Least squares and the singular value decomposition

Ivan Markovsky

University of Southampton

Outline

- QR and SVD decompositions
- Least squares and least norm problems
- Extensions of the least squares problem
  - Recursive
  - Multiobjective
  - Regularized
  - Constrained

QR and SVD decompositions

Orthonormal set of vectors

Consider a finite set of vectors $\mathscr{Q} := \{ q_1, \dots, q_k \} \subset \mathbb{R}^n$

- $\mathscr{Q}$ is orthogonal $:\iff \langle q_i, q_j \rangle := q_i^\top q_j$, for all $i \neq j$
- $\mathscr{Q}$ is normalized $:\iff \|q_i\|_2^2 := \langle q_i, q_i \rangle = 1$, $i = 1, \dots, k$
- $\mathscr{Q}$ is orthonormal $:\iff \mathscr{Q}$ is orthogonal and normalized

with $Q := \begin{bmatrix} q_1 & \cdots & q_k \end{bmatrix}$, $\quad \mathscr{Q}$ orthonormal $\iff Q^\top Q = I_k$

Properties:

- orthonormal vectors are independent
- multiplication with Q preserves inner product and norm

$$\langle Qz, Qy \rangle = z^\top Q^\top Q y = z^\top y = \langle z, y \rangle$$

Orthogonal projectors

Consider orthonormal set $\mathscr{Q} := \{q_1, \dots, q_k\}$ and $\mathscr{L} := \operatorname{span}(\mathscr{Q}) \subseteq \mathbb{R}^n$.

$\mathscr{Q}$ is an orthonormal basis for $\mathscr{L}$.

With $Q := \begin{bmatrix} q_1 & \cdots & q_k \end{bmatrix}$, $Q^\top Q = I_k$, however, for $k < n$, $QQ^\top \neq I_n$.

$\Pi_{\operatorname{span}(\mathscr{Q})} := QQ^\top$ is an orthogonal projector on span$(\mathscr{Q})$, *i.e.*,

$$\Pi_{\mathscr{L}} x = \arg\min_{y} \|x - y\|_2 \quad \text{subject to} \quad y \in \mathscr{L}$$

Properties: $\Pi = \Pi^2$, $\Pi = \Pi^\top$ (necessary and sufficient for Π orth. proj.)

$\Pi^\perp := (I - \Pi)$ is also orthogonal projector, it projects on

$\big(\operatorname{col}\operatorname{span}(\Pi)\big)^\perp \subseteq \mathbb{R}^n$ — orth. complement of the column span of Π

Orthonormal basis for $\mathbb{R}^n$

orthonormal set $\mathcal{Q} := \{ q_1, \dots, q_k \} \subset \mathbb{R}^n$ of $k = n$ vectors

then $Q := \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}$ is called orthogonal and satisfies $Q^\top Q = I_n$

It follows that $Q^{-1} = Q^\top$ and

$$QQ^\top = \sum_{i=1}^{n} q_i q_i^\top = I_n$$

Expansion in orthonormal basis $x = QQ^\top x$

- $\widetilde{x} := Q^\top x$ coordinates of x in the basis $\mathcal{Q}$
- $x = Q\widetilde{x}$ reconstruct x from the coordinates a

Geometrically multiplication by Q (and $Q^\top$) is rotation.

Gram-Schmidt (G-S) procedure

Given independent set $\{a_1, \ldots, a_k\} \subset \mathbb{R}^n$,

G-S produces orthonormal set $\{q_1, \ldots, q_k\} \subset \mathbb{R}^n$ such that

$$\text{span}(a_1, \ldots, a_r) = \text{span}(q_1, \ldots, q_r), \qquad \text{for all } r \leq k$$

G-S procedure: Let $q_1 := a_1 / \|a_1\|_2$. At the ith step $i = 2, \ldots, k$

- orthogonalized a_i w.r.t. $q_1, \ldots, q_{i-1}$:

$$v_i := \underbrace{(I - \Pi_{\text{span}(q_1, \ldots, q_{i-1})}) a_i}_{\text{projection of } a_i \text{ on } (\text{span}(q_1, \ldots, q_{i-1}))^\perp}$$

- normalize the result: $q_i := v_i / \|v_i\|_2$

QR decomposition

G-S procedure gives as a byproduct scalars r_{ji}, $j \leq i$, $i = 1, \ldots, k$, s.t.

$$\begin{aligned} a_i &= (q_1^\top a_i) q_1 + \cdots + (q_{i-1}^\top a_i) q_{i-1} + \|q_i\|_2 q_i \\ &= r_{1i} q_1 + \cdots + r_{ii} q_i \end{aligned}$$

in a matrix form G-S produces the matrix decomposition

$$\underbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} q_1 & q_1 & \cdots & q_k \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ 0 & r_{22} & \cdots & r_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{kk} \end{bmatrix}}_{R}$$

with orthonormal $Q \in \mathbb{R}^{n \times k}$ and upper triangular $R \in \mathbb{R}^{k \times k}$

If $\{a_1, \ldots, a_k\}$ are dependent, $v_i := (I - \Pi_{\text{span}(q_1,\ldots,q_{i-1})})a_i = 0$ for some i

