Projections onto subspaces

Projections

If we have a vector $\mathbf{b}$ and a line determined by a vector $\mathbf{a}$, how do we find the point on the line that is closest to $\mathbf{b}$?

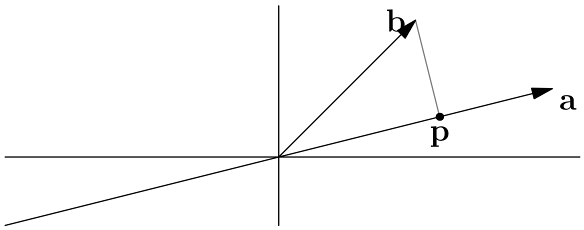

Figure 1: The point closest to $\mathbf{b}$ on the line determined by $\mathbf{a}$.

We can see from Figure 1 that this closest point $\mathbf{p}$ is at the intersection formed by a line through $\mathbf{b}$ that is orthogonal to $\mathbf{a}$. If we think of $\mathbf{p}$ as an approximation of $\mathbf{b}$, then the length of $\mathbf{e} = \mathbf{b} - \mathbf{p}$ is the error in that approximation.

We could try to find $\mathbf{p}$ using trigonometry or calculus, but it's easier to use linear algebra. Since $\mathbf{p}$ lies on the line through $\mathbf{a}$, we know $\mathbf{p} = x\mathbf{a}$ for some number x. We also know that $\mathbf{a}$ is perpendicular to $\mathbf{e} = \mathbf{b} - \mathbf{xa}$:

$$\begin{aligned} \mathbf{a}^T(\mathbf{b} - x\mathbf{a}) &= 0 \\ x\mathbf{a}^T\mathbf{a} &= \mathbf{a}^T\mathbf{b} \\ x &= \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}, \end{aligned}$$

and $\mathbf{p} = \mathbf{a}x = \mathbf{a}\dfrac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$. Doubling $\mathbf{b}$ doubles $\mathbf{p}$. Doubling $\mathbf{a}$ does not affect $\mathbf{p}$.

Projection matrix

We'd like to write this projection in terms of a *projection matrix* P: $\mathbf{p} = P\mathbf{b}$.

$$\mathbf{p} = \mathbf{xa} = \frac{\mathbf{aa}^T\mathbf{a}}{\mathbf{a}^T\mathbf{a}},$$

so the matrix is:

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

Note that $\mathbf{aa}^T$ is a three by three matrix, not a number; matrix multiplication is not commutative.

The column space of P is spanned by $\mathbf{a}$ because for any $\mathbf{b}$, $P\mathbf{b}$ lies on the line determined by $\mathbf{a}$. The rank of P is 1. P is symmetric. $P^2\mathbf{b} = P\mathbf{b}$ because the projection of a vector already on the line through $\mathbf{a}$ is just that vector. In general, projection matrices have the properties:

$$P^T = P \qquad \text{and} \qquad P^2 = P.$$

Why project?

As we know, the equation $A\mathbf{x} = \mathbf{b}$ may have no solution. The vector $A\mathbf{x}$ is always in the column space of A, and $\mathbf{b}$ is unlikely to be in the column space. So, we project $\mathbf{b}$ onto a vector $\mathbf{p}$ in the column space of A and solve $A\hat{\mathbf{x}} = \mathbf{p}$.

Projection in higher dimensions

In $\mathbb{R}^3$, how do we project a vector $\mathbf{b}$ onto the closest point $\mathbf{p}$ in a plane?

If $\mathbf{a}_1$ and $\mathbf{a}_2$ form a basis for the plane, then that plane is the column space of the matrix $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$.

We know that $\mathbf{p} = \hat{x}_1\mathbf{a}_1 + \hat{x}_2\mathbf{a}_2 = A\hat{\mathbf{x}}$. We want to find $\hat{\mathbf{x}}$. There are many ways to show that $\mathbf{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the plane we're projecting onto, after which we can use the fact that $\mathbf{e}$ is perpendicular to $\mathbf{a}_1$ and $\mathbf{a}_2$:

$$\mathbf{a}_1^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \text{and} \quad \mathbf{a}_2^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0.$$

In matrix form, $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$. When we were projecting onto a line, A only had one column and so this equation looked like: $a^T(\mathbf{b} - x\mathbf{a}) = \mathbf{0}$.

Note that $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ is in the nullspace of A^T and so is in the left nullspace of A. We know that everything in the left nullspace of A is perpendicular to the column space of A, so this is another confirmation that our calculations are correct.

We can rewrite the equation $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ as:

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

When projecting onto a line, A^TA was just a number; now it is a square matrix. So instead of dividing by $\mathbf{a}^T\mathbf{a}$ we now have to multiply by $(A^TA)^{-1}$

In n dimensions,

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ \mathbf{p} = A\hat{\mathbf{x}} &= A(A^TA)^{-1}A^T\mathbf{b} \\ P &= A(A^TA)^{-1}A^T. \end{aligned}$$

It's tempting to try to simplify these expressions, but if A isn't a square matrix we can't say that $(A^TA)^{-1} = A^{-1}(A^T)^{-1}$. If A does happen to be a square, invertible matrix then its column space is the whole space and contains $\mathbf{b}$. In this case P is the identity, as we find when we simplify. It is still true that:

$$P^T = P \qquad \text{and} \qquad P^2 = P.$$

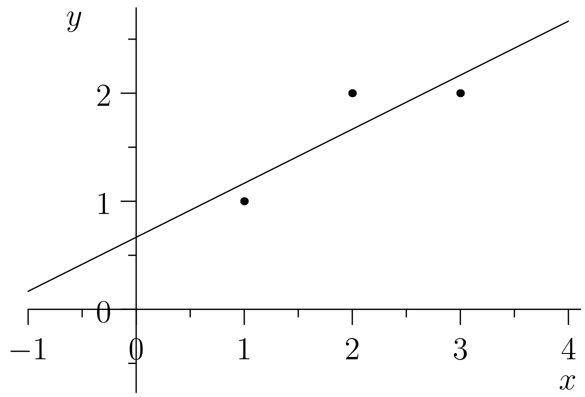

Figure 2: Three points and a line close to them.

Least Squares

Suppose we're given a collection of data points (t, b):

$$\{(1,1),(2,2),(3,2)\}$$

and we want to find the closest line $b = C + Dt$ to that collection. If the line went through all three points, we'd have:

$$\begin{aligned} C + D &= 1 \\ C + 2D &= 2 \\ C + 3D &= 2, \end{aligned}$$

which is equivalent to:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}}_{\mathbf{b}}.$$

In our example the line does not go through all three points, so this equation is not solvable. Instead we'll solve:

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