywhere, except for a 1 in positions (1,5) or (1,2) both have the 5 fold eigenvalue zero, but a 1-dimensional kernel, or only 4 corresponding linearly independent eigenvectors. According to Chapter 9.1 or 9.1.D they cannot be diagonalized.

3. For a non-diagonalizable example, expand the solution of problem 1. to 12 by 12. For a diagonalizable triangular matrix, use an arbitrary 12 by 12 upper triangular matrix with distinct diagonal entries. Such a matrix has 12 distinct eigenvalues and is therefore diagonalizable.

5. Look at the eigenspace for $\lambda = -1$, or at the kernel of $A + I = \begin{pmatrix} 1 & 0 & 1 \\ -3 & 0 & -3 \\ -1 & 0 & -1 \end{pmatrix}$. It has dimension two, so there are two linearly independent eigenvectors for $\lambda = -1$. Now compute trace$(A) = -3 = -1 - 1 + \lambda_3$, making $\lambda = -1$ a triple eigenvalue for A. Therefore the Jordan normal form of A is $J = \text{diag}(J(-1,2), J(-1,1))$.

7. For Example 4:  ;

for Problem 5:  .

9. We can find the eigenvalues of $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ from $\det(A - \lambda I) = \det\begin{pmatrix} a-\lambda & b \\ -b & a-\lambda \end{pmatrix} = (a-\lambda)^2 + b^2 = 0$ as $\lambda_{1,2} = a \pm bi$. We find the complex eigenvectors by looking at $A - (a+bi)I = \begin{pmatrix} -bi & b \\ -b & -bi \end{pmatrix}$ as $\begin{pmatrix} 1 \\ i \end{pmatrix}$, and from $A - (a-bi)I = \begin{pmatrix} bi & b \\ -b & bi \end{pmatrix}$ as $\begin{pmatrix} -1 \\ i \end{pmatrix}$. Now form the eigenvector column matrix $X = \begin{pmatrix} 1 & -1 \\ i & i \end{pmatrix}$ with $X^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -i \\ -1 & -i \end{pmatrix}$. Then $X^{-1}AX = \frac{1}{2}\begin{pmatrix} 1 & -i \\ -1 & -i \end{pmatrix}\begin{pmatrix} a & b \\ -b & a \end{pmatrix}\begin{pmatrix} 1 & -1 \\ i & i \end{pmatrix} = \begin{pmatrix} a+bi & 0 \\ 0 & a-bi \end{pmatrix}$.

11. Since trace$(C) = -6 - 2 + 10 = 2 = \lambda_1 + \lambda_2 + \lambda_3$ and $\lambda_1 = -1$ and $\lambda_2 = 2$ are given, we must have $\lambda_3 = 1$. Therefore C has three distinct eigenvalues and is diagonalizable according to Chapter 9. The Jordan normal form of C is $\operatorname{diag}(-1, 2, 1)$.

13. Let the size of A be n by n. 4 foldness of an eigenvalue refers to the algebraic multiplicity; therefore $n = 4$. Since geom. mult. $\leq$ alg. mult., we may have 1, or 2, or 3, or 4 linearly independent eigenvectors for this single eigenvalue. Giving rise to the following Jordan diagrams in turn: $\begin{array}{l} \bullet \\ \bullet \\ \bullet \\ \bullet \end{array}$ in case of geometric multiplicity 1, or $\begin{array}{ll} \bullet & \\ \bullet & \\ \bullet & \bullet \end{array}$ or $\begin{array}{ll} \bullet & \bullet \\ \bullet & \bullet \end{array}$ in case of geometric multiplicity 2, or $\begin{array}{lll} \bullet & & \\ \bullet & \bullet & \bullet \end{array}$ in case of geometric multiplicity 3, or to $\begin{array}{llll} \bullet & \bullet & \bullet & \bullet \end{array}$ if A is diagonalizable.

15. Here $n = 6$ and $\operatorname{diag}(J(\lambda, n_1), J(\lambda, n_2), ..., J(\lambda, n_k))$ with $k \leq 6$ and $\sum\limits_{i=1}^{k} n_i = 6$.

