Lecture 3: Applications

- Least-squares
- Least-norm
- Total least-squares
- Low-rank approximation

Over/underdetermined linear equations

Consider $Ax = y$ with $A \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$ given and $x \in \mathbb{R}^n$ unknown.

Without loss of generality assume that A is full rank.

- $Ax = y$ is overdetermined if $m > n$ (more eqns than unknowns)
- $Ax = y$ is underdetermined if $m < n$ (more unknowns than eqns)

For most $y \in \mathbb{R}^m$

- overdetermined systems have no solution x
- underdetermined systems have infinitely many solutions x

We are talking about

- Least-squares approximate solution of an overdetermined system
- Least-norm particular solution of an underdetermined system

Notes:

- Least-squares for an underdetermined system, and
- Least-norm for an overdetermined system are meaningless.

- the least-squares approx. solution is (most of the time) not solution
- the least-norm solution is (aways) one of infinitely many solutions

Least-squares

- approach for solving approx. overdetermined system $Ax = y$
- choose x that minimizes 2-norm of the residual (eqn error)

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

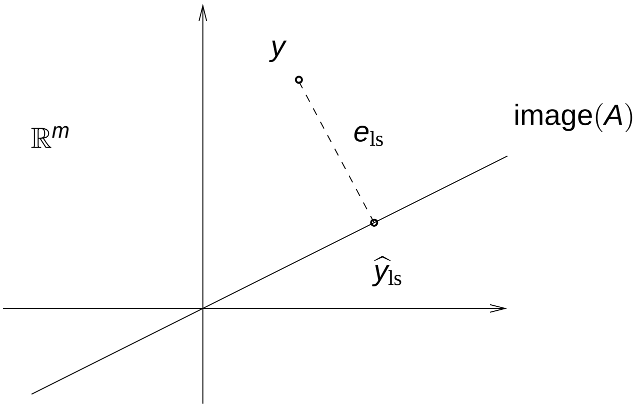
- a minimizing x is called a least-squares approximate solution

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

Geometric interpretation: project y onto the span of A

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

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



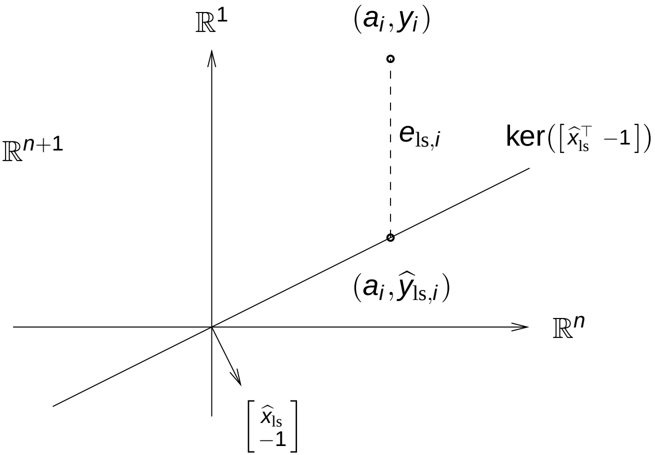
$$A\widehat{x}_{\rm ls} = \widehat{y}_{\rm ls} \quad \iff \quad \begin{bmatrix} A & \widehat{y}_{\rm ls} \end{bmatrix} \begin{bmatrix} \widehat{x}_{\rm ls} \\ -1 \end{bmatrix} = 0$$

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

(a_i is the ith row of A)

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

Another geometric interpretation of the LS approximation:



Derivation of the least-squares solution

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

To minimize the norm of the residual e

$$\|e(x)\|_2^2 = \|y - Ax\|_2^2 = (y - Ax)^\top (y - Ax) = x^\top A^\top Ax - 2y^\top Ax + y^\top y$$

over x, set the gradient with respect to x equal to zero

$$\nabla_x \|e(x)\|_2^2 = \nabla_x (x^\top A^\top Ax - 2y^\top Ax + y^\top y) = 2A^\top Ax - 2A^\top y = 0.$$

This gives the linear equation $A^\top Ax = 2A^\top y$ in x, called normal equation.

A full column rank, implies that $A^\top A$ is nonsingular, so that

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

is the unique least-squares approximate solution.

