Lecture 2: Exact identification

Ivan Markovsky

ELEC doctoral school 2013

Vrije Universiteit Brussel

Outline

Linear static models

Linear time-invariant dynamic models

Exact modeling

Algorithms

Exercises

Linear static model

"Good definition should formalize sensible intuition."
Jan Willems

- linear static model with q variables $=$ subspace of $\mathbb{R}^q$
- model complexity $\quad\leftrightarrow\quad$ subspace dimension
  (the more the model can fit, the less useful it is)
- linear static models with complexity at most $\mathtt{m}$ — $\mathscr{L}^q_{\mathtt{m},0}$
- any $\mathscr{B} \in \mathscr{L}^q_{\mathtt{m},0}$ admits kernel, image, and input/output representations

Representations

- kernel representation with parameter $R \in \mathbb{R}^{\mathtt{p} \times q}$

$$\ker(R) := \{\, d \mid Rd = 0 \,\}$$

- image representation with parameter $P \in \mathbb{R}^{q \times \mathtt{m}}$

$$\operatorname{image}(P) := \{\, d = P\ell \mid \ell \in \mathbb{R}^{\mathtt{m}} \,\}$$

- input/output representation

$$\mathscr{B}_{\text{i/o}}(X, \Pi) := \{\, d = \Pi \begin{bmatrix} u \\ y \end{bmatrix} \mid u \in \mathbb{R}^{\mathtt{m}},\ y = X^{\top} u \,\}$$

with parameters $X \in \mathbb{R}^{\mathtt{m} \times \mathtt{p}}$ and permutation matrix Π

Nonuniqueness of the parameters

- columns of P are generators of the model $\mathscr{B}$
- rows of R are annihilators of $\mathscr{B}$
- the parameters R and P are not unique due to
  - addition of linearly dependent generators/annihilators
  - change of basis transformation

$$\ker(R) = \ker(UR), \qquad \text{for all } U \in \mathbb{R}^{\mathtt{p}\times\mathtt{p}},\ \det(U) \neq 0$$

$$\operatorname{image}(P) = \operatorname{image}(PV), \qquad \text{for all } V \in \mathbb{R}^{\mathtt{m}\times\mathtt{m}},\ \det(V) \neq 0$$

- the smallest number of generators $\mathtt{m} := \dim(\mathscr{B})$
- max. number of annihilators $\mathtt{p} := q - \dim(\mathscr{B})$

Inputs and outputs

- input is a "free" variable

$$\Pi\,\mathrm{col}(u,y) \in \mathscr{B} \text{ and } u \text{ input} \quad \Longleftrightarrow \quad u \in \mathbb{R}^{\mathtt{m}}$$

- output is bound by input and model
- fact: $\mathtt{m} := \dim(\mathscr{B})$ — number of inputs
- $\mathrm{p} := q - \mathtt{m}$ — number of outputs
- generically any I/O partition is possible
- choosing a partition amounts to choosing full rank $\mathrm{p} \times \mathrm{p}$ submatrix of R or full rank $\mathtt{m} \times \mathtt{m}$ submatrix of P

Transition among representations

$$
\begin{tikzcd}
\mathscr{B} = \ker(R) \arrow[rr, leftrightarrow, "RP=0"] \arrow[dr, shift left, "X=-(R_o^{-1}R_i)^\top"] & & \mathscr{B} = \operatorname{image}(P) \arrow[dl, shift right, "X=(P_o P_i^{-1})^\top"'] \\
& \mathscr{B} = \mathscr{B}_{i/o}(X,\Pi) \arrow[ul, shift left, "R=[X^\top \ \ -I]\Pi^\top"] \arrow[ur, shift right, "P^\top=[I \ \ X]\Pi^\top"'] &
\end{tikzcd}
$$

