Numerical methods for data-driven systems theory and signal processing

Ivan Markovsky

ICREA

CIMNE
INTERNATIONAL CENTRE
FOR NUMERICAL METHODS
IN ENGINEERING

Systems theory, signal processing and control are going through third paradigm shift

period	paradigm	types of systems
1940–60	classical	SISO transfer funct.
1960–80	modern	MIMO state space
1980–00	behavioral	system as a set
2000–	data-driven	directly using data

New paradigm brings new notion of system and new techniques for problem solving

system	techniques
transfer funct.	Laplace/Z, Fourier transforms
state-space	Lyapunov, Riccati eqn., LMIs
kernel repr.	polynomial algebra
data-driven	(structured) linear algebra

The course consists of lectures and exercises

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

the way to improve at something is by doing it

$$\begin{aligned} \text{session} &= \text{lecture (you hear and see)} \\ &+ \text{exercises (you do)} \end{aligned}$$

bring your computer

homework

Outline

Example: Free fall prediction

Signals and systems; terminology and notation

Linear time-invariant systems

Data-driven representation

Data-driven signal processing and control

Numerical methods for systems and control

Dealing with noise

Outline

As example consider free fall prediction

object is falling in gravitational field

- w — position
- $w(0), \dot{w}(0)$ — initial condition

task: given initial condition, find the trajectory w

- model-based approach:
  1. physics $\mapsto$ parametric model
  2. model parameter estimation
  3. model + ini. conditions $\mapsto$ w
- data-driven approach: data $w_{\mathrm{d}}^1, \ldots, w_{\mathrm{d}}^N$ + ini. cond. $\mapsto$ w

Free fall is a dynamical system

$w(t) \in \mathbb{R}^2$ — variables of interest

$w \in (\mathbb{R}^2)^{\mathbb{R}_+}$ — trajectory (function of time)

$\mathscr{B} \subset (\mathbb{R}^2)^{\mathbb{R}_+}$ — system (set of trajectories)

how to represent / describe $\mathscr{B}$?

Modeling from first principles yields affine time-invariant system

second law of Newton + the law of gravity

$$m\ddot{w} = m\left[\begin{smallmatrix}0\\-9.81\end{smallmatrix}\right] + f, \quad w(0) = w_{\text{ini}} \text{ and } \dot{w}(0) = \dot{w}_{\text{ini}}$$

- m — mass
- 9.81 — gravitational constant
- $f = -\gamma\dot{w}$ — force due to friction in the air

1st order equation

$$\dot{x} = Ax, \quad w = Cx, \quad x(0) = x_{\text{ini}}$$

- state $x := (w_1, \dot{w}_1, w_2, \dot{w}_2, -9.81)$
- initial state $x_{\text{ini}} := (w_{\text{ini},1}, \dot{w}_{\text{ini},1}, w_{\text{ini},2}, \dot{w}_{\text{ini},2}, -9.81)$
- A, C — model parameters (depend on m and γ)

Data-driven free fall prediction requires solving system of linear equations

data: N, discrete-time trajectories $w_\text{d}^1, \dots, w_\text{d}^N$

$$\operatorname{rank} \begin{bmatrix} w_\text{d}^1 & \cdots & w_\text{d}^N \end{bmatrix} = 5 \qquad \text{"informativity" condition}$$

algorithm for data-driven prediction:

1. solve $\begin{bmatrix} w_\text{d}^1(1) & \cdots & w_\text{d}^N(1) \\ w_\text{d}^1(2) & \cdots & w_\text{d}^N(2) \\ w_\text{d}^1(3) & \cdots & w_\text{d}^N(3) \end{bmatrix} g = \underbrace{\begin{bmatrix} w(1) \\ w(2) \\ w(3) \end{bmatrix}}_{\text{ini. cond.}}$

2. define $w := \begin{bmatrix} w_\text{d}^1 & \cdots & w_\text{d}^N \end{bmatrix} g$

Comparison of methods for free fall prediction

model-based (first principles modeling)

- use Newton's 2nd law, law of gravity, and friction
- and model parameters m, γ, gravitational constant
- lead to autonomous affine time-invariant system

direct data-driven (no model identification)

- bypasses the knowledge of the physical laws
- and prior knowledge or estimation of model parameters
- no hyper-parameters to tune

Exercise 5

Outline

Signals are functions of time

$(\mathbb{R}^q)^{\mathscr{T}}$ — signal space: functions $\mathscr{T} \mapsto (\mathbb{R}^q)$

$w \in (\mathbb{R}^q)^{\mathscr{T}}$ — real vector-valued signal

$w(t) \in \mathbb{R}^q$ is the value of w at time $t \in \mathscr{T}$

Signals are classified according to # of variables q and type of time axis $\mathscr{T}$

$q = 1$ — scalar signal
$q > 1$ — vector signal

$\mathscr{T} = \mathbb{R}$ — continuous-time
$\mathscr{T} = \mathbb{Z}$ — discrete-time

$(\mathbb{R}^q)^{\mathbb{R}} \mapsto (\mathbb{R}^q)^{\mathbb{Z}}$ — sampling / time-discretization

Signals are transformed by operators

$(\sigma w)(t) := w(t+1)$ — unit-shift operator

$\underbrace{R_0 + R_1\sigma + \cdots + R_\ell \sigma^\ell}_{R(\sigma)}$ — polynomial operator

