Least-squares problems

MATH 6610 Lecture 14

October 5, 2020

Trefethen & Bau: Lecture 11

Least-squares problems

If $A \in C^{m \times n}$ and $b \in \mathbb{C}^n$, we are interested in computing the least-squares solution to

$$Ax = b$$

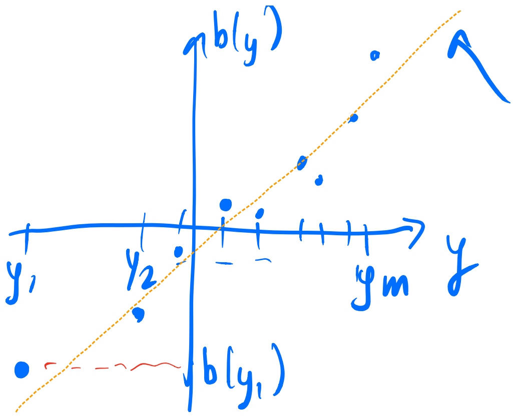
$$\begin{pmatrix} A \end{pmatrix} \begin{pmatrix} x \end{pmatrix} = \begin{pmatrix} b \end{pmatrix}$$

This arises in several situations, e.g., data fitting.

Note: it's not the case (in general) that we find x that satisfies $Ax = b$.

Equivalently: we're actually trying to solve:

$$x = \underset{z \in \mathbb{C}^n}{\operatorname{argmin}} \|Az - b\|_2$$



line that "best fits" the data.

line has equation:

$$x_1 y + x_2 b = x_3$$

(x_1, x_2, x_3 are unknown scalars, y, b are also scalars)

As long as the line is not vertical: same as

$$b = z_1 \cancel{x} y + z_2 \quad (z_1, z_2 \text{ unknowns scalars})$$

"best fit": empirical $\overbrace{\text{root mean square}}^{\text{RMS}}$ difference between line and data.

@ data point $(y_j, b(y_j))$

error: $|\underbrace{z_1 y_j + z_2}_{\text{line estimate}} - \underbrace{b(y_j)}_{\text{data}}|$

$$\text{RMS}: \sum_{j=1}^{m} |z_1 y_j + z_2 - b(y_j)|^2 = \|Az - b\|_2^2$$

b is an m-vector: $b = \begin{pmatrix} b(y_1) \\ \vdots \\ b(y_m) \end{pmatrix}$

A is an $m \times 2$ matrix: $A = \begin{pmatrix} y_1 & 1 \\ y_2 & 1 \\ \vdots & \\ y_m & 1 \end{pmatrix}$

Line fit $\Longleftrightarrow$ solve $\underset{z}{\operatorname{argmin}} \|Az - b\|_2^2$

Least-squares solutions

The following is a result we have essentially already proven: (with the SVD)

Theorem

Suppose $A \in \mathbb{C}^{m\times n}$ has full column rank (n). Then, for any $b \in$ ~~$\mathbb{C}^n$~~ $\mathbb{C}^m$, there is a unique solution x that solves

$$Ax = b$$

in the least-square sense. Furthermore, this solution x is the unique solution to $\underbrace{A^*Ax = A^*b}_{n\times n \text{ square system}}$, and the residual $r := b - Ax$ is orthogonal to $\mathrm{range}(A)$.

Least-squares solutions

The following is a result we have essentially already proven:

Theorem

Suppose $A \in \mathbb{C}^{m \times n}$ *has full column rank* (n). *Then, for any* $b \in \mathbb{C}^n$, *there is a unique solution* x *that solves*

$$Ax = b$$

in the least-square sense. Furthermore, this solution x *is the unique solution to* $A^* Ax = A^* b$, *and the residual* $r := b - Ax$ *is orthogonal to* $\mathrm{range}(A)$.

The system $A^* Ax = A^* b$ is called the *normal equations*.

Proof (sketch)

$$\min_x \|Ax - b\|_2^2$$

Ax in $\mathrm{range}(A) =: V \subset \mathbb{C}^m$.

$V^\perp$: orthogonal complement of V in $\mathbb{C}^m$

P_V, $P_{V^\perp}$: orthogonal projections onto V, $V^\perp$, respectively.

$$\min_x \|Ax - b\|_2^2 = \min_x \|\underbrace{Ax - P_V b}_{\text{orthogonal to } P_{V^\perp} b} - P_{V^\perp} b\|_2^2$$

$$= \underbrace{\min_x \|Ax - P_V b\|_2^2}_{\text{focus on this}} + \underbrace{\|P_{V^\perp} b\|_2^2}_{\text{doesn't depend on } x.}$$

(☆) (this step shows that the least-squares solution produces a residual that's orthogonal to range(A).)

A has full column rank $\Longrightarrow A = QR$

$Q \in \mathbb{C}^{m \times n}$ (orthonormal columns)

$R \in \mathbb{C}^{n \times n}$, (diagonal entries are nonzero)

Also: columns of Q are an orthonormal basis for range(A). $(=V)$

$\Longrightarrow P_V = QQ^*$

(★): $\min_x \|Ax - P_V b\|_2^2 = \min_x \|QRx - QQ^*b\|_2^2$

$= \min_x \|Q(Rx - Q^*b)\|_2^2$

[lemma: if Q has orthonormal columns, then $\|Qx\|_2^2 = \|x\|_2^2$)

$= \min_x \|Rx - Q^*b\|_2^2$

$\rightarrow R$ is invertible $\Longrightarrow x = R^{-1}Q^*b$ solves minimization problem.

What about $A^*Ax = A^*b$?

$\big\downarrow A = QR$

$R^*Q^*QRx = R^*Q^*b.$

$\big\downarrow R$ invertible

$Q^*QRx = Q^*b$

$\big\downarrow Q$ has ON columns

$Rx = Q^*b$

$\downarrow$

$x = R^{-1} Q^* b$ (and this matches the one above.)

$\blacksquare$

Computational solutions

While the normal equations are typically useful for analysis, they are typically not used for computation.

$$A = QR \qquad \Longrightarrow \qquad x = R^{-1}Q^*b.$$

In most cases, the QR decomposition is used, largely for stability reasons.

(Why not SVD? It's expensive.)

Normal equations: $A^*Ax = A^*b$

$$R^*Q^*QRx = R^*Q^*b$$

$$(R^*R)x = R^*Q^*b$$

Recall: relative condition number of $R^*Q^*b \mapsto x$

is $\kappa(R^*R) = \kappa^2(R)$

$$\frac{\sigma_1(R^*R)}{\sigma_n(R^*R)}$$

Via a QR strategy: $Rx = Q^*b$

conditioning $Q^*b \longmapsto x$ is $\kappa(R)$.

I.e. QR strategy: more well-conditioned.