Applied Linear Algebra

OTTO BRETSCHER

http://www.prenhall.com/bretscher

Chapter 5
Orthogonality and Least Squares

Chia-Hui Chang
Email: chia@csie.ncu.edu.tw
National Central University, Taiwan

5.1 ORTHONORMAL BASES AND ORTHOGONAL PROJECTIONS

Not all bases are created equal.

Definition. *5.1.1*

Orthogonality, length, unit vectors
a. Two vectors $\vec{v}$ and $\vec{w}$ in R^n are called perpendicular or orthogonal if $\vec{v} \cdot \vec{w} = 0$.

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

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

Explanation:

If $\vec{v}$ is a nonzero vector in R^n, then

$$\vec{u} = \frac{1}{\|\vec{v}\|}\vec{v}$$

is a unit vector.

Definition. *5.1.2* **Orthonormal vectors**
The vector $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_m}$ *in* R^n *are called orthonormal if they are all unit vectors and orthogonal to one another:*

$$\vec{v_i} \cdot \vec{v_j} = \begin{cases} 1 & if \quad i=j, \\ 0 & if \quad i \neq j. \end{cases}$$

Example. *1.*

The vectors $\vec{e_1}, \vec{e_2}, \ldots, \vec{e_n}$ in R^n are orthonormal.

Example. *2.*

For any scalar α, the vectors $\begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}, \begin{bmatrix} -\sin\alpha \\ \cos\alpha \end{bmatrix}$ are orthonormal.

Example. *3. The vectors*

$$\vec{v_1} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}, \vec{v_3} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix}$$

in R^4 are orthonormal. Can you find a vector $\vec{v_4}$ in R^4 such that all the vectors $\vec{v_1}$,$\vec{v_2}$,$\vec{v_3}$,$\vec{v_4}$ are orthonormal.

The following properties of orthonormal vectors are often useful:

Fact 5.1.3
a. Orthonormal vectors are linearly independent.
b. Orthonormal vectors $\vec{v_1}$, ..., $\vec{v_n}$ in R^n form a basis of R^n.

Proof
a. Consider a relation

$$c_1\vec{v_1}+c_2\vec{v_2}+\cdots+c_i\vec{v_i}+\cdots+c_m\vec{v_m}=\vec{0}$$

Let us form the dot product of each side of this equation with $\vec{v_i}$:

$$(c_1\vec{v_1}+c_2\vec{v_2}+\cdots+c_i\vec{v_i}+\cdots+c_m\vec{v_m})\cdot\vec{v_i}=$$
$$\vec{0}\cdot\vec{v_i}=0.$$

Because the dot product is distributive.

$$c_i(\vec{v_i}\cdot\vec{v_i})=0$$

Therefore, $c_i=0$ for all $i=1$, ..., m.

b. Any n linearly independent vectors in R^n form a basis of R^n.

Definition. *5.1.4* **Orthogonal complement** *Consider a subspace V of R^n. The orthogonal complement $V^\perp$ of V is the set of those vectors $\vec{x}$ in R^n that are orthogonal to all vectors in V:*

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

Fact 5.1.5 If V is a subspace of R^n, then its orthogonal complement $V^\perp$ is a subspace of R^n as well.

Proof

We will verify that $V^\perp$ is closed under scalar multiplication and leave the verification of the two other properties as Exercise 23. Consider a vector $\vec{w}$ in $V^\perp$ and a scalar k. We have to show that $k\vec{w}$ is orthogonal to all vectors $\vec{v}$ in V. Pick an arbitrary vector $\vec{v}$ in V. Then, $(k\vec{w})\cdot\vec{v}=k(\vec{w}\cdot\vec{v})=0$, as claimed.

Orthogonal projections

See Figure 5.

The **orthogonal projection** of a vector $\vec{x}$ onto one-dimentaional subspace V with basis $\vec{v}_1$ (unit vector) is computed by:

$$proj_V\vec{x} = \vec{w} = (\vec{v_1} \cdot \vec{x})\vec{v_1}$$

Now consider a subspace V with arbitrary dimension m. Suppose we have an orthonormal basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ of V. Find $\vec{w}$ in V such that $\vec{x} - \vec{w}$ is in $V^\perp$. Let

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

It is required that

$$\overrightarrow{x - w} = \vec{x} - c_1\vec{v}_1 - c_2\vec{v}_2 - \cdots - c_m\vec{v}_m$$

is perpendicular to V; i.e.:

$$\vec{v}_i \cdot (\vec{x-w}) = \vec{v}_i \cdot (\vec{x} - c_1\vec{v}_1 - c_2\vec{v}_2 - \cdots - c_m\vec{v}_m)$$

$$= \vec{v}_i \cdot \vec{x} - c_1(\vec{v}_i \cdot \vec{v}_1) - \cdots - c_i(\vec{v}_i \cdot \vec{v}_i) - \cdots - c_m(\vec{v}_i \cdot \vec{v}_m)$$

$$= \vec{v}_i \cdot \vec{x} - c_i = 0$$

The equation holds if $c_i = \vec{v}_i \cdot \vec{x}$.
Therefore, there is a unique $\vec{w}$ in V such that $\vec{x} - \vec{w}$ is in $V^\perp$, namely,

$$\vec{w} = (\vec{v}_1 \cdot \vec{x})\vec{v}_1 + (\vec{v}_2 \cdot \vec{x})\vec{v}_2 + \cdots + (\vec{v}_m \cdot \vec{x})\vec{v}_m$$

Fact 5.1.6 Orthogonal projection
Consider a subspace V of R^n with orthonormal basis $\vec{v_1}$,$\vec{v_2}$,. . .,$\vec{v_m}$. For any vector $\vec{x}$ in R^n, there is a unique vector $\vec{w}$ in V such that $\vec{x}$-$\vec{w}$ is in $V^\perp$. This vector $\vec{w}$ is called the orthogonal projection of $\vec{x}$ onto V, denoted by $proj_V\vec{x}$. We have the formula

$$proj_V\vec{x} = (\vec{v_1} \cdot \vec{x})\vec{v_1} + \cdots + (\vec{v_m} \cdot \vec{x})\vec{v_m}.$$

The transformation $T(\vec{x}) = proj_V\vec{x}$ from R^n to R^n is linear.