17. The Jordan normal form J of A has three separate Jordan blocks for λ. One of these must be 4 by 4, while the other two have sizes that will add up to 8. Therefore J will contain the string $J(\lambda, 4), J(\lambda, n_1), J(\lambda, n_2))$ with $n_1 + n_2 = 4$ and $n_i \geq 1$. This gives us two possibilities: $n_1 = 1$, $n_2 = 3$ or $n_1 = 2 = n_2$. The third possible case $n_1 = 3$ and $n_2 = 3$ is only a reordering of the first mentioned one. So for λ there are precisely two variations possible in the Jordan normal form for A.

19. The maximal index must be six since there must be seven linearly independent eigenvectors for our 12 fold eigenvalue. The minimal index is two, provided 5 Jordan blocks of size 2 occur together with two 1 by 1 Jordan blocks. Any index in between 2 and 6 can also occur.

21. (a) A's Jordan normal form has one 2 by 2 block and 1 by 1 blocks else.
(b) A's Jordan normal form has one 3 by 3 block and 1 by 1 blocks else, or it has two 2 by 2 blocks and 1 by 1 blocks else.
(c) There are three Jordan blocks for λ. Their possible sizes are 1, 1, 4; or 1, 2, 3; or 2, 2, 2, respectively.

23. (a) Simply multiply $J^{-1} \cdot J$ out for the matrix J^{-1} as given and you should obtain I.
(b) $J(\lambda, k)^{-1}$ is tridiagonal with its n fold eigenvalue $\frac{1}{\lambda}$ appearing on the diagonal. Since $\operatorname{rank}(J(\lambda, k)^{-1} - \frac{1}{\lambda} I) = k - 1$ from the formula for the inverse in part (a), $J(\lambda, k)^{-1}$ must be similar to a single Jordan block of the form $J\left(\frac{1}{\lambda}, k\right)$.
And if $X^{-1}AX = J$ is a Jordan form matrix, then $J^{-1} = X^{-1}A^{-1}X$, or A^{-1} and J^{-1} are also similar, hence have the same Jordan normal form. But if $J = \operatorname{diag}(J_i)$, then $J^{-1} = \operatorname{diag}(J_i^{-1})$, so the Jordan normal form of the inverse of a nonsingular matrix A has the same Jordan normal form structure as A, except instead of the eigenvalues λ_i for A it has the eigenvalues $\frac{1}{\lambda_i}$.

25. We have $\operatorname{trace}(A) = \lambda_1 + \lambda_2 = 4$ and $\det(A) = \lambda_1 \lambda_2 = 4$, so that $\lambda_1 = \lambda_2 = 2$ is a double eigenvalue for A. Looking at the kernel of $A - 2I$ we find its dimension equal to 1. Therefore A is not diagonalizable with the Jordan normal form $J = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$.

27. (a) Every matrix $A = A(\alpha)$ has only one linearly independent eigenvector for its double eigenvalue λ since $\operatorname{rank}(A - \lambda I) = \operatorname{rank}\begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix} = 1$. Therefore each $A(\lambda)$ is similar to $J(\lambda, 2) = \begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$. If $\alpha = 0$, then $A = \lambda I = \operatorname{diag}(J(\lambda, 1), J(\lambda, 1))$ is not similar to $J(\lambda, 2)$.
(b) Analogous to part (a), if $\prod \alpha_i \neq 0$, then $\operatorname{rank}(A - \lambda I) = 1$ and A is similar to a single Jordan block $J(\lambda, k)$. If at least one off diagonal entry $\lambda_j = 0$, then $\operatorname{rank}(A - \lambda I) \geq 2$ and A is similar to a Jordan form with at least two Jordan blocks for λ.
(c) Both $B - \lambda I$ and $C - \lambda I$ have rank 1 for their double eigenvalue λ. Hence both B and C have the same Jordan normal form

$J = J(\lambda, 2) = \begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$. I.e., $X^{-1}BX = J$ and $Y^{-1}CY = J$ for two matrices X and Y, making $C = YX^{-1}BXY^{-1} = (XY^{-1})^{-1}B(XY^{-1})$ similar.

