

Math 225 Linear Algebra II Lecture Notes

John C. Bowman
University of Alberta
Edmonton, Canada



September 2, 2024

© 2010
John C. Bowman
ALL RIGHTS RESERVED

Reproduction of these lecture notes in any form, in whole or in part, is permitted only for nonprofit, educational use.

Contents

1	Vectors	4
2	Linear Equations	6
3	Matrix Algebra	8
4	Determinants	11
5	Eigenvalues and Eigenvectors	13
6	Linear Transformations	16
7	Dimension	17
8	Similarity and Diagonalizability	18
9	Complex Numbers	23
10	Projection Theorem	28
11	Gram-Schmidt Orthonormalization	29
12	QR Factorization	31
13	Least Squares Approximation	32
14	Orthogonal (Unitary) Diagonalizability	34
15	Systems of Differential Equations	40
16	Quadratic Forms	43
17	Vector Spaces:	46
18	Inner Product Spaces	48
19	General linear transformations	51
20	Singular Value Decomposition	53
21	The Pseudoinverse	57
	Index	58

1 Vectors

- Vectors in \mathbb{R}^n :

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \quad \mathbf{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

- Parallelogram law:

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{bmatrix}.$$

- Multiplication by scalar $c \in \mathbb{R}$:

$$c\mathbf{v} = \begin{bmatrix} cv_1 \\ \vdots \\ cv_n \end{bmatrix}.$$

- Dot (inner) product:

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + \cdots + u_nv_n.$$

- Length (norm):

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2}.$$

- Unit vector:

$$\frac{1}{|\mathbf{v}|}\mathbf{v}.$$

- Distance between (endpoints of) \mathbf{u} and \mathbf{v} (positioned at $\mathbf{0}$):

$$d(\mathbf{u}, \mathbf{v}) = |\mathbf{u} - \mathbf{v}|.$$

- Law of cosines:

$$|\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2|\mathbf{u}||\mathbf{v}|\cos\theta.$$

- Angle θ between \mathbf{u} and \mathbf{v} :

$$\begin{aligned} |\mathbf{u}||\mathbf{v}|\cos\theta &= \frac{1}{2}(|\mathbf{u}|^2 + |\mathbf{v}|^2 - |\mathbf{u} - \mathbf{v}|^2) = \frac{1}{2}(|\mathbf{u}|^2 + |\mathbf{v}|^2 - [\mathbf{u} - \mathbf{v}] \cdot [\mathbf{u} - \mathbf{v}]) \\ &= \frac{1}{2}(|\mathbf{u}|^2 + |\mathbf{v}|^2 - [\mathbf{u} \cdot \mathbf{u} - 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v}]) \\ &= \mathbf{u} \cdot \mathbf{v}. \end{aligned}$$

- *Orthogonal vectors \mathbf{u} and \mathbf{v} :*

$$\mathbf{u} \cdot \mathbf{v} = 0.$$

- *Pythagoras theorem for orthogonal vectors \mathbf{u} and \mathbf{v} :*

$$|\mathbf{u} + \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2.$$

- *Cauchy-Schwarz inequality ($|\cos \theta| \leq 1$):*

$$|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}| |\mathbf{v}|.$$

- *Triangle inequality:*

$$\begin{aligned} |\mathbf{u} + \mathbf{v}| &= \sqrt{(\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v})} = \sqrt{|\mathbf{u}|^2 + 2\mathbf{u} \cdot \mathbf{v} + |\mathbf{v}|^2} \\ &\leq \sqrt{|\mathbf{u}|^2 + 2|\mathbf{u} \cdot \mathbf{v}| + |\mathbf{v}|^2} \leq \sqrt{|\mathbf{u}|^2 + 2|\mathbf{u}| |\mathbf{v}| + |\mathbf{v}|^2} = \sqrt{(|\mathbf{u}| + |\mathbf{v}|)^2} = |\mathbf{u}| + |\mathbf{v}|. \end{aligned}$$

- *Component of \mathbf{v} in the direction of a vector \mathbf{u} at an angle θ :*

$$\mathbf{v} \cdot \frac{\mathbf{u}}{|\mathbf{u}|} = |\mathbf{v}| \cos \theta.$$

- *Equation of line:*

$$Ax + By = C.$$

- *Parametric equation of line through \mathbf{p} and \mathbf{q} :*

$$\mathbf{v} = (1 - t)\mathbf{p} + t\mathbf{q}.$$

- *Equation of plane with unit normal (A, B, C) a distance D from origin:*

$$Ax + By + Cz = D.$$

- *Equation of plane with normal (A, B, C) through (x_0, y_0, z_0) :*

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

- *Parametric equation of plane spanned by \mathbf{u} and \mathbf{w} through \mathbf{v}_0 :*

$$\mathbf{v} = \mathbf{v}_0 + t\mathbf{u} + s\mathbf{w}.$$

2 Linear Equations

- *Linear equation:*

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

- *Homogeneous linear equation:*

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

- *System of linear equations:*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

- Systems of linear equations may have 0, 1, or an infinite number of solutions [Anton & Busby, p. 41].
- *Matrix* formulation:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{n \times m \text{ coefficient matrix}} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

- *Augmented matrix:*

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m. \end{bmatrix}.$$

- *Elementary row operations:*

- Multiply a row by a nonzero constant.
- Interchange two rows.
- Add a multiple of one row to another.

Remark: Elementary row operations do not change the solution!

- *Row echelon form (Gaussian elimination):*

$$\begin{bmatrix} 1 & * & * & \cdots & * \\ 0 & 0 & 1 & \cdots & * \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

- *Reduced row echelon form (Gauss–Jordan elimination):*

$$\begin{bmatrix} 1 & * & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

- For example:

$$\begin{aligned} 2x_1 + x_2 &= 5, \\ 7x_1 + 4x_2 &= 17. \end{aligned}$$

- A *diagonal matrix* is a square matrix whose nonzero values appear only as entries a_{ii} along the diagonal. For example, the following matrix is diagonal:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

- An *upper triangular matrix* has zero entries everywhere below the diagonal ($a_{ij} = 0$ for $i > j$).
- A *lower triangular matrix* has zero entries everywhere above the diagonal ($a_{ij} = 0$ for $i < j$).

Problem 2.1: Show that the product of two upper triangular matrices of the same size is an upper triangle matrix.

Problem 2.2: If \mathbf{L} is a lower triangular matrix, and \mathbf{U} is an upper triangular matrix, show that \mathbf{LU} is a lower triangular matrix and \mathbf{UL} is an upper triangular matrix.

3 Matrix Algebra

- Element a_{ij} appears in row i and column j of the $m \times n$ matrix $\mathbf{A} = [a_{ij}]_{m \times n}$.

- Addition of matrices of the same size:

$$[a_{ij}]_{m \times n} + [b_{ij}]_{m \times n} = [a_{ij} + b_{ij}]_{m \times n}.$$

- Multiplication of a matrix by a scalar $c \in \mathbb{R}$:

$$c[a_{ij}]_{m \times n} = [ca_{ij}]_{m \times n}.$$

- Multiplication of matrices:

$$[a_{ij}]_{m \times n} [b_{jk}]_{n \times \ell} = \left[\sum_{j=1}^n a_{ij} b_{jk} \right]_{m \times \ell}.$$

- Matrix multiplication is *linear*:

$$\mathbf{A}(\alpha \mathbf{B} + \beta \mathbf{C}) = \alpha \mathbf{A}\mathbf{B} + \beta \mathbf{A}\mathbf{C}$$

for all scalars α and β .

- Two matrices \mathbf{A} and \mathbf{B} are said to *commute* if $\mathbf{AB} = \mathbf{BA}$.

Problem 3.1: Give examples of 2×2 matrices that commute and ones that don't.

- *Transpose* of a matrix:

$$[a_{ij}]_{m \times n}^T = [a_{ji}]_{n \times m}.$$

- A *symmetric matrix* A is equal to its transpose.

Problem 3.2: Does a matrix necessarily commute with its transpose? Prove or provide a counterexample.

Problem 3.3: Let \mathbf{A} be an $m \times n$ matrix and \mathbf{B} be an $n \times \ell$ matrix. Prove that

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T.$$

Problem 3.4: Show that the dot product can be expressed as a matrix multiplication:
 $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$. Also, since $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$, we see equivalently that $\mathbf{u} \cdot \mathbf{v} = \mathbf{v}^T \mathbf{u}$.

- Trace of a square matrix:

$$\text{Tr} [a_{ij}]_{n \times n} = \sum_{i=1}^n a_{ii}.$$

Problem 3.5: Let \mathbf{A} and \mathbf{B} be square matrices of the same size. Prove that

$$\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA}).$$

- Identity matrix:

$$\mathbf{I}_n = [\delta_{ij}]_{n \times n}.$$

- Inverse \mathbf{A}^{-1} of an invertible matrix \mathbf{A} :

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Problem 3.6: If $\mathbf{Ax} = \mathbf{b}$ is a linear system of n equations, and the coefficient matrix \mathbf{A} is invertible, prove that the system has the unique solution

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

Problem 3.7: Prove that

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

- Inverse of a 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = (ad - bc) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

\Rightarrow

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \text{exists if and only if } ad - bc \neq 0.$$

- Determinant of a 2×2 matrix:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

- An *elementary matrix* is obtained on applying a *single* elementary row operation to the identity matrix.
- Two matrices are *row equivalent* if one can be obtained from the other by elementary row operations.
- An elementary matrix is row equivalent to the identity matrix.
- A *subspace* is closed under scalar multiplication and addition.
- $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is the subspace formed by the set of linear combinations $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$ generated by all possible scalar multipliers c_1, c_2, \dots, c_n .
- The subspaces in \mathbb{R}^3 are the origin, lines, planes, and all of \mathbb{R}^3 .
- The solution space of the homogeneous linear system $\mathbf{Ax} = \mathbf{0}$ with n unknowns is always a subspace of \mathbb{R}^n .
- A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ in \mathbb{R}^n are *linearly independent* if the equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n = \mathbf{0}$$

has only the trivial solution $c_1 = c_2 = \dots = c_n = 0$. If this were not the case, we say that the set is linearly dependent; for $n \geq 2$, this means we can express at least one of the vectors in the set as a linear combination of the others.

- The following are equivalent for an $n \times n$ matrix \mathbf{A} :
 - (a) \mathbf{A} is invertible;
 - (b) the reduced row echelon form of \mathbf{A} is \mathbf{I}_n ;
 - (c) \mathbf{A} can be expressed as a product of elementary matrices;
 - (d) \mathbf{A} is row equivalent to \mathbf{I}_n ;
 - (e) $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution;
 - (f) $\mathbf{Ax} = \mathbf{b}$ has exactly one solution for all $\mathbf{b} \in \mathbb{R}^n$;
 - (g) the rows of \mathbf{A} are linearly independent;
 - (h) the columns of \mathbf{A} are linearly independent;
 - (i) $\det \mathbf{A} \neq 0$;
 - (j) 0 is not an eigenvalue of \mathbf{A} .

- The elementary row operations that reduce an invertible matrix \mathbf{A} to the identity matrix can be applied directly to the identity matrix to yield \mathbf{A}^{-1} .

Remark: If $\mathbf{BA} = \mathbf{I}$ then the system of equations $\mathbf{Ax} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{Ix} = \mathbf{BAx} = \mathbf{B(Ax)} = \mathbf{B0} = \mathbf{0}$. Hence \mathbf{A} is invertible and $\mathbf{B} = \mathbf{A}^{-1}$.

4 Determinants

- Determinant of a 2×2 matrix:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

- Determinant of a 3×3 matrix:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}.$$

- *Determinant* of a square matrix $\mathbf{A} = [a_{ij}]_{n \times n}$ can be formally defined as:

$$\det \mathbf{A} = \sum_{\substack{\text{permutations} \\ (j_1, j_2, \dots, j_n) \text{ of } (1, 2, \dots, n)}} \text{sgn}(j_1, j_2, \dots, j_n) a_{1j_1} a_{2j_2} \dots a_{nj_n}.$$

- Properties of Determinants:

- Multiplying a single row or column by a scalar c scales the determinant by c .
- Exchanging two rows or columns swaps the sign of the determinant.
- Adding a multiple of one row (column) to another row (column) leaves the determinant invariant.
- Given an $n \times n$ matrix $\mathbf{A} = [a_{ij}]$, the *minor* M_{ij} of the element a_{ij} is the determinant of the submatrix obtained by deleting the i th row and j th column from \mathbf{A} . The signed minor $(-1)^{i+j} M_{ij}$ is called the *cofactor* of the element a_{ij} .
- Determinants can be evaluated by cofactor expansion (signed minors), either along some row i or some column j :

$$\det[a_{ij}]_{n \times n} = \sum_{k=1}^n a_{ik} M_{ik} = \sum_{k=1}^n a_{kj} M_{kj}.$$

- In general, evaluating a determinant by cofactor expansion is inefficient. However, cofactor expansion allows us to see easily that the determinant of a triangular matrix is simply the product of the diagonal elements!

Problem 4.1: Prove that $\det \mathbf{A}^\top = \det \mathbf{A}$.

Problem 4.2: Let \mathbf{A} and \mathbf{B} be square matrices of the same size. Prove that

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}).$$

Problem 4.3: Prove for an invertible matrix \mathbf{A} that $\det(\mathbf{A}^{-1}) = \frac{1}{\det \mathbf{A}}$.

- Determinants can be used to solve linear systems of equations like

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}.$$

- *Cramer's rule:*

$$x = \frac{\begin{vmatrix} u & b \\ v & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}, \quad y = \frac{\begin{vmatrix} a & u \\ c & v \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}.$$

- For matrices larger than 3×3 , row reduction is more efficient than Cramer's rule.
- An upper (or lower) *triangular system* of equations can be solved directly by back substitution:

$$\begin{aligned} 2x_1 + x_2 &= 5, \\ 4x_2 &= 17. \end{aligned}$$

- The transpose of the matrix of cofactors, $\text{adj } \mathbf{A} = [(-1)^{i+j} M_{ji}]$, is called the *adjugate* (formerly sometimes known as the *adjoint*) of \mathbf{A} .
- The inverse \mathbf{A}^{-1} of a matrix \mathbf{A} may be expressed compactly in terms of its adjugate matrix:

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \text{adj } \mathbf{A}.$$

However, this is an inefficient way to compute the inverse of matrices larger than 3×3 . A better way to compute the inverse of a matrix is to row reduce $[\mathbf{A}|\mathbf{I}]$ to $[\mathbf{I}|\mathbf{A}^{-1}]$.

- The *cross product* of two vectors $\mathbf{u} = (u_x, u_y, u_z)$ and $\mathbf{v} = (v_x, v_y, v_z)$ in \mathbb{R}^3 can be expressed as the determinant

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{vmatrix} = (u_y v_z - u_z v_y)\mathbf{i} + (u_z v_x - u_x v_z)\mathbf{j} + (u_x v_y - u_y v_x)\mathbf{k}.$$

- The magnitude $|\mathbf{u} \times \mathbf{v}| = |\mathbf{u}||\mathbf{v}|\sin\theta$ (where θ is the angle between the vectors \mathbf{u} and \mathbf{v}) represents the area of the parallelogram formed by \mathbf{u} and \mathbf{v} .

