Linear Algebra Concise Review

John Yang

April 29, 2022

Contents

1 Vectors 3
- 1.1 The Geometry and Algebra of Vectors 3
- 1.2 Length and Angle: the Dot Product 3
- 1.3 Lines and Planes 4
- 1.4 Applications 5

2 Systems of Linear Equations 5
- 2.1 Introduction to Systems of Linear Equations 5
- 2.2 Direct Methods for Solving Linear Systems 6
- 2.3 Spanning Sets and Linear Independence 7
- 2.4 Applications 7
- 2.5 Iterative Methods for Solving Linear Systems 7

3 Matrices 7
- 3.1 Matrix Operations 7
- 3.2 Matrix Algebra 8
- 3.3 The Inverse of a Matrix 9
- 3.4 The LU Factorization 10
- 3.5 Subspaces, Basis, Dimension, and Rank 10
- 3.6 Introduction to Linear Transformations 12
- 3.7 Applications 13

4 Eigenvalues and Eigenvectors 13
- 4.1 Introduction to Eigenvalues and Eigenvectors 13
- 4.2 Determinants 13
- 4.3 Eigenvalues and Eigenvectors of $n \times n$ Matrices 15
- 4.4 Similarity and Diagonalization 16
- 4.5 Iterative Methods for Computing Eigenvalues 17
- 4.6 Applications and the Perron-Frobenius Theorem 18

5 Orthogonality 19
- 5.1 Orthogonality in $\mathbb{R}^n$ 19
- 5.2 Orthogonal Complements and Orthogonal Projections 20
- 5.3 The Gram-Schmidt Process and the QR Factorization 21
- 5.4 Orthogonal Diagonalization of Symmetric Matrices 21
- 5.5 Applications 21

6 Vector Spaces 22
- 6.1 Vector Spaces and Subspaces 22
- 6.2 Linear Independence, Basis, and Dimension 23
- 6.3 Change of Basis 24
- 6.4 Linear Transformations 25
- 6.5 The Kernel and Range of a Linear Transformation 25

6.6 The Matrix of a Linear Transformation 26
6.7 Applications 27

7 Distance and Approximation 28
7.1 Inner Product Spaces 28
7.2 Norms and Distance Functions 28
7.3 Least Squares Approximation 29
7.4 The Singular Value Decomposition 30
7.5 Applications 31

1 Vectors

1.1 The Geometry and Algebra of Vectors

- A vector is a directed line segment that corresponds to a displacement from one point A to another point B.
- Column vectors and row vectors are different ways to express the same thing:

$$[3,2] = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

- The point is that components of vectors are ordered.
- Two vectors are equal if they have the same magnitude and direction. Two vectors can still be equal if they have different initial and terminal points.
- Standard position of a vector - when the initial point is at the origin.
- Sum $\mathbf{u} + \mathbf{v} = [u_1 + v_1, u_2 + v_2]$
- Place vectors from head to tail.
- Scalar multiples: $c\mathbf{v} = [cv_1, cv_2]$ aka scaling a vector
- Subtraction is just adding the negative.
- Properties of vectors in $\mathbb{R}^n$: let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and let c and d be scalars. Then:
  - $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
  - $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
  - $\mathbf{u} + \mathbf{0} = \mathbf{u}$
  - $\mathbf{u} + (-\mathbf{u}) = 0$
  - $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
  - $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
  - $c(d\mathbf{u}) = (cd)\mathbf{u}$
  - $1\mathbf{u} = \mathbf{u}$
- A vector $\mathbf{v}$ is a linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ if there are scalars $c_1, c_2, \cdots, c_k$ such that $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$. Those scalars are called the coefficients of the linear combination.
- Binary vectors - the components are either 0 or 1.
- Modulus function - divide by a given number and you're left with the remainder.

1.2 Length and Angle: the Dot Product

- dot product: If

  $$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

  then the dot product of $\mathbf{u} \cdot \mathbf{v}$ of $\mathbf{u}$ and $\mathbf{v}$ is defined by

  $$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

- properties of dot product: let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then:
  - $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
  - $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$

- $(c\mathbf{u})\cdot\mathbf{v} = c(\mathbf{u}\cdot\mathbf{v})$
- $\mathbf{u}\cdot\mathbf{u} \geq \mathbf{0}$ and $\mathbf{u}\cdot\mathbf{u} = 0$ IFF $\mathbf{u} = \mathbf{0}$
- Length or norm of a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$ is the nonnegative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

- Normalizing a vector means finding the unit vector.
- Cauchy-Schwarz Inequality: For all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

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

- Triangle inequality: for all vectors $\mathbf{u}$ and $\mathbf{v}$ and $\mathbb{R}^n$,

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

- Distance between two vectors is defined by

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

- Two vectors are orthogonal if $\mathbf{u}\cdot\mathbf{v} = 0$
- For all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ IFF $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.
- If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$ and $\mathbf{u} \neq \mathbf{0}$, then the projection of $\mathbf{v}$ onto $\mathbf{u}$ is the vector defined by

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}$$

1.3 Lines and Planes

- Normal form of the equation of a 2D line:

$$\mathbf{n}\cdot(\mathbf{x}-\mathbf{p}) = \mathbf{0} \qquad \text{or} \qquad \mathbf{n}\cdot\mathbf{x} = \mathbf{n}\cdot\mathbf{p}$$

  where $\mathbf{p}$ is a specific point on the line and $\mathbf{n} \neq \mathbf{0}$ is a normal vector for the line.

- The general form of the equation of the line is $ax + by = c$ where $\mathbf{n} = \begin{bmatrix} a \\ b \end{bmatrix}$ is a normal vector for the line.
- The vector form of the equation of a 2D or 3D line is

$$\mathbf{x} = \mathbf{p} + t\mathbf{d}$$

  where $\mathbf{p}$ is a specific point on the line and $\mathbf{d} \neq \mathbf{0}$ is a direction vector for the line. The equations corresponding to the components of the vector form of the equations are called parametric equations of the line.

- Normal form of the equation of a plane $\mathscr{P}$ in $\mathbb{R}^3$ is

$$\mathbf{n}\cdot(\mathbf{x}-\mathbf{p}) = 0 \qquad \text{or} \qquad \mathbf{n}\cdot\mathbf{x} = \mathbf{n}\cdot\mathbf{p}$$

  where $\mathbf{p}$ is a specific point on $\mathscr{P}$ and $\mathbf{n} \neq \mathbf{0}$ is a normal vector for $\mathscr{P}$.

- The general form of the equation of $\mathscr{P}$ is $ax + by + cz = d$, where $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is a normal vector for $\mathscr{P}$.

- The vector form of the equation of a plane $\mathscr{P}$ in $\mathbb{R}^3$ is

$$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$$

  where $\mathbf{p}$ is a point on $\mathscr{P}$ and $\mathbf{u}$ and $\mathbf{v}$ are direction vectors for $\mathscr{P}$ ($\mathbf{u}$ and $\mathbf{v}$ are nonzero and parallel to $\mathscr{P}$, but not parallel to each other). The equations corresponding to the components of the vector form of the equation are called parametric equations of $\mathscr{P}$.

- Summary of equations of 2D lines:
  - Normal form: $\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$
  - General form: $ax + by = c$
  - Vector form: $\mathbf{x} = \mathbf{p} + t\mathbf{d}$
  - Parametric form:

$$\begin{cases} x = p_1 + td_1 \\ y = p_2 + td_2 \end{cases}$$

- Summary of equations of 3D lines:
  - Normal form:

$$\begin{cases} \mathbf{n_1} \cdot \mathbf{x} = \mathbf{n_1} \cdot \mathbf{p_1} \\ \mathbf{n_2} \cdot \mathbf{x} = \mathbf{n_2} \cdot \mathbf{p_2} \end{cases}$$

  - General form:

$$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \end{cases}$$

  - Vector form: $\mathbf{x} = \mathbf{p} + t\mathbf{d}$
  - Parametric form:

$$\begin{cases} x = p_1 + td_1 \\ y = p_2 + td_2 \\ z = p_3 + td_3 \end{cases}$$

- Summary of equations of 3D planes:
  - Normal form: $\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$
  - General form: $ax + by + cz = d$
  - Vector form: $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$
  - Parametric form:

$$\begin{cases} x = p_1 + su_1 + tv_1 \\ y = p_2 + su_2 + tv_2 \\ z = p_3 + su_3 + tv_3 \end{cases}$$

1.4 Applications

- Force vectors: if the resultant net force is zero, the system is in equilibrium.
- Resolve into components to work with the vectors.

2 Systems of Linear Equations

2.1 Introduction to Systems of Linear Equations

- A linear equation in the n variables $x_1, x_2, x_3, \cdots, x_n$ is an equation that can be written in the form

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

  where the coefficients $a_1, a_2, \cdots, a_n$ and the constant term b are constants.

- A solution of a linear equation $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ is a vector $[s_1, s_2, \cdots, s_n]$ whose components satisfy the equation when we substitute $x_1 = s_1, x_2 = s_2, \cdots, x_n = s_n$.
- A system of linear equations is a finite set of linear equations, each with the same variables. A solution of a system of linear equations is a vector that is simultaneously a solution of each equation in the system. The solution set of a system of linear equations is the set of all solutions of the system.
- A system of linear equations is called consistent if it has at least one solution. A system with no solutions is inconsistent.
- Two linear systems are called equivalent if they have the same solution sets.
- Solving a matrix with a CAS may not always be the best choice.

