ELEC system identification workshop

Behavioral approach

Ivan Markovsky

The course consists of lectures and exercises

- Session 1: behavioral approach to data modeling
- Session 2: subspace identification methods
- Session 3: optimization-based identification methods

"I hear, I forget; I see, I remember; I do, I understand."

session $=$ lecture (you hear and see) $+$ exercises (you do)

Plan

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

Outline

From $Ax = B$ to low-rank approximation

Linear static model representations

Linear time-invariant model representations

Outline

From $Ax = B$ to low-rank approximation

Linear static model representations

Linear time-invariant model representations

A classic line fitting method is solving $Ax \approx B$

problem: fit points $d_1, \ldots, d_N \in \mathbb{R}^2$ by line going through 0

approach: find approximate solution $x \in \mathbb{R}$ of

$$\begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} x = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}, \qquad \text{where} \quad d_i = \begin{bmatrix} a_i \\ b_i \end{bmatrix}$$

the fitting line is $\quad \mathscr{B} := \{ \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathbb{R}^2 \mid ax = b \}$

(x is model parameter)

The choice of a and b is arbitrary

another approach: find approximate solution $x' \in \mathbb{R}$ of

$$\begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} x' = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}$$

the fitting line is $\mathscr{B}' := \{ \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathbb{R}^2 \mid a = bx' \}$

(x' is model parameter)

exceptions:	vertical line	$x = \infty$	$x' = 0$
	horizontal line	$x = 0$	$x' = \infty$

In general, the two solutions differ: $\mathscr{B} \neq \mathscr{B}'$

solving $Ax = B$ and $Bx' = A$ leads to different solutions

the fitting criterion depends on how we choose a and b

the mode representation affects the fitting criterion

$Ax = B$ imposes input/output model structure

functional relations

- $Ax = B$ defined a function $a \mapsto b$
- $Bx = A$ defined a function $b \mapsto a$

in the model $\mathscr{B} := \{ \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathbb{R}^2 \mid ax = b \}$

a is input, b is output (*a* causes *b*)

in the model $\mathscr{B}' := \{ \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathbb{R}^2 \mid bx = a \}$

b is input, a is output (*b* causes *a*)

Model class — set of all candidate models

in the example, the model class is $\mathscr{M} := \{\,\text{lines through } 0\,\}$

separately, $ax = b$ and $bx = a$ don't represent all $\mathscr{B} \in \mathscr{M}$

any $\mathscr{B} \in \mathscr{M}$ is representable as $\mathscr{B} = \{\, \Pi \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \mid ax = b \,\}$

with Π a permutation matrix

Definition of least-squares line fitting problem

given points $\mathscr{D} = \{ d_1, \dots, d_N \} \subset \mathbb{R}^2$ and model class $\mathscr{M}$

$$\text{minimize} \quad \text{over } \mathscr{B} \in \mathscr{M} \quad \text{error}(\mathscr{D}, \mathscr{B})$$

where

$$\text{error}(\mathscr{D}, \widehat{\mathscr{B}}) := \min_{\widehat{\mathscr{D}} \subset \widehat{\mathscr{B}}} \quad \sum_{i=1}^{N} \| d_i - \widehat{d}_i \|_2^2$$

notes:

- $\widehat{\mathscr{D}} \subset \widehat{\mathscr{B}}$ means that $\widehat{\mathscr{B}}$ fits $\{ \widehat{d}_1, \dots, \widehat{d}_N \}$ exactly
- $\widehat{d}_i$ is the projection of d_i on the line $\widehat{\mathscr{B}}$
- $\| d_i - \widehat{d}_i \|_2$ is the orthogonal distance from d_i to $\widehat{\mathscr{B}}$

Any $\mathscr{B} \in \mathscr{M}$ can be represented as kernel

any $\mathscr{B} \in \mathscr{M}$ can be represented as

$$\mathscr{B} = \ker(R) := \{\, d \in \mathbb{R}^2 \mid Rd = 0 \,\}$$

($R \in \mathbb{R}^{1\times 2}$, $R \neq 0$ is a model parameter)

