Math 225 Linear Algebra II Lecture Notes

John C. Bowman
University of Alberta
Edmonton, Canada

May 6, 2022

Contents

1 Vectors

- *Vectors* in $\mathbb{R}^n$:

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

- *Parallelogram law*:

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

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

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

- *Dot* (*inner*) product:

$$\boldsymbol{u}\boldsymbol{\cdot}\boldsymbol{v} = u_1 v_1 + \cdots + u_n v_n.$$

- *Length* (*norm*):

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

- *Unit vector*:

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

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

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

- *Law of cosines*:

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

- *Angle* θ between $\boldsymbol{u}$ and $\boldsymbol{v}$:

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

- *Orthogonal vectors* $\boldsymbol{u}$ and $\boldsymbol{v}$:

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

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

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

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

$$|\boldsymbol{u}\cdot\boldsymbol{v}| \le |\boldsymbol{u}||\boldsymbol{v}|.$$

- *Triangle inequality*:

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

- *Component* of $\boldsymbol{v}$ in direction $\boldsymbol{u}$:

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

- *Equation of line*:

$$Ax + By = C.$$

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

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

- *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 $\boldsymbol{u}$ and $\boldsymbol{w}$ through $\boldsymbol{v}_0$:

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

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

- 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*:

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

  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:

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

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

Problem 3.2: Does a matrix necessarily compute 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

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

.

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

- *Trace* of a square matrix:

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

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

$$\mathrm{Tr}(\mathsf{AB}) = \mathrm{Tr}(\mathsf{BA}).$$

.

- *Identity* matrix:

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

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

$$\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = \mathsf{I}.$$

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

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

Problem 3.7: Prove that

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

- Inverse of a 2×2 matrix:

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

$\Rightarrow$

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

- 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.

- $\text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is the subspace formed by the set of linear combinations $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$ generated by all possible scalar multipliers $c_1, c_2, \ldots, c_n$.

- 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 $\mathbf{A}\boldsymbol{x} = \mathbf{0}$ with n unknowns is always a subspace of $\mathbb{R}^n$.

- A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ in $\mathbb{R}^n$ are *linearly independent* if the equation

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

  has only the trivial solution $c_1 = c_2 = \cdots = c_n = 0$. 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) $\mathbf{A}\boldsymbol{x} = \mathbf{0}$ has only the trivial solution;

(f) $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ has exactly one solution for all $\boldsymbol{b} \in \mathbb{R}^n$;

(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 A to the identity matrix can be applied directly to the identity matrix to yield A^{-1}.

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

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 $\mathsf{A} = [a_{ij}]_{n\times n}$ can be formally defined as:

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

- 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 $\mathsf{A} = [a_{ij}]$, the *minor* M_{ij} of the element a_{ij} is the determinant of the submatrix obtained by deleting the ith row and jth column from 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 $\det \mathbf{A}^\mathsf{T} = \det \mathbf{A}$.

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 $\det(\mathbf{A}^{-1}) = \dfrac{1}{\det \mathbf{A}}$.

- 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*:

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

- 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, $\operatorname{adj} \mathbf{A} = [(-1)^{i+j} M_{ji}]$, 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:

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

  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 $\boldsymbol{u} = (u_x, u_y, u_z)$ and $\boldsymbol{v} = (v_x, v_y, v_z)$ in $\mathbb{R}^3$ can be expressed as the determinant

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

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

5 Eigenvalues and Eigenvectors

- A *fixed point* $\boldsymbol{x}$ of a matrix A satisfies $\mathsf{A}\boldsymbol{x} = \boldsymbol{x}$.

- The zero vector is a fixed point of any matrix.

- An *eigenvector* $\boldsymbol{x}$ of a matrix A is a **nonzero** vector $\boldsymbol{x}$ that satisfies $\mathsf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ for a particular value of λ, known as an *eigenvalue* (characteristic value).

- An eigenvector $\boldsymbol{x}$ is a nontrivial solution to $(\mathsf{A}-\lambda\mathsf{I})\boldsymbol{x} = \mathsf{A}\boldsymbol{x}-\lambda\boldsymbol{x} = 0$. This requires that $(\mathsf{A} - \lambda\mathsf{I})$ be noninvertible; that is, λ must be a root of the *characteristic polynomial* $\det(\lambda\mathsf{I}-\mathsf{A})$: it must satisfy the *characteristic equation* $\det(\lambda\mathsf{I}-\mathsf{A}) = 0$.

- If $[a_{ij}]_{n\times n}$ is a triangular matrix, each of the n diagonal entries a_{ii} are eigenvalues of A:
$$\det(\lambda\mathsf{I} - \mathsf{A}) = (\lambda - a_{11})(\lambda - a_{22})\ldots(\lambda - a_{nn}).$$

- If A has eigenvalues λ_1, λ_2, ..., λ_n then $\det(\mathsf{A} - \lambda\mathsf{I}) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\ldots(\lambda_n - \lambda)$. On setting $\lambda = 0$, we see that $\det\mathsf{A}$ is the product $\lambda_1\lambda_2\ldots\lambda_n$ of the eigenvalues of A.

