MAP 2210 – Aplicações de Álgebra Linear

1º Semestre - 2020

Prof. Dr. Luis Carlos de Castro Santos

lsantos@ime.usp.br

Objetivos

Formação básica de álgebra linear aplicada a problemas numéricos. Resolução de problemas em microcomputadores usando linguagens e/ou software adequados fora do horário de aula.

NÃO ESQUEÇA DE INICIAR A GRAVAÇÃO

MAP 2210 – Aplicações de Álgebra Linear

1º Semestre - 2020

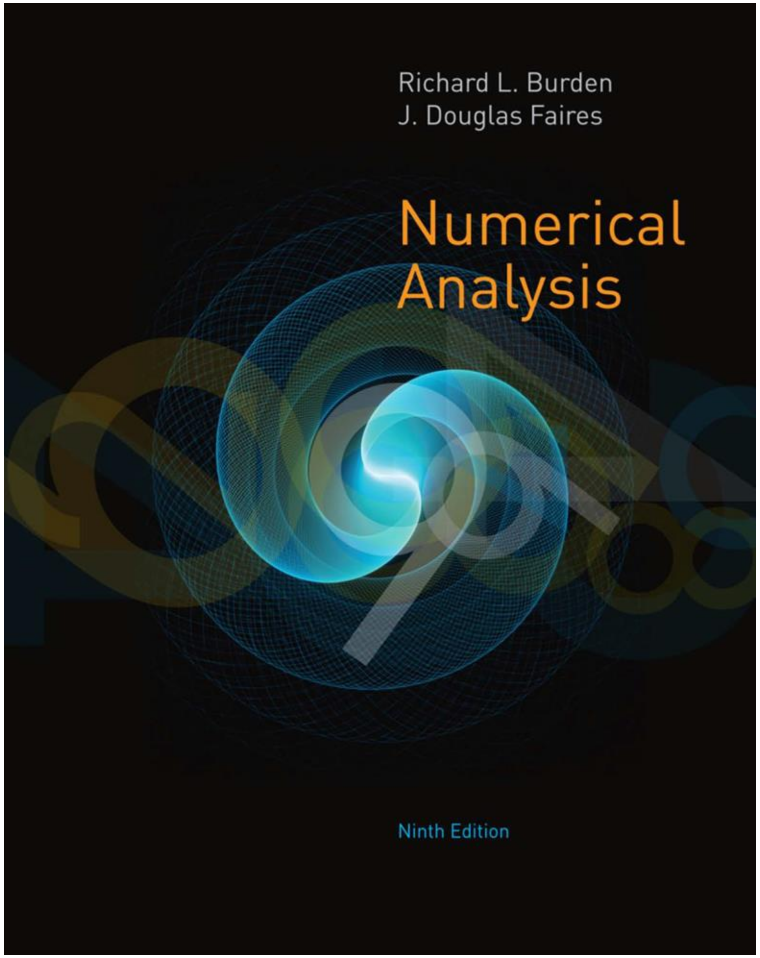
Prof. Dr. Luis Carlos de Castro Santos

lsantos@ime.usp.br

Objetivos

Formação básica de álgebra linear aplicada a problemas numéricos. Resolução de problemas em microcomputadores usando linguagens e/ou software adequados fora do horário de aula.

MAP2210



Numerical Analysis

NINTH EDITION

Richard L. Burden
Youngstown State University

J. Douglas Faires
Youngstown State University

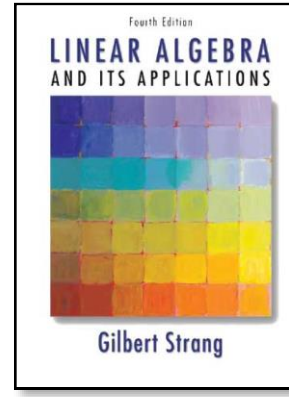
MAP2210

6 Direct Methods for Solving Linear Systems 357

7 Iterative Techniques in Matrix Algebra 431

9 Approximating Eigenvalues 561

+



+

CHAPTER 9

Approximating Eigenvalues

Introduction

The longitudinal vibrations of an elastic bar of local stiffness $p(x)$ and density $\rho(x)$ are described by the partial differential equation

$$\rho(x)\frac{\partial^2 v}{\partial t^2}(x,t) = \frac{\partial}{\partial x}\left[p(x)\frac{\partial v}{\partial x}(x,t)\right],$$

where $v(x,t)$ is the mean longitudinal displacement of a section of the bar from its equilibrium position x at time t. The vibrations can be written as a sum of simple harmonic vibrations:

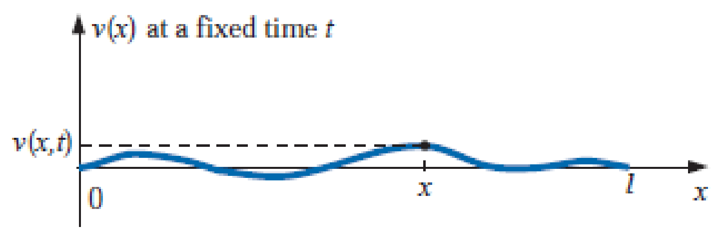
$$v(x,t) = \sum_{k=0}^{\infty} c_k u_k(x)\cos\sqrt{\lambda_k}(t-t_0),$$

where

$$\frac{d}{dx}\left[p(x)\frac{du_k}{dx}(x)\right] + \lambda_k \rho(x) u_k(x) = 0.$$

Discretização por diferenças finitas

If the bar has length l and is fixed at its ends, then this differential equation holds for $0 < x < l$ and $v(0) = v(l) = 0$.



A system of these differential equations is called a Sturm-Liouville system, and the numbers λ_k are eigenvalues with corresponding eigenfunctions $u_k(x)$.

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[p(x)\frac{\mathrm{d}u_k}{\mathrm{d}x}\right] + \lambda_k \rho(x) u_k = 0$$

Considerando *p* e *ρ* constantes, e aproximando a solução em pontos discretos do domínio

$$w_i \approx u_k(x_i)$$

Resulta na equação

$$p\ \frac{d^2\mathrm{w}}{\mathrm{d}x^2} + \lambda_k w = 0$$

Aplicando a aproximação de diferenças finitas do operador obtém-se a equação discretizada:

$$\mathrm{p}\frac{w_{i+1} - 2w_i + w_{i-1}}{\Delta x^2} + \lambda_k w_i = 0$$

$$w_{i+1} - 2w_i + w_{i-1} = -\mathrm{p}\frac{\lambda_k}{\Delta x^2} w_i$$

Aplicando a equação discretizada para cada ponto do domínio produz um problema de autovalor

Suppose the bar is 1 m long with uniform stiffness $p(x) = p$ and uniform density $\rho(x) = \rho$. To approximate u and λ, let $h = 0.2$. Then $x_j = 0.2j$, for $0 \le j \le 5$, and we can use the midpoint formula (4.5) in Section 4.1 to approximate the first derivatives. This gives the linear system

$$A\mathbf{w} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = -0.04\frac{\rho}{p}\lambda \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = -0.04\frac{\rho}{p}\lambda\mathbf{w}.$$

In this system, $w_j \approx u(x_j)$, for $1 \le j \le 4$, and $w_0 = w_5 = 0$. The four eigenvalues of A approximate the eigenvalues of the *Sturm-Liouville system*. It is the approximation of eigenvalues that we will consider in this chapter. A Sturm-Liouville application is discussed in Exercise 13 of Section 9.5.

In[15]:= $A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$

```
N[Eigenvalues[A]]
```

Out[15]= `{{2, -1, 0}, {-1, 2, -1}, {0, -1, 2}}`

Out[16]= `{3.41421, 2., 0.585786}`

In[17]:= $A = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix}$

```
N[Eigenvalues[A]]
```

Out[17]= `{{2, -1, 0, 0, 0}, {-1, 2, -1, 0, 0}, {0, -1, 2, -1, 0}, {0, 0, -1, 2, -1}, {0, 0, 0, -1, 2}}`

Out[18]= `{3.73205, 3., 2., 1., 0.267949}`

In[19]:= $A = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$

```
N[Eigenvalues[A]]
```

Out[19]=
```
{{2, -1, 0, 0, 0, 0, 0, 0, 0, 0}, {-1, 2, -1, 0, 0, 0, 0, 0, 0, 0}, {0, -1, 2, -1, 0, 0, 0, 0, 0, 0},
 {0, 0, -1, 2, -1, 0, 0, 0, 0, 0}, {0, 0, 0, -1, 2, -1, 0, 0, 0, 0},
 {0, 0, 0, 0, -1, 2, -1, 0, 0, 0}, {0, 0, 0, 0, 0, -1, 2, -1, 0, 0},
 {0, 0, 0, 0, 0, 0, -1, 2, -1, 0}, {0, 0, 0, 0, 0, 0, 0, -1, 2, -1}, {0, 0, 0, 0, 0, 0, 0, 0, -1, 2}}
```