2.2 Direct Methods for Solving Linear Systems

- The coefficient matrix contains the coefficients of the variables, and the augmented matrix is the coefficient matrix augmented by an extra column containing the constant terms.
- A matrix is in row echelon form if it satisfies the following properties:
  - Any rows consisting entirely of zeroes are at the bottom
  - In each nonzero row, the first nonzero entry (called the leading entry) is in a column to the left of any leading entries below it.
- Elementary row operations:
  - Interchange two rows
  - Multiply a row by a nonzero constant
  - Add a multiple of a row to another row
- Matrices A and B are row equivalent if there is a sequence of elementary row operations that converts A into B.
- Matrices A and B are row equivalent IFF they can be reduced to the same row echelon form.
- Gaussian elimination:
  - Write the augmented matrix of the system of linear equations.
  - Use elementary row operations to reduce the augmented matrix to row echelon form.
  - Using back substitution, solve the equivalent system that corresponds to the row-reduced matrix.
- The rank of a matrix is the number of nonzero rows in its row echelon form.
- The rank theorem: let A be the coefficient matrix of a system of linear equations with n variables. If the system is consistent, then
$$\text{number of free variables} = n - \text{rank}(A)$$
- A matrix is in reduced row echelon form if it satisfies the following:
  - It is in row echelon form.
  - The leading entry in each nonzero row is a 1 (called a leading 1)
  - Each column containing a leading 1 haas zeroes everywhere else.
- Steps for Gauss-Jordan Elimination:
  - Write the augmented matrix of the system of linear equations.
  - Use elementary row operations to reduce the augmented matrix to RREF
  - If the resulting system is consistent, solve for the leading variables in terms of any remaining free variables.
- A system of linear equations is called homogeneous if the constant term in each equation is zero.
- Theorem: If $[A|\mathbf{0}]$ is a homogeneous system of m linear equations with n variables, where $m < n$, then the system has infinitely many solutions.

2.3 Spanning Sets and Linear Independence

- Theorem: A system of linear equations with the augmented matrix $[A|\mathbf{b}]$ is consistent IFF $\mathbf{b}$ is a linear combination of the columns of A.

- Definition: If $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is a set of vectors in $\mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ is called the span of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ and is denoted by $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$ or $\text{span}(S)$. If $\text{span}(S) = \mathbb{R}^n$, then S is called a spanning set for $\mathbb{R}^n$.

- Definition: A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ is linearly dependent if there are scalars $c_1, c_2, \cdots, c_k$, *at least one of which is not zero*, such that
$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$
A set of vectors that is not linearly dependent is called linearly independent.

- Theorem: Vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ in $\mathbb{R}^n$ are linearly dependent IFF at least one of the vectors can be expressed as a linear combination of the others.

- Theorem: Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ be (column) vectors in $\mathbb{R}^n$ and let A be the $n \times m$ matrix $[\, v_1\mathbf{v}_2 \,\cdots\, \mathbf{v}_m]$ with these vectors as its columns. Then $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ are linearly dependent IFF the homogeneous linear system with augmented matrix $[A|\mathbf{o}]$ has a nontrivial solution.

- Theorem: Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ be (row) vectors in $\mathbb{R}^n$ and let $m \times n$ matrix $\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_m \end{bmatrix}$ with these vectors is its rows. Then $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ are linearly dependent IFF $\text{rank}(A) < m$.

- Theorem: Any set of m vectors in $\mathbb{R}^n$ is linearly dependent if $m > n$.

2.4 Applications

- Applications include:
  - Allocation of resources
  - Balancing chemical equations
  - Network analysis in transportation, economics, electricity and magnetism

2.5 Iterative Methods for Solving Linear Systems

- Two iterative methods: Jacobi's method and Gauss-Seidel method

- Theorem: If a set of n linear equations in n variables has a strictly diagonally dominant coefficient matrix, then it has a unique solution and both the Jacobi and Gauss-Seidel method converge to it.

- Theorem: If the Jacobi or the Gauss-Seidel method converges for a system of n linear equations in n varibales, then it must converge to the solution of the system.

3 Matrices

3.1 Matrix Operations

- A matrix is defined as a rectangular array of numbers called the entries, or elements, of the matrix

- The size of a matrix is based on the number of rows and columns; a matrix with m rows and n columns is an $m \times n$ matrix (m by n).

- Entries of matrix are refered to with double subscripts

- If $m = n$, the matrix is a square matrix. If all nondiagonal entries are 0, the matrix is a diagonal matrix. A diagonal matrix whose diagonal entries are the same is called a scalar matrix. If the scalar on the diagonal is 1 it is an identity matrix.
- Adding matrices: only matrices with the same dimensions can be added. Add each corresponding entry.
- A matrix whose entries are all 0 is a zero matrix denoted by O.
- Multiplying matrices: If A is an $m \times n$ matrix and B is an $n \times r$ matrix, then the product $C = AB$ is an $m \times r$ matrix. The (i, j) entry of the product is computed as follows:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

- Note that if A is an m by n matrix and B is an n by r matrix, AB is an m by r matrix and n must be equal to n.
- The product of two matrices is a dot product.
- Theorem: let A be an $m \times n$ matrix, $\mathbf{e}_i$ a $1 \times m$ standard unit vector, and $\mathbf{e}_j$ an $n \times 1$ standard unit vector. Then:
  - $\mathbf{e}_i A$ is the ith row of A and
  - $A\mathbf{e}_j$ is the jth column of A.
- Partitioned matrices: you can divide a matrix into submatrices by partitioning it into blocks.
- Matrx powers:

$$A^k = AA \cdots A$$

  by k factors
- If A is a square matrix and r and s are nonnegative integers, then

$$A^r A^s = A^{r+s} \qquad (A^r)^s = A^{rs}$$

- The transpose of an $m \times n$ matrix A is the $n \times m$ matrix of A^T obtained by interchanging the rows and columns of A. That is, the ith column of A^T is the ith row of A for all i
- A square matrix A is defined as symmetric if $A^T = A$; that is, if A is equal to its own transpose.
- A square matrix A is symmetric IFF $A_{ij} = A_{ji}$ for all i and j

3.2 Matrix Algebra

- Properties of matrix addition and scalar multiplication
  - $A + B = B + A$
  - $(A + B) + C = A + (B + C)$
  - $A + O = A$
  - $A + (-A) = O$
  - $c(A + B) = cA + cB$
  - $(c + d)A = cA + dA$
  - $c(dA) = (cd)A$
  - $1A = A$
- Linear independence applies to matrices as well
- Properties of matrix multiplication
  - $A(BC) = (AB)C$

  - $A(B+C) = AB + AC$
  - $(A+B)C = AC + BC$
  - $k(AB) = (kA)B = A(kB)$
  - $I_m A = A = AI_n$ if A is $m \times n$

- Properties of the transpose
  - $(A^T)^T = A$
  - $(A+B)^T = A^T + B^T$
  - $(kA)^T = k(A^T)$
  - $(AB)^T = B^T A^T$
  - $(A^r)^T = (A^T)^r$ for all nonnegative integers r

- Transposing a matrix: like flipping it on its side; rows become columns and columns become rows. Order stays the same; left to right, top to bottom

- If A is a square matrix, then $A + A^T$ is a symmetric matrix

- For any matrix A, AA^T and $A^T A$ are symmetric matrices.

3.3 The Inverse of a Matrix

- If A is an $n \times n$ matrix, an inverse of A is an $n \times n$ matrix A' with the property that

$$AA' = I \quad \text{and} \quad A'A = I$$

  where $I = I_n$, the $n \times n$ identity matrix. If A' exists, then A is called invertible.

- If A is an invertible matrix, then its inverse is unique.

- If A is an invertible $n \times n$ matrix, then the system of linear equations given by $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ for any $\mathbf{b}$ in $\mathbb{R}^n$

- If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then A is invertible if $ad - bc \neq 0$, in which case

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

  If $ad - bc = 0$, then A is not invertible.

- The expression $ad - bc$ is the determinant of A, given by $\det A$

- Properties of invertible matrices:
  - If A is an invertible matrix, then A^{-1} is invertible and

$$(A^{-1})^{-1} = A$$

  - If A is an invertible matrix and c is a nonzero scalar, then cA is an invertible matrix and

$$(cA)^{-1} = \frac{1}{c} A^{-1}$$

  - If A and B are invertible matrices of the same size, then AB is invertible, and

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

  - If A is an invertible matrix, then A^T is invertible and

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

- If A is an invertible matrix, then A^n is invertible for all nonnegative integers n and

$$(A^n)^{-1} = (A^{-1})^n$$

- The inverse of a product of invertible matrices is the product of their inverses in reverse order.
- If A is an invertible matrix and n is a positive integer, then A^{-n} is defined by

$$A^{-n} = (A^{-1})^n = (A^n)^{-1}$$

- An elementary matrix is one that can be obtained by performing an elementary row operation on an identity matrix.
- Let E by the elementary matrix obtained by performing an elementary row operation on I_n. If the same elementary row operation is performed on an $n \times r$ matrix A, the result is the same as the matrix EA.
- Each elementary matrix is invertible, and its inverse is an elementary matrix of the same type.
- The fundamental theorem of invertible matrices: version 1. Let A be an $n \times n$ matrix. The following statements are equivalent:
  - A is invertible.
  - $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$
  - $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
  - The RREF of A is I_n
  - A is a product of elementary matrices.
- Let A be a square matrix. If B is a square matrix such that either $AB = I$ or $BA = 1$, then A is invertible and $B = A^{-1}$
- Let A be a square matrix. If a sequence of elementary row operations reduces A to I, then the same sequence of elementary row operations transforms I into A^{-1}

3.4 The LU Factorization

- Let A be a square matrix. A factorization of A as $A = LU$, where L is unit lower triangular and U is upper triangular, is called an LU factorization of A.
- If A is a square matrix that can be reduced to REF without using any row interchanges, then A has an LU factorization.
- If A is an invertible matrix that has an LU factorization, then L and U are unique.
- If P is a permutation matrix, then $P^{-1} = P^T$
- Let A be a square matrix. A factorization of A as $A = P^T LU$, where P is a permutation matrix, L is unit lower triangular, and U is upper triangular, is called a $P^T LU$ factorization of A
- Every square matrix has a $P^T LU$ factorization.

