Lecture 26 § 4.3 Least Squares

<u>Defn</u> A <u>least squares solution</u> to the linear system
$$Ax = b$$
is a vector $x^* \in \mathbb{R}^n$ which minimizes $\|Ax - b\|$, the Euclidean norm.

<u>Rk</u> The name "least squares" is due to the fact that
$$\|Ax - b\| = \sqrt{\sum_{i=1}^{m} ((Ax)_i - b_i)^2},$$
so we search for an x minimizing the sum of the squares of the error.

Let $v = Ax$. Then we seek to find $v \in \operatorname{rng}(A)$ minimizing $\|v - b\|$. Since $\operatorname{rng}(A)$ is a subspace of $\mathbb{R}^m$, we may use the solution of the nearest point problem in order to solve the least squares problem. So we must choose a basis of $\operatorname{rng}(A)$ to complete the solution.

<u>Case 1 $\ker(A) = \{0\}$.</u>

Suppose that $\ker(A) = \{0\}$. Write $A = (v_1 \ldots v_n)$. Since $\ker(A) = \{0\}$, $v_1 \ldots v_n$ are linearly independent and span $\operatorname{rng}(A)$. As before, we set
$$K = A^T A, \quad f = A^T b, \quad c = \|b\|^2,$$
and find that, using the solution to the nearest point problem,
$$\|v - b\|^2 = \|Ax - b\|^2 = x^T K x - 2x^T f + \|b\|^2 =: p(x).$$
Using this notation, we quote Theorem 4.5 to find that the least squares solution x^* is the solution of
$$Kx^* = f.$$
We summarize this below in Theorem 4.8

Thm 4.8 Assume that A is an $m \times n$ matrix with $\ker(A) = \{0\}$ and $b \in \mathbb{R}^m$. Then there is a unique ~~solution~~ least squares solution to $Ax = b$, x^*, which is the solution to

$$A^T A x = A^T b.$$

The least squares error is

$$\|Ax^* - b\| = \sqrt{\|b\|^2 - b^T A x^*}$$

Pf Assume $\ker(A) = \{0\}$. Set $K = A^T A$, $f = A^T b$, $c = \|b\|^2$.
~~By Theorem 4.5, the unique nearest~~ Because $\ker(A) = \{0\}$, $v_1 \dots v_n$ form a basis of $\mathrm{rng}(A) \subseteq \mathbb{R}^m$ where $A = (v_1 \dots v_n)$.
So by Theorem 4.5, the unique nearest point v^* is given by

$$v^* = A x^*$$

where

$$K x^* = f.$$

The least squares error is then the distance (in the Euclidean norm) from v^* to b, which by Theorem 4.5 is

$$\|Ax^* - b\| = \|v^* - b\| = \sqrt{\|b\|^2 - f^T x^*} = \sqrt{\|b\|^2 - (A^T b)^T x^*} = \sqrt{\|b\|^2 - b^T A x^*}.$$

//

Case 2 $\ker(A) = $ anything!
We now study the general case, i.e., where A may have nontrivial kernel. We solve the problem side by side with an example.

Defn Let A be a matrix. The basic submatrix of A is the matrix $\tilde{A}$ consisting of the columns of A which correspond to the basic variables.

ex Let $A = \begin{pmatrix} 1 & 0 & 1 & 3 & 4 \\ -2 & 2 & -1 & -6 & -3 \\ -2 & 4 & 0 & -3 & 3 \\ 1 & -2 & 0 & 6 & -3 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$. ~~Write~~ Let

$A = (v_1 \dots v_5)$, so v_i is the ith column of A.

By row reducing A, we find the row echelon form to be

$$U = \begin{pmatrix} 1 & 0 & 1 & 3 & 4 \\ 0 & 2 & 1 & 0 & 5 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

basic variable columns.

Thus, the basic submatrix of A is

$$\widetilde{A} = (v_1\ v_2\ v_4) = \begin{pmatrix} 1 & 0 & 3 \\ -2 & 2 & -6 \\ -2 & 4 & -3 \\ 1 & -2 & 6 \end{pmatrix}.$$

We continue our solution to the general least squares problem.