Out[20]= `{3.91899, 3.68251, 3.30972, 2.83083, 2.28463, 1.71537, 1.16917, 0.690279, 0.317493, 0.0810141}`

9.1 Linear Algebra and Eigenvalues

Eigenvalues and eigenvectors were introduced in Chapter 7 in connection with the convergence of iterative methods for approximating the solution to a linear system. To determine the eigenvalues of an $n \times n$ matrix A, we construct the characteristic polynomial

$$p(\lambda) = \det(A - \lambda I)$$

and then determine its zeros. Finding the determinant of an $n \times n$ matrix is computationally expensive, and finding good approximations to the roots of $p(\lambda)$ is also difficult. In this chapter we will explore other means for approximating the eigenvalues of a matrix. In Section 9.6 we give an introduction to a technique for factoring a general $m \times n$ matrix into a form that has valuable applications in a number of areas.

In Chapter 7 we found that an iterative technique for solving a linear system will converge if all the eigenvalues associated with the problem have magnitude less than 1. The exact values of the eigenvalues in this case are not of primary importance—only the region of the complex plane in which they lie. An important result in this regard was first discovered by S. A. Geršgorin. It is the subject of a very interesting book by Richard Varga. [Var2]

Theorem 9.1 (Geršgorin Circle)

Let A be an $n \times n$ matrix and R_i denote the circle in the complex plane with center a_{ii} and radius $\sum_{j=1,j\neq i}^{n} |a_{ij}|$; that is,

$$R_i = \left\{ z \in \mathcal{C} \,\middle|\, |z - a_{ii}| \leq \sum_{j=1,j\neq i}^{n} |a_{ij}| \right\},$$

where $\mathcal{C}$ denotes the complex plane. The eigenvalues of A are contained within the union of these circles, $R = \cup_{i=1}^{n} R_i$. Moreover, the union of any k of the circles that do not intersect the remaining $(n - k)$ contains precisely k (counting multiplicities) of the eigenvalues. ■

Proof Suppose that λ is an eigenvalue of A with associated eigenvector $\mathbf{x}$, where $\|\mathbf{x}\|_\infty = 1$. Since $A\mathbf{x} = \lambda\mathbf{x}$, the equivalent component representation is

$$\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i, \quad \text{for each } i = 1, 2, \ldots, n. \tag{9.1}$$

Let k be an integer with $|x_k| = \|\mathbf{x}\|_\infty = 1$. When $i = k$, Eq. (9.1) implies that

$$\sum_{j=1}^{n} a_{kj}x_j = \lambda x_k.$$

Thus

$$\sum_{j=1}^{n} a_{kj}x_j = \lambda x_k - a_{kk}x_k = (\lambda - a_{kk})x_k,$$

and

$$|\lambda - a_{kk}| \cdot |x_k| = \left| \sum_{\substack{j=1,\\ j\neq k}}^{n} a_{kj}x_j \right| \le \sum_{\substack{j=1,\\ j\neq k}}^{n} |a_{kj}||x_j|.$$

But $|x_k| = \|\mathbf{x}\|_\infty = 1$, so $|x_j| \le |x_k| = 1$ for all $j = 1, 2, \ldots, n$. Hence

$$|\lambda - a_{kk}| \le \sum_{\substack{j=1,\\ j\neq k}}^{n} |a_{kj}|.$$

This proves the first assertion in the theorem, that $\lambda \in R_k$. A proof of the second statement is contained in [Var2], p. 8, or in [Or2], p. 48. ■ ■ ■

