Advanced Linear Algebra

References:

[A] Howard Anton & Chris Rorres. (2000). *Elementary Linear Algebra. New York: John Wiley & Sons, Inc*

[B] Setya Budi, W. 1995. *Aljabar Linear*. Jakarta: Gramedia

R. Rosnawati
Jurusan Pendidikan Matematika
FMIPA UNY

.

Inner Product Spaces

- 1 Inner Products
- 2 Angle and Orthogonality in Inner Product Spaces
- 3 Gram-Schmidt Process; QR-Decomposition
- 4 Best Approximation; Least Squares
- 5 Least Squares Fitting to Data
- 6 Function Approximation; Fourier Series

1 Inner Products

DEFINITION 1 An *inner product* on a real vector space V is a function that associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ with each pair of vectors in V in such a way that the following axioms are satisfied for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars k.

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ **[Symmetry axiom]**
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ **[Additivity axiom]**
3. $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$ **[Homogeneity axiom]**
4. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$ **[Positivity axiom]**

A real vector space with an inner product is called a *real inner product space*.

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

2. Algebraic Properties of Inner Products

THEOREM 6.1.2 *If* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *are vectors in a real inner product space* V, *and if* k *is a scalar, then:*

(*a*) $\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{0} \rangle = 0$

(*b*) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

(*c*) $\langle \mathbf{u}, \mathbf{v} - \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{w} \rangle$

(*d*) $\langle \mathbf{u} - \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle$

(*e*) $k\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, k\mathbf{v} \rangle$

θ: the angle between u and v

$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|}\right)$$

DEFINITION 1 Two vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space are called ***orthogonal*** if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.

3 Gram-Schmidt Process; QR-Decomposition

The Gram–Schmidt Process

To convert a basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ into an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$, perform the following computations:

Step 1. $\mathbf{v}_1 = \mathbf{u}_1$

Step 2. $\mathbf{v}_2 = \mathbf{u}_2 - \dfrac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$

Step 3. $\mathbf{v}_3 = \mathbf{u}_3 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2$

Step 4. $\mathbf{v}_4 = \mathbf{u}_4 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3$

$\vdots$

(continue for r steps)

Optional Step. To convert the orthogonal basis into an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_r\}$, normalize the orthogonal basis vectors.

QR-Decomposition

THEOREM 6.3.7 *QR*-Decomposition

If A is an $m \times n$ matrix with linearly independent column vectors, then A can be factored as

$$A = QR$$

where Q is an $m \times n$ matrix with orthonormal column vectors, and R is an $n \times n$ invertible upper triangular matrix.

4. Best Approximation; Least Squares

Least Squares Problem Given a linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns, find a vector $\mathbf{x}$ that minimizes $\|\mathbf{b} - A\mathbf{x}\|$ with respect to the Euclidean inner product on R^m. We call such an $\mathbf{x}$ a ***least squares solution*** of the system, we call $\mathbf{b} - A\mathbf{x}$ the ***least squares error vector***, and we call $\|\mathbf{b} - A\mathbf{x}\|$ the ***least squares error***.

THEOREM 6.4.1 Best Approximation Theorem

*If W is a finite-dimensional subspace of an inner product space V, and if $\mathbf{b}$ is a vector in V, then $\text{proj}_W \mathbf{b}$ is the **best approximation** to $\mathbf{b}$ from W in the sense that*

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$$

for every vector $\mathbf{w}$ in W that is different from $\text{proj}_W \mathbf{b}$.

Least squares solutions to A**x** = **b**

THEOREM 6.4.2 *For every linear system* $A\mathbf{x} = \mathbf{b}$, *the associated normal system*

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{5}$$

is consistent, and all solutions of (5) *are least squares solutions of* $A\mathbf{x} = \mathbf{b}$. *Moreover, if* W *is the column space of* A, *and* $\mathbf{x}$ *is any least squares solution of* $A\mathbf{x} = \mathbf{b}$, *then the orthogonal projection of* $\mathbf{b}$ *on* W *is*

$$\text{proj}_W \mathbf{b} = A\mathbf{x} \tag{6}$$

THEOREM 6.4.6 Equivalent Statements

If A is an $n \times n$ matrix, then the following statements are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

(e) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.

(f) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.

(g) $\det(A) \neq 0$.

(h) The column vectors of A are linearly independent.

(i) The row vectors of A are linearly independent.

(j) The column vectors of A span R^n.

(k) The row vectors of A span R^n.

(l) The column vectors of A form a basis for R^n.

(m) The row vectors of A form a basis for R^n.

(n) A has rank n.

(o) A has nullity 0.

(p) The orthogonal complement of the null space of A is R^n.

(q) The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.

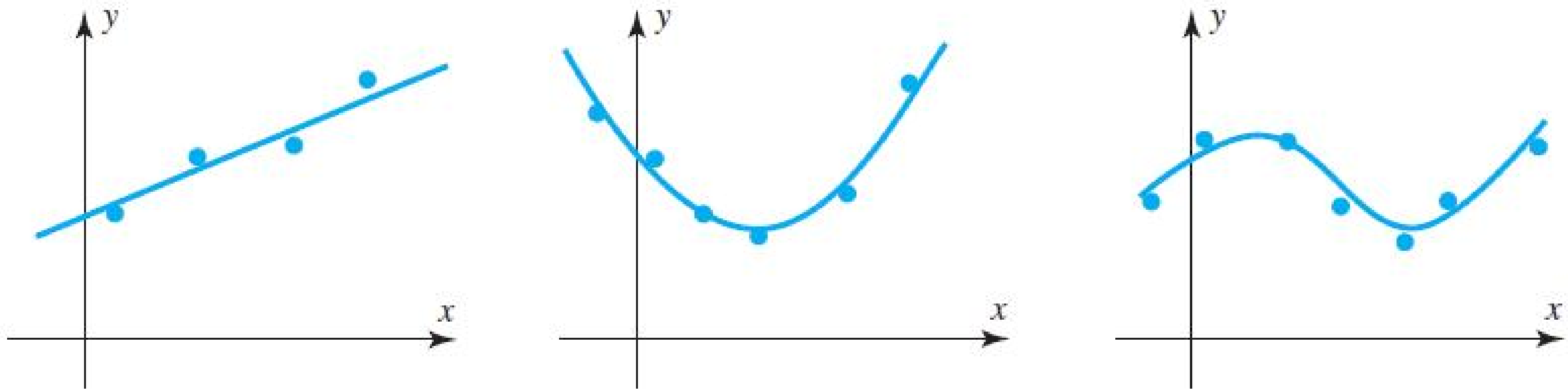
(r) The range of T_A is R^n.

