Outline

Plan

Exact modeling

Algorithms

Exercises

Identification problems

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

- $\mathscr{U}$ — 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

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

- $\mathscr{B}$ — 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 $\mathscr{B}$?)
4. use the software in applications

Notes

- all user choices are set in the problem formulation
- hyper-parameters do not appear in the solutions
- the methods are completely automatic

The problem

- user choices (options) specify

prior knowledge, assumptions, and/or prejudices
about what the true or desirable model is

- model class — imposes hard constraints,
e.g., bound on the model complexity

- fitting criteria — impose soft constraints
e.g., small distance from data to model

- real-life problems are vaguely formulated

"A well defined problem is a half solved problem."

Some user choices

Model class

linear	nonlinear
static	dynamic
time-invariant	time-varying

Fitting criterion

exact	approximate
deterministic	stochastic

Exact identification

we'll consider the simplest (non static) problem:

exact identification of an LTI model

i.e., $\mathscr{M} = \mathscr{L}$ and the fitting criterion is exact match

Why exact identification?

- from simple to complex:

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

- exact identification is ingredient of the other problems

- exact identification leads to effective heuristic approximation methods (subspace methods)

Exact identification in $\mathscr{L}^q$

- given data $\mathscr{D}$
- find $\widehat{\mathscr{B}} \in \mathscr{L}^q$, such that $\mathscr{D} \subset \widehat{\mathscr{B}}$
- nonunique solution always exists

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

- given $(\mathtt{m}, \ell)$ and data $\mathscr{D}$
- find $\widehat{\mathscr{B}} \in \mathscr{L}^q_{\mathtt{m},\ell}$, 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 $(\mathtt{m}, \ell)$, s.t. $\exists\ \widehat{\mathscr{B}} \in \mathscr{L}^{q}_{\mathtt{m},\ell}$, $\mathscr{D} \subset \widehat{\mathscr{B}}$
- J. C. Willems. From time series to linear system—Part II. Exact modelling. *Automatica*, 22(6):675–694, 1986

Why complexity minimization?

- makes the solution unique
- Occam's razor: "simpler $=$ better"

Identifiability question

- Is it possible to recover the data generating system $\overline{\mathscr{B}}$ from exact data

$$w \in \overline{\mathscr{B}} \in \mathscr{L}^q$$

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

- the answer is given by the Fundamantal lemma

- we will assume that upper bounds $\mathtt{n}_{\text{max}}$, ℓ_{max} of the order $\mathtt{n}$ and lag ℓ of $\overline{\mathscr{B}}$ are known

Hankel matrix

- consider the case $\mathscr{D} = w$ (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 $w \in \mathscr{B} \in \mathscr{L}^q$, then image $\big(\mathscr{H}_L(w)\big) \subset \mathscr{B}|_L$
- extra conditions on w and $\mathscr{B}$ are needed for

$$\text{image}\,\big(\mathscr{H}_L(w)\big) = \mathscr{B}|_L$$

Persistency of excitation (PE)

- u is PE of order L if $\mathscr{H}_L(u)$ is full row rank
- system theoretic interpretation:

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

Lemma

1. $\mathscr{B} \in \mathscr{L}^q_{\mathtt{m},\ell}$ controllable and
2. $w \in \mathscr{B}$ admits I/O partition (u, y) with u PE of order $L + \mathtt{p}\ell$

$$\implies \quad \operatorname{image}\big(\mathscr{H}_L(w)\big) = \mathscr{B}|_L$$

Plan

Exact modeling

Algorithms

Exercises

- main idea: any $w \in \mathscr{B}|_L$ can be obtained from $w \in \mathscr{B}$

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

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

Algorithms

- $w \mapsto$ kernel parameter R
- $w \mapsto$ impulse response H
- $w \mapsto$ state/space parameters (A, B, C, D)
  - $w \mapsto R \mapsto (A, B, C, D)$ or $w \mapsto H \mapsto (A, B, C, D)$
  - $w \mapsto$ observability matrix $\mapsto (A, B, C, D)$
  - $w \mapsto$ state sequence $\mapsto (A, B, C, D)$

$w \mapsto R$

- under the assumptions of the lemma

$$\text{image}\left(\mathscr{H}_{\ell+1}(w)\right) = \mathscr{B}|_{\ell+1}$$

- left ker $\left(\mathscr{H}_{\ell+1}(w)\right)$ defines a kernel repr. of $\mathscr{B}$

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

- kernel representation

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

- recursive computation (exploiting Hankel structure)

$w \mapsto H$

- 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, $W = \mathscr{H}_{\ell+t}(w)G$
- define $\mathscr{H}_{\ell+t}(u) =: \begin{bmatrix} U_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}$ and $\mathscr{H}_{\ell+t}(y) =: \begin{bmatrix} Y_\mathrm{p} \\ Y_\mathrm{f} \end{bmatrix}$
- we have

$$\begin{aligned} \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix} G &= \begin{bmatrix} 0 \\ 0 \\ \begin{bmatrix} I_\mathrm{m} \\ 0 \end{bmatrix} \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{zero ini. conditions} \\ \leftarrow \text{ impulse input} \end{matrix} & (1) \\ {\color{red} Y_\mathrm{f} \; G} &{\color{red} = H} & (2) \end{aligned}$$

Block algorithm

- input: u, y, $\ell_{\max}$, and t
- solve (2) and let G_p be a solution
- compute $H = Y_f G_p$
- output: the first t samples of the impulse response H