Example 1 Determine the Geršgorin circles for the matrix

$$A = \begin{bmatrix} 4 & 1 & 1 \\ 0 & 2 & 1 \\ -2 & 0 & 9 \end{bmatrix},$$

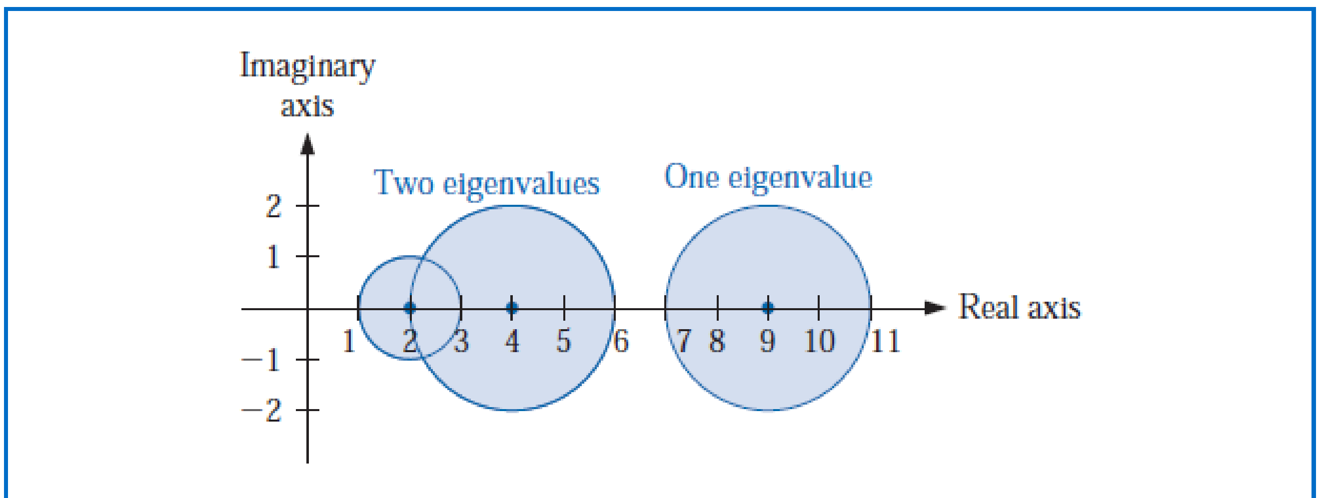
and use these to find bounds for the spectral radius of A.

Solution The circles in the Geršgorin Theorem are (see Figure 9.1)

$$R_1 = \{z \in \mathcal{C} \mid |z-4| \le 2\}, \quad R_2 = \{z \in \mathcal{C} \mid |z-2| \le 1\}, \quad \text{and} \quad R_3 = \{z \in \mathcal{C} \mid |z-9| \le 2\}.$$

Because R_1 and R_2 are disjoint from R_3, there are precisely two eigenvalues within $R_1 \cup R_2$ and one within R_3. Moreover, $\rho(A) = \max_{1 \le i \le 3} |\lambda_i|$, so $7 \le \rho(A) \le 11$. ■

Figure 9.1



Even when we need to find the eigenvalues, many techniques for their approximation are iterative. Determining regions in which they lie is the first step for finding the approximation, because it provides us with an initial approximations.

Definition 9.2 Let $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}, \ldots, \mathbf{v}^{(k)}\}$ be a set of vectors. The set is **linearly independent** if whenever

$$\mathbf{0} = \alpha_1 \mathbf{v}^{(1)} + \alpha_2 \mathbf{v}^{(2)} + \alpha_3 \mathbf{v}^{(3)} + \cdots + \alpha_k \mathbf{v}^{(k)},$$

then $\alpha_i = 0$, for each $i = 0, 1, \ldots, k$. Otherwise the set of vectors is **linearly dependent**. ■

Theorem 9.3 Suppose that $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}, \ldots, \mathbf{v}^{(n)}\}$ is a set of n linearly independent vectors in $\mathbb{R}^n$. Then for any vector $\mathbf{x} \in \mathbb{R}^n$ a unique collection of constants $\beta_1, \beta_2, \ldots, \beta_n$ exists with

$$\mathbf{x} = \beta_1 \mathbf{v}^{(1)} + \beta_2 \mathbf{v}^{(2)} + \beta_3 \mathbf{v}^{(3)} + \cdots + \beta_n \mathbf{v}^{(n)}.$$ ■

Definition 9.4 Any collection of n linearly independent vectors in $\mathbb{R}^n$ is called a **basis for** $\mathbb{R}^n$. ■

Example 2 (a) Show that $\mathbf{v}^{(1)} = (1,0,0)^t$, $\mathbf{v}^{(2)} = (-1,1,1)^t$, and $\mathbf{v}^{(3)} = (0,4,2)^t$ is a basis for $\mathbb{R}^3$, and (b) given an arbitrary vector $\mathbf{x} \in \mathbb{R}^3$ find β_1, β_2, and β_3 with

$$\mathbf{x} = \beta_1 \mathbf{v}^{(1)} + \beta_2 \mathbf{v}^{(2)} + \beta_3 \mathbf{v}^{(3)}.$$

Solution (a) Let α_1, α_2, and α_3 be numbers with $\mathbf{0} = \alpha_1 \mathbf{v}^{(1)} + \alpha_2 \mathbf{v}^{(2)} + \alpha_3 \mathbf{v}^{(3)}$. Then

$$\begin{aligned}(0,0,0)^t &= \alpha_1(1,0,0)^t + \alpha_2(-1,1,1)^t + \alpha_3(0,4,2)^t \\ &= (\alpha_1 - \alpha_2, \alpha_2 + 4\alpha_3, \alpha_2 + 2\alpha_3)^t,\end{aligned}$$

so $\alpha_1 - \alpha_2 = 0$, $\alpha_2 + 4\alpha_3 = 0$, and $\alpha_2 + 2\alpha_3 = 0$.

