SF1684 Algebra and Geometry

Lecture 5

Basis and dimension, orthogonal projections and least squares, orthonormal bases, coordinate vectors and change of basis

Bao Nguyen

Outline

Outline

Basis

Basis of a subspace

Let V be a subspace of $\mathbb{R}^n$. A basis for V is a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in V such that

1. S spans V, and
2. S is linearly independent.

Existence of basis

If V is a nonzero subspace of $\mathbb{R}^n$, that is $V \neq \{\mathbf{0}\}$, then V has a basis and this basis has at most n vectors.

Bases

A subspace generally has infinitely many bases, but they all contain the same number of vectors.

Dimension

Dimension of a subspace

The number of vectors in a basis for V is called the dimension of V and is denoted by $\dim(V)$.

Warning

The zero subspace $\{\mathbf{0}\}$ cannot have a basis. We define $\dim(\{\mathbf{0}\})$ to be 0.

Linear dependence

A set of two or more nonzero vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly dependent if and only if some vector in S is a linear combination of its predecessors.

Determine by inspection whether the vectors are linearly independent

- $\mathbf{v}_1 = (1, 2, -2)$ and $\mathbf{v}_2 = (-2, -4, 4)$ are linearly dependent in $\mathbb{R}^3$, since $\mathbf{v}_2 = -2\mathbf{v}_1$.
- $\mathbf{v}_1 = (1, 2, -2)$, $\mathbf{v}_2 = (-2, -4, 4)$, and $\mathbf{v}_3 = (0, 0, 0)$ are linearly dependent in $\mathbb{R}^3$, since it contains the zero vector.
- $\mathbf{v}_1 = (0, 1, 0)$, $\mathbf{v}_2 = (2, 0, 2)$, and $\mathbf{v}_3 = (-3, 0, 3)$ are linearly independent in $\mathbb{R}^3$, since $\mathbf{v}_2$ is not a scalar multiple of $\mathbf{v}_1$, and $\mathbf{v}_3$ is not a linear combination of $\mathbf{v}_2$ and $\mathbf{v}_1$ (verify).

Basic problems

Given a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$. Let A be an $n \times k$ matrix has these vectors as columns and R a row echelon form of A.

Determining whether S spans $\mathbb{R}^n$

The following are equivalent

a. $\mathrm{span}(S) = \mathbb{R}^n$.
b. $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^n$.
c. R has a leading 1 in every row.

Determining whether S is linearly independent

The following are equivalent

a. S is linearly independent.
b. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.
c. R has a leading 1 in every column.

Corollary

1. If $k = n$, then S spans $\mathbb{R}^n$ if and only if S is linearly independent. In which case, S is a basis for $\mathbb{R}^n$.
2. If $k < n$, then S cannot span $\mathbb{R}^n$.
3. If $k > n$, then S is not linearly independent.

Examples

- $\mathbf{v}_1 = (2, 0, 0)$, $\mathbf{v}_2 = (0, -1, 0)$, and $\mathbf{v}_3 = (0, 0, 3)$ are linearly independent and hence form a basis for $\mathbb{R}^3$.
- $\mathbf{v}_1 = (0, 1)$, $\mathbf{v}_2 = (2, -3)$, and $\mathbf{v}_3 = (-1, 4)$ are linearly dependent in $\mathbb{R}^2$, since it contains 3 vectors in $\mathbb{R}^2$.

Finding a basis for the solution space of $A\mathbf{x} = \mathbf{0}$

1. Solve the system using Gauss-Jordan elimination.
2. If the solution is unique $\mathbf{x} = \mathbf{0}$ then the solution space is $\{\mathbf{0}\}$ and has dimension 0.
3. If the general solution has the vector form

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k$$

then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, called **canonical solutions**, are linearly independent and hence form a basis for the solution space $\rightarrow$ the solution space has dimension k. We call that basis the **canonical basis** for the solution space.

Recall that if $\mathbf{a}$ is a nonzero vector in $\mathbb{R}^n$, then the hyperplane $\mathbf{a}^\perp$ through $\mathbf{0}$ is the set $\mathbf{a}^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{a} = 0\}$.

Dimension of a hyperplane

The hyperplane can be viewed as the solution space of a linear system of one equation in n unknowns. Thus $\dim(\mathbf{a}^\perp) = n - 1$.

