Linear algebra

University of Houston - MATH 2331

Spring 2020

Textbook: Lay - Linear Algebra and Its Applications, 5th edition

Chapters covered: 1.1-5, 1.7-9, 2.1-4, 2.8-9, 3, 4.1-6, 4.9, 5.1-3, 5.5, 6.1-6, 7.1-4

*The contents in these notes do not correspond exactly to the textbook

Contents

1 Matrices

Definition 1.1: Matrix

A **matrix** is a rectangular array of objects. Let A_{ij} denote a matrix with $i = \{1, \ldots, m\}$ and $j = \{1, \ldots, n\}$. Then A is an $m \times n$ matrix with m rows and n columns and (i, j)-entries.

$$A_{i,j} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

Definition 1.2: Zero matrix

A **zero matrix** has all 0s in its entries.

Definition 1.3: Identity matrix

An **identity matrix** I_n is an $n \times n$ matrix such that for any $n \times n$ matrix A, $AI = IA = A$.

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

A **permutation matrix** is an identity matrix that has its rows permuted.

An **elementary matrix** is an identity matrix with exactly one row operation performed.

Definition 1.4: Triangular matrix

A **triangular matrix** is an $n \times n$ matrix with all zeros below or above the diagonal.

An **upper triangular matrix** looks like

$$U_{i,j} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ 0 & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{bmatrix}$$

A **lower triangular matrix** looks like

$$L_{i,j} = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ a_{2,1} & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}$$

1.1 Row reduction

Row reduction (also known as Gaussian elimination) is the process of using elementary row operations to transform a matrix into another form.

The **elementary row operations** are:

- replacement - add a multiple of a row to another row
- exchange - swap the position of two rows
- scaling - multiply a row by a nonzero constant

Definition 1.5: Reduced row echelon form

A matrix is in **echelon form** if it has the following three properties:

1. All rows of only zeros are at the bottom of the matrix.
2. Every leading entry of a row is in a column to the right of the leading entry of the row above it. (Each leading entry is now a **pivot** – their positions do not change.)
3. All entries in a column below a leading are zeros.

A matrix is in **row reduced echelon form** if it also satisfies the following:

4. The leading entry in each nonzero row is 1.
5. Each leading 1 is the only nonzero entry in its column.

Algorithm for row reduction

The row operations can be performed in any order to get to reduced row echelon form. Here is an example of a systematic algorithm for row reduction. Depending on the entries of the matrix, other ways may be more efficient.

1. Put all rows of only zero at the bottom.
2. Make the first row of the matrix have a leading entry of 1.
3. Add multiples of the first row to every row below it so that the rest of the first column is 0.
4. Make the second row of the matrix have a leading entry of 1.
5. Add multiples of the second row to every row above and below it such that the rest of the second column is 0.
6. Repeat the process and exchange rows until the matrix is in reduced row echelon form.

Theorem 1.1

Each matrix is row equivalent to one and only one reduced echelon matrix.

1.2 Solving the matrix equation

Let A be an $m \times n$ matrix and let $\mathbf{x}, \mathbf{b}$ be vectors in $\mathbb{R}^n$. An important equation is

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

The goal is to solve for $\mathbf{x}$ given A and $\mathbf{b}$.

1. Arrange A and $\mathbf{b}$ into an augmented matrix.
$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$$
2. Row reduce the augmented matrix until it is in row reduced echelon form.
3. The last column on the right is the solution.

*If the ending matrix has a row $\begin{bmatrix} 0 & 0 & \cdots & k \end{bmatrix}$, then there is no solution to the equation.

Properties of $A\mathbf{x}$

Let A be an $m \times n$ matrix, let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$, and let c be a scalar.

- $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
- $A(c\mathbf{v}) = c(A\mathbf{v})$

1.3 Matrix algebra

Addition

To add to matrices of the same size, add the corresponding entries.

Scalar multiplication

Multiply each entry in the matrix by the scalar.

