A structured low-rank approximation approach to system identification

Ivan Markovsky

Vrije Universiteit Brussel

Main message: system identification $\subset$ SLRA

$$\begin{array}{ll} \text{minimize} & \text{over } B \quad \text{dist}(A, B) \\ \text{subject to} & \text{rank}(B) \leq r \text{ and } B \text{ structured} \end{array} \qquad \text{(SLRA)}$$

SLRA problem	$\leftrightarrow$	system identification
A	$\leftrightarrow$	observed data
$\text{dist}(\cdot, \cdot)$	$\leftrightarrow$	noise properties
r	$\leftrightarrow$	model complexity
structure	$\leftrightarrow$	model class

Plan of the presentation

System identification $\leftrightarrow$ low-rank approximation

Missing data estimation

Nonlinear system identification

Plan of the presentation

System identification $\leftrightarrow$ low-rank approximation

Missing data estimation

Nonlinear system identification

Identification: finding models from data

$$\begin{array}{ccc} \text{data} & \xrightarrow{\text{identification}} & \text{model} \\ \mathscr{D} & & \mathscr{B} \in \mathscr{M} \end{array}$$

aim: "accurate" and "simple" model

"accurate" $\rightarrow$ smallest approximation error

"simple" $\rightarrow$ Occam's razor principle:
among equally accurate models,
choose the simplest

Data $\mathscr{D}$: set of vector-valued time series

the data $\mathscr{D}$ is a set $\{ w^1, \ldots, w^N \}$

of vector valued $w^k = \begin{bmatrix} w_1^k \\ \vdots \\ w_q^k \end{bmatrix}$

time series $w_i^k = \big(w_i^k(1), \ldots, w_i^k(T_k)\big)$

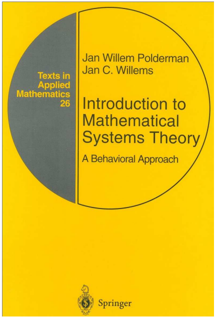
N	—	# of repeated experiments
q	—	# of variables
T_k	—	# of time samples in kth exp.

Model $\mathscr{B}$: subset of the data space

behavioral definition of a model

$$\mathscr{B} = \{ w \mid g(w) = 0 \text{ holds} \}$$

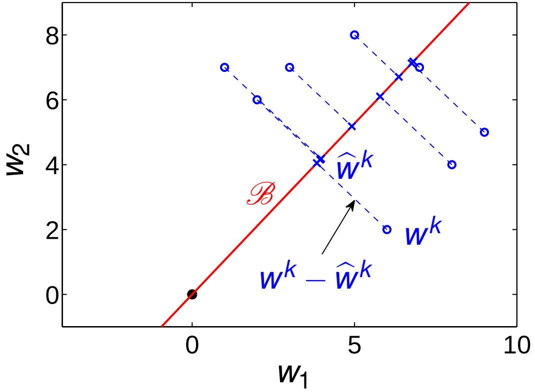
$g(w) = 0$ — representation of $\mathscr{B}$

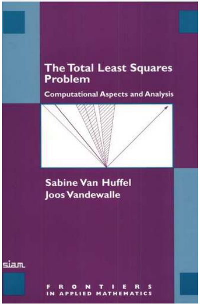


model class $\mathscr{M}$: set of models

$\mathscr{L}$ — set of linear models

$$\mathrm{dist}(\mathscr{D},\mathscr{B}) := \min_{\widehat{\mathscr{D}}\subset\mathscr{B}} \sqrt{\textstyle\sum_k \|w^k - \widehat{w}^k\|_2^2}$$





errors-in-variables model: data = true value + noise

other error measures: output error, ARMAX, ...