29. (a) $Ae_k = ae_k$.
(b) $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & -3 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$ are both diagonalizable, but $\begin{pmatrix} -3 & 0 & 0 & 0 \\ \alpha & -3 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$ is not for any $\alpha \neq 0$.
(c) In our triangular matrix examples, nonzero entries in columns and rows that have the same diagonal entry seem to determine diagonalizability.

31. (a) $A = \begin{pmatrix} i & 1 \\ 0 & 1 \end{pmatrix}$ has one complex eigenvalue $\lambda = i$ and one real eigenvalue $\lambda = 1$.
(b), (c) $A = \begin{pmatrix} -1+i & 3 \\ i & 2-i \end{pmatrix}$ has determinant $(-1+i)(2-i) - 3i = -1$ and trace 1. Thus its characteristic polynomial is $f_A(x) = x^2 + x - 1$ with the roots $x_{1,2} = -\frac{1}{2} \pm \sqrt{\frac{1}{4} + 1} \in \mathbb{R}$. These are the eigenvalues of $A \notin \mathbb{R}^{n,n}$.

14.2.P p. W-34
↓

1. We can use vector iteration of Section 9.1 with $x = \begin{pmatrix} 0 \\ 2 \\ -1 \\ 1 \end{pmatrix}$ for example and obtain $\begin{pmatrix} \vdots & & \vdots & & \vdots \\ x & Ax & A^2x & A^3x & A^4x \\ \vdots & & \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} 0 & -10 & -10 & -20 & -60 \\ 2 & -1 & -2 & 6 & -8 \\ -1 & 3 & 6 & 2 & 24 \\ 1 & -3 & 4 & -2 & -4 \end{pmatrix}$ with the reduced row echelon form $\begin{pmatrix} 1 & 0 & 0 & 4 & 0 \\ 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$. Therefore according to Section 9.1, the roots of the polynomial $x^3 - 2x - 4$ are three eigenvalues of A. The roots are 2 (found by guessing) and $-1 \pm i$ (then found by long division and the quadratic formula). The fourth root is 2 from the trace condition. A row reduction of $A - 2I_4$ shows this matrix to have rank 3. Thus A has only one eigenvector for its double real eigenvalue 2. A therefore has the real Jordan normal form $J = \text{diag}(J(2,2), J(-1,1,2))$.
To start out with computing $\det(A - \lambda I_4)$ seems much harder.

3. For the real matrix A all complex eigenvalues such as λ and μ come doubly as λ, μ and as $\overline{\lambda}, \overline{\mu}$. Thus the real principal subspace associated with λ is 6 dimensional, and the one for μ has dimension 4. Since the size of A is 13 by 13, this makes ν have algebraic multiplicity 3 as a root of the characteristic polynomial of A. Finally ν can have the geometric multiplicities 1, 2, or 3.
The possible real Jordan forms of A consist of any combination from the following three Jordan block groups: for $\lambda = 2 + 3i$: $J(2,3,6)$, $\text{diag}(J(2,3,4), J(2,3,2))$, or $\text{diag}(J(2,3,2), J(2,3,2), J(2,3,2))$;
for $\mu = 1 - 2i$: $J(1,-2,4)$ or $\text{diag}(J(1,-2,2), J(1,-2,2))$;
and for $\nu = 7$: $J(7,3)$, $\text{diag}(J(7,2), J(7,1))$, or $\text{diag}(7,7,7)$.

5. (a) $Je_k = \lambda e_k$.
(b) The index of λ is k.
(c) Since $(J - \lambda I)^k = O_k$ and $(J - \lambda I)^{k-1}$ is the matrix of all zeros except for a 1 in position (k,1), and since a principal vector of index k for A must lie in the kernel of $(J - \lambda I)^k$ but not in the kernel of $(J - \lambda I)^{k-1}$, this principal vector must be a multiple of e_1.
(d) $e_{k-j+1} \in P_j(\lambda)$.