Example. *4*

Consider the subspace V=im(A) of R^4. where

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

Find $proj_V \vec{x}$, for

$$\vec{x} = \begin{bmatrix} 1 \\ 3 \\ 1 \\ 7 \end{bmatrix}.$$

Solution

The two columns of A form a basis of V. Since they happen to be orthogonal, we can construct an orthonormal basis of V merely by dividing these two vectors by their length (2 for both vectors):

$$\vec{v_1} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}$$

Then,

$proj_V\vec{x}$=$(\vec{v_1} \cdot \vec{x})\vec{v_1}$+$(\vec{v_2} \cdot \vec{x})\vec{v_2}$=6$\vec{v_1}$+ 2$\vec{v_2}$=

$$\begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 2 \\ 4 \end{bmatrix}.$$

To check this answer, verify that $\vec{x}$-$proj_V\vec{x}$ is perpendicular to both $\vec{v_1}$ and $\vec{v_2}$.

What happens when we apply Fact 5.1.6 to the subspace $V=R^n$ of R^n with orthonormal basis $\vec{v_1}$, $\vec{v_2}$,$\cdots$,$\vec{v_n}$? Clearly, $proj_V\vec{x}=\vec{x}$, for all $\vec{x}$ in R_n. Therefore,

$$\vec{x} = (\vec{v_1}\cdot\vec{x})\vec{v_1} + \cdots + (\vec{v_n}\cdot\vec{x})\vec{v_n},$$

for all $\vec{x}$ in R^n. See Figure 7.

Fact 5.1.7

Consider an orthonormal basis $\vec{v_1}$,$\cdots$,$\vec{v_n}$ of R^n. Then,

$$\vec{x} = (\vec{v_1}\cdot\vec{x})\vec{v_1} + \cdots + (\vec{v_n}\cdot\vec{x})\vec{v_n},$$

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

This is useful for computing the B-coordinate, since $c_i = \vec{v}_i \cdot \vec{x}$.

Example. *5*

By using paper and pencil, express the vector $\vec{x}= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ as a linear combination of

$$\vec{v_1} = \frac{1}{3}\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \vec{v_2} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \vec{v_3} = \frac{1}{3}\begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}.$$

Solution
Since $\vec{v_1}, \vec{v_2}, \vec{v_3}$ is an orthonormal basis of R^3, we have

$\vec{x} = (\vec{v_1} \cdot \vec{x})\vec{v_1} + (\vec{v_2} \cdot \vec{x})\vec{v_2} + (\vec{v_3} \cdot \vec{x})\vec{v_3} = 3\vec{v_1} + \vec{v_2} + 2\vec{v_3}$.

From Pythagoras to Cauchy

Example. *6*

Consider a line L in R^3 and a vector $\vec{x}$ in R^3. What can you say about the relationship between the lengths of the vectors $\vec{x}$ and $proj_L\vec{x}$?

Solution

Applying the Pythagorean theorem to the shaded right triangle in Figure 8, we find that

$$\| proj_L\vec{x} \| \leq \| \vec{x} \| .$$

The statement is an equality if (and only if) $\vec{x}$ is on L.

Does this inequality hold in higher dimensional cases? We have to examine whether the Pythagorean theorem holds in R^n.

Fact 5.1.8 Pythagorean theorem
Consider two vectors $\vec{x}$ and $\vec{y}$ in R^n. The equation

$$\| \vec{x} + \vec{y} \|^2 = \| \vec{x} \|^2 + \| \vec{y} \|^2$$

holds if (and only if) $\vec{x}$ and $\vec{y}$ are orthogonal. (See Figure 9.)

Proof The verification is straightforward:
$\| \vec{x} + \vec{y} \|^2 = (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y})$
$= \vec{x} \cdot \vec{x} + 2(\vec{x} \cdot \vec{y}) + \vec{y} \cdot \vec{y}$
$= \| \vec{x} \|^2 + 2(\vec{x} \cdot \vec{y}) + \| \vec{y} \|^2$
$= \| \vec{x} \|^2 + \| \vec{y} \|^2$
if (and only if) $\vec{x} \cdot \vec{y} = 0$.

Fact 5.1.9 Consider a subspace V of R^n and a vector $\vec{x}$ in R^n. Then,

$$\| proj_V \vec{x} \| \leq \| \vec{x} \|.$$

The statement is an equality if (and only if) $\vec{x}$ is in V.

Proof we can write $\vec{x} = proj_V \vec{x} + (\vec{x} - proj_V \vec{x})$and apply the Pythagorean theorem(see Figure 10):

$$\| \vec{x} \|^2 = \| proj_V \vec{x} \|^2 + \| \vec{x} - proj_V \vec{x} \|^2.$$

It follows that $\| proj_V \vec{x} \| \leq \| \vec{x} \|$, as claimed.

Let V be a one-dimensional subspace of R^n spanned by a (nonzero) vector $\vec{y}$. We introduce the unit vector

$$\vec{u} = \frac{1}{\|\vec{y}\|}\vec{y}$$

in V. (See Figure 11.)
We know that

$$proj_V\vec{x} = (\vec{u}\cdot\vec{x})\vec{u} = \frac{1}{\|\vec{y}\|^2}(\vec{y}\cdot\vec{x})\vec{y}.$$

for any $\vec{x}$ in R^n. Fact 5.1.9 tells us that

$$\| \vec{x} \| \geq \| proj_V\vec{x} \| = \| \frac{1}{\|\vec{y}\|^2}(\vec{y}\cdot\vec{x})\vec{y} \| = \frac{1}{\|\vec{y}\|^2}|\vec{y}\cdot\vec{x}| \, \| \vec{y} \|.$$

To justify the last step, note that $\| k\vec{v} \| = |k| \, \| \vec{v} \|$, for all vectors $\vec{v}$ in R^n and all scalars k. (See Exercise 25(a).) We conclude that

$$\frac{|\vec{x}\cdot\vec{y}|}{\|\vec{y}\|} \leq \| \vec{x} \|.$$

Fact 5.1.10 Cauchy-Schwarz inequality
If $\vec{x}$ and $\vec{y}$ are vectors in R^n, then

$$|\vec{x} \cdot \vec{y}| \leq \| \vec{x} \| \| \vec{y} \|.$$

The statement is an equality if (and only if) $\vec{x}$ and $\vec{y}$ are parallel.

Definition. *5.1.11*

Angle between two vectors Consider two nonzero vectors $\vec{x}$ and $\vec{y}$ in R^n. The angle α between these vectors is defined as

$$\cos\alpha = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\|\vec{y}\|}.$$

Note that α is between 0 and π, by definition of the inverse cosine functiion.

