Contents

Chapter 9

Eigenanalysis

Contents

9.1 Matrix Eigenanalysis

Studied here is eigenanalysis for matrix equations. The topics are *eigenanalysis*, *eigenvalue*, *eigenvector*, *eigenpair* and *diagonalization*.

What's Eigenanalysis?

The term **eigenanalysis** refers to the identification and computation of a **new coordinate system** and **scale factors**. There is one scale factor per coordinate direction. The new coordinate system has axes with measurement units defined by the scale factors. This coordinate system is employed to simplify the expression of the original mathematical model, be it a matrix model, a differential equation model, or otherwise.

Matrix eigenanalysis is a tool for a matrix equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$.

Eigenanalysis was born from ideas in the 1822 work of J. B. Fourier on heat conduction for an insulated rod, which resulted in a simple algebraic re-scaling formula for the rod temperature: Fourier's idea is explained on page 676. His ideas apply to data analysis matrix equations $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$, systems of linear ordinary differential equations and partial differential equations of mathematical physics.

Larry Page and Sergey Brin in 1996 created from eigenanalysis a search algorithm which became *Google search*. Eigenanalysis is part of the mathematical toolset for research areas like machine learning and data mining.

Simplification of Linear Algebraic Equations

Consider the matrix equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$, where symbol A is a square matrix of constants and symbols $\vec{\mathbf{x}}, \vec{\mathbf{y}}$ are column vectors. The matrix equation is equivalent to simultaneous linear algebraic equations. For a 3×3 matrix $A = (a_{ij})$, $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ is equivalent to linear algebraic equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= y_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= y_2, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= y_3. \end{cases}$$

Table 1. Simplification of 3×3 Linear Algebraic Equations

Matrix eigenanalysis is a tool for $\vec{\mathbf{A}}\vec{\mathbf{x}} = \vec{\mathbf{b}}$, a system of linear simultaneous algebraic equations. It invents a change of variable $\vec{\mathbf{x}} \to \vec{\mathbf{X}}$, $\vec{\mathbf{b}} \to \vec{\mathbf{B}}$ that simplifies the system of equations.

A change of variables $\vec{\mathbf{X}} = P\vec{\mathbf{x}}$, $\vec{\mathbf{B}} = P\vec{\mathbf{b}}$ with eigenanalysis vectors $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ for the columns of P simplifies a 3×3 system of linear algebraic equations $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ into the diagonal form

$$(1) \qquad \begin{cases} \lambda_1 X_1 &= B_1, \\ \lambda_2 X_2 &= B_2, \\ \lambda_3 X_3 &= B_3. \end{cases}$$

Scalar values $\lambda_1, \lambda_2, \lambda_3$ are scale factors (measurement units) corresponding to the directions $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$. Precise definitions are on page 669.

Coordinate Change using Eigenanalysis

Technically, matrix eigenanalysis is an **opportunistic change of coordinates**, which means the analysis must compute a set of **independent** column vectors that span $\mathcal{R}^n$. Linear algebra calls such a set of vectors a **basis**. Eigenanalysis constructs from square matrix A a *special basis*. This special basis defines a change of coordinates $\vec{\mathbf{x}} \to P\vec{\mathbf{x}}$ where P is the augmented matrix of constructed basis vectors.

Consider vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ which form a basis for $\mathcal{R}^3$. To be a *basis* means that *each possible* vector $\vec{\mathbf{x}}$ in $\mathcal{R}^3$ can be uniquely expressed as a linear combination $\vec{\mathbf{x}} = c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3$. Geometrically, the triad $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ must define a parallelepiped of positive volume. For a triad basis $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$, each possible $\vec{\mathbf{x}}$ in $\mathcal{R}^3$ can be constructed from the triad using solely the geometric parallelogram law for vector addition.

The claimed simplifying change of coordinates[1] is defined by:

$$
(2) \qquad \begin{array}{l}
P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle = \text{augmented matrix}, \\
\vec{\mathbf{X}} = P\vec{\mathbf{x}}, \quad \vec{\mathbf{B}} = P\vec{\mathbf{b}}, \quad \text{a change of } A\vec{\mathbf{x}} = \vec{\mathbf{b}} \text{ into } D\vec{\mathbf{X}} = \vec{\mathbf{B}}, \\
D = \textbf{diag}(\lambda_1, \lambda_2, \lambda_3) \quad \text{a diagonal matrix of scale factors}
\end{array}
$$

Details on page 675.

Eigenvalue, Eigenvector and Eigenpair Defined

Eigenanalysis for the matrix equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ when matrix A is 3×3 is an algebraic method for discovering basis vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ and scale factors $\lambda_1, \lambda_2, \lambda_3$. The vectors are called **eigenvectors** and the scale factors are called **eigenvalues**.

A scale factor λ is thought to be a **measurement unit** along an axis $\vec{\mathbf{v}}$, therefore the eigenvectors and eigenvalues occur in pairs, called **eigenpairs**. Pairing is due to fundamental equation (3) below, which is used in references to define and/or compute an eigenpair.

Definition 9.1 (Eigenpair)
An **Eigenpair** $(\lambda, \vec{\mathbf{v}})$ is defined to be a solution of the problem

$$
(3) \qquad A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}, \quad \vec{\mathbf{v}} \neq \vec{\mathbf{0}}.
$$

Vector $\vec{\mathbf{v}}$ is called an **eigenvector**. The value λ is called the **eigenvalue** corresponding to the eigenvector $\vec{\mathbf{v}}$.

Important. Because $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$ in equation (3), then an eigenvector is never the zero vector: an eigenvector is a *direction*. Otherwise stated:

An eigenvector answer of zero signals an algebra error.

Motivation for the rather abstract definition of eigenpair appears below. Excuses aside, definition (3) *must be learned and memorized*, because of explicit use in computations and implicit use in literature.

[1]The triad $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ in **principal coordinate analysis** and **metric scaling** simplifies the data set to find trends and important parameters.

Why the Equation $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$?

The pattern is $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$. However, it is **not** the problem being solved. The maddening historical event of algebraists stripping away the problem from the definition impacts everyone trying to learn eigenanalysis.

The algebraists' Definition 9.1 is a sub-problem. It is madness to try to learn eigenanalysis from it. Learning from it parallels trying to learn about trees by crawling on the ground through the forest examining tree trunks.

Assume matrix A is 3×3. The **problem to be solved** is computation of independent vectors $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ to find an opportunistic change of variables that simplifies the linear algebraic system of equations $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$.

Algebraists were quick to discover that the problem is solved by finding a basis $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ of $\mathcal{R}^3$ satisfying the **three equations** (4) *infra*. They isolated $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$ as a sub-problem to be solved many times, in order to find the basis.

History of Eigenvector and Eigenvalue Terminology

James J. Sylvester in 1883 coined the term **latent root** for what has become the term **eigenvalue**:

> ...the latent roots of a matrix – latent in a somewhat similar sense as vapour may be said to be latent in water or smoke in a tobacco-leaf.

The German term *eigenwert* was coined by David Hilbert in 1904. By 1967, Paul Halmos gave up the battle over which words to use in his new book *A Hilbert Space Problem Book*. The battle: German *eigen* means **proper**, *wert* means **value**.

> For many years I have battled for **proper values** and against the one and a half times translated German-English hybrid (Halmos means **eigenvalue**) that is often used to refer to them. I have now become convinced that the war is over, and eigenvalues have won it; in this book I use them.

No longer used are the historical terms *hidden value*, *proper value*, *characteristic value* and *latent root*. The term *hidden* arose because the vectors and scale factors are generally impossible to determine from matrix A without computation. What has persisted in literature is the *characteristic equation*, the equation which determines eigenvalues. See Theorem 9.2.

Eigenpair Equations and $AP = PD$

Eigenpair equations for a square matrix A can be written by matrix multiply as a single equation.

Theorem 9.1 (Eigenpairs and $AP = PD$)
Assume $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ independent in $\mathcal{R}^3$. Let matrix A be 3×3. Then relations

$$(4) \qquad \left\{ \begin{array}{rcl} A\vec{\mathbf{v}}_1 &=& \lambda_1\vec{\mathbf{v}}_1, \\ A\vec{\mathbf{v}}_2 &=& \lambda_2\vec{\mathbf{v}}_2, \\ A\vec{\mathbf{v}}_3 &=& \lambda_3\vec{\mathbf{v}}_3. \end{array} \right. \qquad \textbf{(Eigenpair Equations)}$$

hold if and only if $AP = PD$ where P and D are defined by equations

$$P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle, \quad D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \tag{5}$$

The result holds for dimension n. Proof on page 693.

Computing Eigenpairs of a Matrix

To compute an eigenpair $(\lambda, \vec{\mathbf{v}})$ of a square matrix A requires finding scalar λ and a nonzero vector $\vec{\mathbf{v}}$ satisfying the homogeneous matrix–vector equation

$$A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}.$$

Write it as $A\vec{\mathbf{x}} - \lambda\vec{\mathbf{x}} = \vec{\mathbf{0}}$, then replace $\lambda\vec{\mathbf{x}}$ by $\lambda I\vec{\mathbf{x}}$ to obtain the standard homogeneous linear algebraic system form[2]

$$(A - \lambda I)\vec{\mathbf{v}} = \vec{\mathbf{0}}, \quad \vec{\mathbf{v}} \neq \vec{\mathbf{0}}.$$

Definition 9.2 (Characteristic Equation)
Determinant equation $|A - \lambda I| = 0$ is called the **characteristic equation**. The **characteristic polynomial** is the polynomial obtained by determinant evaluation on the left, normally by cofactor expansion or the triangular rule.

Theorem 9.2 (Eigenvalues of A)
The eigenvalues of a square matrix A are exactly all the roots λ of the polynomial equation

$$\det(A - \lambda I) = 0.$$

Proof on page 693

Theorem 9.3 (Find Eigenvectors of Matrix A)
For each root λ of the characteristic equation $|A - \lambda I| = 0$, form matrix $B = A - \lambda I$. Write a toolkit sequence to $\mathbf{rref}(B)$. Solve the homogeneous equation $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ for $\vec{\mathbf{v}}$ in terms of invented symbols $t_1, t_2, \ldots$.

A basis of eigenvectors of A for eigenvalue λ is the list of vectors $\partial_{t_1}\vec{\mathbf{v}}, \partial_{t_2}\vec{\mathbf{v}}, \ldots$. They are Strang's **special solutions** of $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$, known to be independent.

These eigenvectors span the nullspace (kernel) of B: if $A\vec{\mathbf{w}} = \lambda\vec{\mathbf{w}}$, then $\vec{\mathbf{w}}$ is a linear combination of these basis vectors.

Proof on page 694.

[2]Identity I is required to factor out the matrix $A - \lambda I$. It is wrong to factor out $A - \lambda$, because A is 3×3 and λ is 1×1, incompatible sizes for matrix addition.

Characteristic Equation Illustration.

$$\begin{aligned}\det\left(\begin{pmatrix}1&3\\1&2\end{pmatrix}-\lambda\begin{pmatrix}1&0\\0&1\end{pmatrix}\right) &= \begin{vmatrix}1-\lambda & 3\\ 1 & 2-\lambda\end{vmatrix}\\ &= (1-\lambda)(2-\lambda)-6\\ &= \lambda^2-3\lambda-4\\ &= (\lambda+1)(\lambda-4).\end{aligned}$$

The characteristic equation $\lambda^2 - 3\lambda - 4 = 0$ has roots $\lambda_1 = -1$, $\lambda_2 = 4$. The characteristic polynomial is $\lambda^2 - 3\lambda - 4$.

Table 2. Shortcut for the Characteristic Polynomial

To find the characteristic polynomial $|A - \lambda I|$, subtract symbol λ from the diagonal of A and then evaluate the determinant.

Key Examples for Finding Eigenvectors

Assume given a 3×3 matrix A. Found after at most 3 applications of Theorem 9.3 is a list of eigenpairs with independent eigenvectors.

There might not be 3 answers!

The amount of work on paper and pencil varies with the number of repeated eigenvalues. Key examples:

$$\boxed{1}\ \begin{pmatrix}1&0&1\\0&2&4\\0&0&3\end{pmatrix},\quad \boxed{2}\ \begin{pmatrix}1&0&1\\0&1&4\\0&0&1\end{pmatrix},\quad \boxed{3}\ \begin{pmatrix}1&0&1\\0&1&4\\0&0&2\end{pmatrix}$$

$\boxed{1}$ Matrix A has eigenvalues 1, 2, 3. Apply Theorem 9.3 three times to write three different matrices B. Each B has a toolkit sequence to $\mathbf{rref}(B)$, a total of 3 toolkit sequences. Each sequence produces one eigenvector: *there are 3 answers*.

$\boxed{2}$ Matrix A has eigenvalues 1, 1, 1. Apply Theorem 9.3 one time to write one matrix B. There is just 1 toolkit sequence to $\mathbf{rref}(B)$. Because of 2 free variables, *there are 2 answers*. In general, the number of free variables is 1, 2 or 3 with correspondingly 1,2 or 3 answers.

$\boxed{3}$ Matrix A has eigenvalues 1, 1, 2. Apply Theorem 9.3 two times to write two matrices B. Each B has a toolkit sequence to $\mathbf{rref}(B)$, a total of 2 toolkit sequences. Eigenvalue 1 has a basis of 2 eigenvectors, caused by 2 free variables. Eigenvalue 2 has a basis of just one eigenvector, caused by only 1 free variable.

In general, the number of answers for a repeated eigenvalue equals the number of free variables for the toolkit sequence B to $\mathbf{rref}(B)$.

Independence of Eigenvectors

Theorem 9.4 (Independence of Eigenvectors)
If $(\lambda_1, \vec{\mathbf{v}}_1)$ and $(\lambda_2, \vec{\mathbf{v}}_2)$ are two eigenpairs of A and $\lambda_1 \neq \lambda_2$, then $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ are linearly independent vectors.

More generally, if $(\lambda_1, \vec{\mathbf{v}}_1)$, ..., $(\lambda_k, \vec{\mathbf{v}}_k)$ are eigenpairs of A corresponding to distinct eigenvalues λ_1, ..., λ_k, then $\vec{\mathbf{v}}_1$, ..., $\vec{\mathbf{v}}_k$ are independent.

Proof on page 694

Theorem 9.5 (Unions of Eigenvectors)
Let A be an $n \times n$ matrix A. Let variable λ denote an arbitrary eigenvalue of A. Let λ_1, ..., λ_k be a list of distinct eigenvalues of A.

Let $\mathcal{B}(\lambda)$ be some basis for the eigenpair equation $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$. Then

(1) For $\lambda \neq \mu$, subspaces $\mathbf{span}(\mathcal{B}(\lambda))$ and $\mathbf{span}(\mathcal{B}(\mu))$ intersect in only the zero vector.

(2) The union U of bases $\mathcal{B}(\lambda_1)$, ..., $\mathcal{B}(\lambda_k)$ is a list of independent vectors in $\mathcal{C}^n$.[3]

(3) If all eigenvalues are real, then $\mathcal{C}^n$ can be replaced by $\mathcal{R}^n$ in results (1), (2).

Proof on page 694

Complete Set of Eigenvectors

Definition 9.3 (Complete Set of Eigenvectors)
A list $U = \{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_k\}$ of independent eigenvectors of an $n \times n$ matrix A is called **complete** provided $k = n$.

Lemma 9.1 (Invertible Change of Variables) Let $U = \{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_n\}$ be a list of independent eigenvectors of an $n \times n$ matrix A. Assume all eigenvalues are real. Define augmented $n \times n$ matrix $P = \left\langle \vec{\mathbf{v}}_1 | \cdots | \vec{\mathbf{v}}_n \right\rangle$. Then:

The eigenvectors span $\mathcal{R}^n$: $\mathbf{span}(U) = \mathcal{R}^n$.

Matrix P is invertible.

Proof: A list U of n independent vectors in $\mathcal{R}^n$ is a basis. Then U spans $\mathcal{R}^n$. An $n \times n$ matrix with independent columns is invertible. ■

[3]Symbol $\mathcal{C}^n$ is the vector space of n-vectors with complex entries.

Theorem 9.6 (Finding Independent Eigenvectors)
Let $n \times n$ matrix A be given. Solve the characteristic equation $|A - \lambda I| = 0$ for all eigenvalues λ. For each λ, let $B = A - \lambda I$ and solve $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ for general solution $\vec{\mathbf{v}}$, which contains invented symbols $t_1, t_2, \ldots$. Let $\mathcal{B}(\lambda)$ be the list of vector partial derivatives $\partial_{t_1}\vec{\mathbf{v}}, \partial_{t_2}\vec{\mathbf{v}}, \ldots$. Then the union U of all lists $\mathcal{B}(\lambda)$ is a set of independent eigenvectors. Examples exist where U is *not* a basis for $\mathcal{R}^n$.

Proof on page 696.

Eigenanalysis Facts

1. An **eigenvalue** λ of a triangular matrix A is one of the diagonal entries. If A is non-triangular, then an eigenvalue is found as a root λ of the characteristic equation $|A - \lambda I| = 0$.
2. An **eigenvalue** of a square matrix A can be zero, positive, negative or even complex. It is a pure number, with a physical meaning inherited from the model, e.g., a scale factor or measurement unit.
3. An **eigenvector** for eigenvalue λ (a scale factor) is a nonzero direction $\vec{\mathbf{v}}$ of application satisfying $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$. It is found from a toolkit sequence starting at $B = A - \lambda I$ and ending at $\mathbf{rref}(B)$. Independent eigenvectors are computed from the general solution of $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ as partial derivatives $\partial\vec{\mathbf{v}}/\partial t_1$, $\partial\vec{\mathbf{v}}/\partial t_2$, $\ldots$.
4. If a 3×3 matrix has three independent real eigenvectors, then they collectively form a **basis** of $\mathcal{R}^3$ (a **coordinate system**).

Diagonalization and Eigenpair Packages

Definition 9.4 (Diagonalizable Matrix)
An $n \times n$ matrix A which has n independent eigenvectors is called **diagonalizable**. The eigenvalues are not required to be distinct.

Given a diagonalizable 3×3 system $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$, the augmented matrix $P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle$ of eigenvectors and diagonal matrix $D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$ provide a variable change $\vec{\mathbf{X}} = P\vec{\mathbf{x}}, \vec{\mathbf{Y}} = P\vec{\mathbf{y}}$ to transform system $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ into the simplified diagonal system $\vec{\mathbf{Y}} = D\vec{\mathbf{X}}$.

Theorem 9.7 (Diagonalization and Diagonal Matrices)

A 3×3 diagonal matrix $A = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$ has eigenvalues on the diagonal. The eigenvectors are the columns of the 3×3 identity matrix:

$$\lambda_1 = a, \qquad \lambda_2 = b, \qquad \lambda_3 = c,$$
$$\vec{\mathbf{v}}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \vec{\mathbf{v}}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \vec{\mathbf{v}}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

The theorem extends to $n\times n$ matrices. Every $n\times n$ diagonal matrix is diagonalizable.

Definition 9.5 (Eigenpair Packages)
Let A be a diagonalizable 3×3 matrix with eigenpairs $(\lambda_1, \vec{\mathbf{v}}_1)$, $(\lambda_2, \vec{\mathbf{v}}_2)$, $(\lambda_3, \vec{\mathbf{v}}_3)$. Define **eigenpair packages** by:[4]

$$P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle, \quad D = \begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{pmatrix}. \tag{6}$$

Package definitions for an $n \times n$ matrix:

$$P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \cdots | \vec{\mathbf{v}}_n \right\rangle, \quad D = \begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ & & \vdots & \\ 0 & 0 & \cdots & \lambda_n\end{pmatrix}$$

If all eigenvalues are real then both P and D are real. Otherwise, matrices P and D will have complex entries.

Theorem 9.8 (Diagonalization)
Let A be a diagonalizable $n \times n$ matrix with eigenpair packages P, D.

1. The matrix A is completely determined by its eigenpairs:
$$A = PDP^{-1}.$$
2. The change of variables $\vec{\mathbf{X}} = P\vec{\mathbf{x}}$, $\vec{\mathbf{Y}} = P\vec{\mathbf{y}}$ transforms the equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ into the diagonal system $\vec{\mathbf{Y}} = D\vec{\mathbf{X}}$.
3. The equation $A(c_1\vec{\mathbf{v}}_1 + \cdots + c_n\vec{\mathbf{v}}_n) = c_1\lambda_1\vec{\mathbf{v}}_1 + \cdots + c_n\lambda_n\vec{\mathbf{v}}_n$ holds for any constants $c_1, \ldots, c_n$ with matrix form
$$AP\vec{\mathbf{c}} = PD\vec{\mathbf{c}}, \quad \vec{\mathbf{c}} = \begin{pmatrix}c_1\\ \vdots\\ c_n\end{pmatrix}. \tag{7}$$

See Fourier Replacement page 676.

[4] Eigenpair packages are not unique. For 3×3, there are six (6) permutations of the pairs, leading to six different packages. In addition, eigenvectors are not unique, leading to infinitely many possible eigenpair packages.

Proof on page 696.

Theorem 9.9 (Distinct Eigenvalues implies Diagonalizable)
If an $n \times n$ matrix A has n distinct eigenvalues, real or complex, then it has n eigenpairs $(\lambda_i, \vec{\mathbf{v}}_i)$, $i = 1, \ldots, n$. The eigenpair packages

$$P = \left\langle \vec{\mathbf{v}}_1 | \cdots | \vec{\mathbf{v}}_n \right\rangle, \quad D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

satisfy $AP = PD$ and matrix A is diagonalizable.

Proof on page 697.

Fourier Replacement

The subject of **eigenanalysis** was popularized by J. B. Fourier in his 1822 publication on the theory of heat, *Théorie analytique de la chaleur*. Fourier's ideas can be summarized for the $n \times n$ matrix equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$:

Vector $A\vec{\mathbf{x}}$ is obtained from $\vec{\mathbf{x}}$ and a complete set of eigenpairs $(\lambda_1, \vec{\mathbf{v}}_1)$, $(\lambda_2, \vec{\mathbf{v}}_2)$, ..., $(\lambda_n, \vec{\mathbf{v}}_n)$ by replacing the eigenvectors by their scaled versions $\lambda_1\vec{\mathbf{v}}_1$, ..., $\lambda_n\vec{\mathbf{v}}_n$:

$$(8) \qquad \begin{array}{rclclclcl} \vec{\mathbf{x}} & = & c_1\vec{\mathbf{v}}_1 & + & c_2\vec{\mathbf{v}}_2 & + & \cdots & + & c_n\vec{\mathbf{v}}_n \quad \text{implies} \\ A\vec{\mathbf{x}} & = & c_1\lambda_1\vec{\mathbf{v}}_1 & + & c_2\lambda_2\vec{\mathbf{v}}_2 & + & \cdots & + & c_n\lambda_n\vec{\mathbf{v}}_n. \end{array}$$

See Example 9.10 page 690 for details about the heat problem.

For the case of $\mathcal{R}^3$, basis vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ are re-scaled by invented **scale factors** $\lambda_1, \lambda_2, \lambda_3$, which we imagine as measurement units along the three directions $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$. Fourier's 1822 idea: vector $\vec{\mathbf{x}}$ is replaced by a new vector $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$, according to the rule

$$(9) \qquad \begin{array}{rcl} \vec{\mathbf{x}} & = & c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3 \text{ implies} \\ \vec{\mathbf{y}} & = & c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\vec{\mathbf{v}}_3. \end{array}$$

Table 3. Fourier's 1822 Re-Scaling Idea

Replace $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ by re-scaled vectors $\lambda_1\vec{\mathbf{v}}_1$, $\lambda_2\vec{\mathbf{v}}_2$, $\lambda_3\vec{\mathbf{v}}_3$.

Criticism: Table 3 makes no mention of a matrix A. Fourier's re-scaling idea does not need a matrix A, but it resurfaces:

Theorem 9.10 (Matrix Form of Fourier Replacement)
Let vectors $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ be independent. Let $\lambda_1, \lambda_2, \lambda_3$ be scalars. Define

$$P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle, \quad D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}, \quad \vec{\mathbf{c}} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

Fourier replacement is defined by

$$\begin{array}{rcl} \vec{\mathbf{x}} & = & c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3 \text{ implies} \\ \vec{\mathbf{y}} & = & c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\vec{\mathbf{v}}_3 \end{array} \qquad \text{for all scalars } c_1, c_2, c_3$$

The statement has vector-matrix forms

$$\begin{array}{l} \vec{\mathbf{x}} = P\vec{\mathbf{c}} \quad \text{implies} \quad \vec{\mathbf{y}} = PD\vec{\mathbf{c}} \\ \vec{\mathbf{y}} = A\vec{\mathbf{x}} \quad \text{where} \quad A = PDP^{-1} \\ A\left(c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3\right) = c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\vec{\mathbf{v}}_3 \end{array}$$

The theorem extends to $n \times n$. Proof on page 697.

Theorem 9.11 (Fourier Re-scaling and Diagonalization)
Let vectors $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ be independent. Let $\lambda_1, \lambda_2, \lambda_3$ be scalars. Define $P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle$, $D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$.

(a) Matrix $A = PDP^{-1}$ has 3 eigenpairs $(\lambda_1, \vec{\mathbf{v}}_1)$, $(\lambda_2, \vec{\mathbf{v}}_2)$, $(\lambda_3, \vec{\mathbf{v}}_3)$ and A is diagonalizable.

(b) If a diagonalizable 3×3 matrix has eigenpairs $(\lambda_1, \vec{\mathbf{v}}_1)$, $(\lambda_2, \vec{\mathbf{v}}_2)$, $(\lambda_3, \vec{\mathbf{v}}_3)$ with independent eigenvectors, then Fourier replacement (8) holds.

(c) Fourier replacement for matrix equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ defined in (8) is equivalent to diagonalizability of matrix A.

Proof on page 697.

Re-scaling Example: Data Conversion

Let $\vec{\mathbf{x}}$ in $\mathcal{R}^3$ be a data set variable with coordinates x_1, x_2, x_3 recorded respectively in units of meters, millimeters and centimeters. Imagine the data being recorded every few milliseconds from three different sensors.

The $\vec{\mathbf{x}}$-data set is converted into a $\vec{\mathbf{y}}$-data set with meter, kilogram, second units (MKS units) via the equations

$$\left\{ \begin{array}{rcl} y_1 & = & x_1, \\ y_2 & = & 0.001x_2, \\ y_3 & = & 0.01x_3. \end{array} \right. \tag{10}$$

Equations (10) are an instance of Fourier's re-scaling process, Table 3. The paired scale factors and vectors are

$$\lambda_1 = 1, \quad \lambda_2 = 0.001, \quad \lambda_3 = 0.01,$$
$$\vec{\mathbf{v}}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \vec{\mathbf{v}}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \vec{\mathbf{v}}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

Then equations (10) can be written as the replacement process

$$\begin{aligned} \vec{\mathbf{x}} &= x_1\begin{pmatrix}1\\0\\0\end{pmatrix} + x_2\begin{pmatrix}0\\1\\0\end{pmatrix} + x_3\begin{pmatrix}0\\0\\1\end{pmatrix} \quad \text{implies} \\ \vec{\mathbf{y}} &= x_1\lambda_1\begin{pmatrix}1\\0\\0\end{pmatrix} + x_2\lambda_2\begin{pmatrix}0\\1\\0\end{pmatrix} + x_3\lambda_3\begin{pmatrix}0\\0\\1\end{pmatrix}. \end{aligned} \tag{11}$$

Vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ are the *data directions* (or *axes*) re-scaled by the measurement units λ_1, λ_2, λ_3, respectively. In particular, data direction $\vec{\mathbf{v}}_2$ is for millimeters and scale factor $\lambda_2 = 0.001$ is the measurement unit along axis $\vec{\mathbf{v}}_2$. Theorem 9.10 applied to (11) gives $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ where $A = \begin{pmatrix}1 & 0 & 0\\ 0 & \frac{1}{1000} & 0\\ 0 & 0 & \frac{1}{100}\end{pmatrix}$, agreeing with conversion of (10) to matrix form.

Fourier Replacement: Matrix Example

Let

$$\begin{aligned} A &= \begin{pmatrix}1 & 3 & 0\\ 0 & 2 & -1\\ 0 & 0 & -5\end{pmatrix} \\ \lambda_1 &= \mathbf{1}, \quad \lambda_2 = \mathbf{2}, \quad \lambda_3 = \mathbf{-5}, \\ \vec{\mathbf{v}}_1 &= \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \vec{\mathbf{v}}_2 = \begin{pmatrix}3\\1\\0\end{pmatrix}, \quad \vec{\mathbf{v}}_3 = \begin{pmatrix}1\\-2\\-14\end{pmatrix}. \end{aligned} \tag{12}$$

Then Fourier's model (9) holds, details in Example 9.3:

$$\begin{aligned} \vec{\mathbf{x}} &= c_1\begin{pmatrix}1\\0\\0\end{pmatrix} + c_2\begin{pmatrix}3\\1\\0\end{pmatrix} + c_3\begin{pmatrix}1\\-2\\-14\end{pmatrix} \\ &\text{implies} \\ A\vec{\mathbf{x}} &= c_1(\mathbf{1})\begin{pmatrix}1\\0\\0\end{pmatrix} + c_2(\mathbf{2})\begin{pmatrix}3\\1\\0\end{pmatrix} + c_3(\mathbf{-5})\begin{pmatrix}1\\-2\\-14\end{pmatrix} \end{aligned}$$

Eigenanalysis and Geometry

In case the matrix A is 2×2 or 3×3, geometry can provide additional intuition about eigenanalysis.

Fourier's 2×2 replacement $A(c_1\vec{v}_1 + c_2\vec{v}_2) = c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2$ can be interpreted as the action of the transformation $T: \ \vec{x} \to A\vec{x}$ between two copies of the plane $\mathcal{R}^2$; see Figure 1.

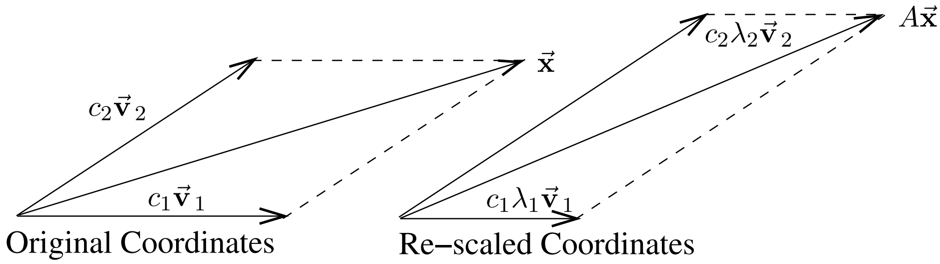


Figure 1. Transformation $T: \ \mathcal{R}^2 \to \mathcal{R}^2$.
Vector $\vec{\mathbf{x}}$ is obtained geometrically from $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2$ by changing their lengths by c_1, c_2, then add with the parallelogram rule. Vector $A\vec{\mathbf{x}}$ is obtained from the two changed vectors by re-scaling by λ_1, λ_2, then apply the parallelogram rule.

Algebraically, A is replaced by the scale factors λ_1, λ_2 and the coordinate system $\vec{v}_1, \vec{v}_2$. The **eigenvalues** are the scale factors λ_1, λ_2. Vectors $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2$ used in the parallelogram rule are the **eigenvectors**.

Shear is not Equivalent to Scaling along Axes

The important geometrical operations are scaling, shears, rotations, projections, reflections and translations. Fourier replacement describes scaling along coordinate axes.

A **planar horizontal shear** $(x_1, x_2) \to (y_1, y_2)$ is a set of equations

$$\begin{aligned} y_1 &= x_1 + kx_2, \quad (k = \text{shear factor} \neq 0), \\ y_2 &= x_2. \end{aligned}$$

The eigenvalues of $A = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ are $\lambda_1 = \lambda_2 = 1$. Assume it is possible to view this shear as a re-scaling. Then it must be feasible to change coordinates to new independent axes $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2$ and express the shear as

$$A = PDP^{-1}, \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 \right\rangle.$$

Then $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} = A = PDP^{-1} = P\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} P^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, a contradiction to the shear factor requirement $k \neq 0$.

Conclusion: A shear is not equivalent to scaling along axes. Fourier replacement fails.

Examples and Methods

Example 9.1 (Computing 2×2 Eigenpairs)

Find all eigenpairs of the 2×2 matrix $A = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}$.

Solution:
The method used to solve for eigenpairs in given in Theorem 9.3 page 671.

