MA 511, Session 15

Inner Products and Projections

Recall from analytic geometry that, for $a, b \in \mathbb{R}^2$,

$$a^T b = \|a\| \|b\| \cos\theta,$$

where $0 \leq \theta \leq \pi$ is the angle between a and b.

We shall find now the closest point on a given line to a given point outside the line. We shall use the geometry of $\mathbb{R}^2$ but the formulas are the same in $\mathbb{R}^n$.

Let $a, b \in \mathbb{R}^2$ be linearly independent, so that b is not on the line $l = \{ca, c \in \mathbb{R}\}$. Clearly, the point p on l that is closest to b makes the segment (vector) $b - p$ perpendicular to l, i.e. $a^T(b - p) = 0$. Since $p \in l$, it follows $p = \bar{x}a$ and, therefore, $a^T(b - \bar{x}a) = 0$ gives $a^T b = \bar{x}a^T a$, so that

$$\bar{x} = \frac{a^T b}{a^T a} \quad \text{and} \quad p = \frac{a^T b}{a^T a} a.$$

From this relation we derive Schwarz's inequality,

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

<u>Proof</u>: Note that

$$0 \leq \left\| b - \frac{a^T b}{a^T a} a \right\|^2 = b^T b - 2 \frac{(a^T b)^2}{a^T a} + \left(\frac{a^T b}{a^T a} \right)^2 a^T a$$
$$= \frac{b^T b a^T a - (a^T b)^2}{a^T a},$$

which implies $0 \leq b^T b a^T a - (a^T b)^2$, as needed.

We also see that equality holds if, and only if b is a multiple of a.

We can easily write the matrix that represents the projection onto the line l of direction a with respect to the standard basis of $\mathbb{R}^n$, $B = \{e_1, \ldots, e_n\}$, where $(e_i)_j = \delta_{ij}$ is the Kronecker symbol (i.e. the standard basis of $\mathbb{R}^n$ consists of the columns of the identity matrix): $P = \frac{aa^T}{a^T a}$.
Indeed, we then have

$$Pb = \frac{aa^T}{a^T a} b = \frac{a(a^T b)}{a^T a} = a \frac{a^T b}{a^T a} = a\bar{x} = \bar{x} a.$$

Example: Find the matrix P that projects $\mathbb{R}^2$ onto the line $y = 3x$.

We take $a = \vec{i} + 3\vec{j} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$. Then,

$$P = \frac{\begin{pmatrix} 1 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 3 \end{pmatrix}}{\begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix}} = \frac{1}{10} \begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix}$$

Remark: Matrices P representing projections onto lines have several important properties.

(i) P is symmetric

(ii) $P^2 = P$

(iii) rank $P = 1$

(iv) $\mathcal{C}(P)$ is the line through a and the origin

(v) $\mathcal{N}(P)$ is the hyperplane (line if $n = 2$) orthogonal to that line

Remark: The identity $a^T b = \|a\|\|b\| \cos\theta$ is valid for vectors in $\mathbb{R}^n$ since, when a and b are not parallel they determine a plane through the origin, and θ is the angle between them in that plane. When a and b are parallel, the relation holds automatically, since $\theta = 0$ makes $\cos\theta = 1$, and $\theta = \pi$ makes $\cos\theta = -1$.

We shall use the notion of projection for least squares approximation. Suppose we have a "model" for a physical process which is linear

$$Ax = b$$

and we want to determine the value of $x \in \mathbb{R}^n$ by collecting data for the coefficients of A and for b. Assume that the measurements are affected by errors, so that the resulting system may be inconsistent. We would then like to find a "solution" x that best fits the data in some sense.

Suppose for a moment that x is a scalar quantity. The problem then looks like

$$\begin{cases} a_1 x = b_1 \\ a_2 x = b_2 \\ \quad \cdots\cdots \\ a_m x = b_m \end{cases}$$

The numbers $a_1, \ldots, a_m$ and $b_1, \ldots, b_m$ may be thought of as observations, and the objective is to find the number x that best fits these observations. The most common way to do this is to use the <u>mean square</u> error E,

$$E^2 = (a_1 x - b_1)^2 + \cdots + (a_m x - b_m)^2 = \|ax - b\|^2.$$

Taking the derivative with respect to x to find the value of x that minimizes E, we obtain

$$2(a_1 x - b_1)a_1 + \cdots + 2(a_m x - b_m)a_m = 0$$

which implies

$$a_1 a_1 x + \cdots + a_m a_m x = a_1 b_1 + \cdots + a_m b_m.$$

We now see that the optimal value of x is given as

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

Let us look at the general case. Now A is a $m \times n$ matrix, with $m > n$ corresponding to having many rows of data corresponding to lots of experimental measurements. Assume that the system $Ax = b$ is inconsistent, but assume also that A is of full rank, i.e. rank $A = n$. To find the best "solution" $x \in \mathbb{R}^n$ to $Ax = b$, we again minimize the error

$$E = \|Ax - b\|.$$

Geometrically, we see that the minimum value for $\|Ax - b\|$ is attained when Ax is the (orthogonal) projection of b onto the column space of A. This means we need $\bar{x} \in \mathbb{R}^n$ such that $b - A\bar{x} \quad \perp \quad \mathcal{C}(A)$, that is $b - A\bar{x} \in \mathcal{N}(A^T)$, so that $A^T(b - A\bar{x}) = 0$. Finally, we obtain the normal equation

$$A^T A\bar{x} = A^T b,$$

which is uniquely solvable when A has full rank, since then the $n \times n$ matrix $A^T A$ has the same rank, n. Then,

$$\bar{x} = (A^T A)^{-1} A^T b,$$

and the projection of b onto $\mathcal{C}(A)$ is

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