Outline

Properties of bases

Expression of vector in basis

If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a basis for a subspace V, then every vector in V can be expressed uniquely as a linear combination of the $\mathbf{v}_i$'s.

Two important facts about bases

Let S be a set of s vectors in a nonzero subspace V of $\mathbb{R}^n$

i. If S spans V, then S is either a basis or contains a basis for V; and thus $\dim(V) \leq s$.
ii. If S is linearly independent, then S is either a basis or can be extended to a basis for V; and thus $\dim(V) \geq s$.

Dimensions of subspaces

If V and W are subspaces of $\mathbb{R}^n$ and if V is contained in W, then

i. $0 \leq \dim(V) \leq \dim(W) \leq n$.
ii. $V = W$ if and only if $\dim(V) = \dim(W)$.

Spanning and linear independence

Let S be a set of s vectors in a subspace V with $\dim(V) = k$.

1. If $s = k$ and S is linearly independent, then S is a basis for V.
2. If $s = k$ and S spans V, then S is a basis for V.
3. If $s < k$, then S cannot span V.
4. If $s > k$, then S is not linearly independent.

Fundamental theorem of invertible matrices (cont.)

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

- d $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- g The column vectors of A are linearly independent.
- h The row vectors of A are linearly independent.
- m The column vectors of A span $\mathbb{R}^n$.
- n The row vectors of A span $\mathbb{R}^n$.
- o The column vectors of A form a basis for $\mathbb{R}^n$.
- p The row vectors of A form a basis for $\mathbb{R}^n$.

Outline

The fundamental spaces of a matrix

There are four fundamental spaces associated to an $m \times n$ matrix A

1. $\mathrm{row}(A)$: the **row space of A** is the span of the m row vectors of A.
2. $\mathrm{col}(A)$: the **column space of A** is the span of the n column vectors of A.
3. $\mathrm{null}(A)$: the **null space of A** is the solution space of $A\mathbf{x} = \mathbf{0}$.
4. $\mathrm{null}(A^T)$: the **null space of A^T** is the solution space of $A^T\mathbf{x} = \mathbf{0}$.

Rank and nullity

The dimension of the row space of A is called the **rank** of A and is denoted by $\mathrm{rank}(A)$. The dimension of the null space of A is called the **nullity** of A and is denoted by $\mathrm{nullity}(A)$.

Since $\mathrm{col}(A) = \mathrm{row}(A^T)$, the dimension of the column space of A is the rank of A^T.

Fundamental spaces of the matrix A

1. $\mathrm{row}(A) = \left\{A^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\right\}$ is a subspace of $\mathbb{R}^n$ and $\mathrm{rank}(A) \leq \min\{m, n\}$.
2. $\mathrm{col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$ is a subspace of $\mathbb{R}^m$ and $\mathrm{rank}(A^T) \leq \min\{m, n\}$.
3. $\mathrm{null}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$ is a subspace of $\mathbb{R}^n$ and $\mathrm{nullity}(A) \leq n$.
4. $\mathrm{null}(A^T) = \left\{\mathbf{x} \in \mathbb{R}^m \mid A^T\mathbf{x} = \mathbf{0}\right\}$ is a subspace of $\mathbb{R}^m$ and $\mathrm{nullity}(A^T) \leq m$.

Orthogonal complements

Orthogonal complement

If S is a nonempty set in $\mathbb{R}^n$, then the set of all vectors in $\mathbb{R}^n$ that are orthogonal to every vector in S is a subspace of $\mathbb{R}^n$. It is called the orthogonal complement of S and is denoted by $S^\perp$.

Properties of orthogonal complements

Let S be a nonempty set in $\mathbb{R}^n$ and W a subspace of $\mathbb{R}^n$.

- i $W^\perp \cap W = \{\mathbf{0}\}$.
- ii $S^\perp = \mathrm{span}(S)^\perp$.
- iii $(W^\perp)^\perp = W$; W and $W^\perp$ are orthogonal complements of one another.
- iv $(S^\perp)^\perp = \mathrm{span}(S)$.

Example

A line L through $\mathbf{0}$ of $\mathbb{R}^3$ and the plane through $\mathbf{0}$ that is perpendicular to L are orthogonal complements of one another.