College Algebra. The eigenvalues are $\lambda_1 = 1$, $\lambda_2 = -1$. Details:

$$\begin{aligned} 0 &= \det(A - \lambda I) && \text{Characteristic equation.} \\ &= \begin{vmatrix} 1-\lambda & 0 \\ 2 & -1-\lambda \end{vmatrix} && \text{Subtract } \lambda \text{ from the diagonal.} \\ &= (1-\lambda)(-1-\lambda) && \text{Sarrus' rule.} \end{aligned}$$

Linear Algebra. The eigenpairs are $\left(1, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right), \left(-1, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$. Details:

Eigenvector for $\lambda_1 = 1$.

$$\begin{aligned} A - \lambda_1 I &= \begin{pmatrix} 1-\lambda_1 & 0 \\ 2 & -1-\lambda_1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 2 & -2 \end{pmatrix} \\ &\approx \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} && \text{Swap and multiply rules.} \\ &= \mathbf{rref}(A - \lambda_1 I) && \text{Reduced echelon form.} \end{aligned}$$

The vector partial derivative $\partial_{t_1}\vec{\mathbf{v}}$ of the scalar general solution $x = t_1$, $y = t_1$ is eigenvector $\vec{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Eigenvector for $\lambda_2 = -1$.

$$\begin{aligned} A - \lambda_2 I &= \begin{pmatrix} 1-\lambda_2 & 0 \\ 2 & -1-\lambda_2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} \\ &\approx \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} && \text{Combination and multiply.} \\ &= \mathbf{rref}(A - \lambda_2 I) && \text{Reduced echelon form.} \end{aligned}$$

The vector partial derivative $\partial_{t_1}\vec{\mathbf{v}}$ of the scalar general solution $x = 0$, $y = t_1$ is eigenvector $\vec{\mathbf{v}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Example 9.2 (Computing 2×2 Complex Eigenpairs)

Find all eigenpairs of the 2×2 matrix $A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$.

Solution:

Reference: Theorem 9.3 page 671.

College Algebra. The eigenvalues are $\lambda_1 = 1 + 2i$, $\lambda_2 = 1 - 2i$. Details:

$$\begin{aligned} 0 &= \det(A - \lambda I) && \text{Characteristic equation.} \\ &= \begin{vmatrix} 1-\lambda & 2 \\ -2 & 1-\lambda \end{vmatrix} && \text{Subtract } \lambda \text{ from the diagonal.} \\ &= (1-\lambda)^2 + 4 && \text{Sarrus' rule.} \end{aligned}$$

The roots $\lambda = 1 \pm 2i$ are found from the quadratic formula after expanding $(1-\lambda)^2 + 4 = 0$. Alternatively, use $(1-\lambda)^2 = -4$ and take square roots.

Linear Algebra. The eigenpairs are $\left(1 + 2i, \begin{pmatrix} -i \\ 1 \end{pmatrix}\right)$, $\left(1 - 2i, \begin{pmatrix} i \\ 1 \end{pmatrix}\right)$.

Eigenvector for $\lambda_1 = 1 + 2i$.

$$\begin{aligned} A - \lambda_1 I &= \begin{pmatrix} 1-\lambda_1 & 2 \\ -2 & 1-\lambda_1 \end{pmatrix} \\ &= \begin{pmatrix} -2i & 2 \\ -2 & -2i \end{pmatrix} \\ &\approx \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} && \text{Multiply rule.} \\ &\approx \begin{pmatrix} 0 & 0 \\ 1 & i \end{pmatrix} && \text{Combination rule, multiplier}{=}{-i}. \\ &\approx \begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix} && \text{Swap rule.} \\ &= \mathbf{rref}(A - \lambda_1 I) && \text{Reduced echelon form.} \end{aligned}$$

The partial derivative $\partial_{t_1}\vec{\mathbf{v}}$ of the general solution $x = -it_1$, $y = t_1$ is eigenvector $\vec{\mathbf{v}}_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}$.

Eigenvector for $\lambda_2 = 1 - 2i$.

The answer is eigenvector $\vec{\mathbf{v}} = \begin{pmatrix} i \\ 1 \end{pmatrix}$. See Lemma 9.2 page 685 for the expected shortcut, which obtains the answer from the eigenvector for $\lambda_1 = 1 + 2i$. The shortcut creates no matrix $B = A - \lambda I$ and no toolkit sequence B to $\mathbf{rref}(B)$.

The shortcut eliminates the following steps:

$$\begin{aligned} A - \lambda_2 I &= \begin{pmatrix} 1-\lambda_2 & 2 \\ -2 & 1-\lambda_2 \end{pmatrix} \\ &= \begin{pmatrix} 2i & 2 \\ -2 & 2i \end{pmatrix} \\ &\approx \begin{pmatrix} i & 1 \\ 1 & -i \end{pmatrix} && \text{Multiply rule.} \\ &\approx \begin{pmatrix} 0 & 0 \\ 1 & -i \end{pmatrix} && \text{Combination rule, multiplier}{=}{-i}. \\ &\approx \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix} && \text{Swap rule.} \\ &= \mathbf{rref}(A - \lambda_2 I) && \text{Reduced echelon form.} \end{aligned}$$

The partial derivative $\partial_{t_1}\vec{\mathbf{v}}$ of the general solution $x = it_1$, $y = t_1$ is eigenvector $\vec{\mathbf{v}}_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}$.

Example 9.3 (Computing 3×3 Eigenpairs: Real Eigenvalues)

Find all eigenpairs of the 3×3 matrix

$$A = \begin{pmatrix} 1 & 3 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & -5 \end{pmatrix}. \tag{13}$$

Solution:

Reference: Theorem 9.3 page 671.

The answers are

$$\begin{array}{lll} \lambda_1 = 1, & \lambda_2 = 2, & \lambda_3 = -5, \\ \vec{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, & \vec{\mathbf{v}}_2 = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, & \vec{\mathbf{v}}_3 = \begin{pmatrix} 1 \\ -2 \\ -14 \end{pmatrix}. \end{array}$$

College Algebra. The eigenvalues are $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = -5$, because matrix A is triangular and the eigenvalues of a triangular matrix appear on the diagonal.

Linear Algebra. There are three toolkit sequences B to $\mathbf{rref}(B)$ to compute, one for each distinct eigenvalue λ where $B = A - \lambda I$.

Eigenvector for $\lambda_1 = 1$.

Subtract $\lambda_1 = 1$ from the diagonal of A to obtain the equation $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$, where

$$B = A - \lambda_1 I = \begin{pmatrix} 0 & 3 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & -6 \end{pmatrix}.$$

A toolkit sequence with swap, combo, multiply will find

$$\mathbf{rref}(B) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The lead variables are v_2, v_3 and the free variable is v_1. Assign invented symbol t_1 to the free variable and back-substitute into $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ to obtain the scalar equations

$$\begin{aligned} v_1 &= t_1, \\ v_2 &= 0, \\ v_3 &= 0. \end{aligned}$$

Take the partial derivative on invented symbol t_1 across these equations to obtain the eigenvector

$$\vec{\mathbf{v}}_1 = \begin{pmatrix} \frac{\partial v_1}{\partial t_1} \\ \frac{\partial v_2}{\partial t_1} \\ \frac{\partial v_3}{\partial t_1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Eigenvector for $\lambda_2 = 2$.

Subtract $\lambda_2 = 2$ from the diagonal of A to obtain the equation $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$, where

$$B = A - \lambda_2 I = \begin{pmatrix} -1 & 3 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & -7 \end{pmatrix}.$$

A toolkit sequence finds

$$\mathbf{rref}(B) = \begin{pmatrix} 1 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The lead variables are v_1, v_3 and the free variable is v_2. Assign invented symbol t_1 to the free variable and back-substitute into $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ to obtain the scalar equations

$$\begin{aligned} v_1 &= 3t_1, \\ v_2 &= t_1, \\ v_3 &= 0. \end{aligned}$$

Take the partial derivative on invented symbol t_1 across these equations to obtain the eigenvector

$$\vec{\mathbf{v}}_2 = \begin{pmatrix} \frac{\partial v_1}{\partial t_1} \\ \frac{\partial v_2}{\partial t_1} \\ \frac{\partial v_3}{\partial t_1} \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}.$$

The eigenpair is $(\lambda_2, \vec{\mathbf{v}}_2) = \left(2, \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}\right)$

Eigenvector for $\lambda_3 = -5$.

Subtract $\lambda_3 = -5$ from the diagonal of A to obtain the equation $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$, where

$$B = A - \lambda_2 I = \begin{pmatrix} 6 & 3 & 0 \\ 0 & 7 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

A toolkit sequence finds

$$\mathbf{rref}(B) = \begin{pmatrix} 1 & 0 & 1/14 \\ 0 & 1 & -1/7 \\ 0 & 0 & 0 \end{pmatrix}.$$

The lead variables are v_1, v_2 and the free variable is v_3. Assign invented symbol t_1 to the free variable and back-substitute into $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ to obtain the scalar equations

$$\begin{aligned} v_1 &= -\tfrac{1}{14}t_1, \\ v_2 &= \tfrac{1}{7}t_1, \\ v_3 &= 0. \end{aligned}$$

Take the partial derivative on invented symbol t_1 across these equations to obtain the eigenvector

$$\vec{\mathbf{v}}_3 = \begin{pmatrix} \frac{\partial v_1}{\partial t_1} \\ \frac{\partial v_2}{\partial t_1} \\ \frac{\partial v_3}{\partial t_1} \end{pmatrix} = \begin{pmatrix} -\frac{1}{14} \\ \frac{1}{7} \\ 1 \end{pmatrix}.$$

It is usual when encountering fractions in an eigenvector to replace the answer $\vec{\mathbf{v}}$ by $c\vec{\mathbf{v}}$ where $c \neq 0$ is chosen to make the answer fraction-free and the first nonzero entry positive. In this case, $c = -14$ is used, and we replace $\vec{\mathbf{v}}_3$ by $-14\vec{\mathbf{v}}_3$. The eigenpair is $(\lambda_3, \vec{\mathbf{v}}_3) = \left(-5, \begin{pmatrix} 1 \\ -2 \\ -14 \end{pmatrix}\right)$.

This completes the computation of all three eigenpairs.

Answer Check. The eigenpair equations are equivalent to the matrix identity $AP = PD$ where P is the matrix of eigenvectors and D is the diagonal matrix of corresponding eigenvalues:

$$P = \begin{pmatrix} 1 & 3 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & -14 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -5 \end{pmatrix}.$$

Eigenpairs are checked by expanding AP and PD, then compare for equality. The two calculations give

$$AP = \begin{pmatrix} 1 & 6 & -5 \\ 0 & 2 & 10 \\ 0 & 0 & 70 \end{pmatrix} = PD.$$

Fourier Replacement page 676 is explicitly

$$\begin{aligned} \vec{\mathbf{x}} &= c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ -2 \\ -14 \end{pmatrix} \\ \text{implies} \\ A\vec{\mathbf{x}} &= c_1(1) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2(2) \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + c_3(-5) \begin{pmatrix} 1 \\ -2 \\ -14 \end{pmatrix} \end{aligned}$$

Example 9.4 (Computing 3×3 Eigenpairs: Complex Eigenvalues)

Find all eigenpairs of the 3×3 matrix $A = \begin{pmatrix} 1 & 2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.

Solution:

Reference: Theorem 9.3 page 671.

College Algebra. The eigenvalues are $\lambda_1 = 1 + 2i$, $\lambda_2 = 1 - 2i$, $\lambda_3 = 3$. Details:

$$\begin{aligned} 0 &= \det(A - \lambda I) && \text{Characteristic equation.} \\ &= \begin{vmatrix} 1-\lambda & 2 & 0 \\ -2 & 1-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{vmatrix} && \text{Subtract } \lambda \text{ from the diagonal.} \\ &= ((1-\lambda)^2 + 4)(3-\lambda) && \text{Cofactor rule and Sarrus' rule.} \end{aligned}$$

Root $\lambda = 3$ is found from the factored form above. The roots $\lambda = 1 \pm 2i$ are found from the quadratic formula after expanding $(1-\lambda)^2 + 4 = 0$. Alternatively, take roots across $(\lambda - 1)^2 = -4$.

Linear Algebra.

The eigenpairs are $\left(1+2i, \begin{pmatrix} -i \\ 1 \\ 0 \end{pmatrix}\right)$, $\left(1-2i, \begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix}\right)$, $\left(3, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right)$.

Eigenvector for $\lambda_1 = 1+2i$.

$$
\begin{aligned}
A - \lambda_1 I &= \begin{pmatrix} 1-\lambda_1 & 2 & 0 \\ -2 & 1-\lambda_1 & 0 \\ 0 & 0 & 3-\lambda_1 \end{pmatrix} & \\
&= \begin{pmatrix} -2i & 2 & 0 \\ -2 & -2i & 0 \\ 0 & 0 & 2-2i \end{pmatrix} & \text{Subtract } \lambda_1 = 1+2i \text{ from the diagonal.} \\
&\approx \begin{pmatrix} i & -1 & 0 \\ 1 & i & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{Multiply rule.} \\
&\approx \begin{pmatrix} 0 & 0 & 0 \\ 1 & i & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{Combination rule, factor}=-i. \\
&\approx \begin{pmatrix} 1 & i & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} & \text{Swap rule.} \\
&= \mathbf{rref}(A - \lambda_1 I) & \text{Reduced echelon form.}
\end{aligned}
$$

The vector partial derivative $\partial_{t_1}\vec{\mathbf{v}}$ of the scalar general solution $x = -it_1$, $y = t_1$, $z = 0$ is eigenvector $\vec{\mathbf{v}}_1 = \begin{pmatrix} -i \\ 1 \\ 0 \end{pmatrix}$.

Eigenvector for $\lambda_2 = 1-2i$.

There is no need for a toolkit sequence to find the eigenvector for a conjugate eigenvalue: see Lemma 9.2 *infra*. Answer: $(1-2i, \vec{\mathbf{v}}_2)$, $\vec{\mathbf{v}}_2 = \begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix}$.

Details. To see why, take conjugates[5] across the equation $(A - \lambda_2 I)\vec{\mathbf{v}}_2 = \vec{\mathbf{0}}$ to give $(\overline{A} - \overline{\lambda_2} I)\overline{\vec{\mathbf{v}}_2} = \vec{\mathbf{0}}$. Then $\overline{A} = A$ (A is real) and $\lambda_1 = \overline{\lambda_2}$ gives $(A - \lambda_1 I)\overline{\vec{\mathbf{v}}_2} = \vec{\mathbf{0}}$. Then $\overline{\vec{\mathbf{v}}_2} = \vec{\mathbf{v}}_1$. Finally, $\vec{\mathbf{v}}_2 = \overline{\overline{\vec{\mathbf{v}}_2}} = \overline{\vec{\mathbf{v}}_1} = \begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix}$. These details prove:

Lemma 9.2 If $(a+ib, \vec{\mathbf{v}})$ is an eigenpair of A, then formally replacing i by $-i$ in this eigenpair finds a second eigenpair for the conjugate eigenvalue.

Eigenvector for $\lambda_3 = 3$.

$$
A - \lambda_3 I = \begin{pmatrix} 1-\lambda_3 & 2 & 0 \\ -2 & 1-\lambda_3 & 0 \\ 0 & 0 & 3-\lambda_3 \end{pmatrix}
$$

[5]The complex conjugate is defined by $\overline{a+ib} = a - ib$ (replace i by $-i$). Two useful rules are $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ and $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$. Conjugation rules extend to vectors and matrices by applying scalar rules componentwise, e.g., $\overline{\vec{\mathbf{u}} + \vec{\mathbf{v}}} = \overline{\vec{\mathbf{u}}} + \overline{\vec{\mathbf{v}}}$ and $\overline{A\vec{\mathbf{x}}} = \overline{A}\,\overline{\vec{\mathbf{x}}}$.

$$
\begin{aligned}
&= \begin{pmatrix} -2 & 2 & 0 \\ -2 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \\
&\approx \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \text{Multiply rule.} \\
&\approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \text{Combination and multiply.} \\
&= \mathbf{rref}(A - \lambda_3 I) & \text{Reduced echelon form.}
\end{aligned}
$$

The partial derivative $\partial_{t_1}\vec{\mathbf{v}}$ of the general solution $x = 0$, $y = 0$, $z = t_1$ is eigenvector $\vec{\mathbf{v}}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

Example 9.5 (Data Conversion)
The data conversion problem

$$
\begin{cases} y_1 &= x_1, \\ y_2 &= 0.001 x_2, \\ y_3 &= 0.01 x_3. \end{cases}
$$

is diagonalizable. The three eigenpairs of A are defined by

$$
\begin{array}{lll}
\lambda_1 = 1, & \lambda_2 = 0.001, & \lambda_3 = 0.01, \\
\vec{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, & \vec{\mathbf{v}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, & \vec{\mathbf{v}}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.
\end{array}
$$

Solution: References: Theorem 9.3 page 671 and Theorem 9.7 page 674.

The example was introduced in equation (10) page 677. The equations can be written as $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$, where $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.001 & 0 \\ 0 & 0 & 0.01 \end{pmatrix}$ is already a diagonal matrix, eigenpairs given by Theorem 9.7 page 674.

Answers can be verified directly from the eigenpair equation $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$ without using theorems. For instance, when $\vec{\mathbf{v}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\lambda = 0.001$, then the two sides $A\vec{\mathbf{v}}$ and $\lambda\vec{\mathbf{v}}$ are computed from matrix multiply, each giving the same answer $\begin{pmatrix} 0 \\ 0.001 \\ 0 \end{pmatrix}$, therefore $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$ is valid and $(\lambda, \vec{\mathbf{v}})$ is an eigenpair of A.

Example 9.6 (Decomposition $A = PDP^{-1}$)

Decompose $A = PDP^{-1}$ where P, D are eigenvector and eigenvalue packages, respectively, for the 3×3 matrix

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

Illustrate Fourier replacement for this matrix.

Solution: By the preceding example, the eigenpairs are

$$\left(1+2i, \begin{pmatrix} -i \\ 1 \\ 0 \end{pmatrix}\right), \quad \left(1-2i, \begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix}\right), \quad \left(3, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right).$$

The packages are therefore

$$D = \textbf{diag}(1+2i, 1-2i, 3), \quad P = \begin{pmatrix} -i & i & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Fourier replacement. The model:

$$A\left(c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3\right) = c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\vec{\mathbf{v}}_3$$

It means $A\vec{\mathbf{x}}$ changes $\vec{\mathbf{x}}$ by replacing the basis $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ by scaled vectors $\lambda_1\vec{\mathbf{v}}_1$, $\lambda_2\vec{\mathbf{v}}_2$, $\lambda_3\vec{\mathbf{v}}_3$. Explicitly,

$$\begin{aligned} \vec{\mathbf{x}} &= c_1\begin{pmatrix} -i \\ 1 \\ 0 \end{pmatrix} + c_2\begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix} + c_3\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \text{ implies} \\ A\vec{\mathbf{x}} &= c_1(1+2i)\begin{pmatrix} -i \\ 1 \\ 0 \end{pmatrix} + c_2(1-2i)\begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix} + c_3(3)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Example 9.7 (Diagonalization I)

Report **diagonalizable** or **non-diagonalizable** for the 4×4 matrix

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

If A is diagonalizable, then report eigenvector and eigenvalue packages P, D.

Solution: Reference: page 674 for definitions and theorems.

The matrix A is **non-diagonalizable**, because it fails to have 4 eigenpairs. The details:

Eigenvalues.

$0 = \det(A - \lambda I)$ Characteristic equation.

$$= \begin{vmatrix} 1-\lambda & 2 & 0 & 0 \\ -2 & 1-\lambda & 0 & 0 \\ 0 & 0 & 3-\lambda & 1 \\ 0 & 0 & 0 & 3-\lambda \end{vmatrix}$$

$$= \begin{vmatrix} 1-\lambda & 2 \\ -2 & 1-\lambda \end{vmatrix} (3-\lambda)^2$$ Cofactor expansion applied twice.

$$= \left((1-\lambda)^2+4\right)(3-\lambda)^2$$ Sarrus' rule.

The roots are $1 \pm 2i$, 3, 3, listed according to multiplicity.

Eigenpairs. They are

$$\left(1+2i, \begin{pmatrix} -i \\ 1 \\ 0 \\ 0 \end{pmatrix}\right), \quad \left(1-2i, \begin{pmatrix} i \\ 1 \\ 0 \\ 0 \end{pmatrix}\right), \quad \left(3, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}\right).$$

Matrix A is **non-diagonalizable**, because only three eigenpairs exist, instead of four. Details:

Eigenvector for $\lambda_1 = 1+2i$.

$$A - \lambda_1 I = \begin{pmatrix} 1-\lambda_1 & 2 & 0 & 0 \\ -2 & 1-\lambda_1 & 0 & 0 \\ 0 & 0 & 3-\lambda_1 & 1 \\ 0 & 0 & 0 & 3-\lambda_1 \end{pmatrix}$$

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

$$\approx \begin{pmatrix} -i & 1 & 0 & 0 \\ -1 & -i & 0 & 0 \\ 0 & 0 & 2-2i & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$ Multiply rule, three times.

$$\approx \begin{pmatrix} -i & 1 & 0 & 0 \\ -1 & -i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$ Combination and multiply rule.

$$\approx \begin{pmatrix} 1 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$ Combination and multiply rule.

$$= \mathbf{rref}(A - \lambda_1 I)$$ Reduced echelon form.

The general solution is $x_1 = -it_1$, $x_2 = t_1$, $x_3 = 0$, $x_4 = 0$. Then ∂_{t_1} applied to this solution gives the reported eigenpair for $\lambda = 1+2i$.

Eigenvector for $\lambda_2 = 1-2i$.
Because λ_2 is the conjugate of λ_1 and A is real, then an eigenpair for λ_2 is found from the eigenpair for λ_1 by replacing i by $-i$ throughout. See Lemma 9.2 page 685.

Eigenvector for $\lambda_3 = 3$. In theory, there can be one or two eigenpairs to report. It turns out there is only one, because of the following details. This single toolkit sequence establishes that A is **non-diagonalizable**. The other toolkit sequences could have been skipped, if only diagonalizability was the issue and we were clever enough to examine this case first.

$$\begin{aligned} A - \lambda_3 I &= \begin{pmatrix} 1-\lambda_3 & 2 & 0 & 0 \\ -2 & 1-\lambda_3 & 0 & 0 \\ 0 & 0 & 3-\lambda_3 & 1 \\ 0 & 0 & 0 & 3-\lambda_3 \end{pmatrix} \\ &= \begin{pmatrix} -2 & 2 & 0 & 0 \\ -2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &\approx \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} && \text{Multiply rule, two times.} \\ &\approx \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} && \text{Combination and multiply rule.} \\ &= \mathbf{rref}(A - \lambda_3 I) && \text{Reduced echelon form.} \end{aligned}$$

Apply ∂_{t_1} to the general solution $x_1 = 0$, $x_2 = 0$, $x_3 = t_1$, $x_4 = 0$ to give the eigenvector matching the eigenpair reported above for $\lambda = 3$.

Example 9.8 (Diagonalization II)

Report **diagonalizable** or **non-diagonalizable** for the 4×4 matrix

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

If A is diagonalizable, then assemble and report eigenvalue and eigenvector packages D, P.

Solution: Reference: page 674 for definitions and theorems.

The matrix A is **diagonalizable**, because it has 4 eigenpairs

$$\left(1+2i, \begin{pmatrix} -i \\ 1 \\ 0 \\ 0 \end{pmatrix}\right), \quad \left(1-2i, \begin{pmatrix} i \\ 1 \\ 0 \\ 0 \end{pmatrix}\right), \quad \left(3, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}\right), \quad \left(3, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right).$$

Then the eigenpair packages are given by

$$D = \begin{pmatrix} -1+2i & 0 & 0 & 0 \\ 0 & 1-2i & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad P = \begin{pmatrix} -i & i & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The details parallel the previous example, except for the calculation of eigenvectors for $\lambda_3 = 3$. In this case, the reduced echelon form of $A - \lambda_3 I$ has two rows of zeros and parameters t_1, t_2 appear in the general solution. The answers given above for eigenvectors correspond to the partial derivatives ∂_{t_1}, ∂_{t_2} applied to the general solution of $(A - 3I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$.

Example 9.9 (Non-diagonalizable Matrices)

Verify that the matrices

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

are all non-diagonalizable.

Solution: References: page 674 for definitions and theorems; Theorem 9.1 page 670 for $AP = PD$ and eigenpair equations.

Let A denote any one of these matrices and let n be its dimension.

Without computing eigenpairs, diagonalization will be decided. Assume, in order to reach a contradiction, that eigenpair packages D, P exist with D diagonal and P invertible such that $AP = PD$. Because A is triangular, its eigenvalues appear already on the diagonal of A. Only 0 is an eigenvalue and its multiplicity is n. Then the package D of eigenvalues is the zero matrix and an equation $AP = PD$ reduces to $AP = 0$. Multiply $AP = 0$ on the right by P^{-1} to obtain $A = 0$. But A is not the zero matrix, a contradiction. Conclusion: A is not diagonalizable.

Secondly, attack the diagonalization question directly, by solving for the eigenvectors corresponding to $\lambda = 0$. The toolkit sequence starts with $B = A - \lambda I$, but B equals $\mathbf{rref}(B)$ and no computations are required. The resulting reduced echelon system is just $x_1 = 0$, giving $n - 1$ free variables. Therefore, the eigenvectors of A corresponding to $\lambda = 0$ are the last $n - 1$ columns of the identity matrix I. Because A does not have n independent eigenvectors, then A is not diagonalizable.

Similar examples of non-diagonalizable matrices A can be constructed with A having from 1 up to $n - 1$ independent eigenvectors. The examples with ones on the super-diagonal and zeros elsewhere have exactly one eigenvector.

Example 9.10 (Fourier's 1822 Heat Model)

Fourier's 1822 treatise *Théorie analytique de la chaleur* studied dissipation of heat from a laterally insulated welding rod with ends held at 0°C (ice-packed ends). Assume the initial heat distribution along the rod at time $t = 0$ is given as a linear combination

$$f = c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3.$$

Symbols $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ are in the vector space V of all twice continuously differentiable functions on $0 \le x \le 1$, given explicitly as

$$\vec{\mathbf{v}}_1 = \sin \pi x, \quad \vec{\mathbf{v}}_2 = \sin 2\pi x, \quad \vec{\mathbf{v}}_3 = \sin 3\pi x.$$

Fourier's heat model re-scales[6] each of these vectors to obtain the temperature $u(t,x)$ at position x along the rod and time $t > 0$ as the model equation

$$u(t,x) = c_1 e^{-\pi^2 t}\vec{\mathbf{v}}_1 + c_2 e^{-4\pi^2 t}\vec{\mathbf{v}}_2 + c_3 e^{-9\pi^2 t}\vec{\mathbf{v}}_3.$$

Verify that $u(t,x)$ solves Fourier's partial differential equation heat model

$$\begin{array}{rcl} \dfrac{\partial u}{\partial t} &=& \dfrac{\partial^2 u}{\partial x^2}, \\ u(0,x) &=& f(x), \quad 0 \le x \le 1, \quad \text{initial temperature}, \\ u(t,0) &=& 0, \quad \text{zero Celsius at rod's left end}, \\ u(t,1) &=& 0, \quad \text{zero Celsius at rod's right end}. \end{array}$$

Solution: First, let's prove that the partial differential equation is satisfied by Fourier's solution $u(t,x)$. This is done by expanding the left side (LHS) and right side (RHS) of the differential equation separately, then comparing the two answers for equality.

Trigonometric functions $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ are solutions of three different linear ordinary differential equations: $u'' + \pi^2 u = 0$, $u'' + 4\pi^2 u = 0$, $u'' + 9\pi^2 u = 0$. Because of these differential equations, calculus derivatives can be computed:

$$\frac{\partial^2 u}{\partial x^2} = -\pi^2 c_1 e^{-\pi^2 t}\vec{\mathbf{v}}_1 - 4\pi^2 c_2 e^{-4\pi^2 t}\vec{\mathbf{v}}_2 - 9\pi^2 c_3 e^{-9\pi^2 t}\vec{\mathbf{v}}_3.$$

Similarly, computing $\partial_t u(t,x)$ involves just the differentiation of exponential functions, giving

$$\frac{\partial u}{\partial t} = -\pi^2 c_1 e^{-\pi^2 t}\vec{\mathbf{v}}_1 - 4\pi^2 c_2 e^{-4\pi^2 t}\vec{\mathbf{v}}_2 - 9\pi^2 c_3 e^{-9\pi^2 t}\vec{\mathbf{v}}_3.$$

Because the second display is exactly the first, then LHS = RHS, proving that the partial differential equation is satisfied.

The relation $u(0,x) = f(x)$ holds because each exponential factor becomes $e^0 = 1$ when $t = 0$.

The two relations $u(t,0) = u(t,1) = 0$ hold because each of $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ vanish at $x = 0$ and $x = 1$. The verification is complete.

Example 9.11 (Powers and Fourier Replacement)

Let 3×3 matrix A have eigenpairs $(\lambda_1, \vec{\mathbf{v}}_i)$, $i = 1,2,3$ and (9) holds. Find the powers $A^k\vec{\mathbf{x}}$ by Fourier's Replacement equation (8) with just the basic vector space toolkit, showing

$$A^k\vec{\mathbf{x}} = x_1\lambda_1^k\vec{\mathbf{v}}_1 + x_2\lambda_2^k\vec{\mathbf{v}}_2 + x_3\lambda_3^k\vec{\mathbf{v}}_3$$

Solution: The vector toolkit for $\mathcal{R}^3$ is used to compute powers:

$$\begin{array}{rcll} A\vec{\mathbf{x}} &=& x_1\lambda_1\vec{\mathbf{v}}_1 + x_2\lambda_2\vec{\mathbf{v}}_2 + x_3\lambda_3\vec{\mathbf{v}}_3 & \\ A^2\vec{\mathbf{x}} &=& A(x_1\lambda_1\vec{\mathbf{v}}_1 + x_2\lambda_2\vec{\mathbf{v}}_2 + x_3\lambda_3\vec{\mathbf{v}}_3) & \\ &=& x_1\lambda_1^2\vec{\mathbf{v}}_1 + x_2\lambda_2^2\vec{\mathbf{v}}_2 + x_3\lambda_3^2\vec{\mathbf{v}}_3 & \text{by (8)} \\ &\vdots& & \\ A^k\vec{\mathbf{x}} &=& x_1\lambda_1^k\vec{\mathbf{v}}_1 + x_2\lambda_2^k\vec{\mathbf{v}}_2 + x_3\lambda_3^k\vec{\mathbf{v}}_3 & \end{array}$$

[6] The scale factors are not constants nor are they eigenvalues, but rather, they are exponential functions of t for fixed t, as is the case for matrix differential equations $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$. See Example 9.13

The calculations do not use matrix multiply and the answer does not depend upon finding previous powers A^2, A^3, A^4, ...

Fourier replacement reduces computational effort. Matrix–vector multiplication to produce $\vec{\mathbf{y}}_k = A^k\vec{\mathbf{x}}$ requires $9k$ multiply operations whereas Fourier replacement gives the answer with $3k + 9$ multiply operations.

Example 9.12 (Change of Variable $\vec{\mathbf{x}} = P\vec{\mathbf{u}}$ for Differential Equations)

Matrix $A = \begin{pmatrix} 1 & 3 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & -5 \end{pmatrix}$ has eigenpairs $(\vec{\mathbf{v}}_1, \lambda_1), (\vec{\mathbf{v}}_2, \lambda_2), (\lambda_3, \vec{\mathbf{v}}_3)$ with three independent eigenvectors given by equation (12). Define $\vec{\mathbf{x}} = P\vec{\mathbf{u}}$, $P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle$, $D = \mathbf{diag}(\lambda_1, \lambda_2, \lambda_3)$. Show that $\vec{\mathbf{x}} = P\vec{\mathbf{u}}$ changes $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ into $\vec{\mathbf{u}}' = D\vec{\mathbf{u}}$, which is the diagonal system of growth-decay equations

$$\begin{cases} u_1' &= u_1, \\ u_2' &= 2u_2, \\ u_3' &= -5u_3. \end{cases}$$

Solution: The calculus derivative of a vector function is performed componentwise. Matrix multiply as a linear combination of columns shows that equation $\vec{\mathbf{x}}(t) = P\vec{\mathbf{u}}(t)$ has derivative $\vec{\mathbf{x}}'(t) = P\vec{\mathbf{u}}'(t)$, because entries of P are constants. Then equation $\vec{\mathbf{x}}(t) = P\vec{\mathbf{u}}(t)$ can change $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ into a differential equation in variable $\vec{\mathbf{u}}$. The details:

$\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t)$	Given.
$P\vec{\mathbf{u}}'(t) = AP\vec{\mathbf{u}}(t)$	Use $\vec{\mathbf{x}}'(t) = P\vec{\mathbf{u}}'(t)$, $\vec{\mathbf{x}}(t) = P\vec{\mathbf{u}}(t)$.
$P\vec{\mathbf{u}}'(t) = PD\vec{\mathbf{u}}(t)$	because $AP = PD$ (A is diagonalizable).
$\vec{\mathbf{u}}'(t) = D\vec{\mathbf{u}}(t)$	because P has an inverse.