Observation 1 The basic submatrix of A has trivial kernel, i.e., $\ker(\widetilde{A}) = \{0\}$. This follows because the linear dependencies of the columns of the row echelon form U are the same as those of the columns of A, and by definition, the basic variable columns contain pivots.

Observation 2 Because $\ker(\widetilde{A}) = \{0\}$, Theorem 4.8 produces the unique least squares solution to $\widetilde{A}\widetilde{x} = b$, namely, $\widetilde{x}^*$ given by

$$\widetilde{A}^T \widetilde{A}\, \widetilde{x}^* = \widetilde{A}^T b.$$

Observation 3 By "setting the free variables to be 0", this produces one solution to the original least squares problem for $Ax = b$.

ex continued Continuing our same example, we set

$$K = \widetilde{A}^T \widetilde{A} = \begin{pmatrix} 10 & -14 & 27 \\ -14 & 24 & -36 \\ 27 & -36 & 90 \end{pmatrix}, \quad f = A^T b = \begin{pmatrix} -5 \\ 8 \\ 6 \end{pmatrix}.$$

Thus, solving for $\widetilde{x}^*$, we find

$K\widetilde{x}^* = f$ yields $\widetilde{x}^* = \begin{pmatrix} -4 \\ -1/4 \\ 7/6 \end{pmatrix}$.

Note that

$$v^* = \tilde{A}\tilde{x}^* = -4v_1 - \tfrac{1}{4}v_2 + 7/6\, v_4 = A\begin{pmatrix} -4 \\ -\frac{1}{4} \\ 0 \\ 7/6 \\ 0 \end{pmatrix},$$

So $x^* = \begin{pmatrix} -4 \\ -\frac{1}{4} \\ 0 \\ \frac{7}{6} \\ 0 \end{pmatrix}$ is a least squares ~~to~~ solution to $Ax = b$.

Observation 4: $v^* = Ax^*$ is the nearest point to b in $\mathrm{rng}(A)$. Moreover, the nearest point is unique. Thus, the general least squares solution will be given by

$$x^{*\prime} = x^* + z \quad \text{for } z \in \ker(A).$$

ex continued

To finish solving the general problem, we now just need to describe the kernel. Because we are looking to "construct" the general solution, the best description will be a basis representation.

The row echelon form of A is

$$U = \begin{pmatrix} 1 & 0 & 1 & 3 & 4 \\ 0 & 2 & 1 & 0 & 5 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

~~We find~~ Setting $Ux = 0$, we find

$$\begin{aligned} x_4 &= -\tfrac{1}{3}x_5 \\ x_2 &= -\frac{x_3}{2} - \frac{5}{2}x_5 \\ x_1 &= -x_3 - 3x_4 - 4x_5 \\ &= -x_3 - x_5 - 4x_5 \\ &= -x_3 - 5x_5. \end{aligned}$$

Setting $x_3 = 1$, $x_5 = 0$, we find one vector in $\ker(A)$ as

$$u_1 = \begin{pmatrix} -1 \\ -\frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

Setting $x_3 = 0$, $x_5 = 1$, we get a second vector in $\ker(A)$

$$u_2 = \begin{pmatrix} -5 \\ -5/2 \\ 0 \\ -1/3 \\ 1 \end{pmatrix}.$$

So the general element in $\ker(A)$ is

$$z = s u_1 + t u_2 \quad \text{for } s, t \in \mathbb{R}.$$

Thus, the general least squares solution to $Ax = b$ is

$$x^* = \begin{pmatrix} -4 \\ -\frac{1}{4} \\ 0 \\ 7/6 \\ 0 \end{pmatrix} + s \begin{pmatrix} -1 \\ -\frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -5 \\ -5/2 \\ 0 \\ -4/3 \\ 1 \end{pmatrix} \quad \text{for } s, t \in \mathbb{R}.$$

Rk The terminology "basic submatrix" is not standard (as far as I know), and should be explained to anyone who hasn't taken this class when solving problems.