5 Eigenvalues and Eigenvectors

- A *fixed point* \mathbf{x} of a matrix \mathbf{A} satisfies $\mathbf{A}\mathbf{x} = \mathbf{x}$.
- The zero vector is a fixed point of any matrix.
- An *eigenvector* \mathbf{x} of a matrix \mathbf{A} is a **nonzero** vector \mathbf{x} that satisfies $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for a particular value of λ , known as an *eigenvalue* (characteristic value).
- An eigenvector \mathbf{x} is a nontrivial solution to $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{A}\mathbf{x} - \lambda\mathbf{x} = 0$. This requires that $(\mathbf{A} - \lambda\mathbf{I})$ be noninvertible; that is, λ must be a root of the *characteristic polynomial* $\det(\lambda\mathbf{I} - \mathbf{A})$: it must satisfy the *characteristic equation* $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$.
- If $[a_{ij}]_{n \times n}$ is a triangular matrix, each of the n diagonal entries a_{ii} are eigenvalues of \mathbf{A} :

$$\det(\lambda\mathbf{I} - \mathbf{A}) = (\lambda - a_{11})(\lambda - a_{22}) \dots (\lambda - a_{nn}).$$

- If \mathbf{A} has eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ then $\det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$. On setting $\lambda = 0$, we see that $\det \mathbf{A}$ is the product $\lambda_1 \lambda_2 \dots \lambda_n$ of the eigenvalues of \mathbf{A} .
- Given an $n \times n$ matrix \mathbf{A} , the terms in $\det(\lambda\mathbf{I} - \mathbf{A})$ that involve λ^n and λ^{n-1} must come from the main diagonal:

$$(\lambda - a_{11})(\lambda - a_{22}) \dots (\lambda - a_{nn}) = \lambda^n - (a_{11} + a_{22} + \dots + a_{nn})\lambda^{n-1} + \dots$$

But

$$\det(\lambda\mathbf{I} - \mathbf{A}) = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n) = \lambda^n - (\lambda_1 + \lambda_2 + \dots + \lambda_n)\lambda^{n-1} + \dots + (-1)^n \lambda_1 \lambda_2 \dots \lambda_n.$$

Thus, $\text{Tr } \mathbf{A}$ is the sum $\lambda_1 + \lambda_2 + \dots + \lambda_n$ of the eigenvalues of \mathbf{A} and the characteristic polynomial of \mathbf{A} always has the form

$$\lambda^n - \text{Tr } \mathbf{A} \lambda^{n-1} + \dots + (-1)^n \det \mathbf{A}.$$

Problem 5.1: Show that the eigenvalues and corresponding eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$$

are -1 , with eigenvector $[1, -1]$, and 4 , with eigenvector $[2, 3]$. Note that the trace (3) equals the sum of the eigenvalues and the determinant (-4) equals the product of eigenvalues. (Any nonzero multiple of the given eigenvectors are also acceptable eigenvectors.)

Problem 5.2: Compute $\mathbf{A}\mathbf{v}$ where \mathbf{A} is the matrix in Problem 5.1 and $\mathbf{v} = [7, 3]^\top$, both directly and by expressing \mathbf{v} as a linear combination of the eigenvectors of \mathbf{A} : $[7, 3] = 3[1, -1] + 2[2, 3]$.

Problem 5.3: Show that the trace and determinant of a 5×5 matrix whose characteristic polynomial is $\det(\lambda\mathbf{I} - \mathbf{A}) = \lambda^5 + 2\lambda^4 + 3\lambda^3 + 4\lambda^2 + 5\lambda + 6$ are given by -2 and -6 , respectively.

- The *algebraic multiplicity* of an eigenvalue λ_0 of a matrix \mathbf{A} is the number of factors of $(\lambda - \lambda_0)$ in the characteristic polynomial $\det(\lambda\mathbf{I} - \mathbf{A})$.
- The *eigenspace* of \mathbf{A} corresponding to an eigenvalue λ is the linear space spanned by all eigenvectors of \mathbf{A} associated with λ . This is the null space of $\mathbf{A} - \lambda\mathbf{I}$.
- The *geometric multiplicity* of an eigenvalue λ of a matrix \mathbf{A} is the dimension of the *eigenspace* of \mathbf{A} corresponding to λ .
- The sum of the algebraic multiplicities of all eigenvalues of an $n \times n$ matrix is equal to n .
- The geometric multiplicity of an eigenvalue is always less than or equal to its algebraic multiplicity.

Problem 5.4: Show that the 2×2 identity matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

has a double eigenvalue of 1 associated with two linearly independent eigenvectors, say $[1, 0]^\top$ and $[0, 1]^\top$. The eigenvalue 1 thus has algebraic multiplicity two and geometric multiplicity two.

Problem 5.5: Show that the matrix

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

has a double eigenvalue of 1 but only a single eigenvector $[1, 0]^T$. The eigenvalue 1 thus has algebraic multiplicity two, while its geometric multiplicity is only one.

Problem 5.6: Show that the matrix

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 3 & 0 \\ -3 & 5 & 3 \end{bmatrix}$$

has [Anton and Busby, p. 458]:

- (a) an eigenvalue 2 with algebraic and geometric multiplicity one;
- (b) an eigenvalue 3 with algebraic multiplicity two and geometric multiplicity one.

6 Linear Transformations

- A mapping $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a *linear transformation* from \mathbb{R}^n to \mathbb{R}^m if for all vectors \mathbf{u} and \mathbf{v} in \mathbb{R}^n and all scalars c :

(a) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$;

(b) $T(c\mathbf{u}) = cT(\mathbf{u})$.

If $m = n$, we say that T is a *linear operator* on \mathbb{R}^n .

- The action of a linear transformation on a vector can be represented by matrix multiplication: $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$.
- An *orthogonal* linear operator T preserves lengths: $|T(\mathbf{x})| = |\mathbf{x}|$ for all vectors \mathbf{x} .
- A square matrix \mathbf{A} with real entries is *orthogonal* if $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$; that is, $\mathbf{A}^{-1} = \mathbf{A}^\top$. This implies that $\det \mathbf{A} = \pm 1$ and preserves lengths:

$$|\mathbf{A}\mathbf{x}|^2 = \mathbf{A}\mathbf{x} \cdot \mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^\top \mathbf{A}\mathbf{x} = \mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{x}^\top \mathbf{x} = \mathbf{x} \cdot \mathbf{x} = |\mathbf{x}|^2.$$

Since the columns of an orthogonal matrix are unit vectors in addition to being mutually orthogonal, they can also be called *orthonormal* matrices.

- The matrix that describes *rotation* about an angle θ is orthogonal:

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

- The *kernel* $\ker T$ of a linear transformation T is the subspace consisting of all vectors \mathbf{u} that are mapped to $\mathbf{0}$ by T .
- The *null space* $\text{null } \mathbf{A}$ (also known as the *kernel* $\ker \mathbf{A}$) of a matrix \mathbf{A} is the subspace consisting of all vectors \mathbf{u} such that $\mathbf{A}\mathbf{u} = \mathbf{0}$.
- Linear transformations map subspaces to subspaces.
- The *range* $\text{ran } T$ of a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the image of the *domain* \mathbb{R}^n under T .
- A linear transformation T is *one-to-one* if $T(\mathbf{u}) = T(\mathbf{v}) \Rightarrow \mathbf{u} = \mathbf{v}$.

Problem 6.1: Show that a linear transformation is *one-to-one* if and only if $\ker T = \{\mathbf{0}\}$.

- A linear operator T on \mathbb{R}^n is *one-to-one* if and only if it is *onto*.
- A set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of linearly independent vectors is called a *basis* for the n -dimensional subspace

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n : (c_1, c_2, \dots, c_n) \in \mathbb{R}^n\}.$$

- If $\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n$ is a vector in $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ with basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ we say that $[a_1, a_2, \dots, a_n]^\top$ are the *coordinates* $[\mathbf{w}]_B$ of the vector \mathbf{w} with respect to the basis B .
- If \mathbf{w} is a vector in \mathbb{R}^n the *transition matrix* $\mathbf{P}_{B' \leftarrow B}$ that changes coordinates $[\mathbf{w}]_B$ with respect to the basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ for \mathbb{R}^n to coordinates $[\mathbf{w}]_{B'}$ = $\mathbf{P}_{B' \leftarrow B}[\mathbf{w}]_B$ with respect to the basis $B' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n\}$ for \mathbb{R}^n is

$$\mathbf{P}_{B' \leftarrow B} = [[\mathbf{v}_1]_{B'} \quad [\mathbf{v}_2]_{B'} \quad \dots \quad [\mathbf{v}_n]_{B'}].$$

- An easy way to find the transition matrix $\mathbf{P}_{B' \leftarrow B}$ from a basis B to B' is to use elementary row operations to reduce the matrix $[B'|B]$ to $[\mathbf{I}|\mathbf{P}_{B' \leftarrow B}]$.
- Note that the matrices $\mathbf{P}_{B' \leftarrow B}$ and $\mathbf{P}_{B \leftarrow B'}$ are inverses of each other.
- If $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a linear operator and if $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ and $B' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n\}$ are bases for \mathbb{R}^n then the *matrix representation* $[T]_B$ with respect to the bases B is related to the *matrix representation* $[T]_{B'}$ with respect to the bases B' according to

$$[T]_{B'} = \mathbf{P}_{B' \leftarrow B} [T]_B \mathbf{P}_{B \leftarrow B'}.$$

7 Dimension

- The number of linearly independent rows (or columns) of a matrix \mathbf{A} is known as its *rank* and written $\text{rank } \mathbf{A}$. That is, the rank of a matrix is the dimension $\dim(\text{row}(\mathbf{A}))$ of the span of its rows. Equivalently, the rank of a matrix is the dimension $\dim(\text{col}(\mathbf{A}))$ of the span of its columns.
- The dimension of the null space of a matrix \mathbf{A} is known as its *nullity* and written $\dim(\text{null}(\mathbf{A}))$ or *nullity* \mathbf{A} .

- The *dimension theorem* states that if \mathbf{A} is an $m \times n$ matrix then

$$\text{rank } \mathbf{A} + \text{nullity } \mathbf{A} = n.$$

- If \mathbf{A} is an $m \times n$ matrix then

$$\begin{aligned} \dim(\text{row}(\mathbf{A})) &= \text{rank } \mathbf{A} & \dim(\text{null}(\mathbf{A})) &= n - \text{rank } \mathbf{A}, \\ \dim(\text{col}(\mathbf{A})) &= \text{rank } \mathbf{A} & \dim(\text{null}(\mathbf{A}^\top)) &= m - \text{rank } \mathbf{A}. \end{aligned}$$

Problem 7.1: Find $\dim(\text{row})$ and $\dim(\text{null})$ for both \mathbf{A} and \mathbf{A}^\top :

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}.$$

- While the row space of a matrix is “invariant to” (unchanged by) elementary row operations, the column space is not. For example, consider the row-equivalent matrices

$$\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Note that $\text{row}(\mathbf{A}) = \text{span}([1, 0]) = \text{row}(\mathbf{B})$. However $\text{col}(\mathbf{A}) = \text{span}([0, 1]^\top)$ but $\text{col}(\mathbf{B}) = \text{span}([1, 0]^\top)$.

- The following are equivalent for an $m \times n$ matrix \mathbf{A} :
 - (a) \mathbf{A} has rank k ;
 - (b) \mathbf{A} has nullity $n - k$;
 - (c) Every row echelon form of \mathbf{A} has k nonzero rows and $m - k$ zero rows;
 - (d) The homogeneous system $\mathbf{A}\mathbf{x} = 0$ has k pivot variables and $n - k$ free variables.

8 Similarity and Diagonalizability

- If \mathbf{A} and \mathbf{B} are square matrices of the same size and $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ for some invertible matrix \mathbf{P} , we say that \mathbf{B} is *similar* to \mathbf{A} .

Problem 8.1: If \mathbf{B} is similar to \mathbf{A} , show that \mathbf{A} is also similar to \mathbf{B} .

Problem 8.2: Suppose $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. Multiplying \mathbf{A} by the matrix \mathbf{P}^{-1} on the left certainly cannot increase its rank (the number of linearly independent rows or columns), so $\text{rank } \mathbf{P}^{-1}\mathbf{A} \leq \text{rank } \mathbf{A}$. Likewise $\text{rank } \mathbf{B} = \text{rank } \mathbf{P}^{-1}\mathbf{A}\mathbf{P} \leq \text{rank } \mathbf{P}^{-1}\mathbf{A} \leq \text{rank } \mathbf{A}$. But since we also know that \mathbf{A} is similar to \mathbf{B} , we see that $\text{rank } \mathbf{A} \leq \text{rank } \mathbf{B}$. Hence $\text{rank } \mathbf{A} = \text{rank } \mathbf{B}$.

Problem 8.3: Show that similar matrices also have the same nullity.

- Similar matrices represent the same linear transformation under the *change of bases* described by the matrix \mathbf{P} .

Problem 8.4: Prove that similar matrices have the following *similarity invariants*:

- rank;
- nullity;
- characteristic polynomial and eigenvalues, including their algebraic and geometric multiplicities;
- determinant;
- trace.

Proof of (c): Suppose $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. We first note that

$$\lambda\mathbf{I} - \mathbf{B} = \lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{P}^{-1}\lambda\mathbf{I}\mathbf{P} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{P}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{P}.$$

The matrices $\lambda\mathbf{I} - \mathbf{B}$ and $\lambda\mathbf{I} - \mathbf{A}$ are thus similar and share the same nullity (and hence geometric multiplicity) for each eigenvalue λ . Moreover, this implies that the matrices \mathbf{A} and \mathbf{B} share the same characteristic polynomial (and hence eigenvalues and algebraic multiplicities):

$$\det(\lambda\mathbf{I} - \mathbf{B}) = \det(\mathbf{P}^{-1}) \det(\lambda\mathbf{I} - \mathbf{A}) \det(\mathbf{P}) = \frac{1}{\det(\mathbf{P})} \det(\lambda\mathbf{I} - \mathbf{A}) \det(\mathbf{P}) = \det(\lambda\mathbf{I} - \mathbf{A}).$$

- A matrix is said to be *diagonalizable* if it is similar to a diagonal matrix.

- An $n \times n$ matrix \mathbf{A} is diagonalizable if and only if \mathbf{A} has n linearly independent eigenvectors. Suppose that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$, where

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Since \mathbf{P} is invertible, its columns are nonzero and linearly independent. Moreover,

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D} = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix}.$$

That is, for $j = 1, 2, \dots, n$:

$$\mathbf{A} \begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix} = \lambda_j \begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix}.$$

This says, for each j , that the nonzero column vector $[p_{1j}, p_{2j}, \dots, p_{nj}]^T$ is an eigenvector of \mathbf{A} with eigenvalue λ_j . Moreover, as mentioned above, these n column vectors are linearly independent. This established one direction of the claim. On reversing this argument, we see that if \mathbf{A} has n linearly independent eigenvectors, we can use them to form an *eigenvector matrix* \mathbf{P} such that $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$.

- Equivalently, an $n \times n$ matrix is diagonalizable if and only if the geometric multiplicity of each eigenvalue is the same as its algebraic multiplicity (so that the sum of the geometric multiplicities of all eigenvalues is n).

Problem 8.5: Show that the matrix

$$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

has a double eigenvalue of λ but only one eigenvector $[1, 0]^T$ (i.e. the eigenvalue λ has algebraic multiplicity two but geometric multiplicity one) and consequently is *not* diagonalizable.