$w|_T$ — restriction to the interval $1, \dots, T$

The classical view of dynamical system is an input/output map (a “signal processor”)

$$\text{input} \longrightarrow \boxed{\text{system}} \longrightarrow \text{output}$$

accepts input signal and produces output signal

intuition: the input *causes* the output

In the behavioral approach to systems theory, dynamical system is a set of signals

$\mathscr{B} \subset (\mathbb{R}^q)^{\mathbb{Z}}$ — q-variate discrete-time system

- $q = 1$ — scalar system
- $q > 1$ — multivariable system

$w \in \mathscr{B}$ — w is a trajectory of $\mathscr{B}$

- w is allowed/predicted by $\mathscr{B}$
- $\mathscr{B}$ is unfalsified by w

$\mathscr{B}|_T$ — restriction of $\mathscr{B}$ to the interval $1, \ldots, T$

$\mathscr{B} = \{ w \mid f(w) = 0 \}$ is a representation of $\mathscr{B}$

a given $\mathscr{B}$ allows different representations

- parametric vs non-parametric representations
- uniqueness of the parameters
- how to switch from one representation to another?

different representations $\rightsquigarrow$ different methods

problems related to a system $\mathscr{B}$

- $\mathscr{B} \mapsto w$ — simulation
- $w \mapsto \mathscr{B}$ — identification
- noise filtering, prediction, control, ...

Outline

Linearity, time-invariance, and complexity are defined in terms of the set $\mathscr{B}$

$\mathscr{B}$ is linear system $\quad :\Longleftrightarrow \quad$ $\mathscr{B}$ is linear subspace

$\mathscr{B}$ is time-invariant $\quad :\Longleftrightarrow \quad$ $\sigma\mathscr{B} = \mathscr{B}$

$\mathscr{L}^q$ linear time-invariant (LTI) model class

Three integers $\big(\mathbf{m}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}), \mathbf{n}(\mathscr{B})\big)$ capture the complexity of $\mathscr{B} \in \mathscr{L}^q$

$$\dim \mathscr{B}|_T = \mathbf{m}(\mathscr{B})T + \mathbf{n}(\mathscr{B}), \quad \text{for } T \geq \boldsymbol{\ell}(\mathscr{B})$$

- $\mathbf{m}(\mathscr{B})$ — number of inputs
- $\boldsymbol{\ell}(\mathscr{B})$ — lag
- $\mathbf{n}(\mathscr{B})$ — order

$\mathbf{c}(\mathscr{B}) := \big(\mathbf{m}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}), \mathbf{n}(\mathscr{B})\big)$ — complexity

class of bounded complexity LTI systems

$$\mathscr{L}_c^q := \{\, \mathscr{B} \in \mathscr{L}^q \mid \mathbf{c}(\mathscr{B}) \leq c \}$$

Nonparameteric representation of $\mathscr{B} \in \mathscr{L}^q$

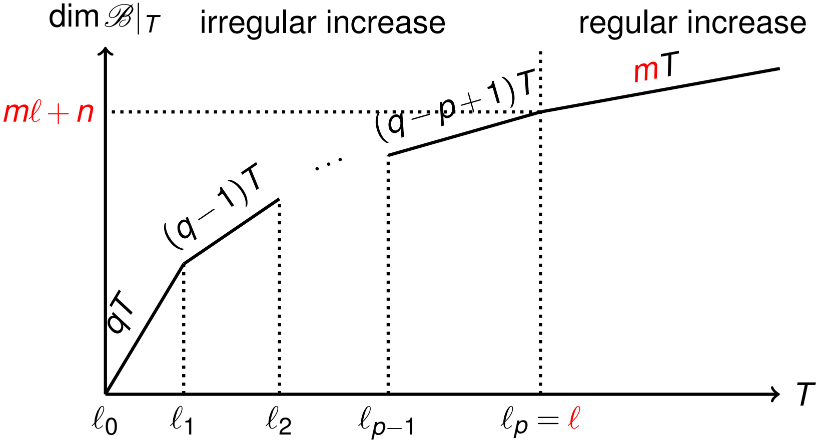
$\mathscr{B}|_T \subset \mathbb{R}^{qT}$ — linear shift-invariant subspace

choose an orthonormal basis for $\mathscr{B}|_T$

$$\mathscr{B}|_T = \operatorname{image} B_T, \quad \text{with } B_T^\top B_T = I_r \qquad (B_T)$$

for $T > \boldsymbol{\ell}(\mathscr{B})$, (B_T) fully defines $\mathscr{B}$

$\dim \mathscr{B}|_T$ is piecewise affine function of T



LTI system's structure is defined by the *structure indices* $(\ell_1, \ldots, \ell_p)$

the structure indices determine the complexity

- $\mathbf{m}(\mathscr{B}) = q - p$
- $\boldsymbol{\ell}(\mathscr{B}) = \ell_p$
- $\mathbf{n}(\mathscr{B}) = \ell_1 + \cdots + \ell_p$

show up in shortest-lag kernel representations

$=$ observability indices in state-space setting

Kernel representation $\mathscr{B} = \ker R(\sigma)$ is ℓ-th order vector difference equation

$$\Big\{\, w \mid R_0 w(t) + R_1 w(t+1) + \cdots + R_\ell w(t+\ell) = 0, \text{ for } t \in \mathscr{T} \,\Big\}$$

$$\Updownarrow$$

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

$$\Updownarrow$$

$$\ker R(\sigma) \qquad \text{(KER)}$$

the parameter is a polynomial matrix $R(z) \in \mathbb{R}^{g\times q}[z]$

A kernel representation is not unique; it has canonical form called *shortest-lag*