The only solution to this system is $\alpha_1 = \alpha_2 = \alpha_3 = 0$, so this set $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}\}$ of 3 linearly independent vectors in $\mathbb{R}^3$ is a basis for $\mathbb{R}^3$.

(b) Let $\mathbf{x} = (x_1, x_2, x_3)^t$ be a vector in $\mathbb{R}^3$. Solving

$$\begin{aligned}\mathbf{x} &= \beta_1 \mathbf{v}^{(1)} + \beta_2 \mathbf{v}^{(2)} + \beta_3 \mathbf{v}^{(3)} \\ &= \beta_1(1,0,0)^t + \beta_2(-1,1,1)^t + \beta_3((0,4,2)^t \\ &= (\beta_1 - \beta_2, \beta_2 + 4\beta_3, \beta_2 + 2\beta_3)^t\end{aligned}$$

is equivalent to solving for β_1, β_2, and β_3 in the system

$$\beta_1 - \beta_2 = x_1, \quad \beta_2 + 4\beta_3 = x_2, \beta_2 + 2\beta_3 = x_3.$$

This system has the unique solution

$$\beta_1 = x_1 - x_2 + 2x_3, \quad \beta_2 = 2x_3 - x_2, \quad \text{and} \quad \beta_3 = \frac{1}{2}(x_2 - x_3).$$ ■

Theorem 9.5 If A is a matrix and $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of A with associated eigenvectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}$, then $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}\}$ is a linearly independent set. ■

Example 3 Show that a basis can be formed for $\mathbb{R}^3$ using the eigenvectors of the 3×3 matrix

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & -1 & 4 \end{bmatrix}.$$

Solution In Example 2 of Section 7.2 we found that A has the characteristic polynomial

$$p(\lambda) = \det(A - \lambda I) = (\lambda - 3)(\lambda - 2)^2.$$

Hence there are two distinct eigenvalues of A: $\lambda_1 = 3$ and $\lambda_2 = 2$. In that example we also found that $\lambda_1 = 3$ has the eigenvector $\mathbf{x}_1 = (0, 1, 1)^t$, and that there are two linearly independent eigenvectors $\mathbf{x}_2 = (0, 2, 1)^t$ and $\mathbf{x}_3 = (-2, 0, 1)^t$ corresponding to $\lambda_2 = 2$.

It is not difficult to show (see Exercise 8) that this set of three eigenvectors

$$\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} = \{(0, 1, 1)^t, (0, 2, 1)^t, (-2, 0, 1)^t\}$$

is linearly independent and hence forms a basis for $\mathbb{R}^3$. ■

Example 4 Show that no collection of eigenvectors of the 3×3 matrix

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

can form a basis for $\mathbb{R}^3$.

Solution This matrix also has the same characteristic polynomial as the matrix A in Example 3:

$$p(\lambda) = \det \begin{bmatrix} 2-\lambda & 1 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{bmatrix} = (\lambda - 3)(\lambda - 2)^2,$$

so its eigenvalues are the same as those of A in Example 3, that is, $\lambda_1 = 3$ and $\lambda_2 = 2$.

To determine eigenvectors for B corresponding to the eigenvalue $\lambda_1 = 3$, we need to solve the system $(B - 3I)\mathbf{x} = \mathbf{0}$, so

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = (B - 3I)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_1 + x_2 \\ -x_2 \\ 0, \end{bmatrix}.$$

Hence $x_2 = 0$, $x_1 = x_2 = 0$, and x_3 is arbitrary. Setting $x_3 = 1$ gives the only linearly independent eigenvector $(0, 0, 1)^t$ corresponding to $\lambda_1 = 3$.

Consider $\lambda_2 = 2$. If

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = (B - 2\lambda)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2 \\ 0 \\ x_3, \end{bmatrix},$$

then $x_2 = 0$, $x_3 = 0$, and x_1 is arbitrary. There is only one linearly independent eigenvector corresponding to $\lambda_2 = 2$, which can be expressed as $(1, 0, 0)^t$, even though $\lambda_2 = 2$ was a zero of multiplicity 2 of the characteristic polynomial of B.