The eigenvalues of triangular matrix A are the diagonal entries: 1,2,-5. Then $D = \mathbf{diag}(1, 2, -5)$ and $\vec{\mathbf{u}}' = D\vec{\mathbf{u}}$ is the reported system of growth-decay differential equations.

Example 9.13 (Differential Equations and Fourier Replacement)

Solve by Fourier re-scaling $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ with $A = \begin{pmatrix} 1 & 3 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & -5 \end{pmatrix}$. The scalar form:

$$\begin{cases} x_1' &= x_1 + 3x_2, \\ x_2' &= \phantom{x_1 + {}} 2x_2 - x_3, \\ x_3' &= - 5x_3. \end{cases}$$

The answer uses the eigenpairs $(\vec{\mathbf{v}}_1, \lambda_1), (\vec{\mathbf{v}}_2, \lambda_2), (\lambda_3, \vec{\mathbf{v}}_3)$ of matrix A in equation (12):

$$(14) \quad \begin{cases} \vec{\mathbf{x}}(t) &= c_1 e^{\lambda_1 t}\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}\vec{\mathbf{v}}_2 + c_3 e^{\lambda_3 t}\vec{\mathbf{v}}_3, \quad \text{realized as} \\ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= c_1 e^{t} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2 e^{2t} \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + c_3 e^{-5t} \begin{pmatrix} 1 \\ -2 \\ -14 \end{pmatrix}. \end{cases}$$

Solution: Fourier's re-scaling idea applies to linear differential equations, as follows. First, expand the initial condition $\vec{\mathbf{x}}(0)$ in terms of basis elements:

$$\vec{\mathbf{x}}(0) = c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3.$$

Fourier's re-scaling replaces each $\vec{\mathbf{v}}_i$ by the re-scaled vector $e^{\lambda_i t}\vec{\mathbf{v}}_i$. The result:

$$\vec{\mathbf{y}} = c_1 e^{\lambda_1 t}\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}\vec{\mathbf{v}}_2 + c_3 e^{\lambda_3 t}\vec{\mathbf{v}}_3 \tag{15}$$

How is this related to Fourier re-scaling? Answer: at each fixed instant t, the basis vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ are replaced by $\Lambda_1\vec{\mathbf{v}}_1$, $\Lambda_2\vec{\mathbf{v}}_2$, $\Lambda_3\vec{\mathbf{v}}_3$ where

$$\Lambda_1 = e^{\lambda_1 t}, \quad \Lambda_2 = e^{\lambda_2 t}, \quad \Lambda_3 = e^{\lambda_3 t}.$$

Why is the solution $\vec{\mathbf{x}}(t) = c_1 e^{\lambda_1 t}\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}\vec{\mathbf{v}}_2 + c_3 e^{\lambda_3 t}\vec{\mathbf{v}}_3$? Answer: Evaluate the LHS and RHS of the differential equation $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ and compare formulas.

$$\begin{aligned}
\text{LHS} &= \vec{\mathbf{x}}'(t) \\
&= c_1\lambda_1 e^{\lambda_1 t}\vec{\mathbf{v}}_1 + c_2\lambda_2 e^{\lambda_2 t}\vec{\mathbf{v}}_2 + c_3\lambda_3 e^{\lambda_3 t}\vec{\mathbf{v}}_3 \\
&= c_1\lambda_1\Lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\Lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\Lambda_3\vec{\mathbf{v}}_3 \\
\text{RHS} &= A\vec{\mathbf{x}}(t) \\
&= A(c_1\Lambda_1\vec{\mathbf{v}}_1 + c_2\Lambda_2\vec{\mathbf{v}}_2 + c_3\Lambda_3\vec{\mathbf{v}}_3) \\
&= c_1\lambda_1\Lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\Lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\Lambda_3\vec{\mathbf{v}}_3 \quad \text{by Theorem 9.10.}
\end{aligned}$$

The last equality is tricky: equation

$$A\left(c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3\right) = c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\vec{\mathbf{v}}_3$$

in Theorem 9.10 is applied with c_1, c_2, c_3 replaced by $c_1\Lambda_1, c_2\Lambda_2, c_3\Lambda_3$.

Justification of the solution is done with Example 9.12 after inserting exponential solutions for the growth-decay equations. A summary of the re-scaling method:

1. Expand $\vec{\mathbf{x}}(0)$ as a linear combination of eigenvectors.
2. Change on the left $\vec{\mathbf{x}}(0)$ to $\vec{\mathbf{x}}(t)$, then re-scale the linear combination on the right with scale factors $\Lambda_1 = e^{\lambda_1 t}$, $\Lambda_2 = e^{\lambda_2 t}$, $\Lambda_3 = e^{\lambda_3 t}$.

Proofs and Details

Proof of Theorem 9.1, Eigenpairs and $AP = PD$:
Let

$$P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle, \quad D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

Write the two matrix multiply equations AP and PD in expanded form

$$AP = \left\langle A\vec{\mathbf{v}}_1 | A\vec{\mathbf{v}}_2 | A\vec{\mathbf{v}}_3 \right\rangle, \quad PD = \left\langle \lambda_1\vec{\mathbf{v}}_1 | \lambda_2\vec{\mathbf{v}}_2 | \lambda_3\vec{\mathbf{v}}_3 \right\rangle. \tag{16}$$

$AP = PD$ **implies equation (4)**. Assume $AP = PD$. Because equal matrices have equal columns, the columns left and right in the equation $AP = PD$ must match, using expansion (16). Then

$$A\vec{\mathbf{v}}_1 = \lambda_1\vec{\mathbf{v}}_1, \quad A\vec{\mathbf{v}}_2 = \lambda_2\vec{\mathbf{v}}_2, \quad A\vec{\mathbf{v}}_3 = \lambda_3\vec{\mathbf{v}}_3,$$

which means equation (4) holds.

Equation (4) implies $AP = PD$. Assume eigenpair equations (4) hold. Then the two matrices AP and PD in expansion (16) have equal columns. Equality of matrices implies $AP = PD$. ■

Proof of Theorem 9.2, Eigenvalues of A:
An eigenvalue λ is a number such that equation $A\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$ has a nonzero solution $\vec{\mathbf{x}}$. Let $B = A - \lambda I$. Then λ is an eigenvalue means $B\vec{\mathbf{x}} = \vec{\mathbf{0}}$ has a nonzero solution $\vec{\mathbf{x}}$. Homogeneous equation $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ has a nonzero solution $\vec{\mathbf{v}}$ if and only if there are infinitely many solutions. Because the matrix is square, infinitely many solutions occur if and only if $\mathbf{rref}(B)$ has a row of zeros. Determinant theory gives a more concise statement: $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ has infinitely many solutions if and only if $\det(B) = 0$. ■

Proof of Theorem 9.3, Find Eigenvectors:
Question: Why does the solution of $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ have invented symbols? Isn't there just one solution?

Answer: According to the *three possibilities*, homogeneous equation $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$ should have unique solution $\vec{\mathbf{v}} = \vec{\mathbf{0}}$ or else infinitely many solutions. An eigenvector cannot be zero. To get infinitely many solutions there has to be at least one free variable, causing the *last frame algorithm* to be applied with invented symbols $t_1, t_2, \ldots$.

The equation $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$ is equivalent to $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$. Because λ is a root of characteristic equation $|A-\lambda I| = 0$, then $\det(B) = 0$ and B has no inverse, equivalent to $\mathbf{rref}(B) \ne I$. Then square matrix $\mathbf{rref}(B)$ must have a row of zeros, which means there is at least one free variable. The last frame algorithm applies with invented symbols t_1, t_2, ... A vector basis $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \ldots$ for the nullspace of B is obtained from the list of vector partial derivatives on symbols t_1, t_2, ... These vectors are Strang's special solutions, which are known to be collectively independent. The nullspace of B is the span of Strang's special solutions $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \ldots$. If $A\vec{\mathbf{w}} = \lambda\vec{\mathbf{w}}$, then $B\vec{\mathbf{w}} = \vec{\mathbf{0}}$, so $\vec{\mathbf{w}}$ belongs to the nullspace of B:

$$\vec{\mathbf{w}} = \text{a linear combination of} \quad \vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \ldots \quad ■$$

Proof of Theorem 9.4, Independence of Eigenvectors:
Let's solve $c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 = \vec{\mathbf{0}}$ for c_1, c_2. The vectors are independent provided the only solution is $c_1 = c_2 = 0$. Apply A to this equation, obtaining $c_1A\vec{\mathbf{v}}_1 + c_2A\vec{\mathbf{v}}_2 = \vec{\mathbf{0}}$. Use $A\vec{\mathbf{v}}_1 = \lambda_1\vec{\mathbf{v}}_1$ and $A\vec{\mathbf{v}}_2 = \lambda_2\vec{\mathbf{v}}_2$ to obtain $c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 = \vec{\mathbf{0}}$. Multiply $c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 = \vec{\mathbf{0}}$ by λ_1 and subtract it from $c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 = \vec{\mathbf{0}}$ to get $c_1(\lambda_1-\lambda_1)\vec{\mathbf{v}}_1 + c_2(\lambda_2-\lambda_1)\vec{\mathbf{v}}_2 = \vec{\mathbf{0}}$. Because $\lambda_2 \ne \lambda_1$, cancel $\lambda_2 - \lambda_1$ to give $c_2\vec{\mathbf{v}}_2 = \vec{\mathbf{0}}$. The assumption $\vec{\mathbf{v}}_2 \ne \vec{\mathbf{0}}$ implies $c_2 = 0$. Return to the first equation and use $c_2 = 0$ to obtain $c_1\vec{\mathbf{v}}_1 = \vec{\mathbf{0}}$. Because $\vec{\mathbf{v}}_1 \ne \vec{\mathbf{0}}$, then $c_1 = 0$. This proves $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ are independent.

The general case is proved by **Mathematical Induction** on k (see the footnote in the proof of Theorem 9.5). The case $k = 1$ follows because a nonzero vector is an independent set. Assume it holds for $k-1$ and let's prove it for k, when $k > 1$. To prove independence, we must solve for $c_1, \ldots, c_k$ in the test equation

$$c_1\vec{\mathbf{v}}_1 + \cdots + c_k\vec{\mathbf{v}}_k = \vec{\mathbf{0}}.$$

Create a second equation by multiplication of the test equation by A, effectively replacing each c_i by $\lambda_i c_i$, due to the eigenpair equation $A\vec{\mathbf{v}}_i = \lambda_i\vec{\mathbf{v}}_i$. Then multiply the test equation by λ_1 and subtract the two equations to get the new equation

$$c_1(\lambda_1 - \lambda_1)\vec{\mathbf{v}}_1 + c_2(\lambda_1 - \lambda_2)\vec{\mathbf{v}}_2 + \cdots + c_k(\lambda_1 - \lambda_k)\vec{\mathbf{v}}_k = \vec{\mathbf{0}}.$$

The first term is zero. Apply the induction hypothesis to the remaining $k-1$ vectors, then independence implies all coefficients $(\lambda_1 - \lambda_i)c_i$ are zero. Because $\lambda_1 - \lambda_i \neq 0$ for $i > 1$, then c_2 through c_k are zero. Substitute the zero values into the test equation to obtain $c_1\vec{\mathbf{v}}_1 = \vec{\mathbf{0}}$. Because $\vec{\mathbf{v}}_1 \neq \vec{\mathbf{0}}$, then $c_1 = 0$. Therefore all $c_i = 0$. The induction is complete. ■

Proof of Theorem 9.5, Unions of Eigenvectors:

Details (1). Assume there is a nonzero vector $\vec{\mathbf{v}}$ in the intersection, which must be an eigenvector for both λ and μ. Then two eigenpairs$(\lambda, \vec{\mathbf{v}}_1)$ and $(\mu, \vec{\mathbf{v}}_2)$ have been found, $\vec{\mathbf{v}}_1 = \vec{\mathbf{v}}_2 = \vec{\mathbf{v}}$, which violates Theorem 9.4, because $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ must be independent.

Details (2). Let's proceed by induction on the number k of eigenvalues used to construct U.[7] Let S_k be the statement that U = union of $\mathcal{B}(\lambda_1)$, ..., $\mathcal{B}(\lambda_k)$ has independent elements, no matter how the k distinct eigenvalues $\{\lambda_i\}_{i=1}^k$ are selected and no matter how the bases are chosen.

Statement S_1 is true, because $\mathcal{B}(\lambda_1)$ is a list of independent elements.

Assume S_k is true. The proof that S_{k+1} is true will be deferred to the exercises. Revealed here are the fundamental ideas, by examining the cases $k = 2$ and $k = 3$.

Case $k = 2$. Then U is a list of vectors, some from $\mathcal{B}(\lambda_1)$ and some from $\mathcal{B}(\lambda_2)$. The test equation for independence of this list of vectors is a linear combination of the vectors equal to the zero vector. The objective is to prove that the coefficients in this linear combination are all zero. Rearrange the test equation in the form

$$\text{Terms using vectors from } \mathcal{B}(\lambda_1) = \text{Terms using vectors from } \mathcal{B}(\lambda_2)$$

The left side of the above equation is an eigenvector $\vec{\mathbf{v}}_1$ for eigenvalue λ_1, giving eigenpair $(\lambda_1, \vec{\mathbf{v}}_1)$. Similarly, the right side determines an eigenpair $(\lambda_2, \vec{\mathbf{v}}_2)$. The previous theorem says that $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ are independent, if nonzero. Analyzing cases, then both $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ are the zero vector. By independence of bases $\mathcal{B}(\lambda_1)$ and $\mathcal{B}(\lambda_2)$, all coefficients are zero, proving independence of the list U.

Case $k = 3$. Let U_2 be the union of bases $\mathcal{B}(\lambda_1), \mathcal{B}(\lambda_2)$, which is a list of vectors $\vec{\mathbf{v}}_1$, ..., $\vec{\mathbf{v}}_q$. Given is U = the union of bases $\mathcal{B}(\lambda_1)$, $\mathcal{B}(\lambda_2)$, $\mathcal{B}(\lambda_3)$. The test equation for independence of the vectors in list U is a linear combination equal to the zero vector. This equation has a summation left and the zero vector on the right. Isolate left in this equation those terms that involve basis vectors from $\mathcal{B}(\lambda_3)$, then move the remaining terms to the right. The rearranged equation looks like

$$\text{Sum of terms from } \mathcal{B}(\lambda_3) = \text{Sum of terms from } U_2$$

The left side is an eigenvector $\vec{\mathbf{v}}$ for λ_3. The right side is a linear combination from U_2, which means $\vec{\mathbf{v}} = \sum_{j=1}^q c_j \vec{\mathbf{v}}_j$. Write two equations for $\lambda_3\vec{\mathbf{v}}$, using the eigenpair equation $A\vec{\mathbf{v}} = \lambda_3\vec{\mathbf{v}}$:

$$\lambda_3\vec{\mathbf{v}} = \sum_{j=1}^q c_j \lambda_3 \vec{\mathbf{v}}_j, \quad \lambda_3\vec{\mathbf{v}} = A\vec{\mathbf{v}} = \sum_{j=1}^q c_j A\vec{\mathbf{v}}_j = \sum_{j=1}^q c_j \lambda(\vec{\mathbf{v}}_j)\vec{\mathbf{v}}_j,$$

[7] Mathematical induction is this theorem:
(1) For each counting number n, S_n is a statement that is either true or false.
(2) Statement S_1 is true.
(3) If statement S_k is true, then statement S_{k+1} is true.
Conclusion: All the statements are true.

where $\lambda(\vec{\mathbf{v}}_j)$ is the eigenvalue for eigenvector $\vec{\mathbf{v}}_j$. Put these two equations together, then move the right side to the left and collect terms:

$$\sum_{j=1}^{q} c_j(\lambda_3 - \lambda(\vec{\mathbf{v}}_j))\vec{\mathbf{v}}_j = \vec{\mathbf{0}}.$$

Because S_2 is true, then the vectors $\{\vec{\mathbf{v}}_j\}_{j=1}^{q}$ are independent. Therefore, all coefficients $c_j(\lambda_3 - \lambda(\vec{\mathbf{v}}_j)) = 0$. **Reminder**: symbols $\lambda_1, \ldots, \lambda_k$ are distinct values and list all eigenvalues of A. Then $\lambda_3 \neq \lambda(\vec{\mathbf{v}}_j)$ implies all $c_j = 0$. This implies $\vec{\mathbf{v}} = \vec{\mathbf{0}}$, which in turn implies that all coefficients in the independence test are zero. Therefore, U is a list of independent vectors. The induction proof is completed by the exercises of this section. ■

Proof of Theorem 9.6, Finding Independent Eigenvectors:
Exercises of this section show that $\partial_{t_1}\vec{\mathbf{v}}, \partial_{t_2}\vec{\mathbf{v}}, \ldots$ are independent vectors which constitute a basis $\mathcal{B}(\lambda)$ for the solution set of the eigenpair equation $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$. These are Strang's special solutions for $B\vec{\mathbf{v}} = \vec{\mathbf{0}}$. Theorem 9.5 says that the union U of bases $\mathcal{B}(\lambda_1)$, ..., $\mathcal{B}(\lambda_k)$ so constructed from the distinct eigenvalues λ_1, ..., λ_k of A is an independent set. For an example where U does not span $\mathcal{R}^n$, let $n = 2$ and $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, a matrix with just one eigenpair. ■

Proof of Theorem 9.8, Diagonalization:
Details 1. To prove $A = PDP^{-1}$, multiply right across $AP = PD$ by matrix P^{-1}, which isolates A on the left. Then $A = AI = APP^{-1} = PDP^{-1}$.

Details 2. Define the change of variables $\vec{\mathbf{X}} = P\vec{\mathbf{x}}$, $\vec{\mathbf{Y}} = P\vec{\mathbf{y}}$. Substitute into the equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ as follows:

$$\vec{\mathbf{Y}} = P\vec{\mathbf{y}} = PA\vec{\mathbf{x}} = PAP^{-1}P\vec{\mathbf{x}} = D\vec{\mathbf{X}}.$$

The result is the diagonal system $\vec{\mathbf{Y}} = D\vec{\mathbf{X}}$.

Details 3. Let column vector $\vec{\mathbf{c}}$ have components $c_1, \ldots, c_n$. To be proved: the left side of $A(c_1\vec{\mathbf{v}}_1 + \cdots + c_n\vec{\mathbf{v}}_n) = c_1\lambda_1\vec{\mathbf{v}}_1 + \cdots + c_n\lambda_n\vec{\mathbf{v}}_n$ is the expansion of $AP\vec{\mathbf{c}}$, while the right side is the expansion of $PD\vec{\mathbf{c}}$. Assume these statements are proved, for the moment, details delayed. Then $AP = PD$ implies $AP\vec{\mathbf{c}} = PD\vec{\mathbf{c}}$ for all vectors $\vec{\mathbf{c}}$, which means (7) holds. It remains to expand $AP\vec{\mathbf{c}}$ and $PD\vec{\mathbf{c}}$, assuming $AP = PD$, or what is the same, the eigenpair equations hold: $A\vec{\mathbf{c}}_i = \lambda_i\vec{\mathbf{v}}_i$ for $1 \le i \le n$.

The expansion of $AP\vec{\mathbf{c}}$:

$$\begin{aligned}
AP\vec{\mathbf{c}} &= A < \vec{\mathbf{v}}_1|\cdots|\vec{\mathbf{v}}_n > \vec{\mathbf{c}} && \text{Use definition } P =< \vec{\mathbf{v}}_1|\cdots|\vec{\mathbf{v}}_n >. \\
&= A(c_1\vec{\mathbf{v}}_1 + \ldots + c_n\vec{\mathbf{v}}_n) && \text{Matrix multiply as a linear combination of the columns.} \\
&= c_1A\vec{\mathbf{v}}_1 + \ldots + c_nA\vec{\mathbf{v}}_n && \text{Linearity of matrix multiply.} \\
&= c_1\lambda_1\vec{\mathbf{v}}_1 + \ldots + c_n\lambda_n\vec{\mathbf{v}}_n && \text{Eigenpair equations } A\vec{\mathbf{v}}_i = \lambda_i\vec{\mathbf{v}}_i \text{ for } 1 \le i \le n.
\end{aligned}$$

The expansion of $PD\vec{\mathbf{c}}$:

$$PD\vec{\mathbf{c}} = P\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}\vec{\mathbf{c}} \qquad \text{Definition of } D.$$

$$=P\begin{pmatrix} c_1\lambda_1 \\ \vdots \\ c_n\lambda_n \end{pmatrix}$$ Matrix multiply as a dot product.

$$= <\vec{\mathbf{v}}_1|\cdots|\vec{\mathbf{v}}_n> \begin{pmatrix} c_1\lambda_1 \\ \vdots \\ c_n\lambda_n \end{pmatrix}$$ Definition of P.

$$=c_1\lambda_1\vec{\mathbf{v}}_1 + \ldots + c_n\lambda_n\vec{\mathbf{v}}_n$$ Matrix multiply as a linear combination of columns.

■

Proof of Theorem 9.9, Distinct Eigenvalues:
Each eigenvalue λ has at least one eigenvector. Because there are n distinct eigenvalues, then there are n eigenvectors. The list of these eigenvectors must be independent, by Theorem 9.5. Therefore, matrix A is diagonalizable. The remaining statements in the theorem are a consequence of Theorem 9.8. ■

Proof of Theorem 9.10, Matrix Form Fourier Replacement:
$\boxed{1}$ Let's prove $\vec{\mathbf{x}} = P\vec{\mathbf{c}}$ implies $\vec{\mathbf{y}} = PD\vec{\mathbf{c}}$, assuming the Fourier replacement equation. Let $\vec{\mathbf{x}} = P\vec{\mathbf{c}}$. Expand the product $P\vec{\mathbf{c}}$ viewing matrix multiply as a linear combination of the columns. Then $\vec{\mathbf{x}} = P\vec{\mathbf{c}} = c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3$. Because Fourier replacement holds, then

$$\vec{\mathbf{y}} = c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\vec{\mathbf{v}}_3$$ Re-scale $\vec{\mathbf{x}}$.

$$= P\begin{pmatrix} c_1\lambda_1 \\ c_2\lambda_2 \\ c_3\lambda_3 \end{pmatrix}$$ Matrix multiply as a linear combination of columns.

$$= P\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$ Matrix multiply as a dot product.

$$= PD\vec{\mathbf{c}}$$ Definition of D.

$\boxed{2}$ Definition $A = PDP^{-1}$ was discovered by solving for A in equation $AP = PD$ ($AP = PD$ means A is diagonalizable). To prove $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$, first solve $\vec{\mathbf{x}} = P\vec{\mathbf{c}}$ for $\vec{\mathbf{c}} = P^{-1}\vec{\mathbf{x}}$. Then $A\vec{\mathbf{x}} = PDP^{-1}\vec{\mathbf{x}} = PD\vec{\mathbf{c}} = \vec{\mathbf{y}}$ by $\boxed{1}$.

$\boxed{3}$ To prove $A(c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3) = c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\vec{\mathbf{v}}_3$, replace its left side by $A\vec{\mathbf{x}}$ and right side by $\vec{\mathbf{y}}$. Then it suffices to prove $A\vec{\mathbf{x}} = \vec{\mathbf{y}}$, which has already been proved in $\boxed{2}$. ■

Proof of Theorem 9.11, Re-scaling and Diagonalization:
(a) Use relation $A(c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + c_3\vec{\mathbf{v}}_3) = c_1\lambda_1\vec{\mathbf{v}}_1 + c_2\lambda_2\vec{\mathbf{v}}_2 + c_3\lambda_3\vec{\mathbf{v}}_3$ from Theorem 9.10. Choose $c_1 = 1, c_2 = c_3 = 0$ to get $A\vec{\mathbf{v}}_1 = \lambda_1\vec{\mathbf{v}}_1$. Similarly, choose zeros and ones for c_1, c_2, c_3 to get $A\vec{\mathbf{v}}_2 = \lambda_2\vec{\mathbf{v}}_2$ and $A\vec{\mathbf{v}}_3 = \lambda_3\vec{\mathbf{v}}_3$. Then three eigenpair equations hold with independent eigenvectors and by definition A is diagonalizable.

(b) By Theorem 9.10 it suffices to prove $\vec{\mathbf{x}} = P\vec{\mathbf{c}}$ implies $A\vec{\mathbf{x}} = PD\vec{\mathbf{c}}$. If A is diagonalizable, then $AP = PD$, which gives $A\vec{\mathbf{x}} = AP\vec{\mathbf{c}} = PD\vec{\mathbf{c}}$ as required.

(c) If A is given and (8) holds, then **(a)** applies to prove A is diagonalizable. Conversely, if A is diagonalizable, then **(b)** applies and Fourier replacement (8) holds. ■

Exercises 9.1

Eigenanalysis

Classify as true or false. If false, then explain.

1. The purpose of eigenanalysis is to discover a new coordinate system.

2. Eigenanalysis can discover an opportunistic change of coordinates.

3. A matrix can have eigenvalue 0.

4. Eigenvalues are scale factors, imagined to be measurement units.

5. Eigenvectors are directions.

6. For each eigenvalue of a matrix A, there always exists at least one eigenpair.

7. If A^{-1} has eigenvalue λ, then A has eigenvalue $1/\lambda$.

8. Eigenvectors cannot be $\vec{\mathbf{0}}$.

9. The transpose of A has the same eigenvalues as A.

10. Eigenpairs $(\lambda, \vec{\mathbf{v}})$ of A satisfy the equation $(A - \lambda I)\vec{\mathbf{v}} = \vec{\mathbf{0}}$.

Eigenpairs of a Diagonal Matrix

Find eigenpairs of A without computation. Use Theorem 9.7.

11. $\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$

12. $\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$

13. $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

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

15. $\begin{pmatrix} 7 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -6 \end{pmatrix}$

16. $\begin{pmatrix} 2 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -1 \end{pmatrix}$

Fourier Replacement

Let symbols c_1, c_2 represent arbitrary constants. Let 2×2 matrix A have Fourier replacement equation

$$A\left(c_1\begin{pmatrix}1\\1\end{pmatrix} + c_2\begin{pmatrix}1\\2\end{pmatrix}\right) = 2c_1\begin{pmatrix}1\\1\end{pmatrix} - 5c_2\begin{pmatrix}1\\2\end{pmatrix}$$

17. Display the eigenpairs of A.

18. Display the replacement equation if the eigenvalues $2, -5$ are replaced by $1, 0$.

19. Display the eigenpair packages P, D such that $AP = PD$.

20. Find A.

Eigenanalysis Facts

Mark as true or false, then explain your answer.

21. If matrix A has all eigenvalues zero, then A is the zero matrix.

22. If 2×2 matrix A has all eigenvalues zero, then Fourier's replacement equation is

$$A\left(c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2\right) = \vec{\mathbf{0}}.$$

23. There are infinitely many 2×2 matrices A with complex eigenvalues $1+i, 1-i$.

24. A real 2×2 matrix A with eigenvalues $2+3i, 2-3i$ cannot have a real eigenvector.

25. A real 2×2 matrix A with real eigenvalues has only real eigenvectors.

26. A real 2×2 matrix A with complex eigenvalues has only complex eigenvectors.

Eigenpair Packages and equation $AP = PD$

27. Suppose A has eigenpair packages. Explain why there are so many different answers for P, D.

28. Suppose $AP = PD$ and $AQ = QD$ hold (same diagonal matrix D). Does $P = Q$?

29. Find one choice of P and D for $A = 2 \times 2$ diagonal matrix.

30. Given $A = 3 \times 3$ zero matrix, find one choice of P and D with column one of P equal to $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.

Matrix Eigenanalysis Method

31. The eigenvalues of $\begin{pmatrix} 1 & 3 \\ 1 & 4 \end{pmatrix}$ satisfy a quadratic equation. Find the equation and solve for the eigenvalues.

32. Find the eigenvalues of $\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$.

33. Find all eigenpairs of $\begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$.

34. A triangular $n \times n$ matrix with distinct diagonal entries has n eigenpairs. Provide a detailed proof for the case $n = 3$.

35. Find all eigenpairs of $\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$.

36. A triangular $n \times n$ matrix may not have n eigenpairs. Provide a series of examples for dimensions $n = 2, 3, 4, 5$.

37. Prove that equations $A\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$ and $(A - \lambda I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$ have exactly the same solutions $\vec{\mathbf{x}}$.

38. Cite basic linear algebra theorems to prove that $(A - \lambda I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$ has a nonzero solution $\vec{\mathbf{x}}$ if and only if λ is a root of the characteristic equation $|A - \lambda I| = 0$.

Basis of Eigenvectors

The problem $A\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$ has a standard general solution $\vec{x}$ with invented symbols $t_1, t_2, t_3, \ldots$. **Strang's special solutions** are defined to be the vector partial derivatives of $\vec{x}$ with respect to the invented symbols.

39. Why are Strang's special solutions independent?

40. Prove that linear combinations of Strang's special solutions provide all possible solutions of $A\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$.

Independence of Eigenvectors

Eigenvectors of matrix A for eigenvalue λ are the nonzero solutions of $A\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$.

41. Invent a 2×2 example A with eigenpairs $\left(2, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$, $\left(2, \begin{pmatrix} 5 \\ 5 \end{pmatrix}\right)$. Then explain why an eigenvector for eigenvalue λ is never unique.

42. Explain: *For a given eigenvalue λ, there are infinitely many eigenvectors.*

43. Explain: *Each solution $\vec{\mathbf{x}}$ of $A\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$ is a linear combination of Strang's special solutions for $B = A - \lambda I$.*

44. Let P be an invertible 3×3 matrix. Construct a matrix A which has eigenvectors equal to the columns of P and corresponding eigenvalues $-1, 0, 0$.

Eigenspaces

Let $\mathcal{B}(\lambda)$ denote some basis of eigenvectors for the eigenpair equation $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$. The **eigenspace** for λ is the subspace $\mathbf{span}(\mathcal{B}(\lambda))$.

45. Explain: The eigenspace of λ does not depend on the choice of basis.

46. Every nonzero vector in eigenspace $\mathbf{span}(\mathcal{B}(\lambda))$ is an eigenvector of A for eigenvalue λ. Provide details of proof.

47. Justify that $\mathbf{span}(\mathcal{B}(\lambda))$ is a vector subspace of $\mathcal{R}^n$, one possible basis being Strang's special solutions for matrix $B = A - \lambda I$.

48. Find a 4×4 matrix A with only one eigenvalue $\lambda = 1$ such that eigenspace $\mathcal{B}(\lambda)$ (defined above) has dimension two.

Independence of Unions of Eigenvectors

Denote by $\mathcal{B}(\lambda)$ some basis for the eigenpair equation $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$.

49. Define U_1 to be the union of lists $\mathcal{B}(\lambda_1)$, $\mathcal{B}(\lambda_2)$ and define U_2 to be the union of lists $\mathcal{B}(\lambda_3)$, $\mathcal{B}(\lambda_4)$, where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ is a list of distinct eigenvalues of A. Prove that subspaces $V_1 = \mathbf{span}(U_1)$ and $V_2 = \mathbf{span}(U_2)$ intersect in only the zero vector.

50. Complete the details of the induction proof of Theorem 9.5, using the textbook details for $k = 3$.

51. Let U^* be a subset of the list U of independent vectors in Theorem 9.5. Explain why U^* is an independent set.

52. Let B_i be a subset of the list of independent vectors in $\mathcal{B}(\lambda_i)$, $i = 1, \ldots, p$. Explain why the union U^* of $B_1, \ldots, B_p$ is an independent set.

Diagonalization Theory

53. Let $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 8 \end{pmatrix}$.

(a) Find Strang's special solutions for each eigenvalue.
(b) Compare to Theorem 9.7 on diagonal matrices.

54. Let $\vec{\mathbf{v}}_!, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ be independent vectors in $\mathcal{R}^3$. Explain why $(0, \vec{\mathbf{v}}_1)$, $(0, \vec{\mathbf{v}}_2)$, $(0, \vec{\mathbf{v}}_3)$ is a complete set of eigenpairs for the 3×3 zero matrix. Does this contradict Theorem 9.7?

55. Write a proof of Theorem 9.7 for $n = 3$.

56. State Theorem 9.7 for $n \times n$ diagonal matrices and outline a proof.

Non-diagonalizable Matrices

Verify that the matrix is not diagonalizable by using the equation $AP = PD$.

57. $A = \begin{pmatrix} 5 & 2 \\ 0 & 5 \end{pmatrix}$

58. $A = \begin{pmatrix} 5 & 2 & 1 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{pmatrix}$

Distinct Eigenvalues

Find the eigenvalues.