$\ker R(\sigma) = \ker UR(\sigma)$, for unimodular matrix U

minimal (KER) — row-dimension $R = p$

shortest-lag (KER) — total degree $R = n$

$$R(z) =: \begin{bmatrix} R^1(z) \\ \vdots \\ R^p(z) \end{bmatrix} \qquad \begin{matrix} \deg R^1(z) = \ell_1 \\ \vdots \\ \deg R^p(z) = \ell_p \end{matrix}$$

Input/state/output representation is 1-st order vector difference equation

$$\mathscr{B}(A,B,C,D,\Pi) := \Big\{ w = \Pi \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \;\Big|\; \text{there is } x \in (\mathbb{R}^n)^{\mathscr{T}}, \\ \text{such that } \sigma x = Ax + Bu,\; y = Cx + Du \Big\} \quad \text{(I/S/O)}$$

x	—	state	,	n	$:=$	$\dim x$	—	order
u	—	input	,	m	$:=$	$\dim u$	—	# of inputs
y	—	output	,	p	$:=$	$\dim y$	—	# of outputs

the parameters are:

- matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$
- permutation matrix $\Pi \in \mathbb{R}^{q\times q}$ and

Summary: linear time-invariant systems

$w \in (\mathbb{R}^q)^{\mathscr{T}}$ signals are functions of time

$\mathscr{B} \subset (\mathbb{R}^q)^{\mathscr{T}}$ systems are sets of signals

$\mathscr{B}$ can be represented by different equations

$\mathscr{L}^q$ LTI model class: shift-invariant subspaces

- complexity: (# of inputs, lag, order)
- $\mathscr{B} = \ker R(\sigma)$ kernel representation
- input/state/output representation

Summary of terminology and notation

$w \in (\mathbb{R}^q)^{\mathbb{N}}$, $w : \mathbb{N} \to \mathbb{R}^q$	q-variate discrete-time signal
$w\vert_T := \big(w(1), \dots, w(T)\big)$	restriction of w to $[1, T]$
$w = w_{\text{ini}} \wedge w_{\text{f}}$	trajectories concatenation
σ, $(\sigma w)(t) := w(t+1)$	unit shift operator
$\mathscr{B} \subset (\mathbb{R}^q)^{\mathbb{N}}$	q-variables discrete-time system
$\mathscr{B}\vert_T := \{\, w\vert_T \mid w \in \mathscr{B} \,\}$	restriction of $\mathscr{B}$ to $[1, T]$
$\mathscr{L}^q$	set of q-variables LTI systems
$\mathbf{m}(\mathscr{B})$ / $\boldsymbol{\ell}(\mathscr{B})$ / $\mathbf{n}(\mathscr{B})$	number of inputs/lag/order of $\mathscr{B}$
$\mathbf{c}(\mathscr{B}) := \big(\mathbf{m}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}), \mathbf{n}(\mathscr{B})\big)$	complexity of $\mathscr{B}$
$\mathscr{L}^q_c := \{\, \mathscr{B} \in \mathscr{L}^q \mid \mathbf{c}(\mathscr{B}) \leq c \,\}$	bounded complexity LTI systems

Exercise 1

Outline

The finite-horizon behavior $\mathscr{B}|_T$ is used for both analysis and computations

restriction of w to finite interval $[1, T]$

$$w|_T := \big(w(1), \dots, w(T)\big) \in (\mathbb{R}^q)^T$$

restriction of $\mathscr{B}$ to $[1, T]$

$$\mathscr{B}|_T := \{\, w|_T \mid w \in \mathscr{B} \,\} \subset (\mathbb{R}^q)^T$$

if $\mathscr{B}$ is linear, $\mathscr{B}|_T$ is a linear subspace of $(\mathbb{R}^q)^T$

$\mathscr{B}|_T$ can be obtained experimentally by collecting "informative" data

collect $N \geq qT$ random trajectories

$$w_{\rm d}^1, \ldots, w_{\rm d}^N \in \mathscr{B}|_T$$

by the linearity of $\mathscr{B}$, we have

$$\operatorname{span}\{\, w_{\rm d}^1, \ldots, w_{\rm d}^N \,\} \subseteq \mathscr{B}|_T$$

with probability one equality holds

Discrete-time LTI systems over finite horizon can be studied using linear algebra only

$$\underbrace{\begin{bmatrix} w_\mathrm{d}^1 & \cdots & w_\mathrm{d}^N \end{bmatrix}}_{W_\mathrm{d}} \in \mathbb{R}^{qT \times N} \text{ — “trajectory matrix”}$$

$\widehat{\mathscr{B}}|_T = \text{image}\, W_\mathrm{d}$ — data-driven representation

how to obtain (B_T) from one trajectory $w_\mathrm{d} \in \mathscr{B}$?