3.5 Subspaces, Basis, Dimension, and Rank

- A subspace of $\mathbb{R}^n$ is any collection of S vectors in $\mathbb{R}^n$ such that
  - The zero vector $\mathbf{0}$ is in S
  - If $\mathbf{u}$ and $\mathbf{v}$ are in S, then $\mathbf{u} + \mathbf{v}$ is in S (That is, S is closed under addition).
  - If vbu is in S and c is a scalar, then $c\mathbf{u}$ is in S (S is closed under scalar multiplication).
  - From the previous two conditions, we conclude that S must then be closed under linear combinations: S includes all linear combinations of all vectors $\mathbf{u}_k$ in S.

- Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ be vectors in $\mathbb{R}^n$. Then, $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$ is a subspace of $\mathbb{R}^n$
- Let A be an $m \times n$ matrix.
  - The row space of A is the subspace $\text{row}(A)$ of $\mathbb{R}^n$ spanned by the rows of A.
  - The column space of A is the subspace $\text{col}(A)$ of $\mathbb{R}^m$ spanned by the columns of A.
- Let B be any matrix that is row equivalent to a matrix A. Then $\text{row}(B) = \text{row}(A)$
- Let A be an $m \times n$ matrix and let N be the set of solutions to the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. Then N is a subspace of $\mathbb{R}^n$
- Let A be an $m \times n$ matrix. The null space of A is the subspace of $\mathbb{R}^n$ consisting of solutions to the homogeneous linear system $A\mathbf{x} = 0$. It is denoted by $\text{null}(A)$
- Let A be a matrix whose entries are real numbers. For any system of linear equations $A\mathbf{x} = \mathbf{b}$, exactly one of the following is true:
  - There is no solution.
  - There is a unique solution.
  - There are infinitely many solutions.
- A basis for a subspace S of $\mathbb{R}^n$ is a set of vectors in S that
  - spans S and
  - Is linearly independent.
- How to find the bases for the row space, column space, and null space of matrix A:
  - Find the rref R of A.
  - Use the nonzero row vectors of R (containing the leading 1s) to form a basis for $\text{row}(A)$
  - Use the column vectors of A that correspond to the columns of R containing the leading 1s (the pivot columns) to form a basis for $\text{col}(A)$
  - Solve for the leading variables of $R\mathbf{x} = 0$ in terms of the free variables, set the free variables equal to parameters, substitute back into $\mathbf{x}$, and write the result as a linear combination of f vectors (where f is the number of free variables). These f vectors form a basis for $\text{null}(A)$
- The Basis Theorem: Let S be a subspace of $\mathbb{R}^n$. Then any two bases for S have the same number of vectors.
- If S is a subspace of $\mathbb{R}^n$, then the number of vectors in a basis for S is called the dimension of S, denoted dim S.
- The row and column spaces of a matrix A have the same dimension.
- The rank of a matrix A is the dimension of its row and column spaces and is denoted by $\text{rank}(A)$.
- For any matrix A,
$$\text{rank}(A^T) = \text{rank}(A)$$
- The nullity of a matrix A is the dimension of its null space and is denoted by $\text{nullity}(A)$.
- The Rank Theorem: If A is an $m \times n$ matrix, then
$$\text{rank}(A) + \text{nullity}(A) = n$$
- The fundamental theorem of invertible matrices: version 2. Let A be an $m \times n$ matrix. The following statements are equivalent:
  - A is invertible.
  - $A\mathbf{x} = b$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$
  - $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

  - The rref of A is I_n
  - A is a product of elementary matrices.
  - $\text{rank}(A) = n$
  - $\text{nullity}(A) = 0$
  - The column vectors of A are linearly independent
  - The column vectors of A span $\mathbb{R}^n$.
  - The column vectors of A form a basis for $\mathbb{R}^n$.
  - The row vectors of A are linearly independent.
  - The row vectors of A span $\mathbb{R}^n$
  - The row vectors of A form a basis for $\mathbb{R}^n$.

- Let A be an $m \times n$ matrix. Then
  - $\text{rank}(A^T A) = \text{rank}(A)$
  - The $n \times n$ matrix $A^T A$ is invertible IFF $\text{rank}(A) = n$

- Let S be a subspace of $\mathbb{R}^n$ and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be a basis for S. For every vector $\mathbf{v}$ in S, there is exactly one way to write $\mathbf{v}$ as a linear combination of the basis vectors in $\mathcal{B}$:
$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

- Let S be a subspace of $\mathbb{R}^n$ and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be a basis for S. Let $\mathbf{v}$ be a vector in S, and write $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$. Then $c_1, c_2, \cdots, c_k$ are called the coordinates of $\mathbf{v}$ with respect to $\mathcal{B}$, and the column vector
$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$
is called the coordinate vector of $\mathbf{v}$ with respect to $\mathcal{B}$.

3.6 Introduction to Linear Transformations

- A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation if:
  - $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all u and v in $\mathbb{R}^n$ and
  - $T(c\mathbf{v}) = cT(\mathbf{v})$ for all $\mathbf{v}$ in $\mathbb{R}^n$ and all scalars c.

- Let A be an $m \times n$ matrix. Then the matrix transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by
$$T_A(\mathbf{x}) = A\mathbf{x} \quad (\text{for } \mathbf{x} \text{ in } \mathbb{R}^n)$$
is a linear transformation.

- Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is a matrix transformation. More specifically, $T = T_A$, where A is the $m \times n$ matrix
$$A = [T(\mathbf{e}_1) \vdots T(\mathbf{e}_2) \vdots \cdots \vdots T(\mathbf{e}_n)]$$

- Let $T : \mathbb{R}^m \to \mathbb{R}^n$ and $S : \mathbb{R}^n \to \mathbb{R}^p$ be linear Transformations. then $S \circ T : \mathbb{R}^m \to \mathbb{R}^p$ is a linear transformation whose standard matrices are related by
$$S \circ T = [S][T]$$

- Let s and T be linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$. Then S and T are inverse transformations if $S \circ T = I_n$ and $T \circ S = I_n$

- Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. Then its standard matrix $[T]$ is an invertible matrix, and
$$[T^{-1}] = [T]^{-1}$$
(“The matrix of the inverse is the inverse of the matrix”).

3.7 Applications

- Markov chain: evolving proccess consisting of a finite number of states.
- Can use linear algebra to analyze probabilities. Consider two-way tables in statistics: working with multiple of these tables as matrices and vectors can allow us to solve probability problems.
- Graphs and digraphs: If G is a graph with n vertices, then its adjacency matrix is the $n \times n$ matrix A (or $A(G)$) defined by

$$a_{ij} = 1 \text{ if there is an edge between vertices i and j, and } a_{ij} = 0 \text{ otherwise.}$$

- If A is the adjacency matrix of a graph G, then the (i, j) entry of A^k is equal to the number of k-paths between vertices i and j.
- If G is a digraph with n vertices, then its adjacency matrix is the $n \times n$ matrix A (or $A(G)$) defined by

$$a_{ij} = 1 \text{ if there is an edge between vertices i and j, and } a_{ij} = 0 \text{ otherwise.}$$

- Error correcting codes: If $k < n$, then any $n \times k$ matrix of the form $G = \begin{bmatrix} I_k \\ A \end{bmatrix}$, where A is an $(n-k) \times k$ matrix over $\mathbb{Z}_2$, is called a standard generator matrix for an (n, k) binary code $T : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$. Any $(n-k) \times n$ matrix of the form $P = [B \ I_{n-k}]$, where B is an $(n-k) \times k$ matrix over $\mathbb{Z}_2$, is called a stardard parity check matrix. The code is said to have length n and dimension k.
- If $G = \begin{bmatrix} I_k \\ A \end{bmatrix}$ is a standard generator matrix and $P = [B \ I_{n-k}]$ is a standard parity check matrix, then P is the parity check matrix associated with G IFF $A = B$. The corresponding n, k binary code is (single) error-correcting IFF the columns of P are nonzero and distinct.
- Summary of error-correcting codes:
  - For $n > k$, and $n \times k$ matrix G and an $(n-k) \times n$ matrix P (with entries in $mathbbZ_2$) are a standard generator matrix and a standard parity check matrix, respectively, for an (n, k) binary code IFF in block form, $G = \begin{bmatrix} I_k \\ A \end{bmatrix}$ and $P = [A \ I_{n-k}]$ for some $(n-k) \times k$ matrix A over $\mathbb{Z}_2$.
  - G encodes a message vector $\mathbf{x}$ in $\mathbb{Z}_2^k$ as a code vector $\mathbf{c}$ in $\mathbb{Z}_2^n$ via $\mathbf{c} = G\mathbf{x}$.
  - G is error-correcting IFF the columns of P are nonzero and distinct. A vector $\mathbf{c}'$ in $\mathbb{Z}_2^n$ is a code vector IFF $P\mathbf{c}' = \mathbf{0}$. In this case, the corresponding message vector is the vector $\mathbf{x}$ in $\mathbb{Z}_2^k$ consisting of the first k components of $\mathbf{c}'$. If $P\mathbf{c}' \neq 0$, then $\mathbf{c}'$ is not a code vector and $P\mathbf{c}'$ is one of the columns of P. If $P\mathbf{c}'$ is the ith column of P, then the error is in the ith component of $\mathbf{c}'$ and we can recover the correct code vector (and hence the message) by changing this component.

4 Eigenvalues and Eigenvectors

4.1 Introduction to Eigenvalues and Eigenvectors

- Let A be an $n \times n$ matrix. A scalar λ is called an eigenvalue of A if there is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Such a vector $\mathbf{x}$ is called an eigenvector of A corresponding to λ.
- Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. The collection of all eigenvectors corresponding to λ, together with the zero vector, is called the eigenspace of λ and is denoted by E_λ.

