Lecture 19: 5.3 Least square solutions. A standard statistical technique is to find a least square fit to data points in the plane by some simple curve e.g. a line. Since there might be errors in the measurements of the data we do not require the curve to pass through the points but instead be such that it is the optimal approximation to the data in the sense that the sum of squares of the error between the y values of the data points and the points on the curve should be minimized.

A least square problem may be formulated as an overdetermined linear system. A system with more equations than unknowns usually is inconsistent. Given a system $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix with $m > n$, we want to find $\mathbf{x}$ that makes $\|A\mathbf{x} - \mathbf{b}\|$ as small as possible. This is called the least square solution:

Def The **least square solution** $\hat{\mathbf{x}}$ of the system $A\mathbf{x} = \mathbf{b}$ is a vector such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \le \|A\mathbf{x} - \mathbf{b}\|, \qquad \text{for all} \quad \mathbf{x} \in \mathbf{R}^n$$

Formulated differently, we want to find the vector $\mathbf{p} \in R(A)$ which is closest to $\mathbf{b}$. Since $R(A)$ can be any subspace we therefore first want to solve the problem of finding the vector $\mathbf{p}$ in a given subspace S of $\mathbf{R}^n$ that is closest to a vector $\mathbf{b} \in \mathbf{R}^n$:

Th Let S be a subspace of $\mathbf{R}^m$ and $\mathbf{b} \in \mathbf{R}^m$. Then there is a unique vector $\mathbf{p} \in S$ such that

$$\|\mathbf{b} - \mathbf{p}\| \le \|\mathbf{b} - \mathbf{y}\|, \qquad \text{for all} \quad \mathbf{y} \in S$$

Furthermore $\mathbf{p} \in S$ is the unique vector such that $\mathbf{b} - \mathbf{p} \in S^\perp$.

Pf Since by section 5.2, $\mathbf{R}^m = S \oplus S^\perp$, each vector $\mathbf{b} \in S$ can be written uniquely as

$$\mathbf{b} = \mathbf{p} + \mathbf{z}, \qquad \text{where} \qquad \mathbf{p} \in S, \quad \text{and} \quad \mathbf{z} \in S^\perp$$

If $\mathbf{y}$ is any other vector in S then $\mathbf{z} - \mathbf{y} \in S$ and since $\mathbf{b} - \mathbf{p} \in S^\perp$ it follows from the Phytagorean law that

$$\|\mathbf{b} - \mathbf{y}\|^2 = \|\mathbf{b} - \mathbf{p} + (\mathbf{p} - \mathbf{y})\|^2 = \|\mathbf{b} - \mathbf{p}\|^2 + \|\mathbf{p} - \mathbf{y}\|^2 \ge \|\mathbf{b} - \mathbf{p}\|^2$$

with equality if and only if $\mathbf{p} - \mathbf{y} = \mathbf{0}$.

We have now shown that there is a unique vector $\mathbf{p} \in R(A)$ that is closest to $\mathbf{b} \in \mathbf{R}^m$ and this is half the solution of the problem but it still remains to find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}} = \mathbf{p}$. We also showed that $\mathbf{b} - \mathbf{p}$ is orthogonal $R(A)$, i.e. in $R(A)^\perp$. But recall from section 5.2 that $R(A)^\perp = N(A^T)$. Hence $\mathbf{b} - A\mathbf{x} \in N(A^T)$, i.e.

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{5.3.1}$$

This so called **normal equation** represents an $n \times n$ system. We have:

Th If A is an $m \times n$ matrix of rank n then (5.3.1) has a unique solution

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

and $\hat{\mathbf{x}}$ is the least square solution of the problem $A\mathbf{x} = \mathbf{b}$.

Note also that we can use the solution of the normal equation to construct the orthogonal projection onto the subspace spanned by the column vectors of A:

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

Ex Find the least square solution to $A\mathbf{x}=\mathbf{b}$ where $A=\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, $\mathbf{b}=\begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}$.

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

and

$$(A^T A)^{-1} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$$

and

$$A^T \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 6 \\ 18 \end{bmatrix}$$

Hence

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 6 \\ 18 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$$

Ex Find the least square fit by a line to the following three points in the plane:

x	1	2	3
y	0	0	6

Sol We want to find the line $y = c_0 + c_1 x$ that is closest to going through the three points, (x_i, y_i), $i = 1, 2, 3$, i.e. such that $\triangle_1^2 + \triangle_2^2 + \triangle_3^2$ is as small as possible, where

$$\triangle_i = y_i - (c_0 + c_1 x_i), \qquad i = 1, 2, 3$$

Or if we plug in the values of (x_i, y_i) and write it in matrix form we want to make the vector

$$\triangle = \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$$

as small as possible. But this is exactly the least square problem in the previous example, and he solution is $(c_0, c_1)^T = (-4, 3)^T$.