- *Eigenvectors of a matrix \mathbf{A} associated with distinct eigenvalues are linearly independent.* If not, then one of them would be expressible as a linear combination of the others. Let us order the eigenvectors so that \mathbf{v}_{k+1} is the first eigenvector that is expressible as a linear combination of the others:

$$\mathbf{v}_{k+1} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k, \quad (1)$$

where the coefficients c_i are not all zero. The condition that \mathbf{v}_{k+1} is the first such vector guarantees that the vectors on the right-hand side are linearly independent. If each eigenvector \mathbf{v}_i corresponds to the eigenvalue λ_i , then on multiplying by \mathbf{A} on the left, we find that

$$\lambda_{k+1}\mathbf{v}_{k+1} = \mathbf{A}\mathbf{v}_{k+1} = c_1\mathbf{A}\mathbf{v}_1 + c_2\mathbf{A}\mathbf{v}_2 + \cdots + c_k\mathbf{A}\mathbf{v}_k = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_k\lambda_k\mathbf{v}_k.$$

If we multiply Eq.(1) by λ_{k+1} we obtain:

$$\lambda_{k+1}\mathbf{v}_{k+1} = c_1\lambda_{k+1}\mathbf{v}_1 + c_2\lambda_{k+1}\mathbf{v}_2 + \cdots + c_k\lambda_{k+1}\mathbf{v}_k,$$

The difference of the previous two equations yields

$$\mathbf{0} = c_1(\lambda_1 - \lambda_{k+1})\mathbf{v}_1 + (\lambda_2 - \lambda_{k+1})c_2\mathbf{v}_2 + \cdots + (\lambda_k - \lambda_{k+1})c_k\mathbf{v}_k,$$

Since the vectors on the right-hand side are linearly independent, we know that each of the coefficients $c_i(\lambda_i - \lambda_{k+1})$ must vanish. But this is not possible since the coefficients c_i are nonzero and eigenvalues are distinct. The only way to escape this glaring contradiction is that all of the eigenvectors of \mathbf{A} corresponding to distinct eigenvalues must in fact be independent!

- An $n \times n$ matrix \mathbf{A} with n distinct eigenvalues is diagonalizable.
- However, the converse is not true (consider the identity matrix). In the words of Anton & Busby, “the key to diagonalizability rests with the dimensions of the eigenspaces,” *not* with the distinctness of the eigenvalues.

Problem 8.6: Show that the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$$

from Problem 5.1 is diagonalizable by evaluating

$$\frac{1}{5} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}.$$

What do you notice about the resulting product and its diagonal elements? What is the significance of each of the above matrices and the factor $1/5$ in front? Show that \mathbf{A} may be decomposed as

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}.$$

Problem 8.7: Show that the characteristic polynomial of the matrix that describes rotation in the x - y plane by $\theta = 90^\circ$:

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

is $\lambda^2 + 1 = 0$. This equation has no real roots, so \mathbf{A} is not diagonalizable over the real numbers. However, it is in fact diagonalizable over a broader number system known as the *complex numbers* \mathbb{C} : we will see that the characteristic equation $\lambda^2 + 1 = 0$ admits two distinct complex eigenvalues, namely $-i$ and i , where i is an *imaginary number* that satisfies $i^2 + 1 = 0$.

Remark: An efficient way to compute high powers and fractional (or even negative powers) of a matrix \mathbf{A} is to first diagonalize it; that is, using the eigenvector matrix \mathbf{P} to express $\mathbf{A} = \mathbf{PDP}^{-1}$ where \mathbf{D} is the diagonal matrix of eigenvalues.

Problem 8.8: To find the 5th power of \mathbf{A} note that

$$\begin{aligned} \mathbf{A}^5 &= (\mathbf{PDP}^{-1})(\mathbf{PDP}^{-1})(\mathbf{PDP}^{-1})(\mathbf{PDP}^{-1})(\mathbf{PDP}^{-1}) \\ &= \mathbf{PDP}^{-1}\mathbf{PDP}^{-1}\mathbf{PDP}^{-1}\mathbf{PDP}^{-1}\mathbf{PDP}^{-1} \\ &= \mathbf{PD}^5\mathbf{P}^{-1} \\ &= \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} (-1)^5 & 0 \\ 0 & 4^5 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1024 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} -1 & 2048 \\ 1 & 3072 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 2045 & 2050 \\ 3075 & 3070 \end{bmatrix} \\ &= \begin{bmatrix} 409 & 410 \\ 615 & 614 \end{bmatrix}. \end{aligned}$$

Problem 8.9: Check the result of the previous problem by manually computing the product \mathbf{A}^5 . Which way is easier for computing high powers of a matrix?

Remark: We can use the same technique to find the square root $\mathbf{A}^{\frac{1}{2}}$ of \mathbf{A} (i.e. a matrix \mathbf{B} such that $\mathbf{B}^2 = \mathbf{A}$, we will again need to work in the complex number system \mathbb{C} , where $i^2 = -1$):

$$\begin{aligned}\mathbf{A}^{\frac{1}{2}} &= \mathbf{P}\mathbf{D}^{\frac{1}{2}}\mathbf{P}^{-1} \\ &= \frac{1}{5} \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} i & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} i & 4 \\ -i & 6 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 4 + 3i & 4 - 2i \\ 6 - 3i & 6 + 2i \end{bmatrix}.\end{aligned}$$

Then

$$\mathbf{A}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}} = \left(\mathbf{P}\mathbf{D}^{\frac{1}{2}}\mathbf{P}^{-1}\right)\left(\mathbf{P}\mathbf{D}^{\frac{1}{2}}\mathbf{P}^{-1}\right) = \mathbf{P}\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}\mathbf{P}^{-1} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \mathbf{A},$$

as desired.

Problem 8.10: Verify explicitly that

$$\frac{1}{25} \begin{bmatrix} 4 + 3i & 4 - 2i \\ 6 - 3i & 6 + 2i \end{bmatrix} \begin{bmatrix} 4 + 3i & 4 - 2i \\ 6 - 3i & 6 + 2i \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} = \mathbf{A}.$$

9 Complex Numbers

In order to diagonalize a matrix \mathbf{A} with linearly independent eigenvectors (such as a matrix with distinct eigenvalues), we first need to solve for the roots of the characteristic polynomial $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$. These roots form the n elements of the diagonal matrix that \mathbf{A} is similar to. However, as we have already seen, a polynomial of degree n does not necessarily have n real roots.

Recall that z is a *root* of the polynomial $P(x)$ if $P(z) = 0$.

Q. Do all polynomials have at least one root $z \in \mathbb{R}$?

A. No, consider $P(x) = x^2 + 1$. It has no real roots: $P(x) \geq 1$ for all x .

The *complex numbers* \mathbb{C} are introduced precisely to circumvent this problem. If we replace “ $z \in \mathbb{R}$ ” by “ $z \in \mathbb{C}$ ”, we can answer the above question affirmatively.

The complex numbers consist of ordered pairs (x, y) together with the usual component-by-component addition rule (e.g. which one has in a vector space)

$$(x, y) + (u, v) = (x + u, y + v),$$

but with the unusual multiplication rule

$$(x, y) \cdot (u, v) = (xu - yv, xv + yu).$$

Note that this multiplication rule is associative, commutative, and distributive. Since

$$(x, 0) + (u, 0) = (x + u, 0) \quad \text{and} \quad (x, 0) \cdot (u, 0) = (xu, 0),$$

we see that $(x, 0)$ and $(u, 0)$ behave just like the real numbers x and u . In fact, we can map $(x, 0) \in \mathbb{C}$ to $x \in \mathbb{R}$:

$$(x, 0) \equiv x.$$

Hence $\mathbb{R} \subset \mathbb{C}$.

Remark: We see that the complex number $z = (0, 1)$ satisfies the equation $z^2 + 1 = 0$. That is, $(0, 1) \cdot (0, 1) = (-1, 0)$.

- Denote $(0, 1)$ by the letter i . Then any complex number (x, y) can be represented as $(x, 0) + (0, 1)(y, 0) = x + iy$.

Remark: Unlike \mathbb{R} , the set $\mathbb{C} = \{(x, y) : x \in \mathbb{R}, y \in \mathbb{R}\}$ is not *ordered*; there is no notion of positive and negative (greater than or less than) on the complex plane. For example, if i were positive or zero, then $i^2 = -1$ would have to be positive or zero. If i were negative, then $-i$ would be positive, which would imply that $(-i)^2 = i^2 = -1$ is positive. It is thus not possible to divide the complex numbers into three classes of negative, zero, and positive numbers.

Remark: The frequently appearing notation $\sqrt{-1}$ for i is misleading and should be avoided, because the rule $\sqrt{xy} = \sqrt{x}\sqrt{y}$ (which one might anticipate) does not hold for negative x and y , as the following contradiction illustrates:

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = i^2 = -1.$$

Furthermore, by definition $\sqrt{x} \geq 0$, but one cannot write $i \geq 0$ since \mathbb{C} is not ordered.

Remark: We may write $(x, 0) = x + i0 = x$ since $i0 = (0, 1) \cdot (0, 0) = (0, 0) = 0$.

- The complex conjugate $\overline{(x, y)}$ of (x, y) is $(x, -y)$. That is,

$$\overline{x + iy} = x - iy.$$

- The *complex modulus* $|z|$ of $z = x + iy$ is given by $\sqrt{x^2 + y^2}$.

Remark: If $z \in \mathbb{R}$ then $|z| = \sqrt{z^2}$ is just the absolute value of z .

We now establish some important properties of the complex conjugate. Let $z = x + iy$ and $w = u + iv$ be elements of \mathbb{C} . Then

(i)
$$z\bar{z} = (x, y)(x, -y) = (x^2 + y^2, yx - xy) = (x^2 + y^2, 0) = x^2 + y^2 = |z|^2,$$

(ii)
$$\overline{z + w} = \bar{z} + \bar{w},$$

(iii)
$$\overline{z\bar{w}} = \bar{z}w.$$

Problem 9.1: Prove properties (ii) and (iii).

Remark: Property (i) provides an easy way to compute *reciprocals* of complex numbers:

$$\frac{1}{z} = \frac{\bar{z}}{z\bar{z}} = \frac{\bar{z}}{|z|^2}.$$

Remark: Properties (i) and (iii) imply that

$$|zw|^2 = zw\bar{z}\bar{w} = z\bar{z}w\bar{w} = |z|^2|w|^2.$$

Thus $|zw| = |z||w|$ for all $z, w \in \mathbb{C}$.

- The *dot product* of vectors $\mathbf{u} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$ and $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ in \mathbb{C}^n is given by

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \bar{\mathbf{v}} = u_1 \bar{v}_1 + \dots + u_n \bar{v}_n.$$

Remark: Note that this definition of the dot product implies that $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}} = \bar{\mathbf{v}}^\top \mathbf{u}$.

Remark: Note however that $\mathbf{u} \cdot \mathbf{u}$ can be written either as $\bar{\mathbf{u}}^\top \mathbf{u}$ or as $\mathbf{u}^\top \bar{\mathbf{u}}$.

Problem 9.2: Prove for $k \in \mathbb{C}$ that $(k\mathbf{u}) \cdot \mathbf{v} = k(\mathbf{u} \cdot \mathbf{v})$ but $\mathbf{u} \cdot (k\mathbf{v}) = \bar{k}(\mathbf{u} \cdot \mathbf{v})$.

- In view of the definition of the complex modulus, it is sensible to define the *length* of a vector $\mathbf{v} = [v_1, v_2, \dots, v_n]$ in \mathbb{C}^n to be

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{\mathbf{v}^\top \bar{\mathbf{v}}} = \sqrt{v_1 \bar{v}_1 + \dots + v_n \bar{v}_n} = \sqrt{|v_1|^2 + \dots + |v_n|^2} \geq 0.$$

Lemma 1 (Complex Conjugate Roots): *Let P be a polynomial with real coefficients. If z is a root of P , then so is \bar{z} .*

Proof: Suppose $P(z) = \sum_{k=0}^n a_k z^k = 0$, where each of the coefficients a_k are real. Then

$$P(\bar{z}) = \sum_{k=0}^n a_k (\bar{z})^k = \sum_{k=0}^n a_k \bar{z}^k = \sum_{k=0}^n \overline{a_k z^k} = \overline{\sum_{k=0}^n a_k z^k} = \overline{P(z)} = \bar{0} = 0.$$

Thus, complex roots of real polynomials occur in *conjugate pairs*, z and \bar{z} .

Remark: Lemma 1 implies that eigenvalues of a matrix \mathbf{A} with real coefficients also occur in conjugate pairs: if λ is an eigenvalue of \mathbf{A} , then so is $\bar{\lambda}$. Moreover, if \mathbf{x} is an eigenvector corresponding to λ , then $\bar{\mathbf{x}}$ is an eigenvector corresponding to $\bar{\lambda}$:

$$\begin{aligned} \mathbf{A}\mathbf{x} = \lambda\mathbf{x} &\Rightarrow \overline{\mathbf{A}\mathbf{x}} = \overline{\lambda\mathbf{x}} \\ &\Rightarrow \mathbf{A}\bar{\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}} \\ &\Rightarrow \mathbf{A}\bar{\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}}. \end{aligned}$$

- A *real* matrix consists only of real entries.

Problem 9.3: Find the eigenvalues and eigenvectors of the real matrix [Anton & Busby, p. 528]

$$\begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix}.$$

- If \mathbf{A} is a real symmetric matrix, it must have real eigenvalues. This follows from the fact that an eigenvalue λ and its eigenvector $\mathbf{x} \neq \mathbf{0}$ must satisfy

$$\begin{aligned} \mathbf{A}\mathbf{x} = \lambda\mathbf{x} &\Rightarrow \bar{\mathbf{x}}^\top \mathbf{A}\mathbf{x} = \bar{\mathbf{x}}^\top \lambda\mathbf{x} = \lambda \mathbf{x} \cdot \mathbf{x} = \lambda |\mathbf{x}|^2 \\ &\Rightarrow \lambda = \frac{\bar{\mathbf{x}}^\top \mathbf{A}\mathbf{x}}{|\mathbf{x}|^2} = \frac{(\mathbf{A}^\top \bar{\mathbf{x}})^\top \mathbf{x}}{|\mathbf{x}|^2} = \frac{(\mathbf{A}\bar{\mathbf{x}})^\top \mathbf{x}}{|\mathbf{x}|^2} = \frac{(\bar{\lambda}\bar{\mathbf{x}})^\top \mathbf{x}}{|\mathbf{x}|^2} = \frac{\bar{\lambda} \mathbf{x} \cdot \mathbf{x}}{|\mathbf{x}|^2} = \bar{\lambda}. \end{aligned}$$

Remark: There is a remarkable similarity between the complex multiplication rule

$$(x, y) \cdot (u, v) = (xu - yv, xv + yu)$$

and the trigonometric angle sum formulae. Notice that

$$\begin{aligned} (\cos \theta, \sin \theta) \cdot (\cos \phi, \sin \phi) &= (\cos \theta \cos \phi - \sin \theta \sin \phi, \cos \theta \sin \phi + \sin \theta \cos \phi) \\ &= (\cos(\theta + \phi), \sin(\theta + \phi)). \end{aligned}$$

That is, multiplication of 2 complex numbers on the unit circle $x^2 + y^2 = 1$ corresponds to addition of their angles of inclination to the x axis. In particular, the mapping $f(z) = z^2$ doubles the angle of $z = (x, y)$ and $f(z) = z^n$ multiplies the angle of z by n .

These statements hold even if z lies on a circle of radius $r \neq 1$,

$$(r \cos \theta, r \sin \theta)^n = r^n (\cos n\theta, \sin n\theta);$$

this is known as *deMoivre's Theorem*.

The power of complex numbers comes from the following important theorem.

