Vv214 Linear Algebra

Mid2

Li Yuzhou

SJTU-UM Joint Institute
Shanghai Jiao-Tong University

July 9, 2019

Table of Contents

Table of Contents

Definition

Coordinates in a subspace of $\mathbb{R}^n$

Consider a basis $\mathfrak{B} = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$ of a subspace V of $\mathbb{R}^n$. By Theorem 3.2.10, any vector $\vec{x}$ in V can be written uniquely as

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m.$$

The scalars $c_1, c_2, \ldots, c_m$ are called the $\mathfrak{B}$-*coordinates* of $\vec{x}$, and the vector

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$$

is the $\mathfrak{B}$-*coordinate vector* of $\vec{x}$, denoted by $[\vec{x}]_{\mathfrak{B}}$. Thus

$$[\vec{x}]_{\mathfrak{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} \quad \text{means that } \vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m.$$

Note that

$$\vec{x} = S\,[\vec{x}]_{\mathfrak{B}}, \quad \text{where } S = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \end{bmatrix}, \text{ an } n \times m \text{ matrix.}$$

Relation between them

$$\begin{tikzcd}
\vec{x} \arrow[r, "A"] & T(\vec{x}) \\
{[\vec{x}]_{\mathfrak{B}}} \arrow[u, "S"] \arrow[r, "B"] & {[T(\vec{x})]_{\mathfrak{B}}} \arrow[u, "S"]
\end{tikzcd}$$

Consider a linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^n$ and a basis $\mathfrak{B} = (\vec{v}_1, \ldots, \vec{v}_n)$ of $\mathbb{R}^n$. Let B be the $\mathfrak{B}$-matrix of T, and let A be the standard matrix of T (such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$ in $\mathbb{R}^n$). Then

$$AS = SB, \quad B = S^{-1}AS, \quad \text{and} \quad A = SBS^{-1}, \quad \text{where} \quad S = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}. \blacksquare$$

Change of basis

Consider two bases $\mathfrak{A}$ and $\mathfrak{B}$ of an n-dimensional linear space V. Consider the linear transformation $L_{\mathfrak{A}} \circ L_{\mathfrak{B}}^{-1}$ from $\mathbb{R}^n$ to $\mathbb{R}^n$, with standard matrix S, meaning that $S\vec{x} = L_{\mathfrak{A}}\left(L_{\mathfrak{B}}^{-1}(\vec{x})\right)$ for all $\vec{x}$ in $\mathbb{R}^n$. This invertible matrix S is called the *change of basis matrix* from $\mathfrak{B}$ to $\mathfrak{A}$, sometimes denoted by $S_{\mathfrak{B}\to\mathfrak{A}}$. See the accompanying diagrams. Letting $f = L_{\mathfrak{B}}^{-1}(\vec{x})$ and $\vec{x} = [f]_{\mathfrak{B}}$, we find that

$$[f]_{\mathfrak{A}} = S\,[f]_{\mathfrak{B}}, \quad \text{for all } f \text{ in } V.$$

If $\mathfrak{B} = (b_1, \ldots, b_i, \ldots, b_n)$, then

$$[b_i]_{\mathfrak{A}} = S\,[b_i]_{\mathfrak{B}} = S\vec{e}_i = (i\text{th column of } S),$$

so that

$$S_{\mathfrak{B}\to\mathfrak{A}} = \begin{bmatrix} [b_1]_{\mathfrak{A}} & \qquad & [b_n]_{\mathfrak{A}} \end{bmatrix}$$

$$\begin{tikzcd} & \mathbb{R}^n \\ V \arrow[ur, "L_{\mathfrak{A}}"] \arrow[dr, "L_{\mathfrak{B}}"'] & \\ & \mathbb{R}^n \arrow[uu, "S"'] \end{tikzcd} \qquad \begin{tikzcd} & {[f]_{\mathfrak{A}}} \\ f \arrow[ur, "L_{\mathfrak{A}}"] \arrow[dr, "L_{\mathfrak{B}}"'] & \\ & {[f]_{\mathfrak{B}}} \arrow[uu, "S"'] \end{tikzcd}$$

Exercise

Let V be the subspace of C^∞ spanned by the functions e^x and e^{-x}, with the bases $\mathfrak{A} = (e^x, e^{-x})$ and $\mathfrak{B} = (e^x + e^{-x}, e^x - e^{-x})$. Find the change of basis matrix $S_{\mathfrak{B}\to\mathfrak{A}}$.