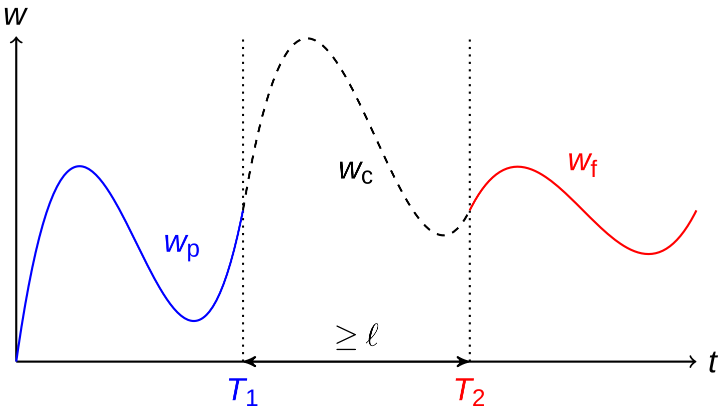
$$
\Pi^\top P =: \begin{array}{cc} \begin{bmatrix} P_i \\ P_o \end{bmatrix} & \begin{array}{c} \mathtt{m} \\ \mathtt{p} \end{array} \end{array} \quad \text{and} \quad R\Pi =: \begin{array}{c} \begin{array}{cc} \mathtt{m} & \mathtt{p} \end{array} \\ \begin{bmatrix} R_i & R_o \end{bmatrix} \end{array}
$$

(for details, see Section 2.1)

Dynamical models

- observations are trajectories: functions $\mathscr{T} \mapsto \mathbb{R}^q$
- universal set: $(\mathbb{R}^q)^{\mathscr{T}}$ — set of functions
- the time axis $\mathscr{T}$ is $\mathbb{Z}$ (discrete) or $\mathbb{R}$ (continuous)
- dynamic model $\mathscr{B}$ is a subset of $(\mathbb{R}^q)^{\mathscr{T}}$
- linearity: $w, v \in \mathscr{B} \implies \alpha w + \beta v \in \mathscr{B}$, $\forall\ \alpha, \beta$
- shift operator: $(\sigma^\tau w)(t) := w(t+\tau)$, for all $t \in \mathscr{T}$
- time-invariance: $\sigma^\tau \mathscr{B} = \mathscr{B}$, for all $\tau \in \mathscr{T}$

Controllability



for all w_p, $w_f \in \mathscr{B}$, $\exists\, w_c$, such that $w_p \wedge w_c \wedge w_f \in \mathscr{B}$

Complexity of an LTI model

- static model $\mathscr{B} \in \mathscr{L}^q_{\mathtt{m},0}$ — complexity $= \mathtt{m}$
  (increasing $\mathtt{m}$ requires increasing # of var. q)

- LTI dynamic model has two aspects:
  - multivariable — number of inputs $\mathtt{m}$
  - dynamics — time memory span ℓ

- complexity of an LTI model is ordered pair $(\mathtt{m}, \ell)$

- notation:
  - $\mathscr{L}^q$ — all LTI models with q variables
  - $\mathscr{L}^q_{\mathtt{m}}$ — at most $\mathtt{m}$ inputs
  - $\mathscr{L}^q_{\mathtt{m},\ell}$ — complexity bounded by $(\mathtt{m}, \ell)$

Restriction of the behavior on an interval

- $w_{\mathrm{p}} \wedge w_{\mathrm{f}}$ — concatenation of w_{p} and w_{f}

  $\mathscr{B}|_T := \{\, w \in (\mathbb{R}^q)^T \mid \exists\, w_{\mathrm{p}},\, w_{\mathrm{f}},\ \text{such that}\ w_{\mathrm{p}} \wedge w \wedge w_{\mathrm{f}} \in \mathscr{B} \,\}$

- for $\mathscr{B} \in \mathscr{L}^q$ and $T > 0$, $\mathscr{B}|_T$ is a subspace

$$\dim(\mathscr{B}|_T) \;\leq\; T\,\mathtt{m} + \mathtt{p}\ell$$

- complexity of $\mathscr{B} \sim \dim(\mathscr{B}|_T)$

- therefore, $(\mathtt{m}, \ell)$ specifies the complexity

Representations

- kernel representation with par. $R(z) \in \mathbb{R}^{g \times q}[z]$

$$\ker(R) = \{\, w \mid R(\sigma)w = R_0 w + R_1 \sigma w + \cdots + R_\ell \sigma^\ell w = 0 \,\}$$