4.2 Determinants

- Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. Then the determinant of A is the scalar

$$\det A = |A| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

- We can simplify this equation as:

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}$$
$$= \sum_{j=1}^{3} (-1)^{1+j} a_{ij} \det A_{ij}$$

- For any square matrix A, $\det A_{ij}$ is called the (i,j)-minor of A.
- Let $A = [a_{ij}]$ be an $n \times n$ matrix, where $n \geq 2$. Then the determinant of A is the scalar

$$\det A = |A| = a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n} a_{1n} \det A_{1n}$$
$$= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j}$$

- The (i,j)-cofactor of A is defined as
$$C_{ij} = (-1)^{i+j} \det A_{ij}$$
- Thus, the definition of the determinant becomes

$$\det A = \sum_{j=1}^{n} a_{1j} C_{1j}$$

- The Laplace Expansion Theorem: The determinant of an $n \times n$ matrix $A = [a_{ij}]$, where $n \geq 2$, can be computed as
$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$
$$\sum_{j=1}^{n} a_{ij}C_{ij}$$
(which is the cofactor expansion along the ith row) and also as
$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$
$$= \sum_{i=1}^{n} a_{ij}C_{ij}$$
(the cofactor expansion along the jth column).
- The determiniant of a triangular matrix is the product of the entries on its main diagonal. Specifically, if $A = [a_{ij}]$ is an $n \times n$ triangular matrix, then

$$\det A = a_{11}a_{22} \cdots a_{nn}$$

- Properties of determinants: let A be a square matrix.
  - If A has a zero row (column), then $\det A = 0$
  - If B is obtained by interchanging two rows (columns) of A, then $\det B = -\det A$
  - If A has two identical rows (columns), then $\det A = 0$
  - If B is obtained by multiplying a row (column) of A by k, then $\det B = k \det A$
  - If A, B, and C are identical except that the ith row (column) of C is the sum of the ith rows (columns) of A and B, then $\det C = \det A + \det B$
  - If B is obtained by adding a multiple of one row (column) of A to another row (column), then $\det B = \det A$
- Let E be an $n \times n$ elementary matrix.

  - If E results from interchanging two rows of I_n, then $\det E = -1$
  - If E results from multiplying one row of I_n by k, then $\det E = k$
  - If E results from adding a multiple of one row of I_n to another row, then $\det E = 1$

- Let B be an $n \times n$ matrix and let E be an $n \times n$ elementary matrix. Then

$$\det(EB) = (\det E)(\det B)$$

- A square matrix A is invertible IFF $\det A \neq 0$
- If A is an $n \times n$ matrix, then

$$\det(kA) = k^n \det A$$

- If A and B are $n \times n$ matrices, then

$$\det(AB) = (\det A)(\det B)$$

- If A is invertible, then

$$\det\left(A^{-1}\right) = \frac{1}{\det A}$$

- For any square matrix A,

$$\det A = \det A^T$$

- Cramer's rule: let A be an invertible $n \times n$ matrix and let $\mathbf{b}$ be a vector in $\mathbb{R}^n$. Then the unique solution $\mathbf{x}$ of the system $A\mathbf{x} = \mathbf{b}$ is given by

$$x_i = \frac{\det(A_i(\mathbf{b}))}{\det A} \qquad \text{for } i = 1, \cdots, n$$

- Let A be an invertible $n \times n$ matrix. Then

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

  where the adjoint of A adj A is defined by

$$[C_{ji}] = [C_{ij}]^T = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

- Let A be an $n \times n$ matrix. Then

$$a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} = \det A = a_{11}C_{11} + A_{21}C_{21} + \cdots + a_{n1}C_{n1}$$

- Let A be an $n \times n$ matrix and let B be obtained by interchanging any two rows (columns) of A. Then

$$\det B = -\det A$$

4.3 Eigenvalues and Eigenvectors of $n \times n$ Matrices

- The eigenvalues of a square matrix A are preciesely the solutions λ of the equation

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

  .

- Finding the eigenvalues and eigenvectors of a matrix: Let A be an $n \times n$ matrix.
  - Compute the characteristic polynomial $\det(A - \lambda I)$ of A.
  - Find the eigenvalues of A by solving the characteristic equation $\det(A - \lambda I) = 0$ for λ

  - For each eigenvalue λ, find the null space of the matrix $A - \lambda I$. This is the eigenspace E_λ, the nonzero vectors of which are the eigenvectors of A corresponding to λ.
  - Find a basis for each eigenspace.

- The algebraic multiplicity of an eigenvalue is its multiplicity as a root of the characteristic equation. An $n \times n$ matrix will always have n eigenvalues, but some will be duplicates due to algebraic multiplicity.
- The eigenvalues of a triangular matrix are the entries on its main diagonal.
- A square matrix A is invertible IFF 0 is not an eigenvalue of A.
- The fundamental theorem of invertible matrices: version 3. Let A be an $n \times n$ matrix. The following statements are equivalent:
  - A is invertible
  - $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$
  - $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
  - The reduced row echelon form of A is I_n.
  - A is the product of elementary matrices.
  - $\text{rank}(A) = n$
  - $\text{nullity}(A) = 0$
  - The column vectors of A are linearly independent
  - The column vectors of A span $\mathbb{R}^n$
  - The column vectors of A form a basis for $\mathbb{R}^n$
  - The row vectors of A are linearly independent
  - The row vectors of A span $\mathbb{R}^n$
  - The row vectors of A form a basis for $\mathbb{R}^n$
  - $\det A \neq 0$
  - 0 is not an eigenvalue of A
- Let A be a square matrix with eigenvalue λ and corresponding eigenvector $\mathbf{x}$.
  - For any positive integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\mathbf{x}$.
  - If A is invertible, then $1/\lambda$ is an eigenvalue of A^{-1} with corresponding eigenvector $\mathbf{x}$.
  - If A is invertible, then for any integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\mathbf{x}$.
- Suppose the $n \times n$ matrix A has eigenvectors $\mathbf{v}_1, \mathbf{v}_2 \cdots, \mathbf{v}_m$ with corresponding eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_m$. If $\mathbf{x}$ is a vector in $\mathbb{R}^n$ that can be expressed as a linear combination of these eigenvectors, then for any integer k,

$$A^k\mathbf{x} = c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_m\lambda_m^k\mathbf{v}_m$$

- Let A be an $n \times n$ matrix and let $\lambda_1, \lambda_2, \cdots, \lambda_m$ be distinct eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$. Then $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ are linearly independent.

4.4 Similarity and Diagonalization

- Let A and B be $n \times n$ matrices. We say that A is similar to B if there is an invertible $n \times n$ matrix P such that $P^{-1}AP = B$. If A is similar to B, we write $A \sim B$
- Let A, B, C be $n \times n$ matrices.
  - $A \sim A$
  - If $A \sim B$ then $B \sim A$
  - If $A \sim B$ and $B \sim C$ then $A \sim C$

- Let A and B be $n \times n$ matrices with $A \sim B$. Then
  - $\det A = \det B$
  - A is invertible IFF B is invertible.
  - A and B have the same rank
  - A and B have the same characteristic polynomial.
  - A and B have the same eigenvalues.
  - $A^m \sim B^m$ for all integers $m \geq 0$
  - If A is invertible, then $A^m \sim B^m$ for all integers m.
- An $n \times n$ matrix A is diagonalizable if there is a diagonal matrix D such that A is similar to D - that is, if there is an invertible $n \times n$ matrix P such that $P^{-1}AP = D$
- Let A be an $n \times n$ matrix. Then A is diagonalizable IFF A has n linearly independent eigenvectors. More precisely, there exist an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$ IFF the columns of P are n linearly independent eigenvectors of A and the diagonal entries of D are the eigenvalues of A corresponding to the eigenvectors in P in the same order.
- Let A be an $n \times n$ matrix and let $\lambda_1, \lambda_2, \cdots, \lambda_k$ be distinct eigenvalues of A. If $\mathcal{B}_i$ is a basis for the eigenspace E_λ, then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ (i.e., the total collection of basis vectors for all the eigenspaces) is linearly independent.
- If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable
- If A is an $n \times n$ matrix, then the geometric multiplicity of each eigenvalue is less than or equal to its algebraic multiplicity.
- The diagonalization theorem: Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \lambda_2, \cdots, \lambda_k$. The following statements are equivalent.
  - A is diagonalizable.
  - The union $\mathcal{B}$ of the bases of the eigenspaces of A (as in theorem 4.24) contains n vectors.
  - The algebraic multiplicity of each eigenvalue equals its geometric multiplicity.

4.5 Iterative Methods for Computing Eigenvalues

- Let A be an $n \times n$ diagonalizable matrix with dominant eigenvalue λ_1. Then there exists a nonzero vector $\mathbf{x}_0$ such that the sequence of vectors $\mathbf{x}_k$ defined by

$$\mathbf{x}_1 = A\mathbf{x}_0, \mathbf{x}_2 = A\mathbf{x}_1, \mathbf{x}_3 = A\mathbf{x}_2, \cdots, \mathbf{x}_k = A\mathbf{x}_{k-1}, \cdots$$

  approaches a dominant eigenvector of A
- Summarization of the power method: Let A be a diagonalizable $n \times n$ matrix with a corresponding dominant eigenvalue λ_1
  - Let $\mathbf{x}_0 = \mathbf{y}_0$ be any initial vector in $\mathbb{R}^n$ whose largest component is 1.
  - Repeat the following steps for $k = 1, 2, \cdots$:
    * Compute $\mathbf{x}_k = A\mathbf{y}_{k-1}$
    * Let m_k be the component of $\mathbf{x}_k$ with the largest absolute value.
    * Set $\mathbf{y}_k = (1/m_k)\mathbf{x}_k$
- For most choices of $\mathbf{x}_0$, m_k converges to the dominant eigenvalue λ_1 and $\mathbf{y}_k$ converges to a dominant eigenvector.
- Let $A = [a_{ij}]$ be a (real or complex) $n \times n$ matrix, and let r_i denote the sum of the absolute values of the off-diagonal entries in the ith row of A; that is, $r_i = \sum_{j \neq i} |a_{ij}|$. The ith Gerschgorin disk is the circular disk D_i in the complex plane with center a_{ii} and radius r_i. That is,

$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i\}$$

- Gerschgorin's Disk Theorem: Let A be an $n \times n$ (real or complex) matrix. Then every eigenvalue of A is contained within a Gerschgorin disk.