- Given an $n\times n$ matrix A, the terms in $\det(\lambda\mathsf{I} - \mathsf{A})$ that involve λ^n and λ^{n-1} must come from the main diagonal:
$$(\lambda - a_{11})(\lambda - a_{22})\ldots(\lambda - a_{nn}) = \lambda^n - (a_{11} + a_{22} + \ldots + a_{nn})\lambda^{n-1} + \ldots.$$
But
$$\det(\lambda\mathsf{I}-\mathsf{A}) = (\lambda-\lambda_1)(\lambda-\lambda_2)\ldots(\lambda-\lambda_n) = \lambda^n-(\lambda_1+\lambda_2+\ldots+\lambda_n)\lambda^{n-1}+\ldots+(-1)^n\lambda_1\lambda_2\ldots\lambda_n.$$
Thus, $\operatorname{Tr}\mathsf{A}$ is the sum $\lambda_1+\lambda_2+\ldots+\lambda_n$ of the eigenvalues of A and the characteristic polynomial of A always has the form
$$\lambda^n - \operatorname{Tr}\mathsf{A}\lambda^{n-1} + \ldots + (-1)^n\det\mathsf{A}.$$

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 $\mathbf{A}\boldsymbol{v}$ where $\mathbf{A}$ is the matrix in Problem 5.1 and $\boldsymbol{v} = [7, 3]^{\mathsf{T}}$, both directly and by expressing $\boldsymbol{v}$ 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 $[1, 0]^{\mathsf{T}}$ and $[0, 1]^{\mathsf{T}}$. 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 $[1, 0]^\intercal$. 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 $T : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* from $\mathbb{R}^n$ to $\mathbb{R}^m$ if for all vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ and all scalars c:

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

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

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: $T(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x}$.

- An *orthogonal* linear operator T preserves lengths: $|T(\boldsymbol{x})| = |\boldsymbol{x}|$ for all vectors $\boldsymbol{x}$b.

- A square matrix $\mathbf{A}$ with real entries is *orthogonal* if $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$; that is, $\mathbf{A}^{-1} = \mathbf{A}^\mathsf{T}$. This implies that $\det \mathbf{A} = \pm 1$ and preserves lengths:

$$|\mathbf{A}\boldsymbol{x}|^2 = \mathbf{A}\boldsymbol{x}\cdot\mathbf{A}\boldsymbol{x} = (\mathbf{A}\boldsymbol{x})^\mathsf{T}\mathbf{A}\boldsymbol{x} = \boldsymbol{x}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{A}\boldsymbol{x} = \boldsymbol{x}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{A}\boldsymbol{x} = \boldsymbol{x}^\mathsf{T}\boldsymbol{x} = \boldsymbol{x}\cdot\boldsymbol{x} = |\boldsymbol{x}|^2.$$

  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 $\boldsymbol{u}$ that are mapped to $\mathbf{0}$ by T.

- The *null space* $\operatorname{null}\mathbf{A}$ (also known as the *kernel* $\ker\mathbf{A}$) of a matrix $\mathbf{A}$ is the subspace consisting of all vectors $\boldsymbol{u}$ such that $\mathbf{A}\boldsymbol{u} = \mathbf{0}$.

- Linear transformations map subspaces to subspaces.

- The *range* $\operatorname{ran} T$ of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the image of the *domain* $\mathbb{R}^n$ under T.

- A linear transformation T is *one-to-one* if $T(\boldsymbol{u}) = T(\boldsymbol{v}) \Rightarrow \boldsymbol{u} = \boldsymbol{v}$.

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 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ of linearly independent vectors is called a *basis* for the n-dimensional subspace

$$\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} = \{c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n : (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n\}.$$

- If $\boldsymbol{w} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \ldots + \boldsymbol{v}_n$ is a vector in $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ with basis $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ we say that $[a_1, a_2, \ldots, a_n]^\intercal$ are the *coordinates* $[\boldsymbol{w}]_B$ of the vector $\boldsymbol{w}$ with respect to the basis B.

- If $\boldsymbol{w}$ is a vector in $\mathbb{R}^n$ the *transition matrix* $\mathbf{P}_{B'\leftarrow B}$ that changes coordinates $[\boldsymbol{w}]_B$ with respect to the basis $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ for $\mathbb{R}^n$ to coordinates $[\boldsymbol{w}]_{B'} = \mathbf{P}_{B'\leftarrow B}[\boldsymbol{w}]_B$ with respect to the basis $B' = \{\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n\}$ for $\mathbb{R}^n$ is

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

- 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 $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator and if $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ and $B' = \{\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n\}$ 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 $\operatorname{rank}\mathbf{A}$. That is, the rank of a matrix is the dimension $\dim(\operatorname{row}(\mathbf{A}))$ of the span of its rows. Equivalently, the rank of a matrix is the dimension $\dim(\operatorname{col}(\mathbf{A}))$ of the span of its columns.

- The dimension of the null space of a matrix $\mathbf{A}$ is known as its *nullity* and written $\dim(\operatorname{null}(\mathbf{A}))$ or $\operatorname{nullity}\mathbf{A}$.

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

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

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

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

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

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

- 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

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

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

- 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 $\mathbf{A}\boldsymbol{x} = 0$ 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 $\operatorname{rank}\mathbf{P}^{-1}\mathbf{A} \le \operatorname{rank}\mathbf{A}$. Likewise $\operatorname{rank}\mathbf{B} = \operatorname{rank}\mathbf{P}^{-1}\mathbf{A}\mathbf{P} \le \operatorname{rank}\mathbf{P}^{-1}\mathbf{A} \le \operatorname{rank}\mathbf{A}$. But since we also know that $\mathbf{A}$ is similar to $\mathbf{B}$, we see that $\operatorname{rank}\mathbf{A} \le \operatorname{rank}\mathbf{B}$. Hence $\operatorname{rank}\mathbf{A} = \operatorname{rank}\mathbf{B}$.