Model complexity = (# inputs, # states)

simple = small $\quad \big(\mathscr{B}_1 \subset \mathscr{B}_2 \implies \mathscr{B}_1$ is simpler than $\mathscr{B}_2\big)$

linear model is subspace, then

$$\text{size of } \mathscr{B} \quad \leftrightarrow \quad \text{dimension of } \mathscr{B}$$

linear time-invariant (LTI) dynamic model

$$\text{dimension of } \mathscr{B} \quad \leftrightarrow \quad (\underbrace{\text{\# inputs}}_{\mathtt{m}}, \underbrace{\text{\# states}}_{\ell})$$

$\mathscr{L}_{\mathtt{m},\ell}$ — LTI systems of bounded complexity

Identification: error-complexity trade-off

$$\underset{\mathscr{D}}{\text{data}} \xrightarrow{\text{identification}} \underset{\mathscr{B} \in \mathscr{M}}{\text{model}}$$

$$\text{minimize} \quad \text{over } \mathscr{B} \in \mathscr{M} \quad \begin{bmatrix} \text{dist}(\mathscr{D}, \mathscr{B}) \\ \text{complexity}(\mathscr{B}) \end{bmatrix}$$

Scalarization of the bi-objective problem

1. minimize $\quad \text{dist}(\mathscr{D}, \mathscr{B}) + \lambda \, \text{complexity}(\mathscr{B})$

2. minimize $\quad \lambda \, \text{complexity}(\mathscr{B})$
   subject to $\quad \text{dist}(\mathscr{D}, \mathscr{B}) \leq \mu$

3. minimize $\quad \text{dist}(\mathscr{D}, \mathscr{B})$
   subject to $\quad \text{complexity}(\mathscr{B}) \leq (\mathtt{m}, \ell)$

describe the same set of Pareto optimal solutions

with $\mathtt{m}$ given, finding ℓ is an order selection problem

LTI identification problem

$$\begin{array}{ll} \text{minimize} & \text{over } \mathscr{B} \quad \text{dist}(\mathscr{D}, \mathscr{B}) \\ \text{subject to} & \mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell} \end{array}$$

with distance measure

$$\text{dist}(\mathscr{D}, \mathscr{B}) = \min_{\widehat{\mathscr{D}} \subset \mathscr{B}} \sqrt{\textstyle\sum_k \| w^k - \widehat{w}^k \|_2^2} = \min_{\widehat{\mathscr{D}} \subset \mathscr{B}} \| \mathscr{D} - \widehat{\mathscr{D}} \|$$

the problem is

$$\begin{array}{ll} \text{minimize} & \text{over } \mathscr{B}, \widehat{\mathscr{D}} \quad \| \mathscr{D} - \widehat{\mathscr{D}} \| \\ \text{subject to} & \widehat{\mathscr{D}} \subset \mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell} \end{array}$$

w exact $\iff$ rank deficient Hankel matrix

$$\text{exact trajectory } w \in \mathscr{B} \in \mathscr{L}_{\mathrm{m},\ell}$$

$$\Updownarrow$$

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_\ell w(t+\ell) = 0$$

$$\Updownarrow$$

$$\underbrace{\begin{bmatrix} R_0 & R_1 & \cdots & R_\ell \end{bmatrix}}_{R} \underbrace{\begin{bmatrix} w(1) & w(2) & \cdots & w(T-\ell) \\ w(2) & w(3) & \cdots & w(T-\ell+1) \\ \vdots & \vdots & & \vdots \\ w(\ell+1) & w(\ell+2) & \cdots & w(T) \end{bmatrix}}_{\mathscr{H}_{\ell+1}(w)} = 0$$

$$\operatorname{rank}\big(\mathscr{H}_{\ell+1}(w)\big) = q(\ell+1) - \operatorname{rank}(R) = q\ell + \mathrm{m}$$

$\mathscr{D}$ exact $\iff$ rank $\big(\mathscr{H}_{\ell+1}(\mathscr{D})\big) \leq q\ell + \mathtt{m}$

$$\text{exact data } \mathscr{D} \subset \mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell}$$

$$\Updownarrow$$

$$w^k \in \mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell} \quad \text{for all } k = 1, \ldots, N$$

$$\Updownarrow$$

$$\text{rank}\Big(\underbrace{\big[\mathscr{H}_{\ell+1}(w^1) \ \cdots \ \mathscr{H}_{\ell+1}(w^N)\big]}_{\text{mosaic-Hankel matrix } \mathscr{H}_{\ell+1}(\mathscr{D})}\Big) \leq q\ell + \mathtt{m}$$