4.6 Applications and the Perron-Frobenius Theorem

- If P is the $n \times n$ transition matrix of a Markov chain, then 1 is an eigenvalue of P.
- Let P be an $n \times n$ transition matrix with eigenvalue λ.
  - $|\lambda| \leq 1$
  - If P is regular and $\lambda \neq 1$, then $|\lambda| < 1$
- Let P be a regular $n \times n$ transition matrix. If P is diagonalizable, then the dominant eigenvalue $\lambda_1 = 1$ has algebraic multiplicity 1
- Let P be a regular $n \times n$ transition matrix. Then as $k \to \infty$, p^k approcahes an $n \times n$ matrix L whose columns are identical, each equal to the same vector $\mathbf{x}$. This vector $\mathbf{x}$ is a steady state probability vector for P.
- Let P be a regular $n \times n$ transition matrix, with $\mathbf{x}$ the steady state probability vector for P, as in the above. Then, for any initial probability vector $\mathbf{x}_0$, the sequence of iterates $\mathbf{x}_k$ approaches $\mathbf{x}$.
- Every Leslie matrix has a unique positive eigenvalue and a corresponding eigenvector with positive components.
- Perron's Theroem: Let A be a positive $n \times n$ matrix. Then A has a real eigenvalue λ_1 with the following properties:
  - $\lambda_1 > 0$
  - λ_1 has a corresponding positive eigenvector.
  - If λ is any other eigenvalue of A, then $|\lambda| \leq \lambda_1$
- The Perron-Frobenius Theorem: Let A be an irreducible nonnegative $n \times n$ matrix. Then A has a real eigenvalue λ_1 with the following properties:
  - $\lambda_1 > 0$
  - λ_1 has a corresponding positive eigenvector.
  - If λ is any other eigenvalue of A, then $|\lambda| \leq \lambda_1$. If A is primitive, then this inequality is strict.
  - If λ is an eigenvalue of A such that $|\lambda| = \lambda_1$, then λ is a (complex) root of the equation $\lambda^n - \lambda_1^n = 0$
  - λ_1 has algebraic multiplicity 1.
- Def: Let $(x_n) = (x_0, x_1, x_2)$ be a sequence of numbers that is defined as follows:
  - $x_0 = a_0, x_1 = a_1, \cdots, x_{k-1} = a_{k-1}$, where $a_0, a_1, \cdots, a_{k-1}$ are scalars.
  - For all $n \geq k, x_n = c_1 x_{n-1} + c_2 x_{n-2} + \cdots + c_k x_{n-k}$, where $c_1, c_2, \cdots, c_k$ are scalars.
- If $c_k \neq 0$, the equation in the second line is called a linear recurrence relation of order k. The equations in the first line are referred to as the initial conditions of the recurrence.
- Let $x_n = ax_{n-1} + bx_{n-2}$ be a recurrence relation that is satisfied by a sequence (x_n). Let λ_1 and λ_2 be the eigenvalues of the associated characteristic equation $\lambda^2 - a\lambda - b = 0$.
  - If $\lambda_1 \neq \lambda_2$, then $x_n = c_1\lambda_1^n + c_2\lambda_2^n$ for some scalars c_1 and c_2.
  - If $\lambda_1 = \lambda_2 = \lambda$, then $x_n = c_1\lambda^n + c_2 n\lambda^n$ for some scalars c_1 and c_2.
- In either case, c_1 and c_2 can be determined using the initial conditions.
- Let $x_n = a_{m-1}x_{n-1} + a_{m-2}x_{n-2} + \cdots + a_0 x_{n-m}$ be a recurrence relation of order m that is satisfied by a sequence (x_n). Suppose the associated characteristic polynomial

$$\lambda^m - a_{m-1}\lambda^{m-1} - a_{m-2}\lambda^{m-2} - \cdots - a_0$$

factors as $(\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_k)^{m_k}$, where $m_1 + m_2 + \cdots + m_k = m$. Then x_n has the form

$$x_n = (c_{11}\lambda_1^n + c_{12}n\lambda_1^n + c_{13}n^2\lambda_1^n + \cdots + c_{1m_1}n^{m_1-1}\lambda_1^n) + \cdots$$
$$+(c_{k1}\lambda_k^n + c_{k2}n\lambda_k^n + c_{k3}n^2\lambda_k^n + \cdots + c_{km_k}n^{m_k-1}\lambda_k^n)$$

- Let A be an $n \times n$ diagonalizable matrix and let $P = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$ be such that
$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$
Then the general solution to the system $\mathbf{x}' = A\mathbf{x}$ is
$$\mathbf{x} = C_1 e^{\lambda_1 t}\mathbf{v}_1 + C_2 e^{\lambda_2 t}\mathbf{v}_2 + \cdots + C_n e^{\lambda_n t}\mathbf{v}_n$$

- Let A be an $n \times n$ diagonalizable matrix with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. Then the general solution to the system $\mathbf{x}' = A\mathbf{x}$ is $\mathbf{x} = e^{At}\mathbf{c}$, where $\mathbf{c}$ is an arbitrary constant vector. If an initial condition $\mathbf{x}(0)$ is specified, then $\mathbf{c} = \mathbf{x}(0)$.

- Let $A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. The eigenvalues of A are $\lambda = a \pm bi$, and if a and b are not both zero, then A can be factored as
$$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$
where $r = |\lambda| = \sqrt{a^2 + b^2}$ and θ is the principal argument of $a + bi$

- Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$ (where $b \neq 0$) and corresponding eigenvector $\mathbf{x}$. Then the matrix $P = [\text{Re}\mathbf{x} \quad \text{Im}\mathbf{x}]$ is invertbible and
$$A = P \begin{bmatrix} a & -b \\ b & a \end{bmatrix} P^{-1}$$

5 Orthogonality

5.1 Orthogonality in $\mathbb{R}^n$

- A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is called an orthogonal set if all pairs of distinct vectors in the set are orthogonal; that is, if
$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \quad \text{whenever} \quad i \neq j \quad \text{for } i, j = 1, 2, \cdots, k$$

- The standard basis of $\mathbb{R}^n$ is an orthogonal set.

- If $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then these vectors are linearly independent.

- An orthogonal basis for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthogonal set.

- Let $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then the unique scalars $c_1, \cdots, c_k$ such that
$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$
are given by
$$c_i = \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \qquad \text{for } i = 1, \cdots, k$$

- A set of vectors $\mathbb{R}^n$ is an orthonormal set if it is an orthogonal set of unit vectors. An orthonormal basis for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthonormal set.

- Let $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then
$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{w} \cdot \mathbf{q}_2)\mathbf{q}_2 + \cdots + (\mathbf{w} \cdot \mathbf{q}_k)\mathbf{q}_k$$
and this representation is unique.

- The columns of an $m \times n$ matrix Q form an orthonormal set IFF $Q^TQ = I_n$

- An $n \times n$ matrix Q whose columns form an orthonormal set is called an orthogonal matrix.
- A square matrix Q is orthogonal IFF $Q^{-1} = Q^T$
- Let Q be an $n \times n$ matrix. The following are equivalent:
  - Q is orthogonal.
  - $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for every $\mathbf{x}$ in $\mathbb{R}^n$
  - $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for every $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$
- If Q is an orthogonal matrix, then its rows form an orthonormal set.
- Let Q be an orthogonal matrix.
  - Q^{-1} is orthogonal.
  - $\det Q = \pm 1$
  - If λ is an eigenvalue of Q, then $|\lambda| = 1$
  - If Q_1 and Q_2 are orthogonal $n \times n$ matrices, then so is $Q_1 Q_2$.

5.2 Orthogonal Complements and Orthogonal Projections

- Let W be a subspace of $\mathbb{R}^n$. We say that a vector $\mathbf{v}$ in $\mathbb{R}^n$ is orthogonal to W if $\mathbf{v}$ is orthogonal to every vector in W. The set of all vectors that are orthogonal to W is called the orthogonal complement of W, denoted $W^\perp$. That is,
$$W^\perp = \{\mathbf{v} \text{ in } \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \quad \text{for all } \mathbf{w} \text{ in } W\}$$
- Let W be a subspace of $\mathbb{R}^n$.
  - $W^\perp$ is a subspace of $\mathbb{R}^n$
  - $(W^\perp)^\perp = W$
  - $W \cap W^\perp = \{\mathbf{0}\}$
  - If $W = \text{span}(\mathbf{w}_1, \ldots, \mathbf{w}_k)$, then $\mathbf{v}$ is in $W^\perp$ IFF $\mathbf{v} \cdot \mathbf{w}_i = 0$ for all $i = 1, \ldots, k$.
- Let A be an $m \times n$ matrix. Then the orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:
$$(\text{row}(A))^\perp = \text{null}(A) \quad \text{and} \quad (\text{col}(A))^\perp = \text{null}(A^T)$$
- Let W be a subspace of $\mathbb{R}^n$ and let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal basis for W. For any vector $\mathbf{v}$ in $\mathbb{R}^n$, the orthogonal projection of $\mathbf{v}$ onto W is defined as
$$\text{proj}_w(\mathbf{v}) = \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 + \cdots + \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)\mathbf{u}_k$$
The complement of $\mathbf{v}$ orthogonal to W is the vector
$$\text{perp}_w(\mathbf{v}) = \mathbf{v} - \text{proj}_w(\mathbf{v})$$
- $\text{proj}_w(\mathbf{v}) = \text{proj}_{\mathbf{u}_1}(\mathbf{v}) + \cdots + \text{proj}_{\mathbf{u}_k}(\mathbf{v})$
- The orthogonal decomposition theorem: Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then there are unique vectors $\mathbf{w}$ in W and $\mathbf{w}^\perp$ in $W^\perp$ such that
$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$$
- If W is a subspace of $\mathbb{R}^n$ then
$$(W^\perp)^\perp = W$$
- If W is a subspace of $\mathbb{R}^n$ then
$$\text{dim}W + \text{dim}W^\perp = n$$
- The Rank Theorem: If A is an $m \times n$ matrix, then
$$\text{rank}(A) + \text{nullity}(A) = n$$

5.3 The Gram-Schmidt Process and the QR Factorization

- The Gram-Schmidt Process: Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be a basis for a subspace W of $\mathbb{R}^n$ and define the following:

$$\mathbf{v}_1 = \mathbf{x}_1; \qquad W_1 = \text{span}(\mathbf{x}_1)$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1, \qquad W_2 = \text{span}(\mathbf{x}_1, \mathbf{x}_2)$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{x}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2, \qquad W_3 = \text{span}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$$

$$\vdots$$

$$\mathbf{v}_k = \mathbf{x}_k - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_k}{\mathbf{v}_1 \cdot = v_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{x}_k}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_{k-1} \cdot \mathbf{x}_k}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}}\right)\mathbf{v}_{k-1}, \qquad W_k = \text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$$

  Then for each $i = 1, \ldots, k, \{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for W_i. In particular, $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for W.