Theorem 1 (Fundamental Theorem of Algebra): *Any non-constant polynomial $P(z)$ with complex coefficients has a root in \mathbb{C} .*

Lemma 2 (Polynomial Factors): *If z_0 is a root of a polynomial $P(z)$ then $P(z)$ is divisible by $(z - z_0)$.*

Proof: Suppose z_0 is a root of a polynomial $P(z) = \sum_{k=0}^n a_k z^k$ of degree n . Consider

$$\begin{aligned} P(z) - P(z_0) &= \sum_{k=0}^n a_k z^k - \sum_{k=0}^n a_k z_0^k = \sum_{k=0}^n a_k (z^k - z_0^k) \\ &= \sum_{k=0}^n a_k (z - z_0)(z^{k-1} + z^{k-2}z_0 + \dots + z_0^{k-1}) = (z - z_0)Q(z), \end{aligned}$$

where $Q(z) = \sum_{k=0}^n a_k (z^{k-1} + z^{k-2}z_0 + \dots + z_0^{k-1})$ is a polynomial of degree $n - 1$.

Corollary 1.1 (Polynomial Factorization): Every complex polynomial $P(z)$ of degree $n \geq 0$ has exactly n complex roots z_1, z_2, \dots, z_n and can be factorized as $P(z) = A(z - z_1)(z - z_2) \dots (z - z_n)$, where $A \in \mathbb{C}$.

Proof: Apply Theorem 1 and Lemma 2 recursively n times. (It is conventional to define the degree of the zero polynomial, which has infinitely many roots, to be $-\infty$.)

10 Projection Theorem

- The *component* of \mathbf{v} in a direction $\hat{\mathbf{u}}$ at an angle θ is

$$\mathbf{v} \cdot \hat{\mathbf{u}} = |\mathbf{v}| \cos \theta.$$

- The *orthogonal projection* (*parallel component*) \mathbf{v}_{\parallel} of \mathbf{v} onto a unit vector $\hat{\mathbf{u}}$ is

$$\mathbf{v}_{\parallel} = (\mathbf{v} \cdot \hat{\mathbf{u}}) \hat{\mathbf{u}} = \frac{\mathbf{v} \cdot \mathbf{u}}{|\mathbf{u}|^2} \mathbf{u}.$$

- The *perpendicular component* of \mathbf{v} relative to the direction $\hat{\mathbf{u}}$ is

$$\mathbf{v}_{\perp} = \mathbf{v} - (\mathbf{v} \cdot \hat{\mathbf{u}}) \hat{\mathbf{u}}.$$

The decomposition $\mathbf{v} = \mathbf{v}_{\perp} + \mathbf{v}_{\parallel}$ is unique. Notice that $\mathbf{v}_{\perp} \cdot \hat{\mathbf{u}} = \mathbf{v} \cdot \hat{\mathbf{u}} - \mathbf{v} \cdot \hat{\mathbf{u}} = 0$.

- In fact, for a general n -dimensional subspace W of \mathbb{R}^m , every vector \mathbf{v} has a unique decomposition as a sum of a vector \mathbf{v}_{\parallel} in W and a vector \mathbf{v}_{\perp} in W^{\perp} (this is the set of vectors that are perpendicular to each of the vectors in W). First, find a basis for W . We can write these basis vectors as the columns of an $m \times n$ matrix \mathbf{A} with *full column rank* (linearly independent columns). Then W^{\perp} is the null space for \mathbf{A}^{\top} . The projection \mathbf{v}_{\parallel} of $\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp}$ onto W satisfies $\mathbf{v}_{\parallel} = \mathbf{A}\mathbf{x}$ for some vector \mathbf{x} and hence

$$\mathbf{0} = \mathbf{A}^{\top} \mathbf{v}_{\perp} = \mathbf{A}^{\top} (\mathbf{v} - \mathbf{A}\mathbf{x}).$$

The vector \mathbf{x} thus satisfies the *normal equation*

$$\mathbf{A}^{\top} \mathbf{A} \mathbf{x} = \mathbf{A}^{\top} \mathbf{v}, \tag{2}$$

so that $\mathbf{x} = (\mathbf{A}^{\top} \mathbf{A})^{-1} \mathbf{A}^{\top} \mathbf{v}$. We know from Question 4 of Assignment 1 that $\mathbf{A}^{\top} \mathbf{A}$ shares the same null space as \mathbf{A} . But the null space of \mathbf{A} is $\{\mathbf{0}\}$ since the columns of \mathbf{A} , being basis vectors, are linearly independent. Hence the square matrix $\mathbf{A}^{\top} \mathbf{A}$ is invertible, and there is indeed a unique solution for \mathbf{x} . The *orthogonal projection* \mathbf{v}_{\parallel} of \mathbf{v} onto the subspace W is thus given by the formula

$$\mathbf{v}_{\parallel} = \mathbf{A} (\mathbf{A}^{\top} \mathbf{A})^{-1} \mathbf{A}^{\top} \mathbf{v}.$$

Remark: This complicated projection formula becomes much easier if we first go to the trouble to construct an *orthonormal basis* for W . This is a basis of unit vectors that are mutual orthogonal to one another, so that $\mathbf{A}^{\top} \mathbf{A} = \mathbf{I}$. In this case the projection of \mathbf{v} onto W reduces to multiplication by the symmetric matrix $\mathbf{A}\mathbf{A}^{\top}$:

$$\mathbf{v}_{\parallel} = \mathbf{A}\mathbf{A}^{\top} \mathbf{v}.$$

Problem 10.1: Find the orthogonal projection of $\mathbf{v} = [1, 1, 1]$ onto the plane W spanned by the orthonormal vectors $[0, 1, 0]$ and $[-\frac{4}{5}, 0, \frac{3}{5}]$.

Compute

$$\mathbf{v}_{\parallel} = \mathbf{A}\mathbf{A}^T\mathbf{v} = \begin{bmatrix} 0 & -\frac{4}{5} \\ 1 & 0 \\ 0 & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -\frac{4}{5} & 0 & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{16}{25} & 0 & -\frac{12}{25} \\ 0 & 1 & 0 \\ -\frac{12}{25} & 0 & \frac{9}{5} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{25} \\ 1 \\ -\frac{3}{25} \end{bmatrix}.$$

As desired, we note that $\mathbf{v} - \mathbf{v}_{\parallel} = [\frac{21}{25}, 0, \frac{28}{25}] = \frac{7}{25}[3, 0, 4]$ is parallel to the normal $[\frac{3}{5}, 0, \frac{4}{5}]$ to the plane computed from the cross product of the given orthonormal vectors.

11 Gram-Schmidt Orthonormalization

The *Gram-Schmidt process* provides a systematic procedure for constructing an orthonormal basis for an n -dimensional subspace $\text{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ of \mathbb{R}^m :

1. Start with the first vector: $\mathbf{q}_1 = \mathbf{a}_1$.
2. Remove the component of the second vector parallel to the first:

$$\mathbf{q}_2 = \mathbf{a}_2 - \frac{\mathbf{a}_2 \cdot \mathbf{q}_1}{|\mathbf{q}_1|^2} \mathbf{q}_1.$$

3. Remove the components of the third vector parallel to the first and second:

$$\mathbf{q}_3 = \mathbf{a}_3 - \frac{\mathbf{a}_3 \cdot \mathbf{q}_1}{|\mathbf{q}_1|^2} \mathbf{q}_1 - \frac{\mathbf{a}_3 \cdot \mathbf{q}_2}{|\mathbf{q}_2|^2} \mathbf{q}_2.$$

4. Continue this procedure by successively defining, for $k = 4, 5, \dots, n$:

$$\mathbf{q}_k = \mathbf{a}_k - \frac{\mathbf{a}_k \cdot \mathbf{q}_1}{|\mathbf{q}_1|^2} \mathbf{q}_1 - \frac{\mathbf{a}_k \cdot \mathbf{q}_2}{|\mathbf{q}_2|^2} \mathbf{q}_2 - \dots - \frac{\mathbf{a}_k \cdot \mathbf{q}_{k-1}}{|\mathbf{q}_{k-1}|^2} \mathbf{q}_{k-1}. \quad (3)$$

5. Finally normalize each of the vectors to obtain the orthogonal basis $\{\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \dots, \hat{\mathbf{q}}_n\}$.

Remark: Notice that \mathbf{q}_2 is orthogonal to \mathbf{q}_1 :

$$\mathbf{q}_2 \cdot \mathbf{q}_1 = \mathbf{a}_2 \cdot \mathbf{q}_1 - \frac{\mathbf{a}_2 \cdot \mathbf{q}_1}{|\mathbf{q}_1|^2} |\mathbf{q}_1|^2 = 0.$$

Remark: At the k th stage of the orthonormalization, if all of the previously created vectors were orthogonal to each other, then so is the newly created vector:

$$\mathbf{q}_k \cdot \mathbf{q}_j = \mathbf{a}_k \cdot \mathbf{q}_j - \frac{\mathbf{a}_k \cdot \mathbf{q}_j}{|\mathbf{q}_j|^2} |\mathbf{q}_j|^2 = 0, \quad j = 1, 2, \dots, k-1.$$

Remark: The last two remarks imply that, at each stage of the orthonormalization, all of the vectors created so far are orthogonal to each other!

Remark: Notice that $\mathbf{q}_1 = \mathbf{a}_1$ is a linear combination of the original $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$.

Remark: At the k th stage of the orthonormalization, if all of the previously created vectors were a linear combination of the original vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, then by Eq. (3), we see that \mathbf{q}_k is as well.

Remark: The last two remarks imply that, at each stage of the orthonormalization, all of the vectors created so far are orthogonal to each vector \mathbf{q}_k can be written as a linear combination of the original basis vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. Since these vectors are given to be linearly independent, we thus know that $\mathbf{q}_k \neq \mathbf{0}$. This is what allows us to normalize \mathbf{q}_k in Step 5. It also implies that $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ spans the same space as $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$.

Problem 11.1: What would happen if we were to apply the Gram-Schmidt procedure to a set of vectors that is not linearly independent?

Problem 11.2: Use the Gram-Schmidt process to find an orthonormal basis for the plane $x + y + z = 0$.

In terms of the two parameters (say) $y = s$ and $z = t$, we see that each point on the plane can be expressed as

$$x = -s - t, \quad y = s, \quad z = t.$$

The parameter values $(s, t) = (1, 0)$ and $(s, t) = (0, 1)$ then yield two linearly independent vectors on the plane:

$$\mathbf{a}_1 = [-1, 1, 0] \quad \text{and} \quad \mathbf{a}_2 = [-1, 0, 1].$$

The Gram-Schmidt process then yields

$$\begin{aligned} \mathbf{q}_1 &= [-1, 1, 0], \\ \mathbf{q}_2 &= [-1, 0, 1] - \frac{[-1, 0, 1] \cdot [-1, 1, 0]}{|[-1, 1, 0]|^2} [-1, 1, 0] \\ &= [-1, 0, 1] - \frac{1}{2} [-1, 1, 0] \\ &= \left[-\frac{1}{2}, -\frac{1}{2}, 1 \right]. \end{aligned}$$

We note that $\mathbf{q}_2 \cdot \mathbf{q}_1 = 0$, as desired. Finally, we normalize the vectors to obtain $\hat{\mathbf{q}}_1 = \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \right]$ and $\hat{\mathbf{q}}_2 = \left[-\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}} \right]$.

12 QR Factorization

- We may rewrite the k th step (Eq. 3) of the Gram-Schmidt orthonormalization process more simply in terms of the unit normal vectors $\hat{\mathbf{q}}_j$:

$$\mathbf{q}_k = \mathbf{a}_k - (\mathbf{a}_k \cdot \hat{\mathbf{q}}_1)\hat{\mathbf{q}}_1 - (\mathbf{a}_k \cdot \hat{\mathbf{q}}_2)\hat{\mathbf{q}}_2 - \cdots - (\mathbf{a}_k \cdot \hat{\mathbf{q}}_{k-1})\hat{\mathbf{q}}_{k-1}. \quad (4)$$

On taking the dot product with $\hat{\mathbf{q}}_k$ we find, using the orthogonality of the vectors $\hat{\mathbf{q}}_j$,

$$\mathbf{0} \neq \mathbf{q}_k \cdot \hat{\mathbf{q}}_k = \mathbf{a}_k \cdot \hat{\mathbf{q}}_k \quad (5)$$

so that $\mathbf{q}_k = (\mathbf{q}_k \cdot \hat{\mathbf{q}}_k)\hat{\mathbf{q}}_k = (\mathbf{a}_k \cdot \hat{\mathbf{q}}_k)\hat{\mathbf{q}}_k$. Equation 4 may then be rewritten as

$$\mathbf{a}_k = \hat{\mathbf{q}}_1(\mathbf{a}_k \cdot \hat{\mathbf{q}}_1) + \hat{\mathbf{q}}_2(\mathbf{a}_k \cdot \hat{\mathbf{q}}_2) + \cdots + \hat{\mathbf{q}}_{k-1}(\mathbf{a}_k \cdot \hat{\mathbf{q}}_{k-1}) + \hat{\mathbf{q}}_k(\mathbf{a}_k \cdot \hat{\mathbf{q}}_k).$$

On varying k from 1 to n we obtain the following set of n equations:

$$[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n] = [\hat{\mathbf{q}}_1 \quad \hat{\mathbf{q}}_2 \quad \cdots \quad \hat{\mathbf{q}}_n] \begin{bmatrix} \mathbf{a}_1 \cdot \hat{\mathbf{q}}_1 & \mathbf{a}_2 \cdot \hat{\mathbf{q}}_1 & \cdots & \mathbf{a}_n \cdot \hat{\mathbf{q}}_1 \\ 0 & \mathbf{a}_2 \cdot \hat{\mathbf{q}}_2 & \cdots & \mathbf{a}_n \cdot \hat{\mathbf{q}}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{a}_n \cdot \hat{\mathbf{q}}_n \end{bmatrix}.$$

If we denote the upper triangular $n \times n$ matrix on the right-hand-side by \mathbf{R} , and the matrices composed of the column vectors \mathbf{a}_k and $\hat{\mathbf{q}}_k$ by the $m \times n$ matrices \mathbf{A} and \mathbf{Q} , respectively, we can express this result as the so-called **QR factorization** of \mathbf{A} :

$$\mathbf{A} = \mathbf{QR}. \quad (6)$$

Remark: Every matrix $m \times n$ matrix \mathbf{A} with full column rank (linearly independent columns) thus has a **QR** factorization.

Remark: Equation 5 implies that each of the diagonal elements of \mathbf{R} is nonzero. This guarantees that \mathbf{R} is invertible.

Remark: If the columns of \mathbf{Q} are orthonormal, then $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$. An efficient way to find \mathbf{R} is to premultiply both sides of Eq. 6 by \mathbf{Q}^T :

$$\mathbf{Q}^T \mathbf{A} = \mathbf{R}.$$

Problem 12.1: Find the **QR** factorization of [Anton & Busby, p. 418]

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

13 Least Squares Approximation

Suppose that a set of data (x_i, y_i) for $i = 1, 2, \dots, n$ is measured in an experiment and that we wish to test how well it fits an *affine* relationship of the form

$$y = \alpha + \beta x.$$

Here α and β are parameters that we wish to vary to achieve the best fit.

If each data point (x_i, y_i) happens to fall on the line $y = \alpha + \beta x$, then the unknown parameters α and β can be determined from the matrix equation

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \quad (7)$$

If the x_i s are distinct, then the $n \times 2$ coefficient matrix \mathbf{A} on the left-hand side, known as the *design matrix*, has full column rank (two). Also note in the case of exact agreement between the experimental data and the theoretical model $y = \alpha + \beta x$ that the vector \mathbf{b} on the right-hand side is in the column space of \mathbf{A} .