Orthogonal complements of $\mathrm{row}(A)$ and $\mathrm{col}(A)$

Let A be an $m \times n$ matrix. The row space and the null space of A are orthogonal complements. Similarly, the column space of A and the null space of A^T are orthogonal complements. In conclusion,

$$\mathrm{row}(A)^\perp = \mathrm{null}(A), \qquad \mathrm{null}(A)^\perp = \mathrm{row}(A),$$
$$\mathrm{col}(A)^\perp = \mathrm{null}(A^T), \quad \mathrm{null}(A^T)^\perp = \mathrm{col}(A).$$

Effect of elementary row operations on the fundamental spaces of a matrix

Effect of elementary row operations

- i) Elementary row operations **do not** change the row space or the null space of a matrix.
- ii) Elementary row operations **do** change the column space of a matrix, but do not change the linear independence or dependence relations between the column vectors.

Some applications

Let A be an $m \times n$ matrix and R a row echelon form of A. We denote by $S = \{\mathbf{r}_1(R), \mathbf{r}_2(R), \ldots, \mathbf{r}_k(R)\}$ the set of successive nonzero row vectors in R and $T = \{\mathbf{c}_{j_1}(R), \mathbf{c}_{j_2}(R), \ldots, \mathbf{c}_{j_k}(R)\}$ the set of column vectors (may be not successive) in R that have the leading 1's. Note that we have the same number of vectors in the two sets.

Basis for $\mathrm{row}(A)$

S is linearly independent (verify) and hence forms a basis for $\mathrm{row}(R)$. Since $\mathrm{row}(A) = \mathrm{row}(R)$, the vectors $\mathbf{r}_1(R), \mathbf{r}_2(R), \ldots, \mathbf{r}_k(R)$ is also a basis for $\mathrm{row}(A)$.

Basis for $\mathrm{col}(A)$

T is linearly independent and any column vector $\mathbf{c}_j(R) \notin T$ is a linear combination of the vectors in T (verify). Since the corresponding column vectors from A satisfy the same linear independence or dependence relations, the vectors $\mathbf{c}_{j_1}(A), \mathbf{c}_{j_2}(A), \ldots, \mathbf{c}_{j_k}(A)$ are also linearly independent and form a basis for $\mathrm{col}(A)$. This is called the **Pivot Theorem**; see in the next slides.

Finding a basis for a span

Given a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$, find a basis for the subspace $W = \mathrm{span}(S)$. Find also a basis for $W^\perp$.

Solution

We start by forming a $k \times n$ matrix A that has $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ as row vectors, then W is the row space of A. A basis for W is the nonzero rows in a row echelon form of A.

Since $W^\perp$ is the null space of A, we wish to find a basis for the solution space of the linear system $A\mathbf{x} = \mathbf{0}$. We will use the canonical basis produced by Gauss-Jordan elimination.

Example

See Example 4 on page 346 of the textbook.

Determining whether a vector is in a subspace

Find conditions in which a vector $\mathbf{b}$ in $\mathbb{R}^n$ will lie in $W = \mathrm{span}(S)$.

Solution 1

We start by forming an $n \times k$ matrix C that has $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ as column vectors, then W is the column space of C. Thus $\mathbf{b}$ lies in W if and only if the linear system $C\mathbf{x} = \mathbf{b}$ is consistent.

Solution 2

We form a $k \times n$ matrix A that has $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ as row vectors and a matrix $A_{\mathbf{b}}$ by adding $\mathbf{b}$ to A as an additional row vector. Our problem reduces to finding conditions on $\mathbf{b}$ under which A have the same rank as $A_{\mathbf{b}}$.

Solution 3

Observe that $\mathbf{b}$ lies in W if and only if $\mathbf{b}$ is orthogonal to every vector in $W^{\perp}$ which is the null space of the matrix A in Solution 2. We find a basis for $\mathrm{null}(A)$ and then determine conditions in which $\mathbf{b}$ is orthogonal to that basis.

Example

See Example 6 on page 347 of the textbook.

Outline

The dimension theorem and its application

Consider a linear system $A\mathbf{x} = \mathbf{0}$ of m equations in n variables (A has size $m \times n$). The number of free variables is the same as the dimension of the solution space and the number of leading variables is the same as the dimension of the row space of A. The above observations, together with

$$\text{number of free variables} + \text{number of leading variables} = n \text{ variables},$$

imply that $\operatorname{nullity}(A) + \operatorname{rank}(A) = n$.