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*:

(a) rank;

(b) nullity;

(c) characteristic polynomial and eigenvalues, including their algebraic and geometric multiplicities;

(d) determinant;

(e) 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 A is diagonalizable if and only if A has n linearly independent eigenvectors.* Suppose that $\mathsf{P}^{-1}\mathsf{A}\mathsf{P} = \mathsf{D}$, where

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

  Since P is invertible, its columns are nonzero and linearly independent. Moreover,

$$\mathsf{AP} = \mathsf{PD} = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix}.$$

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

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

  This says, for each j, that the nonzero column vector $[p_{1j}, p_{2j}, \ldots, p_{nj}]^\intercal$ is an eigenvector of 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 A has n linearly independent eigenvectors, we can use them to form an *eigenvector matrix* P such that $\mathsf{AP} = \mathsf{PD}$.

- 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 $[1, 0]^\intercal$ (i.e. the eigenvalue λ has algebraic multiplicity two but geometric multiplicity one) and consequently is *not* diagonalizable.

- *Eigenvectors of a matrix* 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 $\boldsymbol{v}_{k+1}$ is the first eigenvector that is expressible as a linear combination of the others:
$$\boldsymbol{v}_{k+1} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k, \tag{1}$$
where the coefficients c_i are not all zero. The condition that $\boldsymbol{v}_{k+1}$ is the first such vector guarantees that the vectors on the right-hand side are linearly independent. If each eigenvector $\boldsymbol{v}_i$ corresponds to the eigenvalue λ_i, then on multiplying by A on the left, we find that
$$\lambda_{k+1}\boldsymbol{v}_{k+1} = \mathsf{A}\boldsymbol{v}_{k+1} = c_1\mathsf{A}\boldsymbol{v}_1 + c_2\mathsf{A}\boldsymbol{v}_2 + \cdots + c_k\mathsf{A}\boldsymbol{v}_k = c_1\lambda_1\boldsymbol{v}_1 + c_2\lambda_2\boldsymbol{v}_2 + \cdots + c_k\lambda_k\boldsymbol{v}_k.$$
If we multiply Eq.(1) by λ_{k+1} we obtain:
$$\lambda_{k+1}\boldsymbol{v}_{k+1} = c_1\lambda_{k+1}\boldsymbol{v}_1 + c_2\lambda_{k+1}\boldsymbol{v}_2 + \cdots + c_k\lambda_{k+1}\boldsymbol{v}_k,$$
The difference of the previous two equations yields
$$\mathbf{0} = c_1(\lambda_1 - \lambda_{k+1})\boldsymbol{v}_1 + (\lambda_2 - \lambda_{k+1})c_2\boldsymbol{v}_2 + \cdots + (\lambda_k - \lambda_{k+1})c_k\boldsymbol{v}_k,$$
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 A corresponding to distinct eigenvalues must in fact be independent!

- An $n \times n$ matrix 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
$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$$
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 A may be decomposed as
$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix}\frac{1}{5}\begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}.$$

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

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

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

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

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 $i^2 = -1$:

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

Then

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

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

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

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

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
$$|zw|^2 = zw\overline{zw} = z\overline{z}w\overline{w} = |z|^2 |w|^2.$$
Thus $|zw| = |z|\,|w|$ for all $z, w \in \mathbb{C}$.

- The *dot product* of vectors $\boldsymbol{u} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ and $\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ in $\mathbb{C}^n$ is given by
$$\boldsymbol{u}\boldsymbol{\cdot}\boldsymbol{v} = \boldsymbol{u}^{\mathsf{T}}\bar{\boldsymbol{v}} = u_1\overline{v_1} + \cdots + u_n\overline{v_n}.$$

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

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

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

- In view of the definition of the complex modulus, it is sensible to define the *length* of a vector $\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ in $\mathbb{C}^n$ to be

$$|\boldsymbol{v}| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}} = \sqrt{\boldsymbol{v}^\mathsf{T}\bar{\boldsymbol{v}}} = \sqrt{v_1\bar{v_1} + \cdots + v_n\bar{v_n}} = \sqrt{|v_1|^2 + \cdots + |v_n|^2} \geq 0.$$

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

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

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

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 if $\bar{\lambda}$. Moreover, if $\boldsymbol{x}$ is an eigenvector corresponding to λ, then $\bar{\boldsymbol{x}}$ is an eigenvector corresponding to $\bar{\lambda}$:

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

- 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 $\boldsymbol{x} \neq \mathbf{0}$ must satisfy

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

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

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

where $Q(z) = \sum_{k=0}^{n} a_k (z^{k-1} + z^{k-2}z_0 + \ldots + z_0^{k-1})$ 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, ..., z_n and can be factorized as $P(z) = A(z - z_1)(z - z_2)\ldots(z - z_n)$, 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 a vector $\boldsymbol{v}$ in the direction $\hat{\boldsymbol{u}}$ is given by

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

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

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

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

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

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

- In fact, for a general n-dimensional subspace W of $\mathbb{R}^m$, every vector $\boldsymbol{v}$ has a unique decomposition as a sum of a vector $\boldsymbol{v}_\parallel$ in W and a vector $\boldsymbol{v}_\perp$ 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 $\mathbf{A}^\mathsf{T}$. The projection $\boldsymbol{v}_\parallel$ of $\boldsymbol{v} = \boldsymbol{v}_\parallel + \boldsymbol{v}_\perp$ onto W satisfies $\boldsymbol{v}_\parallel = \mathbf{A}\boldsymbol{x}$ for some vector $\boldsymbol{x}$ and hence