- Recall that if \mathbf{b} is in the column space of \mathbf{A} , the linear system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ is *consistent*, that is, it has at least one solution \mathbf{x} . Recall that this solution is unique iff \mathbf{A} has full column rank.
- In practical applications, however, there may be measurement errors in the entries of \mathbf{A} and \mathbf{b} so that \mathbf{b} does not lie exactly in the column space of \mathbf{A} .
- The *least squares* approximation to an inconsistent system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ is the solution to

$$\mathbf{A}\mathbf{x} = \mathbf{b}_{\parallel}$$

where \mathbf{b}_{\parallel} denotes the projection of \mathbf{b} onto the column space of \mathbf{A} . From Eq. (2) we see that the solution vector \mathbf{x} satisfies the normal equation

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}.$$

Remark: For the normal equation Eq. (7), we see that

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

and

$$\mathbf{A}^T \mathbf{b} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix},$$

where the sums are computed from $i = 1$ to n . Thus the solution to the least-squares fit of Eq. (7) is given by

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}.$$

Problem 13.1: Show that the least squares line of best fit to the measured data $(0, 1)$, $(1, 3)$, $(2, 4)$, and $(3, 4)$ is $y = 1.5 + x$.

- The difference $\mathbf{b}_\perp = \mathbf{b} - \mathbf{b}_\parallel$ is called the *least squares error vector*.
- The least squares method is sometimes called *linear regression* and the line $y = \alpha + \beta x$ can either be called the *least squares line of best fit* or the *regression line*.
- For each data pair (x_i, y_i) , the difference $y_i - (\alpha + \beta x_i)$ is called the *residual*.

Remark: Since the least-squares solution $\mathbf{x} = [\alpha, \beta]$ minimizes the least squares error vector

$$\mathbf{b}_\perp = \mathbf{b} - \mathbf{A}\mathbf{x} = [y_1 - (\alpha + \beta x_1), \dots, y_n - (\alpha + \beta x_n)],$$

the method effectively minimizes the sum $\sum_{i=1}^n [y_i - (\alpha + \beta x_i)]^2$ of the squares of the residuals.

- A numerically robust way to solve the normal equation $\mathbf{A}^T \mathbf{A}\mathbf{x} = \mathbf{A}^T \mathbf{b}$ is to factor $\mathbf{A} = \mathbf{QR}$:

$$\mathbf{R}^T \mathbf{Q}^T \mathbf{Q} \mathbf{R} \mathbf{x} = \mathbf{R}^T \mathbf{Q}^T \mathbf{b}.$$

which, since \mathbf{R} is invertible, simplifies to

$$\mathbf{x} = \mathbf{R}^{-1} \mathbf{Q}^T \mathbf{b}.$$

Problem 13.2: Show that for the same measured data as before, $(0, 1)$, $(1, 3)$, $(2, 4)$, and $(3, 4)$, that

$$\mathbf{R} = \begin{bmatrix} 2 & 3 \\ 0 & \sqrt{5} \end{bmatrix}$$

and from this that the least squares solution again is $y = 1.5 + x$.

Remark: The least squares method can be extended to fit a polynomial

$$y = a_0 + a_1x + \cdots + a_mx^m$$

as close as possible to n given data points (x_i, y_i) :

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Here, the design matrix takes the form of a *Vandermonde* matrix, which has special properties. For example, in the case where $m = n - 1$, its determinant is given by $\prod_{\substack{i,j=1 \\ i>j}}^n (x_i - x_j)$. If the n x_i values are distinct, this determinant is nonzero. The system of equations is consistent and has a unique solution, known as the *Lagrange interpolating polynomial*. This polynomial passes exactly through the given data points. However, if $m < n - 1$, it will usually not be possible to find a polynomial that passes through the given data points and we must find the least squares solution by solving the normal equation for this problem [cf. Example 7 on p. 403 of Anton & Busby].

14 Orthogonal (Unitary) Diagonalizability

- A matrix is said to be *Hermitian* if $\mathbf{A}^\top = \overline{\mathbf{A}}$.
- For real matrices, being Hermitian is equivalent to being symmetric.
- On defining the *Hermitian transpose* $\mathbf{A}^\dagger = \overline{\mathbf{A}^\top}$, we see that a Hermitian matrix \mathbf{A} obeys $\mathbf{A}^\dagger = \mathbf{A}$.
- For real matrices, the Hermitian transpose is equivalent to the usual transpose.
- A complex matrix \mathbf{P} is *orthogonal* (or *orthonormal*) if $\mathbf{P}^\dagger \mathbf{P} = \mathbf{I}$. (Most books use the term *unitary* here but in view of how the dot product of two complex vectors is defined, there is really no need to introduce new terminology.)
- For real matrices, an orthogonal matrix \mathbf{A} obeys $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$.

- If \mathbf{A} and \mathbf{B} are square matrices of the same size and $\mathbf{B} = \mathbf{P}^\dagger \mathbf{A} \mathbf{P}$ for some orthogonal matrix \mathbf{P} , we say that \mathbf{B} is *orthogonally similar* to \mathbf{A} .

Problem 14.1: Show that two orthogonally similar matrices \mathbf{A} and \mathbf{B} are similar.

- Orthogonally similar matrices represent the same linear transformation under the *orthogonal change of bases* described by the matrix \mathbf{P} .
- A matrix is said to be *orthogonally diagonalizable* if it is orthogonally similar to a diagonal matrix.

Remark: An $n \times n$ matrix \mathbf{A} is orthogonally diagonalizable if and only if it has an orthonormal set of n eigenvectors. But which matrices have this property? The following definition will help us answer this important question.

- A matrix \mathbf{A} that commutes with its Hermitian transpose, so that $\mathbf{A}^\dagger \mathbf{A} = \mathbf{A} \mathbf{A}^\dagger$, is said to be *normal*.

Problem 14.2: Show that every diagonal matrix is normal.

Problem 14.3: Show that every Hermitian matrix is normal.

- An orthogonally diagonalizable matrix must be normal. Suppose

$$\mathbf{D} = \mathbf{P}^\dagger \mathbf{A} \mathbf{P}$$

for some diagonal matrix \mathbf{D} and orthogonal matrix \mathbf{P} . Then

$$\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^\dagger,$$

and

$$\mathbf{A}^\dagger = (\mathbf{P} \mathbf{D} \mathbf{P}^\dagger)^\dagger = \mathbf{P} \mathbf{D}^\dagger \mathbf{P}^\dagger.$$

Then $\mathbf{A} \mathbf{A}^\dagger = \mathbf{P} \mathbf{D} \mathbf{P}^\dagger \mathbf{P} \mathbf{D}^\dagger \mathbf{P}^\dagger = \mathbf{P} \mathbf{D} \mathbf{D}^\dagger \mathbf{P}^\dagger$ and $\mathbf{A}^\dagger \mathbf{A} = \mathbf{P} \mathbf{D}^\dagger \mathbf{P}^\dagger \mathbf{P} \mathbf{D} \mathbf{P}^\dagger = \mathbf{P} \mathbf{D}^\dagger \mathbf{D} \mathbf{P}^\dagger$. But $\mathbf{D}^\dagger \mathbf{D} = \mathbf{D} \mathbf{D}^\dagger$ since \mathbf{D} is diagonal. Therefore $\mathbf{A}^\dagger \mathbf{A} = \mathbf{A} \mathbf{A}^\dagger$.

- If \mathbf{A} is a normal matrix and is orthogonally similar to \mathbf{U} , then \mathbf{U} is also normal:

$$\mathbf{U}^\dagger \mathbf{U} = (\mathbf{P}^\dagger \mathbf{A} \mathbf{P})^\dagger \mathbf{P}^\dagger \mathbf{A} \mathbf{P} = \mathbf{P}^\dagger \mathbf{A}^\dagger \mathbf{P} \mathbf{P}^\dagger \mathbf{A} \mathbf{P} = \mathbf{P}^\dagger \mathbf{A}^\dagger \mathbf{A} \mathbf{P} = \mathbf{P}^\dagger \mathbf{A} \mathbf{A}^\dagger \mathbf{P} = \mathbf{P}^\dagger \mathbf{A} \mathbf{P} \mathbf{P}^\dagger \mathbf{A}^\dagger \mathbf{P} = \mathbf{U} \mathbf{U}^\dagger.$$

- if \mathbf{A} is an orthogonally diagonalizable matrix **with real eigenvalues**, it must be Hermitian: if $\mathbf{A} = \mathbf{PDP}^\dagger$, then

$$\mathbf{A}^\dagger = (\mathbf{PDP}^\dagger)^\dagger = \mathbf{PD}^\dagger\mathbf{P}^\dagger = \mathbf{PDP}^\dagger = \mathbf{A}.$$

Moreover, if the entries in \mathbf{A} are also real, then \mathbf{A} must be symmetric.

Q. Are all normal matrices orthogonally diagonalizable?

A. Yes. The following important matrix factorization guarantees this.

- An $n \times n$ matrix \mathbf{A} has a *Schur factorization*

$$\mathbf{A} = \mathbf{PUP}^\dagger,$$

where \mathbf{P} is an orthogonal matrix and \mathbf{U} is an upper triangle matrix. This can be easily seen from the fact that the characteristic polynomial always has at least one complex root, so that \mathbf{A} has at least one eigenvalue λ_1 corresponding to some unit eigenvector \mathbf{x}_1 . Now use the Gram-Schmidt process to construct some orthonormal basis $\{\mathbf{x}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ for \mathbb{R}^n . Note here that the first basis vector is \mathbf{x}_1 itself. In this coordinate system, the eigenvector \mathbf{x}_1 has the components $[1, 0, \dots, 0]^\top$, so that

$$\mathbf{A} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This requires that \mathbf{A} have the form

$$\begin{bmatrix} \lambda_1 & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & * & * & \cdots & * \end{bmatrix}.$$

We can repeat this argument for the $(n-1) \times (n-1)$ submatrix obtained by deleting the first row and column from \mathbf{A} . This submatrix must have an eigenvalue λ_2 corresponding to some eigenvector \mathbf{x}_2 . Now construct an orthogonal basis $\{\mathbf{x}_2, \mathbf{b}_3, \dots, \mathbf{b}_n\}$ for \mathbb{R}^{n-1} . In this new coordinate system $\mathbf{x}_2 = [\lambda_2, 0, \dots, 0]^\top$ and the first column of the submatrix appears as

$$\begin{bmatrix} \lambda_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This means that in the coordinate system $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{b}_3, \dots, \mathbf{b}_n\}$, \mathbf{A} has the form

$$\begin{bmatrix} \lambda_1 & * & * & \cdots & * \\ 0 & \lambda_2 & * & \cdots & * \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & * & \cdots & * \end{bmatrix}.$$

Continuing in this manner we thus construct an orthogonal basis $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n\}$ in which \mathbf{A} appears as an upper triangle matrix \mathbf{U} whose diagonal values are just the n (not necessarily distinct) eigenvalues of \mathbf{A} . Therefore, \mathbf{A} is orthogonally similar to an upper triangle matrix, as claimed.

Remark: Given a normal matrix \mathbf{A} with Schur factorization $\mathbf{A} = \mathbf{P}\mathbf{U}\mathbf{P}^\dagger$, we have seen that \mathbf{U} is also normal.

Problem 14.4: Show that every normal $n \times n$ upper triangular matrix \mathbf{U} is a diagonal matrix. Hint: letting U_{ij} for $i \leq j$ be the nonzero elements of \mathbf{U} , we can write out for each $i = 1, \dots, n$ the diagonal elements $\sum_{k=1}^i U_{ik}^\dagger U_{ki} = \sum_{k=1}^i \bar{U}_{ki} U_{ki}$ of $\mathbf{U}^\dagger \mathbf{U} = \mathbf{U}\mathbf{U}^\dagger$:

$$\sum_{k=1}^i |U_{ki}|^2 = \sum_{k=i}^n |U_{ik}|^2.$$

- Thus \mathbf{A} is orthogonally diagonalizable if and only if it is normal.
- If \mathbf{A} is a Hermitian matrix, it must have real eigenvalues. This follows from the fact that an eigenvalue λ and its eigenvector $\mathbf{x} \neq \mathbf{0}$ must satisfy

$$\begin{aligned} \mathbf{A}\mathbf{x} = \lambda\mathbf{x} &\Rightarrow \mathbf{x}^\dagger \mathbf{A}\mathbf{x} = \mathbf{x}^\dagger \lambda\mathbf{x} = \lambda\mathbf{x} \cdot \mathbf{x} = \lambda|\mathbf{x}|^2 \\ \Rightarrow \lambda &= \frac{\mathbf{x}^\dagger \mathbf{A}\mathbf{x}}{|\mathbf{x}|^2} = \frac{(\mathbf{A}^\dagger \mathbf{x})^\dagger \mathbf{x}}{|\mathbf{x}|^2} = \frac{(\mathbf{A}\mathbf{x})^\dagger \mathbf{x}}{|\mathbf{x}|^2} = \frac{(\lambda\mathbf{x})^\dagger \mathbf{x}}{|\mathbf{x}|^2} = \frac{\bar{\lambda}\mathbf{x} \cdot \mathbf{x}}{|\mathbf{x}|^2} = \bar{\lambda}. \end{aligned}$$

Problem 14.5: Show that the Hermitian matrix

$$\begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$$

is similar to a diagonal matrix with real eigenvalues and find the eigenvalues.

Problem 14.6: Show that the anti-Hermitian matrix

$$\begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

is similar to a diagonal matrix with complex eigenvalues and find the eigenvalues.

- If \mathbf{A} is an $n \times n$ normal matrix, then $|\mathbf{A}\mathbf{x}|^2 = \mathbf{x}^\dagger \mathbf{A}^\dagger \mathbf{A} \mathbf{x} = \mathbf{x}^\dagger \mathbf{A} \mathbf{A}^\dagger \mathbf{x} = |\mathbf{A}^\dagger \mathbf{x}|^2$ for all vectors $\mathbf{x} \in \mathbb{R}^n$.

Problem 14.7: If \mathbf{A} is a normal matrix, show that $\mathbf{A} - \lambda \mathbf{I}$ is also normal.

- If \mathbf{A} is a normal matrix and $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, then $0 = |(\mathbf{A} - \lambda\mathbf{I})\mathbf{x}| = |(\mathbf{A}^\dagger - \bar{\lambda}\mathbf{I})\mathbf{x}|$, so that $\mathbf{A}^\dagger \mathbf{x} = \bar{\lambda}\mathbf{x}$.

- If \mathbf{A} is a normal matrix, the eigenvectors associated with distinct eigenvalues are orthogonal: if $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{A}\mathbf{y} = \mu\mathbf{y}$, then

$$0 = (\mathbf{x}^\dagger \mathbf{A} \mathbf{y})^\top - \mathbf{x}^\dagger \mathbf{A} \mathbf{y} = \mathbf{y}^\top \overline{\mathbf{A}^\dagger \mathbf{x}} - \mathbf{x}^\dagger \mathbf{A} \mathbf{y} = \mathbf{y}^\top \lambda \bar{\mathbf{x}} - \mathbf{x}^\dagger \mu \mathbf{y} = \lambda \mathbf{x}^\dagger \mathbf{y} - \mu \mathbf{x}^\dagger \mathbf{y} = (\lambda - \mu) \mathbf{x} \cdot \mathbf{y},$$

so that $\mathbf{x} \cdot \mathbf{y} = 0$ whenever $\lambda \neq \mu$.

- An $n \times n$ normal matrix \mathbf{A} can therefore be orthogonally diagonalized by applying the Gram-Schmidt process to each of its distinct eigenspaces to obtain a set of n mutually orthogonal eigenvectors that form the columns of an orthonormal matrix \mathbf{P} such that $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$, where the diagonal matrix \mathbf{D} contains the eigenvalues of \mathbf{A} .