- image representation with par. $P(z) \in \mathbb{R}^{q \times g}[z]$

$$\operatorname{image}(P) = \{\, w = P(\sigma)v \mid \text{for some } v \,\}$$

- input/state/output representation

$$\begin{aligned}\mathscr{B}(A,B,C,D,\Pi) := \{\, & w = \Pi \operatorname{col}(u,y) \mid \\ & \exists\, x,\ \text{such that } \sigma x = Ax + Bu \text{ and } y = Cx + Du \,\}\end{aligned}$$

(default $\Pi = I$, in which case it is skipped)

- any $\mathscr{B} \in \mathscr{L}^q$ admits kernel and I/S/O representations
- any controllable $\mathscr{B} \in \mathscr{L}^q$ admits image representation
- lag of $\mathscr{B}$ — minimal ℓ, for which kernel repr. exists
- minimal $\text{row dim}(R) =$ number of outputs
- minimal $\text{col dim}(P) =$ number of inputs

(for details, see Section 2.2)

Nonuniqueness of I/S/O representation

- choice of an input/output partition
- redundant states (nonminimality of representation)
- change of state space basis

$$\mathscr{B}(A,B,C,D) = \mathscr{B}(T^{-1}AT, T^{-1}B, CT, D),$$

for any nonsingular matrix $T \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$

- minimal representation $\implies$ smallest n = order of $\mathscr{B}$

Transition among representations

- using different representations is a powerful idea
- problems are trivial, given suitable representations
- *cf.*, matrix factorizations in numerical linear algebra
- the problem becomes to transform representations

Identification problems

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

- $\mathscr{U}$ — data space, *e.g.*, a function space $(\mathbb{R}^q)^{\mathbb{N}}$
- $\mathscr{D}$ — data, *e.g.*, a set of finite vector time series

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

- $\mathscr{B}$ — 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 $\mathscr{B}$?)
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 first 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. $\exists \widehat{\mathscr{B}} \in \mathscr{L}^{q}_{\mathtt{m},\ell}$, $\mathscr{D} \subset \widehat{\mathscr{B}}$

Why complexity minimization?

- makes the solution unique
- Occam's razor: "simpler $=$ better"
- recovers the data generating system from exact data under some technical conditions (see page 24)

Hankel matrix

- consider the case $\mathscr{D} = w_\text{d}$ (single traj.)
- 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 $\text{image}\big(\mathscr{H}_L(w)\big) \subset \mathscr{B}|_L$
- 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. $w_{\mathrm{d}} \in \mathscr{B}$ admits I/O partition $(u_{\mathrm{d}}, y_{\mathrm{d}})$ with u_{d} PE of order $L + \mathrm{p}\ell$

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

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

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

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

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

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

- 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)

$w_\mathrm{d} \mapsto H$

- 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, $W = \mathscr{H}_{\ell+t}(w_\mathrm{d})G$

- define $\mathscr{H}_{\ell+t}(u_\mathrm{d}) =: \begin{bmatrix} U_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}$ and $\mathscr{H}_{\ell+t}(y_\mathrm{d}) =: \begin{bmatrix} Y_\mathrm{p} \\ Y_\mathrm{f} \end{bmatrix}$

- we have

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

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

- $w_{\mathrm{d}} \mapsto H(0:2\ell)$ or $R(\xi) \xrightarrow{\text{realization}} (A,B,C,D)$
- $w_{\mathrm{d}} \mapsto$ obs. matrix $\mathscr{O}_{\ell+1}(A,C) \xrightarrow{(1)} (A,B,C,D)$

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

- $w_{\mathrm{d}} \mapsto$ state sequence $x_{\mathrm{d}} \xrightarrow{(2)} (A,B,C,D)$

$$\begin{bmatrix} \sigma x_{\mathrm{d}} \\ y_{\mathrm{d}} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_{\mathrm{d}} \\ u_{\mathrm{d}} \end{bmatrix} \tag{2}$$