Data-driven representation (infinite horizon)

data: exact infinite trajectory w_{d} of $\mathscr{B} \in \mathscr{L}$

$\widehat{\mathscr{B}} = \mathscr{B}_{\mathrm{mpum}}(w_{\mathrm{d}}) = \operatorname{span}\{ w_{\mathrm{d}}, \sigma w_{\mathrm{d}}, \sigma^2 w_{\mathrm{d}}, \dots \}$

identifiability condition: $\quad \mathscr{B} = \widehat{\mathscr{B}}$

Consecutive application of σ on finite w_d results in Hankel matrix with missing values

$$
\begin{array}{cccc}
\sigma^0 w_\mathrm{d} & \sigma^1 w_\mathrm{d} & \cdots & \sigma^{T_\mathrm{d}-1} w_\mathrm{d} \\
\hline
w_\mathrm{d}(1) & w_\mathrm{d}(2) & \cdots & w_\mathrm{d}(T_\mathrm{d}) \\
w_\mathrm{d}(2) & \vdots & \iddots & ? \\
\vdots & w_\mathrm{d}(T_\mathrm{d}) & \iddots & \vdots \\
w_\mathrm{d}(T_\mathrm{d}) & ? & \cdots & ?
\end{array}
$$

for $w_\mathrm{d} = \big(w_\mathrm{d}(1), \dots, w_\mathrm{d}(T_\mathrm{d})\big)$ and $1 \leq T \leq T_\mathrm{d}$

$$
\mathscr{H}_T(w_\mathrm{d}) := \Big[(\sigma^0 w_\mathrm{d})|_T \;\; (\sigma^1 w_\mathrm{d})|_T \;\; \cdots \;\; (\sigma^{T_\mathrm{d}-T} w_\mathrm{d})|_T \Big]
$$

Data-driven representation (finite horizon)

the finite horizon data-driven representation

$$\mathscr{B}|_T = \widehat{\mathscr{B}}|_T := \operatorname{image} \mathscr{H}_T(w_\mathrm{d}) \qquad \text{(DD-REPR)}$$

for $T \geq \boldsymbol{\ell}(\mathscr{B})$ holds if and only if

$$\operatorname{rank} \mathscr{H}_T(w_\mathrm{d}) = T\mathbf{m}(\mathscr{B}) + \mathbf{n}(\mathscr{B}) \qquad \text{(GPE)}$$

GPE — generalized persistency of excitation

Identifiability condition verifiable from $w_{\mathrm{d}} \in \mathscr{B}|_{T_{\mathrm{d}}}$ and (m, ℓ, n)

fact: $\mathscr{B} = \mathscr{B}' \iff \mathscr{B}|_{\ell+1} = \mathscr{B}'|_{\ell+1}$ then

$$\begin{aligned} \widehat{\mathscr{B}} = \mathscr{B} \quad &\iff \quad \widehat{\mathscr{B}}|_{\ell+1} = \mathscr{B}|_{\ell+1} \\ &\iff \quad \dim \widehat{\mathscr{B}}|_{\ell+1} = \dim \mathscr{B}|_{\ell+1} \end{aligned}$$

$\mathscr{B}$ is identifiable from $w_{\mathrm{d}} \in \mathscr{B}|_{T_{\mathrm{d}}}$ if and only if

$$\operatorname{rank} \mathscr{H}_{\ell+1}(w_{\mathrm{d}}) = (\ell + 1)m + n \qquad \text{(GPE-ID)}$$

$w_{\mathrm{d}} \mapsto \mathscr{B}$ — system identification

Summary: data-driven representation

assuming $\quad \operatorname{rank} \mathscr{H}_T(w_\mathrm{d}) = \mathbf{m}(\mathscr{B})T + \mathbf{n}(\mathscr{B})$

$\mathscr{B}|_T = \operatorname{image} \mathscr{H}_T(w_\mathrm{d})$ $\quad$ holds

replaces parametric representations

Exercise 2

Outline

Generic data-driven problem: trajectory interpolation/approximation

given: "data trajectory" $w_{\mathrm{d}} \in \mathscr{B}|_{T_{\mathrm{d}}}$
and elements $w|_{I_{\text{given}}}$
of a trajectory $w \in \mathscr{B}|_T$

($w|_{I_{\text{given}}}$ selects the elements of w, specified by I_{given})

aim: minimize over $\widehat{w}$ $\quad \| w|_{I_{\text{given}}} - \widehat{w}|_{I_{\text{given}}} \|$
subject to $\quad \widehat{w} \in \mathscr{B}|_T$

$$\widehat{w} = \mathscr{H}_T(w_{\mathrm{d}}) \big(\mathscr{H}_T(w_{\mathrm{d}})|_{I_{\text{given}}} \big)^{+} w|_{I_{\text{given}}} \qquad \text{(SOL)}$$

Special cases

simulation

- given data: initial condition and input
- to-be-found: output (exact interpolation)

smoothing

- given data: noisy trajectory
- to-be-found: ℓ_2-optimal approximation

tracking control

- given data: to-be-tracked trajectory
- to-be-found: ℓ_2-optimal approximation

Generalizations

multiple data trajectories $w_\mathrm{d}^1, \dots, w_\mathrm{d}^N$

$$\widehat{\mathscr{B}}|_T = \text{image}\, \underbrace{\begin{bmatrix} \mathscr{H}_T(w_\mathrm{d}^1) & \cdots & \mathscr{H}_T(w_\mathrm{d}^N) \end{bmatrix}}_{\text{mosaic-Hankel matrix}}$$

w_d not exact / noisy

maximum-likelihood estimation
$\rightsquigarrow$ Hankel structured low-rank approximation/completion
nuclear norm and ℓ_1-norm relaxations
$\rightsquigarrow$ nonparametric, convex optimization problems

nonlinear systems

results for special classes of nonlinear systems:
Volterra, Wiener-Hammerstein, bilinear, ...

