Cosines and Projections onto Lines; Projections and Least Squares (投影和最小二乘)

Lecture 14 and 15

Dept. of Math., SUSTech

2023.10

Projections and Least Squares

Projection

To find a point p on a subspace that is closest to a given point b, a perpendicular line from b to S meets the subspace at p. Questions:

- Does this projection actually arise in practical applications?
- If we have a basis for the subspace S, is there a formula for the projection p?

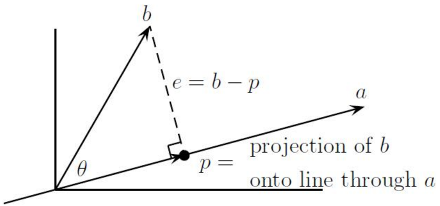

Figure 3.5: The projection p is the point (on the line through a) closest to b.

Inner Products and Cosines

Suppose the vectors $a = (a_1, a_2)$ and $b = (b_1, b_2)$ make angles α and β with the x-axis.

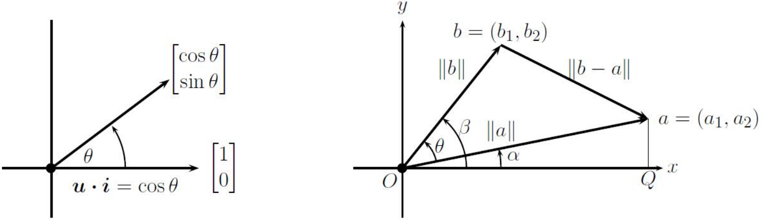

Figure 3.6: The cosine of the angle $\theta = \beta - \alpha$ using inner products.

Cosine Formula: $\cos\theta = \cos(\beta - \alpha) = \frac{a_1 b_1 + a_2 b_2}{||a||||b||}$.

The Cosine Formula

The numerator in this formula is exactly the inner product of a and b. It gives the relationship between $a^T b$ and $\cos\theta$. The cosine of the angle between any nonzero vectors a and b is

$$\textbf{Cosine of } \theta \quad \cos\theta = \frac{a^T b}{||a||||b||}.$$

Remarks:

- This formula is dimensionally correct; if we double the length of b, then both numerator and denominator are doubled, and the cosine is unchanged. Reversing the sign of b, on the other hand, reverses the sign of $\cos\theta$–and changes the angle by 180°.
- There is another law of trigonometry(Law of Cosines) that leads directly to the same result.

Projection onto a Line

The line from b to the closest point $p = \hat{x}a$ is perpendicular to the vector a.

Proposition

The projection of the vector b onto the line in the direction of a is

$$p = \hat{x}a = \frac{a^T b}{a^T a} a.$$

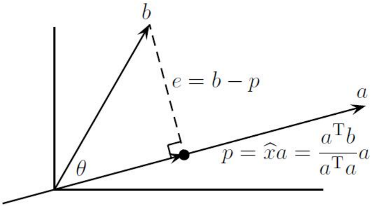

Figure 3.7: The projection p of b onto a, with $\cos\theta = \frac{Op}{Ob} = \frac{a^{\mathrm{T}}b}{\|a\|\|b\|}$.

Schwarz Inequality

Schwarz Inequality is the most important inequality in mathematics. All vectors a and b satisfy the **Schwarz Inequality**, which is $|\cos\theta| \leq 1$ in $\mathbb{R}^n$:

$$|a^T b| \leq ||a||||b||.$$

Remarks:

- The Schwarz inequality is the same as $|\cos\theta| \leq 1$.
- Equality holds if and only if b is a multiple of a.
- The name of Cauchy is also attached to this inequality, and the Russians refer to it as the Cauchy-Schwarz-Buniakowsky inequality! Mathematical historians seen to agree that Buniakowsky's claim is genuine.

Projection Matrix of Rank 1

The projection of b onto the line through a lies at $p = a(a^T b/a^T a)$. That is our formula $p = \hat{x}a$, but it is written with a slight twist: The vector is put before the number $\hat{x} = a^T b/a^T a$.

Projection onto a line is carried out by a projection matrix P, and written in this new order we can see what it is. P is the matrix that multiplies b and produces p:

$$P = \frac{aa^T}{a^T a}.$$

That is a column times a row–a square matrix–divided by the number $a^T a$.

It is also useful to note that the line from b to the closest point $p = \hat{x}a$ is perpendicular to the vector a.

Examples

1. Project $b = (1, 2, 3)$ onto the line through $a = (1, 1, 1)$ to get $\hat{x}$ and p:

$$\hat{x} = \frac{a^T b}{a^T a} = \frac{6}{3} = 2.$$

   The projection is $p = \hat{x} a = (2, 2, 2)$.