These two eigenvectors are clearly not sufficient to form a basis for $\mathbb{R}^3$. In particular, $(0, 1, 0)^t$ is not a linear combination of $\{(0, 0, 1)^t, (1, 0, 0)^t\}$. ■

We will see that when the number of linearly independent eigenvectors does not match the size of the matrix, as is the case in Example 4, there can be difficulties with the approximation methods for finding eigenvalues.

Orthogonal Vectors

Definition 9.6

A set of vectors $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}\}$ is called **orthogonal** if $(\mathbf{v}^{(i)})^t\mathbf{v}^{(j)} = 0$, for all $i \neq j$. If, in addition, $(\mathbf{v}^{(i)})^t\mathbf{v}^{(i)} = 1$, for all $i = 1, 2, \ldots, n$, then the set is called **orthonormal**. ■

Because $\mathbf{x}^t\mathbf{x} = \|\mathbf{x}\|_2^2$ for any $\mathbf{x}$ in $\mathbb{R}^n$, a set of orthogonal vectors $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}\}$ is orthonormal if and only if

$$\|\mathbf{v}^{(i)}\|_2 = 1, \quad \text{for each } i = 1, 2, \ldots, n.$$

Example 5

(a) Show that the vectors $\mathbf{v}^{(1)} = (0, 4, 2)^t$, $\mathbf{v}^{(2)} = (-5, -1, 2)^t$, and $\mathbf{v}^{(3)} = (1, -1, 2)^t$ form an orthogonal set, and **(b)** use these to determine a set of orthonormal vectors.

Solution **(a)** We have $(\mathbf{v}^{(1)})^t\mathbf{v}^{(2)} = 0(-5) + 4(-1) + 2(2) = 0$,

$(\mathbf{v}^{(1)})^t\mathbf{v}^{(3)} = 0(1) + 4(-1) + 2(2) = 0$, and $(\mathbf{v}^{(2)})^t\mathbf{v}^{(3)} = -5(1) - 1(-1) + 2(2) = 0$,

so the vectors are orthogonal, and form a basis for $\mathbb{R}^n$. The l_2 norms of these vectors are

$$\|\mathbf{v}^{(1)}\|_2 = 2\sqrt{5}, \quad \|\mathbf{v}^{(2)}\|_2 = \sqrt{30}, \quad \text{and} \quad \|\mathbf{v}^{(3)}\|_2 = \sqrt{6}.$$

(b) The vectors

$$\mathbf{u}^{(1)} = \frac{\mathbf{v}^{(1)}}{\|\mathbf{v}^{(1)}\|_2} = \left(\frac{0}{2\sqrt{5}}, \frac{4}{2\sqrt{5}}, \frac{2}{2\sqrt{5}}\right)^t = \left(0, \frac{2\sqrt{5}}{5}, \frac{\sqrt{5}}{5}\right)^t,$$

$$\mathbf{u}^{(2)} = \frac{\mathbf{v}^{(2)}}{\|\mathbf{v}^{(2)}\|_2} = \left(\frac{-5}{\sqrt{30}}, \frac{-1}{\sqrt{30}}, \frac{2}{\sqrt{30}}\right)^t = \left(-\frac{\sqrt{30}}{6}, -\frac{\sqrt{30}}{30}, \frac{\sqrt{30}}{15}\right)^t,$$

$$\mathbf{u}^{(3)} = \frac{\mathbf{v}^{(3)}}{\|\mathbf{v}^{(3)}\|_2} = \left(\frac{1}{\sqrt{6}}, \frac{-1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)^t = \left(\frac{\sqrt{6}}{6}, -\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{3}\right)^t$$

form an orthonormal set, since they inherit orthogonality from $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}$, and $\mathbf{v}^{(3)}$, and additionally,

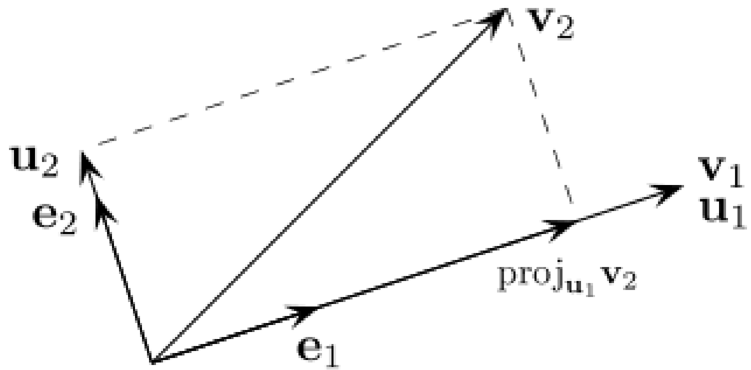
$$\|\mathbf{u}^{(1)}\|_2 = \|\mathbf{u}^{(2)}\|_2 = \|\mathbf{u}^{(3)}\|_2 = 1. \quad \blacksquare$$

