MATH21B – LECTURE 15: LEAST SQUARES AND DATA FITTING
SPRING 2018, HARVARD UNIVERSITY

1. Data fitting

Problem 1. Find the coefficients a, b, c of the quadratic polynomial $p(t) = at^2 + bt + c$ which best fits the four data points $(-1, 8)$, $(0, 8)$, $(1, 4)$ and $(2, 16)$.

Solution. The data points give a system of linear equations for a, b, c by taking $p(-1) = 8$, etc. In matrix form, these look like

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 4 \\ 16 \end{bmatrix}.$$

The matrix A^TA and vector $A^T\vec{b}$ are given by

$$A^TA = \begin{bmatrix} 18 & 8 & 6 \\ 8 & 6 & 2 \\ 6 & 2 & 4 \end{bmatrix}, \quad A^T\vec{b} = \begin{bmatrix} 76 \\ 28 \\ 36 \end{bmatrix}$$

To solve $A^TA\vec{x}^* = A^T\vec{b}$ take $\mathrm{rref}([A^TA \mid A^T\vec{b}])$ to obtain

$$\vec{x}^* = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}.$$

See Figure 1 for the result. ■

Problem 2 (Hard)**.** Show that $\mathrm{im}(AA^T) = \mathrm{im}(A)$.

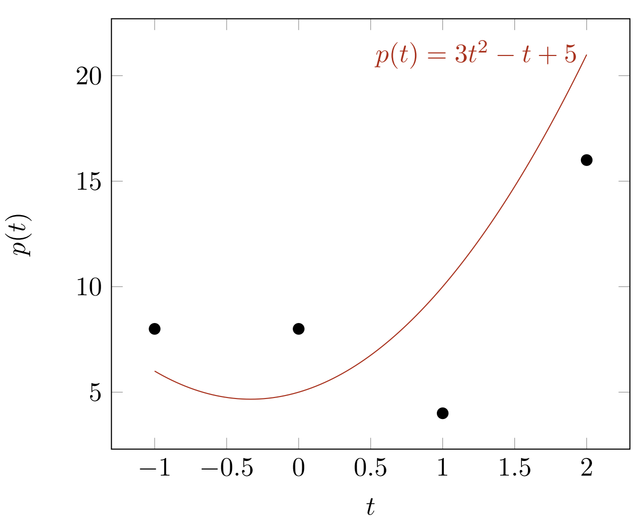

Figure 1. The best fit quadratic polynomial in Problem 1.

Solution. Clearly $\text{im}(AA^T) \subset \text{im}(A)$, so it suffices to show that $\text{im}(A) \subset \text{im}(AA^T)$. This is a consequence of $\ker(A)^\perp = \text{im}(A^T)$, as we can write $A\vec{x}$ as $A(\vec{y} + \vec{z})$ with $\vec{y} \in \ker(A)$ and $\vec{z} = \ker(A)^\perp$ and $A\vec{y} = 0$ while $\vec{z} = A^T\vec{w}$.

So it remains to show that $\ker(A)^\perp = \text{im}(A^T)$. We can take orthogonal complements and instead prove $\ker(A) = \text{im}(A^T)^\perp$. This follows since $\vec{x} \in \ker(A)$ if and only if $\vec{x}$ is orthogonal to the rows of A, which is the same as being orthogonal to the rows of A^T, which is the same as being orthogonal to the image of A^T. ■

2. Orthogonal projections revisited

Problem 3. Let the matrix A be given by

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

Find the matrix for orthogonal projection onto $\text{im}(A)$.

Solution. We use the formula $A(A^TA)^{-1}A^T$. Let us compute

$$A^TA = \begin{bmatrix} 6 & -1 \\ -1 & 1 \end{bmatrix} \qquad (A^TA)^{-1} = \begin{bmatrix} 1/5 & 1/5 \\ 1/5 & 6/5 \end{bmatrix}$$

and thus we have that

$$A(A^TA)^{-1}A^T = \begin{bmatrix} 1/5 & 2/5 & 0 \\ 2/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

■

Summary

- The *least squares solution* to a system of linear equations $A\vec{x} = \vec{b}$ is a solution $\vec{x}^*$ of $A^TA\vec{x}^* = A^T\vec{b}$. These always exist and are unique if $\ker(A) = \{0\}$, in which case A^TA is invertible and then given by $\vec{x}^* = (A^TA)^{-1}A^T\vec{b}$.
- If $\ker(A) = \{0\}$ then A^TA is invertible and the matrix for orthogonal projection onto $\text{im}(A)$ is given by $A(A^TA)^{-1}A^T$. This simplifies to the usual formula when the columns of A are orthonormal.