Conversely, if $v_i = 0$ for some i, a_i is linearly dependent on $\{a_1, \ldots, a_{i-1}\}$

Modified G-S procedure: when $v_i = 0$, skip to the next input vector a_{i+1}

$\implies$ R is in upper staircase form, *e.g.*,

$$\begin{bmatrix} \times & \times & \times & \times & \times & \times & \times \\ & \times & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times \\ & & & & & & \times \\ & & & & & & \end{bmatrix}$$

(empty elements are zeros)

Full QR

$$A = \underbrace{\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}}_{\text{orthogonal}} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

$$\begin{aligned} \operatorname{col span}(A) &= \operatorname{col span}(Q_1) \\ \big(\operatorname{col span}(A)\big)^\perp &= \operatorname{col span}(Q_2) \end{aligned}$$

Procedure for finding Q_2:

complete A to full rank matrix, *e.g.*, $A_\mathrm{m} := \begin{bmatrix} A & I \end{bmatrix}$, and apply G-S on A_m

Application: complete an orthonormal matrix $Q_1 \in \mathbb{R}^{n \times k}$ to an orthogonal matrix $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \in \mathbb{R}^{n \times n}$

(by computing the full QR of $\begin{bmatrix} Q_1 & I \end{bmatrix}$)

Singular value decomposition (SVD)

The SVD is used as both computational and analytical tool.

Any $m \times n$ matrix A of rank r has a reduced SVD

$$A = \underbrace{\begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}}_{U_1} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}}_{\Sigma_1} \underbrace{\begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix}^\top}_{V_1^\top}$$

where U_1 and V_1 are orthonormal

- $\sigma_1 \geq \cdots \geq \sigma_r$ are called singular values
- $u_1, \ldots, u_r$ are called left singular vectors
- $v_1, \ldots, v_r$ are called right singular vectors

Full SVD $\quad A = U\Sigma V^\top$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal and

$$\Sigma = \begin{array}{cc} \begin{array}{cc} r & n-r \end{array} & \\ \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{c} r \\ m-r \end{array} \end{array} \qquad \text{where} \qquad \Sigma_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$$

Note that the singular values of A are

$$\sigma(A) := (\sigma_1, \ldots, \sigma_r, \underbrace{0, \ldots, 0}_{\min(n-r, m-r)})$$

- $\sigma_{\min}(A)$ — smallest singular value of A
- $\sigma_{\max}(A)$ — largest singular value of A

Proof of existence of an SVD

The proof is constructive and uses induction. W.l.o.g. assume $m \geq n$.

- End of induction: vector $A \in \mathbb{R}^{m \times 1}$ has (unique) SVD

$$A = U\Sigma V^\top, \qquad \text{with} \quad U := A/\|A\|_2, \quad \Sigma := \|A\|_2, \quad V := 1$$

- Inductive step: choose $v_i \in \mathbb{R}^n$ with $\|v_i\|_2 = 1$ and let

$$A_i v_i =: \sigma_i u_i, \qquad \text{where} \quad \sigma_i := \|A_i\|_2$$

Complete v_i and u_i to orthogonal matrices (QR decomp.)

$$V_i := \begin{bmatrix} v_i & \star \end{bmatrix} \qquad \text{and} \qquad U_i := \begin{bmatrix} u_i & \star \end{bmatrix}$$

We have that for certain $w \in \mathbb{R}^{n-1}$ and $A_{i+1} \in \mathbb{R}^{(n-1)\times(n-1)}$

$$U_i^\top A_i V_i = \begin{bmatrix} \sigma_i & w^\top \\ 0 & A_{i+1} \end{bmatrix}$$

Next we show that $w = 0$.

Proof of existence of an SVD

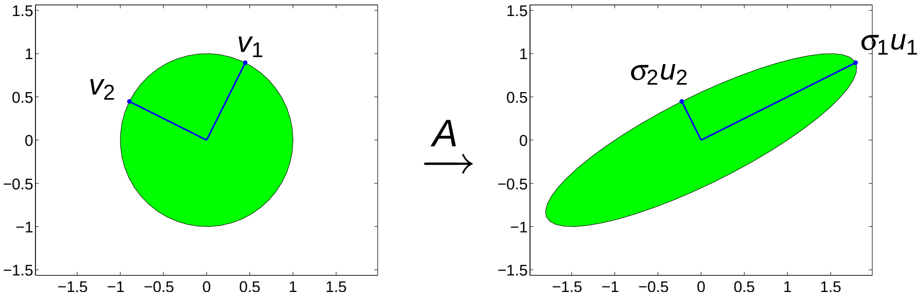
$$
\begin{aligned}
\sigma_i^2 = \|A_i\|_2^2 = \|U_i^\top A_i V_i\|_2^2 \\
&= \max_v \frac{\|A_i v\|_2^2}{\|v\|_2^2} \\
&\geq \frac{\|A_i \left[\begin{smallmatrix}\sigma_i \\ w\end{smallmatrix}\right]\|_2^2}{\|\left[\begin{smallmatrix}\sigma_i \\ w\end{smallmatrix}\right]\|_2^2} \\
&= \frac{1}{\sigma_i^2 + w^\top w}\left\|\begin{bmatrix}\sigma_i^2 + w^\top w \\ A_{i+1} w\end{bmatrix}\right\|_2^2 \\
&\geq \frac{1}{\sigma_i^2 + w^\top w}(\sigma_i^2 + w^\top w)^2 = \sigma_i^2 + w^\top w
\end{aligned}
$$

The inequality $\sigma_i^2 \geq \sigma_i^2 + w^\top w$ can be true only when $w = 0$.