Matrix multiplication

Matrix multiplication is not commutative. The matrices also have to be a certain dimension; the left matrix is $m \times n$, the right matrix is $n \times r$, and the resulting matrix is $m \times r$.

Let A and B be matrices with the appropriate dimensions.

- **right-multiplication**: AB is multiplying A by B from the right.
- **left-multiplication**: BA is multiplying A by B from the left.

Steps to matrix multiplication:

1. Multiply $a_{1,1}$ by $b_{1,1}$ and enter it into $c_{1,1}$.
2. Multiply $a_{2,1}$ by $b_{1,1}$ and enter it into $a_{2,1}$.
3. Repeat until the entire first column of A has been multiplied.
4. Multiply $a_{1,1}$ by $b_{1,2}$ and enter it into $c_{1,2}$.
5. Multiply $a_{2,1}$ by $b_{1,2}$ and enter it into $c_{2,2}$.

6. Repeat until the all columns of the first row of B has been multiplied. This completes the first matrix

7. Repeat the process with the second column of A and the second row of B. This completes the second matrix

8. Repeat until all the columns of A have been completed.

9. Add up all the individual matrices.

2 Vector spaces

Definition 2.1: Vector space

A vector space V is a nonempty set of objects (called **vectors**) on which addition and scalar multiplication is defined. A vector space must fulfill all of the following axioms:

-
-
-
-
-
-
-
-
-
-

Definition 2.2: Vector subspace

Let V be a vector space. Then H is a **subspace** of V if and only if

1. The zero vector of V is also in H.
2. H is closed under vector addition. (For every $\mathbf{u}, \mathbf{v} \in H$, $\mathbf{u} + \mathbf{v} \in H$).
3. H is closed under scalar multiplication. (For every $\mathbf{v} \in H$, $c\mathbf{v} \in H$).

Definition 2.3: Linear independence

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is **linearly independent** if an only if the trivial solution (that is, $\mathbf{0}$) is the only solution to

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

Definition 2.4: Span

Let a set of vectors be $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Then $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is the set of all vectors $\mathbf{x}$ such that

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Definition 2.5: Basis

Let H be a vector space. A **basis** is set of vectors $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ such that $\mathcal{B}$ spans H and is linearly independent.

Every basis can be interpreted as coordinate system for a vector space. Let $\mathbf{x}$ be in the vector space. Then the coordinates of $\mathbf{x}$ relative to $\mathcal{B}$ is the set of weights such that

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$$

Steps to finding coordinates:

1. Organize the basis into a matrix.
2. Solve

$$\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n\end{bmatrix}\begin{bmatrix}c_1 \\ c_2 \\ \vdots \\ c_n\end{bmatrix} = \mathbf{x}$$

2.1 Fundamental subspaces

Definition 2.6: Column space

Let A be an $m \times n$ matrix such that

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}.$$

Then the **column space** of A is the span of the columns of A, that is,

$$\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$$

Since each column vector of A has m entries, the column space of a matrix is a subspace of $\mathbb{R}^m$.

$\operatorname{Col} A = \mathbb{R}^m$ if and only if $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b} \in \mathbb{R}^m$.

$\dim \operatorname{Col} A \leq m$ is always true.

Steps to find a basis for $\operatorname{Col} A$:

1. Row reduce A
2. Find the pivot columns

The corresponding pivot columns in A form a basis for $\operatorname{Col} A$.

Definition 2.7: Null space

Let A be an $m \times n$ matrix. The **null space** of A is the set of all solutions to $A\mathbf{x} = \mathbf{0}$.

Since each row vector of A has n entries, the null space of a matrix is a subspace of $\mathbb{R}^n$.

Steps to find a basis for $\operatorname{Nul} A$:

1. Augment the matrix by $A = \begin{bmatrix} A & \mathbf{0} \end{bmatrix}$.
2. Row reduce the augmented matrix.
3. Solve for $x_1, x_2, \ldots, x_n$.
4. Create the general solution in parametric form.

The vectors attached to the free variables in the general solution form a basis for $\operatorname{Nul} A$.