7. $vA =$

$$\begin{pmatrix} 1 & z & z^2 & \ldots & z^{n-1} \end{pmatrix} \begin{pmatrix} 0 & & & -a_0 \\ 1 & & & \vdots \\ & \ddots & & \vdots \\ & & 1 & -a_{n-1} \end{pmatrix}$$

$= \begin{pmatrix} z & z^2 & \ldots & -a_0 - a_1 z - \ldots - a_{n-1}z^{n-1} \end{pmatrix}$
$= z\begin{pmatrix} 1 & z & z^2 & \ldots & z^{n-1} \end{pmatrix}$ since $z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0 = 0$.

9. The given matrix $B = \begin{pmatrix} -2 & 2i \\ 1 & 3 \end{pmatrix}$ has two distinct complex eigenvalues since its characteristic polynomial is $\lambda^2 - \lambda - 6 - 2i$. Thus B is similar over $\mathbb{C}$ to a complex diagonal matrix diag(λ_1, λ_2), or $X^{-1}BX =$diag(λ_1, λ_2). Thus $B = (X \text{ diag}(\lambda_1, \lambda_2)X^T)(X^{-T}IX^{-1})$ expresses B as the product of two complex symmetric matrices.

11. $A = \begin{pmatrix} -2 & 1 \\ 4 & -4 \end{pmatrix} =$

$$= \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 0 & -2 \\ -2 & 1 \end{pmatrix}.$$

13. Note that $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ b & a \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & -b \\ -b & a \end{pmatrix}$ and that $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$.

14.3.P **p. W-43**
↓

1. (a) For Example 9 we use the product rule of calculus twice in each component function, abbreviate $\cos t$ by C and $\sin t$ by S, and obtain :

$$x'(t) = ce^t \begin{pmatrix} 3S + C + 3tS - tC \\ -tS \\ S + C + tS \\ S - tC \end{pmatrix} + ce^t \begin{pmatrix} 3C - C + 3S - C + 3tC + tS \\ -S - tC \\ C - S + S + tC \\ C - C + tS \end{pmatrix} =$$

$$= ce^t \begin{pmatrix} 5S + 3C + 4tS + 2tC \\ -S - tC - tS \\ S + 2C + tS + tC \\ S - tC + tS \end{pmatrix}.$$

And

$$Ax = ce^t \begin{pmatrix} 4 & 7 & -1 & -6 \\ -2 & -3 & 2 & 3 \\ 0 & 1 & 2 & -1 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

$$\cdot \begin{pmatrix} 3S + C + 3tS - tC \\ -tS \\ S + C + tS \\ S - tC \end{pmatrix} = ce^t \cdot$$

$$\begin{pmatrix} 12S + 4C + 12tS - 4tC - 7tS - S - C - tS - 6S + 6tC \\ -6S - 2C - 6tS + 2tC + 3tS + 2S + 2C + 2tS + 3S - 3tC \\ -tS + 2S + 2C + 2tS - S + tC \\ tS + S - tC \end{pmatrix}$$

$$= ce^t \begin{pmatrix} 5S + 3C + 4tS + 2tC \\ -S - tC - tS \\ S + 2C + tS + tC \\ S - tC + tS \end{pmatrix} \text{ as before.}$$

(b) In Example 10 we have computed the solution as $x(t) = ce^{-t}(1 + t) + ke^{3t}$. Consequently $x'(t) = -cte^{-t} + 3ke^{3t}$, $x''(t) = -ce^{-t} + cte^{-t} + 9ke^{3t}$, and $x'''(t) = 2ce^{-t} - cte^{-t} + 27ke^{3t}$. Therefore $x''' - x'' - 5x' - 3x = 0$.