59. $A = \begin{pmatrix} 2 & 6 \\ 5 & 3 \end{pmatrix}$ Ans: $8, -3$

60. $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ Ans: $0, 5$

61. $A = \begin{pmatrix} 2 & 6 & 2 \\ 9 & 3 & 9 \\ 1 & 3 & 1 \end{pmatrix}$ Ans: $0, 12, -6$

62. $A = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 3 \end{pmatrix}$ Ans: $0, 1, 3$

63. $A = \begin{pmatrix} 7 & 12 & 6 \\ 2 & 2 & 2 \\ -7 & -12 & -6 \end{pmatrix}$ Ans: $0, 1, 2$

64. $A = \begin{pmatrix} 2 & 2 & -6 \\ -3 & -4 & 3 \\ -3 & -4 & -1 \end{pmatrix}$ Ans: $0, 1, 4$

Computing 2×2 Eigenpairs

65. Verify eigenpairs: $\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}$, $\left(-1, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right)$, $\left(5, \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}\right)$

66. Verify eigenpairs: $\begin{pmatrix} 1 & 6 \\ 2 & -3 \end{pmatrix}$, $\left(-5, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right)$, $\left(3, \begin{pmatrix} 3 \\ -1 \end{pmatrix}\right)$

67. Verify eigenpairs: $\begin{pmatrix} 1 & 6 \\ 4 & 3 \end{pmatrix}$, $\left(7, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$, $\left(-3, \begin{pmatrix} -3 \\ 2 \end{pmatrix}\right)$

68. Verify eigenpairs: $\begin{pmatrix} 7 & 4 \\ -1 & 3 \end{pmatrix}$, $\left(5, \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)$, only one eigenpair

Computing 2×2 Complex Eigenpairs

69. Verify eigenpairs: $\begin{pmatrix} -2 & -6 \\ 3 & 4 \end{pmatrix}$, $\left(1 + 3i, \begin{pmatrix} -1 + i \\ 1 \end{pmatrix}\right)$, $\left(1 - 3i, \begin{pmatrix} -1 - i \\ 1 \end{pmatrix}\right)$

70. Verify eigenpairs: $\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}$, $\left(2+3i, \begin{pmatrix} -i \\ 1 \end{pmatrix}\right)$, $\left(2-3i, \begin{pmatrix} i \\ 1 \end{pmatrix}\right)$

71. Let a, b be real with $b \neq 0$. Assume $n \times n$ real matrix A has eigenpair $(a+ib, \vec{\mathbf{v}})$. Replace i by $-i$ throughout expression $\vec{\mathbf{v}}$ to obtain vector $\vec{\mathbf{w}}$. Prove that $(a-ib, \vec{\mathbf{w}})$ is an eigenpair.

72. Explain: Eigenpairs of a 2×2 real matrix A with complex eigenvalues are computed with just one row-reduction sequence.

Computing 3×3 Eigenpairs

73. Show algorithm steps to compute eigenpairs of $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.

Answers: $\left(1, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}\right)$, $\left(3, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right)$

74. Show algorithm steps to compute eigenpairs of $A = \begin{pmatrix} 1 & -2 & 0 \\ 0 & -1 & 0 \\ 4 & -4 & -1 \end{pmatrix}$.

Answers:
$\left(1, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}\right)$, $\left(-1, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right)$, $\left(-1, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right)$

75. Suppose A is row-reduced to a triangular form B. Are the eigenvalues of B also the eigenvalues of A? Give a proof or a counter-example.

76. Suppose $A - \lambda I$ is row-reduced to a triangular form B. Explain: The eigenvalues of A are usually unrelated to the roots λ of $|B| = 0$.

Decomposition $A = PDP^{-1}$

Compute the eigenpairs. If diagonalizable, then display D, P and Fourier's replacement equation.

77. $A = \begin{pmatrix} 7 & 4 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

Ans: only 2 eigenpairs

.

78. $A = \begin{pmatrix} 1 & 6 & 0 \\ 2 & -3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

Ans: $\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -5 \end{pmatrix}$, $\begin{pmatrix} 3 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

Fourier equation: $AP\vec{\mathbf{c}} = PD\vec{\mathbf{c}}$.

Diagonalization

Report **diagonalizable** or not and explain why.

79. $A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & -3 \end{pmatrix}$

Ans: diagonalizable

80. $A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}$

Ans: not diagonalizable

Non-diagonalizable Matrices

81. Verify $A = \begin{pmatrix} 1 & 2 \\ -8 & 9 \end{pmatrix}$ is not diagonalizable.

82. Verify $A = \begin{pmatrix} 1 & 2 & 0 \\ -8 & 9 & 1 \\ 0 & 0 & 5 \end{pmatrix}$ is not diagonalizable.

83. Invent a 3×3 matrix which has exactly one eigenpair.

84. Invent a 4×4 matrix which has exactly two eigenpairs.

Fourier's Heat Model

Define
$\vec{\mathbf{v}}_1 = \sin \pi x, \vec{\mathbf{v}}_2 = \sin 2\pi x, \vec{\mathbf{v}}_3 = \sin 3\pi x$
considered as vectors in the vector space V of twice continuously differentiable functions on $0 \le x \le 1$.

85. Verify that $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ are independent vectors in V.

86. Verify that $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3$ vanish at $x = 0$ and $x = 1$.

87. Define $u(x) = \sin \pi x$ (from $\vec{\mathbf{v}}_1$). Explain: Function u satisfies differential equation $\dfrac{d^2u}{dx^2} + \pi^2 u = 0$.

88. Write vector expression

$$c_1 e^{-\pi^2 t}\vec{\mathbf{v}}_1 + c_2 e^{-4\pi^2 t}\vec{\mathbf{v}}_2 + c_3 e^{-9\pi^2 t}\vec{\mathbf{v}}_3$$

as a scalar function $u(t,x)$. Find initial heat distribution $u(0,x)$. Explain how Fourier replacement (re-scaling) constructs future state $u(t,x)$ from initial state $u(0,x)$.

9.2 Eigenanalysis Applications

Discrete Dynamical Systems

The matrix equation

$$(1) \qquad \vec{\mathbf{y}} = A\vec{\mathbf{x}}, \quad A = \frac{1}{10}\begin{pmatrix} 5 & 4 & 0 \\ 3 & 5 & 3 \\ 2 & 1 & 7 \end{pmatrix}$$

predicts the state $\vec{\mathbf{y}}$ of a system initially in state $\vec{\mathbf{x}}$ after some fixed elapsed time. The 3×3 matrix A in (1) represents the **dynamics** which changes state $\vec{\mathbf{x}}$ into state $\vec{\mathbf{y}}$.

An equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ like equation (1) is called a **discrete dynamical system**. The fixed elapsed time for changing $\vec{\mathbf{x}}$ to $\vec{\mathbf{y}}$ is called the **period** of the discrete dynamical system. Matrix A is called a **transition matrix**, provided A has nonnegative entries and column sums equal to one. See **stochastic matrices** page 704

The eigenpairs of matrix A in (1) are shown on page 713 to be $(1, \vec{\mathbf{v}}_1)$, $(1/2, \vec{\mathbf{v}}_2)$, $(1/5, \vec{\mathbf{v}}_3)$ with eigenvectors

$$(2) \qquad \vec{\mathbf{v}}_1 = \begin{pmatrix} 12 \\ 15 \\ 13 \end{pmatrix}, \quad \vec{\mathbf{v}}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{\mathbf{v}}_3 = \begin{pmatrix} -4 \\ 3 \\ 1 \end{pmatrix}.$$

Market Shares

A model application of discrete dynamical systems is telephone long distance company market shares x_1, x_2, x_3, which are fractions of the total market for long distance service. If three companies provide all the services, then their market fractions add to one: $x_1 + x_2 + x_3 = 1$. Equation $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ in (1) with eigenpairs (2) predicts the market shares of the three companies after a fixed time period, say one year. Market shares after one, two and three years are given by the **iterates**

$$\begin{aligned} \vec{\mathbf{y}}_1 &= A\vec{\mathbf{x}}, \\ \vec{\mathbf{y}}_2 &= A^2\vec{\mathbf{x}}, \\ \vec{\mathbf{y}}_3 &= A^3\vec{\mathbf{x}}. \end{aligned}$$

Fourier's replacement model (8) page 676 gives succinct and useful formulas for the iterates. If $\vec{\mathbf{x}} = a_1\vec{\mathbf{v}}_1 + a_2\vec{\mathbf{v}}_2 + a_3\vec{\mathbf{v}}_3$, then

$$(3) \qquad \begin{aligned} \vec{\mathbf{y}}_1 &= A\vec{\mathbf{x}} &&= a_1\lambda_1\vec{\mathbf{v}}_1 + a_2\lambda_2\vec{\mathbf{v}}_2 + a_3\lambda_3\vec{\mathbf{v}}_3, \\ \vec{\mathbf{y}}_2 &= A^2\vec{\mathbf{x}} &&= a_1\lambda_1^2\vec{\mathbf{v}}_1 + a_2\lambda_2^2\vec{\mathbf{v}}_2 + a_3\lambda_3^2\vec{\mathbf{v}}_3, \\ \vec{\mathbf{y}}_3 &= A^3\vec{\mathbf{x}} &&= a_1\lambda_1^3\vec{\mathbf{v}}_1 + a_2\lambda_2^3\vec{\mathbf{v}}_2 + a_3\lambda_3^3\vec{\mathbf{v}}_3. \end{aligned}$$

The eigenpairs of A in (2) show that $\lambda_1 = 1$ and $\lim_{n\to\infty} |\lambda_2|^n = \lim_{n\to\infty} |\lambda_3|^n = 0$. Then for large n

$$\vec{\mathbf{y}}_n \approx a_1(1)\vec{\mathbf{v}}_1 + a_2(0)\vec{\mathbf{v}}_2 + a_3(0)\vec{\mathbf{v}}_3 = \begin{pmatrix} 12a_1 \\ 15a_1 \\ 13a_1 \end{pmatrix}.$$

The numbers a_1, a_2, a_3 are related to x_1, x_2, x_3 in the expansion $\vec{\mathbf{x}} = a_1\vec{\mathbf{v}}_1 + a_2\vec{\mathbf{v}}_2 + a_3\vec{\mathbf{v}}_3$ by the equations $12a_1 - a_2 - 4a_3 = x_1$, $15a_1 + 3a_3 = x_2$, $13a_1 + a_2 + a_3 = x_3$. Because $x_1 + x_2 + x_3 = 1$, then $a_1 = 1/40$. The three market shares after a long time period are predicted to be $3/10$, $3/8$, $13/40$. The market share identity $\frac{3}{10} + \frac{3}{8} + \frac{13}{40} = 1$ holds because approximating terms from (3) are sums of market shares adding to one.

Stochastic Matrices

The special matrix A in (1) is a **stochastic matrix**[8], defined by the properties

$$\sum_{i=1}^{n} a_{ij} = 1, \quad a_{kj} \geq 0, \quad k, j = 1, \ldots, n.$$

The definition is memorized by the phrase *each column sum is one.*

Leontief input-output models are stochastic models, popularized by 1973 Nobel Prize economist Wassily Leontief. A typical model is $A = R^T$ where

$$R = \begin{pmatrix} 1 & 0 & 0 \\ .2 & .3 & .5 \\ .4 & .4 & .2 \end{pmatrix}.$$

The **rows** of R add to one, therefore the **columns** of A add to one. Row 1 is the bank, Row 2 is Factory 1, Row 3 is Factory 2. Matrix R tracks the money as it is being passed back and forth between the factories and the bank.

Leslie Models in population biology are similar to stochastic models. It is a discrete time model $\vec{\mathbf{v}}_{i+1} = A\vec{\mathbf{v}}_i$ of an age-structured population describing mortality, reproduction and development. The **Leslie matrix** A for $n = 4$ looks like

$$A = \begin{pmatrix} f_1 & f_2 & f_3 & f_4 \\ s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \end{pmatrix}.$$

Neither the row sums nor the column sums are one. However, some stochastic matrix results have analogs for Leslie matrices. Population vector $\vec{\mathbf{v}}_i$ contains counts of age classes. Number $f_i \geq 0$ is the average number of female births for a mother of age class i. Number $s_i \geq 0$ is the fraction of individuals of age class i that survive to age class $i + 1$.

[8]Technically, a **right** stochastic matrix, which means columns add to one. A **left** stochastic matrix has rows adding to one. The term **transition matrix** is also used.

Theorem 9.12 (Stochastic Matrix Properties)
Let A be a stochastic matrix. Then

(a) If $\vec{\mathbf{x}}$ is a vector with $x_1 + \cdots + x_n = 1$, then $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ satisfies $y_1 + \cdots + y_n = 1$.

(b) If the components of $\vec{\mathbf{v}}$ are all 1, then $A^T\vec{\mathbf{v}} = \vec{\mathbf{v}}$. Therefore, $(1, \vec{\mathbf{v}})$ is an eigenpair of A^T.

(c) One root of the characteristic equation $\det(A - \lambda I) = 0$ is $\lambda = 1$. All other roots satisfy $|\lambda| \le 1$.

Proof on page 715.

Theorem 9.13 (Perron-Frobenius: Positive Stochastic Matrix)
Let A be a stochastic matrix all of whose entries are strictly positive. Then

(a) There exists an eigenpair $(1, \vec{\mathbf{w}})$ of A such that $\vec{\mathbf{w}}$ has nonnegative components and $\lim_{n\to\infty} A^n = \left\langle \vec{\mathbf{w}}|\vec{\mathbf{w}}|\cdots|\vec{\mathbf{w}} \right\rangle$.

(b) If $(1, \vec{\mathbf{v}})$ is an eigenpair of A, then $\vec{\mathbf{v}} = c\vec{\mathbf{w}}$ for $c = \sum_{i=1}^n v_i$. Briefly, the eigenspace for $\lambda = 1$ has dimension one.

(c) If $\lambda \ne 1$ is a real or complex eigenvalue of A, then $|\lambda| < 1$.

(d) If $(\lambda, \vec{\mathbf{v}})$ is an eigenpair of A and $\vec{\mathbf{v}}$ has nonnegative components, then all components of $\vec{\mathbf{v}}$ are strictly positive, $\lambda = 1$ and $\vec{\mathbf{v}} = c\vec{\mathbf{w}}$ for some constant c.

Proof on page 715.

Coupled and Uncoupled Systems

The linear system of differential equations

$$
\begin{aligned}
x_1' &= -x_1 - x_3, \\
x_2' &= 4x_1 - x_2 - 3x_3, \\
x_3' &= 2x_1 - 4x_3,
\end{aligned} \tag{4}
$$

is called **coupled**, whereas the linear system of growth-decay equations

$$
\begin{aligned}
y_1' &= -3y_1, \\
y_2' &= -y_2, \\
y_3' &= -2y_3,
\end{aligned} \tag{5}
$$

is called **uncoupled**. The terminology *uncoupled* means that each differential equation in system (5) depends on exactly one variable, e.g., $y_1' = -3y_1$ depends only on variable y_1. In a *coupled* system, one of the differential equations must involve two or more variables.

Matrix Formulation

Coupled system (4) and uncoupled system (5) can be written in matrix form, $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}' = D\vec{\mathbf{y}}$, with coefficient matrices

$$A = \begin{pmatrix} -1 & 0 & -1 \\ 4 & -1 & -3 \\ 2 & 0 & -4 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} -3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

If the coefficient matrix is **diagonal**, then the system is **uncoupled**. If the coefficient matrix is **not diagonal**, then one of the corresponding differential equations involves two or more variables and the system is called **coupled** or **cross-coupled**.

Solving Uncoupled Systems

An uncoupled system consists of independent growth-decay equations of the form $u' = au$. The solution formula $u = ce^{at}$ then leads to the general solution of the system of equations. For instance, system (5) has general solution

$$\begin{aligned} y_1 &= c_1 e^{-3t}, \\ y_2 &= c_2 e^{-t}, \\ y_3 &= c_3 e^{-2t}, \end{aligned} \tag{6}$$

where c_1, c_2, c_3 are **arbitrary constants**. The number of constants equals the dimension of the diagonal matrix D.

Coordinates and Coordinate Systems

If vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ are independent in $\mathcal{R}^3$, then augmented matrix

$$P = \left\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 | \vec{\mathbf{v}}_3 \right\rangle$$

is invertible. The columns $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ of P are called a **coordinate system**. The matrix P is called a **change of coordinates**.

Independence of $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ means every vector $\vec{\mathbf{v}}$ in $\mathcal{R}^3$ can be uniquely expressed as

$$\vec{\mathbf{v}} = t_1\vec{\mathbf{v}}_1 + t_2\vec{\mathbf{v}}_2 + t_3\vec{\mathbf{v}}_3.$$

The values t_1, t_2, t_3 are called the **coordinates** of $\vec{\mathbf{v}}$ relative to the basis $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$, or the coordinates of $\vec{\mathbf{v}}$ relative to P.

Viewpoint of a Driver

The physical meaning of a coordinate system $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ can be understood by considering an auto traveling up a mountain road. Choose orthogonal $\vec{\mathbf{v}}_1$ and

$\vec{\mathbf{v}}_2$ to give positions in the driver's seat and define $\vec{\mathbf{v}}_3$ be the seat-back direction. These are **local coordinates** as viewed from the driver's seat. The road map coordinates x, y and the altitude z define the **global coordinates** for the auto's position $\vec{\mathbf{p}} = x\vec{\imath} + y\vec{\jmath} + z\vec{k}$.

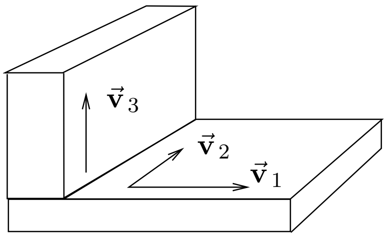


Figure 2. Driver's coordinates.
The vectors $\vec{\mathbf{v}}_1(t)$, $\vec{\mathbf{v}}_2(t)$, $\vec{\mathbf{v}}_3(t)$ form an orthogonal triad which is a local coordinate system from the driver's viewpoint. The orthogonal triad changes continuously in t.

Change of Coordinates $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$

A coordinate change from $\vec{\mathbf{y}}$ to $\vec{\mathbf{x}}$ is a linear algebraic equation $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ where the $n \times n$ matrix P is required to be invertible ($\det(P) \neq 0$). To illustrate, an instance of a change of coordinates from $\vec{\mathbf{y}}$ to $\vec{\mathbf{x}}$ is given by the linear equations

$$(7) \qquad \vec{\mathbf{x}} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix} \vec{\mathbf{y}} \quad \text{or} \quad \begin{cases} x_1 &= y_1 + y_3, \\ x_2 &= y_1 + y_2 - y_3, \\ x_3 &= 2y_1 + y_3. \end{cases}$$

Constructing Coupled Systems

A general method exists to construct rich examples of coupled systems. The idea uses a change of variables for a given uncoupled system. Consider a diagonal system $\vec{\mathbf{y}}' = D\vec{\mathbf{y}}$, like (5), and a change of variables $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$, like (7). Differential calculus applies to give

$$(8) \qquad \begin{aligned} \vec{\mathbf{x}}' &= (P\vec{\mathbf{y}})' \\ &= P\vec{\mathbf{y}}' \\ &= PD\vec{\mathbf{y}} \\ &= PDP^{-1}\vec{\mathbf{x}}. \end{aligned}$$

The matrix $A = PDP^{-1}$ is *not triangular* in general, and therefore the change of variables produces a **cross-coupled** system.

An illustration. To give an example, substitute into uncoupled system (5) the change of variable equations (7). Use equation (8) to obtain

$$(9) \qquad \vec{\mathbf{x}}' = \begin{pmatrix} -1 & 0 & -1 \\ 4 & -1 & -3 \\ 2 & 0 & -4 \end{pmatrix} \vec{\mathbf{x}} \quad \text{or} \quad \begin{cases} x_1' = -x_1 - x_3, \\ x_2' = 4x_1 - x_2 - 3x_3, \\ x_3' = 2x_1 - 4x_3. \end{cases}$$

This **cross-coupled** system (9) can be solved using relations (7), (6) and $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ to give the general solution

$$(10) \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 e^{-3t} \\ c_2 e^{-t} \\ c_3 e^{-2t} \end{pmatrix}.$$

Changing Coupled Systems to Uncoupled

A question, motivated by the above calculations:

> Can every coupled system $\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t)$ be subjected to a change of variables $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ which converts the system into a completely uncoupled system for variable $\vec{\mathbf{y}}(t)$?

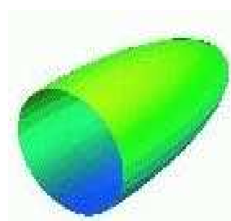
Answer: A coupled system can be so transformed if and only if matrices P and D are eigenpair packages of A. Then $AP = PD$ and A is diagonalizable. Conversely, if A is diagonalizable, then the packages P, D exist and $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ changes $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ into diagonal system $\vec{\mathbf{y}}' = D\vec{\mathbf{y}}$. The connection between $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$ is like (10).

Eigenanalysis provides the opportunity to simultaneously calculate from cross-coupled system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ a change of variable $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ and a diagonal matrix D for an uncoupled system $\vec{\mathbf{y}}' = D\vec{\mathbf{y}}$. System $\vec{\mathbf{y}}' = D\vec{\mathbf{y}}$ consists of uncoupled scalar growth-decay equations like (5).

Matrices A that fail to be diagonalizable present a problem, because eigenanalysis does not apply. The demand to obtain an uncoupled system $\vec{\mathbf{y}}' = D\vec{\mathbf{y}}$ leaves no alternative, because if there is a change of variables $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ into diagonal system $\vec{\mathbf{y}}' = D\vec{\mathbf{y}}$, then $AP = PD$ and A is diagonalizable, a contradiction.

There *does exist* a change of coordinates P to change $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ into a **triangular system** $\vec{\mathbf{y}}' = T\vec{\mathbf{y}}$. This system in scalar form can be solved by the linear integrating factor method. There is again an answer $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ like (5). See page 721.

Eigenanalysis and Footballs



An ellipsoid or *football* is a geometric object described by its **semi-axes** (see Figure 3). In the vector representation, the **semi-axis directions** are unit vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ and the **semi-axis lengths** are the constants a, b, c. The vectors $a\vec{\mathbf{v}}_1$, $b\vec{\mathbf{v}}_2$, $c\vec{\mathbf{v}}_3$ form an **orthogonal triad**.

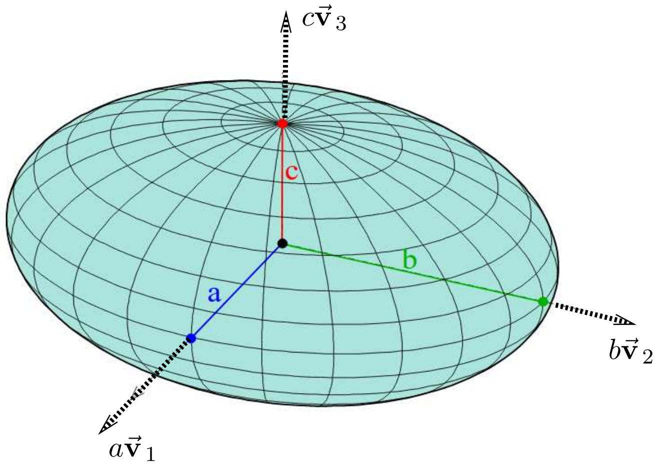


Figure 3. Ellipsoid. An ellipsoid is built from orthonormal semi-axis directions $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ and the semi-axis lengths a, b, c. The semi-axis vectors are $a\vec{\mathbf{v}}_1$, $b\vec{\mathbf{v}}_2$, $c\vec{\mathbf{v}}_3$.

Two vectors $\vec{\mathbf{u}}$, $\vec{\mathbf{w}}$ are *orthogonal* if both are nonzero and their dot product $\vec{\mathbf{u}} \cdot \vec{\mathbf{w}}$ is zero. Vectors are **orthonormal** if each has unit length and they are pairwise orthogonal. The orthogonal triad $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ is an **invariant** of the ellipsoid's algebraic representations. Algebra does not change the triad: the invariants $a\vec{\mathbf{v}}_1$, $b\vec{\mathbf{v}}_2$, $c\vec{\mathbf{v}}_3$ must somehow be **hidden** in the equations that represent the ellipsoid.

Algebraic eigenanalysis finds the hidden invariant triad $a\vec{\mathbf{v}}_1$, $b\vec{\mathbf{v}}_2$, $c\vec{\mathbf{v}}_3$ from the ellipsoid's algebraic equations. Suppose, for instance, that the equation of the ellipsoid is supplied as

$$x^2 + 4y^2 + xy + 4z^2 = 16.$$

A symmetric matrix A is constructed in order to write the equation in the form $\vec{\mathbf{X}}^T A \vec{\mathbf{X}} = 16$, where $\vec{\mathbf{X}}$ has components x, y, z. The replacement equation is[9]

$$(11) \qquad \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 1/2 & 0 \\ 1/2 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 16.$$

It is the 3×3 symmetric matrix A in (11) that is subjected to algebraic eigenanalysis. The matrix calculation will compute the unit semi-axis directions $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$, called the **eigenvectors** or **hidden vectors**. The semi-axis lengths a, b, c are computed at the same time, by finding the **eigenvalues** or **hidden values**[10] λ_1, λ_2, λ_3, known to satisfy the relations

$$\lambda_1 = \frac{16}{a^2}, \quad \lambda_2 = \frac{16}{b^2}, \quad \lambda_3 = \frac{16}{c^2}.$$

For the illustration, the football dimensions are $a = 2$, $b = 1.98$, $c = 4.17$. Details of the computation are delayed until page 711.

[9]Multiply matrices to verify this statement. Halving of the entries corresponding to cross-terms generalizes to any ellipsoid.

[10]The terminology *hidden* arises because neither the semi-axis lengths nor the semi-axis directions are revealed directly by the ellipsoid equation.

Ellipse and Eigenanalysis

An ellipse equation in **standard form** is $\lambda_1 u^2 + \lambda_2 v^2 = 1$, where $\lambda_1 = 1/a^2$, $\lambda_2 = 1/b^2$ are expressed in terms of the semi-axis lengths a, b. The expression $\lambda_1 u^2 + \lambda_2 v^2$ is called a **quadratic form**. The study of the ellipse $\lambda_1 u^2 + \lambda_2 v^2 = 1$ is equivalent to the study of the quadratic form equation

$$\vec{\mathbf{r}}^T D \vec{\mathbf{r}} = 1, \quad \text{where} \quad \vec{\mathbf{r}} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Cross-terms. An ellipse may be represented by an equation in a xy-coordinate system having a cross-term xy, e.g., $4x^2 + 8xy + 10y^2 = 5$. The expression $4x^2 + 8xy + 10y^2$ is again called a quadratic form. Calculus courses provide methods to eliminate the cross-term and represent the equation in standard form, by a **rotation** by angle θ of the xy-system into the uv-system:

$$\begin{pmatrix} u \\ v \end{pmatrix} = R \begin{pmatrix} x \\ y \end{pmatrix}, \quad R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Eigenanalysis computes angle θ through the columns of R, which are the unit semi-axis directions $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ for the ellipse $4x^2 + 8xy + 10y^2 = 5$. If the quadratic form $4x^2 + 8xy + 10y^2$ is represented as $\vec{\mathbf{r}}^T A \vec{\mathbf{r}}$, then

$$\vec{\mathbf{r}} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad A = \begin{pmatrix} 4 & 4 \\ 4 & 10 \end{pmatrix}, \quad R = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix},$$

$$\lambda_1 = 12, \quad \vec{\mathbf{v}}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \lambda_2 = 2, \quad \vec{\mathbf{v}}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Ellipse equations. There are two coordinate systems, the xy-system and the rotated uv-system. The equations in each system, each divided by 5:

$$(12) \qquad \begin{aligned} \tfrac{4}{5}x^2 + \tfrac{8}{5}xy + 2y^2 &= 1, \\ \tfrac{2}{5}u^2 + \tfrac{12}{5}v^2 &= 1. \end{aligned}$$

The rotation relation $\begin{pmatrix} u \\ v \end{pmatrix} = R \begin{pmatrix} x \\ y \end{pmatrix}$ is the set of equations

$$(13) \qquad \begin{cases} u &= \;= \frac{1}{\sqrt{5}}x - \frac{2}{\sqrt{5}}y, \\ v &= \;= \frac{2}{\sqrt{5}}x + \frac{1}{\sqrt{5}}y, \end{cases}$$

which upon substitution into the uv-equation in (12) gives

$$\frac{2}{5}\left(\frac{1}{\sqrt{5}}x - \frac{2}{\sqrt{5}}y\right)^2 + \frac{12}{5}\left(\frac{2}{\sqrt{5}}x + \frac{1}{\sqrt{5}}y\right)^2 = 1.$$

The reader can verify that this is the first equation in (12).

Rotation matrix angle θ. The components of unit eigenvector $\vec{\mathbf{v}}_1$ can be used to determine $\theta = -63.4°$:

$$\begin{pmatrix} \cos\theta \\ -\sin\theta \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{or} \quad \left\{ \begin{array}{rcl} \cos\theta & = & \frac{1}{\sqrt{5}}, \\ -\sin\theta & = & \frac{2}{\sqrt{5}}. \end{array} \right.$$

The interpretation of angle θ: rotate the orthonormal basis $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ by angle $\theta = -63.4°$ in order to obtain the standard unit basis vectors $\vec{\imath}$, $\vec{\jmath}$. Calculus texts might discuss only the inverse rotation, where x, y are given in terms of u, v. In these references, θ is the negative of the value given here, due to a different geometric viewpoint.[11]

Semi-axis lengths. The lengths $a \approx 1.55$, $b \approx 0.63$ for the ellipse $4x^2 + 8xy + 10y^2 = 5$ are computed from the eigenvalues $\lambda_1 = 12$, $\lambda_2 = 2$ of matrix A by the equations

$$\frac{\lambda_1}{5} = \frac{1}{a^2}, \quad \frac{\lambda_2}{5} = \frac{1}{b^2}.$$

Geometry. The ellipse $4x^2 + 8xy + 10y^2 = 5$ is completely determined by the orthogonal semi-axis vectors $a\vec{\mathbf{v}}_1$, $b\vec{\mathbf{v}}_2$. The rotation R is a rigid motion mapping xy-plane vectors $a\vec{\mathbf{v}}_1$, $b\vec{\mathbf{v}}_2$ into uv-plane vectors $a\vec{\imath}$, $b\vec{\jmath}$.

The θ-rotation R maps $4x^2+8xy+10y^2 = 5$ into the uv-equation $\lambda_1 u^2+\lambda_2 v^2 = 5$, where λ_1, λ_2 are the eigenvalues of A. To see why, let $\vec{\mathbf{r}} = \begin{pmatrix} u \\ v \end{pmatrix}$, $\vec{\mathbf{s}} = \begin{pmatrix} x \\ y \end{pmatrix}$ in the equation $\vec{\mathbf{r}} = R\vec{\mathbf{s}}$. Then $\vec{\mathbf{r}}^T A\vec{\mathbf{r}} = \vec{\mathbf{s}}^T(R^T AR)\vec{\mathbf{s}}$. Using $R^T R = I$ gives $R^{-1} = R^T$ and $R^T AR = \textbf{diag}(\lambda_1, \lambda_2)$. Finally, $\vec{\mathbf{r}}^T A\vec{\mathbf{r}} = \lambda_1 u^2 + \lambda_2 v^2$.

Orthogonal Triad Computation

Let's compute the semiaxis directions $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ for the ellipsoid $x^2 + 4y^2 + xy + 4z^2 = 16$. To be applied is Theorem 9.3. As explained on page 709, the starting point is to represent the ellipsoid equation as a quadratic form $\vec{\mathbf{W}}^T A\vec{\mathbf{W}} = 16$, where the symmetric matrix A and vector $\vec{\mathbf{W}}$ are defined by

$$A = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \quad \vec{\mathbf{W}} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

College algebra. The **Characteristic Polynomial** $\det(A-\lambda I) = 0$ determines the eigenvalues or hidden values of the matrix A. By cofactor expansion, this polynomial equation is

$$(4-\lambda)((1-\lambda)(4-\lambda) - 1/4) = 0$$

with roots 4, $5/2 + \sqrt{10}/2$, $5/2 - \sqrt{10}/2$.

[11] Rod Serling, author and playwright for the SciFi series *The Twilight Zone*, enjoyed the view from the other side.