Theorem 9.7 An orthogonal set of nonzero vectors is linearly independent. ■

Theorem 9.8 Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ be a set of k linearly independent vectors in $\mathbb{R}^n$. Then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ defined by

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1, \\
\mathbf{v}_2 &= \mathbf{x}_2 - \left(\frac{\mathbf{v}_1^t \mathbf{x}_2}{\mathbf{v}_1^t \mathbf{v}_1}\right)\mathbf{v}_1, \\
\mathbf{v}_3 &= \mathbf{x}_3 - \left(\frac{\mathbf{v}_1^t \mathbf{x}_3}{\mathbf{v}_1^t \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2^t \mathbf{x}_3}{\mathbf{v}_2^t \mathbf{v}_2}\right)\mathbf{v}_2, \\
&\vdots \\
\mathbf{v}_k &= \mathbf{x}_k - \sum_{i=1}^{k-1}\left(\frac{\mathbf{v}_i^t \mathbf{x}_k}{\mathbf{v}_i^t \mathbf{v}_i}\right)\mathbf{v}_i.
\end{aligned}$$

is set of k orthogonal vectors in $\mathbb{R}^n$. ■



The Gram–Schmidt process [edit]

We define the projection operator by

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u},$$

where $\langle \mathbf{u}, \mathbf{v} \rangle$ denotes the inner product of the vectors **u** and **v**. This operator projects the vector **v** orthogonally onto the line spanned by vector **u**. If **u** = **0**, we define $\text{proj}_{\mathbf{0}}(\mathbf{v}) := \mathbf{0}$. i.e., the projection map $\text{proj}_{\mathbf{0}}$ is the zero map, sending every vector to the zero vector.

The Gram–Schmidt process then works as follows:

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1, & \mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \\
\mathbf{u}_2 &= \mathbf{v}_2 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_2), & \mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} \\
\mathbf{u}_3 &= \mathbf{v}_3 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_3) - \text{proj}_{\mathbf{u}_2}(\mathbf{v}_3), & \mathbf{e}_3 &= \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \\
\mathbf{u}_4 &= \mathbf{v}_4 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_4) - \text{proj}_{\mathbf{u}_2}(\mathbf{v}_4) - \text{proj}_{\mathbf{u}_3}(\mathbf{v}_4), & \mathbf{e}_4 &= \frac{\mathbf{u}_4}{\|\mathbf{u}_4\|} \\
&\ \ \vdots & &\ \ \vdots \\
\mathbf{u}_k &= \mathbf{v}_k - \sum_{j=1}^{k-1} \text{proj}_{\mathbf{u}_j}(\mathbf{v}_k), & \mathbf{e}_k &= \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}.
\end{aligned}$$

Example 6 Use the Gram-Schmidt process to determine a set of orthogonal vectors from the linearly independent vectors

$$\mathbf{x}^{(1)} = (1,0,0)^t, \quad \mathbf{x}^{(2)} = (1,1,0)^t, \quad \text{and} \quad \mathbf{x}^{(3)} = (1,1,1)^t.$$

Solution We have the orthogonal vectors $\mathbf{v}^{(1)}$, $\mathbf{v}^{(2)}$, and $\mathbf{v}^{(3)}$, given by

$$\mathbf{v}^{(1)} = \mathbf{x}^{(1)} = (1,0,0)^t$$

$$\mathbf{v}^{(2)} = (1,1,0)^t - \left(\frac{((1,0,0)^t)^t(1,1,0)^t}{((1,0,0)^t)^t(1,0,0)^t}\right)(1,0,0)^t = (1,1,0)^t - (1,0,0)^t = (0,1,0)^t$$

$$\mathbf{v}^{(3)} = (1,1,1)^t - \left(\frac{((1,0,0)^t)^t(1,1,1)^t}{((1,0,0)^t)^t(1,0,0)^t}\right)(1,0,0)^t - \left(\frac{((0,1,0)^t)^t(1,1,1)^t}{((0,1,0)^t)^t(0,1,0)^t}\right)(0,1,0)^t$$

$$= (1,1,1)^t - (1,0,0)^t - (0,1,0)^t = (0,0,1)^t.$$

The set $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}\}$ happens to be orthonormal as well as orthogonal, but this is not commonly the situation. ■

9.2 Orthogonal Matrices and Similarity Transformations

In this section we will consider the connection between sets of vectors and matrices formed using these vectors as their columns. We first consider some results about a class of special matrices. The terminology in the next definition follows from the fact that the columns of an orthogonal matrix will form an orthogonal set of vectors.