Example. *7*

Find the angle between the vectors

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } \vec{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Solution

$$\cos\alpha = \tfrac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \tfrac{1}{1\cdot 2} = \tfrac{1}{2}$$

$$\alpha = \frac{\pi}{3}$$

Correlation

Consider two characteristics of a population, with deviation vectors $\vec{x}$ and $\vec{y}$. There is a positive correlation between the two characteristics if (and only if) $\vec{x} \cdot \vec{y} > 0$.

Definition. *5.1.12*

Correlation coefficient
The correlation coefficient r between two characteristics of a population is the cosine of the angle α between the deviation vectors $\vec{x}$ and $\vec{y}$ for the two characteristics:

$$r = cos(\alpha) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|}$$

Exercise 5.1: 7, 9, 12, 19, 23, 24, 25, 28

5.2 GRAM-SCHMIDT PROCESS AND QR FACTORIZATION

How can we construct an orthonormal basis? Say, from any basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ of a subspace V?

If V is a line with basis $\vec{v}_1$:

$$\vec{w}_1 = \frac{1}{\|\vec{v}_1\|}\vec{v}_1$$

When V is a plane with basis $\vec{v}_1, \vec{v}_2$, we first get $\vec{w}_1$ as above.

Next find a vector in V orthogonal to $\vec{w}_1$.

$$\vec{v}_2 - proj_L\vec{v}_2 = \vec{v}_2 - (\vec{v}_2 \cdot \vec{w}_1)\vec{w}_1$$

Then Divide the vector by its length to get the second vector $\vec{w}_2$.

$$\vec{w}_2 = \frac{1}{\|\vec{v}_2 - proj_L\vec{v}_2\|}(\vec{v}_2 - proj_L\vec{v}_2)$$

See Figure 1, 2, 3.

EXAMPLE 1 Find an orthonormal basis of the subspace

$$V = span\left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 9 \\ 9 \\ 1 \end{bmatrix}\right)$$

of R^4, with basis

$$\vec{v_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 1 \\ 9 \\ 9 \\ 1 \end{bmatrix}.$$

Solution

Using the terminology just introduced, we find the following results:

$$\| \vec{v_1} \| = \sqrt{1^2 + 1^2 + 1^2 + 1^2} = 2,$$

$$\vec{w_1} = \frac{1}{\|\vec{v_1}\|}\vec{v_1} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}.$$

$$\vec{w_1} \cdot \vec{v_2} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 9 \\ 9 \\ 1 \end{bmatrix} = 10,$$

$$proj_L \vec{v_2} = (\vec{w_1} \cdot \vec{v_2})\vec{w_1} = 10 \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 5 \\ 5 \end{bmatrix}$$

$$\vec{v_2} - proj_L \vec{v_2} = \begin{bmatrix} 1 \\ 9 \\ 9 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ 5 \\ 5 \\ 5 \end{bmatrix} = \begin{bmatrix} -4 \\ 4 \\ 4 \\ -4 \end{bmatrix}.$$

$$\| \vec{v_2} - proj_L \vec{v_2} \| = \sqrt{4 \cdot 16} = 8,$$

$$\vec{w_2} = \frac{1}{\|\vec{v_2} - proj_L \vec{v_2}\|}(\vec{v_2} - proj_L \vec{v_2})$$

$$= \frac{1}{8}\begin{bmatrix} -4 \\ 4 \\ 4 \\ -4 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \\ -1/2 \end{bmatrix}.$$

We have found an orthonormal basis of V:

$$\vec{w_1} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \vec{w_2} = \begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \\ -1/2 \end{bmatrix}$$

We can represent the preceding computations more succinctly in matrix form. Let's solve the equations defining $\vec{w_1}$ and $\vec{w_2}$.

$\vec{w_1} = \frac{1}{\|\vec{v_1}\|}\vec{v_1}$ and $\vec{w_2} = \frac{1}{\|\vec{v_2} - proj_L \vec{v_2}\|}(\vec{v_2} - proj_L \vec{v_2})$, for vectors $\vec{v_1}$ and $\vec{v_2}$:

$$\vec{v_1} = \| \vec{v_1} \| \vec{w_1},$$

and

$$\vec{v_2} = proj_L\vec{v_2} + \| \vec{v_2} - proj_L\vec{v_2} \| \vec{w_2}$$

$$= (\vec{w_1} \cdot \vec{v_2})\vec{w_1} + \| \vec{v_2} - proj_L\vec{v_2} \| \vec{w_2}.$$

We can write the last two equations in matrix form:

$$\begin{bmatrix} \vec{v_1} & \vec{v_2} \end{bmatrix} = \underbrace{\begin{bmatrix} \vec{w_1} & \vec{w_2} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \| \vec{v_1} \| & \vec{w_1} \cdot \vec{v_2} \\ 0 & \| \vec{v_2} - proj_L\vec{v_2} \| \end{bmatrix}}_{R}$$

Note that we have written 4×2 matrix Q with orthonormal columns and the upper triangular 2×2 matrix R with positive entries on the diagonal.

Matrix Q stores the orthonormal basis $\vec{w_1}$, $\vec{w_2}$ we constructed, and matrix R gives the relationship between the "old" basis $\vec{v_1}$, $\vec{v_2}$, and the "new" basis $\vec{w_1}$, $\vec{w_2}$ of V.

Let's plug in numbers (note that we computed all the entries of matrix of matrix R in the process of finding $\vec{w_1}$ and $\vec{w_2}$):

$$\begin{bmatrix} 1 & 1 \\ 1 & 9 \\ 1 & 9 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2 & 10 \\ 0 & 8 \end{bmatrix}$$

Algorithm 5.2.1
The Gram-Schmidt process
Consider a subspace V of R^n with basis $\vec{v_1}$,$\vec{v_2}$,...,$\vec{v_m}$. We wish to construct an orthonormal basis $\vec{w_1}$,$\vec{w_2}$,...,$\vec{w_m}$ of V.

Let $\vec{w_1} = (\frac{1}{\|\vec{v_1}\|})\vec{v_1}$. As we define $\vec{w_j}$ for $j = 2, 3, ..., m$, we may assume that an orthonormal basis $\vec{w_1}$,$\vec{w_2}$,...,$\vec{w_{j-1}}$ of $V_{j-1} = span(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_{j-1}})$ has already been constructed. Let

$$\vec{w_j} = \frac{1}{\|\vec{v_j} - proj_{V_{j-1}}\vec{v_j}\|}(\vec{v_j} - proj_{V_{j-1}}\vec{v_j}).$$

Note that

$$proj_{V_{j-1}}\vec{v_j}$$

$$= (\vec{w_1} \cdot \vec{v_j})\vec{w_1} + (\vec{w_2} \cdot \vec{v_j})\vec{w_2} + \ldots + (\vec{w_{j-1}} \cdot \vec{v_j})\vec{w_{j-1}},$$

by Fact 5.1.6.