Geometric fact motivating the SVD

The image of a unit ball under linear map is a hyperellips.

$$\underbrace{\begin{bmatrix} 1.00 & 1.50 \\ 0 & 1.00 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 0.89 & -0.45 \\ 0.45 & 0.89 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 2.00 & 0 \\ 0 & 0.50 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 0.45 & -0.89 \\ 0.89 & 0.45 \end{bmatrix}}_{V^\top}$$



Low-rank approximation

Given

- a matrix $A \in \mathbb{R}^{m \times n}$, $m \geq n$, and
- an integer r, $0 < r < n$,

find

$$\widehat{A} := \arg\min_{\widehat{A}} \|A - \widehat{A}\| \quad \text{subject to} \quad \operatorname{rank}(\widehat{A}) \leq r$$

Interpretation:

$\widehat{A}^*$ is optimal rank-r approximation of A w.r.t. the norm $\|\cdot\|$, *e.g.*,

$$\|A\|_{\mathrm{F}}^2 := \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \qquad \text{or} \qquad \|A\|_2 := \max_x \frac{\|Ax\|_2}{\|x\|_2}$$

Solution via SVD

$$\widehat{A}^* := \arg\min_{\widehat{A}} \|A - \widehat{A}\|_{\mathrm{F}} \quad \text{subject to} \quad \operatorname{rank}(\widehat{A}) \leq r \qquad \text{(LRA)}$$

Theorem Let $A = U\Sigma V^\top$ be the SVD of A and define

$$U =: \begin{array}{cc} r & r-n \\ \left[U_1\right. & \left.U_2\right] \end{array} \; n \qquad \Sigma =: \begin{array}{cc|c} r & r-n & \\ \Sigma_1 & 0 & r \\ 0 & \Sigma_2 & r-n \end{array} \qquad \text{and} \qquad V =: \begin{array}{cc} r & r-n \\ \left[V_1\right. & \left.V_2\right] \end{array} \; n$$

A solution to (LRA) is

$$\widehat{A}^* = U_1 \Sigma_1 V_1^\top$$

It is unique if and only if $\sigma_r \neq \sigma_{r+1}$.

Proof of the low-rank approximation theorem

Let $\widehat{A}^*$ be solution to (LRA) and let $\widehat{A}^* := U^*\Sigma^*(V^*)^\top$ be an SVD of $\widehat{A}^*$.

$$\|A - \widehat{A}^*\|_{\mathrm{F}} = \|\underbrace{(U^*)^\top A V^*}_{B} - \Sigma^*\|_{\mathrm{F}} \quad \Longrightarrow \quad \Sigma^* \text{ is an opt. approx. of } B$$

Partition $B =: \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ conformably with $\Sigma^* =: \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{bmatrix}$ and observe that

$$\operatorname{rank}\left(\begin{bmatrix} \Sigma_1^* & B_{12} \\ 0 & 0 \end{bmatrix}\right) \leq r \quad \text{and} \quad B_{12} \neq 0 \implies \left\| B - \begin{bmatrix} \Sigma_1^* & B_{12} \\ 0 & 0 \end{bmatrix} \right\|_{\mathrm{F}} < \left\| B - \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{bmatrix} \right\|_{\mathrm{F}}$$

so that $B_{12} = 0$. Similarly $B_{21} = 0$. Observe also that

$$\operatorname{rank}\left(\begin{bmatrix} B_{11} & 0 \\ 0 & 0 \end{bmatrix}\right) \leq r \quad \text{and} \quad B_{11} \neq \Sigma_1^* \implies \left\| B - \begin{bmatrix} B_{11} & 0 \\ 0 & 0 \end{bmatrix} \right\|_{\mathrm{F}} < \left\| B - \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{bmatrix} \right\|_{\mathrm{F}}$$

so that $B_{11} = \Sigma_1^*$. Therefore, $B = \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & B_{22} \end{bmatrix}$.

Proof of the low-rank approximation theorem

Let $B_{22} = U_{22}\Sigma_{22}V_{22}^\top$ be the SVD of B_{22}. Then the matrix

$$\begin{bmatrix} I & 0 \\ 0 & U_{22}^\top \end{bmatrix} B \begin{bmatrix} I & 0 \\ 0 & U_{22} \end{bmatrix} = \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$$

has optimal rank-r approximation $\Sigma^* = \left[\begin{smallmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{smallmatrix}\right]$, so that

$$\min(\text{diag}(\Sigma_1^*)) > \max(\text{diag}(U_{22}))$$

Therefore

$$A = U^* \begin{bmatrix} I & 0 \\ 0 & U_{22} \end{bmatrix} \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & U_{22}^\top \end{bmatrix} (V^*)^\top$$

is an SVD of A.