Problem 14.8: Find a matrix \mathbf{P} that orthogonally diagonalizes

$$\begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}.$$

Remark: If a normal matrix has eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ corresponding to the orthonormal matrix $[\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]$ then \mathbf{A} has the *spectral decomposition*

$$\begin{aligned} \mathbf{A} &= [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^\dagger \\ \mathbf{x}_2^\dagger \\ \vdots \\ \mathbf{x}_n^\dagger \end{bmatrix} \\ &= [\lambda_1 \mathbf{x}_1 \ \lambda_2 \mathbf{x}_2 \ \dots \ \lambda_n \mathbf{x}_n] \begin{bmatrix} \mathbf{x}_1^\dagger \\ \mathbf{x}_2^\dagger \\ \vdots \\ \mathbf{x}_n^\dagger \end{bmatrix} \\ &= \lambda_1 \mathbf{x}_1 \mathbf{x}_1^\dagger + \lambda_2 \mathbf{x}_2 \mathbf{x}_2^\dagger + \dots + \lambda_n \mathbf{x}_n \mathbf{x}_n^\dagger. \end{aligned}$$

- As we have seen earlier, powers of an orthogonally diagonalizable matrix $\mathbf{A} = \mathbf{PDP}^\dagger$ are easy to compute once \mathbf{P} and \mathbf{U} are known.
- The *Caley-Hamilton Theorem* makes a remarkable connection between the powers of an $n \times n$ matrix \mathbf{A} and its characteristic equation $\lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1\lambda + c_0 = 0$. Specifically, the matrix \mathbf{A} satisfies its characteristic equation in the sense that

$$\mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \dots + c_1\mathbf{A} + c_0\mathbf{I} = \mathbf{0}. \quad (8)$$

Problem 14.9: The characteristic polynomial of

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$$

is $\lambda^2 - 3\lambda - 4 = 0$. Verify that the Caley-Hamilton theorem $\mathbf{A}^2 - 3\mathbf{A} - 4\mathbf{I} = \mathbf{0}$ holds by showing that $\mathbf{A}^2 = 3\mathbf{A} + 4\mathbf{I}$. Then use the Caley-Hamilton theorem to show that

$$\mathbf{A}^3 = \mathbf{A}^2\mathbf{A} = (3\mathbf{A} + 4\mathbf{I})\mathbf{A} = 3\mathbf{A}^2 + 4\mathbf{A} = 3(3\mathbf{A} + 4\mathbf{I}) + 4\mathbf{A} = 13\mathbf{A} + 12\mathbf{I} = \begin{bmatrix} 25 & 26 \\ 39 & 38 \end{bmatrix}.$$

Remark: The Caley-Hamilton Theorem can also be used to compute negative powers (such as the inverse) of a matrix. For example, we can rewrite 8 as

$$\mathbf{A} \left(-\frac{1}{c_0}\mathbf{A}^{n-1} - \frac{c_{n-1}}{c_0}\mathbf{A}^{n-2} - \dots - \frac{c_1}{c_0}\mathbf{A} \right) = \mathbf{I}.$$

- We can also easily compute other functions of diagonalizable matrices. Give an analytic function f with a Taylor series

$$f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + f'''(0)\frac{x^3}{3!} + \dots,$$

and a diagonalizable matrix $\mathbf{A} = \mathbf{PDP}^{-1}$, where \mathbf{D} is diagonal and \mathbf{P} is invertible, one defines

$$f(\mathbf{A}) = f(0)\mathbf{I} + f'(0)\mathbf{A} + f''(0)\frac{\mathbf{A}^2}{2!} + f'''(0)\frac{\mathbf{A}^3}{3!} + \dots \quad (9)$$

On recalling that $\mathbf{A}^n = \mathbf{PD}^n\mathbf{P}^{-1}$ and $\mathbf{A}^0 = \mathbf{I}$, one can rewrite this relation as

$$\begin{aligned} f(\mathbf{A}) &= f(0)\mathbf{PD}^0\mathbf{P}^{-1} + f'(0)\mathbf{PDP}^{-1} + f''(0)\frac{\mathbf{PD}^2\mathbf{P}^{-1}}{2!} + f'''(0)\frac{\mathbf{PD}^3\mathbf{P}^{-1}}{3!} + \dots \\ &= \mathbf{P} \left[f(0)\mathbf{D}^0 + f'(0)\mathbf{D} + f''(0)\frac{\mathbf{D}^2}{2!} + f'''(0)\frac{\mathbf{D}^3}{3!} + \dots \right] \mathbf{P}^{-1} \\ &= \mathbf{P}f(\mathbf{D})\mathbf{P}^{-1}. \end{aligned} \quad (10)$$

Here $f(\mathbf{D})$ is simply the diagonal matrix with elements $f(D_{ii})$.

Problem 14.10: Compute e^A for the diagonalizable matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}.$$

15 Systems of Differential Equations

- Consider the *first-order linear differential equation*

$$y' = ay,$$

where $y = y(t)$, a is a constant, and y' denotes the derivative of y with respect to t . For any constant c the function

$$y = ce^{at}$$

is a solution to the differential equation. If we specify an *initial condition*

$$y(0) = y_0,$$

then the appropriate value of c is y_0 .

- Now consider a system of first-order linear differential equations:

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n, \\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n, \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n, \end{aligned}$$

which in matrix notation appears as

$$\begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

or equivalently as

$$\mathbf{y}' = \mathbf{A}\mathbf{y}.$$

Problem 15.1: Solve the system [Anton & Busby p. 543]:

$$\begin{bmatrix} y_1' \\ y_2' \\ y_3' \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

for the initial conditions $y_1(0) = 1$, $y_2(0) = 4$, $y_3(0) = -2$.

Remark: A linear combination of solutions to an ordinary differential equation is also a solution.

- If \mathbf{A} is an $n \times n$ matrix, then $\mathbf{y}' = \mathbf{A}\mathbf{y}$ has a set of n linearly independent solutions. All other solutions can be expressed as linear combinations of such a *fundamental set of solutions*.
- If \mathbf{x} is an eigenvector of \mathbf{A} corresponding to the eigenvalue λ , then $\mathbf{y} = e^{\lambda t}\mathbf{x}$ is a solution of $\mathbf{y}' = \mathbf{A}\mathbf{y}$:

$$\mathbf{y}' = \frac{d}{dt}e^{\lambda t}\mathbf{x} = e^{\lambda t}\lambda\mathbf{x} = e^{\lambda t}\mathbf{A}\mathbf{x} = \mathbf{A}e^{\lambda t}\mathbf{x} = \mathbf{A}\mathbf{y}.$$

- If $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$, are linearly independent eigenvectors of \mathbf{A} corresponding to the (not necessarily distinct) eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, then

$$\mathbf{y} = e^{\lambda_1 t}\mathbf{x}_1, \quad \mathbf{y} = e^{\lambda_2 t}\mathbf{x}_2, \quad \dots, \quad \mathbf{y} = e^{\lambda_k t}\mathbf{x}_k$$

are linearly independent solutions of $\mathbf{y}' = \mathbf{A}\mathbf{y}$: if

$$0 = c_1e^{\lambda_1 t}\mathbf{x}_1 + c_2e^{\lambda_2 t}\mathbf{x}_2 + \dots + c_ke^{\lambda_k t}\mathbf{x}_k$$

then at $t = 0$ we find

$$0 = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \dots + c_k\mathbf{x}_k;$$

the linear independence of the k eigenvectors then implies that

$$c_1 = c_2 = \dots = c_k = 0.$$

- If $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, are linearly independent eigenvectors of an $n \times n$ matrix \mathbf{A} corresponding to the (not necessarily distinct) eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, then every solution to $\mathbf{y}' = \mathbf{A}\mathbf{y}$ can be expressed in the form

$$\mathbf{y} = c_1e^{\lambda_1 t}\mathbf{x}_1 + c_2e^{\lambda_2 t}\mathbf{x}_2 + \dots + c_ne^{\lambda_n t}\mathbf{x}_n,$$

which is known as the *general solution* to $\mathbf{y}' = \mathbf{A}\mathbf{y}$.

- On introducing the matrix of eigenvectors $\mathbf{P} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_n]$, we may write the general solution as

$$\mathbf{y} = c_1e^{\lambda_1 t}\mathbf{x}_1 + c_2e^{\lambda_2 t}\mathbf{x}_2 + \dots + c_ne^{\lambda_n t}\mathbf{x}_n = \mathbf{P} \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

- Moreover, if $\mathbf{y} = \mathbf{y}_0$ at $t = 0$ we find that

$$\mathbf{y}_0 = \mathbf{P} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

If the n eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are linearly independent, then

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{P}^{-1} \mathbf{y}_0.$$

The solution to $\mathbf{y}' = \mathbf{A}\mathbf{y}$ for the initial condition \mathbf{y}_0 may thus be expressed as

$$\mathbf{y} = \mathbf{P} \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n t} \end{bmatrix} \mathbf{P}^{-1} \mathbf{y}_0 = e^{\mathbf{A}t} \mathbf{y}_0,$$

on making use of Eq. (10) with $f(x) = e^x$ and \mathbf{A} replaced by $\mathbf{A}t$.

Problem 15.2: Find the solution to $\mathbf{y}' = \mathbf{A}\mathbf{y}$ corresponding to the initial condition $\mathbf{y}_0 = [1, -1]$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}.$$

Remark: The solution $\mathbf{y} = e^{\mathbf{A}t} \mathbf{y}_0$ to the system of ordinary differential equations

$$\mathbf{y}' = \mathbf{A}\mathbf{y} \quad \mathbf{y}(0) = \mathbf{y}_0$$

actually holds even when the matrix \mathbf{A} isn't diagonalizable (that is, when \mathbf{A} doesn't have a set of n linearly independent eigenvectors). In this case we cannot use Eq. (10) to find $e^{\mathbf{A}t}$. Instead we must find $e^{\mathbf{A}t}$ from the infinite series in Eq. (9):

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} + \dots$$

Problem 15.3: Show for the matrix

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

that the series for $e^{\mathbf{A}t}$ simplifies to $\mathbf{I} + \mathbf{A}t$.

16 Quadratic Forms

- Linear combinations of x_1, x_2, \dots, x_n such as

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = \sum_{i=1}^n a_ix_i,$$

where a_1, a_2, \dots, a_n are fixed but arbitrary constants, are called *linear forms* on \mathbb{R}^n .

- Expressions of the form

$$\sum_{i,j=1}^n a_{ij}x_ix_j,$$

where the a_{ij} are fixed but arbitrary constants, are called *quadratic forms* on \mathbb{R}^n .

- Without loss of generality, we restrict the constants a_{ij} to be the components of a symmetric matrix \mathbf{A} and define $\mathbf{x} = [x_1, x_2, \dots, x_n]$, so that

$$\begin{aligned} \mathbf{x}^\top \mathbf{A} \mathbf{x} &= \sum_{i,j} x_i a_{ij} x_j \\ &= \sum_{i < j} a_{ij} x_i x_j + \sum_{i=j} a_{ij} x_i x_j + \sum_{i > j} a_{ij} x_i x_j \\ &= \sum_{i=j} a_{ij} x_i x_j + \sum_{i < j} a_{ij} x_i x_j + \sum_{i < j} a_{ji} x_j x_i \\ &= \sum_{i=j} a_{ii} x_i^2 + 2 \sum_{i < j} a_{ij} x_i x_j. \end{aligned}$$

Remark: Note that $\mathbf{x}^\top \mathbf{A} \mathbf{x} = \mathbf{x} \cdot \mathbf{A} \mathbf{x} = \mathbf{A} \mathbf{x} \cdot \mathbf{x}$.

- (Principle Axes Theorem) Consider the change of variable $\mathbf{x} = \mathbf{P} \mathbf{y}$, where \mathbf{P} orthogonally diagonalizes \mathbf{A} . Then in terms of the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of \mathbf{A} :

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = (\mathbf{P} \mathbf{y})^\top \mathbf{A} \mathbf{P} \mathbf{y} = \mathbf{y}^\top (\mathbf{P}^\top \mathbf{A} \mathbf{P}) \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2.$$

- (Constrained Extremum Theorem) For a symmetric $n \times n$ matrix \mathbf{A} , the minimum (maximum) value of

$$\{\mathbf{x}^\top \mathbf{A} \mathbf{x} : |\mathbf{x}|^2 = 1\}$$

is given by the smallest (largest) eigenvalue of \mathbf{A} and occurs at the corresponding unit eigenvector. To see this, let $\mathbf{x} = \mathbf{P} \mathbf{y}$, where \mathbf{P} orthogonally diagonalizes \mathbf{A} , and list the eigenvalues in ascending order: $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Then

$$\lambda_1 = \lambda_1(y_1^2 + y_2^2 + \dots + y_n^2) \leq \underbrace{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2}_{\mathbf{x}^\top \mathbf{A} \mathbf{x}} \leq \lambda_n(y_1^2 + y_2^2 + \dots + y_n^2) = \lambda_n.$$

Problem 16.1: Find the minimum and maximum values of the quadratic form

$$z = 5x^2 + 4xy + 5y^2$$

subject to the constraint $x^2 + y^2 = 1$ [Anton & Busby p. 498].

Definition: A symmetric matrix \mathbf{A} is said to be

- *positive definite* if $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$ for $\mathbf{x} \neq 0$;
- *negative definite* if $\mathbf{x}^\top \mathbf{A} \mathbf{x} < 0$ for $\mathbf{x} \neq 0$;
- *indefinite* if $\mathbf{x}^\top \mathbf{A} \mathbf{x}$ exhibits positive **and** negative values for various $\mathbf{x} \neq 0$.

Remark: If \mathbf{A} is a symmetric matrix then $\mathbf{x}^\top \mathbf{A} \mathbf{x}$

- is *positive definite* if and only if all eigenvalues of \mathbf{A} are positive;
 - is *negative definite* if and only if all eigenvalues of \mathbf{A} are negative;
 - is *indefinite* if and only if \mathbf{A} has at least one positive eigenvalue and at least one negative eigenvalue.
- The k th principal submatrix of an $n \times n$ matrix \mathbf{A} is the submatrix composed of the first k rows and columns of \mathbf{A} .
 - (Sylvester's Criterion) If \mathbf{A} is a symmetric matrix then \mathbf{A} is *positive definite* if and only if the determinant of every principal submatrix of \mathbf{A} is positive.

Remark: Sylvester's criterion is easily verified for a 2×2 symmetric matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & d \end{bmatrix};$$

if $a > 0$ and $ad > b^2$ then $d > 0$, so both the trace $a + d$ and determinant are positive, ensuring in turn that both eigenvalues of \mathbf{A} are positive.

Remark: If \mathbf{A} is a symmetric matrix then the following are equivalent:

- (a) \mathbf{A} is positive definite;
- (b) there is a symmetric positive definite matrix \mathbf{B} such that $\mathbf{A} = \mathbf{B}^2$;
- (c) there is an invertible matrix \mathbf{C} such that $\mathbf{A} = \mathbf{C}^\top \mathbf{C}$.