Dimension theorem for matrices

If A is an $m \times n$ matrix, then $\operatorname{rank}(A) + \operatorname{nullity}(A) = n$.

Corollary

Since A^T has m columns, we have $\operatorname{rank}(A^T) + \operatorname{nullity}(A^T) = m$.

Corollary

If A has rank k, then

- i A has nullity $n - k$.
- ii Every row echelon form of A has k nonzero rows and $m - k$ zero rows.
- iii The homogeneous system $A\mathbf{x} = \mathbf{0}$ has k pivot (leading) variables and $n - k$ free variables.

Fundamental theorem of invertible matrices (cont.)

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

- d $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- q $\mathrm{rank}(A) = n$.
- r $\mathrm{nullity}(A) = 0$.

Extending a linearly independent set to a basis

Given a linearly independent set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$, extend S to a basis for $\mathbb{R}^n$.

Solution

1. Form a matrix A that has $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ as row vectors.
2. Find a basis for the null space of A. This basis has $n - k$ vectors, say $\mathbf{w}_{k+1}, \mathbf{w}_{k+2}, \ldots, \mathbf{w}_n$.
3. The set of n vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{w}_{k+1}, \mathbf{w}_{k+2}, \ldots, \mathbf{w}_n\}$ is linearly independent and hence forms a basis for $\mathbb{R}^n$.

Example

See Example 2 on page 353 of the textbook.

Dimension theorem for subspaces

If W is a subspace of $\mathbb{R}^n$, then $\dim(W) + \dim(W^\perp) = n$.

Corollary

If $\dim(W) = n - 1$, then $\dim(W^\perp) = 1$. It implies that $W^\perp$ is a line through the origin of $\mathbb{R}^n$ and its orthogonal complement, the subspace W, is a hyperplane through the origin of $\mathbb{R}^n$.

Rank 1 matrix

If $\mathbf{u}$ is a nonzero $m \times 1$ matrix and $\mathbf{v}$ is a nonzero $n \times 1$ (column vectors), then the outer product $A = \mathbf{u}\mathbf{v}^T$ has rank 1. Conversely, if A has rank 1, then A can be factored into a product of the above form.

Outline

The rank theorem and its applications

Rank theorem

The row space and the column space of a matrix have the same dimension. More precisely, if A is an $m \times n$ matrix, then

$$\begin{aligned} \operatorname{rank}(A) &= k, \quad & \operatorname{nullity}(A) &= n - k, \\ \operatorname{rank}(A^T) &= k, \quad & \operatorname{nullity}(A^T) &= m - k. \end{aligned}$$

Consistency theorem

If A is an $m \times n$ matrix and $\mathbf{b}$ in $\mathbb{R}^m$, then the following statements are equivalent

a. $A\mathbf{x} = \mathbf{b}$ is consistent.
b. $\mathbf{b}$ is in the column space of A.
c. The augmented matrix $[\,A \mid \mathbf{b}\,]$ has the same rank as A.

Definition

A matrix is said to have **full column rank** if its column vectors are linearly independent, and **full row rank** if its row vectors are linearly independent.

Remark

A matrix A has full column rank if and only if A^T has full row rank.

Full column rank and full row rank

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

i. A has full column rank if and only if $\mathrm{rank}(A) = n$.
ii. A has full row rank if and only if $\mathrm{rank}(A) = m$.

Unifying theorem

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

- a $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- b $A\mathbf{x} = \mathbf{b}$ has at most one solution for every $\mathbf{b} \in \mathbb{R}^m$.
- c A has full column rank.
- d The $n \times n$ matrix $A^T A$ is invertible.

Example

See Example 7 on page 366 of the textbook.

Outline

The pivot theorem and its applications

Pivot columns

The columns of a matrix A that contain the leading 1's in the row echelon forms of A are called the pivot columns of A.

Pivot theorem

The pivot columns of A form a basis for the column space of A.

Examples

See Examples 1, 2 on pages 371–373 of the textbook.

Finding a basis for a span

Let W be a subspace of $\mathbb{R}^n$ that is spanned by a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$. Find a basis for W consisting of vectors from S.

Solution

1. Form a matrix A that has $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ as column vectors.
2. Reduce A to a row echelon form U. A basis for W given by the pivot columns of A, identified using U.