Notes

- $A^+ := (A^\top A)^{-1}A^\top$ is called the pseudo-inverse of A
- $\widehat{x}_{\mathrm{ls}}$ is a linear function of y (given by the pseudo inverse matrix A^+)
- If A is square $\widehat{x}_{\mathrm{ls}} = A^{-1}y$ (in other words $A^+ = A^{-1}$)
- $\widehat{x}_{\mathrm{ls}}$ is an exact solution if $Ax = y$ has an exact solution
- $\widehat{y}_{\mathrm{ls}} := A\widehat{x}_{\mathrm{ls}} = A(A^\top A)^{-1}A^\top y$ is a least-squares approximation of y

Projector onto the span of A

The $m \times m$ matrix

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

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

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

Recall that if the columns of Q are an orthonormal basis for $\mathscr{L}$

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

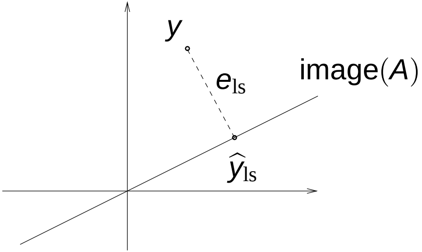
Orthogonality principle

The least-squares residual vector

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

is orthogonal to image(A)

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



Least-squares via QR factorization

Let $A = QR$ be the QR factorization 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 $Q_i := \begin{bmatrix} q_1 & \cdots & q_i \end{bmatrix}$.

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

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 approximate solution

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

The orthogonality principle holds by defining the inner product as

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

Show that

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

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

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

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

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

- $P(0) = 0 \in \mathbb{R}^{n\times n}$, $q(0) = 0 \in \mathbb{R}^n$
- 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, $x_{\text{ls}}(m) = P^{-1}(m)q(m)$.

Notes:

- On 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 for computing $P^{-1}(m+1)$ require $O(n^3)$ operations (for dense matrices)

Multiobjective least-squares

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

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

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

- achievable objectives:

  $\{ (\alpha, \beta) \in \mathbb{R}^2 \mid \exists\, x \in \mathbb{R}^n$ subject to $J_1(x) = \alpha,\ J_2(x) = \beta \}$

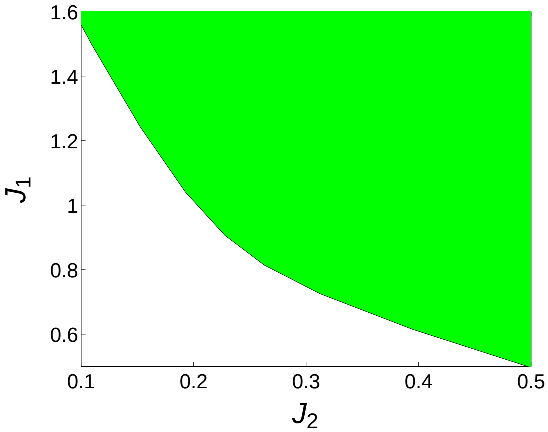
- optimal trade-off curve: boundary of the achievable objectives
- the corresponding x are called Pareto optimal

Scalarization of multiobjective problem

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

Example:



Regularized least-squares

Tychonov regularization

$$\widehat{x} = \arg\min_x \|Ax - b\|_2^2 + \mu \|x\|_2^2$$

the solution

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

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

Trade-off between

- fitting accuracy $\|Ax - b\|_2$, and
- solution size $\|x\|_2$

Least-norm solution

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

The set of solutions is

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

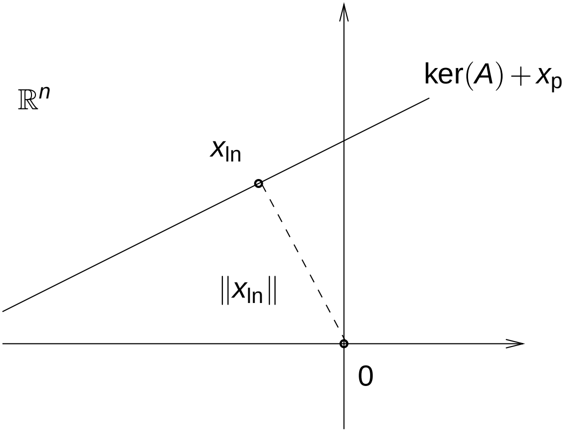
where x_{p} is a particular solution, *i.e.*, $Ax_{\mathrm{p}} = y$.