2.2 Linear transformations

Definition 2.8: Linear transformations

Let $T : V \to W$ be a function. Then T is a **linear transformation** that assigns each $\mathbf{x} \in V$ to a vector $T(\mathbf{x})$ if and only if

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
2. $T(c\mathbf{u}) = \mathrm{c}\; T(\mathbf{u})$

Definition 2.9: Kernel and range

Let $T : V \to W$ be a linear transformation. Then the **kernel** of the transformation is the set of all $\mathbf{x} \in V$ such that $T(\mathbf{u}) = \mathbf{0}$.

If $T(\mathbf{x}) = A(\mathbf{x})$ for some matrix A, then $\operatorname{Ker} T = \operatorname{Nul} A$.

The **range** of the transformation is the set of all vectors in W such that there exists $\mathbf{x} \in V$ with $T(\mathbf{x}) \in W$.

If $T(\mathbf{x}) = A(\mathbf{x})$ for some matrix A, then $\operatorname{Range} T = \operatorname{Col} A$.

Let $T : V \to W$ be a linear transformation. Steps to find a transformation matrix:

1. Determine a basis for V.
2. Transform each basis vector to a vector in W.
3. Put the resulting vectors in the corresponding columns

A basis can be used as coordinates! The standard basis is $\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_n$ is the familiar x and y coordinates extended to n dimensions. However, if a basis spans $\mathbb{R}^n$, that basis can be used as a coordinate system for $\mathbb{R}^n$.

Steps to find the coordinates $(c_1, c2, \ldots, c_n)$ relative to a basis $\mathcal{B}$. Let B be the change of coordinates matrix.

$$P = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$$

Note that P is a bijective (one to one and onto) linear transformation from V to $\mathbb{R}^n$. This is called an **isomorphism** from V onto $\mathbb{R}^n$.

3 Determinants

Definition 3.1: Determinant

The **determinant** is a number that is found by doing operations on the entries of an $n \times n$ matrix. It demonstrates some properties of the matrix. The defining properties of the determinant are:

1.
2.
3.

Other properties of the determinant:

- if two rows of A are equal, then $\det A = 0$
- if B is obtained from A by subtracting a multiple of one row from another row, then $\det B = \det A$.
- if A has a row of zeros, $\det A = 0$
- if A is upper triangular or lower triangular, then $\det A$ is the product of the main diagonal.
- if A is singular, then $\det A = 0$. IfA is invertible then $\det A \neq 0$.
- $\det (AB) = \det A \times \det B$, if A and B are $n \times n matrices.$
- $\det A^\top = \det A$

4 Eigenvector, eigenvalue, eigenspace

Definition 4.1: Eigenvector and eigenvalue

Let A be an $n \times n$ matrix. Let λ be a scalar and $\mathbf{x}$ be a nonzero vector. If

$$A\mathbf{x} = \lambda\mathbf{x}$$

then λ is called an **eigenvalue** of A and $\mathbf{x}$ is an **eigenvector** corresponding to λ. The set of all $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for a certain eigenvalue is called an **eigenspace**. An eigenspace is also the set of all solutions to

$$(A - \lambda I)\,\mathbf{x} = \mathbf{0}$$

or $\mathrm{Nul}\,(A - \lambda I)$

algebraic multiplicity: the multiplicity of the eigenvalue in the characteristic polynomial geometric multiplicity: the dimension of the eigenspace corresponding to the eigenvalue

If a matrix is triangular, then the eigenvalues lie on the diagonal.

If there are two eigenvectors corresponding to distinct eigenvalues, then the eigenvectors are linearly independent.

Steps to finding eigenvalues, eigenvectors, and a basis for the eigenspace:

1. Solve the **characteristic equation** $\det(A - \lambda I) = 0$ for λ. λ is a variable and the solutions are the roots of a polynomial.
2. Plug in a value for λ into $A - \lambda I$.
3. Find the null space of the resulting matrix. This is the eigenspace of A corresponding to λ.
4. Set the free variables in the general parameterized solution to any nonzero constant. The resulting vector is an eigenvector corresponding to λ.

