18. Lecture 18: 5.4 Least-Square Problems

The orthogonal decomposition theorem Let W be a subspace of $\mathbf{R}^n$. Any $\mathbf{x} \in \mathbf{R}^n$ can be written uniquely as

$$\mathbf{x} = \mathbf{x}^{\|} + \mathbf{x}^{\perp}, \qquad \text{with} \quad \mathbf{x}^{\|} \in W, \quad \mathbf{x}^{\perp} \in W^{\perp},$$

where $\mathbf{x}^{\|} = \operatorname{proj}_W \mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto W and $W^{\perp}$ is the orthogonal complement of W (i.e. all vectors orthogonal to every vector in W.)

The best approximation theorem Let W be a subspace of $\mathbf{R}^n$ and $\mathbf{x} \in \mathbf{R}^n$. Then the orthogonal projection $\mathbf{x}^{\|}$ of $\mathbf{x}$ onto W is the point in W closest to $\mathbf{x}$, i.e.

$$\|\mathbf{x} - \mathbf{x}^{\|}\| < \|\mathbf{x} - \mathbf{v}\|, \qquad \text{for all} \quad \mathbf{v} \in W, \quad \mathbf{v} \neq \mathbf{x}^{\|}.$$

Pf We can write

$$\mathbf{x} - \mathbf{v} = \mathbf{x} - \mathbf{x}^{\|} + \mathbf{x}^{\|} - \mathbf{v}$$

where $\mathbf{x} - \mathbf{x}^{\|} \in W^{\perp}$ and $\mathbf{x}^{\|} - \mathbf{v} \in W$ are orthogonal and hence

$$||\mathbf{x} - \mathbf{v}\|^2 = \|\mathbf{x} - \mathbf{x}^{\|}\|^2 + \|\mathbf{x}^{\|} - \mathbf{v}\|^2 > \|\mathbf{x} - \mathbf{x}^{\|}\|^2.$$

by the **Pythagorean theorem:** $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

Ex Find the closest point to $\mathbf{x} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -2 \end{bmatrix}$ to $W = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$.

Sol $\mathbf{x}^{\|} = \dfrac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \dfrac{\mathbf{x} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = 3 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + (-1) \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ -1 \\ -1 \end{bmatrix}$.

We showed last lecture that the transpose satisfy

$$\mathbf{u} \cdot (A^T \mathbf{v}) = (A\mathbf{u}) \cdot \mathbf{v}, \qquad \text{for all} \quad \mathbf{u}, \mathbf{v}. \tag{18.1}$$

From this it follows that
Th We have

$$(\operatorname{Im} A)^{\perp} = \operatorname{Ker} A^T$$

Pf In fact if $\mathbf{v} \in \operatorname{Ker} A^T$ if and only if the left hand side of (18.1) is 0 for all $\mathbf{u}$ and $\mathbf{v} \in (\operatorname{Im} A)^{\perp}$ if and only if the right hand side of (18.1) is 0 for all $\mathbf{u}$.

Least Square Solution

A standard statistical technique is to find a least square fit to data points in the 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 an $\mathbf{x}$ that makes $\|A\mathbf{x}-\mathbf{b}\|$ as small as possible. This is called the least square solution:

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

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

Formulated differently, we want to find the vector $\mathbf{p}\in W=\operatorname{Im}A$ that is closest to $\mathbf{b}$. From the previous section we know that $\mathbf{p}=\operatorname{proj}_W\mathbf{b}$, is the orthogonal projection of $\mathbf{b}$ onto W. Since $\mathbf{p}\in\operatorname{Im}A$, there is $\mathbf{x}^*$ such that $A\mathbf{x}^*=\mathbf{p}$. However we do not know if $\mathbf{x}^*$ is unique.