least-norm solution

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

Geometric interpretation:

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



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 factorization

Let $A^\top = QR$ be the QR factorization 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

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

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

- $\widetilde{A}$ — correction on A, $\qquad \widetilde{y}$ — 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 $(a_1, y_1), \ldots, (a_m, y_m)$

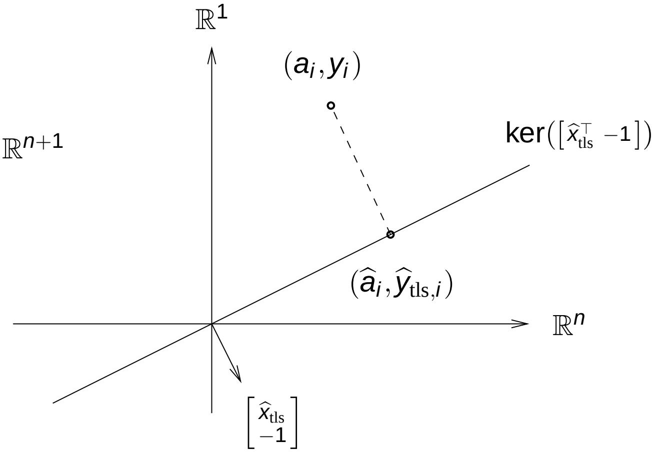
- LS minimizes

  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 $\mathscr{L}(x)$

(Show this algebraically.)


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

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 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

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

and the corresponding TLS corrections are $\begin{bmatrix} \widetilde{A}_{\text{tls}} & \widetilde{y}_{\text{tls}} \end{bmatrix} = -\sigma_{n+1} u_{n+1} v_{n+1}^\top$.

(Corollary of the low-rank approximation theorem, see page 29.)

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

$$U =: \begin{matrix} \scriptstyle r & \scriptstyle r-n \\ \begin{bmatrix} U_1 & U_2 \end{bmatrix} \end{matrix} \; n, \quad \Sigma =: \begin{matrix} \scriptstyle r & \scriptstyle r-n \\ \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \end{matrix} \begin{matrix} r \\ r-n \end{matrix} \quad \text{and} \quad V =: \begin{matrix} \scriptstyle r & \scriptstyle r-n \\ \begin{bmatrix} V_1 & V_2 \end{bmatrix} \end{matrix} \; n.$$

An solution to (LRA) is

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

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

(Outline of the proof.)

Example: linear prediction problem

Future values of w are estimated as linear comb. of past values

$$w(t) = p_1 w(t-1) + p_2 w(t-2) + \cdots + p_\ell w(t-\ell) \qquad \text{(LP)}$$

p_i are the linear prediction coefficients

Given an observed signal w, how do we find the coefficients p_i?

There are many methods for doing this:

- Pisarenko, Prony, Kumaresan–Tufts methods
- subspace methods
- frequency domain methods
- maximum likelihood method

Link to the sum-of-damped-exponentials model

Model the signal w as

$$w(t) = \sum_{i=1}^{\ell} a_i e^{d_i t} e^{\mathbf{i}(\omega_i t + \phi_i)} \qquad \text{(SDE)}$$

where a_i, d_i, ϕ_i, and ω_i are parameters of the model

a_i	—	amplitudes	d_i	—	dampings
ω_i	—	frequencies	ϕ_i	—	initial phases

For all $\{a_i, d_i, \omega_i, \phi_i\}$ there are p_i and $w(-\ell+1), \ldots, w(0)$, s.t. the solution of (LP) coincides with (SDE) and vice verse.

the LP problem $\iff$ modelling by (SDE)

Linear prediction problem as low-rank approx.

$w = \big(w(1),\dots,w(T)\big)$ sum-of-damped-exp. $\implies$ w satisfies

$$p_0 w(t) + p_1 w(t+1) + \cdots + p_\ell w(t+\ell) = 0, \quad \text{for } t = 1,\dots,T-\ell$$