Definition 9.9 A matrix Q is said to be **orthogonal** if its columns $\{\mathbf{q}_1^t, \mathbf{q}_2^t, \ldots, \mathbf{q}_n^t\}$ form an orthonormal set in $\mathbb{R}^n$. ■

Theorem 9.10 Suppose that Q is an orthogonal $n \times n$ matrix. Then

(i) Q is invertible with $Q^{-1} = Q^t$;

(ii) For any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, $(Q\mathbf{x})^t Q\mathbf{y} = \mathbf{x}^t\mathbf{y}$;

(iii) For any $\mathbf{x}$ in $\mathbb{R}^n$, $||Q\mathbf{x}||_2 = ||\mathbf{x}||_2$. ■

In addition, the converse of part (i) holds. (See Exercise 18.) That is,

- any invertible matrix Q with $Q^{-1} = Q^t$ is orthogonal.

As an example, the permutation matrices discussed in Section 6.5 have this property, so they are orthogonal.

Property (iii) of Theorem 9.10 is often expressed by stating that orthogonal matrices are l_2-norm preserving. As an immediate consequence of this property, every orthogonal matrix Q has $||Q||_2 = 1$.

Definition 9.11 Two matrices A and B are said to be **similar** if a nonsingular matrix S exists with $A = S^{-1}BS$. ■

An important feature of similar matrices is that they have the same eigenvalues.

Theorem 9.12 Suppose A and B are similar matrices with $A = S^{-1}BS$ and λ is an eigenvalue of A with associated eigenvector $\mathbf{x}$. Then λ is an eigenvalue of B with associated eigenvector $S\mathbf{x}$. ■

A particularly important use of similarity occurs when an $n \times n$ matrix A is similar to diagonal matrix. That is, when a diagonal matrix D and an invertible matrix S exists with

$$A = S^{-1}DS \quad \text{or equivalently} \quad D = SAS^{-1}.$$

In this case the matrix A is said to be *diagonalizable*. The following result is considered in Exercise 19.

Theorem 9.13 An $n \times n$ matrix A is similar to a diagonal matrix D if and only if A has n linearly independent eigenvectors. In this case, $D = S^{-1}AS$, where the columns of S consist of the eigenvectors, and the ith diagonal element of D is the eigenvalue of A that corresponds to the ith column of S. ■

Corollary 9.14 An $n \times n$ matrix A that has n distinct eigenvalues is similar to a diagonal matrix. ■

In fact, we do not need the similarity matrix to be diagonal for this concept to be useful. Suppose that A is similar to a triangular matrix B. The determination of eigenvalues is easy for a triangular matrix B, for in this case λ is a solution to the equation

$$0 = \det(B - \lambda I) = \prod_{i=1}^{n} (b_{ii} - \lambda)$$

if and only if $\lambda = b_{ii}$ for some i. The next result describes a relationship, called a **similarity transformation**, between arbitrary matrices and triangular matrices.

Theorem 9.15 (Schur)

Let A be an arbitrary matrix. A nonsingular matrix U exists with the property that

$$T = U^{-1}AU,$$

where T is an upper-triangular matrix whose diagonal entries consist of the eigenvalues of A. ■

The matrix U whose existence is ensured in Theorem 9.15 satisfies the condition $\|U\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ for any vector $\mathbf{x}$. Matrices with this property are called **unitary**. Although we will not make use of this norm-preserving property, it does significantly increase the application of Schur's Theorem.

Theorem 9.15 is an existence theorem that ensures that the triangular matrix T exists, but it does not provide a constructive means for finding T, since it requires a knowledge of the eigenvalues of A. In most instances, the similarity transformation U is too difficult to determine.

The following result for symmetric matrices reduces the complication, because in this case the transformation matrix is orthogonal.

Theorem 9.16 The $n \times n$ matrix A is symmetric if and only if there exists a diagonal matrix D and an orthogonal matrix Q with $A = QDQ^t$. ■

Corollary 9.17 Suppose that A is a symmetric $n \times n$ matrix. There exist n eigenvectors of A that form an orthonormal set, and the eigenvalues of A are real numbers. ■

Theorem 9.18 A symmetric matrix A is positive definite if and only if all the eigenvalues of A are positive. ■

Fim...

AULA 15