LTI identification is mosaic-Hankel SLRA

$$\begin{array}{ll} \text{minimize} & \text{over } \mathscr{B} \text{ and } \widehat{\mathscr{D}} \quad \|\mathscr{D} - \widehat{\mathscr{D}}\| \\ \text{subject to} & \widehat{\mathscr{D}} \subset \mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell} \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{\mathscr{D}} \quad \|\mathscr{D} - \widehat{\mathscr{D}}\| \\ \text{subject to} & \text{rank}\left(\mathscr{H}_{\ell+1}(\widehat{\mathscr{D}})\right) \leq q\ell + \mathtt{m} \end{array}$$

Summary: identification $\leftrightarrow$ SLRA

LTI model class	$\Longleftrightarrow$	Hankel structure
bounded complexity	$\Longleftrightarrow$	rank constraint
$(\mathtt{m}, \ell)$	$\leftrightarrow$	$q\ell + \mathtt{m}$

Plan of the presentation

System identification $\leftrightarrow$ low-rank approximation

Missing data estimation

Nonlinear system identification

Motivation goes beyond data corruption

sensor failures

- measurements are accidentally corrupted

compressive sensing

- measurements are intentionally skipped

data-driven estimation and control

- missing data is what we aim to find
- examples: state estimation, control, and realization

1. state estimation
   - given: system $\mathscr{B}$, input u, and output y
   - missing: initial conditions w_{ini} such that

$$\text{minimize} \quad \text{over } w_{\text{ini}},\, \widehat{y} \quad \underbrace{\|y-\widehat{y}\|}_{\text{estimation error}} \quad \text{s.t.} \quad w_{\text{ini}} \wedge (u,\widehat{y}) \in \mathscr{B}$$

2. output tracking control
   - given: $\mathscr{B}$, w_{ini}, and reference output y_{ref}
   - missing: control input u such that

$$\text{minimize} \quad \text{over } \widehat{u},\, \widehat{y} \quad \underbrace{\|y_{\text{ref}}-\widehat{y}\|}_{\text{tracking error}} \quad \text{s.t.} \quad w_{\text{ini}} \wedge (\widehat{u},\widehat{y}) \in \mathscr{B}$$

3. realization
   - given: impulse response $h(1),\dots,h(T)$
   - missing: extension $h(T+1),h(T+2),\dots$

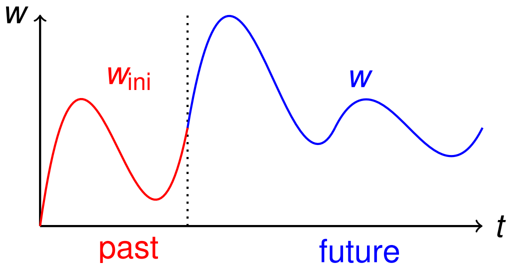
Exact, noisy, and missing data

exact data— kept fixed

inexact / "noisy" data — approximated ($\min \|\text{error}\|_2$)

missing data — interpolated from $w \in \mathscr{B}$

the initial conditions w_{ini} are the "past" of w



1. state estimation

	past	future
input	?	u
output	?	y

2. output tracking control

	past	future
input	u_{ini}	?
output	y_{ini}	y_{ref}

3. (noisy) realization

	past	future
input	0	δ
output	0	$(h, ?)$

black — exact   blue — inexact/noisy   red — missing

SLRA with element-wise weighted 2-norm

$$\begin{array}{lll} \text{minimize} & \text{over } \widehat{\mathscr{D}} & \text{dist}(\mathscr{D}, \widehat{\mathscr{D}}) \\ \text{subject to} & \text{rank}\left(\mathscr{H}_{\ell+1}(\widehat{w})\right) \leq q\ell + \mathtt{m} & \end{array}$$

weighted 2-norm approximation

$$\text{dist}(\mathscr{D}, \widehat{\mathscr{D}}) := \sqrt{\sum_{k,i,t} v_i^k(t) \big(w_i^k(t) - \widehat{w}_i^k(t)\big)^2}$$

with element-wise weights

$v_i^k(t) \in (0, \infty)$	if $w_i^k(t)$ is noisy	approximate $w_i^k(t)$
$v_i^k(t) = 0$	if $w_i^k(t)$ is missing	interpolate $w_i^k(t)$
$v_i^k(t) = \infty$	if $w_i^k(t)$ is exact	$\widehat{w}_i^k(t) = w_i^k(t)$