Proof of the low-rank approximation theorem

SVD of A:

$$A = U^* \begin{bmatrix} I & 0 \\ 0 & U_{22} \end{bmatrix} \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & U_{22}^\top \end{bmatrix} (V^*)^\top$$

Then, if $\sigma_r > \sigma_{r+1}$, the rank-r SVD truncation

$$\widehat{A}^* = U^* \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{bmatrix} (V^*)^\top = U^* \begin{bmatrix} I & 0 \\ 0 & U_{22} \end{bmatrix} \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & U_{22}{}^\top \end{bmatrix} (V^*)^\top$$

is unique and $\widehat{A}^*$ is the unique solution of (LRA).

Note that $\widehat{A}^*$ is simultaneously optimal in any unitarily invariant norm.

Numerical rank

$$\sqrt{\sum_{i=r+1}^{n} \sigma_i^2} = \min_{\widehat{A}} \|A - \widehat{A}\|_{\mathrm{F}} \quad \text{subject to} \quad \operatorname{rank}(\widehat{A}) \le r$$

and

$$\sigma_{r+1} = \min_{\widehat{A}} \|A - \widehat{A}\|_2 \quad \text{subject to} \quad \operatorname{rank}(\widehat{A}) \le r$$

are measures of the distance of A to the manifold of rank-r matrices

In particular, $\sigma_{\min}(A)$ is the distance of A to rank deficiency.

$\operatorname{rank}(A, \varepsilon) :=$ # of singular values $> \varepsilon$ is called numerical rank of A

Note that $\operatorname{rank}(A, \varepsilon)$ depends on an a priori given tolerance ε.

Pseudo-inverse $\quad A^+ := V_1 \Sigma_1^{-1} U_1^\top \in \mathbb{R}^{n \times m}$

$$
\begin{aligned}
\operatorname{rank}(A) = n = m \quad &\Longrightarrow \quad A^+ = A^{-1} \\
\operatorname{rank}(A) = n \quad &\Longrightarrow \quad A^+ = (A^\top A)^{-1} A^\top \\
\operatorname{rank}(A) = m \quad &\Longrightarrow \quad A^+ = A^\top (A A^\top)^{-1}
\end{aligned}
$$

In general, $A^+ y$ is the least squares, least norm solution of $Ax = y$

Note that the pseudo-inverse depends on the rank of A.
In practice the numerical rank $\operatorname{rank}(A, \varepsilon)$ is used.

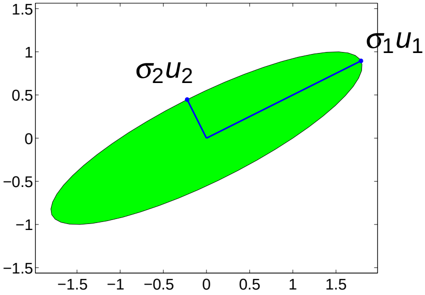
The SVD, using numerical rank and pseudo-inverse, is the most reliable way of solving $Ax = y$.

It should be used in cases when A is ill-conditioned.

Condition number $\kappa(A) := \sigma_{\max}(A)/\sigma_{\min}(A)$

Geometrically $\kappa(A)$ is the eccentricity of the hyperellipsoid

$$A\{x \mid \|x\|_2 = 1\}$$



$\kappa(A)$ measures the sensitivity of A^+y to perturbations in y and A

For large $\kappa(A)$ (above a few 1000) A is called ill-conditioned.

Least squares and least norm

Least squares

- consider an overdetermined system of linear equations $Ax = y$
- problem: given $A \in \mathbb{R}^{m \times n}$, $m > n$ and $y \in \mathbb{R}^m$, find $x \in \mathbb{R}^n$
- for "most" A and y, there is no solution x
- Least squares approximation:

  choose x that minimizes 2-norm of the residual (eqn. error)

$$e(x) := y - Ax$$

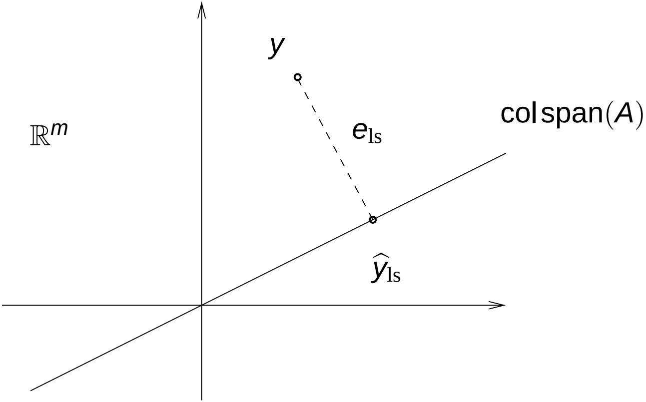
- a minimizing x is called a least squares approximate solution

$$\widehat{x}_{\text{ls}} := \arg\min_x \| \underbrace{y - Ax}_{e(x)} \|_2$$

Geometric interpretation: project y onto the image of A

($\widehat{y}_{\mathrm{ls}} := A\widehat{x}_{\mathrm{ls}}$ is the projection)