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

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

$$\mathbf{A}^\mathsf{T}\mathbf{A}\boldsymbol{x} = \mathbf{A}^\mathsf{T}\boldsymbol{v}, \tag{2}$$

  so that $\boldsymbol{x} = (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}\boldsymbol{v}$. We know from Question 4 of Assignment 1 that $\mathbf{A}^\mathsf{T}\mathbf{A}$ 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 $\mathbf{A}^\mathsf{T}\mathbf{A}$ is invertible, and there is indeed a unique solution for $\boldsymbol{x}$. The *orthogonal projection* $\boldsymbol{v}_\parallel$ of $\boldsymbol{v}$ onto the subspace W is thus given by the formula

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

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 $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$. In this case the projection of $\boldsymbol{v}$ onto W reduces to multiplication by the symmetric matrix $\mathbf{A}\mathbf{A}^\mathsf{T}$:

$$\boldsymbol{v}_\parallel = \mathbf{A}\mathbf{A}^\mathsf{T}\boldsymbol{v}.$$

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

Compute

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

As desired, we note that $\boldsymbol{v} - \boldsymbol{v}_{\|} = \left[\frac{21}{25}, 0, \frac{28}{25}\right] = \frac{7}{25}[3, 0, 4]$ is parallel to the normal $\left[\frac{3}{5}, 0, \frac{4}{5}\right]$ 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 $\operatorname{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ of $\mathbb{R}^m$:

1. Start with the first vector: $\boldsymbol{q}_1 = \boldsymbol{a}_1$.

2. Remove the component of the second vector parallel to the first:
$$\boldsymbol{q}_2 = \boldsymbol{a}_2 - \frac{\boldsymbol{a}_2 \boldsymbol{\cdot} \boldsymbol{q}_1}{|\boldsymbol{q}_1|^2}\boldsymbol{q}_1.$$

3. Remove the components of the third vector parallel to the first and second:
$$\boldsymbol{q}_3 = \boldsymbol{a}_3 - \frac{\boldsymbol{a}_3 \boldsymbol{\cdot} \boldsymbol{q}_1}{|\boldsymbol{q}_1|^2}\boldsymbol{q}_1 - \frac{\boldsymbol{a}_3 \boldsymbol{\cdot} \boldsymbol{q}_2}{|\boldsymbol{q}_2|^2}\boldsymbol{q}_2.$$

4. Continue this procedure by successively defining, for $k = 4, 5, \ldots, n$:
$$\boldsymbol{q}_k = \boldsymbol{a}_k - \frac{\boldsymbol{a}_k \boldsymbol{\cdot} \boldsymbol{q}_1}{|\boldsymbol{q}_1|^2}\boldsymbol{q}_1 - \frac{\boldsymbol{a}_k \boldsymbol{\cdot} \boldsymbol{q}_2}{|\boldsymbol{q}_2|^2}\boldsymbol{q}_2 - \cdots - \frac{\boldsymbol{a}_k \boldsymbol{\cdot} \boldsymbol{q}_{k-1}}{|\boldsymbol{q}_{k-1}|^2}\boldsymbol{q}_{k-1}. \tag{3}$$

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

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

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

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

$$\boldsymbol{q}_k \boldsymbol{\cdot} \boldsymbol{q}_j = \boldsymbol{a}_k \boldsymbol{\cdot} \boldsymbol{q}_j - \frac{\boldsymbol{a}_k \boldsymbol{\cdot} \boldsymbol{q}_j}{|\boldsymbol{q}_j|^2} |\boldsymbol{q}_j|^2 = 0, \qquad j = 1, 2, \ldots k-1.$$

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 $\boldsymbol{q}_1 = \boldsymbol{a}_1$ is a linear combination of the original $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$.

Remark: At the kth stage of the orthonormalization, if all of the previously created vectors were a linear combination of the original vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$, then by Eq. (3), we see that $\boldsymbol{q}_k$ 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 $\boldsymbol{q}_k$ can be written as a linear combination of the original basis vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. Since these vectors are given to be linearly independent, we thus know that $\boldsymbol{q}_k \neq \boldsymbol{0}$. This is what allows us to normalize $\boldsymbol{q}_k$ in Step 5. It also implies that $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n\}$ spans the same space as $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$.

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:

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

The Gram-Schmidt process then yields

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

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

12 QR Factorization

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

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

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

$$\mathbf{0} \neq \boldsymbol{q}_k\cdot\hat{\boldsymbol{q}}_k = \boldsymbol{a}_k\cdot\hat{\boldsymbol{q}}_k \tag{5}$$

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

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

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

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

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 $\boldsymbol{a}_k$ and $\hat{\boldsymbol{q}}_k$ by the $m \times n$ matrices $\mathbf{A}$ and $\mathbf{Q}$, respectively, we can express this result as the so-called $\mathbf{QR}$ *factorization* of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{QR}. \tag{6}$$

Remark: Every matrix $m \times n$ matrix $\mathbf{A}$ with full column rank (linearly independent columns) thus has a $\mathbf{QR}$ 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 $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}$. An efficient way to find $\mathbf{R}$ is to premultiply both sides of Eq. 6 by $\mathbf{Q}^\mathsf{T}$:

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

Problem 12.1: Find the $\mathbf{QR}$ 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 $i = 1, 2, \ldots n$ 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

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

If the x_is are distinct, then the $n \times 2$ coefficient matrix 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 $\boldsymbol{b}$ on the right-hand side is in the column space of A.

- Recall that if $\boldsymbol{b}$ is in the column space of A, the linear system of equations $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$ is *consistent*, that is, it has at least one solution $\boldsymbol{x}$. Recall that this solution is unique iff A has full column rank.