Eigenpairs. It will be shown that three eigenpairs are

$$\lambda_1 = 4, \quad \vec{\mathbf{x}}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

$$\lambda_2 = \frac{5+\sqrt{10}}{2}, \quad \vec{\mathbf{x}}_2 = \begin{pmatrix} \sqrt{10}-3 \\ 1 \\ 0 \end{pmatrix},$$

$$\lambda_3 = \frac{5-\sqrt{10}}{2}, \quad \vec{\mathbf{x}}_3 = \begin{pmatrix} \sqrt{10}+3 \\ -1 \\ 0 \end{pmatrix}.$$

The vector norms of the eigenvectors are given by $\|\vec{\mathbf{x}}_1\| = 1$, $\|\vec{\mathbf{x}}_2\| = \sqrt{20+6\sqrt{10}}$, $\|\vec{\mathbf{x}}_3\| = \sqrt{20-6\sqrt{10}}$. The orthonormal semi-axis directions $\vec{\mathbf{v}}_k = \vec{\mathbf{x}}_k/\|\vec{\mathbf{x}}_k\|$, $k = 1, 2, 3$, are then given by the formulas

$$\vec{\mathbf{v}}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{\mathbf{v}}_2 = \begin{pmatrix} \frac{\sqrt{10}-3}{\sqrt{20-6\sqrt{10}}} \\ \frac{1}{\sqrt{20-6\sqrt{10}}} \\ 0 \end{pmatrix}, \quad \vec{\mathbf{v}}_3 = \begin{pmatrix} \frac{\sqrt{10}+3}{\sqrt{20+6\sqrt{10}}} \\ \frac{-1}{\sqrt{20+6\sqrt{10}}} \\ 0 \end{pmatrix}.$$

Eigenpair Details.

$$\left\langle A - \lambda_1 I, \vec{\mathbf{0}} \right\rangle = \left(\begin{array}{ccc|c} 1-4 & 1/2 & 0 & 0 \\ 1/2 & 4-4 & 0 & 0 \\ 0 & 0 & 4-4 & 0 \end{array}\right)$$

$$\approx \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Used Toolkit rules combination, multiply and swap. Found **rref**.

$$\left\langle A - \lambda_2 I, \vec{\mathbf{0}} \right\rangle = \left(\begin{array}{ccc|c} \frac{-3-\sqrt{10}}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{3-\sqrt{10}}{2} & 0 & 0 \\ 0 & 0 & \frac{3-\sqrt{10}}{2} & 0 \end{array}\right)$$

$$\approx \left(\begin{array}{ccc|c} 1 & 3-\sqrt{10} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Toolkit rules applied. Found **rref**.

$$\left\langle A - \lambda_3 I, \vec{\mathbf{0}} \right\rangle = \left(\begin{array}{ccc|c} \frac{-3+\sqrt{10}}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{3+\sqrt{10}}{2} & 0 & 0 \\ 0 & 0 & \frac{3+\sqrt{10}}{2} & 0 \end{array}\right)$$

$$\approx \left(\begin{array}{ccc|c} 1 & 3+\sqrt{10} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Toolkit rules applied. Found **rref**.

Solving the corresponding reduced echelon systems gives the preceding formulas for the eigenvectors $\vec{\mathbf{x}}_1$, $\vec{\mathbf{x}}_2$, $\vec{\mathbf{x}}_3$. The equation for the ellipsoid is $\lambda_1 X^2 + \lambda_2 Y^2 + \lambda_3 Z^2 = 16$, where the multipliers of the square terms are the eigenvalues of A and X, Y, Z define the new coordinate system determined by the eigenvectors of A. This equation can be re-written in the form $\frac{X^2}{a^2} + \frac{Y^2}{b^2} + \frac{Z^2}{c^2} = 1$, provided the semi-axis lengths a, b, c are defined by the relations $a^2 = 16/\lambda_1$, $b^2 = 16/\lambda_2$, $c^2 = 16/\lambda_3$. After computation, $a = 2$, $b = 1.98$, $c = 4.17$.

Proofs, Methods and Details

Eigenpairs of (1), Telephone Carriers:

To be computed are the eigenvalues λ and eigenvectors $\vec{\mathbf{v}}$ for the 3×3 matrix

$$A = \frac{1}{10}\begin{pmatrix} 5 & 4 & 0 \\ 3 & 5 & 3 \\ 2 & 1 & 7 \end{pmatrix}.$$

The eigenpairs are $(1, \vec{\mathbf{v}}_1)$, $\left(\frac{1}{2}, \vec{\mathbf{v}}_2\right)$, $\left(\frac{1}{5}, \vec{\mathbf{v}}_3\right)$ where

$$\text{(14)} \qquad \vec{\mathbf{v}}_1 = \begin{pmatrix} 12 \\ 15 \\ 13 \end{pmatrix}, \quad \vec{\mathbf{v}}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{\mathbf{v}}_3 = \begin{pmatrix} -4 \\ 3 \\ 1 \end{pmatrix}.$$

Eigenvalues. The roots $\lambda = 1, 1/2, 1/5$ of the characteristic equation $\det(A - \lambda I) = 0$ are found by these details:

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \begin{vmatrix} .5-\lambda & .4 & 0 \\ .3 & .5-\lambda & .3 \\ .2 & .1 & .7-\lambda \end{vmatrix} \\ &= \frac{1}{10} - \frac{8}{10}\lambda + \frac{17}{10}\lambda^2 - \lambda^3 && \text{Expand by cofactors.} \\ &= -\frac{1}{10}(\lambda - 1)(2\lambda - 1)(5\lambda - 1) && \text{Factor the cubic.} \end{aligned}$$

The factorization was found by long division of the cubic by $\lambda - 1$, the idea born from the fact that 1 is a root and therefore $\lambda - 1$ is a factor, by the Factor Theorem of college algebra. The root $\lambda = 1$ was discovered from the Rational Root theorem of college algebra.[12]

Eigenpairs. To each eigenvalue $\lambda = 1, 1/2, 1/5$ corresponds one **rref** calculation, to find the eigenvectors paired to λ. The three eigenvectors are given by (2). The details:

Eigenvalue $\lambda = 1$.

$$\begin{aligned} A - (1)I &= \begin{pmatrix} .5-1 & .4 & 0 \\ .3 & .5-1 & .3 \\ .2 & .1 & .7-1 \end{pmatrix} \\ &\approx \begin{pmatrix} -5 & 4 & 0 \\ 3 & -5 & 3 \\ 2 & 1 & -3 \end{pmatrix} && \text{Multiply rule, multiplier=10.} \end{aligned}$$

[12] A rational root x of $a_n x^n + \cdots + a_0 = 0$ is a rational factor of a_0/a_n.

$$\approx \begin{pmatrix} 0 & 0 & 0 \\ 3 & -5 & 3 \\ 2 & 1 & -3 \end{pmatrix} \quad \text{Combination rule twice.}$$

$$\approx \begin{pmatrix} 0 & 0 & 0 \\ 1 & -6 & 6 \\ 2 & 1 & -3 \end{pmatrix} \quad \text{Combination rule.}$$

$$\approx \begin{pmatrix} 0 & 0 & 0 \\ 1 & -6 & 6 \\ 0 & 13 & -15 \end{pmatrix} \quad \text{Combination rule.}$$

$$\approx \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -\frac{12}{13} \\ 0 & 1 & -\frac{15}{13} \end{pmatrix} \quad \text{Multiply rule and combination rule.}$$

$$\approx \begin{pmatrix} 1 & 0 & -\frac{12}{13} \\ 0 & 1 & -\frac{15}{13} \\ 0 & 0 & 0 \end{pmatrix} \quad \text{Swap rule.}$$

$$= \mathbf{rref}(A - (1)I)$$

An equivalent reduced echelon system is $x - 12z/13 = 0$, $y - 15z/13 = 0$. The free variable assignment is $z = t_1$ and then $x = 12t_1/13$, $y = 15t_1/13$.

An eigenvector can be selected as the partial derivative on variable t_1 across the general solution $x = 12t_1/13$, $y = 15t_1/13$, $z = t_1$ (equivalent here to setting $t_1 = 1$). This computation gives eigenvector $x = 12/13, y = 15/13, z = 1$.

An eigenvector can be multiplied by a constant $c \neq 0$ to obtain another eigenvector. To eliminate fractions in the answer, the practice is to multiply by an integer c to eliminate all fractions. Choose constant $c = 13$ to obtain eigenvector $x = 12, y = 15, z = 13$.

Eigenvalue $\lambda = 1/2$.

$$A - (1/2)I = \begin{pmatrix} .5 - .5 & .4 & 0 \\ .3 & .5 - .5 & .3 \\ .2 & .1 & .7 - .5 \end{pmatrix}$$

$$\approx \begin{pmatrix} 0 & 4 & 0 \\ 3 & 0 & 3 \\ 2 & 1 & 2 \end{pmatrix} \quad \text{Multiply rule, factor=10.}$$

$$\approx \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{Combination and multiply rules.}$$

$$= \mathbf{rref}(A - .5I)$$

An eigenvector is found from the equivalent reduced echelon system $y = 0$, $x + z = 0$ to be $x = -1$, $y = 0$, $z = 1$.

Eigenvalue $\lambda = 1/5$.

$$A - (1/5)I = \begin{pmatrix} .5 - .2 & .4 & 0 \\ .3 & .5 - .2 & .3 \\ .2 & .1 & .7 - .2 \end{pmatrix}$$

$$\approx \begin{pmatrix} 3 & 4 & 0 \\ 1 & 1 & 1 \\ 2 & 1 & 5 \end{pmatrix} \quad \text{Multiply rule.}$$

$$\approx \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix} \qquad \text{Combination rule.}$$
$$= \mathbf{rref}(A - (1/5)I)$$

An eigenvector is found from the equivalent reduced echelon system $x+4z = 0$, $y-3z = 0$ to be $x = -4$, $y = 3$, $z = 1$.

An answer check in maple:

```
with(LinearAlgebra):
A:=(1/10)*Matrix([[5,4,0],[3,5,3],[2,1,7]]);
B:=A-lambda*IdentityMatrix(3);
DD,P:=Eigenvectors(A);
factor(Determinant(B));
```

Proof of Theorem 9.12, Stochastic Matrix Properties:

(a) $\sum_{i=1}^n y_i = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_j = \sum_{j=1}^n \left(\sum_{i=1}^n a_{ij}\right) x_j = \sum_{j=1}^n (1)x_j = 1$.

(b) Entry j of $A^T\vec{\mathbf{v}}$ is given by $\sum_{i=1}^n (a_{ij})(1) = \text{column sum} = 1$.

(c) The determinant rule $\det(B^T) = \det(B)$ applied to $B = A - \lambda I$ implies A and A^T have the same eigenvalues. Apply **(b)** to verify that A has eigenvalue 1. Any other root λ of $|A - \lambda I| = 0$ is also a root of $|A^T - \lambda I| = 0$ with corresponding eigenvector $\vec{\mathbf{x}}$ satisfying $A^T\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$. Because $\vec{\mathbf{x}} \neq \vec{\mathbf{0}}$, then $\vec{\mathbf{x}}$ has a component x_j with largest magnitude $|x_j| > 0$. Isolate index j across equation $\lambda\vec{\mathbf{x}} = A^T\vec{\mathbf{x}}$, then divide by $|x_j|$, to obtain $\lambda = \sum_{i=1}^n a_{ij}\frac{x_i}{x_j}$. Because $a_{ji} \geq 0$ and $0 \leq \left|\frac{x_i}{x_j}\right| \leq 1$, then $|\lambda| \leq 1$, because

$$|\lambda| \leq \sum_{i=1}^n a_{ij}\left|\frac{x_i}{x_j}\right| \leq \sum_{i=1}^n (a_{ij})(1) = \text{column sum} = 1.$$

Proof of Theorem 9.13, Perron-Frobenius:[13]
Proof of (a)

Definition 9.6 (Positive Matrix)
Notation $A > 0$ means all $a_{ij} > 0$. Notation $A \leq B$ means $a_{ij} \leq b_{ij}$, also written $B \geq A$.

Definition 9.7 (Max, Min and Ones Matrices)
Matrix $\mathbf{max_r}(A)$ (resp. $\mathbf{min_r}(A)$) is obtained from A by replacing each entry a_{ij} by the maximum (resp. minimum) element of row i. Symbol $\delta = \min_{i,j} a_{ij}$ is the least element in matrix A. Matrix $\mathcal{O}$ is the $n \times n$ matrix of all ones.

The proof is organized as five lemmas. Assume throughout that $A > 0$ is stochastic with least element δ, $B \geq 0$ and $\mathcal{O}$ is the matrix of all ones.

Lemma 1a. If A, B are stochastic, then BA is stochastic.

Lemma 2a. $\mathbf{min_r}(B) \leq \mathbf{min_r}(BA) \leq BA \leq \mathbf{max_r}(BA) \leq \mathbf{max_r}(B)$.

[13] Perron-Frobenius theory is a basis for the Google Search **PageRank algorithm**.

Proof: The maximum along row i of $C = BA$ is some $c_{ij} = \sum_{k=1}^n b_{ik}a_{kj}$. Let M denote the maximum along row i of B. Because columns of A sum to 1, then $c_{ij} = \sum_{k=1}^n b_{ik}a_{kj} \le \sum_{k=1}^n M a_{kj} = M$. Then $BA \le \mathbf{max_r}(BA) \le \mathbf{max_r}(B)$. Details for inequality $\mathbf{min_r}(B) \le \mathbf{min_r}(BA) \le BA$ are similar.

Lemma 3a. $\mathbf{max_r}(BA) - \mathbf{min_r}(BA) \le (1-\delta)\,(\mathbf{max_r}(B) - \mathbf{min_r}(B))$.

Proof: Let $C = BA$ have row i maximum at c_{ij} and row minimum at c_{ik}. Then all elements in row i of matrix $\mathbf{max_r}(BA) - \mathbf{min_r}(BA)$ have value $S = c_{ij} - c_{ik}$. Let M (resp. m) be the common entry along row i of $\mathbf{max_r}(B)$ (resp. $\mathbf{min_r}(B)$). We'll verify $S \le (1-\delta)\,(M-m)$, which proves the lemma.

Re-write $S = c_{ij} - c_{ik} = \sum_{p=1}^n b_{ip}a_{pj} - \sum_{p=1}^n b_{ip}a_{pk} = \sum_{p=1}^n b_{ip}(a_{pj} - a_{pk})$. Let $p_1, \ldots, p_r$ be the set of indices p such that $a_{pj} - a_{pk} > 0$ and let $q_1, \ldots, q_s$ be the set of indices q such that $a_{qj} - a_{qk} < 0$. Indices p that satisfy $a_{pj} - a_{pk} = 0$ contribute zero to S. In cases $r = 0$ and/or $s = 0$ we have $S \le 0$, so the conclusion follows. Henceforth, assume $r \ge 1$ and $s \ge 1$. The column sums of A are 1, which implies for instance $\sum_{\ell=1}^r a_{p_\ell j} + \sum_{\ell=1}^s a_{q_\ell j} = 1$. We estimate:

$$
\begin{aligned}
S &= \textstyle\sum_{p=1}^n b_{ip}(a_{pj} - a_{pk}) \\
&= \textstyle\sum_{\ell=1}^r b_{ip}(a_{p_\ell j} - a_{p_\ell k}) + \sum_{\ell=1}^s b_{ip}(a_{q_\ell j} - a_{q_\ell k}) \\
&\le \textstyle M\sum_{\ell=1}^r (a_{p_\ell j} - a_{p_\ell k}) + m\sum_{\ell=1}^s (a_{q_\ell j} - a_{q_\ell k}) \\
&= \textstyle M\,(1 - \sum_{\ell=1}^s a_{q_\ell j} - 1 + \sum_{\ell=1}^s a_{q_\ell k}) + m\sum_{\ell=1}^s (a_{q_\ell j} - a_{q_\ell k}) \\
&= \textstyle (M-m)\,(-\sum_{\ell=1}^s a_{q_\ell j} + \sum_{\ell=1}^s a_{q_\ell k}) \\
&\le (M-m)\,(-s\delta + 1) \\
&\le (M-m)\,(-\delta + 1)\,.
\end{aligned}
$$

Lemma 4a. $\mathbf{max_r}(A^{k+1}) - \mathbf{min_r}(A^{k+1}) \le (1-\delta)^k \mathcal{O}$.

Proof: Let $B = A^k$ and apply Lemmas 1a and 3a. Then $\mathbf{max_r}(A^{k+1}) - \mathbf{min_r}(A^{k+1}) \le (1-\delta)\big(\mathbf{max_r}(A^k) - \mathbf{min_r}(A^k)\big)$. Induction on k implies the result, because $\mathbf{max_r}(A) - \mathbf{min_r}(A) \le \mathcal{O}$.

Lemma 5a. There exists a vector $\vec{\mathbf{w}}$ with all positive components such that $\lim_{k\to\infty} A^k = \Big\langle \vec{\mathbf{w}}\,|\vec{\mathbf{w}}\,|\cdots|\vec{\mathbf{w}}\Big\rangle$. Then $A\vec{\mathbf{w}} = \vec{\mathbf{w}}$ and $(1, \vec{\mathbf{w}})$ is an eigenpair.[14]

Proof: The preceding lemmas and the calculus squeeze theorem for limits imply that $\mathbf{max_r}(A^k)$ and $\mathbf{min_r}(A^k)$ converge as $k \to \infty$ to some matrix P. Because $\mathbf{max_r}(A^k)$ has identical elements in each row, then so does P. Therefore, the columns of P are the same vector $\vec{\mathbf{w}}$. Take limits across inequality $\mathbf{min_r}(A^k) \ge \delta\mathcal{O}$ to prove $\vec{\mathbf{w}} > \vec{\mathbf{0}}$. Vector $\vec{\mathbf{w}}$ equals $P\vec{\mathbf{u}}$, where $\vec{\mathbf{u}}$ = column 1 of the identity matrix. Then $\vec{\mathbf{w}} = P\vec{\mathbf{u}} = \lim_{k\to\infty} A^{k+1}\vec{\mathbf{u}} = A\left(\lim_{k\to\infty} A^k\vec{\mathbf{u}}\right) = A\vec{\mathbf{w}}$, which is the eigenpair equation $\vec{\mathbf{w}} = A\vec{\mathbf{w}}$.

Proof of (b)

Eigenpair equation $\vec{\mathbf{v}} = A\vec{\mathbf{v}}$ is multiplied repeatedly by A to give $\vec{\mathbf{v}} = A^{k+1}\vec{\mathbf{v}}$. Take the limit using part (a): $\vec{\mathbf{v}} = P\vec{\mathbf{v}}$, where $P = \Big\langle \vec{\mathbf{w}}\,|\vec{\mathbf{w}}\,|\cdots|\vec{\mathbf{w}}\Big\rangle$. Then $\vec{\mathbf{v}} = P\vec{\mathbf{v}} = (\sum_{i=1}^n v_i)\,\vec{\mathbf{w}}$.

Proof of (c)

Consider an eigenpair $(\lambda, \vec{\mathbf{v}})$. Apply A across $\lambda\vec{\mathbf{v}} = A\vec{\mathbf{v}}$ to obtain $\lambda^k\vec{\mathbf{v}} = A^k\vec{\mathbf{v}}$. Use part (a) to take the limit as $k \to \infty$. Then, as in part (b), $\lim_{k\to\infty}\lambda^k\vec{\mathbf{v}} = (\sum_{i=1}^n v_i)\,\vec{\mathbf{w}}$. This limit exists only in case $|\lambda| \le 1$. If $|\lambda| = 1$, then $\lambda = e^{i\theta}$ for some angle θ. The limit fails to exist unless $\theta = 0$ modulo 2π. Therefore, $\lambda = 1$ and $\vec{\mathbf{v}} = (\sum_{i=1}^n v_i)\,\vec{\mathbf{w}}$.

Proof of (d)

Let's suppose some $v_j = 0$, in order to reach a contradiction. Component j of the identity

[14]The numerical **power method** can be used to approximate eigenvector $\vec{\mathbf{w}}$.

$A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$ says that $\sum_{k=1}^n a_{jk}v_k = 0$. Because $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$, then at least one $v_k \neq 0$. Because $a_{jk} > 0$, then $\sum_{k=1}^n a_{jk}v_k > 0$, a contradiction.

Perron-Frobenius proof completed. ■

Exercises 9.2

Discrete Dynamical Systems

Define matrix A via equation

$$\vec{\mathbf{y}} = \frac{1}{10}\begin{pmatrix} 5 & 1 & 0 \\ 3 & 4 & 3 \\ 2 & 5 & 7 \end{pmatrix}\vec{\mathbf{x}} \tag{15}$$

1. Find eigenpair packages of A.
Answers:

$$D = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$P = \begin{pmatrix} -1 & 1 & 1 \\ 0 & -4 & 5 \\ 1 & 3 & 9 \end{pmatrix}$$

2. Explain: A is a **transition matrix.**[15]

3. Assume $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$ has period one year. Find the system state after two years.

4. Explain: $A^n\vec{\mathbf{x}}$ is the system state after n periods.

Market Shares

Define matrix A via equation

$$\vec{\mathbf{y}} = \frac{1}{10}\begin{pmatrix} 5 & 4 & 0 \\ 3 & 5 & 3 \\ 2 & 1 & 7 \end{pmatrix}\vec{\mathbf{x}} \tag{16}$$

5. Find with software the eigenpairs of A given by equation 2.

6. Compute A^2, A^3, A^4 using software. Predict the limit of A^n as n approaches infinity.

7. Compute with software (rounded)

$$A^{10} = \begin{pmatrix} .30 & .30 & .30 \\ .37 & .38 & .37 \\ .32 & .32 & .33 \end{pmatrix} \tag{17}$$

8. Let $\vec{\mathbf{x}} = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Compute

$$A^{10}\vec{\mathbf{x}} = \begin{pmatrix} 0.30 \\ 0.37 \\ 0.33 \end{pmatrix} \text{ (rounded)}$$

in two ways by calculator:
(1) Fourier replacement (3).
(2) Matrix multiply using (17).

Stochastic Matrices

Reference: Perron-Frobenius proof on page 715.

9. Establish the identity $|A - \lambda I| = |A^T - \lambda I|$.

10. Explain why A and A^T have the same eigenvalues but not necessarily the same eigenvectors.

11. Verify $\mathbf{max_r}(A) = \langle \vec{\mathbf{w}}|\vec{\mathbf{w}}|\cdots|\vec{\mathbf{w}}\rangle$, where $\vec{\mathbf{w}}$ has components $w_i = \max\{a_{ij}, 1 \leq j \leq n\}$.

12. Verify $\mathbf{max_r}(A) = D\mathcal{O}$, where D is the diagonal matrix of row maxima and $\mathcal{O}$ is the matrix of all ones.

Perron-Frobenius Theorem

Let $A > 0$ be $n \times n$ stochastic with unique eigenpair $(1, \vec{\mathbf{w}})$, all $w_i > 0$ and $\sum_{i=1}^n w_i = 1$. Assume $\vec{\mathbf{v}} \geq \vec{\mathbf{0}}$, $\sum_{i=1}^n v_i = 1$ and $\delta = \min_{i,j} a_{ij}$.

13. Apply inequality $\mathbf{min_r}(A^n)\vec{\mathbf{v}} \leq A^n\vec{\mathbf{v}} \leq \mathbf{max_r}(A^n)\vec{\mathbf{v}}$ to prove $\lim_{n\to\infty} A^n\vec{\mathbf{v}} = \left(\sum_{i=1}^n v_i\right)\vec{\mathbf{w}} = \vec{\mathbf{w}}$.

14. Verify Euclidean norm inequality
$\|A^{k+1}\vec{\mathbf{v}} - \vec{\mathbf{w}}\| \leq \sqrt{n}\,(1-\delta)^k$

Weierstrass Proof

These exercises establish existence of an eigenpair $(1, \vec{\mathbf{v}})$ for stochastic matrix A having only nonnegative entries.

[15] Perron-Frobenius theory extensions in the literature apply to transition matrices. See the Weierstrass Proof exercises.

Weierstrass Compactness Theorem
A sequence of vectors $\{\vec{\mathbf{v}}_i\}_{i=1}^{\infty}$ contained in a closed, bounded set K in $\mathcal{R}^n$ has a subsequence converging in the vector norm of $\mathcal{R}^n$ to some vector $\vec{\mathbf{v}}$ in K.

Define set K to be all vectors $\vec{\mathbf{v}}$ with non-negative components adding to 1. Let $\vec{\mathbf{v}}_0$ be any element of K. Assume stochastic A with $a_{ij} \ge 0$ and define $\vec{\mathbf{v}}_N = \frac{1}{N}\sum_{j=0}^{N-1} A^j\vec{\mathbf{v}}_0$.

15. Verify K is closed and bounded in $\mathcal{R}^n$. Then prove $\lambda\vec{\mathbf{x}} + (1-\lambda)\vec{\mathbf{y}}$ is in K for $0 \le \lambda \le 1$ and $\vec{\mathbf{x}}, \vec{\mathbf{y}}$ in K.

16. Prove identity
$\vec{\mathbf{v}}_{N+1} = \lambda\vec{\mathbf{v}}_N + (1-\lambda)A^N\vec{\mathbf{v}}_0$
where $\lambda = \frac{N}{N+1}$ and then prove by induction that $\vec{\mathbf{v}}_N$ is in K.

17. Verify all hypotheses in the Weierstrass theorem applied to $\{\vec{\mathbf{v}}_N\}_{N=0}^{\infty}$. Applying the theorem produces a subsequence $\{\vec{\mathbf{v}}_{N_p}\}_{p=1}^{\infty}$ limiting to some $\vec{\mathbf{v}}$ in K.

18. Verify identity
$\vec{\mathbf{v}}_N - A\vec{\mathbf{v}}_N = \frac{1}{N}(\vec{\mathbf{v}}_0 - A^N\vec{\mathbf{v}}_0)$.

19. Explain why $A\vec{\mathbf{v}} = \lim_{p\to\infty} A\vec{\mathbf{v}}_{N_p}$. Then prove $\vec{\mathbf{v}} = A\vec{\mathbf{v}}$.

20. The claimed eigenpair $(1, \vec{\mathbf{v}})$ has been found, provided $\vec{\mathbf{v}} \ne \vec{\mathbf{0}}$. Explain why $\vec{\mathbf{v}} \ne \vec{\mathbf{0}}$.

Coupled Systems

Find the coefficient matrix A. Identify as coupled or uncoupled and explain why.

21. $x' = 2x + 3y$, $y' = x + y$

22. $x' = 3y$, $y' = x$

23. $x' = 3x$, $y' = 2y$

24. $x' = 3x$, $y' = 2y$, $z' = z$

Solving Uncoupled Systems

Solve for the general solution.

25. $x' = 3x$, $y' = 2y$

26. $x' = 3x$, $y' = 2y$, $z' = z$

Change of Coordinates

Given the change of coordinates $\vec{\mathbf{y}} = A\vec{\mathbf{x}}$, find the matrix B for the inverse change $\vec{\mathbf{x}} = B\vec{\mathbf{y}}$.

27. $\vec{\mathbf{y}} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}\vec{\mathbf{x}}$

28. $\vec{\mathbf{y}} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\vec{\mathbf{x}}$

Constructing Coupled Systems

Given the uncoupled system and change of coordinates $\vec{\mathbf{y}} = P\vec{\mathbf{x}}$, find the coupled system.

29. $x_1' = 2x_1$, $x_2' = 3x_2$, $P = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}$

30. $x_1' = x_1$, $x_2' = -x_2$, $P = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$

Uncoupling a System

Change the given coupled system into an uncoupled system using the eigenanalysis change of variables $\vec{\mathbf{y}} = P\vec{\mathbf{x}}$.

31. $x_1' = 2x_1$, $x_2' = x_1 + x_2$, $x_3' = x_3$
Ans: $P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, $y_1' = 2y_1$, $y_2' = y_2$, $y_3' = y_3$

32. $x_1' = x_1 + x_2$, $x_2' = x_1 + x_2$, $x_3' = x_3$
Ans: $P = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $y_1' = 0$, $y_2' = 2y_2$, $y_3' = y_3$

Solving Coupled Systems

Report the answers for $x(t), y(t)$.

33. $x' = -x - 2y$, $y' = -4x + y$

34. $x' = 8x - y$, $y' = -2x + 7y$

Eigenanalysis and Footballs

The exercises study the ellipsoid
$17x^2 + 8y^2 - 12xy + 80z^2 = 80$.

35. Let $A = \begin{pmatrix} 17 & -6 & 0 \\ -6 & 8 & 0 \\ 0 & 0 & 80 \end{pmatrix}$. Expand equation $\vec{\mathbf{W}}^T A\vec{\mathbf{W}} = 80$, where $\vec{\mathbf{W}}$ has components x, y, z.

36. Find the eigenpairs of
$A = \begin{pmatrix} 17 & -6 & 0 \\ -6 & 8 & 0 \\ 0 & 0 & 80 \end{pmatrix}$.

37. Verify the semi-axis lengths $4, 2, 1$.

38. Verify that the ellipsoid has semi-axis unit directions
$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$

The Ellipse and Eigenanalysis

The exercises study the ellipse
$2x^2 + 4xy + 5y^2 = 24$.

39. Let $A = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$. Expand equation $\vec{\mathbf{W}}^T A \vec{\mathbf{W}} = 24$, where $\vec{\mathbf{W}} = \begin{pmatrix} x \\ y \end{pmatrix}$.

40. Find the eigenpairs of $A = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$.

41. Verify the semi-axis lengths $2, 2\sqrt{6}$.

42. Verify that the ellipse has semi-axis unit directions
$\frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

Orthogonal Triad Computation

The exercises fill in details from page 711. The ellipsoid equation: $x^2+4y^2+xy+4z^2=16$ or $\vec{\mathbf{x}}^T A \vec{\mathbf{x}}=16$,
$A = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$

43. Find the characteristic equation of A. Then verify the roots are 4, $5/2 + \sqrt{10}/2$, $5/2 - \sqrt{10}/2$.

44. Show the steps from **rref** to second eigenvector $\vec{\mathbf{x}}_2$:
$\mathbf{rref} = \begin{pmatrix} 1 & 3-\sqrt{10} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$,
$\vec{\mathbf{x}}_2 = \begin{pmatrix} \sqrt{10}-3 \\ 1 \\ 0 \end{pmatrix}$

9.3 Advanced Topics in Linear Algebra

Diagonalization and Jordan's Theorem

A system of differential equations $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ can be transformed to an uncoupled system $\vec{\mathbf{y}}' = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)\vec{\mathbf{y}}$ by a change of variables $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ if and only if A is **diagonalizable** and P is an invertible matrix of independent eigenvectors of A from eigenpairs $(\lambda_k, \vec{\mathbf{v}}_k)$, $1 \le k \le n$.

If A fails to be diagonalizable, then eigenanalysis does not help. Jordan's theorem 9.14 is a possible generator of a change of coordinates $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$. System $\vec{\mathbf{y}}' = T\vec{\mathbf{y}}$ is not uncoupled, but triangular: the linear integrating factor method applies to solve the triangular system, details forthcoming.

The sad truth about Jordan's theorem: matrix P has no algorithm for construction. The matrix P used as replacement is a matrix of **generalized eigenvectors** constructed from an algorithm for the Jordan normal form page **??**. See page **??** for a `maple` example.

Theoretical existence of P for a change of variables may be enough for proofs. Computation requires a formula for P. What has emerged historically are mathematical algorithms to solve system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ independent of both Jordan's theorem and the Jordan normal form page **??**. The foundation for computer algebra algorithms and low dimensional hand algorithms is the Cayley-Hamilton theorem 9.16 on page 721.

Theorem 9.14 (Jordan's theorem)
Any $n \times n$ matrix A can be represented in the form

$$A = PTP^{-1}$$

where P is invertible and T is upper triangular. The diagonal entries of T are eigenvalues of A.

Proof on page 740.

Theorem 9.15 (Jordan's Extension)
Any $n \times n$ matrix A can be represented in the block triangular form

$$A = PTP^{-1}, \quad T = \mathbf{diag}(T_1, \ldots, T_k),$$

where P is invertible and each matrix T_i is upper triangular with diagonal entries equal to a single eigenvalue of A.

Remarks. An induction proof of the theorem can be based upon Jordan's Theorem 9.14. No proof is supplied. The theorem is presented again in Proposition 9.15 page 720 as a special case of the Jordan decomposition $A = PJP^{-1}$, in which J is the Jordan Form of $n \times n$ matrix A. Because the Jordan form is a triangular matrix, then $T = J$ gives an algorithm for generation of columns in matrix P. Jordan form is largely used in proofs and theoretical investigations and rarely in computation.
Computer algebra systems can find matrices J and P in the Jordan form of matrix A. With limitations, there is a constructible matrix P for Jordan's two theorems 9.14 and 9.15. See page **??** for a `maple` example.

Cayley-Hamilton Identity

A celebrated and deep result for powers of matrices was discovered by Cayley and Hamilton (see Birkhoff–MacLane [**?**]), which says that an $n \times n$ matrix A satisfies its own characteristic equation. More precisely:

Theorem 9.16 (Cayley-Hamilton)
Let $\det(A - \lambda I)$ be expanded as the nth degree polynomial

$$p(\lambda) = \sum_{j=0}^{n} c_j \lambda^j,$$

for some coefficients $c_0, \ldots, c_{n-1}$ and $c_n = (-1)^n$. Then A satisfies the equation $p(\lambda) = 0$, that is,

$$p(A) \equiv \sum_{j=0}^{n} c_j A^j = \mathbf{0}.$$

In factored form in terms of the eigenvalues $\{\lambda_j\}_{j=1}^{n}$ (duplicates possible), the matrix equation $p(A) = \mathbf{0}$ can be written as

$$(-1)^n (A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_n I) = \mathbf{0}.$$

Proof on page 741.