Written in a matrix form these equations are

$$\begin{bmatrix} p_0 & p_1 & \cdots & p_\ell \end{bmatrix} \underbrace{\begin{bmatrix} w(1) & w(2) & \cdots & w(T-\ell) \\ w(2) & w(3) & \cdots & w(T-\ell+1) \\ \vdots & \vdots & & \vdots \\ w(\ell+1) & w(\ell+2) & \cdots & w(T) \end{bmatrix}}_{\mathscr{H}_\ell(w)} = 0$$

which shows that the Hankel matrix $\mathscr{H}_\ell(w)$ is rank deficient

$$\operatorname{rank}\big(\mathscr{H}_\ell(w)\big) \le \ell$$

Structured low-rank approximation

Given

- a vector $p \in \mathbb{R}^{n_p}$,
- a mapping $\mathscr{S} : \mathbb{R}^{n_p} \to \mathbb{R}^{m \times n}$ (structure specification)
- a vector norm $\|\cdot\|$, and
- an integer r, $0 < r < \min(m,n)$,

find

$$\widehat{p}^* := \arg\min_{\widehat{p}} \|p - \widehat{p}\| \quad \text{subject to} \quad \operatorname{rank}\big(\mathscr{S}(\widehat{p})\big) \leq r.$$

Interpretation:

$\widehat{D}^* := \mathscr{S}(\widehat{p}^*)$ is optimal rank-r (or less) approx. of $D := \mathscr{S}(p)$,
within the class of matrices with the same structure as D.

Solution methods for structured low-rank appr.

No closed form solution is known for the general SLRA problem

$$\widehat{p}^* := \arg\min_{\widehat{p}} \|p - \widehat{p}\| \quad \text{subject to} \quad \operatorname{rank}\big(\mathscr{S}(\widehat{p})\big) \leq r.$$

NP-hard, consider solution methods based on local optimization

Representing the constraint in a kernel form, the problem is

$$\min_{R, RR^\top = I_{m-r}} \left(\min_{\widehat{p}} \|p - \widehat{p}\| \quad \text{subject to} \quad R\mathscr{S}(\widehat{p}) = 0 \right)$$

Note: Double minimization with bilinear equality constraint.

There is a matrix $G(R)$, such that $R\mathscr{S}(\widehat{p}) = 0 \iff G(R)p = 0.$

Variable projection vs. alternating projections

Two ways to approach the double minimization:

- Variable projections (VARPRO):
  solve the inner minimization analytically

$$\min_{R,\, RR^\top = I_{m-r}} \operatorname{vec}^\top\left(R\mathscr{S}(\widehat{p})\right)\left(G(R)G^\top(R)\right)^{-1}\operatorname{vec}\left(R\mathscr{S}(\widehat{p})\right)$$

$\rightsquigarrow$ a nonlinear least squares problem for R only.

- Alternating projections (AP):
  alternate between solving two least squares problems

VARPRO is globally convergent with a super linear conv. rate.

AP is globally convergent with a linear convergence rate.

Software implementation

The structure of $\mathscr{S}$ can be exploited for efficient $O(\dim(p))$ cost function and first derivative evaluations.

SLICOT library includes high quality FORTRAN implementation of algorithms for block Toeplitz matrices.

VARPRO approach based on the Levenberg–Marquardt alg. implemented in MINPACK.

Another extension: weighted low-rank approx.

The basic low-rank approximation

$$\widehat{D}^* := \arg\min_{\widehat{D}} \|D - \widehat{D}\| \quad \text{subject to} \quad \text{rank}(\widehat{D}) \leq \mathtt{m}.$$

is a maximum liklihood estimate assuming $\text{cov}(\text{vec}(\widetilde{D})) = I$.

If $\text{cov}(\text{vec}(\widetilde{D})) = W$, the maximum liklihood estimate is given by

$$\min_{\widehat{D}} \text{vec}^\top(D - \widehat{D}) W \, \text{vec}(D - \widehat{D}) \quad \text{subject to} \quad \text{rank}(\widehat{D}) \leq \mathtt{m}$$

weighted low-rank approximation (maximum likelihood PCA)

NP-hard problem

Another extension: nonnegative low-rank approx.