- QR Factorization: Let A be an $m \times n$ matrix with linearly independent columns. Then A can be factored as $A = QR$, where Q is an $m \times n$ matrix with orthonormal columns and R is an invertible upper triangular matrix.

- Finding the QR factorization: find an orthonormal basis for col(A) using the Gram-Schmidt Process. Then, $Q = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k]$. Then, use the fact that $A = QR$ and $Q^TQ = I$ since Q has orthonormal columns. Therefore $Q^TA = Q^TQR = IR = R$

5.4 Orthogonal Diagonalization of Symmetric Matrices

- A square matrix A is orthogonally diagonalizable if there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^TAQ = D$
- If A is orthogonally diagonalizable, then A is symmetric.
- If A is a real symmetric matrix, then the eigenvalues of A are real.
- If A is a symmetric matrix, then any two eigenvectors corresponding to distinct eigenvalues of A are orthonormal.
- The spectral theorem: Let A be an $n \times n$ real matrix. Then A is symmetric IFF it is orthogonally diagonalizable.
- Spectral decomposition:

$$A = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T$$

5.5 Applications

- A quadratic form in n variables is a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$

  where A is a symmetric $n \times n$ matrix and $\mathbf{x}$ is in $\mathbb{R}^n$. We refer to A as the matrix associated with f.

- The principal axes theroem: Every quadratic form can be diagonalized. Specifically, if A is the $n \times n$ symmetric matrix associated with the quadratic form $\mathbf{x}^TA\mathbf{x}$ and if Q is an orthogonal matrix such that $Q^TAQ = D$ is a diagonal matrix, then the change of variable $\mathbf{x} = Q\mathbf{y}$ transforms the quadratic form $\mathbf{x}^TA\mathbf{x}$ into the quadratic form $\mathbf{y}^TD\mathbf{y}$, which has no cross-product terms. If the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$ and $\mathbf{y} = [y_1 \cdots y_n]^T$, then

$$\mathbf{x}^TA\mathbf{x} = \mathbf{y}^TD\mathbf{y} = \lambda_1 y_a^2 + \cdots + \lambda_n y_n^2$$

- A quadratic form $f(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$ is classified as one of the following:
  - positive definite if $f(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$

  - positive semidefinite if $f(\mathbf{x}) \geq 0$ for all $\mathbf{x}$
  - negative definite if $f(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$
  - negative semidefinite if $f(\mathbf{x}) \leq 0$ for all $\mathbf{x}$
  - indefinite if $f(\mathbf{x})$ takes on both positive and negative values

- A symmetric matrix A is called positive definite, positive semidefinite, negative definite, negative semidefinite, or indefinite if the associated quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ has the corresponding property.

- Let A be an $n \times n$ symmetric matrix. The quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is
  - Positive definite IFF all eigenvalues of A are positive.
  - positive semidefinite IFF all eigenvalues are nonnegative.
  - negative definite IFF all eigenvalues are negative
  - negative semidefinite IFF all eigenvalues are nonpositive.
  - indefinite IFF A has both positive and negative eigenvalues.

- Let $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form with associated $n \times n$ symmetric matrix A. Let the eigenvalues of A be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then the following are true, with the constraint of $\|\mathbf{x}\| = 1$:
  - $\lambda_1 \geq f(\mathbf{x}) \geq \lambda_n$
  - The max value of $f(\mathbf{x})$ is λ_1 and occurs when $\mathbf{x}$ is a unit eigenvector corresponding to λ_1
  - The min value of $f(\mathbf{x})$ is λ_n and occurs when $\mathbf{x}$ is a unit eigenvector corresponding to λ_n

- The general form of a quadratic equation in two variables x and y is
$$ax^2 + by^2 + cxy + dx + ey + f = 0$$

- The general form of a quadratic equation in three variables x, y, and z is
$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0$$

6 Vector Spaces

6.1 Vector Spaces and Subspaces

- In the past, we studied vectors in a concrete situation, $\mathbb{R}^n$. Now, we generalize "vectors" by abstracting them into a general setting.

- Let V be a set on which two operations, called addition and scalar multiplication, have been defined. If $\mathbf{u}$ and $\mathbf{v}$ are in V, the sum of $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\mathbf{u} + \mathbf{v}$, and if c is a scalar, the scalar multiple of $\mathbf{u}$ by c is denoted by $c\mathbf{u}$. If the following axioms hold for all $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and for all scalars c and d, then V is called a vector space and its elements are vectors.

1. $\mathbf{u} + \mathbf{v}$ is in V.
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
4. There exists an element $\mathbf{0}$ in V, called a zero vector, such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$
5. For each $\mathbf{u}$ in V, there is an element $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
6. $c\mathbf{u}$ is in V.
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$

10. $1\mathbf{u} = \mathbf{u}$

- Let V be a vector space $\mathbf{u}$ a vector in V, and c a scalar.
  - $0\mathbf{u} = \mathbf{0}$
  - $c\mathbf{0} = \mathbf{0}$
  - $(-1)\mathbf{u} = -\mathbf{u}$
  - If $c\mathbf{u} = \mathbf{0}$, then $c = 0$ or $\mathbf{u} = \mathbf{0}$
- A subset W of a vector space V is called a subspace of V if W is itself a vector space with the same scalars, addition, and scalar multiplication as V.
- Let V be a vector space and let W be a nonempty subset of V. Then W is a subspace of V IFF the following conditions hold:
  - If $\mathbf{u}$ and $\mathbf{v}$ are in W, then $\mathbf{u} + \mathbf{v}$ is in W
  - If $\mathbf{u}$ is in W and c is a scalar, then $c\mathbf{u}$ is in W.
- If W is a subspace of a vector space V, then W contains the zero vector $\mathbf{0}$ of V.
- If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a set of vectors in a vector space V, then the set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is called the span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ and is denoted by $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ or $\text{span}(S)$. If $V = \text{span}(S)$, then S is called a spanning set of V and V is said to be spanned by S.
- Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors in a vector space V.
  - $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is a subspace of V.
  - $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is the smallest subspace of V that contains $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$

6.2 Linear Independence, Basis, and Dimension

- A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in a vector space V is linearly dependent if there are scalars $c_1, c_2, \ldots, c_k$, at least one of which is not zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

  A set of vectors that is not linearly dependent is said to be linearly independent.
- A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in a vector space V is linearly dependent IFF at least one of the vectors can be expressed as a linear combination of the others.
- A set S of vectors in a vector space V is linearly dependent if it contains finitely many linearly dependent vectors. A set of vectors that is not linearly dependent is said to be linearly independent.
- A subset $\mathcal{B}$ of a vector space V is a basis for V if
  - $\mathcal{B}$ spans V and
  - $\mathcal{B}$ is linearly indepnedent.
- Let V be a vector space and let $\mathcal{B}$ be a basis for V. For every vector $\mathbf{v}$ in V, there is exactly one way to write $\mathbf{v}$ as a linear combination of the basis vectors in $\mathcal{B}$
- Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a basis for the vector space V. Let $\mathbf{v}$ be a vector in V, and write $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$. Then $c_1, c_2, \ldots, c_n$ are called the coordinates of $\mathbf{v}$ with respect to $\mathcal{B}$, and the column vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

  is called the coordinate vector of $\mathbf{v}$ with respect to $\mathcal{B}$.

- Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a basis for a vector space V. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in V and let c be a scalar. Then
  - $[\mathbf{u}+\mathbf{v}]_\mathcal{B} = [\mathbf{u}]_\mathcal{B} + [\mathbf{v}]_\mathcal{B}$
  - $[c\mathbf{u}]_\mathcal{B} = c[\mathbf{u}]_\mathcal{B}$
- $$[c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k]_\mathcal{B} = c_1[\mathbf{u}_1]_\mathcal{B} + \cdots + c_k[\mathbf{u}_k]_\mathcal{B}$$
- Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a basis for a vector space V and let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be vectors in V. Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent in V IFF $\{[\mathbf{u}_1]_\mathcal{B}, \ldots, [\mathbf{u}_k]_\mathcal{B}\}$ is linearly independent in $\mathbb{R}^n$.
- Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a basis for a vector space V.
  - Any set of more than n vectors in V must be linearly dependent.
  - Any set of fewer than n vectors in V cannot span V.
- The Basis Theorem: If a vector space V has a basis with n vectors, then every basis for V has exactly n vectors.
- A vector space V is called finite-dimensional if it has a basis consisting of finitely many vectors. The dimension of V, denoted by dim V, is the number of vectors in a basis for V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. A vector space that has no finite basis is called infinite-dimensional.
- Let V be a vector space with dim $V = n$. Then:
  - Any linearly independent set in V contains at most n vectors.
  - Any spanning set for V contains at least n vectors.
  - Any linearly indpendent set of exactly n vectors in V is a basis for V.
  - Any spanning set for V consisting of exactly n vectors is a basis for V.
  - Any linearly independent set in V can be extended to a basis for V.
  - Any spanning set for V can be reduced to a basis for V.
- Let W be a subspace of a finite-dimensional vector space V. Then:
  - W is finite-dimensional and dim $W \leq$ dim V.
  - dim W = dim V IFF $W = V$