$Rd = 0$ defines a relation (implicit fucntion) between a and b

exact modeling condition

$$\{\, d_1, \dots, d_N \,\} \subset \ker(R) \quad \iff \quad R \underbrace{\begin{bmatrix} d_1 & \cdots & d_N \end{bmatrix}}_{D} = 0$$

Any $\mathscr{B} \in \mathscr{M}$ can be represented as image

any $\mathscr{B} \in \mathscr{M}$ can be represented as

$$\mathscr{B} = \text{image}(P) := \{\, d = P\ell \mid \ell \in \mathbb{R} \,\}$$

($P \in \mathbb{R}^{2\times 1}$ is a model parameter)

$d = P\ell$ also defines a relation between a and b

exact modeling condition

$$\mathscr{D} \subset \text{image}(P) \quad \Longleftrightarrow \quad \begin{bmatrix} d_1 & \cdots & d_N \end{bmatrix} = PL$$

($L \in \mathbb{R}^{1\times N}$ is a latent variable)

For exact data, $\operatorname{rank}\left(\begin{bmatrix} d_1 & \cdots & d_N \end{bmatrix}\right) \leq 1$

common feature of the representations considered

$$\left.\begin{array}{lcl} \exists\ x \in \mathbb{R}, \Pi \text{ permut.} & \begin{bmatrix} x & -1 \end{bmatrix} \Pi D = 0 & \iff \\ \exists\ R \in \mathbb{R}^{1\times 2}, R \neq 0 & RD = 0 & \iff \\ \exists\ P \in \mathbb{R}^{2\times 1}, L \in \mathbb{R}^{1\times N} & D = PL & \iff \end{array}\right\} \operatorname{rank}(D) \leq 1$$

representation free characterization of exact data

$$\mathscr{D} \subset \mathscr{B} \in \mathscr{M}$$
$$\Updownarrow$$
$$\operatorname{rank}(D) = 1$$

Approximate modeling of data is equivalent to low-rank approximation

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{\mathscr{D}} \quad \text{error}(\mathscr{D}, \widehat{\mathscr{D}}) \\ \text{subject to} & \text{exact model for } \widehat{\mathscr{D}} \text{ exists} \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{D} \quad \text{error}(D, \widehat{D}) \\ \text{subject to} & \widehat{D} \text{ is rank deficient} \end{array}$$

Low-rank approximation is a general concept

1. multivariable data fitting $\mathscr{U} = \mathbb{R}^q$
   - linear static model $\leftrightarrow$ subspace
   - model complexity $\leftrightarrow$ subspace dimension
   - rank(D) $\leftrightarrow$ upper bound on the model complexity
2. nonlinear static modeling
   - $\mathscr{D} \mapsto D$ — nonlinear function
   - nonlinearly structured low-rank approximation
3. linear time-invariant dynamical models
   - $\mathscr{D} \mapsto$ Hankel matrix D
   - Hankel structured low-rank approximation

The matrix structure corresponds to the model class

structure $\mathscr{S}$	model class $\mathscr{M}$
unstructured	linear static
Hankel	scalar LTI
$q \times 1$ Hankel	q-variate LTI
$q \times N$ Hankel	N equal length traj.
mosaic Hankel	N general trajectory
$\begin{bmatrix} \text{Hankel} & \text{unstructured} \end{bmatrix}$	finite impulse response
block-Hankel Hankel-block	2D linear shift-invariant

EIV, PCA, and factor analysis are related

errors-in-variables modeling

- all variables are perturbed by noise
- maximum likelihood estimation $\leftrightarrow$ LRA

principal component analysis

- another statistical setting for LRA

factor analysis

- factors $\leftrightarrow$ latent variables in an image representation

Outline

From $Ax = B$ to low-rank approximation

Linear static model representations

Linear time-invariant model representations

A linear static model is a subspace

linear static model with q variables $=$ subspace of $\mathbb{R}^q$

model complexity $\quad\leftrightarrow\quad$ subspace dimension

- $\mathscr{L}_{\mathtt{m},0}$ — linear static models with complexity at most $\mathtt{m}$