(s) T_A is one-to-one.

(t) $\lambda = 0$ is not an eigenvalue of A.

(u) A^TA is invertible.

5 Least Squares Fitting to Data



(a) $y = a + bx$ (b) $y = a + bx + cx^2$ (c) $y = a + bx + cx^2 + dx^3$

The Least Squares Solution

THEOREM 6.5.1 Uniqueness of the Least Squares Solution

Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ *be a set of two or more data points, not all lying on a vertical line, and let*

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Then there is a unique least squares straight line fit

$$y = a^* + b^* x$$

to the data points. Moreover,

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix}$$

is given by the formula

$$\mathbf{v}^* = (M^T M)^{-1} M^T \mathbf{y} \tag{6}$$

which expresses the fact that $\mathbf{v} = \mathbf{v}^*$ *is the unique solution of the normal equations*

$$M^T M \mathbf{v} = M^T \mathbf{y} \tag{7}$$

6 Function Approximation; Fourier Series

THEOREM 6.6.1 *If* $\mathbf{f}$ *is a continuous function on* $[a, b]$, *and* W *is a finite-dimensional subspace of* $C[a, b]$, *then the function* $\mathbf{g}$ *in* W *that minimizes the mean square error*

$$\int_a^b [f(x) - g(x)]^2 \, dx$$

is $\mathbf{g} = \text{proj}_W \mathbf{f}$, *where the orthogonal projection is relative to the inner product*

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x) \, dx$$

The function $\mathbf{g} = \text{proj}_W \mathbf{f}$ *is called the* ***least squares approximation*** *to* $\mathbf{f}$ *from* W.

Fourier Coefficients & Series

A function of the form

$$T(x) = c_0 + c_1 \cos x + c_2 \cos 2x + \cdots + c_n \cos nx + d_1 \sin x + d_2 \sin 2x + \cdots + d_n \sin nx \quad (2)$$

is called a ***trigonometric polynomial***; if c_n and d_n are not both zero, then $T(x)$ is said to have ***order n***. For example,

$$T(x) = 2 + \cos x - 3\cos 2x + 7\sin 4x$$

is a trigonometric polynomial of order 4 with

$$c_0 = 2,\quad c_1 = 1,\quad c_2 = -3,\quad c_3 = 0,\quad c_4 = 0,\quad d_1 = 0,\quad d_2 = 0,\quad d_3 = 0,\quad d_4 = 7$$

$$\text{proj}_W \mathbf{f} = \frac{a_0}{2} + [a_1 \cos x + \cdots + a_n \cos nx] + [b_1 \sin x + \cdots + b_n \sin nx]$$

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx, \quad b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx$$

The numbers $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ are called the ***Fourier coefficients*** of **f**.

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx)$$

Fourier Approximation to y = x

$$x \approx \pi - 2\left(\sin x + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \cdots + \frac{\sin nx}{n}\right)$$

The graphs of $y = x$ and some of these approximations are shown in Figure 6.6.4.

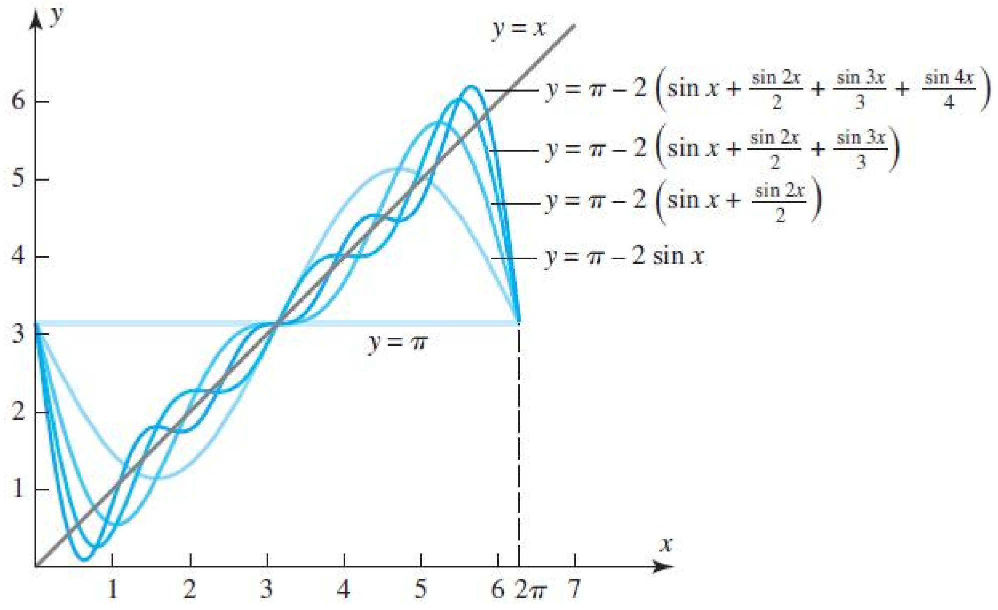


Figure 6.6.4

Diagonalization & Quadratic Forms

- 1 Orthogonal Matrices
- 2 Orthogonal Diagonalization
- 3 Quadratic Forms
- 4 Optimization Using Quadratic Forms
- 5 Hermitian, Unitary, and Normal Matrices

1 Orthogonal Matrices

DEFINITION 1 A square matrix A is said to be ***orthogonal*** if its transpose is the same as its inverse, that is, if

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

or, equivalently, if

$$AA^T = A^TA = I \tag{1}$$

THEOREM 7.1.1 *The following are equivalent for an $n \times n$ matrix A.*

(*a*) *A is orthogonal.*

(*b*) *The row vectors of A form an orthonormal set in R^n with the Euclidean inner product.*

(*c*) *The column vectors of A form an orthonormal set in R^n with the Euclidean inner product.*

THEOREM 7.1.2

(*a*) *The inverse of an orthogonal matrix is orthogonal.*

(*b*) *A product of orthogonal matrices is orthogonal.*

(*c*) *If A is orthogonal, then* $\det(A) = 1$ *or* $\det(A) = -1$.