Answer

$$S = \begin{bmatrix} [e^x + e^{-x}]_{\mathfrak{A}} & [e^x - e^{-x}]_{\mathfrak{A}} \end{bmatrix}.$$

$$e^x + e^{-x} = 1 \cdot e^x + 1 \cdot e^{-x} \qquad e^x - e^{-x} = 1 \cdot e^x + (-1) \cdot e^{-x}$$

$$[e^x + e^{-x}]_{\mathfrak{A}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad [e^x - e^{-x}]_{\mathfrak{A}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$S_{\mathfrak{B} \to \mathfrak{A}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

What you should know about this part

1. What function S,B,A has?
2. How to find them?
3. If x is not a simple vector, instead it is a function, a matrix, a complex number...Just handle it in the same way!
4. Be clear what the question demands you to get.

Table of Contents

Isomorphism

An invertible linear transformation T is called an *isomorphism*. We say that the linear space V is isomorphic to the linear space W if there exists an isomorphism T from V to W.

Properties

a. A linear transformation T from V to W is an isomorphism if (and only if) $\ker(T) = \{0\}$ and $\operatorname{im}(T) = W$.

In parts (b) through (d), the linear spaces V and W are assumed to be finite dimensional.

b. If V is isomorphic to W, then $\dim(V) = \dim(W)$.

c. Suppose T is a linear transformation from V to W with $\ker(T) = \{0\}$. If $\dim(V) = \dim(W)$, then T is an isomorphism.

d. Suppose T is a linear transformation from V to W with $\operatorname{im}(T) = W$. If $\dim(V) = \dim(W)$, then T is an isomorphism.

To prove a transformation is a isomorphism, first prove it is linear!

What you should know about this part

1. Definition of isomorphism
2. Prove it via properties(Im, Ker, Dim...)

Table of Contents

Inner Product

An *inner product* in a linear space V is a rule that assigns a real scalar (denoted by $\langle f, g \rangle$) to any pair f, g of elements of V, such that the following properties hold for all f, g, h in V, and all c in $\mathbb{R}$:

a. $\langle f, g \rangle = \langle g, f \rangle$ (symmetry)

b. $\langle f + h, g \rangle = \langle f, g \rangle + \langle h, g \rangle$

c. $\langle cf, g \rangle = c\langle f, g \rangle$

d. $\langle f, f \rangle > 0$, for all nonzero f in V (positive definiteness)

A linear space endowed with an inner product is called an *inner product space*.

Properties b and c express the fact that $T(f) = \langle f, g \rangle$ is a linear transformation from V to $\mathbb{R}$, for a fixed g in V.

Compare these rules with those for the dot product in $\mathbb{R}^n$, listed in the Appendix, Theorem A.5. Roughly speaking, an inner product space behaves like $\mathbb{R}^n$ as far as addition, scalar multiplication, and the dot product are concerned.

Norm and orthogonality

The *norm* (or magnitude) of an element f of an inner product space is

$$\|f\| = \sqrt{\langle f, f\rangle}.$$

Two elements f and g of an inner product space are called *orthogonal* (or perpendicular) if

$$\langle f, g\rangle = 0.$$

We can define the *distance* between two elements of an inner product space as the norm of their difference:

$$\text{dist}(f, g) = \|f - g\|.$$

What you should know about this part

1. A redefined inner product changes the definition of everything including norm(magnitude), orthogonality, distance...

Exercise

Consider a linear space $P_2(\mathbb{R})$.

a. Prove that $(f_1, f_2) = f_1(-1)f_2(-1) + f_1(0)f_2(0) + f_1(1)f_2(1)$ satisfies all the requirements for an inner product.

b. Take any nonzero element $f \in P_2(\mathbb{R})$ and calculate $||f||$.

c. Find a pair of orthonormal elements $u_1(t), u_2(t) \in P_2(\mathbb{R})$. Calculate the distance between $u_1(t)$ and $u_2(t)$.

d. Let $V = span\ (u_1(t), u_2(t))$. Describe $V^\perp$, calculate $\dim V^\perp$ and find any nonzero element in $V^\perp$.

e. Find an element $f \in P_2(\mathbb{R})$ such that $f \notin V \cup V^\perp$ and calculate its orthogonal projection onto V.

Table of Contents

Orthogonality

a. Two vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^n$ are called *perpendicular* or *orthogonal*[1] if $\vec{v} \cdot \vec{w} = 0$.

b. The *length* (or magnitude or norm) of a vector $\vec{v}$ in $\mathbb{R}^n$ is $\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$.

c. A vector $\vec{u}$ in $\mathbb{R}^n$ is called a *unit vector* if its length is 1, (i.e., $\|\vec{u}\| = 1$, or $\vec{u} \cdot \vec{u} = 1$).

The vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$ in $\mathbb{R}^n$ are called *orthonormal* if they are all unit vectors and orthogonal to one another:

$$\vec{u}_i \cdot \vec{u}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

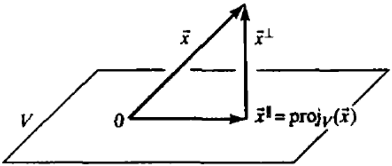
a. Orthonormal vectors are linearly independent.

b. Orthonormal vectors $\vec{u}_1, \ldots, \vec{u}_n$ in $\mathbb{R}^n$ form a basis of $\mathbb{R}^n$.

Orthogonal projection

The vector $\vec{x}^{\parallel}$ is called the *orthogonal projection* of $\vec{x}$ onto V, denoted by $\text{proj}_V(\vec{x})$. See Figure 4.

The transformation $T(\vec{x}) = \text{proj}_V(\vec{x}) = \vec{x}^{\parallel}$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ is linear.



If V is a subspace of $\mathbb{R}^n$ with an orthonormal basis $\vec{u}_1, \ldots, \vec{u}_m$, then

$$\text{proj}_V(\vec{x}) = \vec{x}^{\parallel} = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_m \cdot \vec{x})\vec{u}_m.$$

for all $\vec{x}$ in $\mathbb{R}^n$.

Consider an orthonormal basis $\vec{u}_1, \ldots, \vec{u}_n$ of $\mathbb{R}^n$. Then

$$\vec{x} = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_n \cdot \vec{x})\vec{u}_n.$$

for all $\vec{x}$ in $\mathbb{R}^n$.

Orthogonal complement

Consider a subspace V of $\mathbb{R}^n$. The *orthogonal complement* $V^\perp$ of V is the set of those vectors $\vec{x}$ in $\mathbb{R}^n$ that are orthogonal to all vectors in V:

$$V^\perp = \{\vec{x} \text{ in } \mathbb{R}^n : \vec{v} \cdot \vec{x} = 0, \text{ for all } \vec{v} \text{ in } V\}.$$

Note that $V^\perp$ is the kernel of the orthogonal projection onto V.

Consider a subspace V of $\mathbb{R}^n$.

a. The orthogonal complement $V^\perp$ of V is a subspace of $\mathbb{R}^n$.

b. The intersection of V and $V^\perp$ consists of the zero vector alone: $V \cap V^\perp = \{\vec{0}\}$.

c. $\dim(V) + \dim(V^\perp) = n$

d. $(V^\perp)^\perp = V$

Gram-Schmidt process

Definition

It gives the way to construct an orthonormal bases given an arbitrary basis of the space.
The construction process is as follow:

1. $$\vec{v}_j = \vec{v}_j^{\parallel} + \vec{v}_j^{\perp}, \qquad \text{with respect to } \operatorname{span}(\vec{v}_1, \ldots, \vec{v}_{j-1}).$$

$$\vec{v}_j^{\perp} = \vec{v}_j - \vec{v}_j^{\parallel} = \vec{v}_j - (\vec{u}_1 \cdot \vec{v}_j)\,\vec{u}_1 - \cdots - (\vec{u}_{j-1} \cdot \vec{v}_j)\,\vec{u}_{j-1}$$

2. $$\vec{u}_1 = \frac{1}{\|\vec{v}_1\|}\vec{v}_1, \quad \vec{u}_2 = \frac{1}{\left\|\overrightarrow{v_2}^{\perp}\right\|}\vec{v}_2^{\perp}, \ldots, \overline{u}_m = \frac{1}{\left\|\overrightarrow{v_m}^{\perp}\right\|}\vec{v}_m^{\perp}$$

The calculation will not be more than 4 dimension.

QR factorization

Definition

Consider an $n \times m$ matrix M with linearly independent columns $\vec{v}_1, \dots, \vec{v}_m$. Then there exists an $n \times m$ matrix Q whose columns $\vec{u}_1, \dots, \vec{u}_m$ are orthonormal and an upper triangular matrix R with positive diagonal entries such that

$$M = QR.$$

This representation is unique. Furthermore, $r_{11} = \|\vec{v}_1\|$, $r_{jj} = \left\|\vec{v}_j^{\perp}\right\|$ (for $j = 2, \dots, m$), and $r_{ij} = \vec{u}_i \cdot \vec{v}_j$ (for $i < j$). ■

Just perform the Gram-Schmidt process for the set of columns of M and all the information about matrix Q and R will be shown in the process.

Table of Contents

Least squares approximation

Method

The least-squares solutions of the system

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

are the exact solutions of the (consistent) system

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

The system $A^T A\vec{x} = A^T\vec{b}$ is called the *normal equation* of $A\vec{x} = \vec{b}$.

If $\ker(A) = \{\vec{0}\}$, then the linear system

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

has the unique least-squares solution

$$\vec{x}^* = (A^T A)^{-1} A^T\vec{b}.$$

Note that $\vec{x}^*$ is not the orthogonal projection of $\vec{b}$, $A\vec{x}^*$ is.