Exercise 3

Outline

We want to obtain B_T from other representations or data

$$(A, B, C, D) \mapsto B_T = \Pi \begin{bmatrix} 0_{mT \times n} & I_{mT} \\ \mathscr{O}_T(A,C) & \mathscr{T}_T(H) \end{bmatrix} \qquad (\texttt{ss2BT})$$

$$\text{kernel: } R \xmapsto{???} B_T \qquad (\texttt{R2BT})$$

$$\text{data: } \mathscr{W}_{\mathrm{d}} = \{ w_{\mathrm{d}}^1, \ldots, w_{\mathrm{d}}^N \} \xmapsto{???} B_T \qquad (\texttt{w2BT})$$

$$\texttt{w2BT:}\ \mathscr{W}_\mathrm{d} \xmapsto{\ \texttt{hank}\ } \mathscr{H}_T(\mathscr{W}_\mathrm{d}) \xmapsto{\ \texttt{orth}\ } B_T$$

$$\mathscr{W}_\mathrm{d} = \{\, w_\mathrm{d}^1, \dots, w_\mathrm{d}^N \,\}, \quad w_\mathrm{d}^i = \big(w_\mathrm{d}^i(1), \dots, w_\mathrm{d}^i(T_\mathrm{d}^i)\big) \in \mathscr{B}|_{T_\mathrm{d}^i}$$

Hankel matrix

$$\mathscr{H}(\mathscr{W}_\mathrm{d}) := \begin{bmatrix} \mathscr{H}(w_\mathrm{d}^1) & \cdots & \mathscr{H}(w_\mathrm{d}^N) \end{bmatrix}$$

$$\mathscr{H}_T(w_\mathrm{d}^i) := \begin{bmatrix} (\sigma^0 w_\mathrm{d}^i)|_T & (\sigma^1 w_\mathrm{d}^i)|_T & \cdots & (\sigma^{T_\mathrm{d}^i - T} w_\mathrm{d})|_T \end{bmatrix}$$

theorem

$$\mathscr{B}|_T = \operatorname{image} \mathscr{H}_T(\mathscr{W}_\mathrm{d}) \qquad \text{(DDR)}$$

$$\Updownarrow$$

$$\operatorname{rank} \mathscr{H}_T(\mathscr{W}_\mathrm{d}) = \mathbf{m}(\mathscr{B})T + \mathbf{n}(\mathscr{B}) \qquad \text{(GPE)}$$

R2BT: $R \xmapsto{\texttt{multmat}} \mathscr{M}_T(R) \xmapsto{\texttt{null}} B_T$

$$\mathscr{B} = \ker R(\sigma), \quad R(z) = \begin{bmatrix} R^1(z) \\ \vdots \\ R^g(z) \end{bmatrix} = \begin{bmatrix} R^1_0 + R^1_1 z + \cdots + R^1_{\ell_1} z^{\ell_1} \\ \vdots \\ R^g_0 + R^g_1 z + \cdots + R^g_{\ell_g} z^{\ell_g} \end{bmatrix}$$

multiplication matrix

$$\mathscr{M}_T(R) := \begin{bmatrix} \mathscr{M}_T(R^1) \\ \vdots \\ \mathscr{M}_T(R^g) \end{bmatrix}, \quad \mathscr{M}_T(R^i) := \begin{bmatrix} R^i_0 & R^i_1 & \cdots & R^i_{\ell_i} & & \\ & \ddots & \ddots & & \ddots & \\ & & R^i_0 & R^i_1 & \cdots & R^i_{\ell_i} \end{bmatrix}$$

theorem

$$\mathscr{B}|_T = \ker \mathscr{M}_T(R) \qquad \text{(FHK)}$$

$$\Updownarrow$$

$$\operatorname{rank} R = \mathbf{p}(\mathscr{B})\, T - \mathbf{n}(\mathscr{B}) \qquad \text{(DGPE)}$$

B_T is structured, due to the LTI dynamics

the structure is fully revealed when $T \geq \boldsymbol{\ell}(\mathscr{B})$

$\mathscr{B}|_T$ is $(mT+n)$-dimensional shift-invariant subspace

complexity bounded $\iff$ $\dim \mathscr{B}|_T < qT$

$$\mathscr{B} \in \mathscr{L}^q_{(m,\ell,n)} \implies \dim \mathscr{B}|_T = mT + n = \operatorname{rank} B_T$$

time-invariance $\iff$ shift-invariance

- autonomous case: sum-of-exponentials
- open systems: the structure is hidden
- image $\mathscr{H}_T(w_\mathrm{d})$ and $\ker \mathscr{M}_T(R)$ impose shift-invariance