THE QR Factorization

The Gram-Schmidt process can be presented succinctly in matrix form, as illustrated in Example 1. Using the terminology introduced in Algorithm 5.2.1, we can write

$$\vec{v_1} = \|\vec{v_1}\|\vec{w_1}$$

and

$$\vec{v_j} = proj_{V_{j-1}}\vec{v_j} + \|\vec{v_j} - proj_{V_{j-1}}\vec{v_j}\|\vec{w_j}$$

$$= (\vec{w_1}\vec{v_j})\vec{w_1} + \cdots + (\vec{w}_{j-1}\vec{v_j})\vec{w}_{j-1} + \|\vec{v_j} - proj_{V_{j-1}}\vec{v_j}\|\vec{w_j}$$

(for j=2,3,...,m).

Let

$$\begin{aligned} r_{11} &= \|\vec{v_1}\| \\ r_{jj} &= \|\vec{v_j} - proj_{V_{j-1}}\vec{v_j}\| \quad (j = 2, 3, ..., m), \\ r_{ij} &= \vec{w_i} \cdot \vec{v_j} \qquad\qquad\qquad (i < j). \end{aligned}$$

Then,

$$\vec{v_1} = r_{11}\vec{w_1}$$
$$\vec{v_2} = r_{12}\vec{w_1} + r_{22}\vec{w_2}$$
$$\vdots$$
$$\vec{v_m} = r_{1m}\vec{w_1} + r_{2m}\vec{w_2} + \cdots + r_{mm}\vec{w_m}.$$

We can write these equations in matrix form:

$$\begin{bmatrix} | & | & & | \\ \vec{v_1} & \vec{v_2} & \cdots & \vec{v_m} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{w_1} & \vec{w_2} & \cdots & \vec{w_m} \\ | & | & & | \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ 0 & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{mm} \end{bmatrix}$$

$$M = QR$$

Note that M is an $n \times m$ matrix with linearly independent columns, Q is an $n \times m$ matrix with orthonormal columns, and R is an upper triangular $m \times m$ matrix with positive entires on the diagonal.

Fact 5.2.2 QR factorization

Consider an $n \times m$ matrix M with linearly independent columns $\vec{v_1}, ..., \vec{v_m}$. Then there is an $n \times m$ matrix Q whose columns $\vec{w_1}, ..., \vec{w_m}$ are orthonormal and an upper triangular $m \times m$ matrix R with positive diagonal entries such that

$$M = QR.$$

This representation is unique. Furthermore, $r_{11} = \|\vec{v_1}\|$, $r_{ij} = \|\vec{v_j} - proj_{V_{j-1}}\vec{v_j}\|$(for $j > 1$),

and $r_{ij} = \vec{w_i} \cdot \vec{v_j}$ (for $i < j$),

where $V_{j-1} = span(\vec{v_1}, \vec{v_2}, ..., \vec{v}_{j-1})$.

EXAMPLE 2 Find the QR factorization of the shear matrix M= $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

Solution

Here

$$\vec{v_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

As in Example 1, the QR factorization of M will have the form

$$M = \begin{bmatrix} \vec{v_1} & \vec{v_2} \end{bmatrix} = \begin{bmatrix} \vec{w_1} & \vec{w_2} \end{bmatrix} \begin{bmatrix} \| \vec{v_1} \| & \vec{w_1} \cdot \vec{v_2} \\ 0 & \| \vec{v_2} - proj_{V_1} \vec{v_2} \| \end{bmatrix}$$

We will compute the columns of W and the entries of R step by step:

$$r_{11} = \|\vec{v_1}\| = \sqrt{2}$$

$$\vec{w_1} = \frac{1}{\|\vec{v_1}\|}\vec{v_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$r_{12} = \vec{w_1} \cdot \vec{v_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}$$

$$\vec{v_2} - proj_{v_1}\vec{v_2} = \vec{v_2} - (\vec{w_1} \cdot \vec{v_2})\vec{w_1}$$

$$= \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix}$$

$$r_{22} = \|\vec{v_2} - proj_{v_1}\vec{v_2}\| = \sqrt{\frac{1}{4} + \frac{1}{4}} = \frac{1}{\sqrt{2}}$$

$$\vec{w_2} = \frac{1}{\|\vec{v_2} - proj_{v_1}\vec{v_2}\|}(\vec{v_2} - proj_{v_1}\vec{v_2})$$

$$= \sqrt{2} \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Now,

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = M = QR = \begin{bmatrix} \vec{w_1} & \vec{w_2} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{bmatrix}$$

$$= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \right).$$

Draw pictures analogous to Figures 1 through 3 to illustrate these computations!

Exercise 5.2 5, 11, 13, 19, 27, 31, 33, 37

5.3 ORTHOGONAL TRANSFORMATIONS AND ORTHOGONAL MATRICES

Definition 5.3.1 Orthogonal transformations and orthogonal matrices

A linear transformation T from R^n to R^n is called orthogonal if it preserves the length of vectors:

$$\|T(\vec{x})\| = \|\vec{x}\|, \text{ for all } \vec{x} \text{ in } R^n.$$

If $T(\vec{x}) = A\vec{x}$ is an orthogonal transformation, we say that A is an orthogonal matrix.

EXAMPLE 1 The rotation

$$T(\vec{x}) = \begin{bmatrix} cos\phi & -sin\phi \\ sin\phi & cos\phi \end{bmatrix} \vec{x}$$

is an orthogonal transformation from R^2 to R^2, and

$$A = \begin{bmatrix} cos\phi & -sin\phi \\ sin\phi & cos\phi \end{bmatrix} \vec{x}$$

is an orthogonal matrix, for all angles ϕ.

EXAMPLE 2 Reflection
Consider a subspace V of R^n. For a vector $\vec{x}$ in R^n, the vector $R(\vec{x}) = 2proj_V\vec{x} - \vec{x}$ is called the reflection of $\vec{x}$ in V. (see Figure 1).
Show that reflections are orthogonal transformations.