$$e_{\mathrm{ls}} := \widehat{y}_{\mathrm{ls}} - A\widehat{x}_{\mathrm{ls}}$$

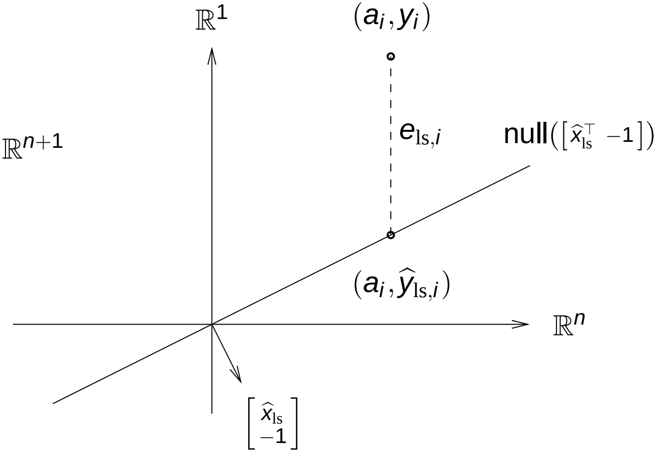


$$\begin{aligned} A\widehat{x}_{\mathrm{ls}} = \widehat{y}_{\mathrm{ls}} \quad &\iff \quad \begin{bmatrix} A & \widehat{y}_{\mathrm{ls}} \end{bmatrix} \begin{bmatrix} \widehat{x}_{\mathrm{ls}} \\ -1 \end{bmatrix} = 0 \\ &\iff \quad \begin{bmatrix} a_i & \widehat{y}_{\mathrm{ls},i} \end{bmatrix} \begin{bmatrix} \widehat{x}_{\mathrm{ls}} \\ -1 \end{bmatrix} = 0, \quad \text{for } i = 1, \ldots, m \end{aligned}$$

(a_i is the ith row of A)

- $(a_i, \widehat{y}_{\mathrm{ls},i})$, for all i, lies on the subspace perpendicular to $(\widehat{x}_{\mathrm{ls}}, -1)$
- "data point" $(a_i, y_i) = (a_i, \widehat{y}_{\mathrm{ls},i}) + (0, e_{\mathrm{ls},i})$
- the approximation error $(0, e_{\mathrm{ls},i})$ is the vertical distance from (a_i, y_i) to the subspace

Another geometric interpretation of the LS approximation:



Notes

Assuming $m \geq n = \text{rank}(A)$, *i.e.*, A is full column rank,

$$\widehat{x}_{\text{ls}} = (A^\top A)^{-1} A^\top y$$

is the unique least squares approximate solution.

- $\widehat{x}_{\text{ls}}$ is a linear function of y
- If A is square $\widehat{x}_{\text{ls}} = A^{-1} y$
- $\widehat{x}_{\text{ls}}$ is an exact solution if $Ax = y$ has an exact solution
- $\widehat{y}_{\text{ls}} := A\widehat{x}_{\text{ls}} = A(A^\top A)^{-1} A^\top y$ is a least squares approximation of y

Projector onto the span of A

The $m \times m$ matrix

$$\Pi_{\text{col span}(A)} := A(A^\top A)^{-1}A^\top$$

is the orthogonal projector onto $\mathscr{L} := \text{col span}(A)$.

The columns of A are an arbitrary basis for $\mathscr{L}$.

If the columns of Q form an orthonormal basis for $\mathscr{L}$

$$\Pi_{\text{col span}(Q)} := QQ^\top$$

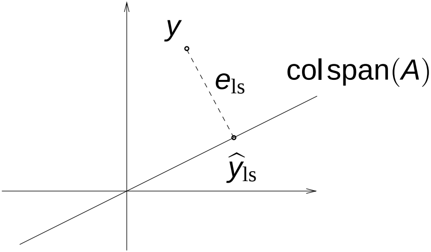
Orthogonality principle

The least squares residual vector

$$e_{\text{ls}} := y - A\widehat{x}_{\text{ls}} = \underbrace{\big(I_m - A(A^\top A)^{-1}A^\top\big)}_{\Pi_{(\text{col span}(A))^\perp}} y$$

is orthogonal to $\text{col span}(A)$

$$\langle e_{\text{ls}}, A\widehat{x}_{\text{ls}}\rangle = y^\top\big(I_m - A(A^\top A)^{-1}A^\top\big)A\widehat{x}_{\text{ls}} = 0, \quad \text{for all } x \in \mathbb{R}^n$$



Least squares via QR decomposition

Let $A = QR$ be the QR decomposition of A.

$$(A^\top A)^{-1}A^\top = (R^\top Q^\top QR)^{-1}R^\top Q^\top$$
$$= (R^\top Q^\top QR)^{-1}R^\top Q^\top = R^{-1}Q^\top$$

so that

$$\widehat{x}_{\text{ls}} = R^{-1}Q^\top y \quad \text{and} \quad \widehat{y}_{\text{ls}} := Ax_{\text{ls}} = QQ^\top y$$

Let $A =: \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$ and consider the sequence of LS problems

$$A^i x^i = y, \qquad \text{where } A^i := \begin{bmatrix} a_1 & \cdots & a_i \end{bmatrix}, \quad \text{for } i = 1, \dots, n$$

Define R_i as the leading $i \times i$ submatrix of R and $Q_i := \begin{bmatrix} q_1 & \cdots & q_i \end{bmatrix}$.

$$\widehat{x}^i_{\text{ls}} = R_i^{-1}Q_i^\top y$$

Least norm solution

Consider an underdetermined system $Ax = y$, with full rank $A \in \mathbb{R}^{m \times n}$.