7. The solution $w(t)$ of $w'(t) =$ diag$(J(2, 2), J(-1, 1, 2))w(t)$ in terms of the Jordan basis $\mathcal{U}$ of A is given according to (14.7) and (14.10) by $w(t) = \begin{pmatrix} ce^{2t} \\ cte^{2t} \\ ke^{-t}\sin t \\ ke^{-t}\cos t \end{pmatrix}$. Since the solution $x(t)$ of $x' = Ax$ is related to w and the Jordan normal form $J = U^{-1}AU$ of A according to the formula $x = Uw$, we need to compute the eigenvector/principal vector matrix U for A and J next.
With $B = A - 2I$ we compute $B^2 = \begin{pmatrix} -12 & -10 & -36 & 14 \\ -14 & -20 & -42 & 8 \\ 14 & 20 & 42 & -8 \\ 2 & 10 & 6 & 6 \end{pmatrix}$ and its RREF $\begin{pmatrix} 1 & 0 & 3 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. Clearly $x^{(2)} =$

$\begin{pmatrix} 2 \\ -1 \\ 0 \\ 1 \end{pmatrix} \in \ker((A-2I)^2)$ is a principal vector for $\lambda = 2$ of order two since $(A-2I)x^{(2)} = \begin{pmatrix} 3 \\ 0 \\ -1 \\ 0 \end{pmatrix} = x^{(1)}$ is an eigenvector of A.

For the complex root $\lambda = -1+i$ we look at the kernel of $A-(-1+i)I$ via the RREF of this matrix: $\begin{pmatrix} 1 & 0 & 0 & -1+i \\ 0 & 1 & 0 & i \\ 0 & 0 & 1 & -i \\ 0 & 0 & 0 & 0 \end{pmatrix}$. A complex eigenvector for $\lambda = -1+i$ is the vector $\begin{pmatrix} 1-i \\ -i \\ i \\ 1 \end{pmatrix}$.

Thus $U = \begin{pmatrix} 2 & 3 & 1 & -1 \\ -1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$ transforms A to its Jordan form J. Note that the columns of U contain the principal vector chain in descending order for $\lambda = 2$ first and then the real and the complex part vectors for the complex eigenvalue follow.

Therefore $x(t) = \begin{pmatrix} 2 & 3 & 1 & -1 \\ -1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} w(t) = Uw(t)$ solves the original DE $x' = Ax$. Here $x(t)$ computes to be $\begin{pmatrix} ce^{2t}(2+3t) + ke^{-t}(\sin t - \cos t) \\ -ce^{2t} - ke^{-t}\cos t \\ -cte^{2t} + ke^{-t}\cos t \\ ce^{2t} + ke^{-t}\sin t \end{pmatrix}$.

9. [The solution to this problem goes along the same path as that of Problem 7 above (and built on that of Problem 1 of 14.2.P). For brevity, here are the main steps of a solution:]

B has the double complex conjugate eigenvalue pair $1 \pm 2i$.

The RREF of $(B-(1+2i)I)^2$ is $\begin{pmatrix} 1 & 0 & 1+i & -4-2i \\ 0 & 1 & 1 & -5+i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and $x^{(2)} = \begin{pmatrix} 4+2i \\ 5-i \\ 0 \\ 1 \end{pmatrix}$ is a complex principal vector of order 2 for $1+2i$ since $(B-(1+2i)I)x^{(2)} = \begin{pmatrix} 5-i \\ 12-5i \\ 12-5i \\ 5-i \end{pmatrix} =: x^{(1)} \in \ker(B-(1+2i)I)$ is an eigenvector for $1+2i$.

From (14.5) we obtain the real Jordan basis for B from the real and imaginary parts vectors of $x^{(2)}$ and $x^{(1)}$: $X = \begin{pmatrix} 4 & 2 & 5 & -1 \\ 5 & -1 & 12 & -5 \\ 0 & 0 & 12 & -5 \\ 1 & 0 & 5 & -1 \end{pmatrix}$ transforms B to its real Jordan normal form $J = X^{-1}BX = J(1,2,4)$ that is comprised of a single 4-dimensional real Jordan block for $1+2i$.

Next, according to (14.10), the solution to the Jordan form DE $w'(t) = J(1,2,4)w(t)$ is given as $w(t) = ce^t \begin{pmatrix} \sin(2t) \\ \cos(2t) \\ t\sin(2t) \\ t\cos(2t) \end{pmatrix}$. Finally

$$x(t) = Xw(t) = ce^t \begin{pmatrix} 4S+2C+5tS-tC \\ 5S-C+12tS-5tC \\ 12tS-5tC \\ S+5tS-tC \end{pmatrix}$$

solves the original DE $x'(t) = Bx(t)$. Here we have abbreviated $\sin(2t)$ by S and $\cos(2t)$ by C.