THEOREM 7.1.3 *If A is an* $n \times n$ *matrix, then the following are equivalent.*

(*a*) *A is orthogonal.*

(*b*) $\|A\mathbf{x}\| = \|\mathbf{x}\|$ *for all* $\mathbf{x}$ *in* R^n.

(*c*) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ *for all* $\mathbf{x}$ *and* $\mathbf{y}$ *in* R^n.

Orthonormal Basis

THEOREM 7.1.4 *If S is an orthonormal basis for an n-dimensional inner product space V, and if*

$$(\mathbf{u})_S = (u_1, u_2, \ldots, u_n) \quad \text{and} \quad (\mathbf{v})_S = (v_1, v_2, \ldots, v_n)$$

then:

(a) $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$

(b) $d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}$

(c) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$

THEOREM 7.1.5 *Let V be a finite-dimensional inner product space. If P is the transition matrix from one orthonormal basis for V to another orthonormal basis for V, then P is an orthogonal matrix.*

Orthogonal Diagonalization

DEFINITION 1 If A and B are square matrices, then we say that A and B are ***orthogonally similar*** if there is an orthogonal matrix P such that $P^TAP = B$.

If A is orthogonally similar to some diagonal matrix, say

$$P^TAP = D$$

then we say that A is ***orthogonally diagonalizable*** and that P ***orthogonally diagonalizes*** A.

THEOREM 7.2.1 *If A is an $n \times n$ matrix, then the following are equivalent.*

(*a*) *A is orthogonally diagonalizable.*

(*b*) *A has an orthonormal set of n eigenvectors.*

(*c*) *A is symmetric.*

Symmetric Matrices

THEOREM 7.2.2 *If A is a symmetric matrix, then:*

(a) *The eigenvalues of A are all real numbers.*

(b) *Eigenvectors from different eigenspaces are orthogonal.*

Orthogonally Diagonalizing an $n \times n$ Symmetric Matrix

Step 1. Find a basis for each eigenspace of A.

Step 2. Apply the Gram–Schmidt process to each of these bases to obtain an orthonormal basis for each eigenspace.

Step 3. Form the matrix P whose columns are the vectors constructed in Step 2. This matrix will orthogonally diagonalize A, and the eigenvalues on the diagonal of $D = P^TAP$ will be in the same order as their corresponding eigenvectors in P.

Schur's Theorem

THEOREM 7.2.3 Schur's Theorem

If A is an $n \times n$ matrix with real entries and real eigenvalues, then there is an orthogonal matrix P such that P^TAP is an upper triangular matrix of the form

$$P^TAP = \begin{bmatrix} \lambda_1 & \times & \times & \cdots & \times \\ 0 & \lambda_2 & \times & \cdots & \times \\ 0 & 0 & \lambda_3 & \cdots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{11}$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of the matrix A repeated according to multiplicity.

Hessenberg's Theorem

THEOREM 7.2.4 Hessenberg's Theorem

If A is an $n \times n$ matrix, then there is an orthogonal matrix P such that P^TAP is a matrix of the form

$$P^TAP = \begin{bmatrix} \times & \times & \cdots & \times & \times & \times \\ \times & \times & \cdots & \times & \times & \times \\ 0 & \times & \ddots & \times & \times & \times \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \times & \times & \times \\ 0 & 0 & \cdots & 0 & \times & \times \end{bmatrix} \tag{13}$$

It is common to denote the upper Hessenberg matrix in (13) by H (for Hessenberg), in which case that equation can be rewritten as

$$A = PHP^T \tag{14}$$

which is called an ***upper Hessenberg decomposition*** of A.

3 Quadratic Forms

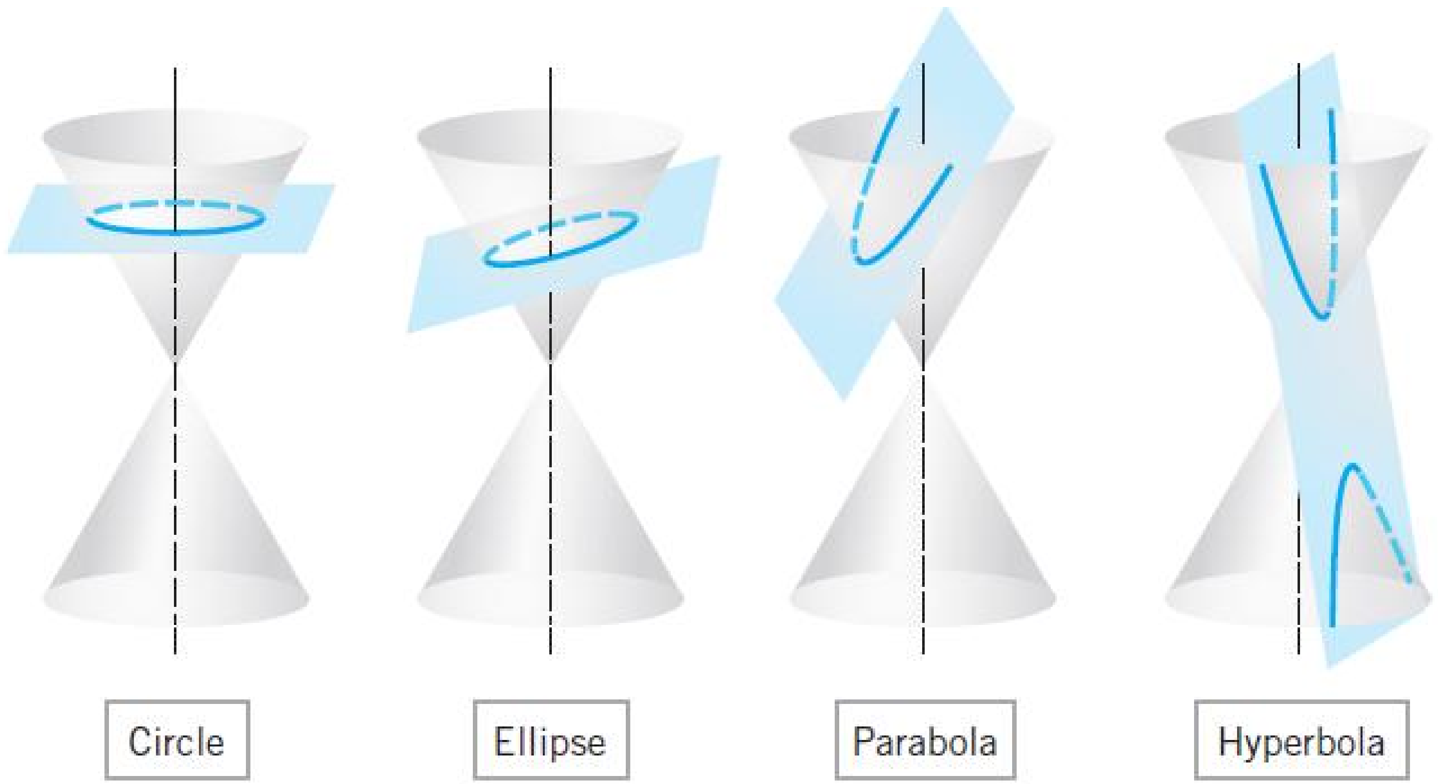
There are three important kinds of problems that occur in applications of quadratic forms:

Problem 1 If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on R^2 or R^3, what kind of curve or surface is represented by the equation $\mathbf{x}^T A\mathbf{x} = k$?

Problem 2 If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on R^n, what conditions must A satisfy for $\mathbf{x}^T A\mathbf{x}$ to have positive values for $\mathbf{x} \neq \mathbf{0}$?

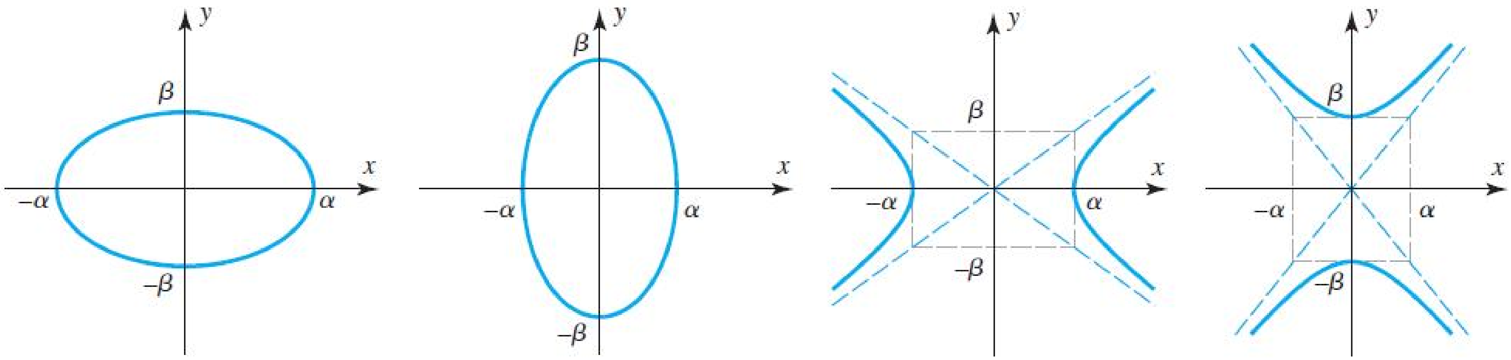
Problem 3 If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on R^n, what are its maximum and minimum values if $\mathbf{x}$ is constrained to satisfy $\|\mathbf{x}\| = 1$?

Conic Sections



Central conics in standard position

Table 1



$\dfrac{x^2}{\alpha^2}+\dfrac{y^2}{\beta^2}=1$ $(\alpha \geq \beta > 0)$	$\dfrac{x^2}{\alpha^2}+\dfrac{y^2}{\beta^2}=1$ $(\beta \geq \alpha > 0)$	$\dfrac{x^2}{\alpha^2}-\dfrac{y^2}{\beta^2}=1$ $(\alpha > 0, \beta > 0)$	$\dfrac{y^2}{\beta^2}-\dfrac{x^2}{\alpha^2}=1$ $(\alpha > 0, \beta > 0)$

Definite quadratic forms

DEFINITION 1 A quadratic form $\mathbf{x}^T A\mathbf{x}$ is said to be

positive definite if $\mathbf{x}^T A\mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$

negative definite if $\mathbf{x}^T A\mathbf{x} < 0$ for $\mathbf{x} \neq \mathbf{0}$

indefinite if $\mathbf{x}^T A\mathbf{x}$ has both positive and negative values

THEOREM 7.3.2 *If A is a symmetric matrix, then:*

(*a*) $\mathbf{x}^T A\mathbf{x}$ *is positive definite if and only if all eigenvalues of A are positive.*

(*b*) $\mathbf{x}^T A\mathbf{x}$ *is negative definite if and only if all eigenvalues of A are negative.*

(*c*) $\mathbf{x}^T A\mathbf{x}$ *is indefinite if and only if A has at least one positive eigenvalue and at least one negative eigenvalue.*

Ellipse? Hyperbola? Neither?

THEOREM 7.3.3 *If A is a symmetric 2×2 matrix, then:*

(*a*) $\mathbf{x}^T A\mathbf{x} = 1$ *represents an ellipse if A is positive definite.*

(*b*) $\mathbf{x}^T A\mathbf{x} = 1$ *has no graph if A is negative definite.*

(*c*) $\mathbf{x}^T A\mathbf{x} = 1$ *represents a hyperbola if A is indefinite.*

THEOREM 7.3.4 *A symmetric matrix A is positive definite if and only if the determinant of every principal submatrix is positive.*

4 Optimization Using Quadratic Forms

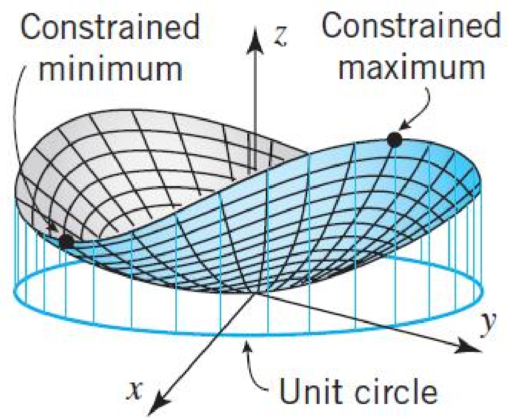
THEOREM 7.4.1 Constrained Extremum Theorem