The set of solutions is

$$\{ x \in \mathbb{R}^n \mid Ax = y \} = \{ x_{\mathrm{p}} + z \mid z \in \mathrm{null}(A) \}$$

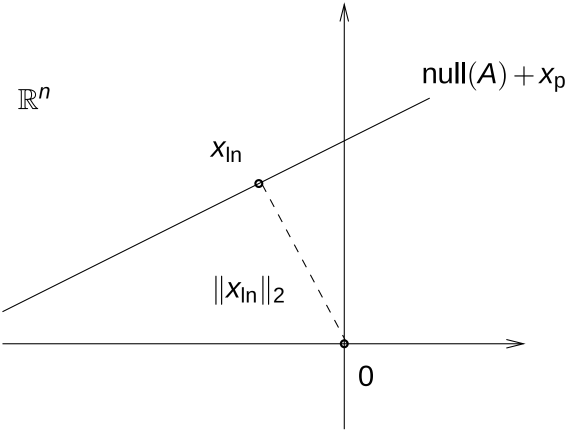
where x_{p} is a particular solution, *i.e.*, $Ax_{\mathrm{p}} = y$.

Least norm problem

$$x_{\mathrm{ln}} := \arg\min_x \|x\|_2 \quad \text{subject to} \quad Ax = y$$

Geometric interpretation:

- x_{ln} is the projection of 0 onto the solution set
- orthogonality principle $x_{\text{ln}} \perp \text{null}(A)$



Derivation of the solution: Lagrange multipliers

Consider the least norm problem with A full rank

$$\min_x \|x\|_2^2 \quad \text{subject to} \quad Ax = y$$

introduce Lagrange multipliers $\lambda \in \mathbb{R}^m$

$$L(x, \lambda) = xx^\top + \lambda^\top (Ax - y)$$

the optimality conditions are

$$\begin{aligned} \nabla_x L(x, \lambda) &= 2x + A^\top \lambda = 0 \\ \nabla_\lambda L(x, \lambda) &= Ax - y = 0 \end{aligned}$$

from the first condition $x = -A^\top \lambda / 2$, substituting into the second

$$\lambda = -2(AA^\top)^{-1} y \quad \Longrightarrow \quad x_{\text{ln}} = A^\top (AA^\top)^{-1} y$$

Solution via QR decomposition

Let $A^\top = QR$ be the QR decomposition of $A^\top$.

$$A^\top(AA^\top)^{-1} = QR(R^\top Q^\top QR)^{-1} = Q(R^\top)^{-1}$$

is a right inverse of A. Then

$$x_{\mathrm{ln}} = Q(R^\top)^{-1}y$$

Extensions

Weighted least squares

Given a positive definite matrix $W \in \mathbb{R}^{m \times m}$, define wighted 2-norm

$$\|e\|_W^2 := e^\top W e$$

Weighted least squares approximation problem

$$\widehat{x}_{W,\mathrm{ls}} := \arg\min_x \|y - Ax\|_W$$

The orthogonality principle holds by defining the inner product as

$$\langle e, y \rangle_W := e^\top W y$$

and

$$\widehat{x}_{W,\mathrm{ls}} = (A^\top W A)^{-1} A^\top W y$$

Recursive least squares

Let $a_i^\top$ be the ith row of A

$$A = \begin{bmatrix} — & a_1^\top & — \\ & \vdots & \\ — & a_m^\top & — \end{bmatrix}$$

with this notation, $\|y - Ax\|_2^2 = \sum_{i=1}^m (y_i - a_i^\top x)^2$ and

$$\widehat{x}_{\text{ls}} = \widehat{x}_{\text{ls}}(m) := \left(\sum_{i=1}^m a_i a_i^\top \right)^{-1} \sum_{i=1}^m a_i y_i$$

- (a_i, y_i) correspond to a measurement
- often the measurements (a_i, y_i) come sequentially (*e.g.*, in time)

Recursive computation of $\widehat{x}_{\mathrm{ls}}(m) = \left(\sum_{i=1}^{m} a_i a_i^\top\right)^{-1} \sum_{i=1}^{m} a_i y_i$

- $P(0) = 0 \in \mathbb{R}^{n\times n}$, $q(0) = 0 \in \mathbb{R}^n$
- For $m = 0, 1, \ldots$
- $P(m+1) := P(m) + a_{m+1}a_{m+1}^\top$, $q(m+1) := q(m) + a_{m+1}y_{m+1}$.
- If $P(m)$ is invertible, $x_{\mathrm{ls}}(m) = P^{-1}(m)q(m)$.