6.3 Change of Basis

- Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be bases for a vector space V. The $n \times n$ matrix whose columns are the coordinate vectors $[\mathbf{u}_1]_\mathcal{C}, \ldots, [\mathbf{u}_n]_\mathcal{C}$ of the vectors in $\mathcal{B}$ with respect to $\mathcal{C}$ is denoted by $P_{\mathcal{C}\leftarrow\mathcal{B}}$ and is called the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$. That is,
$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[\mathbf{u}_1]_\mathcal{C}[\mathbf{u}_2]_\mathcal{C} \cdots [\mathbf{u}_n]_\mathcal{C}]$$
- Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be bases for a vector space V and let $P_{\mathcal{C}\leftarrow\mathcal{B}}$ be the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$. Then
  - $P_{\mathcal{C}\leftarrow\mathcal{B}[\mathbf{x}]_\mathcal{B}=[\mathbf{x}]_\mathcal{C}}$ for all $\mathbf{x}$ in V.
  - $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is the unique matrix P with the property that $P[\mathbf{x}]_\mathcal{B} = [\mathbf{x}]_\mathcal{C}$ for all $\mathbf{x}$ in V.
  - $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible and $(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$.
- Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be bases for a vector space V. Let $B = [[\mathbf{u}_1]_\mathcal{E} \cdots [\mathbf{u}_n]_\mathcal{E}]$ and $C = [[\mathbf{v}_1]_\mathcal{E} \cdots [\mathbf{v}_n]_\mathcal{E}]$, where $\mathcal{E}$ is any basis for V. Then the row reduction applied to the $n \times 2n$ augmented matrix $[C|B]$ produces
$$[C|B] \to [I|P_{\mathcal{C}\leftarrow\mathcal{B}}]$$

6.4 Linear Transformations

- A linear transformation from a vector space V to a vector space W is a mapping $T : V \to W$ such that, for all $\mathbf{u}$ and $\mathbf{v}$ in V and for all scalars c,
  - $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
  - $T(c\mathbf{u}) = cT(\mathbf{u})$
- $T : V \to W$ is a linear transformation IFF
$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_kT(\mathbf{v}_k)$$
for all $\mathbf{v}_1, \dots, \mathbf{v}_k$ in V and scalars $c_1, \dots, c_k$.
- Let $T : V \to W$ be a linear transformation. Then:
  - $T(\mathbf{0}) = \mathbf{0}$
  - $T(-\mathbf{v}) = -T(\mathbf{v})$ for all $\mathbf{v}$ in V.
  - $T(\mathbf{u}-\mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$ for all $\mathbf{u}$ and $\mathbf{v}$ in V.
- Let $T : V \to W$ be a linear transformation and let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a spanning set for V. Then $T(\mathcal{B}) = \{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ spans the range of T.
- If $T : U \to V$ and $S : V \to W$ are linear transformations, then the composition of S with T is the mapping $S \circ T$, defined by
$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u}))$$
where $\mathbf{u}$ is in U.
- If $T : U \to V$ and $S : V \to W$ are linear transformations, then $S \circ T : U \to W$ is a linear transformation.
- $R \circ (S \circ T) = (R \circ S) \circ T$
- A linear transformation $T : V \to W$ is invertible if there is a linear transformation $T' : W \to V$ such that
$$T' \circ T = I_V \quad \text{and} \quad T \circ T' = I_W$$
In this case, T' is called an inverse for T.
- If T is an invertible linear transformation, then its inverse is unique.

6.5 The Kernel and Range of a Linear Transformation

- Let $T : V \to W$ be a linear transformation. The kernel of T, denoted $\ker(T)$, is the set of all vectors in V that are mapped by T to $\mathbf{0}$ in W. That is,
$$\ker(T) = \{\mathbf{v} \text{ in } V : T(\mathbf{v}) = \mathbf{0}\}$$
The range of T, denoted $\text{range}(T)$, is the set of all vectors in W that are images of vectors in V under T. That is,
$$\text{range}(T) = \{T(\mathbf{v}) : \mathbf{v} \text{ in } V\} = \{\mathbf{w} \text{ in } W : \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \text{ in } V\}$$
- Let $T : V \to W$ be a linear transformation. Then:
  - The kernel of T is a subspace of V.
  - The range of T is a subspace of W.
- Let $T : V \to W$ be a linear transformation. The rank of T is the dimension of the range of T and is denoted by $\text{rank}(T)$. The nullity of T is the dimension of the kernel of T and is denoted by $\text{nullity}(T)$.
- The rank theorem: Let $T : V \to W$ be a linear transformation from a finite-dimensional vector space V into a vector space W. Then
$$\text{rank}(T) + \text{nullity}(T) = \dim V$$

- A linear transformation $T : V \to W$ is called one-to-one if T maps distinct vectors in V to distinct vectors in W. If range(T) $= W$, then T is called onto.
- $T : V \to W$ is one-tocloftone if, for all $\mathbf{u}$ and $\mathbf{v}$ in V,

$$\mathbf{u} \neq \mathbf{v} \text{ implies that } T(\mathbf{u}) \neq T(\mathbf{v})$$

- Which is to say, if $T : V \to W$ is one-to-one if, for all $\mathbf{u}$ and $\mathbf{v}$ in V,

$$T(\mathbf{u}) = T(\mathbf{v}) \text{ implies that } \mathbf{u} = \mathbf{v}$$

- $T : V \to W$ is onto if, for all $\mathbf{w}$ in W, there is at leeast one $\mathbf{v}$ in V such that

$$\mathbf{w} = T(\mathbf{v})$$

- A linear transformation $T : V \to W$ is one-to-one IFF $\ker(T) = \{\mathbf{0}\}$.
- Let $\dim V = \dim W = n$. Then a linear transformation $T : V \to W$ is one-to-one IFF it is onto
- Let $T : V \to W$ be a one-to-one linear transformation. If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly indepnedent set in V, then $T(S) = \{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_k)\}$ is a linearly independent set in W.
- Let $\dim V = \dim W = n$. Then a one-to-one linear transformation $T : V \to W$ maps a basis for V to a basis for W.
- A linear transformation $T : V \to W$ is invertible IFF it is one-to-one and onto.
- A lineart transformation $T : V \to W$ is called an isomorphism if it is one-to-one and onto. If V and W are two vector spaces such that there is an isomorphism from V to W, then we say that V is isomorphic to W and write $V \cong W$.
- Let V and W be two finite-dimensional vector spaces (over the same field of scalars). Then V is isomorphic to W IFF $\dim V = \dim W$.

6.6 The Matrix of a Linear Transformation

- Let V and W be two finite-dimensional vector spaces with bases $\mathcal{B}$ and $\mathcal{C}$, respectively, where $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. If $T : V \to W$ is a linear transformation, then the $m \times n$ matrix A defined by

$$A = [[T(\mathbf{v}_1)]_{\mathcal{C}}|[T(\mathbf{v}_2)]_{\mathcal{C}}| \cdots |[T(\mathbf{v}_n)]_{\mathcal{C}}]$$

satisfies

$$A[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{C}}$$

for every vector $\mathbf{v}$ in V.

- $[T]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{C}}$
- $[T]_{\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{B}}$
- Let U, V, and W be finite-dimensional vector spaces with bases $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$, respectively. Let $T : U \to V$ and $S : V \to W$ be linear transformations. Then

$$[S \circ T]_{\mathcal{D}\leftarrow\mathcal{B}} = [S]_{\mathcal{D}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}}$$

- Let $T : V \to W$ be a linear transformation between n-dimensional vector spaces V and W and let $\mathcal{B}$ and $\mathcal{C}$ be bases for V and W, respectively. Then T is invertible IFF the matrix $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible. In this case,

$$([T]_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = [T^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}}$$

- Let V be a finite dimensional vector space with bases $\mathcal{B}$ and $\mathcal{C}$ and let $T : V \to V$ be a linear transformation. Then
$$[T]_{\mathcal{C}} = P^{-1}[T]_{\mathcal{B}}P$$
where P is the change-of-basis matrix from $\mathcal{C}$ to $\mathcal{B}$.
- Let V be a finite-dimensional vector space and let $T : V \to V$ be a linear transformation. Then T is called diagonalizable if there is a basis $\mathcal{C}$ for V such that the matrix $[T]_{\mathcal{C}}$ is a diagonal matrix.
- The Fundamental Theorem of invertible matrices: version 4.
  - A is invertible
  - $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$
  - $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
  - The reduced row echelon form of A is I_n.
  - A is the product of elementary matrices.
  - $\text{rank}(A) = n$
  - $\text{nullity}(A) = 0$
  - The column vectors of A are linearly independent
  - The column vectors of A span $\mathbb{R}^n$
  - The column vectors of A form a basis for $\mathbb{R}^n$
  - The row vectors of A are linearly independent
  - The row vectors of A span $\mathbb{R}^n$
  - The row vectors of A form a basis for $\mathbb{R}^n$
  - $\det A \neq 0$
  - 0 is not an eigenvalue of A
  - T is invertible.
  - T is one-to-one.
  - T is onto.
  - $\ker(T) = \{\mathbf{0}\}$
  - $\text{range}(T) = W$

6.7 Applications

- The set S of all solutions to $y' + ay = 0$ is a subspace of $\mathscr{F}$
- If S is the solution space of $y' + ay = 0$, then $\dim S = 1$ and $\{e^{-at}\}$ is a basis for S.
- Let S be the solution space of
$$y'' + ay' + by = 0$$
and let λ_1 and λ_2 be the roots of the characteristic equation $\lambda^2 + a\lambda + b = 0$.
  - If $\lambda_1 \neq \lambda_2$, then $\{e^{\lambda_1 t}, e^{\lambda_2 t}\}$ is a basis for S.
  - If $\lambda_1 = \lambda_2$, then $\{e^{\lambda_1 t}, te^{\lambda_1 t}\}$ is a basis for S.