Let A be a symmetric $n \times n$ matrix whose eigenvalues in order of decreasing size are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then:

(a) *the quadratic form $\mathbf{x}^T A\mathbf{x}$ attains a maximum value and a minimum value on the set of vectors for which $\|\mathbf{x}\| = 1$;*

(b) *the maximum value attained in part (a) occurs at a unit vector corresponding to the eigenvalue λ_1;*

(c) *the minimum value attained in part (a) occurs at a unit vector corresponding to the eigenvalue λ_n.*



THEOREM 7.4.2 Second Derivative Test

Suppose that (x_0, y_0) is a critical point of $f(x, y)$ and that f has continuous second-order partial derivatives in some circular region centered at (x_0, y_0). Then:

(a) *f has a relative minimum at (x_0, y_0) if*

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \quad \textit{and} \quad f_{xx}(x_0, y_0) > 0$$

(b) *f has a relative maximum at (x_0, y_0) if*

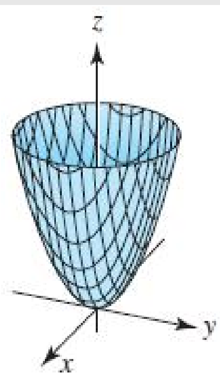
$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \quad \textit{and} \quad f_{xx}(x_0, y_0) < 0$$

(c) *f has a saddle point at (x_0, y_0) if*

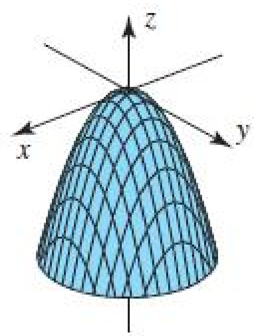
$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) < 0$$

(d) *The test is inconclusive if*

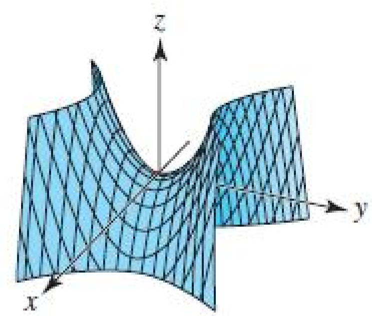
$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) = 0$$



Relative minimum at (0, 0)



Relative maximum at (0, 0)



Saddle point at (0, 0)

Hessian Form of the second derivative test

THEOREM 7.4.3 Hessian Form of the Second Derivative Test

Suppose that (x_0, y_0) is a critical point of $f(x, y)$ and that f has continuous second-order partial derivatives in some circular region centered at (x_0, y_0). If $H(x_0, y_0)$ is the Hessian of f at (x_0, y_0), then:

(a) *f has a relative minimum at (x_0, y_0) if $H(x_0, y_0)$ is positive definite.*

(b) *f has a relative maximum at (x_0, y_0) if $H(x_0, y_0)$ is negative definite.*

(c) *f has a saddle point at (x_0, y_0) if $H(x_0, y_0)$ is indefinite.*

(d) *The test is inconclusive otherwise.*

5 Hermitian, Unita and Normal Matrices

DEFINITION 1 If A is a complex matrix, then the ***conjugate transpose*** of A, denoted by A^*, is defined by

$$A^* = \overline{A}^T \tag{1}$$

THEOREM 7.5.1 *If k is a complex scalar, and if A, B, and C are complex matrices whose sizes are such that the stated operations can be performed, then:*

(*a*) $(A^*)^* = A$

(*b*) $(A + B)^* = A^* + B^*$

(*c*) $(A - B)^* = A^* - B^*$

(*d*) $(kA)^* = \overline{k}A^*$

(*e*) $(AB)^* = B^*A^*$

Hermitian Matrices

DEFINITION 2 A square complex matrix A is said to be ***unitary*** if

$$A^{-1} = A^* \tag{3}$$

and is said to be ***Hermitian**** if

$$A^* = A \tag{4}$$

THEOREM 7.5.2 *The eigenvalues of a Hermitian matrix are real numbers.*

THEOREM 7.5.3 *If A is a Hermitian matrix, then eigenvectors from different eigenspaces are orthogonal.*

Unitary Matrices

THEOREM 7.5.4 *If A is an $n \times n$ matrix with complex entries, then the following are equivalent.*

(a) *A is unitary.*

(b) $\|A\mathbf{x}\| = \|\mathbf{x}\|$ *for all* $\mathbf{x}$ *in* C^n.

(c) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ *for all* $\mathbf{x}$ *and* $\mathbf{y}$ *in* C^n.

(d) *The column vectors of A form an orthonormal set in C^n with respect to the complex Euclidean inner product.*

(e) *The row vectors of A form an orthonormal set in C^n with respect to the complex Euclidean inner product.*

DEFINITION 3 A square complex matrix is said to be ***unitarily diagonalizable*** if there is a unitary matrix P such that $P^*AP = D$ is a complex diagonal matrix. Any such matrix P is said to ***unitarily diagonalize*** A.

Unitarily Diagonalizing a Hermitian Matrix

THEOREM 7.5.5 *Every $n \times n$ Hermitian matrix A has an orthonormal set of n eigenvectors and is unitarily diagonalized by any $n \times n$ matrix P whose column vectors form an orthonormal set of eigenvectors of A.*

Unitarily Diagonalizing a Hermitian Matrix

Step 1. Find a basis for each eigenspace of A.

Step 2. Apply the Gram–Schmidt process to each of these bases to obtain orthonormal bases for the eigenspaces.

Step 3. Form the matrix P whose column vectors are the basis vectors obtained in Step 2. This will be a unitary matrix (Theorem 7.5.4) and will unitarily diagonalize A.

I. Linear Transformations

- General Linear Transformations
- Isomorphisms
- Compositions and Inverse Transformations
- Matrices for General Linear Transformations
- Similarity

General Linear Transformations

DEFINITION 1 If $T: V \rightarrow W$ is a function from a vector space V to a vector space W, then T is called a ***linear transformation*** from V to W if the following two properties hold for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V and for all scalars k:

(i) $T(k\mathbf{u}) = kT(\mathbf{u})$ **[Homogeneity property]**

(ii) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ **[Additivity property]**

In the special case where $V = W$, the linear transformation T is called a ***linear operator*** on the vector space V.