2. The matrix that projects onto the line through $a = (1, 1, 1)$ is

$$P = \frac{a a^T}{a^T a} = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

Projection matrix

The matrix has two properties that we will see as typical of projections:

1. P is a symmetric matrix.
2. Its square is itself: $P^2 = P$.

Example

Projection onto the "θ-direction" in the x-y plane. The line goes through $a = (\cos\theta, \sin\theta)$ and the matrix is symmetric with $P^2 = P$:

$$P = \frac{aa^T}{a^T a} = \frac{\begin{bmatrix} c \\ s \end{bmatrix} \begin{bmatrix} c & s \end{bmatrix}}{\begin{bmatrix} c & s \end{bmatrix} \begin{bmatrix} c \\ s \end{bmatrix}} = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix}.$$

Project onto the "θ-direction" in the x-y plane

In the above example, c is $\cos\theta$, s is $\sin\theta$, and $c^2 + s^2 = 1$ in the denominator. This matrix was discovered in Section 2.6 on linear transformation.

To repeat: P is the matrix that multiplies b and produces p.

Transposes from Inner Products

Definition

The transpose A^T can be defined by the following property: The inner product of Ax with y equals the inner product of x with A^Ty. Formally, this simply means that

$$(Ax)^Ty = x^TA^Ty = x^T(A^Ty).$$

The definition gives us another (better) way to verify that formula $(AB)^T = B^TA^T$. Use the above equation twice:

$$\textbf{Move } A \textbf{ then move } B \qquad (ABx)^Ty = (Bx)^T(A^Ty) = x^T(B^TA^Ty).$$

The transposes turn up in reverse order on the right side, just as the inverses do in the formula $(AB)^{-1} = B^{-1}A^{-1}$.

Normal Equations

Now we are ready for the serious step, to project b onto a subspace–rather than just onto a line. When $Ax = b$ is inconsistent, its least-squares solution minimizes $||Ax - b||^2$, and normal equations are

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

$A^T A$ is invertible exactly when the columns of A are linearly independent! Then, the best estimate $\hat{x}$ is

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

The projection of b onto the column space is the nearest point $A\hat{x}$:

$$p = A\hat{x} = A(A^T A)^{-1} A^T b.$$

Example

Example

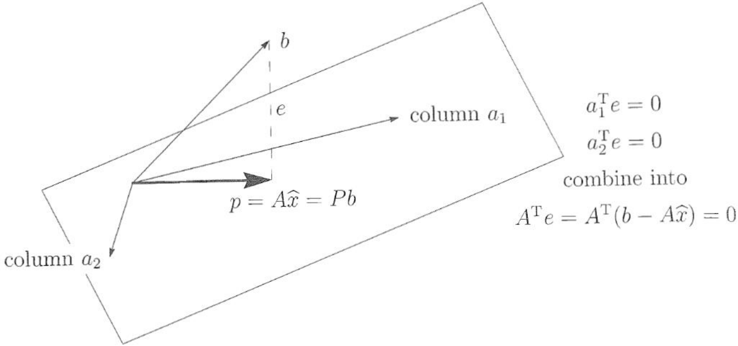

Figure 3.8: Projection onto the column space of a 3 by 2 matrix.

Example

Example

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix},$$

$Ax = b$ has no solution

$A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$ gives the best x.

Remarks

Remarks:

- If b is in the column space of A–it is a combination $b = Ax$ of the columns. Then the projection of b is still b: $Pb = b$.
- If b is perpendicular to every column of A, so $A^T b = 0$. In this case b projects to the zero vector: $Pb = 0$.
- If A is square and invertible, the column space is the whole space. Every vector projects to itself: $Pb = b$.
- Suppose A has only one column, containing a. Then we return to the earlier formula:

$$\hat{x} = \frac{a^T b}{a^T a}.$$

The Cross-Product Matrix of $A^T A$

The matrix $A^T A$ is certainly symmetric, the key question is the invertibility of it, and fortunately

Proposition

$A^T A$ has the same nullspace as A.

If A has independent columns and only 0 is in its nullspace. The same is true for $A^T A$:

Proposition

If A has independent columns, then $A^T A$ is square, symmetric, and invertible.

Projection Matrices

Projection matrix

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

Remarks:

- The projection matrix $P = A(A^TA)^{-1}A^T$ has two basic properties: (i) It equals its square: $P^2 = P$.
  (ii) It equals its transpose: $P^T = P$.
- Conversely, any symmetric matrix with $P^2 = P$ represents a projection.
- We deduce from $P^2 = P$ and $P^T = P$ that Pb is the projection of b onto the column space of P.
- The error vector $b - Pb$ is orthogonal to the column space. For any vector Pc in the column space, the inner product is zero.