Summary: SLRA solves control problems

the given data is exact or noisy

what we want to compute is missing data

exact/noisy/missing data is handled by ∞/finite/0 weights

Plan of the presentation

System identification $\leftrightarrow$ low-rank approximation

Missing data estimation

Nonlinear system identification

Conic section fitting

the points $(x_1, y_1), \ldots, (x_N, y_N)$ lie on a conic section

$$\Updownarrow$$

there are $A = A^\top$, b, c, at least one of them nonzero, s.t.

$$\begin{bmatrix} x_i & y_i \end{bmatrix} A \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} x_i & y_i \end{bmatrix} b + c = 0, \quad \text{for } i = 1, \ldots, N$$

$$\Updownarrow$$

there is $\theta = \begin{bmatrix} a_{11} & a_{12} & a_{22} & b_1 & b_2 & c \end{bmatrix} \neq 0$, such that

$$\theta \begin{bmatrix} x_1^2 & \cdots & x_N^2 \\ x_1 y_1 & \cdots & x_N y_N \\ x_1 & \cdots & x_N \\ y_1^2 & \cdots & y_N^2 \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix} = 0$$

Conic section fitting $\iff$ rank deficiency

the points $(x_1, y_1), \dots, (x_N, y_N)$ lie on a conic section

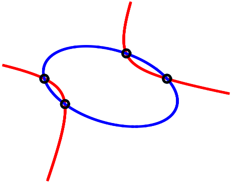
$$\mathscr{B}(\theta) = \{\, w \mid w^\top A w + w^\top b + c = 0 \,\}$$

$$\Updownarrow$$

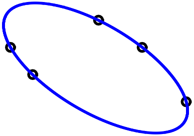
$$\operatorname{rank} \left(\begin{bmatrix} x_1^2 & \cdots & x_N^2 \\ x_1 y_1 & \cdots & x_N y_N \\ x_1 & \cdots & x_N \\ y_1^2 & \cdots & y_N^2 \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix} \right) \leq 5$$

Examples

rank $< 5 \quad \Longrightarrow \quad$ nonunique fit



rank $= 5 \quad \Longrightarrow \quad$ unique fit



rank $= 6 \quad \Longrightarrow \quad$ no exact fit by a conic section

Nonlinear system identification

discrete-time nonlinear system

$$\mathscr{B} := \{\, w \mid R\big(w(t), w(t-1), \dots, w(t-\ell)\big) = 0 \,\}$$

special case: input/output NARX system

$$\mathscr{B} = \{\, w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid y(t) = f\big(u(t), w(t-1), \dots, w(t-\ell)\big) \,\}$$

linear parameterization: $\mathscr{B}_\theta$

$$R(x) = \sum \theta_i \phi_i(x) = \theta\phi(x), \qquad \begin{array}{ccl} \phi & — & \text{model structure} \\ \theta & — & \text{parameter vector} \end{array}$$

$$x(t) := \big(w(t), w(t-1), \dots, w(t-\ell)\big)$$

Link to SLRA

parameter estimation problem

$$\begin{array}{lll} \text{minimize} & \text{over } \theta \text{ and } \widehat{w} & \|w - \widehat{w}\| \\ \text{subject to} & \widehat{w} \in \mathscr{B}_\theta & \end{array} \qquad \text{(NL SYSID)}$$

$$\widehat{w} \in \mathscr{B}_\theta \iff \operatorname{rank}\Big(\underbrace{\begin{bmatrix} \phi\big(\widehat{x}(1)\big) & \cdots & \phi\big(\widehat{x}(T-\ell)\big) \end{bmatrix}}_{\text{polynomially structured matrix } \Phi(\widehat{w})} \Big) \leq r$$