Solving Block Triangular Differential Systems

A matrix differential system $\vec{\mathbf{y}}\,'(t) = T\vec{\mathbf{y}}\,(t)$ with T block upper triangular splits into scalar equations which can be solved by elementary methods for first order scalar differential equations. To illustrate, consider the system

$$\begin{aligned} y_1' &= 3y_1 + x_2 + y_3, \\ y_2' &= 3y_2 + y_3, \\ y_3' &= 2y_3. \end{aligned}$$

The techniques that apply are the growth-decay formula for $u' = ku$ and the integrating factor method for $u' = ku + p(t)$. Working backwards from the last equation with back-substitution gives

$$\begin{array}{rcl} y_3 &=& c_3 e^{2t}, \\ y_2 &=& c_2 e^{3t} - c_3 e^{2t}, \\ y_1 &=& (c_1 + c_2 t) e^{3t}. \end{array}$$

What has been said here applies to any triangular system $\vec{\mathbf{y}}'(t) = T\vec{\mathbf{y}}(t)$, in order to write an exact formula for the solution $\vec{\mathbf{y}}(t)$.

If A is an $n \times n$ matrix, then Jordan's theorem gives $A = PTP^{-1}$ with T block upper triangular and P invertible. The change of variable $\vec{\mathbf{x}}(t) = P\vec{\mathbf{y}}(t)$ changes $\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t)$ into the block triangular system $\vec{\mathbf{y}}'(t) = T\vec{\mathbf{y}}(t)$.

There is no special condition on A, to effect the change of variable $\vec{\mathbf{x}}(t) = P\vec{\mathbf{y}}(t)$. The solution $\vec{\mathbf{x}}(t)$ of $\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t)$ is a product of the invertible matrix P and a column vector $\vec{\mathbf{y}}(t)$; the latter is the solution of the block triangular system $\vec{\mathbf{y}}'(t) = T\vec{\mathbf{y}}(t)$, obtained by growth-decay and integrating factor methods.

The *importance of this idea* is to provide a theoretical method for solving any system $\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t)$.

Matrices P and T in Jordan's extension $A = PTP^{-1}$ can be found using computer algebra systems. See page **??** for a `maple` example in which T is the Jordan normal form of A and P is the matrix of generalized eigenvectors.

Symmetric Matrices and Orthogonality

A **symmetric matrix** A is defined by the identity $A^T = A$. In applications the symmetric matrix A might be obtained as $A = B^T B$ for some non-square matrix B. Studied here is the eigenanalysis of symmetric matrices, which reproduces $AP = PD$ from classical eigenanalysis with a difference: the eigenvectors in columns of P are of **unit length**, meaning $\|\vec{\mathbf{x}}\| = 1$, and also **orthogonal**, meaning dot product zero or 90 degrees apart. See Chapter 5 Section 1.

Definition 9.8 (Unitize)
A vector $\vec{\mathbf{x}}$ is said to be **unitized** into vector $\vec{\mathbf{y}}$ if $\vec{\mathbf{y}} = c\vec{\mathbf{x}}$ for some scalar c and $\|\vec{\mathbf{y}}\| = 1$.

An eigenpair $(\lambda, \vec{\mathbf{x}})$ of A can always be selected so that $\|\vec{\mathbf{x}}\| = 1$: replace eigenvector $\vec{\mathbf{x}}$ by $\frac{1}{\|\vec{\mathbf{x}}\|}\vec{\mathbf{x}}$.

Theorem 9.17 (Orthogonality of Eigenvectors)
Assume that $n \times n$ matrix A is **symmetric**, $A^T = A$. If $(\alpha, \vec{\mathbf{x}})$ and $(\beta, \vec{\mathbf{y}})$ are eigenpairs of A with $\alpha \neq \beta$, then $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$ are orthogonal: $\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = 0$. Proof on page 741.

Theorem 9.18 (Real Eigenvalues)
If $A^T = A$, then all eigenvalues of A are real. Consequently, matrix A has n real eigenvalues counted according to multiplicity. Proof on page 741.

Proposition 9.1 (Independence of Orthogonal Sets) Let $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_k$ be a set of nonzero orthogonal vectors. Then this set is independent.

Duplicated by the orthogonal vector test Chapter 5 Section 3, Theorem 5.33.

The Gram-Schmidt process

The eigenvectors of a symmetric matrix A may be constructed to be orthogonal. First of all, observe that eigenvectors corresponding to distinct eigenvalues are orthogonal by Theorem 9.17. It remains to construct from k independent eigenvectors $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_k$, corresponding to a single eigenvalue λ, another set of independent eigenvectors $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_k$ for λ which are pairwise orthogonal. The idea, due to Gram-Schmidt, applies to any set of k independent vectors.

Theorem 9.19 (Gram-Schmidt)
Let $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_k$ be independent n-vectors. The set of vectors $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_k$ constructed below as linear combinations of $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_k$ are pairwise orthogonal, independent and $\mathbf{span}(\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_k) = \mathbf{span}(\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_k)$.

$$
\begin{aligned}
\vec{\mathbf{y}}_1 &= \vec{\mathbf{x}}_1 \\
\vec{\mathbf{y}}_2 &= \vec{\mathbf{x}}_2 - \frac{\vec{\mathbf{x}}_2 \cdot \vec{\mathbf{y}}_1}{\vec{\mathbf{y}}_1 \cdot \vec{\mathbf{y}}_1}\vec{\mathbf{y}}_1 \\
\vec{\mathbf{y}}_3 &= \vec{\mathbf{x}}_3 - \frac{\vec{\mathbf{x}}_3 \cdot \vec{\mathbf{y}}_1}{\vec{\mathbf{y}}_1 \cdot \vec{\mathbf{y}}_1}\vec{\mathbf{y}}_1 - \frac{\vec{\mathbf{x}}_3 \cdot \vec{\mathbf{y}}_2}{\vec{\mathbf{y}}_2 \cdot \vec{\mathbf{y}}_2}\vec{\mathbf{y}}_2 \\
&\vdots \\
\vec{\mathbf{y}}_k &= \vec{\mathbf{x}}_k - \frac{\vec{\mathbf{x}}_k \cdot \vec{\mathbf{y}}_1}{\vec{\mathbf{y}}_1 \cdot \vec{\mathbf{y}}_1}\vec{\mathbf{y}}_1 - \cdots - \frac{\vec{\mathbf{x}}_k \cdot \vec{\mathbf{y}}_{k-1}}{\vec{\mathbf{y}}_{k-1} \cdot \vec{\mathbf{y}}_{k-1}}\vec{\mathbf{y}}_{k-1}
\end{aligned}
$$

Proof on page 742.

Example 9.14 (Gram-Schmidt on Four Eigenvectors)

Let $(-1, \vec{\mathbf{v}}_1)$, $(2, \vec{\mathbf{v}}_2)$, $(2, \vec{\mathbf{v}}_3)$, $(2, \vec{\mathbf{v}}_4)$ be eigenpairs of a 4×4 symmetric matrix A. Apply the Gram-Schmidt process to find 4 pairwise orthogonal eigenvectors of A.

Solution: Because eigenvector $\vec{\mathbf{v}}_1$ is for eigenvalue 1 and the others are for eigenvalue 2, then Theorem 9.17 implies that $\vec{\mathbf{v}}_1$ is orthogonal to $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$, $\vec{\mathbf{v}}_4$. Eigenvectors $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$, $\vec{\mathbf{v}}_4$ belong to eigenvalue $\lambda = 2$, but they are not assumed orthogonal. The Gram-Schmidt process applied to eigenvectors $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$, $\vec{\mathbf{v}}_4$ finds pairwise orthogonal vectors $\vec{\mathbf{y}}_2$, $\vec{\mathbf{y}}_3$, $\vec{\mathbf{y}}_4$ that are linear combinations of eigenvectors $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$, $\vec{\mathbf{v}}_4$. Then $\vec{\mathbf{y}}_2$, $\vec{\mathbf{y}}_3$, $\vec{\mathbf{y}}_4$ are also eigenvectors for $\lambda = 2$. The four eigenvectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{y}}_2$, $\vec{\mathbf{y}}_3$, $\vec{\mathbf{y}}_4$ are pairwise orthogonal, as required.

Orthogonal Projection

Reproduced here for reference is the basic material on shadow projection. The ideas are then extended to obtain the orthogonal projection onto a subspace V of $\mathcal{R}^n$. Finally, the orthogonal projection formula will be related to the Gram-Schmidt equations.

The **shadow projection** of vector $\vec{X}$ onto the direction of vector $\vec{Y}$ is the number d defined by

$$d = \frac{\vec{X}\cdot\vec{Y}}{|\vec{Y}|}.$$

The triangle determined by $\vec{X}$ and $d\dfrac{\vec{Y}}{|\vec{Y}|}$ is a right triangle.

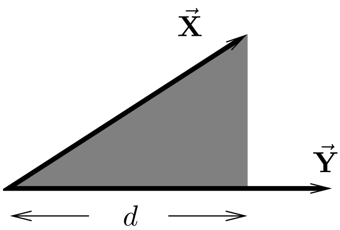


Figure 4. Shadow projection d of vector $\vec{\mathbf{X}}$ onto the direction of vector $\vec{\mathbf{Y}}$.

The **vector shadow projection** of $\vec{X}$ onto the line L through the origin in the direction of $\vec{Y}$ is the vector representing the shadow, direction $\vec{Y}$ and length d, defined by

$$\mathbf{proj}_{\vec{Y}}(\vec{X}) = d\frac{\vec{Y}}{|\vec{Y}|} = \frac{\vec{X}\cdot\vec{Y}}{\vec{Y}\cdot\vec{Y}}\vec{Y}.$$

Definition 9.9 (One-Dimensional Orthogonal Projection)
Let V be the line through the origin in the direction of nonzero vector $\vec{\mathbf{Y}}$. Then $V = \mathbf{span}\{\vec{\mathbf{Y}}\}$. Define the **orthogonal projection**:

$$\mathbf{Proj}_V(\vec{\mathbf{x}}) = (\vec{\mathbf{u}}\cdot\vec{\mathbf{x}})\vec{\mathbf{u}}, \quad \vec{\mathbf{u}} = \vec{\mathbf{Y}}/\|\vec{\mathbf{Y}}\|$$

Is Definition 9.9 the same as vector shadow projection? Yes. Does the definition depend on $\vec{\mathbf{Y}}$? No, because of Theorem 9.20 below.

Definition 9.10 (Orthogonal Projection onto a Subspace)
Let subspace V of $\mathcal{R}^n$ be spanned by orthonormal vectors $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_k$. Define the **orthogonal projection** of vector $\vec{\mathbf{x}}$ in $\mathcal{R}^n$ onto subspace V by the formula (justified in Theorem 9.20):

$$(1) \qquad \begin{aligned}\mathbf{Proj}_V(\vec{\mathbf{x}}) &= \textstyle\sum_{j=1}^{k}(\vec{\mathbf{u}}_j\cdot\vec{\mathbf{x}})\vec{\mathbf{u}}_j, \\ &= \textstyle\sum_{j=1}^{k} \text{vector shadow projection } \vec{\mathbf{x}} \text{ onto } \vec{\mathbf{u}}_j\end{aligned}$$

Theorem 9.20 (Formula $\mathbf{Proj}_V(\vec{\mathbf{x}})$ is Well-Defined)
Orthogonal projection formula $\mathbf{Proj}_V(\vec{\mathbf{x}}) = \sum_{j=1}^{k}(\vec{\mathbf{u}}_j \cdot \vec{\mathbf{x}})\vec{\mathbf{u}}_j$ is independent of the choice of orthonormal vectors $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_k$ that span V.

Proof on page 742

Important: Formula $\mathbf{Proj}_V(\vec{\mathbf{x}}) = \sum_{j=1}^{k}(\vec{\mathbf{u}}_j \cdot \vec{\mathbf{x}})\vec{\mathbf{u}}_j$ requires a basis which is **orthonormal**. An **orthogonal** basis suffices with the shadow projection summation in (1). Applications might use either formula.

Orthogonal Projection and Gram-Schmidt. Define $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_k$ by the Gram-Schmidt relations on page 723. Define

$$\vec{\mathbf{u}}_j = \vec{\mathbf{y}}_j / \|\vec{\mathbf{y}}_j\|$$

for $j = 1, \ldots, k$. Then $V_{j-1} = \mathbf{span}\{\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_{j-1}\}$ is a subspace of $\mathcal{R}^n$ of dimension $j-1$ with orthonormal basis $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_{j-1}$ and

$$\begin{aligned}
\vec{\mathbf{y}}_j &= \vec{\mathbf{x}}_j - \left(\frac{\vec{\mathbf{x}}_j \cdot \vec{\mathbf{y}}_1}{\vec{\mathbf{y}}_1 \cdot \vec{\mathbf{y}}_1}\vec{\mathbf{y}}_1 + \cdots + \frac{\vec{\mathbf{x}}_k \cdot \vec{\mathbf{y}}_{j-1}}{\vec{\mathbf{y}}_{j-1} \cdot \vec{\mathbf{y}}_{j-1}}\vec{\mathbf{y}}_{j-1}\right) \\
&= \vec{\mathbf{x}}_j - \mathbf{Proj}_{V_{j-1}}(\vec{\mathbf{x}}_j)
\end{aligned}$$

The Gram-Schmidt relations are memorized by the formula

$$\vec{\mathbf{y}}_j = \vec{\mathbf{x}}_j - \sum_{k<j}(\text{vector shadow projection of } \vec{\mathbf{x}}_j \text{ onto } \vec{\mathbf{y}}_k)$$

Near Point Theorem

Developed here is the characterization of the orthogonal projection of a vector $\vec{\mathbf{x}}$ onto a subspace V as the unique point $\vec{\mathbf{v}}$ in V which minimizes $\|\vec{\mathbf{x}} - \vec{\mathbf{v}}\|$, that is, the point in V which is nearest to $\vec{\mathbf{x}}$.

Theorem 9.21 (Orthogonal Projection Properties)
Let subspace V be the span of orthonormal vectors $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_k$.

(a) Each vector $\vec{\mathbf{v}}$ in V has an orthogonal expansion $\vec{\mathbf{v}} = \sum_{j=1}^{k}(\vec{\mathbf{u}}_j \cdot \vec{\mathbf{v}})\vec{\mathbf{u}}_j$.

(b) The orthogonal projection $\mathbf{Proj}_V(\vec{\mathbf{x}})$ is a vector in V.

(c) Vector $\vec{\mathbf{w}} = \vec{\mathbf{x}} - \mathbf{Proj}_V(\vec{\mathbf{x}})$ is orthogonal to every vector in V.

(d) Among all vectors $\vec{\mathbf{v}}$ in V, the minimum value of $\|\vec{\mathbf{x}} - \vec{\mathbf{v}}\|$ is uniquely obtained by the orthogonal projection $\vec{\mathbf{v}} = \mathbf{Proj}_V(\vec{\mathbf{x}})$.

(e) Let $n \times k$ matrix A have independent columns that span V. If vector $\vec{\mathbf{w}}$ is orthogonal to every vector in V, then $A^T\vec{\mathbf{w}} = \vec{\mathbf{0}}$.

Proof on page 743.

Theorem 9.22 (Near Point to a Subspace)
Let V be a subspace of $\mathcal{R}^n$ and $\vec{\mathbf{x}}$ a vector not in V. The **near point** to $\vec{\mathbf{x}}$ in V is the orthogonal projection of $\vec{\mathbf{x}}$ onto V. This point is characterized as the minimum of $\|\vec{\mathbf{x}} - \vec{\mathbf{v}}\|$ over all vectors $\vec{\mathbf{v}}$ in the subspace V.

Proof by part **(d)** of Theorem 9.21.

Theorem 9.23 (Cross Product and Projections)
The cross product $\vec{\mathbf{a}} \times \vec{\mathbf{b}}$ is a constant multiple of $\vec{\mathbf{c}} - \mathbf{Proj}_V(\vec{\mathbf{c}})$, where vector $\vec{\mathbf{c}}$ is not in $V = \mathbf{span}\{\vec{\mathbf{a}}, \vec{\mathbf{b}}\}$.

Proof: The cross product makes sense only in $\mathcal{R}^3$. Subspace V is two dimensional when $\vec{\mathbf{a}}$, $\vec{\mathbf{b}}$ are independent, and Gram-Schmidt applies to find an orthonormal basis $\vec{\mathbf{u}}_1$, $\vec{\mathbf{u}}_2$. By (c) of Theorem 9.21, the vector $\vec{\mathbf{c}} - \mathbf{Proj}_V(\vec{\mathbf{c}})$ has the same or opposite direction to the cross product. ■

Linear Least Squares

A primary application of linear least squares is fitting of large data sets to an equation. Desired is a simple equation which can be used to interpolate or extrapolate missing data items or to find trends in the data.

Example 9.15 (Height-Weight Data)
Verify that slope $m = 61.27$ and intercept $b = -39, 05$ best fit equation $y = mx + b$ to the 15 data items in Table 4, where x=height, y=weight. Graphic in Figure 5.

The solution is on page 744.

Table 4. Height-Weight Data for 15 women ages $30 - 39$ years.
Source: The World Almanac and Book of Facts, 1975.

Height (m)	1.47	1.50	1.52	1.55	1.57	1.60	1.63	1.65
Weight (kg)	52.21	53.12	54.48	55.84	57.20	58.57	59.93	61.29
Height (m)	1.68	1.70	1.73	1.75	1.78	1.80	1.83	
Weight (kg)	63.11	64.47	66.28	68.10	69.92	72.19	74.46	

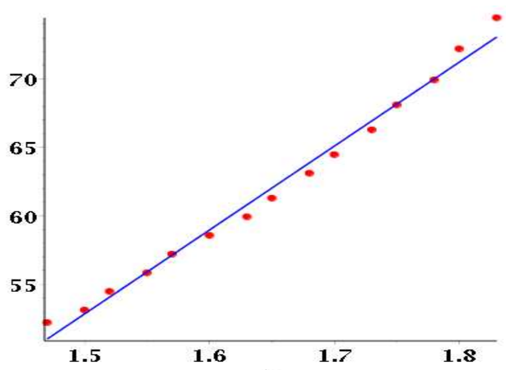


Figure 5. Best fit
The least squares fit straight line in blue $y = 61.27x - 39.06$.
Red dots are the 15 data points from Table 4.

Least Squares Normal Equation

Let $m \times n$ matrix A and vector $\vec{\mathbf{b}}$ be given. Assume hereafter that the problem $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ has *no solution*. The discussion will be guided by the unsolvable system

$$\begin{cases} x + y = 1, \\ x + y = 0. \end{cases} \tag{2}$$

System (2) is the special case $b_1^* = 1$, $b_2^* = 0$ for system

$$\begin{cases} x + y = b_1^*, \\ x + y = b_2^*. \end{cases} \tag{3}$$

Least squares chooses two values b_1^*, b_2^* such that system (3) is solvable for x, y. The choice requires that substitution of x, y into the original unsolvable equation (2) gives the **least error**, in some well-defined sense.

Mathematically, the least error for a trial solution x, y in (5) might be realized [16] by choosing x, y to minimize the vector norm $\|\vec{\mathbf{E}}\|$ for error vector

$$\vec{\mathbf{E}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{4}$$

Minimization leads to the geometry problem solved in Figure 6. By geometry, points on the line $x + y = \frac{1}{2}$ are half way between the two lines of the original system. Point $x = \frac{1}{2}$, $y = 0$ is isolated as a proper candidate for a *best solution* to original unsolvable problem (2).

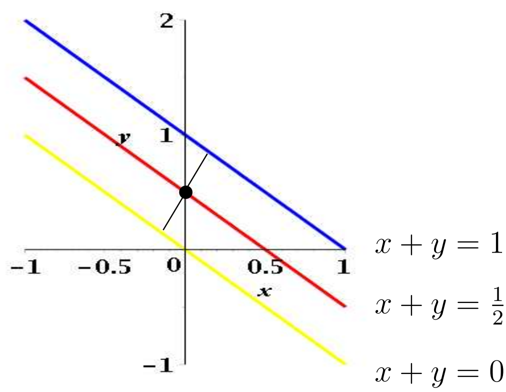


Figure 6. Black dot $x^* = \frac{1}{2}$, $y^* = 0$ is one best solution to system $x + y = 1$, $x + y = 0$. Any point along the red line $x + y = \frac{1}{2}$ makes minimum vector norm error between the two lines $x + y = 1$, $x + y = 0$.

Warning: The isolated point $x = \frac{1}{2}$, $y = 0$ does not actually work in the original equations! The geometrical solution invents one possible solvable replacement

[16] The expression to minimize is controversial: at the very least, it depends on the intended application.

system (3) with best fit to the original unsolvable equations (2):

$$\begin{cases} x + y = \frac{1}{2}, \\ x + y = \frac{1}{2}. \end{cases} \tag{5}$$

System (5) has a name:

Definition 9.11 (Normal Equation for Linear Least Squares)
The **normal equation** for unsolvable problem $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ is the solvable system

$$A^T A\vec{\mathbf{x}} = A^T\vec{\mathbf{b}} \tag{6}$$

It is **not** implied that a solution $\vec{\mathbf{x}}$ of (6) is also a solution of $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$: the original equation is assumed to have *no solution*.

System (3) has matrix form $A\vec{\mathbf{x}} = \vec{\mathbf{b}}^*$. If vector $\vec{\mathbf{x}}$ solves $A\vec{\mathbf{x}} = \vec{\mathbf{b}}^*$, then $\vec{\mathbf{b}}^*$ equals $A\vec{\mathbf{x}}$, which means $\vec{\mathbf{b}}^*$ is a linear combination of the columns of A, or $\vec{\mathbf{b}}^*$ belongs to subspace $S = \mathbf{colspace}(A)$. Overloaded symbol $\vec{\mathbf{x}}$ is not the same as in equation $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$: the latter has no solution.

Geometrically, $\vec{\mathbf{b}}^*$ is a specific given vector in S and equation $A\vec{\mathbf{x}} = \vec{\mathbf{b}}^*$ can have infinitely many solutions $\vec{\mathbf{x}}$, or just one. Important: the no solution case has been eliminated from the *three possibilities*.

Error minimization seeks a *best solution* $\vec{\mathbf{x}} = \vec{\mathbf{x}}^*$ to the unsolvable problem $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$. Applied literature suggests to find $\vec{\mathbf{x}} = \vec{\mathbf{x}}^*$ as a minimizer for a function which measures the error between $A\vec{\mathbf{x}}$ and $\vec{\mathbf{b}}$.

Proposition 9.2 $A\vec{\mathbf{x}} = \vec{\mathbf{b}}^*$ has a solution $\vec{\mathbf{x}}$ if and only if $\vec{\mathbf{b}}^*$ belongs to subspace $S = \mathbf{colspace}(A)$.

Proposition 9.3 Let $\vec{\mathbf{x}} = \vec{\mathbf{x}}^*$ achieve the minimum for vector norm $\|A\vec{\mathbf{x}} - \vec{\mathbf{b}}\|$, taken over all $\vec{\mathbf{x}}$ in $\mathcal{R}^n$. Then $\vec{\mathbf{x}} = \vec{\mathbf{x}}^*$ is a *best possible* solution of unsolvable equation $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$, because it minimizes the vector norm error $\|A\vec{\mathbf{x}} - \vec{\mathbf{b}}\|$ over all possible $\vec{\mathbf{x}}$.

Theorem 9.24 (Least Squares Solution of Unsolvable $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$)
Let $\vec{\mathbf{x}}^*$ satisfy

$$\|A\vec{\mathbf{x}}^* - \vec{\mathbf{b}}\| = \min_{\vec{\mathbf{x}}} \|A\vec{\mathbf{x}} - \vec{\mathbf{b}}\|$$

Then $\vec{\mathbf{x}} = \vec{\mathbf{x}}^*$ is a solution of **Normal Equation** $A^T A\vec{\mathbf{x}} = A^T\vec{\mathbf{b}}$. Vector $\vec{\mathbf{x}} = \vec{\mathbf{x}}^*$ is a *best possible solution* for unsolvable equation $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$.
Proof on page 743.

Data Fitting

Assume given experimentally measured values $y_1, y_2, \ldots, y_m$ taken at independent variable values $x_1, x_2, \ldots, x_m$. Data fitting invents a model equation of the form[17]

$$y = \sum_{j=1}^{n} c_j f_j(x).$$

The invented functions f_j will have additional requirements, for example they could be polynomials $1, x, x^2, \ldots$ or trigonometric functions, e.g., a model motivated by truncation of Taylor series or Fourier series. The problem: find values for the constants $c_1, \ldots, c_n$.

Ideally, the model equation fits the data exactly. What actually holds is an exact equation with error terms $E_1, \ldots, E_m$:

$$y_i = \sum_{j=1}^{n} c_j f_j(x_i) + E_j$$

Linear least squares minimizes the sum of squares of the errors:

$$\min \sum_{j=1}^{m} |E_j|^2 \quad \text{over all choices of } c_1, \ldots, c_n$$

Minimization is assumed to return special values $c_1^*, \ldots, c_n^*$ giving the **best fit**. The predicted model for the data set is then:

$$y = \sum_{j=1}^{n} c_j^* f_j(x).$$

The QR Decomposition

Matrix multiply can express Gram-Schmidt formulas as $A = QR$, where A has independent columns $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n$ and the columns of Q are the unitized Gram-Schmidt orthonormal vectors $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_n$.

Definition 9.12 (Orthogonal Matrix)
A matrix Q having pairwise orthogonal columns of unit length is called **orthogonal**. Alternatively, $Q^TQ = I$. If Q is square, then $QQ^T = I$.[18]

[17] Statistical experiments might use vector variables. For instance a 3-vector $\vec{\mathbf{x}}$ with components of sex, age and height replaces scalar variable x. Scalars y_j could be vectors. Symbol c_j is replaced by symbol β_j, these parameters called **regressors**.

[18] Non-square matrices with *orthonormal columns* certainly exist. A warning: terminology **orthonormal matrix** usually means the matrix A is **square** and has orthonormal columns: $A^TA = AA^T = I$.

Theorem 9.25 (The QR-Decomposition)
Let the $m \times n$ matrix A have independent columns $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n$. Then there is an upper triangular matrix R with positive diagonal entries and an orthogonal matrix Q such that

$$A = QR.$$

Proof: Let $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_n$ be the Gram-Schmidt orthogonal vectors given by relations on page 723. Define $\vec{\mathbf{u}}_k = \vec{\mathbf{y}}_k/\|\vec{\mathbf{y}}_k\|$ and $r_{kk} = \|\vec{\mathbf{y}}_k\|$ for $k = 1, \ldots, n$, and otherwise $r_{ij} = \vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}}_j$. Let $Q = \left\langle \vec{\mathbf{u}}_1 | \cdots | \vec{\mathbf{u}}_n \right\rangle$. Then

$$(7)\qquad \begin{array}{rcl} \vec{\mathbf{x}}_1 &=& r_{11}\vec{\mathbf{u}}_1, \\ \vec{\mathbf{x}}_2 &=& r_{22}\vec{\mathbf{u}}_2 + r_{21}\vec{\mathbf{u}}_1, \\ \vec{\mathbf{x}}_3 &=& r_{33}\vec{\mathbf{u}}_3 + r_{31}\vec{\mathbf{u}}_1 + r_{32}\vec{\mathbf{u}}_2, \\ &\vdots& \\ \vec{\mathbf{x}}_n &=& r_{nn}\vec{\mathbf{u}}_n + r_{n1}\vec{\mathbf{u}}_1 + \cdots + r_{nn-1}\vec{\mathbf{u}}_{n-1}. \end{array}$$

It follows from (7) and matrix multiplication that $A = QR$. The columns of Q have unit length and they are pairwise orthogonal: Q is orthogonal. ■

Theorem 9.26 (Matrices Q and R in $A = QR$)
Let $m \times n$ matrix A have independent columns $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n$. Let $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_n$ be the Gram-Schmidt orthogonal vectors from page 723. Define $\vec{\mathbf{u}}_k = \vec{\mathbf{y}}_k/\|\vec{\mathbf{y}}_k\|$. Then $AQ = QR$ is satisfied by $Q = \left\langle \vec{\mathbf{u}}_1 | \cdots | \vec{\mathbf{u}}_n \right\rangle$ and

$$R = \begin{pmatrix} \|y_1\| & \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{x}}_2 & \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{x}}_3 & \cdots & \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{x}}_n \\ 0 & \|y_2\| & \vec{\mathbf{u}}_2 \cdot \vec{\mathbf{x}}_3 & \cdots & \vec{\mathbf{u}}_2 \cdot \vec{\mathbf{x}}_n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \|y_n\| \end{pmatrix}.$$

Proof: Details are contained in the proof of Theorem 9.25 above. ■

Some references cite the diagonal entries as $\|\vec{\mathbf{x}}_1\|, \|\vec{\mathbf{x}}_2^{\perp}\|, \ldots, \|\vec{\mathbf{x}}_n^{\perp}\|$, where $\vec{\mathbf{x}}_j^{\perp} = \vec{\mathbf{x}}_j - \mathbf{Proj}_{V_{j-1}}(\vec{\mathbf{x}}_j)$, $V_{j-1} = \mathbf{span}\{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_{j-1}\}$. Because $\vec{\mathbf{y}}_1 = \vec{\mathbf{x}}_1$ and $\vec{\mathbf{y}}_j = \vec{\mathbf{x}}_j - \mathbf{Proj}_{V_{j-1}}(\vec{\mathbf{x}}_j)$, the formulas for the entries of R are identical.

Theorem 9.27 (Uniqueness of Q and R)
Let $m \times n$ matrix A have independent columns and satisfy the decomposition $A = QR$. If Q is $m \times n$ orthogonal and R is $n \times n$ upper triangular with positive diagonal elements, then Q and R are uniquely determined.

Proof: The problem is to show that $A = Q_1R_1 = Q_2R_2$ implies $R_2R_1^{-1} = I$ and $Q_1 = Q_2$. We start with $Q_1 = Q_2R_2R_1^{-1}$. Define $P = R_2R_1^{-1}$. Then $Q_1 = Q_2P$. Because $I = Q_1^TQ_1 = P^TQ_2^TQ_2P = P^TP$, then P is orthogonal. Matrix P is the product of square upper triangular matrices with positive diagonal elements, which implies P itself is square upper triangular with positive diagonal elements. The only orthogonal matrix with these properties is the identity matrix I. Then $R_2R_1^{-1} = P = I$, which implies $R_1 = R_2$ and $Q_1 = Q_2$. ■

Theorem 9.28 (The QR Decomposition and Least Squares)
Let $m{\times}n$ matrix A have independent columns and satisfy the decomposition $A = QR$ with Q orthogonal and R invertible. Then the normal equation

$$A^T A\vec{\mathbf{x}} = A^T\vec{\mathbf{b}}$$

in the theory of least squares can be represented as

$$R\vec{\mathbf{x}} = Q^T\vec{\mathbf{b}}.$$

Proof: Because Q is orthogonal, then $Q^TQ = I$. Let's use the identity $(CD)^T = D^TC^T$, the equation $A = QR$, and assumed R^T invertible to obtain

$A^TA\vec{\mathbf{x}} = A^T\vec{\mathbf{b}}$	Normal equation
$R^TQ^TQR\vec{\mathbf{x}} = R^TQ^T\vec{\mathbf{x}}$	Substitute $A = QR$.
$R\vec{\mathbf{x}} = Q^T\vec{\mathbf{x}}$	Multiply by the inverse of R^T.

■

The formula $R\vec{\mathbf{x}} = Q^T\vec{\mathbf{b}}$ can be solved by back-substitution, which accounts for its popularity in numerical solution of least squares problems.

Theorem 9.29 (Spectral Theorem)
Let A be a given $n \times n$ real matrix. Then $A = QDQ^{-1}$ with Q orthogonal and D diagonal if and only if $A^T = A$.

Proof: Requirement Q *is orthogonal* means that the columns of Q are **orthonormal** and $n \times n$. The equation $A = A^T$ means A is **symmetric**.

Assume first that $A = QDQ^{-1}$ with $Q = Q^T$ orthogonal ($Q^TQ = I$) and D diagonal. Then $Q^T = Q = Q^{-1}$. This implies $A^T = (QDQ^{-1})^T = (Q^{-1})^TD^TQ^T = QDQ^{-1} = A$.

Conversely, assume $A^T = A$. Then the eigenvalues of A are real and eigenvectors corresponding to distinct eigenvalues are orthogonal. The proof proceeds by induction on the dimension n of the $n \times n$ matrix A.

For $n = 1$, let Q be the 1×1 identity matrix. Then Q is orthogonal and $AQ = QD$ where D is 1×1 diagonal.