$\mathscr{B} \in \mathscr{L}_{\mathtt{m},0}$ admits kernel, image, and I/O representations

A linear static model admits kernel, image, and input/output representations

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

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

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

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

input/output representation with parameters $X \in \mathbb{R}^{\mathrm{m} \times \mathrm{p}}$, Π

$$\mathscr{B}_{\mathrm{i/o}}(X, \Pi) := \{\, d = \Pi \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid u \in \mathbb{R}^{\mathrm{m}},\ y = X^\top u \,\}$$

The parameters R and P are not unique

addition of linearly dependent

- rows of R
- columns of P

minimal representations

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

change of basis transformation

- $\ker(R) = \ker(UR), \quad U \in \mathbb{R}^{\mathtt{p}\times\mathtt{p}}, \det(U) \neq 0$
- $\operatorname{image}(P) = \operatorname{image}(PV), \quad V \in \mathbb{R}^{\mathtt{m}\times\mathtt{m}}, \det(V) \neq 0$

Inputs and outputs can be deduced from $\mathscr{B}$

definition

- input is a "free" variable

$$\Pi \begin{bmatrix} u \\ y \end{bmatrix} \in \mathscr{B} \text{ and } u \text{ input} \quad \iff \quad u \in \mathbb{R}^{\mathtt{m}}$$

- output is bound by input and model

fact:
$$\mathtt{m} := \dim(\mathscr{B}) \text{ — number of inputs}$$
$$\mathtt{p} := q - \mathtt{m} \text{ — number of outputs}$$

choosing an I/O partition amounts to choosing

- full rank $\mathtt{p} \times \mathtt{p}$ submatrix of R
- full rank $\mathtt{m} \times \mathtt{m}$ submatrix of P

It is possible to convert a given representation into an equivalent one

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

$$\Pi^\top P =: \begin{array}{cc} \begin{bmatrix} P_{\mathrm{i}} \\ P_{\mathrm{o}} \end{bmatrix} & \begin{array}{c} \mathtt{m} \\ \mathtt{p} \end{array} \end{array} \quad \text{and} \quad R\Pi =: \begin{array}{c} \begin{array}{cc} \quad\mathtt{m} & \quad\mathtt{p} \end{array} \\ \begin{bmatrix} R_{\mathrm{i}} & R_{\mathrm{o}} \end{bmatrix} \end{array}$$

Outline

From $Ax = B$ to low-rank approximation

Linear static model representations

Linear time-invariant model representations

Dynamical models are sets of functions

observations are trajectories w

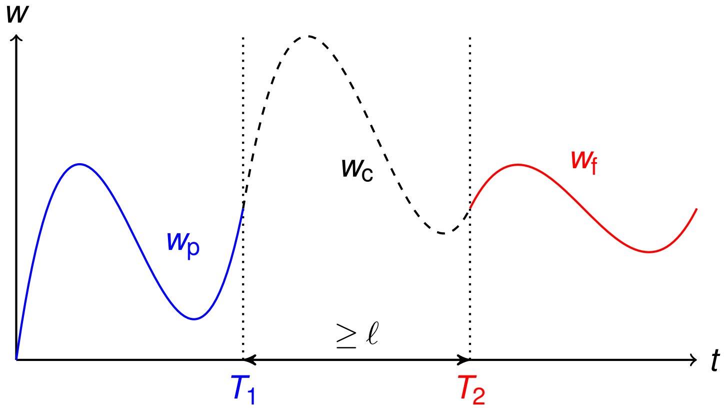
- $(\mathbb{R}^q)^{\mathbb{N}}$ — set of functions from $\mathbb{N}$ to $\mathbb{R}^q$
- shift operator: $(\sigma^\tau w)(t) := w(t+\tau)$, for all $t \in \mathbb{N}$

discrete-time dynamic model $\mathscr{B}$ is a subset of $(\mathbb{R}^q)^{\mathbb{N}}$

properties

- linearity: $w, v \in \mathscr{B} \implies \alpha w + \beta v \in \mathscr{B}$, for all α, β
- time-invariance: $\sigma^\tau \mathscr{B} = \mathscr{B}$, for all $\tau \in \mathbb{N}$