Express the vectors in S that are not in the basis as linear combinations of the basis vectors

Solution

3. Continue reducing U to the reduced row echelon form R of A.
4. By inspection, express each vector of R that does not contain a leading 1 as a linear combination of preceding columns that contain leading 1's. Then those same linear combinations will apply to the corresponding columns of A.

Bases for the fundamental spaces of a matrix

Finding bases for the four fundamental spaces of a matrix A

All four bases can be found using a single row reduction procedure. Let U be a row echelon form of A and let R be the reduced row echelon form. Then bases are given by the following vectors:

- I $\mathrm{row}(A)$: the nonzero rows of U or R.
- II $\mathrm{col}(A)$: the pivot columns of A, identified using U or R.
- III $\mathrm{null}(A)$: the canonical solutions of $A\mathbf{x} = \mathbf{0}$; and these are readily obtained from the system $R\mathbf{x} = \mathbf{0}$.
- IV $\mathrm{null}(A^T)$: form an $m \times (n+m)$ matrix with left half A and right half the identity matrix I_m. Reduce A to R applying the same operations to the whole matrix. A basis given by rows of resulting right half matrix which are beside the zero rows of left half R.

Example

See Example 3 on page 374 of the textbook.

Outline

The projection theorem and its applications

Orthogonal projection onto a line

If $\mathbf{a} \in \mathbb{R}^n$ is a nonzero vector, then every $\mathbf{x} \in \mathbb{R}^n$ can be expressed in exactly one way as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1$ is parallel to $\mathbf{a}$ and $\mathbf{x}_2$ is orthogonal to $\mathbf{a}$. We have

$$\mathbf{x}_1 = \operatorname{proj}_{\mathbf{a}} \mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} \quad \text{and} \quad \mathbf{x}_2 = \mathbf{x} - \mathbf{x}_1 = \mathbf{x} - \frac{\mathbf{x} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a},$$

where $\operatorname{proj}_{\mathbf{a}} \mathbf{x}$ is called the **orthogonal projection of $\mathbf{x}$ onto $\operatorname{span}\{\mathbf{a}\}$**.

Projection operator on $\mathbb{R}^n$

Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be an operator defined by $T(\mathbf{x}) = \operatorname{proj}_{\mathbf{a}} \mathbf{x}$. Then T is a linear operator and its standard matrix is given by

$$P = \frac{1}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T, \quad \text{with } \mathbf{a} \text{ is in column form.}$$

This matrix is symmetric ($P = P^T$) and idempotent ($P^2 = P$) and has rank 1.

Examples

See Examples 1–5 on pages 380–383 of the textbook.

Projecton theorem for subspaces

If W is a subspace of $\mathbb{R}^n$, then every $\mathbf{x} \in \mathbb{R}^n$ can be expressed in exactly one way as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1$ is in W and $\mathbf{x}_2$ is in $W^\perp$.

$\mathbf{x}_1$ is the **orthogonal projection of $\mathbf{x}$ on W** and $\mathbf{x}_2$ is the **orthogonal projection of $\mathbf{x}$ on $W^\perp$**. Thus, we can write

$$\mathbf{x} = \operatorname{proj}_W \mathbf{x} + \operatorname{proj}_{W^\perp} \mathbf{x}.$$

Orthogonal projection onto W

If M is any matrix whose column vectors form a basis for W, then $M^T M$ is invertible and

$$\operatorname{proj}_W \mathbf{x} = M(M^T M)^{-1} M^T \mathbf{x}.$$

If we define the **orthogonal projection of $\mathbb{R}^n$ onto W**: $T(\mathbf{x}) = \operatorname{proj}_W \mathbf{x}$, then its standard matrix is

$$P = M(M^T M)^{-1} M^T.$$

This matrix is symmetric and idempotent and has rank equal to the dimension of W.

Orthogonal projection onto $W^\perp$

The standard matrix for $\mathrm{proj}_{W^\perp}\,\mathbf{x}$ can be expressed in terms of the standard matrix P for $\mathrm{proj}_W\,\mathbf{x}$ as

$$I - P = I - M(M^T M)^{-1} M^T.$$

Orthogonal projection matrix

If P is an $n \times n$ symmetric, idempotent matrix, then $T_P(\mathbf{x}) = P\mathbf{x}$ is the orthogonal projection onto the column space of P. Moreover, since P is idempotent, the dimension of the column space is equal to the trace of P.