Remark: If \mathbf{A} is a symmetric 2×2 matrix then the equation $\mathbf{x}^\top \mathbf{A} \mathbf{x} = 1$ can be expressed in the orthogonal coordinate system $\mathbf{y} = \mathbf{P}^\top \mathbf{x}$ as $\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1$. Thus $\mathbf{x}^\top \mathbf{A} \mathbf{x}$ represents

- an ellipse if \mathbf{A} is positive definite;
 - no graph if \mathbf{A} is negative definite;
 - a hyperbola if \mathbf{A} is indefinite.
- A *critical point* of a function f is a point in the domain of f where either f is not differentiable or its derivative is 0.
 - The *Hessian* $H(x_0, y_0)$ of a twice-differentiable function f is the matrix of second partial derivatives

$$\begin{bmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{bmatrix}.$$

Remark: If f has continuous second-order partial derivatives at a point (x_0, y_0) then $f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$, so that the Hessian is a symmetric matrix.

Remark (Second Derivative Test): If f has continuous second-order partial derivatives near a critical point (x_0, y_0) of f then

- f has a *relative minimum* at (x_0, y_0) if H is positive definite.
- f has a *relative maximum* at (x_0, y_0) if H is negative definite.
- f has a *saddle point* at (x_0, y_0) if H is indefinite.
- anything can happen at (x_0, y_0) if $\det H(x_0, y_0) = 0$.

Problem 16.2: Determine the relative minima, maxima, and saddle points of the function [Anton & Busby p. 498]

$$f(x, y) = \frac{1}{3}x^3 + xy^2 - 8xy + 3.$$

17 Vector Spaces:

- A *vector space* V over R (\mathbb{C}) is a set containing an element $\mathbf{0}$ that is closed under a vector addition and scalar multiplication operation such that for all vectors \mathbf{u} , \mathbf{v} , \mathbf{w} in V and scalars c , d in \mathbb{R} (\mathbb{C}) the following *axioms* hold:

$$(A1) \quad (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) \quad (\text{associative});$$

$$(A2) \quad \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \quad (\text{commutative});$$

$$(A3) \quad \mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u} \quad (\text{additive identity});$$

$$(A4) \quad \text{there exists an element } (-\mathbf{u}) \in V \text{ such that } \mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = \mathbf{0} \\ (\text{additive inverse});$$

$$(A5) \quad 1\mathbf{u} = \mathbf{u} \quad (\text{multiplicative identity});$$

$$(A6) \quad c(d\mathbf{u}) = (cd)\mathbf{u} \quad (\text{scalar multiplication});$$

$$(A7) \quad (c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u} \quad (\text{distributive});$$

$$(A8) \quad c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v} \quad (\text{distributive}).$$

Problem 17.1: Show that axiom (A2) in fact follows from the other seven axioms.

Remark: In addition to \mathbb{R}^n , other vector spaces can also be constructed:

- Since it satisfies axioms A1–A8, the set of $m \times n$ matrices is in fact a vector space.
- The set of $n \times n$ symmetric matrices is a vector space.
- The set P_n of polynomials $a_0 + a_1x + \dots + a_nx^n$ of degree less than or equal to n (where a_0, a_1, \dots, a_n are real numbers) is a vector space.
- The set P_∞ of all polynomials is an infinite-dimensional vector space.
- The set $F(\mathbb{R})$ of functions defined on \mathbb{R} form an infinite-dimensional vector space if we define vector addition by

$$(f + g)(x) = f(x) + g(x) \text{ for all } x \in \mathbb{R}.$$

and scalar multiplication by

$$(cf)(x) = cf(x) \text{ for all } x \in \mathbb{R}.$$

- The set $C(\mathbb{R})$ of continuous functions on \mathbb{R} is a vector space.
- The set of differentiable functions on \mathbb{R} is a vector space.
- The set $C^1(\mathbb{R})$ of functions with continuous first derivatives on \mathbb{R} is a vector space.
- The set $C^m(\mathbb{R})$ of functions with continuous m th derivatives on \mathbb{R} is a vector space.
- The set $C^\infty(\mathbb{R})$ of functions with continuous derivatives of all orders on \mathbb{R} is a vector space.

Definition: A nonempty subset of a vector space V that is itself a vector space under the same vector addition and scalar multiplication operations is called a *subspace* of V .

- A nonempty subset of a vector space V is a subspace of V iff it is closed under vector addition and scalar multiplication.

Remark: The concept of *linear independence* can be carried over to infinite-dimensional vector spaces.

- If the *Wronskian*

$$W(x) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}$$

of functions $f_1(x), f_2(x), \dots, f_n(x)$ in $C^{n-1}(x)$ is nonzero for some $x \in \mathbb{R}$, then these n functions are linearly independent on \mathbb{R} . This follows on observing that if there were constants c_1, c_2, \dots, c_n such that the function

$$g(x) = c_1 f_1(x) + c_2 f_2(x) + \dots + c_n f_n(x)$$

is zero for all $x \in \mathbb{R}$, so that $g^{(k)}(x) = 0$ for $k = 0, 1, 2, \dots, n-1$ for all $x \in \mathbb{R}$, then the linear system

$$\begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

would have a nontrivial solution for all $x \in \mathbb{R}$.

Problem 17.2: Show that the functions $f_1(x) = 1$, $f_2(x) = e^x$, and $f_3(x) = e^{2x}$ are linearly independent on \mathbb{R} .

18 Inner Product Spaces

- An *inner product* on a vector space V over \mathbb{R} (\mathbb{C}) is a function that maps each pair of vectors \mathbf{u} and \mathbf{v} in V to a unique real (complex) number $\langle \mathbf{u}, \mathbf{v} \rangle$ such that, for all vectors \mathbf{u} , \mathbf{v} , and \mathbf{w} in V and scalars c in \mathbb{R} (\mathbb{C}):

$$(I1) \quad \langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle} \quad (\text{conjugate symmetry});$$

$$(I2) \quad \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \quad (\text{additivity});$$

$$(I3) \quad \langle c\mathbf{u}, \mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle \quad (\text{homogeneity});$$

$$(I4) \quad \langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \text{ and } \langle \mathbf{v}, \mathbf{v} \rangle = 0 \text{ iff } \mathbf{v} = \mathbf{0} \quad (\text{positivity}).$$

- The usual Euclidean dot product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \bar{\mathbf{v}}$ provides an inner product on \mathbb{C}^n .

Problem 18.1: If w_1, w_2, \dots, w_n are positive numbers, show that

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + \dots + w_n u_n v_n$$

provides an inner product on the vector space \mathbb{R}^n .

- Every inner product on \mathbb{R}^n can be uniquely expressed as $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\top \mathbf{A} \mathbf{v}$ for some positive definite symmetric matrix \mathbf{A} , so that $\langle \mathbf{v}, \mathbf{v} \rangle = \mathbf{v}^\top \mathbf{A} \mathbf{v} > 0$ for $\mathbf{v} \neq \mathbf{0}$.
- In a vector space V with inner product $\langle \mathbf{u}, \mathbf{v} \rangle$, the *norm* $|\mathbf{v}|$ of a vector \mathbf{v} is given by $\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, the *distance* $d(\mathbf{u}, \mathbf{v})$ is given by $|\mathbf{u} - \mathbf{v}| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle}$, and the *angle* θ between two vectors \mathbf{u} and \mathbf{v} satisfies $\cos \theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{|\mathbf{u}| |\mathbf{v}|}$. Two vectors \mathbf{u} and \mathbf{v} are *orthogonal* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Analogues of the **Pythagoras theorem**, **Cauchy-Schwarz inequality**, and **triangle inequality** follow directly from (I1)–(I4).

Problem 18.2: Show that

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

defines an inner product on the vector space $C[a, b]$ of continuous functions on $[a, b]$, with norm

$$|f| = \sqrt{\int_a^b f^2(x) dx}.$$

- The *Fourier theorem* states that an orthonormal basis for the infinite-dimensional vector space of differentiable periodic functions on $[-\pi, \pi]$ with inner product $\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x) dx$ is given by $\{u_n\}_{n=0}^{\infty} = \{\frac{c_0}{\sqrt{2}}, c_1, s_1, c_2, s_2, \dots\}$, where

$$c_n(x) = \frac{1}{\sqrt{\pi}} \cos nx$$

and

$$s_n(x) = \frac{1}{\sqrt{\pi}} \sin nx.$$

The orthonormality of this basis follows from the trigonometric addition formulae

$$\sin(nx \pm mx) = \sin nx \cos mx \pm \cos nx \sin mx$$

and

$$\cos(nx \pm mx) = \cos nx \cos mx \mp \sin nx \sin mx,$$

from which we see that

$$\sin(nx + mx) - \sin(nx - mx) = 2 \cos nx \sin mx, \quad (11)$$

$$\cos(nx + mx) + \cos(nx - mx) = 2 \cos nx \cos mx \quad (12)$$

$$\cos(nx - mx) - \cos(nx + mx) = 2 \sin nx \sin mx \quad (13)$$

Integration of Eq. (11) from $-\pi$ to π for distinct non-negative integers n and m yields

$$2 \int_{-\pi}^{\pi} \cos nx \sin mx dx = \left[-\frac{\cos(nx + mx)}{n + m} + \frac{\cos(nx - mx)}{n - m} \right]_{-\pi}^{\pi} = 0,$$

since the cosine function is periodic with period 2π . When $n = m > 0$ we find

$$2 \int_{-\pi}^{\pi} \cos nx \sin nx dx = \left[-\frac{\cos(2nx)}{2n} \right]_{-\pi}^{\pi} = 0,$$

Likewise, for distinct non-negative integers n and m , Eq. (12) leads to

$$2 \int_{-\pi}^{\pi} \cos nx \cos mx dx = \left[\frac{\sin(nx + mx)}{n + m} + \frac{\sin(nx - mx)}{n - m} \right]_{-\pi}^{\pi} = 0,$$

but when $n = m > 0$ we find, since $\cos(nx - nx) = 1$,

$$2 \int_{-\pi}^{\pi} \cos^2 nx dx = \left[\frac{\sin(2nx)}{2n} + x \right]_{-\pi}^{\pi} = 2\pi.$$

Note when $n = m = 0$ that $\int_{-\pi}^{\pi} \cos^2 nx dx = 2\pi$.

Similarly, Eq. (13) yields for distinct non-negative integers n and m ,

$$2 \int_{-\pi}^{\pi} \sin nx \sin mx \, dx = 0,$$

but

$$2 \int_{-\pi}^{\pi} \sin^2 nx \, dx = \left[x - \frac{\sin(2nx)}{2n} \right]_{-\pi}^{\pi} = 2\pi.$$

- A differentiable periodic function f on $[-\pi, \pi]$ may thus be represented exactly by its infinite *Fourier series*:

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \langle f, u_n \rangle u_n \\ &= \left\langle f, \frac{c_0}{\sqrt{2}} \right\rangle \frac{c_0}{\sqrt{2}} + \sum_{n=1}^{\infty} [\langle f, c_n \rangle c_n + \langle f, s_n \rangle s_n] \\ &= \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos nx + b_n \sin nx], \end{aligned}$$

in terms of the *Fourier coefficients*

$$a_n = \frac{1}{\sqrt{\pi}} \langle f, c_n \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx \quad (n = 0, 1, 2, \dots),$$

and

$$b_n = \frac{1}{\sqrt{\pi}} \langle f, s_n \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx \quad (n = 1, 2, \dots).$$

Problem 18.3: By using integration by parts, show that the Fourier series for $f(x) = x$ on $[-\pi, \pi]$ is

$$2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \sin nx}{n}.$$

For $x \in (-\pi, \pi)$ this series is guaranteed to converge to x . For example, at $x = \pi/2$, we find

$$4 \sum_{m=0}^{\infty} \frac{(-1)^{2m+2} \sin \left((2m+1) \frac{\pi}{2} \right)}{2m+1} = 4 \sum_{m=0}^{\infty} \frac{(-1)^m}{2m+1} = \pi.$$

Problem 18.4: By using integration by parts, show that the Fourier series for $f(x) = |x|$ on $[-\pi, \pi]$ is

$$\frac{\pi}{2} - \frac{4}{\pi} \sum_{m=1}^{\infty} \frac{\cos(2m-1)x}{(2m-1)^2}.$$

This series can be shown to converge to $|x|$ for all $x \in [-\pi, \pi]$. For example, at $x = 0$, we find

$$\sum_{m=1}^{\infty} \frac{1}{(2m-1)^2} = \frac{\pi^2}{8}.$$

Remark: Many other concepts for Euclidean vector spaces can be generalized to function spaces.

- The functions $\cos nx$ and $\sin nx$ in a Fourier series can be thought of as the *eigenfunctions* of the differential operator d^2/dx^2 :

$$\frac{d^2}{dx^2}y = -n^2y.$$

19 General linear transformations

- A *linear transformation* $T : V \rightarrow W$ from one vector space to another is a mapping that for all vectors \mathbf{u}, \mathbf{v} in V and all scalars satisfies c

(a) $T(c\mathbf{u}) = cT(\mathbf{u})$;

(b) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$;

Problem 19.1: Show that every linear transformation T satisfies

(a) $T(\mathbf{0}) = \mathbf{0}$;

(b) $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$;

Problem 19.2: Show that the transformation that maps each $n \times n$ matrix to its trace is a linear transformation.

Problem 19.3: Show that the transformation that maps each $n \times n$ matrix to its determinant is **not** a linear transformation.

- The *kernel* of a linear transformation $T : V \rightarrow W$ is the set of all vectors in V that are mapped by T to the zero vector.
- The *range* of a linear transformation T is the set of all vectors in W that are the image under T of at least one vector in V .

Problem 19.4: Given an inner product space V containing a fixed nonzero vector \mathbf{u} , let $T(\mathbf{x}) = \langle \mathbf{x}, \mathbf{u} \rangle$ for $\mathbf{x} \in V$. Show that the kernel of T is the set of vectors that are orthogonal to \mathbf{u} and that the range of T is \mathbb{R} .

Problem 19.5: Show that the derivative operator, which maps continuously differentiable functions $f(x)$ to continuous functions $f'(x)$, is a linear transformation from $C^1(\mathbb{R})$ to $C(\mathbb{R})$.

Problem 19.6: Show that the antiderivative operator, which maps continuous functions $f(x)$ to continuously differentiable functions $\int_0^x f(t) dt$, is a linear transformation from $C(\mathbb{R})$ to $C^1(\mathbb{R})$.

Problem 19.7: Show that the kernel of the derivative operator on $C^1(\mathbb{R})$ is the set of constant functions on \mathbb{R} and that its range is $C(\mathbb{R})$.

Problem 19.8: Let T be a linear transformation T . Show that T is one-to-one if and only if $\ker T = \{\mathbf{0}\}$.

Definition: A linear transformation $T : V \rightarrow W$ is called an *isomorphism* if it is one-to-one and onto. If such a transformation exists from V to W we say that V and W are *isomorphic*.

- Every real n -dimensional vector space V is *isomorphic* to \mathbb{R}^n : given a basis $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$, we may uniquely express $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \dots + c_n\mathbf{u}_n$; the linear mapping $T(\mathbf{v}) = (c_1, c_2, \dots, c_n)$ from V to \mathbb{R}^n is then easily shown to be one-to-one and onto. This means V differs from \mathbb{R}^n only in the notation used to represent vectors.
- For example, the set P_n of polynomials of degree n is isomorphic to \mathbb{R}^{n+1} .

- Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear operator and $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a basis for \mathbb{R}^n . The matrix

$$\mathbf{A} = [[T(\mathbf{v}_1)]_B \ [T(\mathbf{v}_2)]_B \ \cdots \ [T(\mathbf{v}_n)]_B]$$

is called the *matrix for T with respect to B* , with

$$[T(\mathbf{x})]_B = \mathbf{A}[\mathbf{x}]_B$$

for all $\mathbf{x} \in \mathbb{R}^n$. In the case where B is the standard Cartesian basis for \mathbb{R}^n , the matrix \mathbf{A} is called the *standard matrix* for the linear transformation T . Furthermore, if $B' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n\}$ is any basis for \mathbb{R}^n , then

$$[T(\mathbf{x})]_{B'} = \mathbf{P}_{B' \leftarrow B} [T(\mathbf{x})]_B \mathbf{P}_{B \leftarrow B'}.$$

20 Singular Value Decomposition

- Given an $m \times n$ matrix \mathbf{A} , consider the square Hermitian $n \times n$ matrix $\mathbf{A}^\dagger \mathbf{A}$. Each eigenvalue λ associated with an eigenvector \mathbf{v} of $\mathbf{A}^\dagger \mathbf{A}$ must be real and non-negative since

$$0 \leq |\mathbf{A}\mathbf{v}|^2 = \mathbf{A}\mathbf{v} \cdot \mathbf{A}\mathbf{v} = (\mathbf{A}\mathbf{v})^\dagger \mathbf{A}\mathbf{v} = \mathbf{v}^\dagger \mathbf{A}^\dagger \mathbf{A}\mathbf{v} = \mathbf{v}^\dagger \lambda \mathbf{v} = \lambda \mathbf{v} \cdot \mathbf{v} = \lambda |\mathbf{v}|^2.$$

Problem 20.1: Show that $\text{null}(\mathbf{A}^\dagger \mathbf{A}) = \text{null}(\mathbf{A})$ and conclude from the **dimension theorem** that $\text{rank}(\mathbf{A}^\dagger \mathbf{A}) = \text{rank}(\mathbf{A})$.

- Since $\mathbf{A}^\dagger \mathbf{A}$ is Hermitian, it can be orthogonally diagonalized as $\mathbf{A}^\dagger \mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^\dagger$, where the diagonal matrix \mathbf{D} contains the n (non-negative) eigenvalues σ_j^2 of $\mathbf{A}^\dagger \mathbf{A}$ listed in decreasing order (taking $\sigma_j \geq 0$) and the column vectors of the matrix $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n]$ are a set of orthonormal eigenvectors for $\mathbf{A}^\dagger \mathbf{A}$. Now if \mathbf{A} has rank k , so does \mathbf{D} since it is similar to $\mathbf{A}^\dagger \mathbf{A}$, which we have just seen has the same rank as \mathbf{A} . In other words $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > 0$ but $\sigma_{k+1} = \sigma_{k+2} = \dots = \sigma_n = 0$. Since

$$\overline{\mathbf{A}\mathbf{v}_i \cdot \mathbf{A}\mathbf{v}_j} = (\mathbf{A}\mathbf{v}_i)^\dagger \mathbf{A}\mathbf{v}_j = \mathbf{v}_i^\dagger \mathbf{A}^\dagger \mathbf{A}\mathbf{v}_j = \mathbf{v}_i^\dagger \sigma_j^2 \mathbf{v}_j = \sigma_j^2 \overline{\mathbf{v}_i \cdot \mathbf{v}_j},$$

we see that the vectors

$$\mathbf{u}_j = \frac{\mathbf{A}\mathbf{v}_j}{|\mathbf{A}\mathbf{v}_j|} = \frac{\mathbf{A}\mathbf{v}_j}{\sigma_j}, \quad j = 1, 2, \dots, k$$

are also orthonormal, so that the matrix

$$[\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_k]$$

is orthogonal. We also see that $\mathbf{A}\mathbf{v}_j = \mathbf{0}$ for $k < j \leq n$. We can then extend this set of vectors to an orthonormal basis for \mathbb{R}^m , which we write as the column vectors of an $m \times m$ matrix

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_k \quad \dots \quad \mathbf{u}_m].$$

The product \mathbf{U} and the $m \times n$ matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_k & 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & \dots & 0 \end{bmatrix},$$

is the $m \times n$ matrix

$$\begin{aligned} \mathbf{U}\mathbf{\Sigma} &= [\sigma_1\mathbf{u}_1 \quad \sigma_2\mathbf{u}_2 \quad \dots \quad \sigma_k\mathbf{u}_k \quad \mathbf{0} \quad \dots \quad \mathbf{0}] \\ &= [\mathbf{A}\mathbf{v}_1 \quad \mathbf{A}\mathbf{v}_2 \quad \dots \quad \mathbf{A}\mathbf{v}_k \quad \mathbf{A}\mathbf{v}_{k+1} \quad \dots \quad \mathbf{A}\mathbf{v}_n] \\ &= \mathbf{A}\mathbf{V}. \end{aligned}$$

Since the $n \times n$ matrix \mathbf{V} is an orthogonal matrix, we then obtain a *singular value decomposition* of the $m \times n$ matrix \mathbf{A} :

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger.$$

The positive values σ_j for $j = 1, 2, \dots, k$ are known as the *singular values* of \mathbf{A} .

Remark: By allowing the orthogonal matrices \mathbf{U} and \mathbf{V} to be different, the concept of orthogonal diagonalization can thus be replaced by a more general concept that applies even to nonnormal matrices.

Remark: The singular value decomposition plays an important role in numerical linear algebra, image compression, and statistics.

Problem 20.2: Show that the singular values of a positive-definite Hermitian matrix are the same as its eigenvalues.

Problem 20.3: Find a singular value decomposition of [Anton & Busby p. 505]

$$\mathbf{A} = \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix}.$$

Problem 20.4: Find a singular value decomposition of [Anton & Busby p. 509]

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

We find

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

which has characteristic polynomial $(2 - \lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$. The matrix $\mathbf{A}^T \mathbf{A}$ has eigenvalues 3 and 1; their respective eigenvectors

$$\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

form the columns of \mathbf{V} .

The singular values of \mathbf{A} are $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$ and

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix},$$

$$\mathbf{u}_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

We now look for a third vector which is perpendicular to scaled versions of \mathbf{u}_1 and \mathbf{u}_2 , namely $\sqrt{6}\mathbf{u}_1 = [2, 1, 1]$ and $\sqrt{2}\mathbf{u}_2 = [0, -1, 1]$. That is, we want to find a vector in the null space of the matrix

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix},$$

or equivalently, after row reduction,

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

We then see that $[-1, 1, 1]$ lies in the null space of this matrix. On normalizing this vector, we find

$$\mathbf{u}_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}.$$

A singular value decomposition of \mathbf{A} is thus given by

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Remark: An alternative way of computing a singular value decomposition of the nonsquare matrix in Problem 20.4 is to take the transpose of a singular value decomposition for

$$\mathbf{A}^\top = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

- The first k columns of the matrix \mathbf{U} form an orthonormal basis for $\text{col}(\mathbf{A})$, whereas the remaining $m - k$ columns form an orthonormal basis for $\text{col}(\mathbf{A})^\perp = \text{null}(\mathbf{A}^\top)$.
- The first k columns of the matrix \mathbf{V} form an orthonormal basis for $\text{row}(\mathbf{A})$, whereas the remaining $n - k$ columns form an orthonormal basis for $\text{row}(\mathbf{A})^\perp = \text{null}(\mathbf{A})$.
- If \mathbf{A} is a nonzero matrix, the singular value decomposition may be expressed more efficiently as $\mathbf{A} = \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^\dagger$, in terms of the $m \times k$ matrix

$$\mathbf{U}_1 = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_k],$$

the diagonal $k \times k$ matrix

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k \end{bmatrix},$$

and the $n \times k$ matrix

$$\mathbf{V}_1 = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_k].$$

This *reduced singular value decomposition* avoids the superfluous multiplications by zero in the product $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger$.

Problem 20.5: Show that the matrix $\boldsymbol{\Sigma}_1$ is always invertible.

Problem 20.6: Find a reduced singular value decomposition for the matrix

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

considered previously in Problem 20.4.

- The singular value decomposition can equivalently be expressed as the *singular value expansion*

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\dagger + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^\dagger + \dots + \sigma_k \mathbf{u}_k \mathbf{v}_k^\dagger.$$

In contrast to the *spectral decomposition*, which applies only to Hermitian matrices, the singular value expansion applies to all matrices.

- If \mathbf{A} is an $n \times n$ matrix of rank k , then \mathbf{A} has the *polar decomposition*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\dagger\mathbf{U}\mathbf{V}^\dagger = \mathbf{P}\mathbf{Q},$$

where $\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\dagger$ is a positive semidefinite matrix of rank k and $\mathbf{Q} = \mathbf{U}\mathbf{V}^\dagger$ is an $n \times n$ orthogonal matrix.

21 The Pseudoinverse

- If \mathbf{A} is an $n \times n$ matrix of rank n , then we know it is invertible. In this case, the singular value decomposition and the reduced singular value decomposition coincide:

$$\mathbf{A} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^\dagger,$$

where the diagonal matrix $\mathbf{\Sigma}_1$ contains the n positive singular values of \mathbf{A} . Moreover, the orthogonal matrices \mathbf{U}_1 and \mathbf{V}_1 are square, so they are invertible. This yields an interesting expression for the inverse of \mathbf{A} :

$$\mathbf{A}^{-1} = \mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^\dagger,$$

where $\mathbf{\Sigma}_1^{-1}$ contains the reciprocals of the n positive singular values of \mathbf{A} along its diagonal.

Definition: For a general nonzero $m \times n$ matrix \mathbf{A} of rank k , the $n \times m$ matrix

$$\mathbf{A}^+ = \mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^\dagger$$

is known as the *pseudoinverse* of \mathbf{A} . It is also convenient to define $\mathbf{0}^+ = \mathbf{0}$, so that the pseudoinverse exists for all matrices. In the case where \mathbf{A} is invertible, then $\mathbf{A}^{-1} = \mathbf{A}^+$.

Remark: Equivalently, if we define $\mathbf{\Sigma}^+$ to be the $m \times n$ matrix obtained by replacing all nonzero entries of $\mathbf{\Sigma}$ with their reciprocals, we can define the pseudoinverse directly in terms of the unreduced singular value decomposition:

$$\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^\dagger.$$

Problem 21.1: Show that the pseudoinverse of the matrix in Problem 20.4 is [cf. Anton & Busby p. 520]:

$$\begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}.$$

Problem 21.2: Prove that the pseudoinverse \mathbf{A}^+ of an $m \times n$ matrix \mathbf{A} satisfies the following properties:

- (a) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$;
- (b) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$;
- (c) $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^+ = \mathbf{A}^\dagger$;
- (d) $\mathbf{A}^+\mathbf{A}$ and $\mathbf{A}\mathbf{A}^+$ are Hermitian;
- (e) $(\mathbf{A}^+)^\dagger = (\mathbf{A}^\dagger)^+$.

Remark: If \mathbf{A} has full column rank, then $\mathbf{A}^\dagger\mathbf{A}$ is invertible and property (c) in Problem 21.2 provides an alternative way of computing the pseudoinverse:

$$\mathbf{A}^+ = (\mathbf{A}^\dagger\mathbf{A})^{-1}\mathbf{A}^\dagger.$$

- An important application of the pseudoinverse of a matrix \mathbf{A} is in solving the least-squares problem

$$\mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}$$

since the solution \mathbf{x} of minimum norm can be conveniently expressed in terms of the pseudoinverse:

$$\mathbf{x} = \mathbf{A}^+\mathbf{b}.$$

Index

- \mathbb{C} , 23
- span, 10
- QR** factorization, 31

- Addition of matrices, 8
- additive identity, 46
- additive inverse, 46
- additivity, 48
- adjoint, 13
- adjugate, 13
- affine, 32
- Angle, 4
- angle, 48
- associative, 46
- Augmented matrix, 6
- axioms, 46

- basis, 17

- Caley-Hamilton Theorem, 39
- Cauchy-Schwarz inequality, 5
- change of bases, 19
- characteristic equation, 14
- characteristic polynomial, 14
- cofactor, 12
- commutative, 46
- commute, 8
- Complex Conjugate Roots, 26
- complex modulus, 25
- complex numbers, 22, 23
- Component, 5
- component, 28
- conjugate pairs, 26
- conjugate symmetry, 48
- consistent, 32
- coordinates, 17
- Cramer's rule, 12
- critical point, 45
- cross product, 13

- deMoivre's Theorem, 27

- design matrix, 32
- Determinant, 11
- diagonal matrix, 7
- diagonalizable, 19
- dimension theorem, 18
- Distance, 4
- distance, 48
- distributive, 46
- domain, 16
- Dot, 4
- dot product, 25

- eigenfunctions, 51
- eigenspace, 15
- eigenvalue, 13
- eigenvector, 13
- eigenvector matrix, 20
- elementary matrix, 10
- Elementary row operations, 7
- Equation of line, 5
- Equation of plane, 5

- first-order linear differential equation, 40
- fixed point, 13
- Fourier coefficients, 50
- Fourier series, 50
- Fourier theorem, 49
- full column rank, 28
- fundamental set of solutions, 41
- Fundamental Theorem of Algebra, 27

- Gauss-Jordan elimination, 7
- Gaussian elimination, 7
- general solution, 41
- Gram-Schmidt process, 29

- Hermitian, 34
- Hermitian transpose, 34
- Hessian, 45
- homogeneity, 48
- Homogeneous linear equation, 6

Identity, 9
 imaginary number, 22
 indefinite, 44
 initial condition, 40
 inner, 4
 inner product, 48
 Inverse, 9
 invertible, 9
 isomorphic, 52
 isomorphism, 52

 kernel, 16, 52

 Lagrange interpolating polynomial, 34
 Law of cosines, 4
 least squares, 32
 least squares error vector, 33
 least squares line of best fit, 33
 Length, 4
 length, 26
 linear, 8
 Linear equation, 6
 linear forms, 43
 linear independence, 47
 linear operator, 16
 linear regression, 33
 linear transformation, 16, 51
 linearly independent, 10
 lower triangular matrix, 7

 Matrix, 6
 matrix, 8
 matrix for T with respect to B , 53
 matrix representation, 17
 minor, 12
 Multiplication by scalar, 4
 Multiplication of a matrix by a scalar, 8
 Multiplication of matrices, 8
 multiplicative identity, 46

 negative definite, 44
 norm, 4, 48
 normal, 35
 normal equation, 28

 null space, 16
 nullity, 17

 one-to-one, 16, 17
 onto, 17
 ordered, 24
 orthogonal, 16, 34, 48
 orthogonal change of bases, 35
 orthogonal projection, 28
 Orthogonal vectors, 5
 orthogonally diagonalizable, 35
 orthogonally similar, 35
 orthonormal, 16, 34
 orthonormal basis, 28

 parallel component, 28
 Parallelogram law, 4
 Parametric equation of line, 5
 Parametric equation of plane, 6
 perpendicular component, 28
 polar decomposition, 57
 Polynomial Factorization, 27
 positive definite, 44
 positivity, 48
 pseudoinverse, 57
 Pythagoras theorem, 5

 quadratic forms, 43

 range, 16, 52
 rank, 17
 real, 26
 reciprocals, 25
 Reduced row echelon form, 7
 reduced singular value decomposition, 56
 regression line, 33
 relative maximum, 45
 relative minimum, 45
 residual, 33
 root, 23
 rotation, 16
 Row echelon form, 7
 row equivalent, 10

saddle point, 45
scalar multiplication, 46
Schur factorization, 36
similar, 18
similarity invariants, 19
singular value decomposition, 54
singular value expansion, 56
singular values, 54
spectral decomposition, 38
standard matrix, 53
subspace, 10, 47
symmetric matrix, 8
System of linear equations, 6

Trace, 9
transition matrix, 17
Transpose, 8
Triangle inequality, 5
triangular system, 13

Unit vector, 4
unitary, 34
upper triangular matrix, 7

Vandermonde, 34
vector space, 46
Vectors, 4

Wronskian, 47