Controllability can be defined in a representation free manner



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

("$\wedge$" denotes "concatenation" of trajectories)

An LTI model admits kernel and input/state/output representations

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

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

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

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

input/state/output representation

$$\mathscr{B}(A,B,C,D,\Pi) := \{\, w = \Pi \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid \\ \text{exists } \ x, \text{ such that } \sigma x = Ax + Bu \text{ and } y = Cx + Du \}$$

Minimal kernel and image representations have full rank R and P parameters

minimal $\operatorname{row dim}(R) =$ number of outputs

minimal $\operatorname{col dim}(P) =$ number of inputs

lag of $\mathscr{B}$ — minimal ℓ, for which kernel repr. exists

The I/S/O representation is not unique

choice of an input/output partition

redundant states (nonminimality of the representation)

- minimal representation $\iff$ $\mathtt{n} =$ order of $\mathscr{B}$

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 $T \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}$

The complexity of an LTI model is determined by the number of inputs and the order

restriction of $\mathscr{B}$ on an interval $[1, T]$

$$\mathscr{B}|_T = \{\, w = \big(w(1), \dots, w(T)\big) \mid \text{there are } w_\text{p}, w_\text{f}, \\ \text{such that } w_\text{p} \wedge w \wedge w_\text{f} \in \mathscr{B} \,\}$$

for sufficiently large T

$$\dim(\mathscr{B}|_T) = (\text{\# of inputs}) \cdot T + (\text{order})$$

$$\text{complexity}(\mathscr{B}) = \begin{bmatrix} \mathtt{m} \\ \ell \end{bmatrix} \begin{matrix} \to & \text{\# of inputs} \\ \to & \text{order or lag} \end{matrix}$$

$\mathscr{L}^q_{\mathtt{m},\ell}$ — LTI models with q variables and complexity bounded by $(\mathtt{m}, \ell)$

Transition among different representations is a powerful problem solving tool

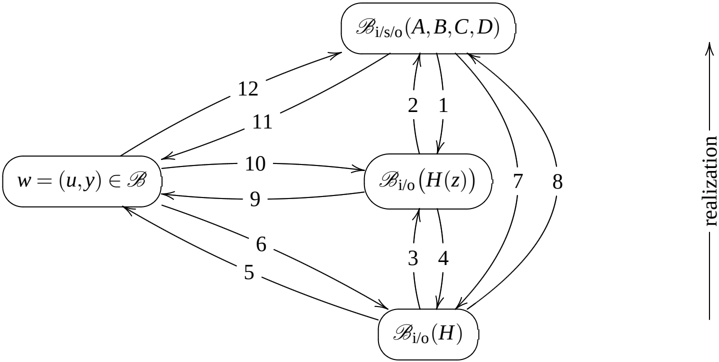
a problem is easier, when suitable representation is used

examples:

- decoupling of a MIMO system
- diagonalization in linear algebra
- pole placement using canonical forms

the problem becomes to transform the representation

$\textit{data} \xrightarrow{\quad\text{identification}\quad} \textit{model}$



1. $H(z) = C(Iz - A)^{-1}B + D$
2. realization of a transfer function
3. Z or Laplace transform of $H(t)$
4. inverse transform of $H(z)$
5. convolution $y_{\mathrm{d}} = H \star u_{\mathrm{d}}$
6. exact identification
7. $H(0) = D$, $H(t) = CA^{t-1}B$ (discrete-time), $H(t) = Ce^{At}B$ (continuous-time), for $t > 0$
8. realization of an impulse response
9. simulation with input u_{d} and $x(0) = 0$
10. exact identification
11. simulation with input u_{d} and $x(0) = x_{\mathrm{ini}}$
12. exact identification