- In practical applications, however, there may be measurement errors in the entries of A and $\boldsymbol{b}$ so that $\boldsymbol{b}$ does not lie exactly in the column space of A.

- The *least squares* approximation to an inconsistent system of equations $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$ is the solution to

  $$\mathsf{A}\boldsymbol{x} = \boldsymbol{b}_{\|}$$

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

  $$\mathsf{A}^{\mathsf{T}}\mathsf{A}\boldsymbol{x} = \mathsf{A}^{\mathsf{T}}\boldsymbol{b}.$$

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

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

and

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

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 $\boldsymbol{b}_\perp = \boldsymbol{b} - \boldsymbol{b}_\parallel$ 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 $\boldsymbol{x} = [\alpha, \beta]$ minimizes the least squares error vector

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

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

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

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

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

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

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

$$y = a_0 + a_1 x + \cdots + a_m x^m$$

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 $\mathbf{A}^T = \overline{\mathbf{A}}$.

- For real matrices, being Hermitian is equivalent to being symmetric.

- On defining the *Hermitian transpose* $\mathbf{A}^\dagger = \overline{\mathbf{A}^T}$, 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 $\mathbf{A}^T \mathbf{A} = \mathbf{I}$.

- 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:

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

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

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

  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*

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

  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 $\boldsymbol{x}_1$. Now use the Gram-Schmidt process to construct some orthonormal basis $\{\boldsymbol{x}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$ for $\mathbb{R}^n$. Note here that the first basis vector is $\boldsymbol{x}_1$ itself. In this coordinate system, the eigenvector $\boldsymbol{x}_1$ has the components $[1, 0, \ldots, 0]^\mathsf{T}$, 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 $\boldsymbol{x}_2$. Now construct an orthogonal basis $\{\boldsymbol{x}_2, \boldsymbol{b}_3, \ldots, \boldsymbol{b}_n\}$ for $\mathbb{R}^{n-1}$. In this new coordinate system $\boldsymbol{x}_2 = [\lambda_2, 0, \ldots, 0]^\mathsf{T}$ 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 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{b}_3, \ldots, \boldsymbol{b}_n\}$, $\mathbf{A}$ 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 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_n\}$ 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 $\mathbf{A} = \mathbf{PUP}^\dagger$, 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 $i \le j$ be the nonzero elements of $\mathbf{U}$, we can write out for each $i = 1, \ldots, n$ the diagonal elements $\sum_{k=1}^{i} U^\dagger_{ik} U_{ki} = \sum_{k=1}^{i} \overline{U_{ki}} U_{ki}$ 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 $\boldsymbol{x} \neq \mathbf{0}$ must satisfy

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

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 $|\mathbf{A}\boldsymbol{x}|^2 = \boldsymbol{x}^\dagger \mathbf{A}^\dagger \mathbf{A}\boldsymbol{x} = \boldsymbol{x}^\dagger \mathbf{A}\mathbf{A}^\dagger \boldsymbol{x} = \left|\mathbf{A}^\dagger \boldsymbol{x}\right|^2$ for all vectors $\boldsymbol{x} \in \mathbb{R}^n$.

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 $\mathbf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, then $0 = |(\mathbf{A} - \lambda\mathbf{I})\boldsymbol{x}| = |(\mathbf{A}^\dagger - \overline{\lambda}\mathbf{I})\boldsymbol{x}|$, so that $\mathbf{A}^\dagger \boldsymbol{x} = \overline{\lambda}\boldsymbol{x}$.

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

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

  so that $\boldsymbol{x}\boldsymbol{\cdot}\boldsymbol{y} = 0$ 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 $\mathbf{AP} = \mathbf{PD}$, 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 $\lambda_1, \lambda_2, \ldots, \lambda_n$ corresponding to the orthonormal matrix $[\,\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_n\,]$ then $\mathbf{A}$ has the *spectral decomposition*

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

- 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 $\lambda^n + c_{n-1}\lambda^{n-1} + \ldots c_1\lambda + c_0 = 0$. Specifically, the matrix $\mathbf{A}$ satisfies its characteristic equation in the sense that
$$\mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \ldots + c_1\mathbf{A} + c_0\mathbf{I} = \mathbf{0}. \tag{8}$$

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

- We can also easily compute other functions of diagonalizable matrices. Give an analytic function f with a Taylor series
$$f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + f'''(0)\frac{x^3}{3!} + \ldots,$$
and a diagonalizable matrix $\mathbf{A} = \mathbf{PDP}^{-1}$, where $\mathbf{D}$ is diagonal and $\mathbf{P}$ is invertible, one defines
$$f(\mathbf{A}) = f(0)\mathbf{I} + f'(0)\mathbf{A} + f''(0)\frac{\mathbf{A}^2}{2!} + f'''(0)\frac{\mathbf{A}^3}{3!} + \ldots. \tag{9}$$
On recalling that $\mathbf{A}^n = \mathbf{PD}^n\mathbf{P}^{-1}$ and $\mathbf{A}^0 = \mathbf{I}$, one can rewrite this relation as
$$\begin{aligned} f(\mathbf{A}) &= f(0)\mathbf{PD}^0\mathbf{P}^{-1} + f'(0)\mathbf{PDP}^{-1} + f''(0)\frac{\mathbf{PD}^2\mathbf{P}^{-1}}{3!} + f'''(0)\frac{\mathbf{PD}^3\mathbf{P}^{-1}}{3!} + \ldots \\ &= \mathbf{P}\left[f(0)\mathbf{D}^0 + f'(0)\mathbf{D} + f''(0)\frac{\mathbf{D}^2}{3!} + f'''(0)\frac{\mathbf{D}^3}{3!} + \ldots\right]\mathbf{P}^{-1} \\ &= \mathbf{P}f(\mathbf{D})\mathbf{P}^{-1}. \end{aligned} \tag{10}$$
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:

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

  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

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

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 $\boldsymbol{y}' = \mathbf{A}\boldsymbol{y}$ 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 $\boldsymbol{x}$ is an eigenvector of $\mathbf{A}$ corresponding to the eigenvalue λ, then $\boldsymbol{y} = e^{\lambda t}\boldsymbol{x}$ is a solution of $\boldsymbol{y}' = \mathbf{A}\boldsymbol{y}$:

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

- If $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, …, $\boldsymbol{x}_k$, are linearly independent eigenvectors of $\mathbf{A}$ corresponding to the (not necessarily distinct) eigenvalues λ_1, λ_2, …, λ_k, then

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

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

$$0 = c_1 e^{\lambda_1 t}\boldsymbol{x}_1 + c_2 e^{\lambda_2 t}\boldsymbol{x}_2 + \ldots + c_k e^{\lambda_k t}\boldsymbol{x}_k$$

  then at $t = 0$ we find

$$0 = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \ldots + c_k\boldsymbol{x}_k;$$

  the linear independence of the k eigenvectors then implies that

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

- If $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, …, $\boldsymbol{x}_n$, are linearly independent eigenvectors of an $n \times n$ matrix $\mathbf{A}$ corresponding to the (not necessarily distinct) eigenvalues λ_1, λ_2, …, λ_n, then every solution to $\boldsymbol{y}' = \mathbf{A}\boldsymbol{y}$ can be expressed in the form

$$\boldsymbol{y} = c_1 e^{\lambda_1 t}\boldsymbol{x}_1 + c_2 e^{\lambda_2 t}\boldsymbol{x}_2 + \cdots + c_n e^{\lambda_n t}\boldsymbol{x}_n,$$

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

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

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

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

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

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

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

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

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

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 $\boldsymbol{y}' = \mathbf{A}\boldsymbol{y}$ corresponding to the initial condition $\boldsymbol{y}_0 = [1, -1]$, where

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

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

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

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):

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

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, ..., x_n such as
$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = \sum_{i=1}^{n} a_ix_i,$$
where a_1, a_2, ..., 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 $\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_n]$, so that
$$\begin{aligned}
\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x} &= \sum_{i,j} x_ia_{ij}x_j \\
&= \sum_{i<j} a_{ij}x_ix_j + \sum_{i=j} a_{ij}x_ix_j + \sum_{i>j} a_{ij}x_ix_j \\
&= \sum_{i=j} a_{ij}x_ix_j + \sum_{i<j} a_{ij}x_ix_j + \sum_{i<j} a_{ji}x_jx_i \\
&= \sum_{i=j} a_{ii}x_i^2 + 2\sum_{i<j} a_{ij}x_ix_j.
\end{aligned}$$

Remark: Note that $\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x} = \boldsymbol{x}\cdot\mathbf{A}\boldsymbol{x} = \mathbf{A}\boldsymbol{x}\cdot\boldsymbol{x}$.

- (Principle Axes Theorem) Consider the change of variable $\boldsymbol{x} = \mathbf{P}\boldsymbol{y}$, where $\mathbf{P}$ orthogonally diagonalizes $\mathbf{A}$. Then in terms of the eigenvalues λ_1, λ_2, ..., λ_n of $\mathbf{A}$:
$$\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x} = (\mathbf{P}\boldsymbol{y})^\mathsf{T}\mathbf{A}\mathbf{P}\boldsymbol{y} = \boldsymbol{y}^\mathsf{T}(\mathbf{P}^\mathsf{T}\mathbf{A}\mathbf{P})\boldsymbol{y} = \lambda_1y_1^2 + \lambda_2y_2^2 + \ldots + \lambda_ny_n^2.$$

- (Constrained Extremum Theorem) For a symmetric $n \times n$ matrix $\mathbf{A}$, the minimum (maximum) value of
$$\{\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x} : |\boldsymbol{x}|^2 = 1\}$$
is given by the smallest (largest) eigenvalue of $\mathbf{A}$ and occurs at the corresponding unit eigenvector. To see this, let $\boldsymbol{x} = \mathbf{P}\boldsymbol{y}$, where $\mathbf{P}$ orthogonally diagonalizes $\mathbf{A}$, and list the eigenvalues in ascending order: $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$. Then
$$\lambda_1 = \lambda_1(y_1^2+y_2^2+\ldots+y_n^2) \le \underbrace{\lambda_1y_1^2 + \lambda_2y_2^2 + \ldots + \lambda_ny_n^2}_{\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x}} \le \lambda_n(y_1^2+y_2^2+\ldots+y_n^2) = \lambda_n.$$

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 A is said to be

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

Remark: If A is a symmetric matrix then $\boldsymbol{x}^\mathsf{T}\mathsf{A}\boldsymbol{x}$

- is *positive definite* if and only if all eigenvalues of A are positive;
- is *negative definite* if and only if all eigenvalues of A are negative;
- is *indefinite* if and only if A has at least one positive eigenvalue and at least one negative eigenvalue.

- The kth principal submatrix of an $n \times n$ matrix A is the submatrix composed of the first k rows and columns of A.

- (Sylvester's Criterion) If A is a symmetric matrix then A is *positive definite* if and only if the determinant of every principal submatrix of A is positive.

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

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

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 A are positive.