(NL SYSID) $\iff$ polynomially structured LRA

(NL SYSID) is nonconvex and yields biased estimator

Bias correction

ignoring the structure of $\Phi(\widehat{w})$ leads to kernel PCA

easy to compute, but biased in the EIV model

$$w = \bar{w} + \widetilde{w}, \quad \text{where} \quad \bar{w} \in \bar{\mathscr{B}} \quad \text{and} \quad \widetilde{w} \sim \mathrm{N}(0, \sigma^2 I)$$

define $\quad \Psi := \Phi(w)\Phi^\top(w) \quad$ and $\quad \bar{\Psi} := \Phi(\bar{w})\Phi^\top(\bar{w})$

goal: construct "corrected" matrix Ψ_c, such that

$$\mathbf{E}(\Psi_c) = \bar{\Psi}$$

Derivation of the correction

Hermite polynomials $h_k(x)$ have the property

$$\mathbf{E}\left(h_k(\bar{x}+\widetilde{x})\right) = \bar{x}^k, \quad \text{where} \quad \widetilde{x} \sim \mathrm{N}(0,\sigma^2) \qquad (*)$$

with $w = (x, y)$, the (i,j)th element of $\Psi = \Phi\Phi^\top$ is

$$\sum (\bar{x}+\widetilde{x})^{n_{x,i}+n_{x,j}} (\bar{y}+\widetilde{y})^{n_{y,i}+n_{y,j}}$$

then, by $(*)$

$$\phi_{c,ij} := \sum h_{n_{x,i}+n_{x,j}}(x) h_{n_{y,i}+n_{y,j}}(y)$$

has the desired property

$$\mathbf{E}(\psi_{c,ij}) = \sum \bar{x}^{n_{x,i}+n_{x,j}} \bar{y}^{n_{y,i}+n_{y,j}} =: \bar{\psi}_{ij}$$

Unbiased estimator

the corrected Ψ_c is an even polynomial in σ

$$\Psi_c(\sigma^2) = \Psi_{c,0} + \sigma^2 \Psi_{c,1} + \cdots + \sigma^{2n_\psi} \Psi_{c,n_\psi}$$

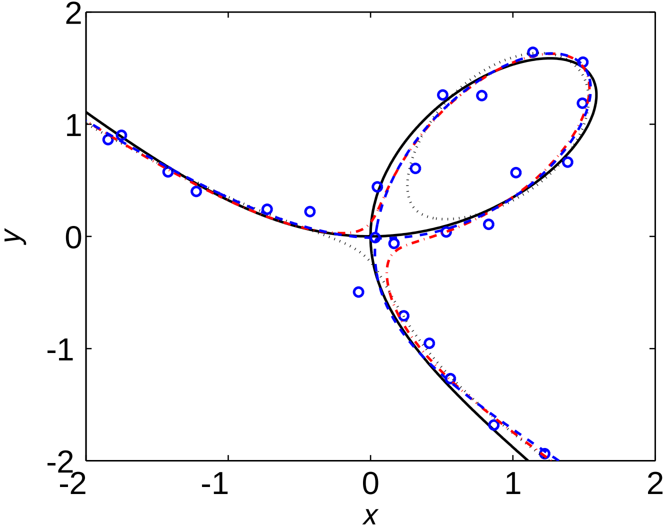
estimate: $\Psi_c(\sigma^2)\theta = 0$

computing simultaneously σ and θ is polynomial EVP

examples of static nonlinear model fitting

KPCA	dotted
PLRA	dashed-dotted
bias corrected	dashed

Example: $x^3 + y^3 - 3xy = 0$



Conclusion: system identification $\subset$ SLRA

LTI model class $\leftrightarrow$ mosaic-Hankel structure

solving control problems as missing data estimation

bias correction procedure for polynomial SLRA

Conclusion: system identification $\subset$ SLRA

LTI model class $\leftrightarrow$ mosaic-Hankel structure

solving control problems as missing data estimation

bias correction procedure for polynomial SLRA

papers, course materials, and code at:

`http://slra.github.io/`