THEOREM 8.1.1 *If $T: V \rightarrow W$ is a linear transformation, then:*

(*a*) $T(\mathbf{0}) = \mathbf{0}$.

(*b*) $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$ *for all* $\mathbf{u}$ *and* $\mathbf{v}$ *in* V.

Dilation and Contraction Operators

If V is a vector space and k is any scalar, then the mapping $T : V \rightarrow V$ given by $T(\mathbf{x}) = k\mathbf{x}$ is a linear operator on V, for if c is any scalar and if $\mathbf{u}$ and $\mathbf{v}$ are any vectors in V, then

$$T(c\mathbf{u}) = k(c\mathbf{u}) = c(k\mathbf{u}) = cT(\mathbf{u})$$
$$T(\mathbf{u} + \mathbf{v}) = k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$$

If $0 < k < 1$, then T is called the ***contraction*** of V with factor k, and if $k > 1$, it is called the ***dilation*** of V with factor k (Figure 8.1.1).

Image, Kernel and Range

THEOREM 8.1.2 *Let $T : V \to W$ be a linear transformation, where V is finite dimensional. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V, then the image of any vector $\mathbf{v}$ in V can be expressed as*

$$T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + c_2 T(\mathbf{v}_2) + \cdots + c_n T(\mathbf{v}_n) \tag{3}$$

where $c_1, c_2, \ldots, c_n$ are the coefficients required to express $\mathbf{v}$ as a linear combination of the vectors in S.

DEFINITION 2 If $T : V \to W$ is a linear transformation, then the set of vectors in V that T maps into $\mathbf{0}$ is called the ***kernel*** of T and is denoted by $\ker(T)$. The set of all vectors in W that are images under T of at least one vector in V is called the ***range*** of T and is denoted by $R(T)$.

THEOREM 8.1.3 *If $T : V \to W$ is a linear transformation, then:*

(*a*) *The kernel of T is a subspace of V.*

(*b*) *The range of T is a subspace of W.*

Rank, Nullity and Dimension

DEFINITION 3 Let $T : V \to W$ be a linear transformation. If the range of T is finite-dimensional, then its dimension is called the ***rank of T***; and if the kernel of T is finite-dimensional, then its dimension is called the ***nullity of T***. The rank of T is denoted by $\text{rank}(T)$ and the nullity of T by $\text{nullity}(T)$.

THEOREM 8.1.4 Dimension Theorem for Linear Transformations

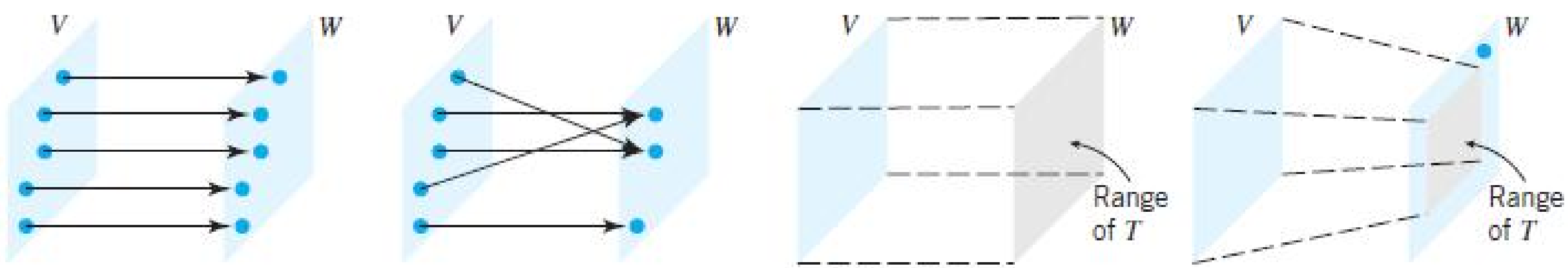
If $T : V \to W$ is a linear transformation from an n-dimensional vector space V to a vector space W, then

$$\text{rank}(T) + \text{nullity}(T) = n \tag{7}$$

Isomorphism

DEFINITION 1 If $T: V \to W$ is a linear transformation from a vector space V to a vector space W, then T is said to be ***one-to-one*** if T maps distinct vectors in V into distinct vectors in W.

DEFINITION 2 If $T: V \to W$ is a linear transformation from a vector space V to a vector space W, then T is said to be ***onto*** (or ***onto W***) if every vector in W is the image of at least one vector in V.



One-to-one. Distinct vectors in V have distinct images in W.

Not one-to-one. There exist distinct vectors in V with the same image.

Onto W. Every vector in W is the image of some vector in V.

Not onto W. Not every vector in W is the image of some vector in V.

Isomorphism

THEOREM 8.2.1 *If $T: V \to W$ is a linear transformation, then the following statements are equivalent.*

(*a*) *T is one-to-one.*

(*b*) $\ker(T) = \{\mathbf{0}\}$.

THEOREM 8.2.2 *If V is a finite-dimensional vector space, and if $T: V \to V$ is a linear operator, then the following statements are equivalent.*

(*a*) *T is one-to-one.*

(*b*) $\ker(T) = \{\mathbf{0}\}$.

(*c*) *T is onto* [*i.e.*, $R(T) = V$].