Since $\mathbf{p}=\operatorname{proj}_W\mathbf{b}$ it follows that $\mathbf{b}-\mathbf{p}$ is orthogonal $W=\operatorname{Im}A$, i.e. in $(\operatorname{Im}A)^\perp$. But recall that $(\operatorname{Im}A)^\perp=\operatorname{Ker}A^T$. Hence $\mathbf{b}-A\mathbf{x}^*\in\operatorname{Ker}A^T$, i.e.

$$A^TA\mathbf{x}^*=A^T\mathbf{b} \tag{18.2}$$

This so called **normal equation** represents a much smaller $n\times n$ system. It still remains to find out if the solution to this system is unique. For this we have the help of

Th We have

$$\operatorname{Ker}A=\operatorname{Ker}(A^TA)$$

Pf If $A^TA\mathbf{x}=\mathbf{0}$ then $A\mathbf{x}$ is in the image of A and in the kernel of A^T. But recall that $(\operatorname{Im}A)^\perp=\operatorname{Ker}A^T$. Therefore $A\mathbf{x}$ would be both in the image of A and in it is orthogonal complement, i.e. it would be orthogonal to itself so $\|A\mathbf{x}\|^2=(A\mathbf{x})\cdot(A\mathbf{x})=0$, but this implies that $A\mathbf{x}=\mathbf{0}$, i.e. $\mathbf{x}\in\operatorname{Ker}A$.

We have now proven:
Th If A is an $m\times n$ matrix of rank n (i.e. $\operatorname{Ker}A=\{0\}$ by the rank-nullity theorem) then (18.2) has a unique solution

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

and $\mathbf{x}^*$ 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:

$$\mathbf{p}=\operatorname{proj}_W\mathbf{b}=A\mathbf{x}^*=A(A^TA)^{-1}A^T\mathbf{b}$$

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

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

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. (Note that the 1s in the matrix comes from that c_0 is multiplied by 1.) But this is exactly the least square problem in the previous example, and the solution is $(c_0, c_1)^T = (-4, 3)^T$.

Summary

Let A be an $m \times n$ matrix with $m \geq n$. If $m > n$ the system $A\mathbf{x} = \mathbf{b}$ is over determined and may not have a solution. Instead we will look for an approximate solution.

The best approximation theorem Let W be a subspace of $\mathbf{R}^n$ and $\mathbf{x} \in \mathbf{R}^n$. Then the orthogonal projection $\mathbf{x}^{\parallel}$ of $\mathbf{x}$ onto W is the point in W closest to $\mathbf{x}$, i.e.

$$\|\mathbf{x} - \mathbf{x}^{\parallel}\| < \|\mathbf{x} - \mathbf{v}\|, \qquad \text{for all} \quad \mathbf{v} \in W, \quad \mathbf{v} \neq \mathbf{x}^{\parallel}.$$

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

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

Formulated differently, we want to find the vector $\mathbf{p} \in W = \operatorname{Im} A$ that is closest to $\mathbf{b}$. From the previous section we know that $\mathbf{p} = \operatorname{proj}_W \mathbf{b}$, is the orthogonal projection of $\mathbf{b}$ onto W. Since $\mathbf{p} \in \operatorname{Im} A$, there is $\mathbf{x}^*$ such that $A\mathbf{x}^* = \mathbf{p}$. However we do not know if $\mathbf{x}^*$ is unique.

Th We have $(\operatorname{Im} A)^{\perp} = \operatorname{Ker} A^T$
Pf This follows from that $\mathbf{u} \cdot (A^T\mathbf{v}) = (A\mathbf{u}) \cdot \mathbf{v}$, for all $\mathbf{u}, \mathbf{v}$ (which we proved in last lecture).

Since $\mathbf{p} = \operatorname{proj}_W \mathbf{b}$ it follows that $\mathbf{b} - \mathbf{p}$ is orthogonal $W = \operatorname{Im} A$, i.e. in $(\operatorname{Im} A)^{\perp}$. But recall that $(\operatorname{Im} A)^{\perp} = \operatorname{Ker} A^T$. Hence $\mathbf{b} - A\mathbf{x}^* \in \operatorname{Ker} A^T$, i.e.

$$A^T A\mathbf{x}^* = A^T\mathbf{b} \tag{18.3}$$

This so called **normal equation** represents a much smaller $n \times n$ system.
Th If A is an $m \times n$ matrix of rank n (i.e. $\operatorname{Ker} A = \{0\}$) then (18.3) has a unique solution

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

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

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

$$\triangle = \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \mathbf{b} - A\mathbf{c},$$

as small as possible. (Note that the 1s in the matrix comes from that c_0 is multiplied by 1.) This is exactly a least square problem. The least square solution $\mathbf{c}^*$ is the solution to the normal equation $A^T A\mathbf{c}^* = A^T\mathbf{b}$. The solution is $(c_0, c_1)^T = (-4, 3)^T$.