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...

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

Example:

$$\min \|Ax - b\|, \quad A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 3 \\ 1 \\ 2 \end{bmatrix}.$$

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

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

Full column rank definition

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

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 $Az = 0$.

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 $z \neq 0$

Least squares problems — solution

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

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

Min of a quadratic function — sounds familiar?

Some optimization

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

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

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

$$A^T Ax = A^T b.$$

$A^T A$ 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...
(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').

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?