Computing B_T has hidden dangers

the (GPE) / (DGPE) condition has to hold

rank computation is needed

$$\dim \mathscr{B}|_T = qT - \operatorname{rank} \mathscr{M}_T(R) = \operatorname{rank} \mathscr{H}_T(\mathscr{W}_\mathrm{d}) \qquad (\dim \mathscr{B}|_T)$$

the numerical rank is tolerance dependent

$$\widehat{r} := \#\text{ of singular values} \geq \varepsilon \qquad (\widehat{r})$$

The SVD approximation doesn't impose time-invariant structure on B_T

two options for complexity estimation:

1. specify ε, in which case r is found from ($\widehat{r}$), or
2. $c := (m, \ell, n)$, in which case r is found from ($\dim \mathscr{B}|_T$)

SVD approximation

when "small" singular values are discarded,
B_T is not shift-invariant and therefore
does not represent $\mathscr{B}|_T$ of $\mathscr{B} \in \mathscr{L}_c^q$

The horizon T depends on the usage of (**??**)

in system analysis, where $\mathscr{B} \in \mathscr{L}_c^q$ is given

- T is of the order of $\boldsymbol{\ell}(\mathscr{B})$
- B_T is constructed from another repr. or data $w_\mathrm{d} \in \mathscr{B}|_{T_\mathrm{d}}$

$$T_\mathrm{d} \geq T_\mathrm{d,min} := \big(\mathbf{m}(\mathscr{B}) + 1\big)\, T + \mathbf{n}(\mathscr{B}) - 1$$

- the results are exact up to numerical errors ($\sim 10^{-10}$)

in SYSID/data-driven control, where $\mathscr{W}_\mathrm{d}$ is given

- T may be large (*e.g.*, control horizon in DeePC)
- $\mathscr{W}_\mathrm{d} \mapsto B_T$ is nontrivial (it is an identification problem)
  typically $T_\mathrm{d} \gg T$ and generally it is best to use all the data
- the results are approximate and the errors “large” (∼20%)

Computing kernel representation from B_T is trivial, however, making it minimal is nontrivial

`R = null(BT')'` — kernel repr. of $\mathscr{B}$

for multi-output systems, it is not minimal

finding minimal (shortest-lag) kernel repr.

start with $R = \emptyset$ *(a trivial model* $(\mathbb{R}^q)^{\mathbb{N}}$*)*
for $T = 1, \dots, \ell + 1$*, if* B_T *has new annihilators, add them to* R

data-driven approach for reducing nonminimal R

$$R \mapsto B_T \mapsto R_{\text{min}}$$

Minimal state-space representation can be computed by past/future intersection

1. compute a state X by past/future intersection

$$\begin{bmatrix} B_{\rm p} \\ B_{\rm f} \end{bmatrix} := B_T, \quad \text{row span}\, X = \text{row span}\, B_{\rm p} \cap \text{row span}\, B_{\rm f}$$

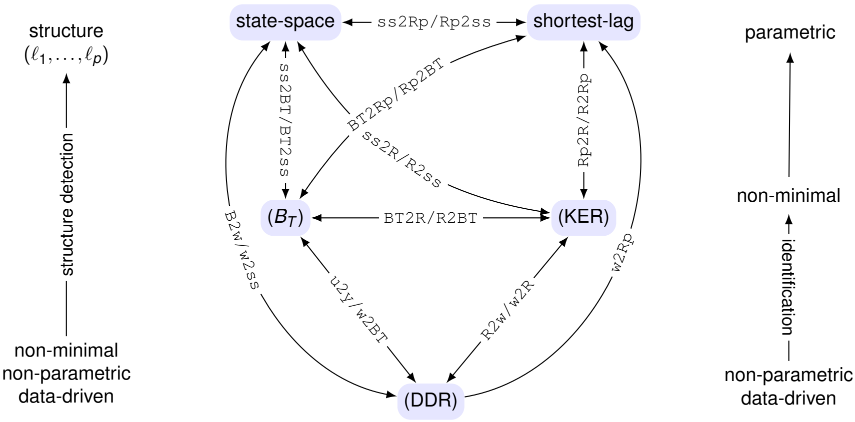
2. make it minimal by low-rank approximation: $X \mapsto X_{\min}$
3. compute the state-space parameters

$$\begin{bmatrix} \sigma X_{\min} \\ Y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_{\min} \\ U \end{bmatrix}$$

important details:

- how to compute the intersection?
- the resulting X is not sequential, how to find σX?
- minimality is achieved by low-rank approximation of X

Transitions among representations



Exercise 4

Outline

The data w_d being exact vs inexact / “noisy”

w_d exact and satisfying (GPE)

- “systems theory” problems
- image $\mathscr{H}_T(w_\mathrm{d})$ is nonparametric finite-horizon model
- data-driven solution $=$ model-based solution

w_d inexact, due to noise and/or nonlinearities

- naive approach: apply the solution (SOL) for exact data
- rigorous: assume noise model $\rightsquigarrow$ ML estimation problem
- heuristics: convex relaxations of the ML estimator