Remark: If A is a symmetric matrix then the following are equivalent:

(a) A is positive definite;

(b) there is a symmetric positive definite matrix B such that $\mathsf{A} = \mathsf{B}^2$;

(c) there is an invertible matrix C such that $\mathsf{A} = \mathsf{C}^\mathsf{T}\mathsf{C}$.

Remark: If $\mathbf{A}$ is a symmetric 2×2 matrix then the equation $\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x} = 1$ can be expressed in the orthogonal coordinate system $\boldsymbol{y} = \mathbf{P}^\mathsf{T}\boldsymbol{x}$ as $\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1$. Thus $\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x}$ 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 $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ in V and scalars c, d in $\mathbb{R}$ ($\mathbb{C}$) the following *axioms* hold:

(A1) $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$ (*associative*);

(A2) $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$ (*commutative*);

(A3) $\boldsymbol{u}+\mathbf{0}=\mathbf{0}+\boldsymbol{u}=\boldsymbol{u}$ (*additive identity*);

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

(A5) $1\boldsymbol{u}=\boldsymbol{u}$ (*multiplicative identity*);

(A6) $c(d\boldsymbol{u})=(cd)\boldsymbol{u}$ (*scalar multiplication*);

(A7) $(c+d)\boldsymbol{u}=c\boldsymbol{u}+d\boldsymbol{u}$ (*distributive*);

(A8) $c(\boldsymbol{u}+\boldsymbol{v})=c\boldsymbol{u}+c\boldsymbol{v}$ (*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 $a_0+a_1x+\ldots+a_nx^n$ of degree less than or equal to n (where a_0, a_1, $\ldots$, 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 mth 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 $f_1(x)$, $f_2(x)$, …, $f_n(x)$ 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, …, c_n such that the function

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

is zero for all $x \in \mathbb{R}$, so that $g^{(k)}(x) = 0$ for $k = 0, 1, 2, \ldots n - 1$ 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 $\boldsymbol{u}$ and $\boldsymbol{v}$ in V to a unique real (complex) number $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ such that, for all vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ in V and scalars c in $\mathbb{R}$ ($\mathbb{C}$):

(I1) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \overline{\langle \boldsymbol{v}, \boldsymbol{u} \rangle}$ (*conjugate symmetry*);

(I2) $\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$ (*additivity*);

(I3) $\langle c\boldsymbol{u}, \boldsymbol{v} \rangle = c \langle \boldsymbol{u}, \boldsymbol{v} \rangle$ (*homogeneity*);

(I4) $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$ and $\langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0$ iff $\boldsymbol{v} = 0$ (*positivity*).

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

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

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

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

- Every inner product on $\mathbb{R}^n$ can be uniquely expressed as $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^\mathsf{T}\mathbf{A}\boldsymbol{v}$ for some positive definite symmetric matrix $\mathbf{A}$, so that $\langle \boldsymbol{v}, \boldsymbol{v} \rangle = \boldsymbol{v}^\mathsf{T}\mathbf{A}\boldsymbol{v} > 0$ for $\neq 0$.

- In a vector space V with inner product $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$, the *norm* $|\boldsymbol{v}|$ of a vector $\boldsymbol{v}$ is given by $\sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$, the *distance* $d(\boldsymbol{u}, \boldsymbol{v})$ is given by $|\boldsymbol{u} - \boldsymbol{v}| = \sqrt{\langle \boldsymbol{u} - \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v} \rangle}$, and the *angle* θ between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfies $\cos\theta = \langle \boldsymbol{u}, \boldsymbol{v} \rangle$. Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are *orthogonal* if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$. 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 than 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 $\{u_n\}_{n=0}^{\infty} = \{\frac{c_0}{\sqrt{2}}, c_1, s_1, c_2, s_2, \ldots\}$, 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

$$\sin(nx + mx) - \sin(nx - mx) = 2\cos nx \sin mx, \tag{11}$$

$$\cos(nx + mx) + \cos(nx - mx) = 2\cos nx \cos mx \tag{12}$$

$$\cos(nx - mx) - \cos(nx + mx) = 2\sin nx \sin mx \tag{13}$$

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*:

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

in terms of the *Fourier coefficients*

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

and

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

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:

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

19 General linear transformations

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

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

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

Problem 19.1: Show that every linear transformation T satisfies

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

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

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 $T : V \to W$ 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 $\boldsymbol{u}$, let $T(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{u} \rangle$ for $\boldsymbol{x} \in V$. Show that the kernel of T is the set of vectors that are orthogonal to $\boldsymbol{u}$ and that the range of T is 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 $T : V \to W$ 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 $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$, we may uniquely express $\boldsymbol{v} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \ldots + c_n\boldsymbol{u}_n$; the linear mapping $T(\boldsymbol{v}) = (c_1, c_2, \ldots, c_n)$ 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 $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator and $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a basis for $\mathbb{R}^n$. The matrix
$$\mathbf{A} = [[T(\boldsymbol{v}_1)]_B \ [T(\boldsymbol{v}_2)]_B \ \cdots \ [T(\boldsymbol{v}_n)]_B]$$
is called the *matrix for T with respect to B*, with
$$[T(\boldsymbol{x})]_B = \mathbf{A}[\boldsymbol{x}]_B$$
for all $\boldsymbol{x} \in \mathbb{R}^n$. 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 $B' = \{\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n\}$ is any basis for R^n, then
$$[T(\boldsymbol{x})]_{B'} = \mathbf{P}_{B' \leftarrow B}[T(\boldsymbol{x})]_B \mathbf{P}_{B \leftarrow B'}.$$