$w_\mathrm{d} \mapsto$ observability matrix

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

$$\begin{bmatrix} \mathscr{H}_t(u_\mathrm{d}) \\ \mathscr{H}_t(y_\mathrm{d}) \end{bmatrix} G = \begin{bmatrix} 0 \\ Y_0 \end{bmatrix} \quad \begin{array}{ll} \leftarrow & \text{zero inputs} \\ \leftarrow & \text{free responses} \end{array}$$

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

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

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

$$\begin{bmatrix} U_{\mathrm{p}} \\ Y_{\mathrm{p}} \\ U_{\mathrm{f}} \end{bmatrix} G = \begin{bmatrix} U_{\mathrm{p}} \\ Y_{\mathrm{p}} \\ 0 \end{bmatrix} \begin{array}{l} \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{sequential ini. conditions} \\ \leftarrow \ \text{zero inputs} \end{array}$$

$$Y_{\mathrm{f}} \quad G = Y_0$$

- rank revealing factorization

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

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 comput.
- under controllability and u_{d} persistently exciting

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

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

Exercise 1: Check whether $w_\text{d} \overset{?}{\in} \mathscr{B}$

- $w_\text{d} = (u_\text{d}, y_\text{d}) = \big((0,1),(0,1),(0,1),(0,1)\big)$

```
w = [0 0 0 0; 1 1 1 1];
```

- $\mathscr{B} = \ker\big(R(\sigma)\big)$, where $R(z) = \begin{bmatrix}1 & -1\end{bmatrix} + \begin{bmatrix}-1 & 1\end{bmatrix} z$

```
R = [1 -1 -1 1]; ell = 1;
```

Exercise 2: affine time-invariant system

- an LTI system $\mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell}$ admits a kernel repr.

  $$\mathscr{B} = \ker\big(R(\sigma)\big) := \{w \mid R(\sigma)w = 0\}$$

  for some $R(z) = R_0 z^0 + R_1 z^1 + \cdots + R_\ell z^\ell$

- show that

  $$\mathscr{B}_{\mathrm{c}} := \{\, w \mid R(\sigma)w = c \,\}$$

  is an affine time-invariant system, *i.e.*, $\mathscr{B}_{\mathrm{c}} = \mathscr{B} + w_{\mathrm{p}}$
  for LTI model $\mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell}$ and trajectory w_{p}

- find $\mathscr{B}$ and w_{p}, s.t. $\mathscr{B} + w_{\mathrm{p}} = \{\, w \mid (0.5 + \sigma)w = 1 \,\}$

Exercise 3: transfer function $\mapsto$ kernel repr.

- what model $\mathscr{B}_{\text{tf}}(H)$ is specified by a transfer function

$$H(z) = \frac{q(z)}{p(z)} = \frac{q_0 + q_1 z^1 + \cdots + q_\ell z^\ell}{p_0 + p_1 z^1 + \cdots + p_\ell z^\ell}$$

- find R, such that

$$\mathscr{B}_{\text{tf}}(H) = \ker(R)$$

- write a function `tf2ker` converting H (`tf` object) to R

Exercise 4: Specification of initial conditions

- initial conditions are explicitly specified in I/S/O repr.
- in MATLAB

```
LSIM(SYS,U,T,X0) specifies the initial
state vector X0 at time T(1)
  (for state-space models only).
```

- in transfer function representation initial conditions are often set to 0
- explain how to specify initial conditions in a representation free manner
- what is the link to $x_{\text{ini}} = x(1)$ in I/S/O repr?

Exercise 5: Output matching

- given y_f and $\mathscr{B}$
- find u_f, such that $(u_\mathrm{f}, y_\mathrm{f}) \in \mathscr{B}$

Setup

- random SISO unstable system $\mathscr{B}$

```
clear all, n = 3;
Br = drss(n); [Qr, Pr] = tfdata(Br, 'v');
B = ss(tf(fliplr(Qr), fliplr(Pr), -1));
```

- reference output

```
T = 100; yf = ones(T, 1);
```