Solution
We can write

$$R(\vec{x}) = proj_V\vec{x} + (proj_V\vec{x} - \vec{x})$$

and

$$\vec{x} = proj_V\vec{x} + (\vec{x} - proj_V\vec{x}).$$

By the pythagorean theorem, we have

$$\|R(\vec{x})\|^2 = \|proj_V\vec{x}\|^2 + \|proj_V\vec{x} - \vec{x}\|^2$$

$$= \|proj_V\vec{x}\|^2 + \|\vec{x} - proj_V\vec{x}\|^2 = \|\vec{x}\|^2.$$

Fact 5.3.2 Orthogonal transformations preserve orthogonality

Consider an orthogonal transformation T from R^n to R^n. If the vectors $\vec{v}$ and $\vec{w}$ in R^n are orthogonal, then so are $T(\vec{v})$ and $T(\vec{w})$.

Proof

By the theorem of Pythagoras, we have to show that

$$\|T(\vec{v}) + T(\vec{w})\|^2 = \|T(\vec{v})\|^2 + \|T(\vec{w})\|^2.$$

Let's see:

$\|T(\vec{v}) + T(\vec{w})\|^2 = \|T(\vec{v} + \vec{w})\|^2$ (T is linear)

$= \|\vec{v} + \vec{w}\|^2$ (T is orthogonal)

$= \|\vec{v}\|^2 + \|\vec{w}\|^2$ ($\vec{v}$ and $\vec{w}$ are orthogonal)

$= \|T(\vec{v})\|^2 + \|T(\vec{w})\|^2$.
($T(\vec{v})$ and $T(\vec{w})$ are orthogonal)

Fact 5.3.3 Orthogonal transformations and orthonormal bases

a. A linear transformation T from R^n to R^n is orthogonal iff the vectors $T(\vec{e_1})$, $T(\vec{e_2})$,...,$T(\vec{e_n})$ form an orthonormal basis of R^n.

b. An $n \times n$ matrix A is orthogonal iff its columns form an orthonormal basis of R^n.

Proof Part(a):

$\Rightarrow$ If T is orthogonal, then, by definition, the $T(\vec{e_i})$ are unit vectors, and by Fact 5.3.2, since $\vec{e_1}$, $\vec{e_2}$,...,$\vec{e_n}$ are orthogonal, $T(\vec{e_1})$, $T(\vec{e_2})$,...,$T(\vec{e_n})$ are orthogonal.

$\Leftarrow$ Conversely, suppose the $T(\vec{e_i})$ form an orthonormal basis.

Consider a vector

$$\vec{x} = x_1\vec{e_1} + x_2\vec{e_2} + \cdots + x_n\vec{e_n}$$

in R^n. Then,

$$\|T(\vec{x})\|^2 = \|x_1T(\vec{e_1}) + x_2T(\vec{e_2}) + \cdots + x_nT(\vec{e_n})\|^2$$

$$= \|x_1T(\vec{e_1})\|^2 + \|x_2T(\vec{e_2})\|^2 + \cdots + \|x_nT(\vec{e_n})\|^2$$
(by Pythagoras)

$$= x_1^2 + x_2^2 + \cdots + x_n^2$$

$$= \|\vec{x}\|^2$$

Part(b) then follows from Fact 2.1.2.

Warning: A matrix with orthogonal columns need not be orthogonal matrix.

As an example, consider the matrix $A = \begin{bmatrix} 4 & -3 \\ 3 & 4 \end{bmatrix}$.

EXAMPLE 3 Show that the matrix A is orthogonal:

$$A = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}.$$

Solution

Check that the columns of A form an orthonoraml basis of R^4.

Fact 5.3.4
Products and inverses of orthogonal matrices

a. The product AB of two orthogonal $n \times n$ matrices A and B is orthogonal.

b. The inverse A^{-1} of an orthogonal $n \times n$ matrix A is orthogonal.

Proof
In part (a), the linear transformation $T(\vec{x}) = AB\vec{x}$ preserves length, because $\|T(\vec{x})\| = \|A(B\vec{x})\| = \|B\vec{x}\| = \|\vec{x}\|$. Figure 4 illustrates property (a).

In part (b), the linear transformation $T(\vec{x}) = A^{-1}\vec{x}$ preserves length, because $\|A^{-1}\vec{x}\| = \|A(A^{-1}\vec{x})\|$.

The Transpose of a Matrix

EXAMPLE 4 Consider the orthogonal matrix

$$A = \tfrac{1}{7}\begin{bmatrix} 2 & 6 & 3 \\ 3 & 2 & -6 \\ 6 & -3 & 2 \end{bmatrix}.$$

Form another 3×3 matrix B whose ijth entry is the jith entry of A:

$$B = \tfrac{1}{7}\begin{bmatrix} 2 & 3 & 6 \\ 6 & 2 & -3 \\ 3 & -6 & 2 \end{bmatrix}$$

Note that the rows of B correspond to the columns of A. Compute BA, and explain the result.

Solution

$$BA = \frac{1}{49}\begin{bmatrix} 2 & 6 & 3 \\ 6 & 2 & -3 \\ 3 & -6 & 2 \end{bmatrix}\begin{bmatrix} 2 & 6 & 3 \\ 3 & 2 & -6 \\ 6 & -3 & 2 \end{bmatrix} =$$

$$\frac{1}{49}\begin{bmatrix} 49 & 0 & 0 \\ 0 & 49 & 0 \\ 0 & 0 & 49 \end{bmatrix} = I_3$$

This result is no coincidence: The ijth entry of BA is the dot product of the ith row of B and the jth column of A. By definition of B, this is just the dot product of the ith column of A and the jth column of A. Since A is orthogonal, this product is 1 if $i = j$ and 0 otherwise.

Definition 5.3.5 The transpose of a matrix; symmetric and skew-symmetric matrices

Consider an $m \times n$ matrix A.

The transpose A^T of A is the $n \times m$ matrix whose ijth entry is the jith entry of A: The roles of rows and columns are reversed.

We say that a square matrix A is symmetric if $A^T = A$, and A is called skew-symmetric if $A^T = -A$.

EXAMPLE 5 If $A = \begin{bmatrix} 1 & 2 & 3 \\ 9 & 7 & 5 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & 9 \\ 2 & 7 \\ 3 & 5 \end{bmatrix}$.

EXAMPLE 6 The symmetric 2×2 matrices are those of the form $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$, for example, $A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$.

The symmetric 2×2 matrices form a three-dimensional subspace of $R^{2\times 2}$, with basis $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

The skew-symmetric 2×2 matrices are those of the form $A = \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}$, for example, $A = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$. These form a one-dimmensional space with basis $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

Note that the transpose of a (column) vector $\vec{v}$ is a row vector: If

$$\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \text{ then } \vec{v}^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}.$$

The transpose give us a convenient way to express the dot product of two (cloumn) vectors as a matrix product.