4.0.1 Complex eigenvalues

A matrix can have complex eigenvalues. Let A be a 2×2 matrix with complex eigenvalues.

The process of finding eigenvalues is the same as normal, except there will be imaginary components to the eigenvalues. (Solve the characteristic equation for the eigenvalues).

To find the eigenvectors,

1. Find $A - \lambda I$
2. Find the null space of $A - \lambda I$.
3. Solve the resulting linear equation by inspection.
4. Choose one of the equations. Set a variable equation to 1 or another constant.
5. Solve for the other variable. The resulting 2 variables are the components of the eigenvectors.

Similarity

Definition 4.2: Similarity of matrices

Let A and B be matrices. A is **similar** to B if and only if there exists an invertible matrix P such that

$$A = PBP^{-1}.$$

Similarity is an equivalence relation.

Theorem 4.1

Let A and B be matrices. If A is similar to B, then A and B have the same eigenvalues

Note: this is not a biconditional! If A and B have the same eigenvalues, it does NOT necessarily mean that A and B are similar!

5 Orthogonality

Definition 5.1: Inner product

Let V be a vector space and let F be $\mathbb{R}$ or $\mathbb{C}$. Let $\mathbf{u}, \mathbf{v} \in V$. The **inner product** of $\mathbf{u}$ and $\mathbf{v}$ is a function such that

$$\langle \mathbf{u}, \mathbf{v} \rangle : V \times V \to F$$

The inner product must fulfill the following properties

- positive definite

$$\langle \mathbf{v}, \mathbf{v} \rangle > 0 \text{ and} \langle \mathbf{v}, \mathbf{v} \rangle = 0 \text{ if and only if } \mathbf{v} = 0$$

- linear in the first argument

$$\langle a\mathbf{u}, \mathbf{v} \rangle = a \langle \mathbf{u}, \mathbf{v} \rangle$$

$$\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$$

- conjugate symmetry

$$\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$$

In $\mathbb{R}^n$, the inner product is the **dot product**. It is defined as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots u_n v_n$$

The length (or norm) of a vector is

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

Note also that

$$\|vectv\|^2 = \mathbf{v} \cdot \mathbf{v}$$

A unit vector is

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

The distance bewtween two points is

$$\|\mathbf{u} - \mathbf{v}\|$$

Definition 5.2: Orthogonal and orthonormal

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}$. Then $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if and only if

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

If $\mathbf{u}, \mathbf{v}$ are orthogonal and they have magnitude 1, then they are **orthonormal**.

Definition 5.3: Projection

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}$. The **projection** of $\mathbf{u}$ onto $\mathbf{v}$ is defined as

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v}$$

Theorem 5.1

Let W be a subspace of $\mathbb{R}^n$ with an orthogonal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_n\}$. Then every $\mathbf{y} \in W$ can be written as

$$\mathbf{y} = \sum_{i=1}^{n} \operatorname{proj}_{\mathbf{u}_i} \mathbf{y}$$

And every $\mathbf{x} \in \mathbb{R}$ can be written as

$$\mathbf{x} = \sum_{i=1}^{n} \operatorname{proj}_{\mathbf{u}_i} \mathbf{x} + \mathbf{w}$$

where $\mathbf{w} \perp W$ and is the orthogonal projection of $\mathbf{x}$ onto W.

The shortest distance from $\vec{x}$ to W is $\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{w}\|$.

Definition 5.4: Orthonormal matrix

Let U be an $m \times n$ matrix. U is an **orthogonal matrix** if the columns of U are all othogonal. U is an **orthonormal matrix** if the columns of U are orthonormal.