11. Following Section 14.3 (b), we form the associated 5^{th} degree polynomial $p(r) = r^5 - 5r^4 + 5r^3 + 5r^2 - 6r$. Its roots are -1, 0, 1, 2, and 3. The associated transposed companion matrix $C = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 6 & -5 & -5 & 5 \end{pmatrix}$ is diagonalizable since p has distinct roots. Thus $J = \operatorname{diag}(-1,0,1,2,3) = X^{-1}CX$ for the matrix X with the eigenvectors of C as its columns. For $\lambda = -1$, an eigenvector is $x_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}$; for $\lambda = 2$ it is $x_2 = e_1$; for $\lambda = 1$ it is x_3, the vector of all ones; for $\lambda = 2$ it is $x_4 = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 8 \\ 16 \end{pmatrix}$; and for $\lambda = 3$ we have $x_5 = \begin{pmatrix} 1 \\ 3 \\ 9 \\ 27 \\ 81 \end{pmatrix}$ as the corresponding eigenvector of C.

The Jordan normal form solution $w(t)$ of $w'(t) = Jw(t)$ has the totally separated form $w(t) = \begin{pmatrix} ce^{-t} \\ d \\ ke^{t} \\ \ell e^{2t} \\ me^{3t} \end{pmatrix}$ since J is diagonal. And the solution to the original 5^{th} order DE is equal to the first entry $x(t) = y_1(t)$ of $y(t) = Xw(t)$, or to $x(t) = ce^{-t}+d+ke^{t}+\ell e^{2t}+me^{3t}$ for arbitrary constants c, d, k, ℓ, m, since the first row of X is the vector of all ones.

14.R **p. W-44**

↓

1. (a) See Theorem 7.2.
(b) See the Determinant Proposition, part (12), in Section 8.1.
(c), ..., (h) Similar matrices have the same characteristic and minimal polynomials.

3. (a) Each individual Jordan block is associated with a principal vector of order equal to the size of the block.
(b) If $\lambda \in \mathbb{R}$, then the above statement remains true. But if $\lambda \notin \mathbb{R}$, then the index of $\lambda = a + bi$ is equal to half the size m of the largest real Jordan block $J(a, b, 2m)$ associated with λ.

5. Since $(A - \lambda I)^k x = 0$ and $(A - \lambda I)^{k-1} x \neq 0$, we likewise have $(A - \lambda I)^{k-2} y = (A - \lambda I)^k x = 0$ and $(A - \lambda I)^{k-3} y = (A - \lambda I)^{k-1} x \neq 0$, provided $k > 2$.

7. 1 by 1 matrices are the only such matrices.

9. $A = \text{diag}(J(0, 3), J(-2, 4))$ and $B = \text{diag}(J(0, 3), J(0, 2), J(-2, 4), J(-2, 3))$ for example, or any matrices similar to these.

11. To find the minimal sized matrix B with the same minimal polynomial $(x - 2)^4 (x + 1)^5 x$ as J, we must only allow one properly sized Jordan block for each distinct eigenvalue, or B must be similar to $\text{diag}(J(2, 4), J(-1, 5), J(0, 1))$. This B has size $4 + 5 + 1 = 10$ by 10.

13. Clearly A^2 is similar to J^2 if A is to J, which we assume to be the Jordan normal form of A. If all eigenvalues of A are nonzero, i.e., if A is nonsingular, then A^2 has the same Jordan structure as A for the squared eigenvalues of A from the previous problem. If A is singular, i.e., if the Jordan normal form J of A has a Jordan block of the form $J(0, k)$, then $J(0, k)^2$ is similar to the matrix $\text{diag}(J(0, k/2), J(0, k/2))$ if k is even, and to $\text{diag}(J(0, (k + 1)/2), J(0, (k - 1)/2))$ if k is odd. This specifies the Jordan form of A^2 in terms of the Jordan form J of A completely.