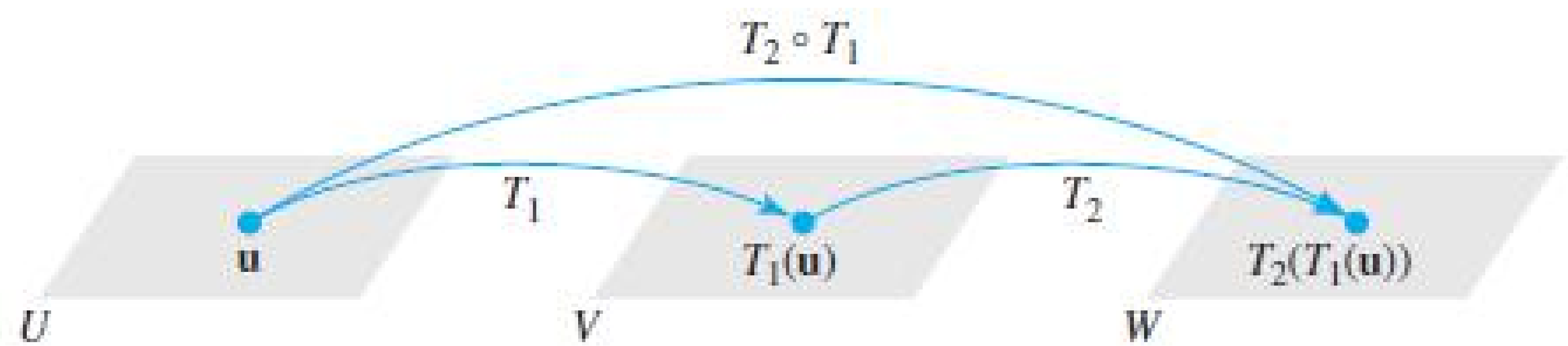
DEFINITION 3 If a linear transformation $T: V \to W$ is both one-to-one and onto, then T is said to be an ***isomorphism***, and the vector spaces V and W are said to be ***isomorphic***.

Compositions and Inverse Transformations

DEFINITION 1 If $T_1: U \rightarrow V$ and $T_2: V \rightarrow W$ are linear transformations, then the ***composition of T_2 with T_1***, denoted by $T_2 \circ T_1$ (which is read "T_2 circle T_1"), is the function defined by the formula

$$(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u})) \tag{1}$$

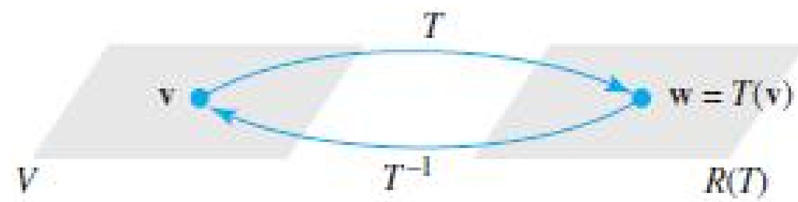
where $\mathbf{u}$ is a vector in U.



Inverses

$$T^{-1}(T(\mathbf{v})) = T^{-1}(\mathbf{w}) = \mathbf{v}$$

$$T(T^{-1}(\mathbf{w})) = T(\mathbf{v}) = \mathbf{w}$$



THEOREM 8.3.2 *If $T_1: U \to V$ and $T_2: V \to W$ are one-to-one linear transformations, then*

(a) $T_2 \circ T_1$ *is one-to-one.*

(b) $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$.

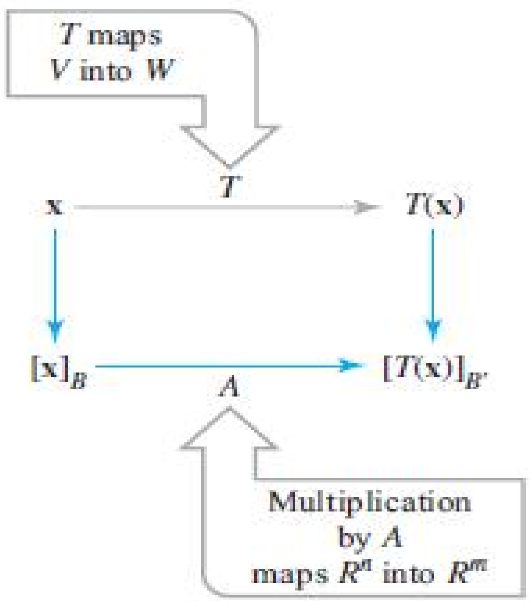
Matrices for General Linear Transformations

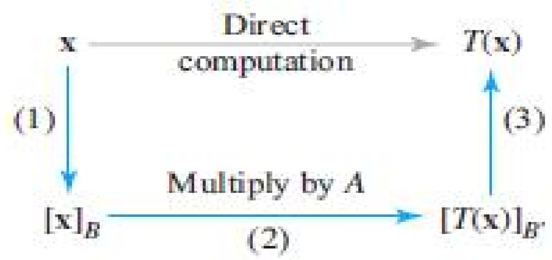
Finding $T(\mathbf{x})$ Indirectly

Step 1. Compute the coordinate vector $[\mathbf{x}]_B$.

Step 2. Multiply $[\mathbf{x}]_B$ on the left by A to produce $[T(\mathbf{x})]_{B'}$.

Step 3. Reconstruct $T(\mathbf{x})$ from its coordinate vector $[T(\mathbf{x})]_{B'}$.





Matrix of Compositions and Inverse Transformations

THEOREM 8.4.1 *If $T_1: U \to V$ and $T_2: V \to W$ are linear transformations, and if B, B'', and B' are bases for U, V, and W, respectively, then*

$$[T_2 \circ T_1]_{B',B} = [T_2]_{B',B''}[T_1]_{B'',B} \tag{10}$$

THEOREM 8.4.2 *If $T: V \to V$ is a linear operator, and if B is a basis for V, then the following are equivalent.*

(*a*) T *is one-to-one.*

(*b*) $[T]_B$ *is invertible.*

Moreover, when these equivalent conditions hold,

$$[T^{-1}]_B = [T]_B^{-1} \tag{11}$$

Similarity

THEOREM 8.5.3 *Two matrices, A and B, are similar if and only if they represent the same linear operator. Moreover, if $B = P^{-1}AP$, then P is the transition matrix from the basis relative to matrix B to the basis relative to matrix A.*

Table 1 Similarity Invariants

Property	Description
Determinant	A and $P^{-1}AP$ have the same determinant.
Invertibility	A is invertible if and only if $P^{-1}AP$ is invertible.
Rank	A and $P^{-1}AP$ have the same rank.
Nullity	A and $P^{-1}AP$ have the same nullity.
Trace	A and $P^{-1}AP$ have the same trace.
Characteristic polynomial	A and $P^{-1}AP$ have the same characteristic polynomial.
Eigenvalues	A and $P^{-1}AP$ have the same eigenvalues.
Eigenspace dimension	If λ is an eigenvalue of A and $P^{-1}AP$, then the eigenspace of A corresponding to λ and the eigenspace of $P^{-1}AP$ corresponding to λ have the same dimension.