Fact 5.3.6
If $\vec{v}$ and $\vec{w}$ are two (column) vectors in R^n, then

$$\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w}.$$

For example,

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 2.$$

Fact 5.3.7

Consider an $n \times n$ matrix A. The matrix A is orthogonal if (and only if) $A^T A = I_n$ or, equivalently, if $A^{-1} = A^T$.

Proof

To justify this fact, write A in terms of its columns:

$$A = \begin{bmatrix} | & | & & | \\ \vec{v_1} & \vec{v_2} & \dots & \vec{v_n} \\ | & | & & | \end{bmatrix}$$

Then,

$$A^T A = \begin{bmatrix} - & \vec{v_1}^T & - \\ - & \vec{v_2}^T & - \\ & \vdots & \\ - & \vec{v_n}^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \vec{v_1} & \vec{v_2} & \dots & \vec{v_n} \\ | & | & & | \end{bmatrix} =$$

$$\begin{bmatrix} \vec{v_1} \cdot \vec{v_1} & \vec{v_1} \cdot \vec{v_2} & \dots & \vec{v_1} \cdot \vec{v_n} \\ \vec{v_2} \cdot \vec{v_1} & \vec{v_2} \cdot \vec{v_2} & \dots & \vec{v_2} \cdot \vec{v_n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v_n} \cdot \vec{v_1} & \vec{v_n} \cdot \vec{v_2} & \dots & \vec{v_n} \cdot \vec{v_n} \end{bmatrix}.$$

By Fact 5.3.3(b) this product is I_n if (and only if) A is orthogonal.

Summary 5.3.8 Orthogonal matrices
Consider an $n \times n$ matrix A. Then, the following statements are equivalent:

1. A is an orthogonal matrix.

2. The transformation $L(\vec{x}) = A\vec{x}$ preserves length, that is, $\|A\vec{x}\| = \|\vec{x}\|$ for all $\vec{x}$ in R^n.

3. The columns of A form an orthonormal basis of R^n.

4. $A^T A = I_n$.

5. $A^{-1} = A^T$.

Fact 5.3.9 Properties of the transpose

a. If A is an $m \times n$ matrix and B an $n \times p$ matrix, then

$$(AB)^T = B^T A^T.$$

Note the order of the factors.

b. If an $n \times n$ matrix A is invertible, then so is A^T, and

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

c. For any matrix A,

$$rank(A) = rank(A^T).$$

Proof

a. Compare entries:

ijth entry of $(AB)^T$=jith entry of AB
=(jth row of A)·(ith column of B)

ijth entry of $B^T A^T$=(ith row of B^T)·(jth column of A^T)
=(ith column of B)·(jth row of A)

b. We know that

$$AA^{-1} = I_n$$

Transposing both sides and using part(a), we find that

$$(AA^{-1})^T = (A^{-1})^T A^T = I_n.$$

By Fact 2.4.9, it follows that

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

c. Consider the row space of A (i.e., the span of the rows of A). It is not hard to show that the dimmension of this space is rank(A) (see Exercise 49-52 in section 3.3):

rank(A^T)=dimension of the span of the columns of A^T
=dimension of the span of the rows of A
=rank(A)

The Matrix of an Orthogonal projection

The transpose allows us to write a formula for the matrix of an orthogonal projection. Consider first the orthogonal projection

$$proj_L\vec{x} = (\vec{v_1} \cdot \vec{x})\vec{v_1}$$

onto a line L in R^n, where $\vec{v_1}$ is a unit vector in L. If we view the vector $\vec{v_1}$ as an $n \times 1$ matrix and the scalar $\vec{v_1} \cdot \vec{x}$ as a 1×1, we can write

$proj_L\vec{x} = \vec{v_1}(\vec{v_1} \cdot \vec{x})$
$= \vec{v_1}\vec{v_1}^T\vec{x}$
$= M\vec{x}$,

where $M = \vec{v_1}\vec{v_1}^T$. Note that $\vec{v_1}$ is an $n \times 1$ matrix and $\vec{v_1}^T$ is $1 \times n$, so that M is $n \times n$, as expected.

More generally, consider the projection

$$proj_v\vec{x} = (\vec{v_1} \cdot \vec{x})\vec{v_1} + \cdots + (\vec{v_m} \cdot \vec{x})\vec{v_m}$$

onto a subspace V of R^n with orthonormal basis $\vec{v_1}$,...,$\vec{v_m}$. We can write

$$proj_v\vec{x} = \vec{v_1}\vec{v_1}^T\vec{x} + \cdots + \vec{v_m}\vec{v_m}^T\vec{x}$$
$$= (\vec{v_1}\vec{v_1}^T + \cdots + \vec{v_m}\vec{v_m}^T)\vec{x}$$

$$= \begin{bmatrix} | & & | \\ \vec{v_1} & \ldots & \vec{v_m} \\ | & & | \end{bmatrix} \begin{bmatrix} - & \vec{v_1}^T & - \\ & \vdots & \\ - & \vec{v_m}^T & - \end{bmatrix} \vec{x}$$

Fact 5.3.10 The matrix of an orthogonal projection

Consider a subspace V of R^n with orthonormal basis $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_m}$. The matrix of the orthogonal projection onto V is

$$AA^T, \text{ where } A = \begin{bmatrix} | & | & & | \\ \vec{v_1} & \vec{v_2} & \ldots & \vec{v_m} \\ | & | & & | \end{bmatrix}.$$

Pay attention to the order of the factors (AA^T as opposed to A^TA).

EXAMPLE 7 Find the matrix of the orthogonal projection onto the subspace of R^4 spanned by

$$\vec{v_1} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \vec{v_2} = \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}.$$

Solution

Note that the vectors $\vec{v_1}$ and $\vec{v_2}$ are orthonormal. Therefore, the matrix is

$$AA^T = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Exercises 5.3: 1, 3, 5, 11, 13, 15, 20

5.4 LEAST SQUARES AND DATA FITTING

ANOTHER CHARACTERIZATION OF ORTHOGONAL COMPLEMENTS

Consider a subspace V = im(A) of R^n, where $A = \left[\begin{array}{cccc} \vec{v_1} & \vec{v_2} & \ldots & \vec{v_m} \end{array} \right]$. Then,

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

$$= \{\ \vec{x} \text{ in } R^n : \vec{v_i} \cdot \vec{x} = 0, \text{ for } i = 1, \ldots, \text{m}\}$$

$$= \{\ \vec{x} \text{ in } R^n : \vec{v_i}^T \vec{x} = 0, \text{ for } i = 1, \ldots, \text{m}\}$$

