5.4 LEAST SQUARES AND DATA FITTING

ANOTHER CHARACTERIZATION OF ORTHOGONAL COMPLEMENTS

Consider a subspace V = im(A) of R^n, where $A = \begin{bmatrix} \vec{v_1} & \vec{v_2} & ... & \vec{v_m} \end{bmatrix}$. Then,

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

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

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

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

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

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

$$V^{\perp} = ker[123]$$

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 T($\vec{x}$) = $proj_V$ be the orthogonal projection onto V. Note that im(A) = V and ker(T) = $V^{\perp}$. Fact 3.3.9 tells us that n = dim(ker T) + dim(im T) = dim(V) + dim($V^{\perp}$).

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):

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

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

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

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 $\vec{a}^*$ 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

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

$\Updownarrow$ Def 5.4.4

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

$\Updownarrow$ Fact 5.1.6 and 5.4.1

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

$$\Updownarrow$$

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

$$\Updownarrow$$

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

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

$$A = \left[\ \vec{v}_1,\ \ \vec{v}_2,\ \ ...,\ \ \vec{v}_m\ \right]$$

Then the matrix of the orthogonal projection onto V is

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

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 $A^TA = I_m$ and the formula simplifies to A^TA. (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

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

and compute

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

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

$$\text{f(t)} = c_0 + c_1\text{t} + c_2t^2 + c_3t^3$$

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

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

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 $f^*(\mathrm{t})$ fits the data points best, in that the vector

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

is close as possible to

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

This means that

$\|\vec{b} - \mathsf{A}\vec{c}^*\|^2 = (b_1 - f^*(a_1))^2 + (b_2 - f^*(a_2))^2 + (b_3 - f^*(a_3))^2 + (b_4 - f^*(a_4))^2$

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

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

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.