# Application of low-rank approximation in nonlinear system identification

Ivan Markovsky

# General setup: linearly parameterized discrete-time nonlinear systems

kernel: $R\big(\underbrace{w(t), w(t-1), \ldots, w(t-\ell)}_{x(t)}\big) = 0$

special case: input/output NARX system

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

linearly parameterized model $\mathscr{B}_\theta$

$$R(x) = \sum \theta_i \phi_i(x) = \theta\phi(x),$$

$\phi$ — model structure

$\theta$ — parameter vector

# Example: single-input single-output polynomially time-invariant model

$\phi$ is a vector of monomials $\phi_i := \Pi_j x_j^{\mathrm{n}_{ij}}$

the structure $\phi$ is defined by the degrees matrix

$$\phi \;\leftrightarrow\; \mathrm{N} := \left[\mathrm{n}_{ij}\right] \in \mathbb{N}^{\mathrm{n}_\phi \times \mathrm{n}_x}$$

polynomially time-invariant (PTI) model class

$$\mathscr{P}_{\ell,\mathrm{n}} := \{\, \mathscr{B}_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \mathbb{R}^{\mathrm{n}_\theta} \,\},$$

$\ell$ — lag

$\mathrm{n} := \max_{i,j}\ \mathrm{n}_{ij}$

# Our goal is to find PTI model from data:

$$\big(w(1),\ldots,w(T)\big) \mapsto \mathscr{B} \in \mathscr{P}_{\ell,\mathrm{n}}$$

1. structure selection: find $\phi$
2. parameter estimation: find $\theta$

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

# Link to low-rank approximation

$$w \in \mathscr{B}_\theta$$

$$\Updownarrow$$

$$R\big(x(t)\big) = \theta^\top \phi\big(x(t)\big) = 0, \quad \text{for } t = 1, \ldots, T - \ell$$

$$\Updownarrow$$

$$\theta^\top \begin{bmatrix} \phi\big(x(1)\big) & \cdots & \phi\big(x(T-\ell)\big) \end{bmatrix} = 0$$

$$\Updownarrow$$

$$\operatorname{rank}\Big(\Phi(w)\Big) \leq n_\phi - 1$$

# (NL SYSID) $\iff$ low-rank approximation

$$\begin{array}{ll} \text{minimize} & \text{over } \theta \text{ and } \widehat{w} \quad \|w - \widehat{w}\|_2 \\ \text{subject to} & \operatorname{rank}\big(\Phi(\widehat{w})\big) \leq n_\phi - 1 \end{array} \qquad \text{(SLRA)}$$

non-convex optimization problem

there are no efficient solution methods

heuristic method: ignore the structure of $\Phi(\widehat{w})$

$$\text{minimize} \quad \text{over } \theta \neq 0 \quad \|\theta^\top \Phi(w)\|_2 \qquad \text{(LRA)}$$

# Structure selection via sparsity regularization

select "large" model class $\mathscr{P}_{\ell,\mathrm{n}}$ and impose sparsity on $\theta$

$$\text{minimize} \quad \text{over } \theta \quad \|\theta^\top \Phi(w)\|_2 + \gamma \|\theta\|_1$$

$\gamma$ controls the sparsity level

- $\gamma = 0 \quad \rightsquigarrow \quad$ (LRA) $\quad \rightsquigarrow \quad$ full $\theta$
- $\gamma \to \infty \quad \rightsquigarrow \quad \theta \to 0$

selected, so that # nonzero elements $=$ given number

# Perspectives / future work

bias correction procedure

*Consistent least squares fitting of ellipsoids.*
*Numerische Mathematik, 98(1):177-194, 2004*

*Adjusted least squares fitting of algebraic hypersurfaces.*
*Linear Algebra Appl., 502:243–274, 2016.*

conditions under which the $\ell_1$-regularizer "works"

benchmarking and comparison with alternative methods

# 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

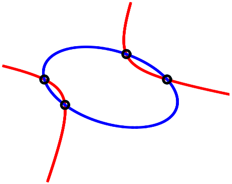
$$\mathscr{B}(\theta) = \{\, w \mid w^\top A w + w^\top b + c = 0 \,\}$$

$$\Updownarrow$$

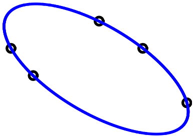
$$\operatorname{rank}\left( \begin{bmatrix} x_1^2 & \cdots & x_N^2 \\ x_1 y_1 & \cdots & x_N y_N \\ x_1 & \cdots & x_N \\ y_1^2 & \cdots & y_N^2 \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix} \right) \leq 5$$

# Examples

rank $< 5$ $\Longrightarrow$ nonunique fit



rank $= 5$ $\Longrightarrow$ unique fit



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

# Unstructured LRA is biased

easy to compute, but biased in the EIV setup

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

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

goal: construct "corrected" matrix $\Psi_\mathrm{c}$, such that

$$\mathbf{E}(\Psi_\mathrm{c}) = \bar{\Psi}$$

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

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

then, by $(*)$

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

has the desired property

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

# Unbiased estimator

the corrected $\Psi_c$ is an even polynomial in $\sigma$

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

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

computing simultaneously $\sigma$ and $\theta$ is polynomial EVP