The maximum-likelihood estimation problem in the errors-in-variables setup is nonconvex

errors-in-variables setup: $\quad w_{\rm d} = \overline{w}_{\rm d} + \widetilde{w}_{\rm d}$

- $\overline{w}_{\rm d}$ — true data, $\overline{w}_{\rm d} \in \mathscr{B}|_{T_{\rm d}}$, $\mathscr{B} \in \mathscr{L}_c^q$
- $\widetilde{w}_{\rm d}$ — zero mean, white, Gaussian measurement noise

ML problem: given $w_{\rm d}$, c, and $w|_{I_{\rm given}}$

$$\begin{aligned} \underset{g}{\text{minimize}} \quad & \| w|_{I_{\rm given}} - \mathscr{H}_T(\widehat{w}_{\rm d}^*)|_{I_{\rm given}} g \| \\ \text{subject to} \quad & \widehat{w}_{\rm d}^* = \arg\min_{\widehat{w}_{\rm d}, \widehat{\mathscr{B}}} \quad \| w_{\rm d} - \widehat{w}_{\rm d} \| \\ & \qquad \text{subject to} \quad \widehat{w}_{\rm d} \in \widehat{\mathscr{B}}|_{T_{\rm d}} \text{ and } \widehat{\mathscr{B}} \in \mathscr{L}_c^q \end{aligned}$$

The ML estimation problem is equivalent to Hankel structured low-rank approximation

$$\begin{aligned}
&\underset{g}{\text{minimize}} \quad \| w|_{I_{\text{given}}} - \mathscr{H}_T(\widehat{w}_{\text{d}}^*)|_{I_{\text{given}}} g \| \\
&\text{subject to} \quad \widehat{w}_{\text{d}}^* = \arg\min_{\widehat{w}_{\text{d}}, \widehat{\mathscr{B}}} \quad \| w_{\text{d}} - \widehat{w}_{\text{d}} \| \\
&\qquad\qquad\qquad \text{subject to} \quad \widehat{w}_{\text{d}} \in \widehat{\mathscr{B}}|_{T_{\text{d}}} \text{ and } \widehat{\mathscr{B}} \in \mathscr{L}_c^q
\end{aligned}$$

$$\Updownarrow$$

$$\begin{aligned}
&\underset{g}{\text{minimize}} \quad \| w|_{I_{\text{given}}} - \mathscr{H}_T(\widehat{w}_{\text{d}}^*)|_{I_{\text{given}}} g \| \\
&\text{subject to} \quad \widehat{w}_{\text{d}}^* = \arg\min_{\widehat{w}_{\text{d}}} \quad \| w_{\text{d}} - \widehat{w}_{\text{d}} \| \\
&\qquad\qquad\qquad \text{subject to} \quad \operatorname{rank} \mathscr{H}_{\ell+1}(\widehat{w}_{\text{d}}) \leq (\ell+1)m + n
\end{aligned}$$

Solution methods

local optimization

- choose a parametric representation of $\widehat{\mathscr{B}}(\theta)$
- optimize over $\widehat{w}$, $\widehat{w_{\mathrm{d}}}$, and θ
- depends on the initial guess

convex relaxation based on the nuclear norm

$$\text{minimize} \quad \text{over } \widehat{w}_{\mathrm{d}} \text{ and } \widehat{w} \quad \|w|_{I_{\text{given}}} - \widehat{w}|_{I_{\text{given}}}\| + \|w_{\mathrm{d}} - \widehat{w}_{\mathrm{d}}\| + \gamma \cdot \left\| \begin{bmatrix} \mathscr{H}_{\Delta}(\widehat{w}_{\mathrm{d}}) & \mathscr{H}_{\Delta}(\widehat{w}) \end{bmatrix} \right\|_{*}$$

convex relaxation based on ℓ_1-norm (LASSO)

$$\text{minimize} \quad \text{over } g \quad \|w|_{I_{\text{given}}} - \mathscr{H}_{T}(w_{\mathrm{d}})|_{I_{\text{given}}} g\| + \lambda \|g\|_1$$

Summary: convex relaxations

w_d exact $\rightsquigarrow$ systems theory

- exact analytical solution
- current work: efficient real-time algorithms

w_d inexact $\rightsquigarrow$ nonconvex optimization

- subspace methods
- local optimization
- convex relaxations

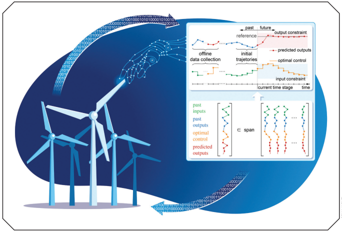
empirical validation

- the naive approach works (surprisingly) well
- parametric local optimization is not robust
- ℓ_1-norm regularization gives the best results