Least squares approximation

Exercise

Consider the data in the following table.

Reading score, r	Shoe size, s
73	9
92	10
67	5
48	4

Suppose we model the relationship by $r = a + bs + cs^2$ for unknown a, b, c.

(a) (3 points) Set up a system of linear equations for a, b, c that uses all the data above. (Your system may be consistent or inconsistent, depending on the data. Your technique should result in the correct a, b, c when the data allows.)

(b) (3 points) Sketch a graph of r against s.

(c) (3 points) Is the system above consistent? Explain. Hint: You might want to refer to the graph and to consider $92 - 73, 73 - 67$, and $67 - 48$.

Least squares approximation

Exercise

(d) (3 points) Set up a system of linear equations whose solution is the least squares solution for a, b, c. Show work. You need not solve the system. Explain why there is a unique solution.

(e) (3 points) Instead of the above, now assume that s is a constant that doesn't depend on r, so $s = d$. Find the least squares solution for d, showing work.

Least squares approximation

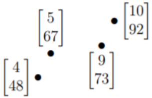
Answer

Solution:
We have

$$\begin{bmatrix} 1 & 9 & 81 \\ 1 & 10 & 100 \\ 1 & 5 & 25 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 73 \\ 92 \\ 67 \\ 48 \end{bmatrix}.$$

(b) (3 points) Sketch a graph of r against s.

Solution:



(c) (3 points) Is the system above consistent? Explain. Hint: You might want to refer to the graph and to consider $92 - 73, 73 - 67$, and $67 - 48$.

Solution:
No. As s increases, r increases, but with the pattern "big jump, little jump, big jump." This is inconsistent with r a quadratic function of s. That is, r is neither convex nor concave.

Least squares approximation

Answer

(d) (3 points) Set up a system of linear equations whose solution is the least squares solution for a, b, c. Show work. You need not solve the system. Explain why there is a unique solution.

Solution:

With A equal to

$$\begin{bmatrix} 1 & 9 & 81 \\ 1 & 10 & 100 \\ 1 & 5 & 25 \\ 1 & 4 & 16 \end{bmatrix},$$

the normal equation is

$$(A^T A) \begin{bmatrix} a \\ b \\ c \end{bmatrix} = A^T \begin{bmatrix} 73 \\ 92 \\ 67 \\ 48 \end{bmatrix}.$$

Since the matrix A is Vandermonde with different parameters and at least as many rows as columns (i.e., since three distinct points determine a quadratic and we have data for at least three distinct points), the matrix A has full rank, 3. It follows that

Least squares approximation

Answer

A^TA is invertible. Thus

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = (A^TA)^{-1}A^T \begin{bmatrix} 73 \\ 92 \\ 67 \\ 48 \end{bmatrix}$$

is the unique solution.

(e) (3 points) Instead of the above, now assume that s is a constant that doesn't depend on r, so $s = d$. Find the least squares solution for d, showing work.

Solution:
Note: There was a typo in the question and so it was not scored in Fall 2016. It should read "$r = d$."
Algebraicly, we have $A = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$, so $A^TA = 4$ and

$$d = (A^TA)^{-1}A^T \begin{bmatrix} 73 \\ 92 \\ 67 \\ 48 \end{bmatrix}.$$

This is the average, $\frac{(73+67)+(92+48)}{4} = 70$.

Table of Contents

Riesz-Fischer

Refer to the slide "Riesz-Fischer" on Canvas. This is important!

Table of Contents

Remark

1. Know how to find determinant using properties of determinant, patterns, Laplace expansion...Properties are especially important!!
2. Try to make the determinant easy to solve: using Laplace expansion to reduce the calculation of 4×4 matrix to 3×3 matrix, use properties of determinant and elementary row operation to make the determinant obvious...

Exercise

$$\det\begin{pmatrix} 1+x & 1 & 1 & 1 \\ 1 & 1-x & 1 & 1 \\ 1 & 1 & 1+z & 1 \\ 1 & 1 & 1 & 1-z \end{pmatrix}, \qquad \det\begin{pmatrix} 3 & 2 & 2 & 2 & \dots & 2 & 2 \\ 2 & 3 & 2 & 2 & \dots & 2 & 2 \\ 2 & 2 & 3 & 2 & \dots & 2 & 2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 2 & 2 & 2 & 2 & \dots & 2 & 3 \end{pmatrix}$$

$$\det\begin{pmatrix} 2 & 2 & 2 & \dots & 2 & 2 & 1 \\ 2 & 2 & 2 & \dots & 2 & 2 & 2 \\ 2 & 2 & 2 & \dots & 3 & 2 & 2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 2 & n-1 & 2 & \dots & 2 & 2 & 2 \\ n & 2 & 2 & \dots & 2 & 2 & 2 \end{pmatrix}$$