Appendix D

A.D.P **p. W-58**

↓

1. $f(x, y) = x^T \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 1/2 \\ 0 & 1/2 & 4 \end{pmatrix} = x^T Sx.$
But S is not positive definite since $e_2^T Se_2 =$

-1 and therefore the bilinear form f does not define a norm on $\mathbb{R}^3$.

3. (a) All example functions, except (5) are bilinear forms.
(b) We need to check whether $...(x,x) > 0$ for all $x \neq 0$:
In (1), $h(e_1, e_1) = 0$, i.e., h does not define an inner product.
In (2) the defining matrix $\begin{pmatrix} 4 & 2 \\ 1 & 1 \end{pmatrix}$ can be replaced by the symmetric matrix $\begin{pmatrix} 4 & 3/2 \\ 3/2 & 1 \end{pmatrix}$ which has two positive real eigenvalues, hence is positive definite. This makes h an inner product.
In (3) $\ell(e_2, e_2) = -1$ contradicts ℓ being an inner product.
In (4) the defining matrix is positive definite and thus m is an inner product.
In (6) $p(e_1, e_1) = 0$, hence p cannot be an inner product.

5. If the weight function $w > 0$ and $f \neq 0$, then $\langle f, f\rangle = \int_0^1 f^2 w\, dx > 0$ since $f^2 w > 0$ on at least one subinterval of $[0,1]$, establishing property (d) of an inner product. The other defining properties (a), (b), and (c) follow readily from the linearity of integrals.

7. We take $w(x) := x + 1 > 0$ on $[0, 1)$. Then $\int_0^1 f(x)g(x)w(x)\, dx = \int_0^1 (x-1/2)(x+1)\, dx \neq 0$ and thus with respect to the inner product $\langle f, g\rangle = \int_0^1 fgw\, dx$ the two given functions f and g are not perpendicular.

9. First level Gram–Schmidt:
$v_1 = f = 1$;
$v_2 = \langle v_1, v_1\rangle g - \langle g, v_1\rangle v_1 = x - 1/2$;
$v_3 = \langle v_1, v_1\rangle h - \langle h, v_1\rangle v_1 = x^2 - 1/3$;
Second level Gram–Schmidt:
$v_3 = \langle v_2, v_2\rangle v_3 - \langle v_3, v_2\rangle v_2 = \frac{x^2}{12} - \frac{3x}{40} + \frac{7}{720}$.
Students should check that the finally computed v_i are mutually orthogonal.

11. We verify conditions (1), (2), and (4) of Proposition 4:
(1) and (2) are obvious; for (4) we use the triangle inequality for numbers n times:

$$\|x+y\|_1 = \sum_i |x_i + y_i| \leq \sum_i (|x_i| + |y_i|) = \\ = \sum_i |x_i| + \sum_i |y_i| = \|x\|_1 + \|y\|_1.$$

$\|x\|_1$ is not an induced norm, because it violates the parallelogram law: For $x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ we have $x + y = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $x - y = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ with $\|x\|_1^2 = 4$, $\|y\|_1^2 = 1$, $\|x+y\|_1^2 = 9$, and $\|x-y\|_1^2 = 9$. Therefore $\frac{1}{2}\left(\|x+y\|_1^2 + \|x-y\|_1^2\right) = 9 \neq 5 = \|x\|_1^2 + \|y\|_1^2$.

13. (a) From the parellelogram identity $\|v\|^2 = \frac{1}{2}\left(\|u+v\|^2 + \|u-v\|^2\right) - \|u\|^2 = \frac{1}{2}(36 + 25) - 16 = 15.5$, or $\|v\| = \sqrt{15.5}$. Note that $d(u, v) = \|u - v\| = 5$ as given.
(b) If $\|u\| = 7$, then from part (a) $\|v\|^2 = \frac{61}{2} - 49 < 0$ which is impossible.

