NLA 2021-2022
Special linear systems

Martin Sombra

13 October 2021

The least square problem (LSP)

LSP: for an $m \times n$ matrix A and an m-vector b, find the n-vector $x_{\min}$ minimizing the quantity

$$\|Ax - b\|_2$$

$\rightsquigarrow$ gives the linear combination of the columns

$$Ax = \sum_{j=1}^{n} x_j \operatorname{col}_j(A) \in \mathbb{R}^m$$

that best approaches (in the 2-norm) the m-vector b

- If $m = n$ and A is nonsingular then $Ax_{\min} = b$
- If $m > n$ then typically $Ax = b$ has no solution, and the minimum is positive

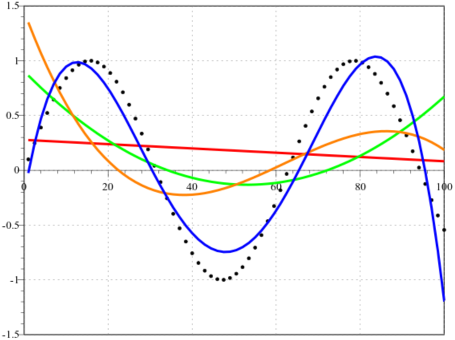
Central case: $m \geqslant n$ and $\operatorname{rank}(A) = n$

Example

Suppose we have pairs

$$(y_i, b_i) \in \mathbb{R}^2, \quad i = 1, \ldots, m,$$

and we want to find the polynomial of degree d that best fits the b_i's as a function of the y_i's



Example (cont.)

Boils down to computing the polynomial

$$p(y) = \sum_{j=0}^{d} x_j \, y^j$$

minimizing the 2-norm of the *residual* $p(y_i) - b_i$, $i = 1, \ldots, m$:

$$\sum_{i=1} (p(y_i) - b_i)^2$$

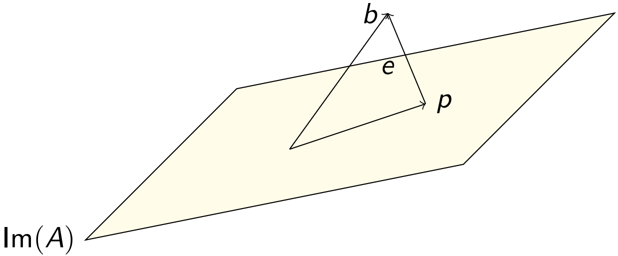
LSP for the data

$$A = \begin{bmatrix} 1 & y_1 & \cdots & y_1^d \\ \vdots & \vdots & & \vdots \\ 1 & y_m & \cdots & y_m^d \end{bmatrix} \in \mathbb{R}^{m \times (d+1)} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m$$

Three ways

- normal equations
- QR factorization
- SVD

LSP and its solution



- $p = A\,x_{\text{min}}$ the orthogonal projection of b into $\mathrm{Im}(A)$
- $e = b - p$ the error vector

Normal equations

The m-vector e is perpendicular to $\mathrm{Im}(A)$: for all $x \in \mathbb{R}^n$

$$0 = \langle Ax, e \rangle = (Ax)^T (b - Ax_{\mathrm{min}}) = x^T A^T (b - Ax_{\mathrm{min}})$$

Hence

$$\boxed{A^T A x_{\mathrm{min}} = A^T b} \tag{1}$$

The $n \times n$ matrix $A^T A$ is symmetric and positive definite

$\Rightarrow$ it is nonsingular and x_{min} is the only solution of (1)

By Pythagoras theorem, the 2-norm of the error is

$$\|e\|_2 = (\|b\|_2 - \|Ax\|_2^2)^{1/2}$$

Normal equations algorithm

We can solve the normal equations with Cholesky factorization:

Compute the lower triangular part of $C \leftarrow A^T A$
Compute $d \leftarrow A^T b$
Compute the Cholesky factorization $C = G\, G^T$
Solve $G\, y = d$ and $G^T x_{\min} = y$

Its *complexity* is

$$\left(m + \frac{n}{3}\right) n^2 + O(n^2) \text{ flops}$$

For $m \gg n$ the complexity $m\, n^2$ of computing the product $A^T A$ dominates

- uses standard algorithms
- reliable when A is far from rank deficient (but unstable otherwise)