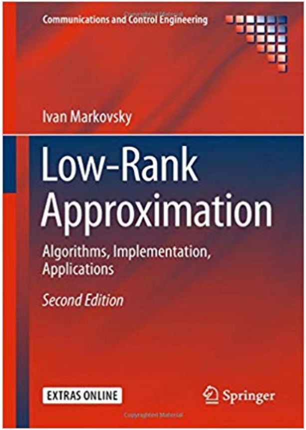
References

Data-Driven Control Based on the Behavioral Approach

FROM THEORY TO APPLICATIONS IN POWER SYSTEMS



IVAN MARKOVSKY, LINBIN HUANG, and FLORIAN DÖRFLER



Research topics

systems theory

- descriptor systems
- distance to uncontrollability
- dynamic networks: interconnection of systems

signal processing

- input estimation
- missing data estimation
- data-driven smoothing and control

Research topics

numerical linear algebra

- polynomial algebra
- structure-exploiting methods
- uncertainty quantification

applications

- fault detection
- dynamic measurements
- optimal sensor / actuator placement

Exercises

1. Checking if a signal is a trajectory

2. Finding the data-generating system

3. Projection on $\mathscr{B}|_T$

4. Equality of systems

5. Case study: free fall prediction

Given $w \in (\mathbb{R}^q)^T$ and $R(z) \in \mathbb{R}^{g \times q}[z]$, checking if $w \in \mathscr{B}|_T$, where $\mathscr{B} := \ker R(\sigma)$

1. propose a method for solving the problem
2. implement it in a function `w_in_ker(w, R)`
3. test it with the signal

$$w = (u, y) = \big((0,1),(0,1),(0,1),(0,1)\big)$$

and the system

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

Homework

1. checking $\quad w \overset{?}{\in} \mathscr{B}(A, B, C, D)$

2. checking $\quad w \overset{?}{\in} \text{image}\, B_T$

as in the exercise

1. propose method
2. implement it
3. test it on examples

Given $w_\mathrm{d} \in \mathscr{B}|_{T_\mathrm{d}}$ and T,
where $\mathscr{B} \in \mathscr{L}^q$, find (B_T)

1. assuming (GPE) holds,
   propose a method for finding (B_T)

2. implement the method in a function

   ```
   BT = w2BT(w, T)
   ```

3. test it on examples

Homework

how to check the result of `w2BT`?

how to convert (I/S/O) to ($\mathcal{B}_T$)?

- propose different methods
- implement the simplest one in function

```
BT = ss2BT(B, T)
```

Given $w_{\rm d} \in \mathscr{B}|_{T_{\rm d}}$ and ℓ,
where $\mathscr{B} \in \mathscr{L}^q_{(m,\ell,n)}$, find (KER)

1. assuming that (GPE-ID) holds,
   propose a method for finding (KER)

2. implement the method in a function

   ```
   R = w2R(w, ell)
   ```

3. test it on

   $\ell = 3, \ w_{\rm d} = (1, 2, 4, 7, 13, 24, 44, 81)$

Homework

generalize the method for unknown $\ell \geq \boldsymbol{\ell}(\mathscr{B})$

compare the methods of finding (B_T) and (KER)

Projection of $w \in (\mathbb{R}^q)^T$ on $\mathscr{B}|_T$

1. using (B_T) solve the optimization problem

$$\text{minimize} \quad \text{over } \widehat{w} \quad \|w - \widehat{w}\| \quad \text{subject to} \quad \widehat{w} \in \mathscr{B}|_T \quad \text{(KS)}$$

2. implement the solution in a function

```
wh = proj_BT(w, BT)
```

3. test it on examples

4. discuss possible applications

Homework

the trajectory approximation/interpolation problem on slide 43 is generalization of (KS)

implement the solution (SOL)

test it on examples

Checking if two systems are equal

we would like this code to give 'true'

```
m = 2; p = 2; n = 3;
sys1 = drss(n, p, m);
sys2 = ss2ss(sys1, rand(n));
sys1 == sys2
```

instead it gives an error:

```
Operator '==' is not supported
for operands of type 'ss'.
```

1. propose methods for checking systems equality

2. implement and test them

Free fall as a dynamical system

setup: mass m falling in gravitational field

- $w \in (\mathbb{R}^2)^{\mathbb{R}_+}$ — position in 2D plane
- $v := \dot{w} \in (\mathbb{R}^2)^{\mathbb{R}_+}$ — velocity
- $w(0), v(0) \in \mathbb{R}^2$ — initial condition

task: given initial condition, find the trajectory w

first work out the model-based approach

1. using physics, derive model (include friction force $-\gamma\dot{w}$)
2. write a function `w = fall(w0, v0, t, m, gamma)`

simulate $T_{\mathrm{d}} = 100$-samples trajectories

- $N = 10$ "data" trajectories $w_{\mathrm{d}}^1, \dots, w_{\mathrm{d}}^N$ and
- one "to-be-predicted" trajectory w

verify the data "informativity" condition

$$\operatorname{rank} \begin{bmatrix} w_{\mathrm{d}}^1 & \cdots & w_{\mathrm{d}}^N \end{bmatrix} = 5$$

implement and verify the data-driven method

1. solve $\begin{bmatrix} w_{\mathrm{d}}^1(1) & \cdots & w_{\mathrm{d}}^N(1) \\ w_{\mathrm{d}}^1(2) & \cdots & w_{\mathrm{d}}^N(2) \\ w_{\mathrm{d}}^1(3) & \cdots & w_{\mathrm{d}}^N(3) \end{bmatrix} g = \begin{bmatrix} w(1) \\ w(2) \\ w(3) \end{bmatrix}$
2. define $w := \begin{bmatrix} w_{\mathrm{d}}^1 & \cdots & w_{\mathrm{d}}^N \end{bmatrix} g$