In other words, $V^\perp$ is the kernel of the matrix

$$A^T = \left[\begin{array}{ccc} - & \vec{v_1}^T & - \\ - & \vec{v_2}^T & - \\ & \vdots & \\ - & \vec{v_m}^T & - \end{array} \right]$$

Fact 5.4.1 For any matrix A,

$$(\text{im A})^{\perp} = \text{ker } (A^T).$$

Example: consider the line

$$V = im \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Then

$$V^{\perp} = ker[\ 1 \quad 2 \quad 3\ ]$$

is the plan with equation $x_1 + 2x_2 + 3x_3 = 0$. (See Figure 1)

Fact 5.4.2 Consider a subspace V of R^n. then,

a. $dim(V) + dim(V^{\perp}) = n$

b. $(V^{\perp})^{\perp} = V$

c. $V \cap V^{\perp} = \{\vec{0}\}$

Proof

a. Let V = im(A) and $ker(A^T) = V^{\perp}$. Fact 3.3.9 tells us that n = dim(imA^T) + dim($kerA^T$) = rank(A^T) + dim($V^{\perp}$) =rank(A)+dim($V^{\perp}$) =dim(V)+dim($V^{\perp}$) by Fact 5.3.9.

b. First observe that $V \subseteq (V^{\perp})^{\perp}$, since a vector in V is orthogonal to every vector in $V^{\perp}$ (by definition of $V^{\perp}$). Furthermore, the dimensions of the two spaces are equal, by part(a):

$$dim(V^{\perp})^{\perp} = n - dim(V^{\perp})$$

$$= n - (n - dim(V))$$

$$= dim(V).$$

It follows that the two spaces are equal. (See Exercise 3.3.41.)

c. If $\vec{x}$ is in V and in $V^{\perp}$, then $\vec{x}$ is orthogonal to itself; that is, $\vec{x} \cdot \vec{x} = \|\vec{x}\|^2 = 0$, and thus $\vec{x} = \vec{0}$.

Fact 5.4.3
a. If A is an $m * n$ matrix, then

$$ker(A) = ker(A^T A).$$

b. If A is an $m * n$ matrix with $ker(A) = \{\vec{0}\}$, then $A^T A$ is invertible.

Proof
a. Clearly, the kernel of A is contained in the kernel kernel of $A^T A$. Conversely, consider a vector $\vec{x}$ in the kernel of $A^T A$, so that $A^T A\vec{x} = \vec{0}$ Then, $A\vec{x}$ is in the image of A and in the kernel of A^T. Since ker(A^T) is the orthogonal complement of $im(A)$ by Fact 5.4.1, the vector $A\vec{x}$ is $\vec{0}$ by Fact 5.4.2(c), that is, $\vec{x}$ is in the kernel of A.

b. Note that $A^T A$ is an $n * n$ matrix. By part (a), ker($A^T A$) $= \{\vec{0}\}$, and $A^T A$ is therefore invertible.(See Summary 3.3.11)

An Alternative Characterization of Orthogonal Projections

Fact 5.4.4
Consider a vector $\vec{x}$ in R^n and a subspace V of R^n. Then, the orthogonal projection proj$_V\vec{x}$ is the vector in V $closest$ to $\vec{x}$, in that

$$\|\vec{x} - proj_V\vec{x}\| < \|\vec{x} - \vec{v}\|,$$

for all $\vec{v}$ in V different from $proj_V\vec{x}$

Least-Squares Approximations

Definition 5.4.5 Least-squares solution
Consider a linear system

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

where A is an $m*n$ matrix. A vector $\vec{x}^*$ in R^n is called a $least - squares\ solution$ of this system if $\|\vec{b} - A\vec{x}^*\| \leq \|\vec{b} - A\vec{x}\|$ for all $\vec{x}$ in R^n.

The vector $\vec{x}^*$ is a least-square solution of the system $A\vec{x} = \vec{b}$

$\Updownarrow$ Def 5.4.5

$\|\vec{b} - A\vec{x}^*\| \leq \|\vec{b} - A\vec{x}\|$ for all $\vec{x}$ in R^n.

$\Updownarrow$ Def 5.4.4

$A\vec{x}^* = proj_V b$, where $V = im(A)$

$\Updownarrow$ Fact 5.1.6 and 5.4.1

$\vec{b} - A\vec{x}^*$ is in $V^\perp = im(A)^\perp = ker(A^T)$

$\Updownarrow$

$A^T(\vec{b} - A\vec{x}^*) = \vec{0}$

$\Updownarrow$

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

Fact 5.4.6 The normal equation
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 $A\vec{x} = \vec{b}$

Fact 5.4.7
If ker(A) $= \{\vec{0}\}$, then the linear system

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

has the unique least-squares solution

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

Example 1 Use Fact 5.4.7 to find the least-squares solution $\vec{x}^*$ of the system

$$A\vec{x} = \vec{b}, \text{ where } A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \text{ and } \vec{b} \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}$$

what is the geometric relationship between $A\vec{x}^*$ and $\vec{b}$?

Solution We compute

$$\vec{x}^* = (A^TA)^{-1}A^T\vec{b} = \begin{bmatrix} -4 \\ 3 \end{bmatrix} \text{ and } A\vec{x}^* = \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$$

Recall that A$\vec{x}^*$ is the orthogonal projection of $\vec{b}$ onto the image of A.

Fact 5.4.8 The matrix of an orthogonal projection

Consider a subspace V of R^n with basis $\vec{v}_1$, $\vec{v}_2$,..., $\vec{v}_m$. Let

$$A = \left[\begin{array}{cccc} \vec{v}_1, & \vec{v}_2, & ..., & \vec{v}_m \end{array} \right]$$

Then the matrix of the orthogonal projection onto V is

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

This means we are not required to find an *orthonormal* basis of V here. If the vectors $\vec{v}_i$ happen to be orthonormal, then $A^T A = I_m$ and the formula simplifies to $A^T A$. (See Fact 5.3.10.)

Example 2 Find the matrix of the orthogonal projection onto the subspace of R^4 spanned by the vector

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} and \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Solution Let

$$\mathrm{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix},$$

and compute

$$\mathrm{A}(A^T A)^{-1} A^T = \begin{bmatrix} 7 & 4 & 1 & -2 \\ 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ -2 & 1 & 4 & 7 \end{bmatrix}$$

Data Fitting Scientists are often interested in fitting a function of a certain type to data they have gathered. The functions considered could be linear, polynomial, relational' trigonometric, or exponential. The equations we have to solve as we fit data are frequently linear. (See Exercises 29 and 30 of section 1.1, and Exercises 30 through 33 of Section 1.2.)