The Problem

Suppose we do a series of experiments, and expect the output b to be a linear function of the input t. We look for a straight line $b = C + Dt$. For example:

1. At different times we measure the distance to a satellite on its way to Mars. In this case t is the time and b is the distance. Unless the motor was left on or gravity is strong, the satellite should move with nearly constant velocity v: $b = b_0 + vt$.
2. We vary the load on a structure, and measure the movement it produces. In this experiment t is the load and b is the reading from the strain gauge. Unless the load is so great that the material becomes plastic, a linear relation $b = C + Dt$ is normal in the theory of elasticity.

Question

3. The cost of producing t books like this one is nearly linear, $b = C + Dt$, with editing and typesetting in C and then printing and binding in D. C is the set up cost and D is the cost for each additional book.

How to compute C and D? If there is no experimental error, then two measurements of b will determine the line $b = C + Dt$. But if there is error, we must be prepared to "average" the experiments and find an optimal line.

Example

Example

Three measurements b_1, b_2, b_3 are marked on Figure 3.9a:

$$b = 1 \text{ at } t = -1, \quad b = 1 \text{ at } t = 1, \quad b = 3 \text{ at } t = 2.$$

Note that the values $t = -1, 1, 2$ are not required to be equally spaced. The first step is to write the equations that would hold if a line could go through all three points. Then every $C + Dt$ would agree exactly with b:

$$Ax = b \text{ is } \left\{ \begin{array}{rcrl} C & - & D & = 1 \\ C & + & D & = 1 \\ C & + & 2D & = 3 \end{array} \right. \text{ or } \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}.$$

Solution

If those equations $Ax = b$ could be solved , there would be no errors. They can't be solved because the points are not on a line. Therefore they are solved by least-squares:

$$A^T A\hat{x} = A^T b \text{ is } \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}.$$

The best solution is $\hat{C} = \frac{9}{7}, \hat{D} = \frac{4}{7}$ and the best line is

$$\frac{9}{7} + \frac{4}{7}t.$$

See the following figure for more details.

Figure 3.9

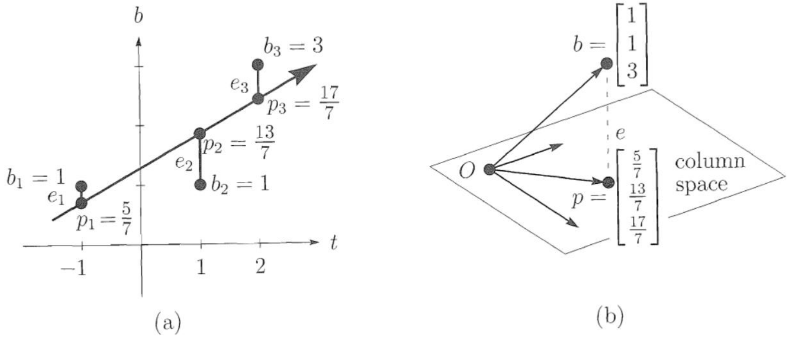

Figure 3.9: Straight-line approximation matches the projection p of b.

Note the beautiful connections between the two figures.

Conclusions

- The line $\frac{9}{7}+\frac{4}{7}t$ has heights $\frac{5}{7}, \frac{13}{7}, \frac{17}{7}$ at the measurements times $-1, 1, 2$. Those points do lie on a line.
- Subtracting p from b, the errors are $e=(\frac{2}{7}, -\frac{6}{7}, \frac{4}{7})$. Those are the vertical errors in Figure 3.9b. It is orthogonal to both columns of A.

Least-Squares Fitting of Data

We can quickly summarize the equations for fitting by a straight line. The first column of A contains 1s, and the second column contains the times t_i. Therefore $A^T A$ contains the sum of the 1s and the t_i and the t_i^2:

Theorem

The measurements $b_1, b_2, \cdots, b_m$ are given at distinct points $t_1, t_2, \cdots, t_m$. Then the straight line $\widehat{C} + \widehat{D}t$ which minimizes E^2 comes from least-squares:

$$A^T A \begin{bmatrix} \widehat{C} \\ \widehat{D} \end{bmatrix} = A^T b \text{ or } \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} \widehat{C} \\ \widehat{D} \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}$$

Weighted Least Squares

If the observations are not trusted to the same degree, we need to minimize the weighted sum of least squares:

Theorem

The least squares solution to $WAx = Wb$ is $\hat{x}_W$, and the weighted normal equations:

$$(A^T W^T WA)\hat{x}_W = A^T W^T Wb.$$

- Weighted average, Variance, Covariance matrix, etc.
- For any invertible matrix W, we can define a new inner product and length as follows:

$$(x,y)_W = (Wy)^T(Wx) \quad \text{and} \quad ||x||_W = ||Wx||.$$

Homework Assignment 14 and 15

3.2: 1, 3, 10, 11, 16, 17, 18, 19, 21, 24.

3.3: 1, 3, 6, 11, 13, 15, 21, 35.