ELEC system identification workshop

Subspace methods

Ivan Markovsky

Plan

1. Behavioral approach
2. Subspace methods
3. Optimization methods

Outline

Plan

Exact modeling

Algorithms

The goal is to obtain a model $\mathscr{B}$ from data $\mathscr{D}$

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

$\mathscr{U}$ — data space $(\mathbb{R}^q)^{\mathbb{N}}$: functions from $\mathbb{N}$ to $\mathbb{R}^q$

$\mathscr{D}$ — data: set of finite vector-valued time series

$$\mathscr{D} = \{ w_\text{d}^1, \dots, w_\text{d}^N \}, \quad w_\text{d}^i = \big(w_\text{d}^i(1), \dots, w_\text{d}^i(T_i)\big)$$

$\mathscr{B}$ — model: subset of the data space $\mathscr{U}$

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

Work plan

1. define a modeling problem (What is $\mathscr{D} \mapsto \mathscr{B}$?)
2. find an algorithm that solves the problem
3. implement the algorithm (How to compute $\mathscr{B}$?)

State the aim without hidden assumptions

all user choices should enter in the problem formulation

hyper-parameters should not appear in the solutions

the resulting methods should be automatic

User choices reflect prior knowledge; they determine the model class and fitting criterion

the "true model" assumption

$$\mathscr{D} = \underbrace{\bar{\mathscr{D}}}_{\text{true data}} + \underbrace{\widetilde{\mathscr{D}}}_{\text{noise}} \qquad \text{where} \quad \bar{\mathscr{D}} \subset \underbrace{\bar{\mathscr{B}}}_{\text{true model}} \in \mathscr{M}$$

assuming, in addition, that $\widetilde{\mathscr{D}}$ is a stochastic process

noise distribution $\leftarrow$ maximum likelihood principle $\rightarrow$ fitting criterion

we can specify $\mathscr{M}$ and $\|\cdot\|$ as deterministic approximation

Examples of user choices for $\mathscr{M}$ and $\|\cdot\|$

Model class

linear	nonlinear
static	dynamic
time-invariant	time-varying

Fitting criterion

exact	approximate
deterministic	stochastic

Why exact identification?

from simple to complex:

exact $\mapsto$ approx. $\mapsto$ stoch. $\mapsto$ approx. stoch.

exact identification is ingredient of the other problems

exact methods lead to effective approximation heuristics

Exact identification in $\mathscr{L}$

given data $\mathscr{D}$

find $\widehat{\mathscr{B}} \in \mathscr{L}$, such that $\mathscr{D} \subset \widehat{\mathscr{B}}$

nonunique solution always exists

Exact identification in $\mathscr{L}_{\mathtt{m},\ell}$

given $(\mathtt{m}, \ell)$ and data $\mathscr{D}$

find $\widehat{\mathscr{B}} \in \mathscr{L}_{\mathtt{m},\ell}$, such that $\mathscr{D} \subset \widehat{\mathscr{B}}$

solution may not exist

Most powerful unfalsified model $\mathscr{B}_{\text{mpum}}(\mathscr{D})$

given data $\mathscr{D}$

find the smallest $(\mathtt{m}, \ell)$, such that $\exists\ \widehat{\mathscr{B}} \in \mathscr{L}^{q}_{\mathtt{m},\ell}$, $\mathscr{D} \subset \widehat{\mathscr{B}}$

Why complexity minimization?

makes the solution unique

Occam's razor: "simpler $=$ better"

Identifiability question

Recover the data generating system $\overline{\mathscr{B}}$ from exact data $\mathscr{D}$

$$\mathscr{D} \subset \overline{\mathscr{B}} \in \mathscr{L}^{\mathtt{q}}$$

Under what conditions $\mathscr{B}_{\text{mpum}}(\mathscr{D}) = \overline{\mathscr{B}}$?

the answer is given by the "fundamental lemma"