Examples

See Examples 6–8 on pages 385–386 of the textbook.

Outline

Best approximation

Best approximation theorem

If W is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a point in $\mathbb{R}^n$, then $\widehat{\mathbf{w}} = \operatorname{proj}_W \mathbf{b}$ is the unique best approximation to $\mathbf{b}$ from W.
That is, for any other $\mathbf{w}$ in W, $\|\mathbf{b} - \widehat{\mathbf{w}}\| < \|\mathbf{b} - \mathbf{w}\|$.

Distance from point to subspace

The distance from a point $\mathbf{b}$ to W is defined to be

$$d = \|\mathbf{b} - \operatorname{proj}_W \mathbf{b}\| = \|\operatorname{proj}_{W^\perp} \mathbf{b}\| .$$

Distance from a point to a hyperplane

The hyperplane $W = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{a} = 0\}$ has its orthogonal complement $W^\perp = \operatorname{span}\{\mathbf{a}\}$. Thus, we have

$$d = \|\operatorname{proj}_{W^\perp} \mathbf{b}\| = \frac{|\mathbf{a} \cdot \mathbf{b}|}{\|\mathbf{a}\|} = \frac{|a_1 b_1 + a_2 b_2 + \cdots + a_n b_n|}{\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}}.$$

Least squares

Let A be an $m \times n$ matrix and $\mathbf{b}$ a vector in $\mathbb{R}^m$, not necessarily in the column space of A. Since $A\mathbf{x} = \mathbf{b}$ might be inconsistent, we wish to find a vector $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that $A\widehat{\mathbf{x}}$ is a best approximation $\mathbf{b}$.

Definition

A vector $\widehat{\mathbf{x}} \in \mathbb{R}^n$ is called a **least squares solution** of $A\mathbf{x} = \mathbf{b}$ if it minimizes the error $\|\mathbf{b} - A\mathbf{x}\|$. The vector $\mathbf{b} - A\widehat{\mathbf{x}}$ is called the **least squares error vector** and the scalar $\|\mathbf{b} - A\widehat{\mathbf{x}}\|$ is called the **least squares error**.

By the Best Approximation Theorem, $A\widehat{\mathbf{x}}$ is the projection of $\mathbf{b}$ onto $\mathrm{col}(A)$.

Least squares solution of linear system

- i The least squares solutions are the solutions of the equation $A\mathbf{x} = \mathrm{proj}_{\mathrm{col}(A)}\, \mathbf{b}$. This system is certainly consistent, since $\mathrm{proj}_{\mathrm{col}(A)}\, \mathbf{b}$ is in the column space of A.
- ii Every least squares solution $\widehat{\mathbf{x}}$ has the same error vector, namely $\mathbf{b} - A\widehat{\mathbf{x}} = \mathrm{proj}_{\mathrm{null}(A^T)}\, \mathbf{b}$.

Normal equation

The system $A^T A\mathbf{x} = A^T\mathbf{b}$ is called the normal equation or normal system associated with $A\mathbf{x} = \mathbf{b}$.

Finding least squares solution

The least squares solutions of $A\mathbf{x} = \mathbf{b}$ are the exact solutions of the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$.

1. If A has full column rank (that is, the columns are linearly independent), then $A^T A$ is invertible and the unique solution of the normal equation is

$$\widehat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}.$$

2. If A does not has full column rank, then the normal equation has infinitely many solutions.

Examples

See Examples 3–4 on pages 386–387 of the textbook.

Fitting a curve to data

Given data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that are supposed to be related by a **linear equation** $y = ax + b$ (called **linear regression model**).
We have

$$M\mathbf{v} = \mathbf{y} \quad \text{where} \quad M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \; \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}, \; \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

If the data do not exactly lie on a line, then the linear system will be inconsistent. In this case we look for a least squares approximation to a and b by solving the normal system

$$M^T M\mathbf{v} = M^T\mathbf{y} \quad \text{or alternatively,} \quad \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}.$$

The line $y = ax + b$ is called the **least squares line of best fit** to the data (or the **regression line**). Note that $M^T M$ is invertible unless x_i's are all the same.