A property of **orthonormal matrix**. Let U be an orthonormal matrix. then

$$U^\top U = I$$

Theorem 5.2

Let W be a subspace of $\mathbb{R}^n$ and U be a matrix with orthonormal columns that form a basis for W. Let $\mathbf{y} \in \mathbb{R}^n$. Then

$$\text{proj}_W \mathbf{y} = UU^\top \mathbf{y}$$

Definition 5.5: Gram-Schmidt process

The Gram-Schmidt process takes any basis of a subspace of $\mathbb{R}^n$ and produces an orthogonal basis that spans the same subspace.

Let $\beta = \{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_k\}$ be a basis for a subspace S of $\mathbb{R}^n$.

Steps for Gram-Schmidt:

1. Let $\mathbf{v}_1 = \mathbf{x}_1$.
2. Project $\mathbf{x}_2$ onto $\mathbf{v}_1$ and define $\mathbf{v}_2 = \mathbf{x}_2 - \text{proj}_{\mathbf{v}_1} \mathbf{x}_2$.
3. Keep projecting the next vector $\mathbf{x}_p$ onto each every orthogonal vector before it and substract all the projections from $\mathbf{x}_p$.

In general, for $1 < p \leq k$, the process is

$$\mathbf{v}_p = \mathbf{x}_p - \sum_{i=1}^{p-1} \text{proj}_{\mathbf{v}_i} \mathbf{x}_p$$

6 Symmetric matrices

Definition 6.1: Symmetric matrix

Let A be an $n \times n$ matrix such that $A = A^\top$. Then A is a **symmetric** matrix.

Theorem 6.1

Let A be a matrix. If A is a symmetric matrix, then any two eigenvectors from distinct eigenspaces are orthogonal (A has mutually orthogonal eigenspaces).

Theorem 6.2

Let A be a matrix. A is a symmetric matrix if and only if it is orthogonally diagonalizable.

Theorem 6.3: Spectral theorem

Let A be an $n \times n$ symmetric matrix. Then A has the following properties:

- It has n real eigenvalues (sometimes called **spectra**).
- The dimension of every eigenspace is equivalent to the algebraic multiplicity of the corresponding eigenvalue.
- A has mutually orthogonal eigenspaces.

6.1 Quadratic forms

Definition 6.2: Quadratic forms

Let A be an $n \times n$ symmetric matrix. A quadratic form is a function $Q : \mathbb{R}^n \to \mathbb{R}$ such that

$$Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$$

Classification of quadratic forms

classification	description	eigenvalues of A
positive definite	$Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$	positive
positive semidefinite	$Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$	nonnegative
negative definite	$Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$	negative
negative semidefinite	$Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$	nonpositive
indefinite	$Q(\mathbf{x})$ is positive/negative	positive/negative

7 Applications

7.1 Systems of linear equations

a linear system has

- no solutions
- a unique solutions
- infinite solutions

Definition 7.1: inconsistent

a linear system with no solutions

solving linear systems is the same as solving the equation

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

coefficient matrix is a matrix with all the coefficients of the equations.

augmented matric is the matric with the coefficients and the other side of the equation.

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$$

7.2 Markov chains

Definition 7.2: Stochastic matrix

An $n \times n$ matrix whose columns are probability vectors (vectors whose components are nonnegative and add to 1).

Definition 7.3: Markov chain

Let P be a stochastic matrix. Then a **Markov chain** is a sequence of probability vectors such that given an initial vector $\mathbf{x}_0$,

$$\mathbf{x}_{k+1} = P\mathbf{x}_k$$

Construction of a stochastic matrix for a Markov chain:

Definition 7.4: Steady-state vector

A **steady-state vector** is a probability vector such that

$$P\mathbf{x}_s = \mathbf{x}_s$$

Let $\mathbf{x}_0$ be any probability vector. Then, the **steady-state vector** is

$$\lim_{n\to\infty} P^n\mathbf{x}_0 = \mathbf{x}_s$$

Steps to calculate the steady-state vector:

1. Calculate $P - I$.
2. To simplify calculations, multiply $P - I$ by some power of 10.
3. Row reduce and find a basis for null space of the resulting matrix (should be one-dimensional).
4. Take the vector in the basis and divide all the components by the sum of the components to turn it into a probability vector.