7 Distance and Approximation

7.1 Inner Product Spaces

- An inner product on a vector space V is an operation that assigns to every pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ such that the following properties hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars c:
  - $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
  - $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
  - $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
  - $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ IFF $\mathbf{u} = \mathbf{0}$
- A vector space with an inner product is called an inner product space.
- Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in an inner product space V and let c be a scalar.
  - $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v} + \mathbf{w} \rangle$
  - $\langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
  - $\langle \mathbf{u}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle = 0$
- Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V.
  - The length (or norm) of $\mathbf{v}$ is $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.
  - The distance between $\mathbf{u}$ and $\mathbf{v}$ is $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$
  - $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.
- Pythagoras' Theorem: Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal IFF

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

- The Cauchy-Schwarz Inequality: Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

  with equality holding IFF $\mathbf{u}$ and $\mathbf{v}$ are scalar multiples of each other.
- The triangle inequality: Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Then

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

7.2 Norms and Distance Functions

- A norm on a vector space V is a mapping that associates with each vector $\mathbf{v}$ a real number $\|\mathbf{v}\|$, called the norm of $\mathbf{v}$, such that the following properties are satisfied for all vectors $\mathbf{u}$ and $\mathbf{v}$ and all scalars c:
  - $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ IFF $\mathbf{v} = \mathbf{0}$
  - $\|c\mathbf{v}\| = |c$
    $|\mathbf{v}\|$
  - $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
- A vector space with a norm is called a normed vector space.
- We define a distance function for any norm as:

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

- Let d be a distance function defined on a normed linear space V. The following properties hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V:
  - $d(\mathbf{u}, \mathbf{v}) \geq 0$, and $d(\mathbf{u}, \mathbf{v}) = 0$ IFF $\mathbf{u} = \mathbf{v}$

  - $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
  - $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$

- A matrix norm on M_{nn} is a mapping that associates with each $n \times n$ matrix A a real number $\|A\|$, called the norm of A, such that the following properties are satisfied for all $n \times n$ matrices A and B and all scalars c.
  - $\|A\| \geq 0$ and $\|A\| = 0$ IFF $A = O$.
  - $\|cA\| = |c$
    $|A\|$
  - $\|A + B\| \leq \|A\| + \|B\|$
  - $\|AB\| \leq \|A\|\|B\|$

- A matrix norm on M_{nn} is said to be compatible with a vector norm on $\|\mathbf{x}\|$ on $\mathbb{R}^n$ if, for all $n \times n$ matrices A and all vectors $\mathbf{x}$ in $\mathbb{R}^n$, we have
$$\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$$

- The Frobenius norm is given by
$$\|A\|_F = \sqrt{\sum_{i,j=1}^{n} a_{ij}^2}$$

- If $\|\mathbf{x}\|$ is a vector norm on $\mathbb{R}^n$, then $\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$ defines a matrix norm on M_{nn} that is compatible with the vector norm that induces it.

- The matrix norm $\|A\|$ in the previous is called the operator norm induced by the vector norm $\|\mathbf{x}\|$

- Let A be an $n \times n$ matrix with column vectors $\mathbf{a}_i$ and row vectors $\mathbf{A}_i$ for $i = 1, \ldots, n$.
$$\text{a. } \|A\|_1 = \max_{j=1,\ldots,n} \left\{\|\mathbf{a}_j\|_s\right\} = \max_{j=1,\ldots,n} \left\{\sum_{i=1}^{n} |a_{ij}|\right\}$$
$$\text{b. } \|A\|_\infty = \max_{i=1,\ldots,n} \{\|\mathbf{A}_i\|_s\} = \max_{i=1,\ldots,n} \left\{\sum_{j=1}^{n} |a_{ij}|\right\}$$

- A matrix A is ill-conditioned if small changes in its entries can produce large changes in the solutions to $A\mathbf{x} = \mathbf{b}$. If small changes in the entries of A produce only small changes in the solutions to $A\mathbf{x} = \mathbf{b}$, then A is called well-conditioned.

7.3 Least Squares Approximation

- If A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathbb{R}^m$, a least squares solution of $\overline{A\mathbf{x}} = \mathbf{b}$ is a vector $\overline{\mathbf{x}}$ in $\mathbb{R}^n$ such that
$$\|\mathbf{b} - A\overline{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$
for all $\mathbf{x}$ in $\mathbb{R}^n$.

- The Least Squares Theorem: Let A be an $m \times n$ matrix and let $\mathbf{b}$ be in $\mathbb{R}^m$. Then $A\mathbf{x} = \mathbf{b}$ always has at least one least squares solution $\overline{\mathbf{x}}$. Moreover:
  - $\overline{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\overline{\mathbf{x}}$ is a solution of the normal equations $A^T A\overline{\mathbf{x}} = A^T\mathbf{b}$.
  - A has linearly independent columns if and only if $A^T A$ is invertible. In this case, the least squares solution of $A\mathbf{x} = \mathbf{b}$ is unique and is given by
$$\overline{\mathbf{x}} = \left(A^T A\right)^{-1} A^T\mathbf{b}$$

- Let A be an $m \times n$ matrix with linearly independent columns and let $\mathbf{b}$ be in $\mathbb{R}^m$. If $A = QR$ is a QR factorization of A, then the unique least squares solution $\overline{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ is
$$\overline{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$$

- Let W be a subspace of $\mathbb{R}^m$ and let A be an $m \times n$ matrix whose columns form a basis for W. If $\mathbf{v}$ is any vector in $\mathbb{R}^m$, then the orthogonal projection of $\mathbf{v}$ onto W is the vector

$$\operatorname{proj}_W(\mathbf{v}) = A\left(A^T A\right)^{-1} A^T \mathbf{v}$$

  The linear transformation $P : \mathbb{R}^m \to \mathbb{R}^m$ that projects $\mathbb{R}^m$ onto W has $A\left(A^T A\right)^{-1} A^T$ as its standard matrix.

- If A is a matrix with linearly independent columns, then the pseudoinverse of A is the matrix A^+defined by

$$A^+ = \left(A^T A\right)^{-1} A^T$$

- Let A be a matrix with linearly independent columns. Then the pseudoinverse $A^+$of A satisfies the following properties, called the Penrose conditions for A:
  - $AA^+A = A$
  - $A^+AA^+ = A^+$
  - $AA^+$and A^+A are symmetric.

7.4 The Singular Value Decomposition

- If A is an $m \times n$ matrix, the singular values of A are the square roots of the eigenvalues of $A^T A$ and are denoted by $\sigma_1, \dots, \sigma_n$. It is conventional to arrange the singular values so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

- The Singular Value Decomposition: Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Then there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ matrix Σ of the form shown in Equation (1) such that

$$A = U\Sigma V^T$$

- The Outer Product Form of the SVD: Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Let $\mathbf{u}_1, \dots, \mathbf{u}_r$ be left singular vectors and let $\mathbf{v}_1, \dots, \mathbf{v}_r$ be right singular vectors of A corresponding to these singular values. Then

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

- Let $A = U\Sigma V^T$ be a singular value decomposition of an $m \times n$ matrix A. Let $\sigma_1, \dots, \sigma_r$ be all the nonzero singular values of A. Then:
  - The rank of A is r.
  - $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ is an orthonormal basis for $\operatorname{col}(A)$.
  - $\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ is an orthonormal basis for $\operatorname{null}\left(A^T\right)$.
  - $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is an orthonormal basis for $\operatorname{row}(A)$.
  - $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is an orthonormal basis for $\operatorname{null}(A)$.

- Let A be an $m \times n$ matrix with rank r. Then the image of the unit sphere in $\mathbb{R}^n$ under the matrix transformation that maps $\mathbf{x}$ to $A\mathbf{x}$ is
  - the surface of an ellipsoid in $\mathbb{R}^m$ if $r = n$.
  - a solid ellipsoid in $\mathbb{R}^m$ if $r < n$.

- Let A be an $m \times n$ matrix and let $\sigma_1, \dots, \sigma_r$ be all the nonzero singular values of A. Then

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

- If A is an $m \times n$ matrix and Q is an $m \times m$ orthogonal matrix, then

$$\|QA\|_F = \|A\|_F$$

- Let $A = U\Sigma V^T$ be an SVD for an $m \times n$ matrix A, where $\Sigma = \begin{bmatrix} D & O \\ O & O \end{bmatrix}$ and D is an $r \times r$ diagonal matrix containing the nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ of A. The pseudoinverse (or Moore-Penrose inverse) of A is the $n \times m$ matrix A^+defined by

$$A^+ = V\Sigma^+U^T$$

where $\Sigma^+$is the $n \times m$ matrix

$$\Sigma^+ = \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}$$

- The least squares problem $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution $\overline{\mathbf{x}}$ of minimal length that is given by

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

- The Fundamental Theorem of invertible matrices: Final Version.
  - A is invertible
  - $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$
  - $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
  - The reduced row echelon form of A is I_n.
  - A is the product of elementary matrices.
  - $\text{rank}(A) = n$
  - $\text{nullity}(A) = 0$
  - The column vectors of A are linearly independent
  - The column vectors of A span $\mathbb{R}^n$
  - The column vectors of A form a basis for $\mathbb{R}^n$
  - The row vectors of A are linearly independent
  - The row vectors of A span $\mathbb{R}^n$
  - The row vectors of A form a basis for $\mathbb{R}^n$
  - $\det A \neq 0$
  - 0 is not an eigenvalue of A
  - T is invertible.
  - T is one-to-one.
  - T is onto.
  - $\ker(T) = \{\mathbf{0}\}$
  - $\text{range}(T) = W$
  - 0 is not a singular value of A.

7.5 Applications

- General problem of approximating functions can be stated as: Given a continuous function f on an interval $[a, b]$ and a subspace W of $\mathscr{C}[a, b]$, find the function "closest" to f in W.
- The n'th order Fourier approximation to f on $[-\pi, \pi]$:

$$a_0 = \frac{\langle 1, f \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)dx$$
$$a_k = \frac{\langle \cos kx, f \rangle}{\langle \cos kx, \cos kx \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx dx$$
$$b_k = \frac{\langle \sin kx, f \rangle}{\langle \sin kx, \sin kx \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx dx$$