Notes:

- In each step, the algorithm requires inversion of an $n \times n$ matrix
- $P(m)$ invertible $\implies$ $P(m')$ invertible, for all $m' > m$

Rank-1 update formula

$$(P + aa^\top)^{-1} = P^{-1} - \frac{1}{1 + a^\top P^{-1} a}(P^{-1}a)(P^{-1}a)^\top$$

Notes:

- gives an $O(n^2)$ method for computing $P^{-1}(m+1)$ from $P^{-1}(m)$
- standard methods based on dense LU, QR, or SVD for computing $P^{-1}(m+1)$ require $O(n^3)$ operations

Multiobjective least squares

least squares minimizes the cost function $J_1(x) := \|y - Ax\|_2^2$.

Consider a second cost function $J_2(x) := \|z - Bx\|_2^2$,

which we want to minimize together with J_1.

Usually the criteria $\min_x J_1(x)$ and $\min_x J_2(x)$ are competing.

Common example: $J_2(x) := \|x\|_2^2$ — minimize J_1 with small x

- feasible objectives:

  $\{ (\alpha, \beta) \in \mathbb{R}^2 \mid \exists\, x \in \mathbb{R}^n \text{ subject to } J_1(x) = \alpha,\ J_2(x) = \beta \}$
- optimal trade-off curve: boundary of the feasible objectives
- the corresponding x is called Pareto optimal

Set of Pareto optimal solutions

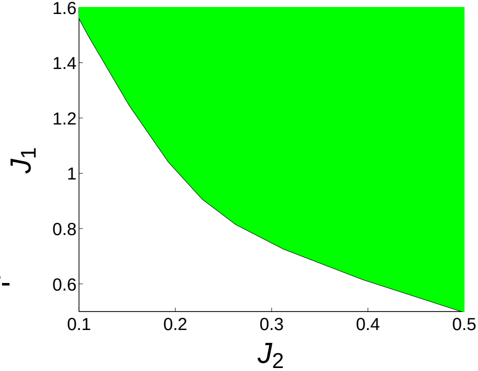
Example:

green area — feasible

white area — infeasible

black line — marginally feasible

Pareto optimal solutions correspond to points on the line



For any $\mu \geq 0$, $\widehat{x}(\mu) = \arg\min_x J_1(x) + \mu J_2(x)$ is Pareto optimal.

By varying $\mu \in [0, \infty)$, $\widehat{x}(\mu)$ sweeps all Pareto optimal solutions

Regularized least squares

Tychonov regularization

$$\widehat{x}_{\text{tych}}(\mu) = \arg\min_x \|y - Ax\|_2^2 + \mu\|x\|_2^2$$

the solution

$$\widehat{x}_{\text{tych}}(\mu) = (A^\top A + \mu I_n)^{-1} A^\top y$$

exists for any $\mu > 0$, independent on size and rank of A.

Trade-off between

- fitting accuracy $J_1(x) = \|y - Ax\|_2$, and
- solution size $J_2(x) = \|x\|_2$.

Quadratically constrained least squares

Consider again the biobjective LS problem $\min_x J_1(x)$ and $J_2(x)$

Scalarization approach:

$$\widehat{x}_{\text{tych}}(\mu) = \arg\min_x \; J_1(x) + \mu J_2(x)$$

where μ is trade-off parameter

Constrained optimization approach:

$$\widehat{x}_{\text{constr}}(\gamma) = \arg\min_x \; J_1(x) \quad \text{subject to} \quad J_2(x) \leq \gamma$$

where γ is upper bound on the J_2 objective

Regularized least squares

Tychonov regularization corresponds to the scalarization approach for

- fitting accuracy $J_1(x) = \|y - Ax\|_2$, and
- solution size $J_2(x) = \|x\|_2$.

The constrained optimization approach leads in this case to

$$\widehat{x}_{\text{constr}}(\gamma) = \arg\min_x \|y - Ax\|_2^2 \quad \text{subject to} \quad \|x\|_2^2 \leq \gamma^2$$

least squares minimization over the ball $\mathscr{U}_{\gamma^2} := \{\, x \mid \|x\|_2^2 \leq \gamma^2 \,\}$.

The solution to the latter problem involves scalar nonlinear equation.

Secular equation

If $\|A^+y\|_2^2 \leq \gamma^2$, then $\widehat{x}_{\text{constr}}(\gamma) = \|A^+y\|_2^2$.

If $\|A^+y\|_2^2 > \gamma^2$, then it can be shown that $\widehat{x}_{\text{constr}}(\gamma) \in \mathscr{U}_{\gamma^2}$.

The Lagrangian of

$$\text{minimize}_x \quad \|y - Ax\|_2^2 \quad \text{subject to} \quad \|x\|_2^2 = \gamma^2$$

is $\|y - Ax\|_2^2 + \mu(\|x\|_2^2 - \gamma^2)$, where μ is a Lagrange multiplier.