This technique can be generalized to fitting a polynomial to a set of data points. Suppose that we want to find a **polynomial of degree m** with coefficients $a_0, a_1, \ldots, a_m$ that is a best fit (in a least squares sense) for the n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. We have

$$M\mathbf{v} = \mathbf{y} \quad \text{where} \quad M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}, \; \mathbf{v} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}, \; \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

In the special case where $m = n - 1$ and the x_i's are distinct, the linear system has unique solution. If $m < n - 1$, the system will usually be inconsistent, so we solve for a_i's using the normal system

$$M^T M\mathbf{v} = M^T\mathbf{y}.$$

If at least $m + 1$ of x_i's are distinct, then M has full column rank. Thus, $M^T M$ is invertible and the unique solution of the normal equation is

$$\mathbf{v} = (M^T M)^{-1} M^T \mathbf{y}.$$

There are three important models in application

1. Exponential model: $y = ae^{bx}$.
2. Power function model: $y = ax^b$.
3. Logarithmic model: $y = a + b\ln(x)$.

The data can be transformed to a linear form in which a linear regression can be used to approximated the constants a and b. For example, if we take the natural log of both sides of the equation $y = ae^{bx}$, then we have the equivalent equation $\ln(y) = \ln(a) + bx$. This expresses $\ln(y)$ as a linear function of x and hence we can use the least squares line of best fit to the transformed data points $(x_1, \ln y_1), (x_2, \ln y_2), \ldots, (x_n, \ln y_n)$ to estimate $\ln(a)$ and b, and then computing a from $\ln(a)$.

Examples

See Examples 5–7 on pages 400–403 and Exercise T5–T7 on page 406 of the textbook.

Outline

Orthogonal and orthonormal bases

Definition

$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an **orthogonal basis** for a subspace V of $\mathbb{R}^n$ if

1. S is a basis for V, and
2. S is orthogonal, that is $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for all $i \neq j$.

S is an **orthonormal basis** if it is an orthogonal basis, and in addition

3. Each vector in S has length 1, that is $\|\mathbf{v}_i\| = 1$ for all $1 \leq i \leq k$.

Orthogonality and linear independence

An orthogonal set of nonzero vectors in $\mathbb{R}^n$ is linearly independent.

Existence of orthonormal basis

Every nonzero subspace of $\mathbb{R}^n$ has an orthonormal basis.

Examples

See Examples 1–3 on page 407 of the textbook.

Finding orthogonal and orthogonal bases

Gram-Schmidt process

This process is an algorithm for finding an orthonormal basis for any nonzero subspace W. Start with any basis $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ for W. Let

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{w}_1 \\
\mathbf{v}_2 &= \mathbf{w}_2 - \frac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_k &= \mathbf{w}_k - \frac{\mathbf{w}_k \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \cdots - \frac{\mathbf{w}_k \cdot \mathbf{v}_{k-1}}{\|\mathbf{v}_{k-1}\|^2}\mathbf{v}_{k-1}.
\end{aligned}$$

The set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W.

To obtain an orthonormal basis, we divide each vector $\mathbf{v}_i$ by its length. Thus, if we set $\mathbf{q}_i = \dfrac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$, then $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k\}$ is an orthonormal basis for W.

Examples

See Examples 9, 10 on pages 412–413 of the textbook.

Orthogonal projection using orthornormal basis

Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a basis for a subspace W of $\mathbb{R}^n$. We consider the orthogonal projection of a vector $\mathbf{x} \in \mathbb{R}^n$ onto W.

Orthogonal projection using orthornormal basis

i. If the basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is orthogonal, then

$$\mathrm{proj}_W\, \mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\mathbf{x} \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\mathbf{x} \cdot \mathbf{v}_k}{\|\mathbf{v}_k\|^2}\mathbf{v}_k.$$

ii. If the basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is orthonormal, then

$$\mathrm{proj}_W\, \mathbf{x} = (\mathbf{x} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{x} \cdot \mathbf{v}_2)\mathbf{v}_2 + \cdots + (\mathbf{x} \cdot \mathbf{v}_k)\mathbf{v}_k.$$

Examples

See Examples 5, 6 on page 409 of the textbook.

Linear combination of orthogonal basis

If $\mathbf{w}$ is a vector in W, then $\mathrm{proj}_W\,\mathbf{w} = \mathbf{w}$. Thus, we have

Linear combination of orthogonal basis

i. If the basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is orthogonal, then

$$\mathbf{w} = \frac{\mathbf{w} \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\mathbf{w} \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\mathbf{w} \cdot \mathbf{v}_k}{\|\mathbf{v}_k\|^2}\mathbf{v}_k.$$

ii. If the basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is orthonormal, then

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{w} \cdot \mathbf{v}_2)\mathbf{v}_2 + \cdots + (\mathbf{w} \cdot \mathbf{v}_k)\mathbf{v}_k.$$

Example

See Example 8 on page 411 of the textbook.

