Lecture 26 (hrs. 52,53) - November 25, 2025, 16:30 - 18:30 F3

(3.16) <u>Definition</u> (least squares solution of a linear system).

Let $A \in R^{r \times c}$ where $r > c$, and let $b \in R^r$. An element $x^* \in R^c$ is called *a least squares solution* of the system $A x = b$ if x^* is the absolute minimum point of the function $SQ: R^c \to R$ defined by:

$$SQ(x) = \| A x - b \|^2$$

Note that: if $y \in R^c$ is a solution of $A x = b$ then y is *also* a solution of $A x = b$ in the least squares sense (why?) <u>but</u>, except in special cases, a solution of $A x = b$ in the least squares sense *is not* a solution of $A x = b$.

Let us see *how we can find* the least squares solutions of $A x = b$.

(3.17) <u>Remark</u> (orthogonal decomposition of a vector).

Let $A = (a_1,...,a_c) \in R^{r \times c}$ where $r > c$, and let $b \in R^r$. Denoted by b_* <u>the</u> *orthogonal projection*[1] of b onto span{ $a_1,...,a_c$ } = $C(A)$[2], and set $b_\perp = b - b_*$ we get the orthogonal decomposition:

$$b = b_* + b_\perp$$

Observe that:

(1) Since $b_* \in C(A)$, there exists $y \in R^c$ such that $b_* = A y$;

(2) By the definition of orthogonal projection, the column $b_\perp = b - b_*$ is orthogonal to all the elements of $C(A)$.

(3.18) <u>Remark</u>.

To find the least squares solutions of $A x = b$, note that, using the orthogonal decomposition of b introduced in the previous remark, for each $x \in R^c$ we have:

$$SQ(x) = \| A x - b \|^2 = \| A x - b_* + b_\perp \|^2 = \| A x - A y + b_\perp \|^2 = \| A (x - y) + b_\perp \|^2$$

Since $A (x - y) \in C(A)$ and $b_\perp$ is orthogonal to all the elements of $C(A)$, by Pythagoras's theorem[3] we have:

$$\| A (x - y) + b_\perp \|^2 = \| A (x - y) \|^2 + \| b_\perp \|^2$$

Then:

- For every $x \in R^c$ it is: $SQ(x) = \| A (x - y) \|^2 + \| b_\perp \|^2 \geqslant \| b_\perp \|^2$
- $SQ(x) = \| b_\perp \|^2 \Leftrightarrow \| A (x - y) \|^2 = 0 \Leftrightarrow A (x - y) = 0$, i.e. $x - y \in \ker A$.[4]

---

1 The orthogonal projection of $v \in R^n$ onto a subspace $W \subset R^n$ is the *unique* element $v_* \in W$ such that the difference $v - v_*$ is orthogonal to all elements of W.

2 $C(A)$ is also called the *column space of* A and coincides with the image of the linear application from R^c to R^r defined by $x \to A x$.

3 Let a,b be elements of R^n, and let $<a,b> = b^t a$ be the scalar product of a and b. If a and b are orthogonal (that is, if $<a,b> = 0$) then we have:

$$\| a + b \|^2 = <a + b,a + b> = <a,a> + 2 <a,b> + <b,b> = <a,a> + <b,b> = \| a \|^2 + \| b \|^2$$

4 If $A \in R^{r \times c}$, we denote by $\ker A$ the set of the solutions of the homogeneous system $A z = 0$. $\ker A$ is a vector subspace of R^c

Hence, the set $S_{MQ}(A,b)$ of the least squares solutions of $A\,x = b$ is:

$$S_{MQ}(A,b) = y + \ker A$$

(3.19) <u>Remark</u> (normal equations).

To find all the columns $y \in R^c$ such that $b_* = A\,y$ observe that, by the definition of orthogonal projection onto $C(A)$:

$$y \in R^c \text{ is such that } A\,y = b_* \Leftrightarrow b - b_* = b - A\,y \text{ is orthogonal to } \textit{all} \text{ the elements of } C(A)$$

<u>But</u>: a column $v \in R^r$ is orthogonal to all the elements of $C(A)$ *if and only if* v is orthogonal to the columns of A (prove it!). Hence: v is orthogonal to all the elements of $C(A) \Leftrightarrow \langle v,a_1\rangle = a_1^t\,v = 0,\ldots,\langle v,a_c\rangle = a_c^t\,v = 0 \Leftrightarrow A^t\,v = 0$. Then:

$$y \in R^c \text{ is such that } A\,y = b_* \;\Leftrightarrow\; A^t\,(b - A\,y) = 0 \;\Leftrightarrow\; A^t\,A\,y = A^t\,b$$

The system of liner equations $A^t\,A\,x = A^t\,b$ is called the *system of the normal equations* relative to the system $A\,x = b$.

Observe that: $\ker A = \ker A^t\,A$ (indeed: $x \in \ker A \Rightarrow A\,x = 0 \Rightarrow A^t\,(A\,x) = 0 \Rightarrow A^t\,A\,x = 0 \Rightarrow x \in \ker A^t\,A$; on the contrary: $x \in \ker A^t\,A \Rightarrow A^t\,A\,x = 0 \Rightarrow x^t\,(A^t\,A\,x) = 0 \Rightarrow (x^t\,A^t)(A\,x) = 0 \Rightarrow (A\,x)^t\,(A\,x) = 0 \Rightarrow \|A\,x\|^2 = 0 \Rightarrow A\,x = 0 \Rightarrow x \in \ker A$). Then:

$$S_{MQ}(A,b) = y + \ker A = y + \ker A^t\,A = \{ \text{ least squares solutions of } A\,x = b \}$$

Moreover:

- The matrix $A^t\,A \in R^{c \times c}$ is *symmetric* and *positive semidefinite* (indeed, for every column $x \neq 0$ in R^c it is: $x^t\,(A^t\,A)\,x = (x^t\,A^t)(A\,x) = (A\,x)^t\,(A\,x) = \|A\,x\|^2 \geqslant 0$) and it is *positive definite* if and only if the columns of A are *linearly independent* (prove it!).
- The columns of A are *linearly independent* $\Leftrightarrow \ker A = \ker A^t\,A = \{0\} \Leftrightarrow$ the set of the solutions of the system of the normal equations has *only one element* $\Leftrightarrow$ the matrix $A^t\,A$ *is invertible*.
- The columns of A are *linearly dependent* $\Leftrightarrow \dim\ker A = \dim\ker A^t\,A > 0 \Leftrightarrow$ the set of the solutions of the system of the normal equations has *infinite elements* $\Leftrightarrow$ the matrix $A^t\,A$ *is not invertible*.