Constrained LRA arise in Markov chains and image mining

$$\min_{\widehat{D}} \|D - \widehat{D}\| \quad \text{subject to} \quad \operatorname{rank}(\widehat{D}) \leq \mathtt{m} \text{ and } \widehat{D}_{ij} \geq 0 \text{ for all } i,j.$$

Using an image representation, an equivalent problem is

$$\min_{P \in \mathbb{R}^{\mathtt{d}\times\mathtt{m}}, L \in \mathbb{R}^{\mathtt{m}\times N}} \|D - PL\| \quad \text{subject to} \quad P_{ik}, L_{kj} \geq 0 \text{ for all } i,k,j.$$

Alternating projections algorithm:

- Choose an initial approximation $P^{(0)} \in \mathbb{R}^{\mathtt{d}\times\mathtt{m}}$ and set $k := 0$.
- Solve: $L^{(k)} = \arg\min_L \|D - P^{(k)}L\|$ subject to $L \geq 0$.
- Solve: $P^{(k+1)} = \arg\min_P \|D - PL^{(k)}\|$ subject to $P \geq 0$.
- Repeat until convergence.

Data fitting by a second order model

$$\mathscr{B}(A,b,c) := \{ d \in \mathbb{R}^{\mathrm{d}} \mid d^\top A d + b^\top d + c = 0 \}, \quad \text{with } A = A^\top$$

Consider first exact data:

$$d \in \mathscr{B}(A,b,c) \iff d^\top A d + b^\top d + c = 0$$
$$\iff \langle \underbrace{\mathrm{col}(d \otimes_{\mathrm{s}} d, d, 1)}_{d_{\mathrm{ext}}}, \underbrace{\mathrm{col}\big(\mathrm{vec_s}(A), b, c\big)}_{\theta} \rangle = 0$$

$$\{ d_1, \dots, d_N \} \in \mathscr{B}(\theta) \iff \theta \in \mathrm{left\,ker} \underbrace{\begin{bmatrix} d_{\mathrm{ext},1} & \cdots & d_{\mathrm{ext},N} \end{bmatrix}}_{D_{\mathrm{ext}}}, \quad \theta \neq 0$$

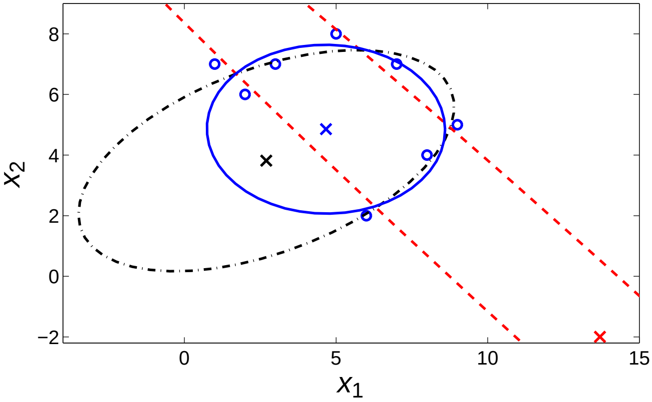
$$\iff \mathrm{rank}(D_{\mathrm{ext}}) \leq \mathrm{d} - 1$$

Therefore, for measured data $\rightsquigarrow$ LRA of D_{ext}.

Notes:

- Special case $\mathscr{B}$ an ellipsoid (for $A > 0$ and $4c < b^\top A^{-1} b$).
- Related to kernel PCA

Example: ellipsoid fitting



dashed — kernel PCA solid — modified method

dashed-dotted — orthogonal regression (geometric fitting)

○ — data points × — centers

Rank minimization

Approximate modeling is a tradeoff between:

- fitting accuracy and
- model complexity

Two possible scalarizations of the bi-objective optimization are:

- LRA: minimize misfit under a constraint on complexity
- RM: minimize complexity under a constraint ($\mathscr{C}$) on misfit

$$\text{minimize}_X \ \text{rank}(X) \quad \text{subject to} \quad X \in \mathscr{C}$$

RM is also NP-hard, however, there are effective heuristics, *e.g.*,

with $X = \text{diag}(x)$, $\text{rank}(X) = \text{card}(x)$,

$$\ell_1 \text{ heuristic:} \quad \text{minimize}_x \ \|x\|_1 \quad \text{subject to} \quad \text{diag}(x) \in \mathscr{C}$$

References

- S. Boyd, EE263: Linear dynamical systems
- Golub & Van Loan, An analysis of the total least-squares problem, *SIAM J. Numer. Anal.*, volume 17, pages 883–893, 1980
- Van Huffel & Vandewalle, The total least-squares problem: Computational aspects and analysis, SIAM, 1991
- Markovsky & Van Huffel, Overview of total least-squares methods, *Signal Processing*, volume 87, pages 2283–2302, 2007
- Markovsky, Structured low-rank approximation and its applications, *Automatica*, 2008