Hankel matrix

consider the case $\mathscr{D} = w_\mathrm{d}$ (single trajectory)

main tool

$$\mathscr{H}_L(w) := \begin{bmatrix} w(1) & w(2) & w(3) & \cdots & w(T-L+1) \\ w(2) & w(3) & w(4) & \cdots & w(T-L+2) \\ w(3) & w(4) & w(5) & \cdots & w(T-L+3) \\ \vdots & \vdots & \vdots & & \vdots \\ w(L) & w(L+1) & w(L+2) & \cdots & w(T) \end{bmatrix}$$

if $w_\mathrm{d} \in \mathscr{B} \in \mathscr{L}^q$, then $\operatorname{image}\big(\mathscr{H}_L(w_\mathrm{d})\big) \subset \mathscr{B}|_L$

extra conditions on w_d and $\mathscr{B}$ are needed for

$$\operatorname{image}\big(\mathscr{H}_L(w_\mathrm{d})\big) = \mathscr{B}|_L$$

Persistency of excitation (PE)

u is PE of order L if $\mathscr{H}_L(u)$ is full row rank

system theoretic interpretation:

$$u \in (\mathbb{R}^{\mathrm{m}})^T \text{ is PE of order } L \iff \text{there is no } \mathscr{B} \in \mathscr{L}_{\mathrm{m}-1,L}, \text{ such that } u \in \mathscr{B}$$

Lemma

1. $\mathscr{B} \in \mathscr{L}^q_{\mathrm{m},\ell}$ controllable and
2. $w_\mathrm{d} = (u_\mathrm{d}, y_\mathrm{d}) \in \mathscr{B}$ with u_d PE of order $L + \mathrm{p}\ell$

$$\implies \quad \text{image}\,\big(\mathscr{H}_L(w_\mathrm{d})\big) = \mathscr{B}|_L$$

Plan

Exact modeling

Algorithms

The main idea is that a desired trajectory w can be constructed directly from the data w_{d}

any $w \in \mathscr{B}|_L$ can be obtained from $w_{\text{d}} \in \mathscr{B}$

$$w = \mathscr{H}_L(w_{\text{d}})g, \qquad \text{for some } g$$

$g \sim$ input and initial conditions, *cf.*, image representation

Algorithms

$w_{\mathrm{d}} \mapsto$ kernel parameter R

$w_{\mathrm{d}} \mapsto$ impulse response H

$w_{\mathrm{d}} \mapsto$ state/space parameters (A, B, C, D)

- $w_{\mathrm{d}} \mapsto R \mapsto (A, B, C, D)$ or $w_{\mathrm{d}} \mapsto H \mapsto (A, B, C, D)$
- $w_{\mathrm{d}} \mapsto$ observability matrix $\mapsto (A, B, C, D)$
- $w_{\mathrm{d}} \mapsto$ state sequence $\mapsto (A, B, C, D)$

$w_{\mathrm{d}} \mapsto R$

under the assumptions of the lemma

$$\operatorname{image}\big(\mathscr{H}_{\ell+1}(w_{\mathrm{d}})\big) = \mathscr{B}|_{\ell+1}$$

$\operatorname{left\,ker}\big(\mathscr{H}_{\ell+1}(w_{\mathrm{d}})\big)$ defines a kernel repr. of $\mathscr{B}$

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_\ell \end{bmatrix} \mathscr{H}_{\ell+1}(w_{\mathrm{d}}) = 0, \quad R_i \in \mathbb{R}^{g \times q}$$

kernel representation

$$\mathscr{B} = \ker\big(R(\sigma)\big), \qquad \text{with} \quad R(z) = \sum_{i=0}^{\ell} R_i z^i$$

recursive computation (exploiting Hankel structure)

$w_d \mapsto H$

impulse response (matrix values trajectory)

$$W = \Big(\underbrace{0,\dots,0}_{\ell}, \begin{bmatrix} I \\ H(0) \end{bmatrix}, \begin{bmatrix} 0 \\ H(1) \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ H(t) \end{bmatrix}\Big)$$

by the lemma, $W = \mathscr{H}_{\ell+t}(w_d)G$

define $\mathscr{H}_{\ell+t}(u_d) =: \begin{bmatrix} U_p \\ U_f \end{bmatrix}$ and $\mathscr{H}_{\ell+t}(y_d) =: \begin{bmatrix} Y_p \\ Y_f \end{bmatrix}$

we have

$$\begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix} G = \begin{bmatrix} 0 \\ 0 \\ \begin{bmatrix} I_m \\ 0 \end{bmatrix} \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} 0 \\ 0 \end{matrix}}\right\} \text{ zero ini. conditions} \\ \leftarrow \text{ impulse input} \end{matrix} \tag{1}$$