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 $\boldsymbol{v}$ of $\mathbf{A}^\dagger\mathbf{A}$ must be real and non-negative since
$$0 \leq |\mathbf{A}\boldsymbol{v}|^2 = \mathbf{A}\boldsymbol{v}\boldsymbol{\cdot}\mathbf{A}\boldsymbol{v} = (\mathbf{A}\boldsymbol{v})^\dagger\mathbf{A}\boldsymbol{v} = \boldsymbol{v}^\dagger\mathbf{A}^\dagger\mathbf{A}\boldsymbol{v} = \boldsymbol{v}^\dagger\lambda\boldsymbol{v} = \lambda\boldsymbol{v}\boldsymbol{\cdot}\boldsymbol{v} = \lambda|\boldsymbol{v}|^2.$$

Problem 20.1: Show that null($\mathbf{A}^\dagger\mathbf{A}$) = null($\mathbf{A}$) and conclude from the dimension theorem that rank($\mathbf{A}^\dagger\mathbf{A}$) = 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 $\mathbf{V} = [\,\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \ldots \quad \boldsymbol{v}_n\,]$ 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 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$ but $\sigma_{k+1} = \sigma_{k+2} = \cdots = \sigma_n = 0$. Since
$$\overline{\mathbf{A}\boldsymbol{v}_i\boldsymbol{\cdot}\mathbf{A}\boldsymbol{v}_j} = (\mathbf{A}\boldsymbol{v}_i)^\dagger\mathbf{A}\boldsymbol{v}_j = \boldsymbol{v}_i^\dagger\mathbf{A}^\dagger\mathbf{A}\boldsymbol{v}_j = \boldsymbol{v}_i^\dagger\sigma_j^2\boldsymbol{v}_j = \sigma_j^2\overline{\boldsymbol{v}_i\boldsymbol{\cdot}\boldsymbol{v}_j},$$
we see that the vectors
$$\boldsymbol{u}_j = \frac{\mathbf{A}\boldsymbol{v}_j}{|\mathbf{A}\boldsymbol{v}_j|} = \frac{\mathbf{A}\boldsymbol{v}_j}{\sigma_j}, \qquad j = 1, 2, \ldots, k$$
are also orthonormal, so that the matrix
$$[\,\boldsymbol{u}_1 \quad \boldsymbol{u}_2 \quad \ldots \quad \boldsymbol{u}_k\,]$$

is orthogonal. We also see that $\mathbf{A}\boldsymbol{v}_j = \mathbf{0}$ for $k < j \le n$. 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

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

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

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

is the $m \times n$ matrix

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

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 $\mathbf{A}$:

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

The positive values σ_j for $j = 1, 2, \ldots k$ 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

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

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

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

form the columns of $\mathbf{V}$.

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

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

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

We now look for a third vector which is perpendicular to scaled versions of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, namely $\sqrt{6}\boldsymbol{u}_1 = [2, 1, 1]$ and $\sqrt{2}\boldsymbol{u}_2 = [0, -1, 1]$. 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

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

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

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

- The first k columns of the matrix $\mathbf{U}$ form an orthonormal basis for $\mathrm{col}(\mathbf{A})$, whereas the remaining $m-k$ columns form an orthonormal basis for $\mathrm{col}(\mathbf{A})^{\perp} = \mathrm{null}(\mathbf{A}^{\mathsf{T}})$.

- The first k columns of the matrix $\mathbf{V}$ form an orthonormal basis for $\mathrm{row}(\mathbf{A})$, whereas the remaining $n-k$ columns form an orthonormal basis for $\mathrm{row}(\mathbf{A})^{\perp} = \mathrm{null}(\mathbf{A})$.

- If $\mathbf{A}$ is a nonzero matrix, the singular value decomposition may be expressed more efficiently as $\mathbf{A} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^{\dagger}$, in terms of the $m \times k$ matrix

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

the diagonal $k \times k$ matrix

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

and the $n \times k$ matrix

$$\mathbf{V}_1 = [\,\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \ldots \quad \boldsymbol{v}_k\,].$$

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

Problem 20.5: Show that the matrix $\mathbf{\Sigma}_1$ 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*

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

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*

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

where $\mathbf{P} = \mathbf{U\Sigma U}^\dagger$ is a positive semidefinite matrix of rank k and $\mathbf{Q} = \mathbf{UV}^\dagger$ 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 decompostion:

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

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 pseudinverse A^+ of an $m \times n$ matrix A satisfies the following properties:

(a) $\mathsf{A}\mathsf{A}^+\mathsf{A} = \mathsf{A}$;

(b) $\mathsf{A}^+\mathsf{A}\mathsf{A}^+ = \mathsf{A}^+$;

(c) $\mathsf{A}^\dagger\mathsf{A}\mathsf{A}^+ = \mathsf{A}^\dagger$;

(d) $\mathsf{A}^+\mathsf{A}$ and $\mathsf{A}\mathsf{A}^+$ are Hermitian;

(e) $(\mathsf{A}^+)^\dagger = (\mathsf{A}^\dagger)^+$.

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

$$\mathsf{A}^+ = \left(\mathsf{A}^\dagger\mathsf{A}\right)^{-1}\mathsf{A}^\dagger.$$

- An important application of the pseudoinverse of a matrix A is in solving the least-squares problem
$$\mathsf{A}^\dagger\mathsf{A}\boldsymbol{x} = \mathsf{A}^\dagger\boldsymbol{b}$$
since the solution $\boldsymbol{x}$ of minimum norm can be conveniently expressed in terms of the pseudoinverse:
$$\boldsymbol{x} = \mathsf{A}^+\boldsymbol{b}.$$

Index