Solvability of least squares problems

Linear systems: $Ax = b$ with A square: unique solution if A nonsingular

Linear least squares problems: $\min\|Ax - b\|$ with A tall thin: unique solution if...

$\| A x - b \|$

Observation: if I can obtain 0 as a linear combination of the columns of A, then I have multiple solutions.

Example:

$\|Ax - b\| = \left\| \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \end{bmatrix} \right\| = 2$

$$\min\|Ax - b\|, \quad A = \begin{array}{ccc} a_1 & a_2 & a_3 = a_1 + a_2 \\ \hline \end{array} \begin{bmatrix} 1 & -1 & 0 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 3 \\ 1 \\ 2 \end{bmatrix}. \qquad \begin{bmatrix} 0 \\ 3 \\ 1 \\ 0 \end{bmatrix}$$

Solution: We can 'match' the first three entries (but not the 4th).

$x_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ solves the problem. But also $x_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Or $x_3 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$...

$Ax_1 = Ax_2 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ 0 \end{bmatrix}$

Full column rank definition

$Az = 0$

What is going on: there is a vector $z \neq 0$ in $\ker A$: $A \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = 0$.

$Ax = A(x+z) = A(x+2z) = A(x-37z)$

If x is a solution, then so is $x + z, x + 2z, x - 37z \ldots$

Definition

We say that $A \in \mathbb{R}^{m \times n}$ has full column rank if $\ker A = \{0\}$, or, equivalently: $\operatorname{rank} A = n$, or, equivalently: there is no $z \in \mathbb{R}^n, z \neq 0$ such that $Az = 0$.

$A(x+z) = Ax + \underbrace{Az}_{=0} = Ax$

$a_1 + a_2 - a_3 = 0$

rank A = (max. number of linearly independent columns)

Criterion for full column rank

Theorem

A has full column rank if and only if $A^T A$ is positive definite.

We already saw (lecture on orthogonal matrices) that $A^T A$ is symmetric and positive semidefinite.

For each $z \neq 0$, $z^T A^T A z = \|Az\|^2 \geq 0$.

Proof: A full column rank $\iff$ $Az \neq 0$ for all $z \neq 0$ $\iff$ $z^T A^T A z = \|Az\|^2 \neq 0$ for all $z \neq 0$

(Def: $A^T A$ is positive semidef if $z^T (A^T A) z \geq 0$ &

" " positive definite if $z^T (A^T A) z > 0$

for all $z \neq 0$

$$z^T A^T A z = (Az)^T (Az) = \|Az\|^2 \geq 0$$

$\Downarrow$

> 0 if A is full c.r.

A has full col. rank $\Leftrightarrow$ for all $z \neq 0$, $Az \neq 0$

$\Leftrightarrow$ for all $z \neq 0$, $\|Az\| > 0$ $\Leftrightarrow$ for all $z \neq 0$, $\|Az\|^2 > 0$

$\Leftrightarrow$ for all $z \neq 0$ $(Az)^T (Az) > 0$

$\Leftrightarrow$ for all $z \neq 0$ $z^T A^T A z > 0$ $\Leftrightarrow$ $A^T A$ is pos. def.

Least squares problems — solution

Suppose A has full column rank. Then $\min \|Ax - b\|$ can also be written as

$$\begin{aligned}
&\min_{x\in\mathbb{R}^n} \|Ax - b\|^2 = \min_{x\in\mathbb{R}^n} (Ax - b)^T (Ax - b) \\
&= \min_{x\in\mathbb{R}^n} x^T A^T A x - b^T A x - x^T A^T b + b^T b \\
&= \min_{x\in\mathbb{R}^n} x^T A^T A x - 2 b^T A x + b^T b
\end{aligned}$$

Min of a quadratic function — sounds familiar?

$$\underbrace{\min_{x\in\mathbb{R}^n} \|Ax-b\|^2}_{} = \min_{x\in\mathbb{R}^n} (Ax-b)^T(Ax-b) =$$

$$\min (x^TA^T - b^T)(Ax-b) = \min x^TA^TAx - b^TAx - x^TA^Tb + b^Tb =$$

$$x^TA^Tb = (x^TAb)^T = b^TA^Tx$$

$$= \min_{x\in\mathbb{R}^n} \underbrace{x^TA^TAx - 2b^TAx + b^Tb}_{\text{quadratic function}}$$

$$f(x) = x^TQx + q^Tx + c$$

$$\nabla f(x) = 2Qx + q$$

Gradient: $\underline{2A^TAx - 2A^Tb}$

Hessian: $2A^TA$ $\leftarrow$ always positive semidef.

pos. def. if A has full col. rank

Objective function is $\underline{\text{convex}}$ (strictly if A has f.c.r.)

Some optimization

$$\min_{x \in \mathbb{R}^n} x^T A^T A x - 2b^T A x + b^T b$$

Gradient $2A^T A x - 2A^T b$.
Hessian $2A^T A$. $\rightarrow$ strictly convex!

Solution exists unique, and can be found as
$0 = \text{gradient} = 2A^T A x - 2A^T b$, or

$\underset{n \times n}{A^T A}$ $\underset{n \times 1}{A^T b}$

$$\boxed{A^T A x = A^T b.} \quad \cancel{A}$$

$A^T A$ is square invertible (because it's positive definite), so this is just a linear system.

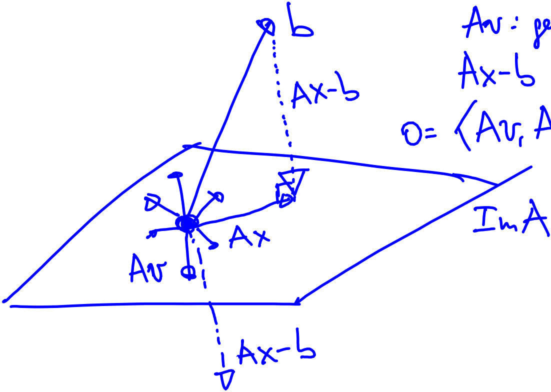
Can be solved with 'classical' methods: Gaussian elimination, or QR factorization. . . . $A^T A = QR$ $\quad x = R^{-1} Q^T b$
(We will see a faster/better method for positive definite matrices, Cholesky factorization, $A^T A = R^T R$.)

Geometric idea

TL;DR: can't solve $Ax = b$? Multiply both sides by A^T and try again!

Geometric idea The residual $Ax - b$ is orthogonal to any vector $Av \in \operatorname{span} A$: $(Av)^T(Ax - b) = 0$.

This method to solve LS problems is known as method of normal equations ('normal' is a fancy word for 'perpendicular/orthogonal').



Av : generic vector on Im A

Ax-b : difference

$$0 = \langle Av, Ax-b \rangle = v^T \underbrace{A^T(Ax-b)}_{=0}$$

Exercises

1. Can a short-fat matrix $A \in \mathbb{R}^{m \times n}$, $n > m$, have full column rank?
2. Take a 'simple' linear least squares problem, e.g. $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 1 \end{bmatrix} x = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. Try to solve $\min \|Ax - b\|^2$ with one of the techniques seen in prof. Frangioni's lectures (for instance: gradient descent), and compare the solution with the closed-form one that we found in this lecture. Do they have the same objective function? Which is larger?