$$Y_f \, G = H \tag{2}$$

Block algorithm

input: u_{d}, y_{d}, ℓ, and t

solve (2) and let G_{p} be a solution

compute $H = Y_{\mathrm{f}} G_{\mathrm{p}}$

output: the first t samples of the impulse response H

Exercise: implement and test the algorithm

Refinements

solve (2) efficiently exploiting the Hankel structure

do the computations iteratively for pieces of H

automatically choose t, for a sufficient decay of H

Exercise: try the improvements

application for noisy data

$w_\mathrm{d} \mapsto (A, B, C, D)$

$$w_\mathrm{d} \mapsto H(0:2\ell) \text{ or } R(\xi) \xrightarrow{\text{realization}} (A, B, C, D)$$

$$w_\mathrm{d} \mapsto \text{obs. matrix } \mathscr{O}_{\ell+1}(A, C) \xrightarrow{(3)} (A, B, C, D)$$

$$\mathscr{O}_{\ell+1}(\mathrm{A,C}) \mapsto (A, C), \quad (u_\mathrm{d}, y_\mathrm{d}, A, C) \mapsto (B, C, x_\mathrm{ini}) \tag{3}$$

$$w_\mathrm{d} \mapsto \text{state sequence } x_\mathrm{d} \xrightarrow{(4)} (A, B, C, D)$$

$$\begin{bmatrix} \sigma x_\mathrm{d} \\ y_\mathrm{d} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_\mathrm{d} \\ u_\mathrm{d} \end{bmatrix} \tag{4}$$

$\mathscr{O}_{\ell+1}(A, C) \mapsto (A, B, C, D)$

C is the first block entry of $\mathscr{O}_{\ell+1}(A, C)$

A is given by the shift equation

$$\big(\sigma^* \mathscr{O}_{\ell+1}(A, C)\big) A = \big(\sigma \mathscr{O}_{\ell+1}(A, C)\big)$$

(σ / σ^* removes first / last block entry)

Once C and A are known, the system of equations

$$y_{\mathrm{d}}(t) = CA^t x_{\mathrm{d}}(1) + \sum_{\tau=1}^{t-1} CA^{t-1-\tau} B u_{\mathrm{d}}(\tau) + D\delta(t+1)$$

is linear in D, B, $x_{\mathrm{d}}(1)$

$w_d \mapsto$ observability matrix

columns of $\mathscr{O}_t(A, C)$ are $\mathtt{n}$ indep. free resp. of $\mathscr{B}$

under the conditions of the lemma,

$$\begin{bmatrix} \mathscr{H}_t(u_d) \\ \mathscr{H}_t(y_d) \end{bmatrix} G = \begin{bmatrix} 0 \\ Y_0 \end{bmatrix} \quad \begin{matrix} \leftarrow & \text{zero inputs} \\ \leftarrow & \text{free responses} \end{matrix}$$

lin. indep. free responses $\implies$ G maximal rank

rank revealing factorization

$$Y_0 = \mathscr{O}_t(A, C) \underbrace{\begin{bmatrix} x_{\text{ini},1} & \cdots & x_{\text{ini},j} \end{bmatrix}}_{X_{\text{ini}}}$$

$w_d \mapsto$ state sequence

sequential free responses $\implies$ Y_0 block-Hankel

then X_{ini} is a state sequence of $\mathscr{B}$

computation of sequential free responses

$$\begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix} G = \begin{bmatrix} U_p \\ Y_p \\ 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} U_p \\ Y_p \end{matrix}}\right\} \text{sequential ini. conditions} \\ \leftarrow \quad \text{zero inputs} \end{matrix} \tag{5}$$

$$Y_f \quad G = Y_0$$

rank revealing factorization

$$Y_0 = \mathscr{O}_t(A, C) \begin{bmatrix} x_d(1) & \cdots & x_d(\mathtt{n}+\mathtt{m}+1) \end{bmatrix}$$

Refinements

solve (5) efficiently exploiting the Hankel structure

iteratively compute pieces of $Y_0 \rightsquigarrow$ iterative algorithm

requires smaller persistency of excitation of u_d

could be more efficient
(solve a few smaller systems of eqns than one big)