Example 3 Find a cubic polyonmial whose graph passes through the points (1, 3), (-1, 13), (2, 1), (-2, 33).

Solution We are looking for a function

$$\mathrm{f(t)} = c_0 + c_1\mathrm{t} + c_2 t^2 + c_3 t^3$$

such that f(1) = 3, f(-1) = 13, f(2) = 1, f(-2) = 33; that is, we have to solve the linear system

$$\left| \begin{array}{rcrcrcrcr} c_0 & + & c_1 & + & c_2 & + & c_3 & = & 3 \\ c_0 & - & c_1 & + & c_2 & - & c_3 & = & 13 \\ c_0 & + & 2c_1 & + & 4c_2 & + & 8c_3 & = & 1 \\ c_0 & - & 2c_1 & + & 4c_2 & - & 8c_3 & = & 33 \end{array} \right|$$

This linear system has the unique solution

$$\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -4 \\ 3 \\ -1 \end{bmatrix}.$$

Thus, the cubic polynomial whose graph passes through the four given data points if $f(t) = 5 - 4t + 3t^2 - t^3$, as shown in Figure 6.

Example 4 Fit a quadratic function to the four data points (a_1, b_1) = (-1, 8), (a_2, b_2) = (0, 8), (a_3, b_3) = (1, 4), and (a_4, b_4) = (2, 16).

Solution We are looking for a function f(t) = $c_0 + c_1$t $+ C_2 t^2$ such that

$$\left| \begin{array}{l} f(a_1) = b_1 \\ f(a_2) = b_2 \\ f(a_3) = b_3 \\ f(a_4) = b_4 \end{array} \right| \text{ or } \left| \begin{array}{llll} c_0 & -c_1 + c_2 & & = 8 \\ c_0 & & & = 8 \\ c_0 & +c_1 + c_2 & & = 4 \\ c_0 & +2c_1 + 4c_2 & & = 16 \end{array} \right| \text{ or } A \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}$$

where

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} 8 \\ 8 \\ 4 \\ 16 \end{bmatrix}.$$

We have four equations, corresponding to the four data points, but only three unknowns, the three coefficients of a quadratic polynomial. Check that this system is indeed inconsistent. The least-squares solution is

$$\vec{x}^* = \begin{bmatrix} c_0^* \\ c_1^* \\ c_2^* \end{bmatrix} = (A^T A)^{-1} A^T \vec{b} = \begin{bmatrix} 5 \\ -1 \\ 3 \end{bmatrix}$$

The least-squares approximation is $f^*(t) = 5 - t + 3t^2$, as shown in Figure 7. This quadratic function f^*(t) fits the data points best, in that the vector

$$\text{A}\vec{x}^* = \begin{bmatrix} f^*(a_1) \\ f^*(a_2) \\ f^*(a_3) \\ f^*(a_4) \end{bmatrix}$$

is close as possible to

$$A = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}.$$

This means that
$\|\vec{b} - A\vec{c}^*\|^2 = (b_1 - f^*(a_1))^2 + (b_2 - f^*(a_2))^2 + (b_3 - f^*(a_3))^2 + (b_4 - f^*(a_4))^2$
is minimal: The sum of the squares of the vertical distances between graph and data points is minimal. (See Figure 8.)

Example 5 Find the linear function $c_0 + c_1$t that best fits the data points (a_1, b_1), (a_2, b_2),...,(a_n, b_n), use least squares. Assume that $a_1 \neq a_2$.

Solution We attempt to solve the system

$$\left| \begin{array}{ccccc} c_0 & + & c_1 a_1 & = & b_1 \\ c_0 & + & c_1 a_2 & = & b_2 \\ \vdots & & \vdots & & \vdots \\ c_0 & + & c_1 a_n & = & b_n \end{array} \right|$$

or

$$\begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

or

$$\mathsf{A} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \vec{b}$$

Note that rank(A) = 2, since $a_1 \neq a_2$. The least-squares solution is

$$\left| \begin{array}{c} c_0^* \\ c_1^* \end{array} \right| = (A^T A)^{-1} A^T \vec{b} =$$

$$\left(\begin{bmatrix} 1 & \dots & 1 \\ a_1 & \dots & a_n \end{bmatrix} \begin{bmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \dots & 1 \\ a_1 & \dots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

$$= \begin{bmatrix} n & \Sigma_i a_i \\ \Sigma_i a_i & \Sigma_i a_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_i b_i \\ \Sigma_i a_i b_i \end{bmatrix}$$

(where $\sum_i$ refers to the sum for i = 1,...,n)

We have found that

$$C_0^* = \frac{(\sum_i a_i^2)(\sum_i b_i) - (\sum_i a_i)(\sum_i a_i b_i)}{n(\sum_i a_i^2) - (\sum_i a_i)^2},$$

$$C_1^* = \frac{n(\sum_i a_i b_i) - (\sum_i a_i)(\sum_i b_i)}{n(\sum_i a_i^2) - (\sum_i a_i)^2}.$$

There formulas are well known to statisticians. There is no need to memorize them.

Example 6 In the accompanying table, we list the scores of five students in the three exams given in a class.

Find the function of the form $f = c_0 + c_1 h + c_2 m$ that best fits these data, using least squares. what score f does your formula predict for Marlisa, another student, whose scores in the first two exams were $h = 92$ and $m = 72$?

Solution

We attempt to solve the system

$$\left| \begin{array}{ccccccc} c_0 & + & 76c_1 & + & 48c_2 & = & 43 \\ c_0 & + & 92c_1 & + & 92c_2 & = & 90 \\ c_0 & + & 68c_1 & + & 82c_2 & = & 64 \\ c_0 & + & 86c_1 & + & 68c_2 & = & 69 \\ c_0 & + & 54c_1 & + & 70c_2 & = & 50 \end{array} \right|.$$

The least-squares solution is

$$\left| \begin{array}{c} c_0^* \\ c_1^* \\ c_2^* \end{array} \right| . = (A^T A)^{-1} A^T \vec{b} \approx \left| \begin{array}{c} -42.4 \\ 0.639 \\ 0.799 \end{array} \right| .$$

The function which gives the best fit is approximately

$$f = \text{-}42.4 + 0.639h + 0.799m.$$

The formula predicts the score

$$f = \text{-}42.4 + 0.639 \cdot 92 + 0.799 \cdot 72 \approx 74.$$

for Marlisa.