Necessary and sufficient optimality condition is

$$x_{\text{tych}}^\top(\mu) x_{\text{tych}}(\mu) = \gamma^2, \qquad \text{where} \quad x_{\text{tych}}(\mu) := (A^\top A + \mu I)^{-1} y$$

The nonlinear equation in μ

$$y^\top (A^\top A + \mu I)^{-2} y = \gamma^2$$

is called secular equation. It has unique positive solution because $\|x_{\text{tych}}(\mu)\|$ is monotonically decreasing on the interval $\mu \in [0, \infty)$ and by assumption $\|x_{\text{tych}}(0)\|_2^2 > \gamma^2$.

Total least squares (TLS)

The LS method minimizes 2-norm of the equation error $e(x) := y - Ax$.

$$\min_{x,e} \|e\|_2 \quad \text{subject to} \quad Ax = y - e$$

alternatively the equation error e can be viewed as a correction on y.

The TLS method is motivated by the asymmetry of the LS method:

both A and y are given data, but only y is corrected.

TLS problem: $\displaystyle \min_{x,\Delta A,\Delta y} \left\| \begin{bmatrix} \Delta A & \Delta y \end{bmatrix} \right\|_{\mathrm{F}} \quad \text{subject to} \quad (A + \Delta A)x = y + \Delta y$

- ΔA — correction on A, Δy — correction on y
- Frobenius matrix norm: $\|C\|_{\mathrm{F}} := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}^2}$, where $C \in \mathbb{R}^{m \times n}$

Geometric interpretation of the TLS criterion

In the case $n = 1$, the problem of solving approximately $Ax = y$ is

$$\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} x = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \qquad x \in \mathbb{R}$$

Geometric interpretation:

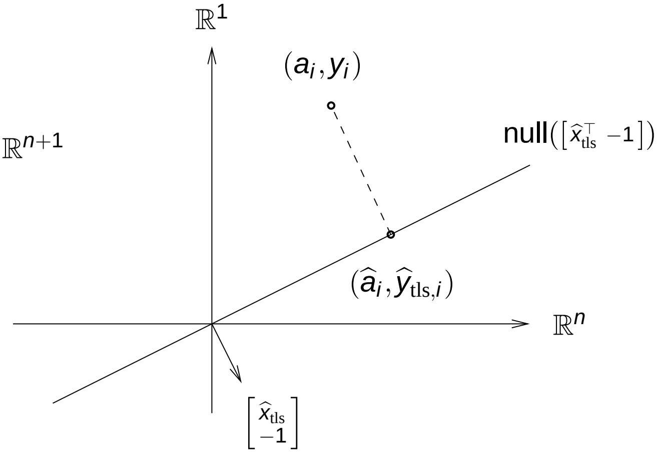
fit a line $\mathscr{L}(x)$ passing through 0 to the points $(a_1, y_1), \dots, (a_m, y_m)$

- LS minimizes

  sum of squared vertical distances from (a_i, y_i) to $\mathscr{L}(x)$

- TLS minimizes

  sum of squared orthogonal distances from (a_i, y_i) to $\mathscr{L}(x)$


$\mathbb{R}^1$
(a_i, y_i)
$\mathbb{R}^{n+1}$
$\mathrm{null}([\widehat{x}_{\mathrm{tls}}^\top \; -1])$
$(\widehat{a}_i, \widehat{y}_{\mathrm{tls},i})$
$\mathbb{R}^n$
$\begin{bmatrix} \widehat{x}_{\mathrm{tls}} \\ -1 \end{bmatrix}$

Solution of the TLS problem

Let $\begin{bmatrix} A & y \end{bmatrix} = U\Sigma V^\top$ be the SVD of the data matrix $\begin{bmatrix} A & y \end{bmatrix}$ and

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_{n+1}), \quad U = \begin{bmatrix} u_1 & \cdots & u_{n+1} \end{bmatrix}, \quad V = \begin{bmatrix} v_1 & \cdots & v_{n+1} \end{bmatrix}.$$

A TLS solution of $Ax = y$ exists iff $v_{n+1,n+1} \neq 0$ (last element of v_{n+1}) and is unique iff $\sigma_n \neq \sigma_{n+1}$.

In the case when a TLS solution exists and is unique, it is given by

$$\widehat{x}_{\mathrm{tls}} = -\frac{1}{v_{n+1,n+1}} \begin{bmatrix} v_{1,n+1} \\ \vdots \\ v_{n,n+1} \end{bmatrix}$$

and the corresponding TLS corrections are $\begin{bmatrix} \Delta A_{\mathrm{tls}} & \Delta y_{\mathrm{tls}} \end{bmatrix} = -\sigma_{n+1} u_{n+1} v_{n+1}^\top$

(Corollary of the low-rank approximation theorem, see page 17.)

References

1. S. Boyd.
   EE263: Introduction to linear dynamical systems.
2. G. Golub and C. Van Loan.
   Matrix Computations.
   Johns Hopkins, 1996.
3. L. Trefethen and D. Bau.
   Numerical Linear Algebra.
   SIAM, 1997.
4. B. Vanluyten, J. C. Willems, and B. De Moor.
   Model reduction of systems with symmetries.
   In *Proc. of the CDC*, pages 826–831, 2005.
5. I. Markovsky and S. Van Huffel
   Overview of total least squares methods
   Signal Processing, 87, pages 2283–2302, 2007