Assume the decomposition $AQ = QD$ for dimension n. Let's prove it for A of dimension $n + 1$. Choose a real eigenvalue λ of A and eigenvector $\vec{\mathbf{v}}_1$ with $\|\vec{\mathbf{v}}_1\| = 1$. Complete a basis $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_{n+1}$ of $\mathcal{R}^{n+1}$. By Gram-Schmidt, we assume as well that this basis is orthonormal. Define $P = \left\langle \vec{\mathbf{v}}_1 | \cdots | \vec{\mathbf{v}}_{n+1} \right\rangle$. Then P is square, orthogonal and satisfies $P^T = P^{-1}$. Define $B = P^{-1}AP$. Then B is symmetric ($B^T = B$) and $\mathbf{col}(B, 1) = \lambda\, \mathbf{col}(I, 1)$. These facts imply that B is a block matrix

$$B = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & C \end{array}\right)$$

where C is symmetric ($C^T = C$). The induction hypothesis applies to C to obtain the existence of an orthogonal matrix Q_1 such that $CQ_1 = Q_1D_1$ for some diagonal matrix

D_1. Define block diagonal matrix D, block matrix W and square matrix Q as follows:

$$D = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & D_1 \end{array}\right),$$
$$W = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1 \end{array}\right),$$
$$Q = PW.$$

Then Q is the product of two orthogonal matrices, which makes Q orthogonal. Compute

$$W^{-1}BW = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1^{-1} \end{array}\right)\left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & C \end{array}\right)\left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1 \end{array}\right) = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & D_1 \end{array}\right).$$

Then $Q^{-1}AQ = W^{-1}P^{-1}APW = W^{-1}BW = D$. This completes the induction, ending the proof of the theorem. ■

> **Spectral Theorem Consequence**: The eigenpair equation $AP = PD$ with $A \neq A^T$ (A not symmetric) cannot be converted to $AQ = QD$ with Q orthogonal.

Theorem 9.30 (Schur's Theorem)
Given any real $n \times n$ matrix A, possibly non-symmetric, there is an upper triangular matrix T, whose diagonal entries are the eigenvalues of A, and a complex matrix Q satisfying $\overline{Q}^T = Q^{-1}$ (Q is unitary), such that

$$AQ = QT.$$

If $A = A^T$, then Q is real orthogonal ($Q^T = Q$).

Schur's theorem can be proved by induction, following the induction proof of Jordan's theorem, or the induction proof of the Spectral Theorem. The result can be used to prove the Spectral Theorem in two steps. Indeed, Schur's Theorem implies Q is real, T equals its transpose, and T is triangular. Then T must equal a diagonal matrix D.

Theorem 9.31 (Eigenpairs of a Symmetric A)
Let A be a symmetric $n \times n$ real matrix. Then A has n eigenpairs $(\lambda_1, \vec{\mathbf{v}}_1)$, ..., $(\lambda_n, \vec{\mathbf{v}}_n)$, with independent eigenvectors $\vec{\mathbf{v}}_1$, ..., $\vec{\mathbf{v}}_n$.

Proof: Apply the Spectral Theorem 9.29, page 731, to prove the existence of an orthogonal matrix Q and a diagonal matrix D such that $AQ = QD$. The diagonal entries of D are the eigenvalues of A, in some order. For a diagonal entry λ of D appearing in row j, the relation $A\,\mathbf{col}(Q, j) = \lambda\,\mathbf{col}(Q, j)$ holds, which implies that A has n eigenpairs. The eigenvectors are the columns of Q, which are orthogonal and hence independent. ■

Theorem 9.32 (Diagonalization of Symmetric A)
Let A be a symmetric $n \times n$ real matrix. Then A has n eigenpairs $(\lambda_i, \vec{\mathbf{x}}_i)$. Assume the eigenvalues are listed with duplicates grouped together. For each distinct eigenvalue

λ, replace its eigenvectors by orthonormal eigenvectors, using the Gram-Schmidt process. Let $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_n$ be the orthonormal vectors so obtained and define

$$Q = \left\langle \vec{\mathbf{u}}_1 | \cdots | \vec{\mathbf{u}}_n \right\rangle \quad D = \mathbf{diag}(\lambda_1, \ldots, \lambda_n).$$

Then Q is an orthogonal matrix and $AQ = QD$.

Proof: Theorem 9.31 justifies the eigenanalysis result. Already, eigenpairs corresponding to distinct eigenvalues are orthogonal. Within the set of eigenpairs with the same eigenvalue λ, the Gram-Schmidt process produces a replacement basis of orthonormal eigenvectors. Then the union of all the eigenvectors is orthonormal. The process described here does not disturb the ordering of eigenpairs, because it only replaces an eigenvector. ■

The Singular Value Decomposition

Coined the **SVD** in literature, the singular value decomposition $A = U\Sigma V^T$ has some interesting algebraic properties and it conveys important geometrical and theoretical insights about linear transformations.

Data science uses the SVD as a compression algorithm. Machine vision uses the SVD to find the nearest orthogonal matrix to A. Linear regression modeling uses the SVD to find the pseudo-inverse. Signal processing noise reduction and image processing size reduction use the SVD. Latent semantic indexing in natural-language text processing uses the SVD to identify patterns in unstructured text. Geometric interpretations of the SVD appear in a later subsection.

Theorem 9.33 (Positive Eigenvalues of A^TA)
Given an $m \times n$ real matrix A, then A^TA is a real symmetric matrix whose eigenpairs $(\lambda, \vec{\mathbf{v}})$ satisfy[19]

$$\lambda = \frac{\|A\vec{\mathbf{v}}\|^2}{\|\vec{\mathbf{v}}\|^2} \geq 0. \tag{8}$$

Proof: Symmetry follows from $(A^TA)^T = A^T(A^T)^T = A^TA$. An eigenpair $(\lambda, \vec{\mathbf{v}})$ satisfies $\lambda\vec{\mathbf{v}}^T\vec{\mathbf{v}} = \vec{\mathbf{v}}^TA^TA\vec{\mathbf{v}} = (A\vec{\mathbf{v}})^T(A\vec{\mathbf{v}}) = \|A\vec{\mathbf{v}}\|^2$, hence (8). ■

Definition 9.13 (Singular Values of A)
Let the real symmetric matrix A^TA have real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$. The numbers

$$\sigma_k = \sqrt{\lambda_k}, \quad 1 \leq k \leq n,$$

are called the **singular values** of the matrix A. The ordering of the singular values is always with decreasing magnitude.

[19] Can a real symmetric matrix have negative or complex eigenvalues?
The answer is **NO**.

Theorem 9.34 (Orthonormal Set $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_m$)
Let the real symmetric matrix A^TA have real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$ and corresponding orthonormal eigenvectors $\vec{\mathbf{v}}_1,\ldots,\vec{\mathbf{v}}_n$, obtained by the Gram-Schmidt process. Define the vectors

$$\vec{\mathbf{u}}_1 = \frac{1}{\sigma_1} A\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{u}}_r = \frac{1}{\sigma_r} A\vec{\mathbf{v}}_r.$$

Because $\|A\vec{\mathbf{v}}_k\| = \sigma_k$, then $\{\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_r\}$ is orthonormal. Gram-Schmidt can extend this set to an orthonormal basis $\{\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_m\}$ of $\mathcal{R}^m$.

Proof of Theorem 9.34: Compute $\|\vec{\mathbf{u}}_k\|^2 = \vec{\mathbf{v}}_k \cdot (A^TA\vec{\mathbf{v}}_k)/\lambda_k = \|\vec{\mathbf{v}}_k\|^2 = 1$, because $\{\vec{\mathbf{v}}_k\}_{k=1}^n$ is an orthonormal set. Then the vectors $\vec{\mathbf{u}}_k$ are nonzero. Given $i \neq j$, then $\sigma_i\sigma_j\vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_j = (A\vec{\mathbf{v}}_i)^T(A\vec{\mathbf{v}}_j) = \lambda_j\vec{\mathbf{v}}_i^T\vec{\mathbf{v}}_j = 0$, showing that the vectors $\vec{\mathbf{u}}_k$ are orthogonal.

The extension of the $\vec{\mathbf{u}}_k$ to an orthonormal basis of $\mathcal{R}^m$ is not unique, because it depends upon a choice of independent spanning vectors $\vec{\mathbf{y}}_{r+1}, \ldots, \vec{\mathbf{y}}_m$ for the set $\{\vec{\mathbf{x}} \;:\; \vec{\mathbf{x}} \cdot \vec{\mathbf{u}}_k = 0, \quad 1 \leq k \leq r\}$. Once selected, Gram-Schmidt is applied to $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_r, \vec{\mathbf{y}}_{r+1}, \ldots, \vec{\mathbf{y}}_m$ to obtain the desired orthonormal basis.

Computer algebra systems can compute the orthonormal basis $\{\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_m\}$ of $\mathcal{R}^m$ by appending all columns of the identity matrix to columns $\{\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_r\}$ to define an augmented matrix Z. Then the reduced row echelon form of Z identifies the pivot columns of Z. The first r pivot columns are $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_r$. The remaining pivot columns are columns of the identity. Apply Gram-Schmidt to the pivot columns to obtain the orthonormal basis $\{\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_m\}$. ■

Theorem 9.35 (The Singular Value Decomposition (svd))
Let A be a given real $m \times n$ matrix. Let $(\lambda_1, \vec{\mathbf{v}}_1),\ldots,(\lambda_n, \vec{\mathbf{v}}_n)$ be a set of orthonormal eigenpairs for A^TA such that $\sigma_k = \sqrt{\lambda_k}$ $(1 \leq k \leq r)$ defines the positive singular values of A and $\lambda_k = 0$ for $r < k \leq n$. Complete $\vec{\mathbf{u}}_1 = (1/\sigma_1)A\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{u}}_r = (1/\sigma_r)A\vec{\mathbf{v}}_r$ to an orthonormal basis $\{\vec{\mathbf{u}}_k\}_{k=1}^m$for $\mathcal{R}^m$. Define

$$U = \left\langle \vec{\mathbf{u}}_1 | \cdots | \vec{\mathbf{u}}_m \right\rangle, \quad \Sigma = \left(\begin{array}{c|c} \mathbf{diag}(\sigma_1, \ldots, \sigma_r) & 0 \\ \hline 0 & 0 \end{array} \right),$$
$$V = \left\langle \vec{\mathbf{v}}_1 | \cdots | \vec{\mathbf{v}}_n \right\rangle.$$

Then the columns of U and V are orthonormal and

$$\begin{aligned} A &= U\Sigma V^T \\ &= \sigma_1\vec{\mathbf{u}}_1\vec{\mathbf{v}}_1^T + \cdots + \sigma_r\vec{\mathbf{u}}_r\vec{\mathbf{v}}_r^T \\ &= A(\vec{\mathbf{v}}_1)\vec{\mathbf{v}}_1^T + \cdots + A(\vec{\mathbf{v}}_r)\vec{\mathbf{v}}_r^T \end{aligned}$$

Proof of Theorem 9.35: The product of U and Σ is the $m \times n$ matrix

$$\begin{aligned} U\Sigma &= \left\langle \sigma_1\vec{\mathbf{u}}_1 | \cdots | \sigma_r\vec{\mathbf{u}}_r | \vec{\mathbf{0}} | \cdots | \vec{\mathbf{0}} \right\rangle \\ &= \left\langle A(\vec{\mathbf{v}}_1) | \cdots | A(\vec{\mathbf{v}}_r) | \vec{\mathbf{0}} | \cdots | \vec{\mathbf{0}} \right\rangle. \end{aligned}$$

Let $\vec{\mathbf{v}}$ be any vector in $\mathcal{R}^n$. It will be shown that $U\Sigma V^T\vec{\mathbf{v}}$, $\sum_{k=1}^r A(\vec{\mathbf{v}}_k)(\vec{\mathbf{v}}_k^T\vec{\mathbf{v}})$ and $A\vec{\mathbf{v}}$ are the same column vector. We have the equalities

$$\begin{aligned}
U\Sigma V^T\vec{\mathbf{v}} &= U\Sigma\begin{pmatrix}\vec{\mathbf{v}}_1^T\vec{\mathbf{v}}\\ \vdots\\ \vec{\mathbf{v}}_n^T\vec{\mathbf{v}}\end{pmatrix}\\
&= \left\langle A(\vec{\mathbf{v}}_1)|\cdots|A(\vec{\mathbf{v}}_r)|\vec{\mathbf{0}}|\cdots|\vec{\mathbf{0}}\right\rangle\begin{pmatrix}\vec{\mathbf{v}}_1^T\vec{\mathbf{v}}\\ \vdots\\ \vec{\mathbf{v}}_n^T\vec{\mathbf{v}}\end{pmatrix}\\
&= \sum_{k=1}^r(\vec{\mathbf{v}}_k^T\vec{\mathbf{v}})A(\vec{\mathbf{v}}_k).
\end{aligned}$$

Because $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_n$ is an orthonormal basis of $\mathcal{R}^n$, then $\vec{\mathbf{v}} = \sum_{k=1}^n(\vec{\mathbf{v}}_k^T\vec{\mathbf{v}})\vec{\mathbf{v}}_k$. Additionally, $A(\vec{\mathbf{v}}_k) = \vec{\mathbf{0}}$ for $r < k \le n$ implies

$$\begin{aligned}
A\vec{\mathbf{v}} &= A\left(\sum_{k=1}^n(\vec{\mathbf{v}}_k^T\vec{\mathbf{v}})\vec{\mathbf{v}}_k\right)\\
&= \sum_{k=1}^r(\vec{\mathbf{v}}_k^T\vec{\mathbf{v}})A(\vec{\mathbf{v}}_k)
\end{aligned}$$

Then $A\vec{\mathbf{v}} = U\Sigma V^T\vec{\mathbf{v}} = \sum_{k=1}^r A(\vec{\mathbf{v}}_k)(\vec{\mathbf{v}}_k^T\vec{\mathbf{v}})$, which proves the theorem. ■

Singular Values and Geometry

Discussed here is how to interpret singular values geometrically, especially in low dimensions 2 and 3. Conics will be reviewed, adopting the viewpoint of eigenanalysis.

Standard Equation of an Ellipse

Calculus courses consider ellipse equations like

$$85x^2 - 60xy + 40y^2 = 2500$$

and discuss removal of the cross term $-60xy$. The objective is to obtain a **standard** ellipse equation

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1.$$

We re-visit this old problem from a different point of view, and in the derivation establish a connection between the ellipse equation, the symmetric matrix A^TA, and the singular values of A.

Example 9.16 (Image of the Unit Circle)
Let $A = \begin{pmatrix}-2 & 6\\ 6 & 7\end{pmatrix}$.
Verify that the invertible matrix A maps the unit circle into the ellipse

$$85x^2 - 60xy + 40y^2 = 2500.$$

Solution: The unit circle has parameterization $\theta \to (\cos\theta, \sin\theta)$, $0 \le \theta \le 2\pi$.
The unit circle is mapped by matrix A via the set of dual relations

$$\begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = A^{-1} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The Pythagorean identity $\cos^2\theta + \sin^2\theta = 1$ used on the vector norm of second relation implies

$$85x^2 - 60xy + 40y^2 = 2500.$$

Example 9.17 (Removing the xy-Term in an Ellipse Equation)
After a rotation $(x, y) \to (X, Y)$ to remove the xy-term in

$$85x^2 - 60xy + 40y^2 = 2500,$$

verify that the ellipse equation in the new XY-coordinates is

$$\frac{X^2}{25} + \frac{Y^2}{100} = 1.$$

Solution: The xy-term removal is accomplished by a change of variables $(x, y) \to (X, Y)$ which transforms the ellipse equation $85x^2 - 60xy + 40y^2 = 2500$ into the ellipse equation $100X^2 + 25Y^2 = 2500$, details below. It's standard form is obtained by dividing by 2500, to give

$$\frac{X^2}{25} + \frac{Y^2}{100} = 1.$$

Analytic geometry says that the semi-axis lengths are $\sqrt{25} = 5$ and $\sqrt{100} = 10$.
In previous discussions of the ellipse, the equation $85x^2 - 60xy + 40y^2 = 2500$ was represented by the vector-matrix identity

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 85 & -30 \\ -30 & 40 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2500.$$

The program used earlier to remove the xy-term was to diagonalize the coefficient matrix $B = \begin{pmatrix} 85 & -30 \\ -30 & 40 \end{pmatrix}$ by calculating the eigenpairs of B:

$$\left(100, \begin{pmatrix} -2 \\ 1 \end{pmatrix}\right), \quad \left(25, \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right).$$

Because B is symmetric, then the eigenvectors are orthogonal. The eigenpairs above are replaced by unitized pairs:

$$\left(100, \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}\right), \quad \left(25, \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right).$$

Then the diagonalization theory for B can be written as

$$BQ = QD, \quad Q = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 & 1 \\ 1 & 2 \end{pmatrix}, \quad D = \begin{pmatrix} 100 & 0 \\ 0 & 25 \end{pmatrix}.$$

The single change of variables

$$\begin{pmatrix} x \\ y \end{pmatrix} = Q \begin{pmatrix} X \\ Y \end{pmatrix}$$

then transforms the ellipse equation $85x^2 - 60xy + 40y^2 = 2500$ into $100X^2 + 25Y^2 = 2500$ as follows:

$85x^2 - 60xy + 40y^2 = 2500$	Ellipse equation.
$\vec{\mathbf{u}}^T B \vec{\mathbf{u}} = 2500$	Where $B = \begin{pmatrix} 85 & -30 \\ -30 & 40 \end{pmatrix}$ and $\vec{\mathbf{u}} = \begin{pmatrix} x \\ y \end{pmatrix}$.
$(Q\vec{\mathbf{w}})^T B (Q\vec{\mathbf{w}}) = 2500$	Change $\vec{\mathbf{u}} = Q\vec{\mathbf{w}}$, where $\vec{\mathbf{w}} = \begin{pmatrix} X \\ Y \end{pmatrix}$.
$\vec{\mathbf{w}}^T (Q^T BQ)\vec{\mathbf{w}} = 2500$	Expand, ready to use $BQ = QD$.
$\vec{\mathbf{w}}^T (D\vec{\mathbf{w}}) = 2500$	Because $D = Q^{-1}BQ$ and $Q^{-1} = Q^T$.
$100X^2 + 25Y^2 = 2500$	Expand $\vec{\mathbf{w}}^T D\vec{\mathbf{w}}$.

Rotations, Reflections and Scaling

The 2×2 singular value decomposition $A = U\Sigma V^T$ can be used to decompose the change of variables $(x, y) \to (X, Y)$ into three distinct changes of variables, each with a geometrical meaning:

$$(x, y) \longrightarrow (x_1, y_1) \longrightarrow (x_2, y_2) \longrightarrow (X, Y).$$

Table 5. Three Changes of Variable

Domain	Equation	Image	Meaning
Circle 1	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = V^T \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$	Circle 2	Proper Rotation
Circle 2	$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \Sigma \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$	Ellipse 1	Scale axes
Ellipse 1	$\begin{pmatrix} X \\ Y \end{pmatrix} = U \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$	Ellipse 2	Improper Rotation

Proper Rotation. Matrix $R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ satisfies $R^T R = I$ and $|R| = 1$, called a **proper rotation**. The rotation is clockwise about the origin, following use in computer graphics. Replace θ by $-\theta$ for a counterclockwise rotation about the origin.

Improper Rotation. Matrix $R = \begin{pmatrix} 0.936 & 0.352 \\ 0.352 & -0.936 \end{pmatrix}$ is orthogonal with $|R| = -1$, called an **improper rotation**. It represents a reflection, which inverts orientation. Reference:

https://en.wikipedia.org/wiki/Rotation_matrix

Geometry

Figure 7 provides a geometrical interpretation for the singular value decomposition

$$A = U\Sigma V^T.$$

For illustration, the matrix A is assumed 2×2 and invertible.

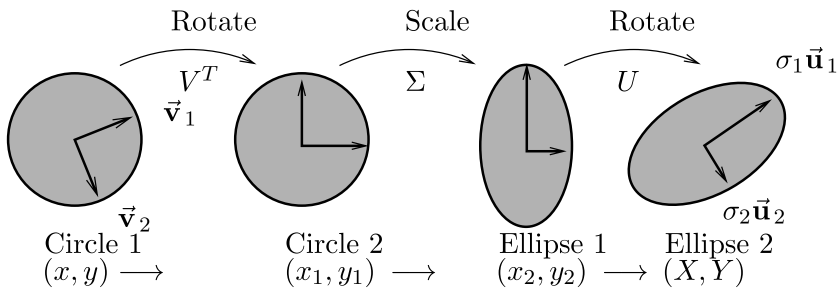


Figure 7. Mapping the unit circle.

- Invertible matrix A maps Circle 1 into Ellipse 2.
- Orthonormal vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ are mapped by matrix $A = U\Sigma V^T$ into orthogonal vectors $A\vec{\mathbf{v}}_1 = \sigma_1\vec{\mathbf{u}}_1$, $A\vec{\mathbf{v}}_2 = \sigma_2\vec{\mathbf{u}}_2$, which are exactly the semi-axes vectors of Ellipse 2.
- The semi-axis lengths of Ellipse 2 equal the singular values σ_1, σ_2 of matrix A.
- The semi-axis directions of Ellipse 2 are equal to the basis vectors $\vec{\mathbf{u}}_1$, $\vec{\mathbf{u}}_2$.
- The process is a rotation $(x, y) \to (x_1, y_1)$, followed by an axis-scaling $(x_1, y_1) \to (x_2, y_2)$, followed by $(x_2, y_2) \to (X, Y)$, a rotation.

Example 9.18 (Mapping and the SVD)
The singular value decomposition $A = U\Sigma V^T$ for $A = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}$ is given by

$$U = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}, \quad V = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}.$$

- Invertible matrix $A = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}$ maps the unit circle into an ellipse.
- The columns of V are orthonormal vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, computed as eigenpairs $(\lambda_1, \vec{\mathbf{v}}_1)$, $(\lambda_2, \vec{\mathbf{v}}_2)$ of A^TA, ordered by $\lambda_1 \geq \lambda_2$.

$$\left(100, \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right), \quad \left(25, \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right).$$

- The singular values are $\sigma_1 = \sqrt{\lambda_1} = 10$, $\sigma_2 = \sqrt{\lambda_2} = 5$.
- The image of $\vec{\mathbf{v}}_1$ is $A\vec{\mathbf{v}}_1 = U\Sigma V^T \vec{\mathbf{v}}_1 = U \begin{pmatrix} \sigma_1 \\ 0 \end{pmatrix} = \sigma_1 \vec{\mathbf{u}}_1$.
- The image of $\vec{\mathbf{v}}_2$ is $A\vec{\mathbf{v}}_2 = U\Sigma V^T \vec{\mathbf{v}}_2 = U \begin{pmatrix} 0 \\ \sigma_2 \end{pmatrix} = \sigma_2 \vec{\mathbf{u}}_2$.

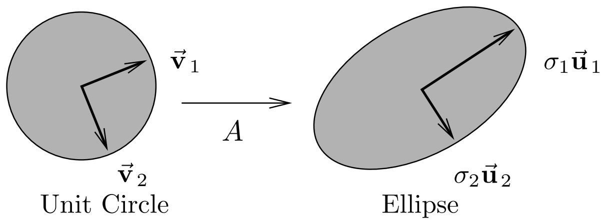


Figure 8.
Mapping the unit circle into an ellipse.

The Four Fundamental Subspaces

The subspaces appearing in the Fundamental Theorem of Linear Algebra are called the **Four Fundamental Subspaces**. They are:

Subspace	**Notation**
Row Space of A	**Image** (A^T)
Nullspace of A	**kernel**(A)
Column Space of A	**Image**(A)
Nullspace of A^T	**kernel** (A^T)

The singular value decomposition $A = U\Sigma V^T$ computes orthonormal bases for the row and column spaces of of A. In the table below, symbol $r = \mathbf{rank}(A)$. Matrix A is assumed $m \times n$, which implies A maps $\mathcal{R}^n$ into $\mathcal{R}^m$.

Table 6. Four Fundamental Subspaces and the SVD

Orthonormal Basis	**Subspace**	**Name**
First r columns of U $(m \times n)$	**Image**(A)	Column Space of A
Last $n - r$ columns of U	**kernel** (A^T)	Nullspace of A^T
First r columns of V $(n \times m)$	**Image** (A^T)	Row Space of A
Last $m - r$ columns of V	**kernel**(A)	Nullspace of A

Table 7. Fundamental Subspaces by Columns of U and V

$m \times n \quad A = U\Sigma V^T \quad$ Singular Value Decomposition

$$m \times m \quad U = \left\| \begin{array}{c|c} \textbf{colspace}(A) & \textbf{nullspace}(A^T) \\ r \text{ columns} & m-r \text{ columns} \end{array} \right\|$$

$$m \times n \quad \Sigma = \left\| \begin{array}{c|c} \begin{pmatrix} \sigma_1 & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & \sigma_r \end{pmatrix} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right\|$$

$$n \times n \quad V = \left\| \begin{array}{c|c} \textbf{rowspace}(A) & \textbf{nullspace}(A) \\ r \text{ columns} & n-r \text{ columns} \end{array} \right\|$$

A Change of Basis Interpretation of the SVD

The singular value decomposition can be described as follows:

For every $m \times n$ matrix A of rank r, orthonormal bases

$$\{\vec{\mathbf{v}}_i\}_{i=1}^n \text{ and } \{\vec{\mathbf{u}}_j\}_{j=1}^m$$

can be constructed such that

- Matrix A maps basis vectors $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_r$ to nonnegative multiples of basis vectors $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_r$, respectively.
- The $n-r$ left-over basis vectors $\vec{\mathbf{v}}_{r+1}, \ldots \vec{\mathbf{v}}_n$ map by A into the zero vector.
- With respect to these bases, matrix A is represented by a real diagonal matrix Σ with non-negative entries.

Proofs, Methods and Details

Proof of Theorem 9.14, Jordan's Theorem:
Proceed by induction on the dimension n of A. For $n = 1$ there is nothing to prove. Assume the result for dimension n. Assume A is $(n+1)\times(n+1)$. To prove the induction step, choose an eigenpair $(\lambda_1, \vec{\mathbf{v}}_1)$ of A with $\vec{\mathbf{v}}_1 \neq \vec{\mathbf{0}}$. Complete a basis $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_{n+1}$ for

$\mathcal{R}^{n+1}$ and define $V = \left\langle \vec{\mathbf{v}}_1 | \cdots | \vec{\mathbf{v}}_{n+1} \right\rangle$. Then $V^{-1}AV = \left(\begin{array}{c|c} \lambda_1 & B \\ \hline \vec{\mathbf{0}} & A_1 \end{array}\right)$ for some matrices B and A_1. The induction hypothesis implies there is an invertible $n \times n$ matrix P_1 and an upper triangular matrix T_1 such that $A_1 = P_1 T_1 P_1^{-1}$. Let $R = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & P_1 \end{array}\right)$ and $T = \left(\begin{array}{c|c} \lambda_1 & BT_1 \\ \hline 0 & T_1 \end{array}\right)$. Then T is upper triangular and $(V^{-1}AV)R = RT$, which implies $A = PTP^{-1}$ for $P = VR$. The induction is complete. ■

Proof of Theorem 9.16, Cayley-Hamilton:
An algebraic proof was given in Chapter 5 Section 3. It depended on the adjugate identity $\mathbf{adj}(A)\,A = A\,\mathbf{adj}(A) = |A|I$. Below is a different proof which suggests how the theorem might have been discovered.

If A is diagonalizable, $AP = P\,\mathbf{diag}(\lambda_1, \ldots, \lambda_n)$, then the proof is obtained from the expansion

$$A^j = P\,\mathbf{diag}(\lambda_1^j, \ldots, \lambda_n^j)P^{-1},$$

because summing across this identity leads to

$$\begin{array}{rcl} p(A) & = & \sum_{j=0}^n c_j A^j \\ & = & P\left(\sum_{j=0}^n c_j\,\mathbf{diag}(\lambda_1^j, \ldots, \lambda_n^j)\right)P^{-1} \\ & = & P\,\mathbf{diag}(p(\lambda_1), \ldots, p(\lambda_n))P^{-1} \\ & = & P\,\mathbf{diag}(0, \ldots, 0)P^{-1} \\ & = & \mathbf{0}. \end{array}$$

If A is not diagonalizable, then this proof fails. To handle the general case, apply **Jordan's theorem** 9.14 to write $A = PTP^{-1}$ where T is *upper triangular* (instead of *diagonal*) and the not necessarily distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ of A appear on the diagonal of T. Define

$$A_\epsilon = P(T + \epsilon\,\mathbf{diag}(1, 2, \ldots, n))P^{-1}.$$

For small $\epsilon > 0$, the matrix A_ϵ has distinct eigenvalues $\lambda_j + j\epsilon$, $1 \le j \le n$. Then the diagonalizable case implies that A_ϵ satisfies its characteristic equation. Let $p_\epsilon(\lambda) = \det(A_\epsilon - \lambda I)$. Use $\mathbf{0} = \lim_{\epsilon \to 0} p_\epsilon(A_\epsilon) = p(A)$ to complete the proof.

Proof of Theorem 9.17, orthogonality: Compute $\alpha\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = (A\vec{\mathbf{x}})^T\vec{\mathbf{y}} = \vec{\mathbf{x}}^T A^T \vec{\mathbf{y}} = \vec{\mathbf{x}}^T A \vec{\mathbf{y}}$. Analogously, $\beta\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = \vec{\mathbf{x}}^T A \vec{\mathbf{y}}$. Subtract the relations, then $(\alpha - \beta)\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = 0$. Because $\alpha \ne \beta$, then $\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = 0$. ■

Proof of Theorem 9.18, real eigenvalues: The second statement is due to the fundamental theorem of algebra. To prove the eigenvalues are real, it suffices to prove $\lambda = \overline{\lambda}$ when $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$ with $\vec{\mathbf{v}} \ne \vec{\mathbf{0}}$. A complex conjugate is computed by replacing i by $-i$. Conjugates of vectors and matrices are found componentwise. Assume that $\vec{\mathbf{v}}$ may have complex entries. Because A is real, then $\overline{A} = A$. Take the complex conjugate across $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$ to obtain $A\overline{\vec{\mathbf{v}}} = \overline{\lambda}\overline{\vec{\mathbf{v}}}$. Transpose $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$ to obtain $\vec{\mathbf{v}}^T A^T = \lambda\vec{\mathbf{v}}^T$ and then conclude $\vec{\mathbf{v}}^T A = \lambda\vec{\mathbf{v}}^T$ from $A^T = A$. Multiply this equation by $\overline{\vec{\mathbf{v}}}$ on the right to obtain $\vec{\mathbf{v}}^T A\overline{\vec{\mathbf{v}}} = \lambda\vec{\mathbf{v}}^T\overline{\vec{\mathbf{v}}}$. Then multiply $A\overline{\vec{\mathbf{v}}} = \overline{\lambda}\overline{\vec{\mathbf{v}}}$ by $\vec{\mathbf{v}}^T$ on the left to obtain $\vec{\mathbf{v}}^T A\overline{\vec{\mathbf{v}}} = \overline{\lambda}\vec{\mathbf{v}}^T\overline{\vec{\mathbf{v}}}$. The result:

$$\lambda\vec{\mathbf{v}}^T\overline{\vec{\mathbf{v}}} = \overline{\lambda}\vec{\mathbf{v}}^T\overline{\vec{\mathbf{v}}}.$$

Because $\vec{\mathbf{v}}^T\overline{\vec{\mathbf{v}}} = \sum_{j=1}^n |v_j|^2 > 0$, then it cancels: $\lambda = \overline{\lambda}$ and λ is real. ■

Proof of Theorem 9.19, Gram-Schmidt relations: Induction will be applied on k to show that $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_k$ are nonzero and orthogonal. If $k = 1$, then there is just one nonzero vector constructed $\vec{\mathbf{y}}_1 = \vec{\mathbf{x}}_1$. Orthogonality for $k = 1$ is not discussed because there are no pairs to test. Assume the result holds for $k - 1$ vectors. Let's verify that it holds for k vectors, $k > 1$. Assume orthogonality $\vec{\mathbf{y}}_i \cdot \vec{\mathbf{y}}_j = 0$ for $i \neq j$ and $\vec{\mathbf{y}}_i \neq \vec{\mathbf{0}}$ for $1 \leq i, j \leq k - 1$. It remains to test $\vec{\mathbf{y}}_i \cdot \vec{\mathbf{y}}_k = 0$ for $1 \leq i \leq k - 1$ and $\vec{\mathbf{y}}_k \neq \vec{\mathbf{0}}$. The test depends upon the identity

$$\vec{\mathbf{y}}_i \cdot \vec{\mathbf{y}}_k = \vec{\mathbf{y}}_i \cdot \vec{\mathbf{x}}_k - \sum_{j=1}^{k-1} \frac{\vec{\mathbf{x}}_k \cdot \vec{\mathbf{y}}_j}{\vec{\mathbf{y}}_j \cdot \vec{\mathbf{y}}_j} \vec{\mathbf{y}}_i \cdot \vec{\mathbf{y}}_j,$$

which is obtained from the formula for $\vec{\mathbf{y}}_k$ by taking the dot product with $\vec{\mathbf{y}}_i$. In the identity, $\vec{\mathbf{y}}_i \cdot \vec{\mathbf{y}}_j = 0$ by the induction hypothesis for $1 \leq j \leq k - 1$ and $j \neq i$. Therefore, the summation in the identity contains just the term for index $j = i$, and the contribution is $\vec{\mathbf{y}}_i \cdot \vec{\mathbf{x}}_k$. This contribution cancels the leading term on the right in the identity, resulting in the orthogonality relation $\vec{\mathbf{y}}_i \cdot \vec{\mathbf{y}}_k = 0$. If $\vec{\mathbf{y}}_k = \vec{\mathbf{0}}$, then $\vec{\mathbf{x}}_k$ is a linear combination of $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_{k-1}$. But each $\vec{\mathbf{y}}_j$ is a linear combination of $\{\vec{\mathbf{x}}_i\}_{i=1}^{j}$, therefore $\vec{\mathbf{y}}_k = \vec{\mathbf{0}}$ implies $\vec{\mathbf{x}}_k$ is a linear combination of $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_{k-1}$, a contradiction to the independence of $\{\vec{\mathbf{x}}_i\}_{i=1}^{k}$. ■

Proof of Theorem 9.20, Formula $\mathbf{Proj}_V(\vec{\mathbf{x}})$ is Well-Defined:
Suppose that $\{\vec{\mathbf{w}}_j\}_{j=1}^{k}$ is another orthonormal basis of V. Define $\vec{\mathbf{u}} = \sum_{i=1}^{k} (\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}}) \vec{\mathbf{u}}_j$ and $\vec{\mathbf{w}} = \sum_{j=1}^{k} (\vec{\mathbf{w}}_j \cdot \vec{\mathbf{x}}) \vec{\mathbf{w}}_j$. It will be established that $\vec{\mathbf{u}} = \vec{\mathbf{w}}$, which justifies that the projection formula is independent of basis. First, two lemmas.