- Exerise: 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
- Exerise: try the improvements
- application for noisy data

  E. Reynders, R. Pintelon, and G. De Roeck. Consistent impulse-response estimation and system realization from noisy data. *IEEE Trans. Signal Proc.*, 56:2696–2705, 2008

$w \mapsto (A, B, C, D)$

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

- $w \mapsto$ obs. matrix $\mathscr{O}_{\ell+1}(A, C) \xrightarrow{(3)} (A, B, C, D)$

$$\mathscr{O}_{\ell+1}(\mathrm{A,C}) \mapsto (A, C), \quad (u, y, A, C) \mapsto (B, C, x_{\text{ini}}) \tag{3}$$

- $w \mapsto$ state sequence $x \xrightarrow{(4)} (A, B, C, D)$

$$\begin{bmatrix} \sigma x \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \tag{4}$$

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

- C is the first block entry of $\mathscr{O}_{\ell_{\max}+1}(A,C)$
- A is given by the shift equation

$$\left(\sigma^{*}\mathscr{O}_{\ell_{\max}+1}(A,C)\right)A = \left(\sigma\mathscr{O}_{\ell_{\max}+1}(A,C)\right)$$

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

- Once C and A are known, the system of equations

$$y(t) = CA^{t}x(1) + \sum_{\tau=1}^{t-1} CA^{t-1-\tau}Bu(\tau) + D\delta(t+1)$$

is linear in D, B, $x(1)$

$w \mapsto$ observability matrix

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

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

- lin. indep. free responses $\implies$ G maximal rank
- rank revealing factorization

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

$w \mapsto$ 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

$$\begin{bmatrix} U_\text{p} \\ Y_\text{p} \\ U_\text{f} \end{bmatrix} G = \begin{bmatrix} U_\text{p} \\ Y_\text{p} \\ 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} U \\ Y \end{matrix}}\right\} \text{sequential ini. conditions} \\ \leftarrow \quad \text{zero inputs} \end{matrix} \qquad (5)$$
$$Y_\text{f} \; G = Y_0$$

- rank revealing factorization

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

Refinements

- solve (5) efficiently exploiting the Hankel structure
- iteratively compute pieces of $Y_0 \rightsquigarrow$ iterative algorithm
- requires smaller persistency of excitation of u
- could be more efficient

  (solve a few smaller systems of eqns than one big)

References

- N4SID methods

  P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer, Boston, 1996

- MOESP methods

  M. Verhaegen and P. Dewilde. Subspace model identification, Part 1: The output-error state-space model identification class of algorithms. *Int. J. Control*, 56:1187–1210, 1992

MOESP-type algorithms

project the rows of $\mathscr{H}_{\mathrm{n}_{\max}}(y)$ on $\text{row span}^{\perp}\left(\mathscr{H}_{\mathrm{n}_{\max}}(u)\right)$

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

where

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

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

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

$\Longrightarrow$ the orthogonal projection computes free responses

Comments

- $T - \mathtt{n}_{\max} + 1$ free responses are computed via the orth. proj. while $\mathtt{n}_{\max}$ such responses suffice for the purpose of exact identification
- the orth. proj. is a geometric operation, whose system theoretic meaning is not revealed
- the condition for $\operatorname{rank}(Y_0) = \mathtt{n}$, given in the MOESP literature

  $$\operatorname{rank}\left(\begin{bmatrix} X_{\text{ini}} \\ \mathscr{H}_{\mathtt{n}_{\max}}(u) \end{bmatrix}\right) = \mathtt{n} + \mathtt{n}_{\max}\mathtt{m}$$

  is not verifiable from the data (u, y) $\implies$ can not be checked whether the computation gives $\mathscr{O}(A, C)$

N4SID-type algorithms

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

$$\mathscr{H}_{2\mathrm{n}_{\mathrm{max}}}(u) =: \begin{bmatrix} U_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}, \qquad \mathscr{H}_{2\mathrm{n}_{\mathrm{max}}}(y) =: \begin{bmatrix} Y_\mathrm{p} \\ Y_\mathrm{f} \end{bmatrix}$$

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 $\operatorname{row span}(U_\mathrm{f})$ onto $\operatorname{row span}(W_\mathrm{p})$

N4SID-type algorithms

Observe that

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

(in fact Π_obl is the least-norm, least-squares solution)

$\Longrightarrow$ the oblique proj. computes sequential free responses

Comments

- $T - 2\mathtt{n}_{\max} + 1$ sequential free responses are computed via the oblique projection while $\mathtt{n}_{\max} + \mathtt{m} + 1$ such responses suffice for exact ident.
- The oblique proj.\ is a geometric operation, whose system theoretic meaning is not revealed
- The conditions for $\text{rank}(Y_0) = \mathtt{n}$, given in the N4SID literature,
  1. u persistently exciting of order $2\mathtt{n}_{\max}$ and
  2. $\text{row span}(X_{\text{ini}}) \cap \text{row span}(U_{\text{f}}) = \{0\}$

  are not verifiable from the data (u, y)

Summary

- transitions among representations $\approx$ system theory
- exact identification aims at $\mathscr{B}_{\text{mpum}}(w)$
- $\mathscr{H}_t(w)$ plays key role in both analysis and comput.
- under controllability and u persistently exciting

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

- subspace algorithms can be viewed as construction of special responses from data

Plan

Exact modeling

Algorithms

Exercises