Eigenvalues and Eigenvectors

Definition 1: A nonzero vector **x** is an ***eigenvector*** (or *characteristic vector*) of a square matrix **A** if there exists a scalar **λ** such that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Then **λ** is an ***eigenvalue*** (or *characteristic value*) of **A**.

Note: The zero vector can not be an eigenvector even though $\mathbf{A}\mathbf{0} = \lambda\mathbf{0}$. But $\lambda = 0$ can be an eigenvalue.

Example:

Geometric interpretation of Eigenvalues and Eigenvectors

An $n \times n$ matrix **A** multiplied by $n \times 1$ vector **x** results in another $n \times 1$ vector **y=Ax**. Thus **A** can be considered as a transformation matrix.

In general, a matrix acts on a vector by changing both its magnitude and its direction. However, a matrix may act on certain vectors by changing only their magnitude, and leaving their direction unchanged (or possibly reversing it). These vectors are the **eigenvectors** of the matrix.

A matrix acts on an eigenvector by multiplying its magnitude by a factor, which is positive if its direction is unchanged and negative if its direction is reversed. This factor is the **eigenvalue** associated with that eigenvector.

6.2 Eigenvalues

Let x be an eigenvector of the matrix **A.** Then there must exist an eigenvalue λ such that $\mathbf{Ax = \lambda x}$ or, equivalently,

$$Ax - \lambda x = 0 \quad \text{or}$$

$$(A - \lambda I)x = 0$$

If we define a new matrix $\mathbf{B = A - \lambda I}$, then

$$Bx = 0$$

If **B** has an inverse then $\mathbf{x = B^{-1}0 = 0}$. But an eigenvector cannot be zero.

Thus, it follows that x will be an eigenvector of **A** if and only if **B** does not have an inverse, or equivalently $\mathbf{det(B)=0}$, or

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

This is called the **characteristic equation** of **A**. Its roots determine the eigenvalues of **A**.

6.2 Eigenvalues: examples

Example 1: Find the eigenvalues of $A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}$

$$|aI - A| = \begin{vmatrix} a - 2 & 12 \\ -1 & a + 5 \end{vmatrix} = (a - 2)(a + 5) + 12$$

$$= a^2 + 3a + 2 = (a + 1)(a + 2)$$

two eigenvalues: –1, – 2

Note: The roots of the characteristic equation can be repeated. That is, $\lambda_1 = \lambda_2$ =…= λ_k. If that happens, the eigenvalue is said to be of multiplicity k.

Example 2: Find the eigenvalues of

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

$$|aI - A| = \begin{vmatrix} a - 2 & -1 & 0 \\ 0 & a - 2 & 0 \\ 0 & 0 & a - 2 \end{vmatrix} = (a - 2)^3 = 0$$

λ = 2 is an eigenvector of multiplicity 3.

6.3 Eigenvectors

To each distinct eigenvalue of a matrix **A** there will correspond at least one eigenvector which can be found by solving the appropriate set of homogenous equations. If $\boldsymbol{\lambda_i}$ is an eigenvalue then the corresponding eigenvector $\mathbf{x_i}$ is the solution of **$(A - \lambda_i I)x_i = 0$**

Example 1 (cont.):

$$a = -1 : (-1)I - A = \begin{bmatrix} -3 & 12 \\ -1 & 4 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & -4 \\ 0 & 0 \end{bmatrix}$$

$$x_1 - 4x_2 = 0 \Rightarrow x_1 = 4t,\ x_2 = t$$

$$\mathbf{x}_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t\begin{bmatrix} 4 \\ 1 \end{bmatrix},\ t \neq 0$$

$$a = -2 : (-2)I - A = \begin{bmatrix} -4 & 12 \\ -1 & 3 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{x}_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = s\begin{bmatrix} 3 \\ 1 \end{bmatrix},\ s \neq 0$$

6.3 Eigenvectors

Example 2 (cont.): Find the eigenvectors of $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Recall that λ = 2 is an eigenvector of multiplicity 3.

Solve the homogeneous linear system represented by

$$(2I - A)\mathbf{x} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Let $x_1 = s,\ x_3 = t$. The eigenvectors of λ = 2 are of the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} s \\ 0 \\ t \end{bmatrix} = s \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

s and *t* not both zero.

6.4 Properties of Eigenvalues and Eigenvectors

Definition: The trace of a matrix A, designated by tr(A), is the sum of the elements on the main diagonal.

Property 1: The sum of the eigenvalues of a matrix equals the trace of the matrix.

Property 2: A matrix is singular if and only if it has a zero eigenvalue.

Property 3: The eigenvalues of an upper (or lower) triangular matrix are the elements on the main diagonal.

Property 4: If **λ** is an eigenvalue of **A** and **A** is invertible, then **1/λ** is an eigenvalue of matrix $\mathbf{A}^{-1}$.

6.4 Properties of Eigenvalues and Eigenvectors

Property 5: If **λ** is an eigenvalue of **A** then **kλ** is an eigenvalue of **kA** where **k** is any arbitrary scalar.

Property 6: If **λ** is an eigenvalue of **A** then $\boldsymbol{\lambda^k}$ is an eigenvalue of $\mathbf{A^k}$ for any positive integer **k**.

Property 8: If **λ** is an eigenvalue of **A** then **λ** is an eigenvalue of $\mathbf{A^T}$.

Property 9: The product of the eigenvalues (counting multiplicity) of a matrix equals the determinant of the matrix.

6.5 Linearly independent eigenvectors

Theorem: Eigenvectors corresponding to distinct (that is, different) eigenvalues are linearly independent.

Theorem: If **λ** is an eigenvalue of multiplicity **k** of an **n** × **n** matrix **A** then **the number of linearly independent** eigenvectors of **A** associated with **λ** is given by **m = n - r(A- λI)**. Furthermore, **1 ≤ m ≤ k**.

Example 2 (cont.): The eigenvectors of $\lambda = 2$ are of the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} s \\ 0 \\ t \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

s and *t* not both zero.

$\lambda = 2$ has ***two linearly independent eigenvectors***