Lemma 9.3 (Orthonormal Basis Expansion)
Let $\{\vec{\mathbf{v}}_j\}_{j=1}^{k}$ be an orthonormal basis of a subspace V in $\mathcal{R}^n$. Then each vector $\vec{\mathbf{v}}$ in V is represented as

$$\vec{\mathbf{v}} = \sum_{j=1}^{k} (\vec{\mathbf{v}}_j \cdot \vec{\mathbf{v}}) \vec{\mathbf{v}}_j.$$

Proof: First, $\vec{\mathbf{v}}$ has a basis expansion $\vec{\mathbf{v}} = \sum_{j=1}^{k} c_j \vec{\mathbf{v}}_j$ for some constants $c_1, \ldots, c_k$. Take the inner product of this equation with vector $\vec{\mathbf{v}}_i$ to prove that $c_i = \vec{\mathbf{v}}_i \cdot \vec{\mathbf{v}}$, hence the claimed expansion is proved.

Lemma 9.4 (Orthogonality) Let $\{\vec{\mathbf{u}}_i\}_{i=1}^{k}$ be an orthonormal basis of a subspace V in $\mathcal{R}^n$. Let $\vec{\mathbf{x}}$ be any vector in $\mathcal{R}^n$ and define $\vec{\mathbf{u}} = \sum_{i=1}^{k} (\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}}) \vec{\mathbf{u}}_i$. Then $\vec{\mathbf{y}} \cdot (\vec{\mathbf{x}} - \vec{\mathbf{u}}) = 0$ for all vectors $\vec{\mathbf{y}}$ in V.

Proof: The first lemma implies $\vec{\mathbf{u}}$ can be written a second way as a linear combination of $\vec{\mathbf{u}}_1, \ldots. \vec{\mathbf{u}}_k$. Independence implies equal basis coefficients, which gives $\vec{\mathbf{u}}_j \cdot \vec{\mathbf{u}} = \vec{\mathbf{u}}_j \cdot \vec{\mathbf{x}}$. Then $\vec{\mathbf{u}}_j \cdot (\vec{\mathbf{x}} - \vec{\mathbf{u}}) = 0$. Because $\vec{\mathbf{y}}$ is in V, then $\vec{\mathbf{y}} = \sum_{j=1}^{k} c_j \vec{\mathbf{u}}_j$, which implies $\vec{\mathbf{y}} \cdot (\vec{\mathbf{x}} - \vec{\mathbf{u}}) = \sum_{j=1}^{k} c_j \vec{\mathbf{u}}_j \cdot (\vec{\mathbf{x}} - \vec{\mathbf{u}}) = 0$. ■

Justification of $\vec{\mathbf{w}} = \vec{\mathbf{u}}$

The justification of Formula (1) is concluded here, showing that $\vec{\mathbf{w}} = \vec{\mathbf{u}}$.

$\vec{\mathbf{w}} = \sum_{j=1}^{k} (\vec{\mathbf{w}}_j \cdot \vec{\mathbf{x}}) \vec{\mathbf{w}}_j$	
$= \sum_{j=1}^{k} (\vec{\mathbf{w}}_j \cdot \vec{\mathbf{u}}) \vec{\mathbf{w}}_j$	Because $\vec{\mathbf{w}}_j \cdot (\vec{\mathbf{x}} - \vec{\mathbf{u}}) = 0$ by the second lemma.

$= \sum_{j=1}^k \left(\vec{\mathbf{w}}_j \cdot \sum_{i=1}^k (\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}})\vec{\mathbf{u}}_i\right) \vec{\mathbf{w}}_j$	Definition of $\vec{\mathbf{u}}$.
$= \sum_{j=1}^k \sum_{i=1}^k (\vec{\mathbf{w}}_j \cdot \vec{\mathbf{u}}_i)(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}})\vec{\mathbf{w}}_j$	Dot product properties.
$= \sum_{i=1}^k \left(\sum_{j=1}^k (\vec{\mathbf{w}}_j \cdot \vec{\mathbf{u}}_i)\vec{\mathbf{w}}_j\right) (\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}})$	Switch summations.
$= \sum_{i=1}^k \vec{\mathbf{u}}_i(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}})$	First lemma with $\vec{\mathbf{v}} = \vec{\mathbf{u}}_i$.
$= \vec{\mathbf{u}}$	Definition of $\vec{\mathbf{u}}$.

Proof of Theorem 9.21, Projection properties: Properties (a), (b) and (c) were proved in preceding lemmas. Details are outlined here, in case the lemmas were skipped.

(a): Write a basis expansion $\vec{\mathbf{v}} = \sum_{j=1}^k c_j \vec{\mathbf{u}}_j$ for some constants $c_1, \ldots, c_k$. Take the inner product of this equation with vector $\vec{\mathbf{u}}_i$ to prove that $c_i = \vec{\mathbf{u}}_i \cdot \vec{\mathbf{v}}$.

(b): Vector $\mathbf{Proj}_V(\vec{\mathbf{x}})$ is a linear combination of basis elements of V.

(c): Represent a given vector $\vec{\mathbf{v}}$ in V by the orthogonal expansion of$\vec{\mathbf{v}}$ from (a). Let's compute the dot product of $\vec{\mathbf{w}}$ and $\vec{\mathbf{v}}$:

$$\begin{aligned}
\vec{\mathbf{w}} \cdot \vec{\mathbf{v}} &= (\vec{\mathbf{x}} - \mathbf{Proj}_V(\vec{\mathbf{x}})) \cdot \vec{\mathbf{v}} \\
&= \vec{\mathbf{x}} \cdot \vec{\mathbf{v}} - \left(\sum_{j=1}^k (\vec{\mathbf{x}} \cdot \vec{\mathbf{u}}_j)\vec{\mathbf{u}}_j\right) \cdot \vec{\mathbf{v}} \\
&= \sum_{j=1}^k (\vec{\mathbf{u}}_j \cdot \vec{\mathbf{v}})(\vec{\mathbf{x}} \cdot \vec{\mathbf{u}}_j) - \sum_{j=1}^k (\vec{\mathbf{x}} \cdot \vec{\mathbf{u}}_j)(\vec{\mathbf{u}}_j \cdot \vec{\mathbf{v}}) \\
&= 0.
\end{aligned}$$

(d): Begin with the Pythagorean identity

$$\|\vec{\mathbf{a}}\|^2 + \|\vec{\mathbf{b}}\|^2 = \|\vec{\mathbf{a}} + \vec{\mathbf{b}}\|^2$$

valid exactly when $\vec{\mathbf{a}} \cdot \vec{\mathbf{b}} = 0$ (a right triangle, $\theta = 90°$). Using an arbitrary $\vec{\mathbf{v}}$ in V, define $\vec{\mathbf{a}} = \mathbf{Proj}_V(\vec{\mathbf{x}}) - \vec{\mathbf{v}}$ and $\vec{\mathbf{b}} = \vec{\mathbf{x}} - \mathbf{Proj}_V(\vec{\mathbf{x}})$. By (b), vector $\vec{\mathbf{a}}$ is in V. Because of (c), then $\vec{\mathbf{a}} \cdot \vec{\mathbf{b}} = 0$. This gives the identity

$$\| \mathbf{Proj}_V(\vec{\mathbf{x}}) - \vec{\mathbf{v}}\|^2 + \|\vec{\mathbf{x}} - \mathbf{Proj}_V(\vec{\mathbf{x}})\|^2 = \|\vec{\mathbf{x}} - \vec{\mathbf{v}}\|^2,$$

which establishes $\|\vec{\mathbf{x}} - \mathbf{Proj}_V(\vec{\mathbf{x}})\| < \|\vec{\mathbf{x}} - \vec{\mathbf{v}}\|$ except for the unique $\vec{\mathbf{v}}$ such that $\| \mathbf{Proj}_V(\vec{\mathbf{x}}) - \vec{\mathbf{v}}\| = 0$.

(e): Let $\vec{\mathbf{w}}$ be orthogonal to all vectors in V. Because the columns of A are in V, then $\vec{\mathbf{w}}$ is orthogonal to the columns of A, which are rows of A^T. Equation $A^T\vec{\mathbf{w}} = \vec{\mathbf{0}}$ means $\vec{\mathbf{w}}$ is orthogonal to the rows of A^T. ■

Proof of Theorem 9.24, Least Squares Solution:
Let $V = \mathbf{colspace}(A)$. Let $\vec{\mathbf{y}} = \mathbf{proj}_V(\vec{\mathbf{b}})$. Let $\vec{\mathbf{w}} = \vec{\mathbf{b}} - \vec{\mathbf{y}}$. Because $\vec{\mathbf{y}}$ is in the column space of A, then $\vec{\mathbf{y}} = A\vec{\mathbf{x}}^*$ for some $\vec{\mathbf{x}}^*$. By Theorem 9.21 (c), $\vec{\mathbf{w}} \cdot \vec{\mathbf{u}} = 0$ for every vector $\vec{\mathbf{u}}$ in V. This means $\vec{\mathbf{w}}\,\vec{\mathbf{u}}^T = 0$ for every column $\vec{\mathbf{u}}$ of A, which in turn means $A^T\vec{\mathbf{w}} = \vec{\mathbf{0}}$. Then $A^T(\vec{\mathbf{b}} - \vec{\mathbf{y}}) = \vec{\mathbf{0}}$ or equivalently $A^T\vec{\mathbf{b}} = A^TA\vec{\mathbf{x}}^*$. The Normal Equation has been verified for any $\vec{\mathbf{x}}^*$ such that $A\vec{\mathbf{x}}^* = \vec{\mathbf{y}} = \mathbf{proj}_V(\vec{\mathbf{b}})$. Theorem 9.21 (d) says that $\vec{\mathbf{x}} = \vec{\mathbf{x}}^*$ is a minimizer for $\|A\vec{\mathbf{x}} - \vec{\mathbf{b}}\|$ over all $\vec{\mathbf{x}}$. Then

$$\|A\vec{\mathbf{x}}^* - \vec{\mathbf{b}}\| = \min_{\vec{\mathbf{x}}} \|A\vec{\mathbf{x}} - \vec{\mathbf{b}}\|$$

■

Solution to Example 9.15, Height-Weight Best Fit:
The answer with 10 digits is $y = 61.2721865421106x - 39.0619559188439$, using assumptions made below for what is meant by **best fit**. A plot of the data and this straight line are in Figure 5.

Literature for the example might use the statistical term **simple linear regression**. The **regressors** in the example are unknowns m, b. The regression line $y = mx + b$ gives the expected weight y for height x, so y is the average or mean weight for a given height. Historically, **regression** abbreviates **regress back to the mean**, attributed to Sir Francis Galton (1822-1911) in work on genetics.

Linear algebraic equations in the unknowns m, b are discovered by inserting Table 4 data into $y = mx + b$, x=height, y=weight:

$$\begin{array}{llll} 52.21 = 1.47m + b & 53.12 = 1.50m + b & 54.48 = 1.52m + b & 55.84 = 1.55m + b \\ 57.20 = 1.57m + b & 58.57 = 1.60m + b & 59.93 = 1.63m + b & 61.29 = 1.65m + b \\ 63.11 = 1.68m + b & 64.47 = 1.70m + b & 66.28 = 1.73m + b & 68.10 = 1.75m + b \\ 69.92 = 1.78m + b & 72.19 = 1.80m + b & 74.46 = 1.83m + b & \end{array}$$

Define height vector $\vec{\mathbf{H}}$ and weight vector $\vec{\mathbf{W}}$ from Table 4, both vectors in $\mathcal{R}^{15}$. Let vector $\vec{\mathbf{O}}$ in $\mathcal{R}^{15}$ have all entries 1. Define augmented matrix $A = \left\langle \vec{\mathbf{H}} | \vec{\mathbf{O}} \right\rangle$. The fifteen linear algebraic equations become:

$$A \begin{pmatrix} m \\ b \end{pmatrix} = \vec{\mathbf{W}} \tag{9}$$

Among the *three possibilities* for a system of linear algebraic equations (Chapter 3 Section 1), system (9) has **no solution**. Terminology **best fit** has multiple possibilities, from which a single interpretation is isolated:

Best Fit

Find m, b to minimize the error between vectors $A \begin{pmatrix} m \\ b \end{pmatrix}$ and $\vec{\mathbf{W}}$.

The two vectors are the LHS and RHS of equation (9).

The answers $m = 61.2721865421106$, $b = -39.0619559188439$ are found by solving 2×2 matrix equation (12) on page 745. Details follow.

The idea for *solving* the unsolvable equation (9) is geometric: replace it with a solvable equation:

$$A \begin{pmatrix} m \\ b \end{pmatrix} = \vec{\mathbf{Z}} \tag{10}$$

Mystery vector $\vec{\mathbf{Z}}$ in (10) is the *unique* near point in $V = \mathbf{span}(\vec{\mathbf{H}}, \vec{\mathbf{O}})$ to $\vec{\mathbf{W}}$ given by near point Theorem 9.22.

Uniqueness of $\vec{\mathbf{Z}}$ means that the new equation $A \begin{pmatrix} m \\ b \end{pmatrix} = \vec{\mathbf{Z}}$ has a unique solution for m, b. The solution is efficiently found by multiplication of equation (10) by A^T:

$$A^T A \begin{pmatrix} m \\ b \end{pmatrix} = A^T \vec{\mathbf{Z}} = A^T \vec{\mathbf{W}}. \tag{11}$$

Equality $A^T \vec{\mathbf{Z}} = A^T \vec{\mathbf{W}}$ results from Theorem 9.21 (c): vector $\vec{\mathbf{w}} = \vec{\mathbf{W}} - \mathbf{Proj}_V(\vec{\mathbf{W}})$ is orthogonal to the columns of A and by Theorem 9.21 (e) $A^T \vec{\mathbf{w}} = \vec{\mathbf{0}}$. The simplified equation

(12) $$A^T A \begin{pmatrix} m \\ b \end{pmatrix} = A^T \vec{\mathbf{W}}$$

is called the **normal equation** for unsolvable system (9).

The new system is a 2×2 system with a unique solution m, b given by matrix inversion:

$$\begin{aligned} \begin{pmatrix} m \\ b \end{pmatrix} &= \left(A^T A\right)^{-1} A^T \vec{\mathbf{W}} \\ &= \begin{pmatrix} 41.0532 & 24.76 \\ 24.76 & 15 \end{pmatrix}^{-1} \begin{pmatrix} 1548.245 \\ 931.17 \end{pmatrix} \\ &= \begin{pmatrix} 61.2721865421106 \\ -39.0619559188439 \end{pmatrix}. \end{aligned}$$

```
with(LinearAlgebra):# Maple check
H:=Vector([ 1.47,1.50,1.52,1.55,1.57,1.60,1.63,1.65,
            1.68,1.70,1.73,1.75,1.78,1.80,1.83]);
W:=Vector([ 52.21,53.12,54.48,55.84,57.20,58.57,59.93,61.29,
            63.11,64.47,66.28,68.10,69.92,72.19,74.46]);
ONE:=Vector([1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]);
A:=<H|ONE>;
LinearSolve(A,W);Rank(A);# fail expected
B:=A^+ . A; Z:=A^+ . W; (1/B) . Z;
```

Exercises 9.3

Diagonalization

Find the eigenpair packages P and D in the relation $AP = PD$.

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

2. $A = \begin{pmatrix} 7 & 5 \\ 10 & -7 \end{pmatrix}$

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

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

5. $A = \begin{pmatrix} -1 & 0 & 3 \\ 3 & 4 & -9 \\ -1 & 0 & 3 \end{pmatrix}$

6. $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -3 \end{pmatrix}$

7. $A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$

8. $A = \begin{pmatrix} 4 & 0 & 0 & 1 \\ 12 & -2 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 21 & -6 & 1 & 0 \end{pmatrix}$

Jordan's Theorem

Given matrices P and T, verify Jordan's relation $AP = PT$.

9. $A = \begin{pmatrix} -4 & 2 \\ 0 & -1 \end{pmatrix}$, $P = I$, $T = A$.

10. $A = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}$, $P = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $T = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$

Cayley-Hamilton Theorem

11. Verify that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ satisfies
$A^2 = (a+d)A - (ad - bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

12. Verify $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}^{20} = \begin{pmatrix} 1 & 0 \\ 40 & 1 \end{pmatrix}$ by induction using Cayley-Hamilton.

Gram-Schmidt Process

Find the Gram–Schmidt orthonormal basis from the given independent set.

13. $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.
Ans: Columns of I.

14. $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 1 \end{pmatrix}$.

15. $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}$.

16. $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.
Ans: Columns of I.

Gram-Schmidt on Polynomials

Define $V = \mathbf{span}(1, x, x^2)$ with inner product $\int_0^1 f(x)g(x)dx$. Find a Gram–Schmidt orthonormal basis.

17. 1, $1 + x$, x^2

18. $1 - x$, $1 + x$, $1 + x^2$

Gram-Schmidt: Coordinate Map

Define $V = \mathbf{span}(1, x, x^2)$ with inner product $\int_0^1 f(x)g(x)dx$. The coordinate map is

$$T : \; c_1 + c_2 x + c_3 x^2 \to \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

19. Find the images of $1 - x$, $1 + x$, $1 + x^2$ under T.

20. Assume column vectors $\vec{\mathbf{x}}_1$, $\vec{\mathbf{x}}_2$, $\vec{\mathbf{x}}_3$ in $\mathcal{R}^3$ orthonormalize under Gram-Schmidt to $\vec{\mathbf{u}}_1$, $\vec{\mathbf{u}}_2$, $\vec{\mathbf{u}}_3$. Are the pre-images $T^{-1}(\vec{\mathbf{u}}_1)$, $T^{-1}(\vec{\mathbf{u}}_2)$, $T^{-1}(\vec{\mathbf{u}}_3)$ orthonormal in V?

Shadow Projection

Compute shadow vector $(\vec{\mathbf{x}} \cdot \vec{\mathbf{u}})\vec{\mathbf{u}}$ for direction $\vec{\mathbf{u}} = \frac{\vec{\mathbf{v}}}{|\vec{\mathbf{v}}|}$. Illustrate with a hand–drawn figure.

21. $\vec{\mathbf{x}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\vec{\mathbf{v}} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$
Ans: $-\frac{1}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

22. $\vec{\mathbf{x}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\vec{\mathbf{v}} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$

23. $\vec{\mathbf{x}} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$, $\vec{\mathbf{v}} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$
Ans: $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$

24. $\vec{\mathbf{x}} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 1 \end{pmatrix}$, $\vec{\mathbf{v}} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}$

Orthogonal Projection

Find an orthonormal basis $\{\vec{\mathbf{u}}_k\}_{k=1}^n$ for $V = \mathbf{span}(1 + x, x, x + x^2)$, inner product $\int_0^1 f(x)g(x)dx$. Then compute the orthogonal projection $\vec{\mathbf{p}} = \sum_{k=1}^n (\vec{\mathbf{x}} \cdot \vec{\mathbf{u}}_k)\vec{\mathbf{u}}_k$.

25. $\vec{\mathbf{x}} = 1 + x + x^2$

26. $\vec{\mathbf{x}} = 1 + 2x + x^2 + x^3$

Orthogonal Projection: Theory

27. Prove that the orthogonal projection $\mathbf{Proj}_V(\vec{\mathbf{x}})$ on $V = \mathbf{span}\{\vec{\mathbf{Y}}\}$ is the vector shadow projection $\mathbf{proj}_{\vec{\mathbf{Y}}}(\vec{\mathbf{x}})$.

28. **(Gram-Schmidt Construction)** Define $\vec{\mathbf{x}}_j^{\perp} = \vec{\mathbf{x}}_j - \mathbf{Proj}_{W_{j-1}}(\vec{\mathbf{x}}_j)$, and $W_{j-1} = \mathbf{span}(\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_{j-1})$. Prove these properties.

(a) Subspace W_{j-1} is equal to the Gram-Schmidt $V_{j-1} = \mathbf{span}(\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_j)$.

(b) Vector $\vec{\mathbf{x}}_j^\perp$ is orthogonal to all vectors in W_{j-1}.

(c) The vector $\vec{\mathbf{x}}_j^\perp$ is not zero.

(d) The Gram-Schmidt vector is

$$\vec{\mathbf{u}}_j = \frac{\vec{\mathbf{x}}_j^\perp}{\|\vec{\mathbf{x}}_j^\perp\|}.$$

Near Point Theorem

Find the near point to the subspace V.

29. $\vec{\mathbf{x}} = \begin{pmatrix}1\\1\end{pmatrix}$, $V = \mathbf{span}\left(\begin{pmatrix}1\\2\end{pmatrix}\right)$

30. $\vec{\mathbf{x}} = \begin{pmatrix}1\\1\end{pmatrix}$, $V = \mathbf{span}\left(\begin{pmatrix}0\\1\end{pmatrix}\right)$

31. $\vec{\mathbf{x}}=\begin{pmatrix}1\\1\\0\end{pmatrix}$,$V=\mathbf{span}\left(\begin{pmatrix}1\\2\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}\right)$

32. $\vec{\mathbf{x}}=\begin{pmatrix}1\\0\\1\end{pmatrix}$,$V=\mathbf{span}\left(\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}\right)$

QR-Decomposition

Give A, find an orthonormal matrix Q and an upper triangular matrix R such that $A = QR$.

33. $A=\begin{pmatrix}5&9\\1&7\\1&5\\3&5\end{pmatrix}$, Ans: $R = \begin{pmatrix}6&12\\0&6\end{pmatrix}$

34. $A=\begin{pmatrix}2&1\\2&0\\2&0\\2&1\end{pmatrix}$, Ans: $R = \begin{pmatrix}4&1\\0&1\end{pmatrix}$

35. $A=\begin{pmatrix}1&0&0\\1&1&0\\1&1&0\\1&0&0\end{pmatrix}$, Ans: $R=\begin{pmatrix}2&1&0\\0&1&0\end{pmatrix}$

36. $A=\begin{pmatrix}1&0&0\\1&1&1\\1&1&1\\1&0&0\end{pmatrix}$, Ans: $R=\begin{pmatrix}2&1&1\\0&1&1\end{pmatrix}$

Linear Least Squares: 3×2

Let $A=\begin{pmatrix}2&0\\0&2\\1&1\end{pmatrix}$, $\vec{\mathbf{b}}=\begin{pmatrix}1\\0\\5\end{pmatrix}$.

37. Find the normal equations for $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$.

38. Solve $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ by least squares.

Linear Least Squares: 4×3

Let $A=\begin{pmatrix}4&0&1\\1&0&1\\0&1&0\\1&1&1\end{pmatrix}$, $\vec{\mathbf{b}}=\begin{pmatrix}3\\0\\0\\0\end{pmatrix}$.

39. Find the normal equations for $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$.

40. Solve $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ by least squares.

Orthonormal Diagonal Form

Let $A = A^T$. The **spectral theorem** implies $AQ = QD$ where D is diagonal and Q has orthonormal columns. Find Q and D.

41. $A=\begin{pmatrix}7&2\\2&4\end{pmatrix}$

42. $A=\begin{pmatrix}1&5\\5&1\end{pmatrix}$

43. $A=\begin{pmatrix}1&5&0\\5&1&0\\0&0&2\end{pmatrix}$

Ans: Eigenvalues $-4, 2, 6$, orthonormal eigenvectors $\frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\\0\end{pmatrix}$,

$minicolvectorC001$, $\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix}$

44. $A=\begin{pmatrix}1&5&0\\5&1&1\\0&1&1\end{pmatrix}$

Eigenpairs of Symmetric Matrices: **Spectral Theorem**.

45. Let $A=\begin{pmatrix}3&-1&1\\-1&3&-1\\1&-1&3\end{pmatrix}$. Eigenvalues are $2, 2, 5$. Find three orthonormal eigenpairs.

46. Let $A = \begin{pmatrix} 5 & -1 & 1 \\ -1 & 5 & -1 \\ 1 & -1 & 5 \end{pmatrix}$. Then $|A - \lambda I| = (4-\lambda)^2(7-\lambda)$. Find three orthonormal eigenpairs.

47. Let $A = \begin{pmatrix} 6 & -1 & 1 \\ -1 & 6 & -1 \\ 1 & -1 & 6 \end{pmatrix}$. Eigenvectors $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ are for $\lambda = 5, 5, 8$. Illustrate $AQ = QD$ with D diagonal and Q orthogonal.

48. Matrix A for $\lambda = 1, 1, 4$ has orthogonal eigenvectors $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$. Find A and directly verify $A = A^T$.

Singular Value Decomposition

Find the SVD $A = U\Sigma V^T$.

49. $A = \begin{pmatrix} -1 & 1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix}$. Ans: $U = 3 \times 3$, $V = 2 \times 2$. Matrix $\Sigma = \begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = 3 \times 2$, the size of A.

50. $A = \begin{pmatrix} -1 & 1 \\ -2 & 2 \\ 1 & 1 \end{pmatrix}$. Ans: $\sigma_1 = \sqrt{10}, \sigma_2 = \sqrt{2}$.

51. $A = \begin{pmatrix} -3 & 3 \\ 0 & 0 \\ 1 & 1 \end{pmatrix}$.

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

Ellipse and the SVD

Repeat Example 9.17, page 736 for the given ellipse equation.

53. $50x^2 - 30xy + 10y^2 = 2500$

54. $40x^2 - 16xy + 10y^2 = 2500$

Mapping and the SVD

Reference: Example 9.18, page 738.

Let $\vec{\mathbf{w}} = \begin{pmatrix} x \\ y \end{pmatrix} = c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2$,

$U = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}$, $V = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$,

$A = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}$. Then $A = U\Sigma V^T$.

55. Verify $\|\vec{\mathbf{w}}\|^2 = \vec{\mathbf{w}} \cdot \vec{\mathbf{w}} = c_1^2 + c_2^2$.

56. Verify $V^T\vec{\mathbf{w}} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ from the general identity $V^TV = I$. Then show that $\Sigma V^T\vec{\mathbf{w}} = \begin{pmatrix} 10c_1 \\ 5c_2 \end{pmatrix}$.

Therefore, coordinate map $\vec{\mathbf{w}} \to \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ undergoes re-scaling by 10 in direction $\vec{\mathbf{v}}_1$ and 5 in direction $\vec{\mathbf{v}}_2$.

57. Find the angle θ of rotation for V^T and the reflection axis for U.

58. Assume $|\vec{\mathbf{w}}\| = 1$, a point on the unit circle. Is $A\vec{\mathbf{w}}$ on an ellipse with semi-axes 10 and 5? Justify your answer geometrically, no proof expected. Check your answer with a computer plot.

Four Fundamental Subspaces

Compute matrices S_1, S_2 such that the column spaces of S_1, S_2 are the nullspaces of A and A^T. Verify the two orthogonality relations of the four subspaces page 739 from the matrix identities $AS_1 = 0$, $A^TS_2 = 0$.

59. $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 1 & 0 \end{pmatrix}$. Answer:

$S_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, $S_2 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$.

60. $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 0 \end{pmatrix}$. Answer:

$S_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, $S_2 = \begin{pmatrix} -1 & -1 \\ -2 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$

61. $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 2 & 2 & 0 & 2 \end{pmatrix}$ Answer:

$$S_1 = \begin{pmatrix} 0 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \; S_2 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

62. $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$ Answer:

$$S_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \; S_2 = \begin{pmatrix} 2 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix},$$

Fundamental Theorem of Linear Algebra

Strang's Theorem says that the four subspaces built from $n \times m$ matrix A and $m \times n$ matrix A^T satisfy

$$\begin{array}{l} \textbf{colspace}(A^T) \perp \textbf{nullspace}(A), \\ \textbf{colspace}(A) \perp \textbf{nullspace}(A^T). \end{array}$$

Let $r = \textbf{rank}(A) = \textbf{rank}(A^T)$. The four subspace dimensions are:

$$\begin{array}{l} \dim(\textbf{colspace}(A)) = r, \\ \dim(\textbf{nullspace}(A)) = n - r, \\ \dim(\textbf{colspace}(A^T)) = r, \\ \dim(\textbf{nullspace}(A^T)) = m - r. \end{array}$$

63. Explain why $\dim(\textbf{colspace}(A)) = \dim(\textbf{colspace}(A^T)) = r$ from the Pivot Theorem.

64. Suppose A is 10×4. What are the dimensions of the four subspaces?

65. Invent a 4×4 matrix A where one of the four subspaces is the zero vector alone.

66. Prove that the only vector in common with **rowspace**(A) and **nullspace**(A) is the zero vector.

67. Prove that each vector $\vec{\mathbf{x}}$ in $\mathcal{R}^n$ can be uniquely written as $\vec{\mathbf{x}} = \vec{\mathbf{x}}_1 + \vec{\mathbf{x}}_2$ where $\vec{\mathbf{x}}_1$ is in **colspace**(A^T) and $\vec{\mathbf{x}}_2$ is in **nullspace**(A). See **direct sum** in exercise **??** page **??**.

68. Prove that each vector $\vec{\mathbf{y}}$ in $\mathcal{R}^m$ can be uniquely written as $\vec{\mathbf{y}} = \vec{\mathbf{y}}_1 + \vec{\mathbf{y}}_2$ where $\vec{\mathbf{y}}_1$ is in **colspace**(A) and $\vec{\mathbf{y}}_2$ is in **nullspace**(A^T).

PDF Sources

Text, Solutions and Corrections

Author: Grant B. Gustafson, University of Utah, Salt Lake City 84112.

Paperback Textbook: There are 12 chapters on differential equations and linear algebra, book format 7 x 10 inches, 1077 pages. Copies of the textbook are available in two volumes at **Amazon** Kindle Direct Publishing for Amazon's cost of printing and shipping. No author profit. Volume I chapters 1-7, ISBN 9798705491124, 661 pages. Volume II chapters 8-12, ISBN 9798711123651, 479 pages. Both paperbacks have extra pages of backmatter: background topics Chapter A, the whole book index and the bibliography.

Textbook PDF with Solution Manual: Packaged as one PDF (13 MB) with hyperlink navigation to displayed equations and theorems. The header in an exercise set has a blue hyperlink to the same section in the solutions. The header of the exercise section within a solution Appendix has a red hyperlink to the textbook exercises. Solutions are organized by chapter, e.g., Appendix 5 for Chapter 5. Odd-numbered exercises have a solution. A few even-numbered exercises have hints and answers. Computer code can be mouse-copied directly from the PDF. Free to use or download, no restrictions for educational use.

Sources at Utah:

https://math.utah.edu/g̃ustafso/indexUtahBookGG.html

Sources for a Local Folder No Internet: The same PDF can be downloaded to a tablet, computer or phone to be viewed locally. After download, no internet is required. Best for computer or tablet using a PDF viewer (Adobe Reader, Evince) or web browser with PDF support (Chrome, FireFox). Smart phones can be used in landscape mode.

Sources at GitHub and GitLab Projects: Utah sources are duplicated at

`https://github.com/ggustaf/github.io` and mirror
`https://gitlab.com/ggustaf/answers`.

Communication: To contribute a solution or correction, ask a question or request an answer, click the link below, then create a GitHub issue and post. Contributions and corrections are credited, privacy respected.

`https://github.com/ggustaf/github.io/issues`