The resulting vector is the steady state vector of P.

7.3 Least-squares

Let A be an $m \times n$ matrix and let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$. If $A\mathbf{x} = \mathbf{b}$ has no solution, an approximation for $\mathbf{x}$ can be made by using least-squares. The goal is to minimize $\|A\mathbf{x} - \mathbf{b}\|$.

Steps for least squares:

1. Left multiply both sides of the equation by $A^\top$. (The resulting equation is called the **normal equation**.)

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

2. Solve for $\mathbf{x}$ in the normal equation.

This will give the closest solution for the original $\mathbf{x}$. The error is found by doing $\|A\mathbf{x} - \mathbf{b}\|$.

An alternative method using QR factorization.

Let A be an $m \times n$ matrix with linearly independent columns.

1. Factorize A using QR factorization.
2. Set up the equation

$$R\mathbf{x} = Q^\top \mathbf{b} \text{ (This is easily derived from the normal equation)}$$

3. Solve for $\mathbf{x}$ in the new equation.

Curve fitting

Let $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ be a set of points in $\mathbb{R}^2$ and let $F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \cdots + \alpha_k f_k(x)$ be the curve to be fitted to.

Steps to curve fitting:

1. Set up the following matrix. (The leftmost matrix is not necessarily square, and the goal is to solve for α)
2. Construct the normal equation.
3. Solve the normal equaton for $\alpha_1, \alpha_2, \ldots, \alpha_k$.

7.4 Constrained optimization

Let A be an $n \times n$ symmetric matrix with diagonalization $A = PDP^\top$, D arranged from greatest to least eigenvalue of A and P such that $P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$. Let $Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ and constrained by $\mathbf{x}^\top \mathbf{x} = 1$ (the unit sphere in $\mathbb{R}^n$)).

The maximum of Q is the greatest eigenvalue of A. The minimum of Q is the least eigenvalue of A. If Q is constrained also by $\mathbf{x}^\top \mathbf{u}_k, 1 \leq k \leq n$, then the maximum of Q is λ_{k+1}.

8 Theorems

Theorem 8.1: Rank-nullity

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

$$\operatorname{rank} A + \dim \operatorname{Nul} A = n.$$

Theorem 8.2: Invertible matrices

An $n \times n$ matrix is **invertible** if and only if any of the following is true.

-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
- $\dim \operatorname{Nul} A = 0$
- $\operatorname{Nul} A = \{\mathbf{0}\}$

- $\operatorname{rank} A = n$
- $\dim \operatorname{Col} A = n$
- $\operatorname{Col} A = \mathbb{R}^n$
- The columns of A form a basis for $\mathbb{R}^n$
- $\det A \neq 0$
- 0 is not an eigenvalue of A
- $(\operatorname{Col} A)^\perp = \{\mathbf{0}\}$
- $(\operatorname{Nul} A)^\perp = \mathbb{R}^n$
- $\operatorname{Row} A = \mathbb{R}^n$
- A has n nonzero singular values.

9 Matrix factorizations

Definition 9.1: QR factorization

Let A be an $m \times n$ matrix with linearly independent columns. Then

$$A = QR \qquad \text{where } Q \text{ is orthonormal and } R \text{ is upper triangular.}$$

Steps for QR factorization:

Let $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$.

1. Start with $\mathbf{v}_1$ and construct an orthonormal basis for $\operatorname{Col} A$ using Gram-Schmidt.

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \longrightarrow \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$$

2. Construct Q by putting each vector into the columns of Q.

$$Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

.

3. Solve for R by $R = Q^\top A$.

Putting it all together,

$$A = QR.$$

Definition 9.2: Diagonalization

Let A be an $n \times n$ matrix with n linearly independent eigenvectors. Then

$$A = PDP^{-1} \qquad \text{where } D \text{ is a diagonal matrix.}$$

Steps for diagonalization:

1. Find the eigenvalues λ_i and the corresponding eigenvectors $\mathbf{v}_i$ of A.
2. Construct D by placing the eigenvalues on the diagonal.