MOESP-type algorithms

project the rows of $\mathscr{H}_{\mathrm{n}}(y_{\mathrm{d}})$ on rowspan$^{\perp}\left(\mathscr{H}_{\mathrm{n}}(u)\right)$

$$Y_0 := \mathscr{H}_{\mathrm{n}}(y_{\mathrm{d}})\Pi_u^{\perp}$$

where

$$\Pi_u^{\perp} := \left(I - \mathscr{H}_{\mathrm{n}}^{\top}(u)\left(\mathscr{H}_{\mathrm{n}}(u)\mathscr{H}_{\mathrm{n}}^{\top}(u)\right)^{-1}\mathscr{H}_{\mathrm{n}}(u)\right)$$

Observe that $\Pi_u^{\perp}$ is maximal rank and

$$\begin{bmatrix} \mathscr{H}_{\mathrm{n}}(u) \\ \mathscr{H}_{\mathrm{n}}(y_{\mathrm{d}}) \end{bmatrix} \Pi_u^{\perp} = \begin{bmatrix} 0 \\ Y_0 \end{bmatrix}$$

$\implies$ the orthogonal projection computes free responses

Comments

$\mathscr{H}_{\mathtt{n}}(y_{\mathrm{d}})\Pi_u^{\perp}$ are $T - \mathtt{n} + 1$ free responses
($\mathtt{n}$ such responses suffice for exact identification)

a geometric operation has system theoretic meaning

condition for $\mathrm{rank}(Y_0) = \mathtt{n}$ given in the literature

$$\mathrm{rank}\left(\begin{bmatrix} X_{\mathrm{ini}} \\ \mathscr{H}_{\mathtt{n}}(u) \end{bmatrix}\right) = \mathtt{n} + \mathtt{nm}$$

is not verifiable from the data $(u_{\mathrm{d}}, y_{\mathrm{d}})$

N4SID-type algorithms

splitting of the data into "past" and "future"

$$\mathscr{H}_{2\mathrm{n}}(u_\mathrm{d}) =: \begin{bmatrix} U_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}, \qquad \mathscr{H}_{2\mathrm{n}}(y_\mathrm{d}) =: \begin{bmatrix} Y_\mathrm{p} \\ Y_\mathrm{f} \end{bmatrix}$$

and define $W_\mathrm{p} := \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \end{bmatrix}$

oblique projection

$$Y_0 := Y_\mathrm{f} /_{U_\mathrm{f}} W_\mathrm{p} := Y_\mathrm{f} \underbrace{\begin{bmatrix} W_\mathrm{p}^\top & U_\mathrm{f}^\top \end{bmatrix} \begin{bmatrix} W_\mathrm{p} W_\mathrm{p}^\top & W_\mathrm{p} U_\mathrm{f}^\top \\ U_\mathrm{f} W_\mathrm{p}^\top & U_\mathrm{f} U_\mathrm{f}^\top \end{bmatrix}^+ \begin{bmatrix} W_\mathrm{p} \\ 0 \end{bmatrix}}_{\Pi_\mathrm{obl}}$$

of the rows of Y_f along row span(U_f) onto row span(W_p)

N4SID-type algorithms

Observe that

$$\begin{bmatrix} W_{\mathrm{p}} \\ U_{\mathrm{f}} \\ Y_{\mathrm{f}} \end{bmatrix} \Pi_{\mathrm{obl}} = \begin{bmatrix} W_{\mathrm{p}} \\ 0 \\ Y_0 \end{bmatrix}$$

(Π_{obl} gives the least-norm, least-squares solution)

$\Longrightarrow$ oblique proj. computes sequential free responses

Comments

$Y_0 := Y_f /_{U_f} W_p$ are $T - 2\mathtt{n} + 1$ sequential free responses ($\mathtt{n} + \mathtt{m} + 1$ such responses suffice for exact identification)

geometric operation has system theoretic meaning

conditions for $\operatorname{rank}(Y_0) = \mathtt{n}$ given in the literature

1. u_d persistently exciting of order $2\mathtt{n}$ and
2. $\operatorname{row span}(X_{ini}) \cap \operatorname{row span}(U_f) = \{0\}$

are not verifiable from the data (u_d, y_d)

Summary

transitions among representations $\approx$ system theory

exact identification aims at $\mathscr{B}_{\text{mpum}}(w_\text{d})$

$\mathscr{H}_t(w_\text{d})$ plays key role in both analysis and computation

under controllability and u_d persistently exciting

$$\text{image}\,\big(\mathscr{H}_t(w_\text{d})\big) = \mathscr{B}|_t$$

subspace methods construct special responses from data