15. If $x = \sum_{i=1}^k \alpha_i u_i$ then $\|x\|^2 = \langle \sum_{i=1}^k \alpha_i u, \sum_{j=1}^k \alpha_j u_j\rangle = \sum_{i=1}^k \alpha_i \langle u_i, \sum_{j=1}^k \alpha_j u_j\rangle = \sum_{i=1}^k \alpha_i \langle u_i, \alpha_i u_i\rangle = \sum_{i=1}^k |\alpha_i|^2$. And $\langle x, u_j\rangle = \langle \sum_{i=1}^k \alpha_i u_i, u_j\rangle = \alpha_j$ completes one direction of the statement.
Conversely assume that V has an ONB of the form $\{u_1, ..., u_k, u_{k+1}, ..., u_n\}$ where the first k basis vectors are as given. Then for any $x \in V$ we have $x = \sum_{i=1}^n \alpha_i u_i$. Consequently $\|x\|^2 = \langle x, x\rangle = \langle \sum_{i=1}^n \alpha_i u, \sum_{j=1}^n \alpha_j u_j\rangle = \sum_{i=1}^n |\alpha_i|^2$. If we assume that $\|x\|^2 = \sum_{i=1}^k |\alpha_i|^2$, then $\alpha_j = 0$ for $j = k+1, ..., n$, or $x \in \operatorname{span}\{u_1, ..., u_k\}$.

17. (a) $d(x, y) = \|x - y\| = |-1|\ \|x - y\| = \|y - x\| = d(y, x)$.
(b) $d(x, y) = \|x - y\| = \|(x - z) - (y - z)\| \leq \|x - z\| + \|y - z\| = d(x, z) + d(y, z)$.
(c) $\|x - y\|_2 = \sqrt{1^2 + 4^2 + 5^2 + 0^2} = \sqrt{52}$.
$\|x - y\|_\infty = \max\{|x_i - y_i|\} = 5$.

19. (a) $\|x\| = 4,\ \|y\| = \sqrt{6}$.
(b) $d(x,y) = \|x - y\| = \sqrt{20}$.
(c) $\|y - x\| = d(x,y) = \sqrt{20}$ from part (b).
(d) $\|x + y\|^2 = 24$.
(e) $\cos\angle(x,y) = \dfrac{x^*y}{\|x\|\ \|y\|} = \dfrac{1-3i}{4\sqrt{6}}$ and $\angle(x,y) =$ complex arccos $\left(\frac{1-3i}{4\sqrt{6}}\right)$. [Complex trigonometric functions are defined over $\mathbb{C}$ in books and courses on complex analysis.]

21. (a) Write out $f(p+r,q)$, $f(p,q)$, and $f(r,q)$ and compare. likewise for $f(p,q+s)$.
(b) Since a polynomial of degree 0, 1, or 2 can have at most two zeros unless it is the zero polynomial, the induced function $N(p) := \sqrt{p^2(1) + 2p^2(2) + 3p^2(3)} : \mathcal{P}_2 \to \mathbb{R}$ satisfies property (d) of Definition 1. The properties (a) through (c) are obviously true. Thus N is a norm on $\mathcal{P}_2$ by Proposition 4.
(c) N is no longer definite for polynomials of degree exceeding 3 since such polynomials may have the zeros 1, 2, and 3 without being the zero function themselves. So $N(p) = 0$ does no longer imply $p = 0 \in \mathcal{P}_m$ for $m > 2$.
(d) $N(2x^2 - 3x) = \|2x_2 - 3x\| = \sqrt{(2-3)^2 + 2(6-6)^2 + 3(18-9)^2} = \sqrt{28}$.
$1 - x^2 - x^3 + 4x \in \mathcal{P}_3$ and $1 - x^2 - x^3 + 4x \notin \mathcal{P}_2$. But N is no longer a norm in $\mathcal{P}_3$; therefore asking for the length or the "induced norm" of the second polynomial makes no sense.

Index

J

L

M

N

O

P

Q

R

S

T

V

W

List of Photographs

Frank Uhlig
Department of Mathematics
Auburn University
Auburn, AL 36849–5310
uhligfd@auburn.edu
http://www.auburn.edu/~uhligfd