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

3. Construct P by placing the corresponding eigenvectors into the columns.

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$$

4. Calculate P^{-1}

Putting it all together,

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

Of note is that the columns of P form an **eigenvector basis** of $\mathbb{R}^n$.

A useful property of diagonalizable matrices is that

$$A^k = PD^kP^{-1} = P \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix} P^{-1}.$$

Theorem 9.1

Let A be an $n \times n$ matrix. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.

- If A has n distinct eigenvalues, then A is diagonalizable.
- If A does not have distinct eigenvalues, then the algebraic multiplicity of each λ_k must equal the dimension of the eigenspace corresponding to λ_k.

Definition 9.3: Complex diagonalization

Let A be an 2×2 matrix with complex eigenvalues – that is, $\lambda = a \pm b\mathrm{i}$ for some $a, b \in \mathbb{R}$. Then A can be written as

$$A = PCP^{-1}.$$

Steps for diagonalization:

1. Find the eigenvalues λ_i and the corresponding eigenvectors $\mathbf{v}_i$ of A.
2. Using the eignvalue of the form $\lambda = a - b\mathrm{i}$, construct C.
$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$
3. Construct P with the corresponding eigenvector.
$$P = \begin{bmatrix} \operatorname{Re} \mathbf{v} & \operatorname{Im} \mathbf{v} \end{bmatrix}$$
4. Calculate P^{-1}

Putting it all together,
$$A = PCP^{-1}$$

Definition 9.4: Orthonormal diagonalization

Let A be an $n \times n$ symmetric matrix. Then A can be factored into

$$A = PDP^\top \qquad \text{where } D \text{ is a diagonal matrix.}$$

Steps to orthonormal diagonalization:

1. Find the eigenvalues λ_i and the corresponding eigenvectors $\mathbf{v}_i$ of A.
2. Construct D by placing the eigenvalues on the diagonal.

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

3. Orthonormalize each of eigenvectors by $\mathbf{u}_i = \frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i$ and construct P by placing the corresponding eigenvectors into the columns.

$$P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

4. Construct $P^\top$.

Putting it all together,

$$A = PDP^\top.$$

Definition 9.5: Spectral decomposition

Let A be an $n \times n$ symmetric matrix with an orthonormal diagonalization

$$A = PDP^\top \qquad \text{where } P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

Then A can be written as

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^\top + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^\top + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^\top = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$$

Definition 9.6: Singular value decomposition

Steps for singular value decomposition:

1. Calculate $A^\top A$
2. Find the eigenvalues λ_i and the corresponding eigenvectors $\mathbf{v}_i$ of $A^\top A$.
3. Find the singular values by $\sigma_i = \sqrt{\lambda_i}$ and organize them such that $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r$.
4. Construct an $r \times r$ diagonal matrix D with the singular values.

$$D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix}$$

5. Construct the $m \times n$ matrix Σ by "cushioning" the above matrix with zeros to be the same size as A.

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

6. Normalize each $\mathbf{v}_i$ in the order of the corresponding eigenvalues and construct the matrix

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix}$$

7. Construct the columns of U by $\mathbf{u}_i = \dfrac{1}{\sigma_i} A\mathbf{v}_i$. If necessary, use the Gram-Schmidt process to orthogonalize the columns of U.

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

Putting it all together,

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

About the spaces of A.

$$U = \begin{bmatrix} \operatorname{Col} A & (\operatorname{Col} A)^\perp \end{bmatrix} \qquad V = \begin{bmatrix} \operatorname{Row} A & \operatorname{Nul} A \end{bmatrix}$$

Also, U spans $\mathbb{R}^m$ and is orthonormal.

U is $m \times m$.

The singular values of of A are related to V by $\sigma_i = \|A\mathbf{v}_i\|$

V is $n \times n$.

V is an orthonormal basis for $\mathbb{R}^n$