Outline

Coordinates with respect to a basis

Coordinate vector

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be an ordered basis for a subspace W of $\mathbb{R}^n$ and let w be in W. If $w = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_k\mathbf{v}_k$, then we call $a_1, a_2, \ldots, a_k$ the **ordered coordinates of $\mathbf{w}$ with respect to B**. The ordered k-tuple of coordinates $(\mathbf{w})_B = (a_1, a_2, \ldots, a_k)$ is called the **coordinate vector** for $\mathbf{w}$ with respect to B, and

$$[\mathbf{w}]_B = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}$$

is called the **coordinate matrix** for $\mathbf{w}$ with respect to B.

Coordinates with respect to an orthonormal basis

If $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthonormal basis for a subspace W, and if $\mathbf{w} \in W$ then the coordinate vector of $\mathbf{w}$ with respect to B is

$$(\mathbf{w})_B = \big((\mathbf{w} \cdot \mathbf{v}_1), (\mathbf{w} \cdot \mathbf{v}_2), \ldots, (\mathbf{w} \cdot \mathbf{v}_k)\big).$$

Computing with coordinates

Let B be an orthonormal basis for a k-dimensional subspace W of $\mathbb{R}^n$. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in W with coordinate vectors $(\mathbf{u})_B = (u_1, u_2, \ldots, u_k)$ and $(\mathbf{v})_B = (v_1, v_2, \ldots, v_k)$, then

1. $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_k^2} = \|(\mathbf{u})_B\|$.
2. $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_k v_k = (\mathbf{u})_B \cdot (\mathbf{v})_B$.

Change of basis

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $B' = \{\mathbf{v}_1', \mathbf{v}_2', \ldots, \mathbf{v}_n'\}$ be bases for $\mathbb{R}^n$. If $\mathbf{w} \in \mathbb{R}^n$ then the coordinate matrices of $\mathbf{w}$ with respect to the two bases are related by the equation

$$[\mathbf{w}]_{B'} = P_{B \to B'} [\mathbf{w}]_B,$$

where

$$P_{B \to B'} = \begin{bmatrix} [\mathbf{v}_1]_{B'} & [\mathbf{v}_2]_{B'} & \cdots & [\mathbf{v}_n]_{B'} \end{bmatrix}$$

is the **transition matrix** (or the **change of coordinates matrix**) from B to B'. In the case $B' = S$ is the standard basis for $\mathbb{R}^n$, then $P_{B \to S} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$.

Invertibility of transition matrix

The transition matrices $P_{B\to B'}$ and $P_{B'\to B}$ are invertible and inverse of one another, that is
$$P_{B\to B'} = (P_{B'\to B})^{-1}.$$

Procedure for computing $P_{B\to B'}$

1. Form the matrix $[\,B' \mid B\,]$.
2. Reduce the matrix above to reduced row echelon form using Gauss-Jordan elimination.
3. The resulting matrix will be $[\,I \mid P_{B\to B'}\,]$. Just extract the matrix $P_{B\to B'}$.

Examples

See Examples 1–7 on pages 429–435 of the textbook.

Coordinate map

Let B be a basis for $\mathbb{R}^n$. The transformation $\mathbf{x} \to (\mathbf{x})_B$ (or $\mathbf{x} \to [\mathbf{x}]_B$), called the **coordinate map** for B, is a one-to-one (and hence also onto) linear operator on $\mathbb{R}^n$. Moreover, if B is an orthonormal basis for $\mathbb{R}^n$, then the coordinate map is an orthogonal operator.

Transition between orthonormal bases

If B and B' are orthonormal bases for $\mathbb{R}^n$, then the transition matrices $P_{B \to B'}$ and $P_{B' \to B}$ are orthogonal.

Invertible matrix as transition matrix

If $P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix}$ is an invertible $n \times n$ matrix then P is the transition matrix from the basis $B = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ to the standard basis $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$.

Examples

See Examples 8, 9 on pages 436–437 of the textbook.