# Data-driven structured noise filtering via common dynamics estimation

Ivan Markovsky

VRIJE UNIVERSITEIT BRUSSEL

# Motivation: separate signal from structured as well as random noises

## classical setup

- signal — trajectory of (low-order) LTI system
- noise — zero-mean (white, Gaussian) random process

## structured noise: offset, ramp, sine, ...

also trajectory of low-order LTI system

## how to distinguish signal from structured noise?

# In multiple experiments, the signal dynamics is fixed, while the noise dynamics varies

new setup: data from multiple experiments

filtering principle: select "common dynamics"

two problems:

- common dynamics detection (no random noise)
- common dynamics approximation (with random noise)

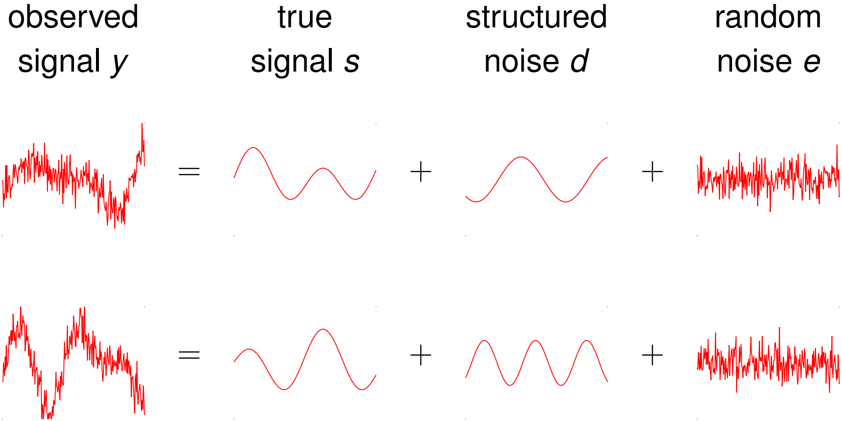
# Plan


Example

Subspace methods

Optimization methods

# The observed signals are sums of true values, structured noises, and random noises



# A dynamical system $\mathscr{B}$ is a set of signals $y$ Common dynamics — intersection of subsets

$$y \in \mathscr{B} \quad \iff \quad y \text{ is trajectory of } \mathscr{B}$$
$$\iff \quad \mathscr{B} \text{ is exact model for } y$$

$$\mathscr{B} \text{ subspace} \quad \iff \quad \mathscr{B} \text{ is a linear system}$$

$$\mathscr{B}_1 \cap \cdots \cap \mathscr{B}_N \quad — \quad \text{common dynamics}$$

# We consider the set of autonomous linear time-invariant systems $\mathscr{L}_n$

$$\mathscr{B} \in \mathscr{L}_n \quad \Longrightarrow \quad \dim(\mathscr{B}) = n$$

$$\mathscr{B} \in \mathscr{L}_n \quad \Longleftrightarrow \quad \mathscr{B} \text{ has kernel representation}$$

$$\mathscr{B} = \ker p(\sigma) := \{\, y \mid p_0 y + p_1 \sigma y + \cdots + p_n \sigma^n y = 0 \,\}$$

shift operator $\quad (\sigma y)(t) := y(t+1)$

# The common dynamics computation with given models is a GCD problem

given: $\mathscr{B}_1 = \ker p^1(\sigma),\ \dots,\ \mathscr{B}_N = \ker p^N(\sigma)$

$\mathscr{B} = \mathscr{B}_1 \cap \cdots \cap \mathscr{B}_N = \ker p(\sigma)$ is given by

the greatest common divisor

$$p(z) = \mathrm{GCD}\big(p^1(z), \dots, p^N(z)\big)$$

# Data-driven common dynamics computation: given are trajectories instead of models

## given

- (noisy) trajectories $y_i = \bar{y}_i + e_i$, where $\bar{y}_i \in \mathscr{B}_i \in \mathscr{L}_n$
- model orders $n = \dim \mathscr{B}_i$
- common dynamics order $n_{\mathrm{s}}$

## we aim to find $\mathscr{B} = \mathscr{B}_1 \cap \cdots \cap \mathscr{B}_N$

## indirect solution

1. identification: $(y_i, n) \mapsto p^i$, for $i = 1, \ldots, N$
2. (approximate) GCD computation: $(p^1, \ldots, p^N, n_{\mathrm{s}}) \mapsto p$

# Hankel matrix of data specifies the system

$\mathscr{B}|_L$ — restriction of $\mathscr{B}$ to the interval $[1, L]$

$\bar{y}_i$ is persistently exciting of order $n$ if

$$\operatorname{rank} \underbrace{\begin{bmatrix} \bar{y}_i(1) & \bar{y}_i(2) & \cdots & \bar{y}_i(T-n+1) \\ \bar{y}_i(2) & \bar{y}_i(3) & \cdots & \bar{y}_i(T-n+2) \\ \vdots & \vdots & & \vdots \\ \bar{y}_i(n) & \bar{y}_i(n+1) & \cdots & \bar{y}_i(T) \end{bmatrix}}_{\mathscr{H}_n(\bar{y}_i)} = n$$

if $\bar{y}_i \in \mathscr{B}_i \in \mathscr{L}_n$ is persistently exciting of order $n$

$$\mathscr{B}_i|_L = \operatorname{image} \mathscr{H}_L(\bar{y}_i), \quad \text{for} \quad n \leq L \leq T-n+1$$

# Persistency of excitation + no common disturbance dynamics implies identifiability

## structured noise filtering setup

$$y_i = \bar{y}_i + e_i, \quad \bar{y}_i \in \mathscr{B}_i \in \mathscr{L}_n, \quad e_i \sim \mathrm{N}(0, s^2 I)$$

$$\bar{y}_i = s_i + d_i, \quad s_i \in \mathscr{B}_\mathrm{s} \in \mathscr{L}_{n_\mathrm{s}}, \quad d_i \in \mathscr{B}_{\mathrm{d,i}} \in \mathscr{L}_{n_\mathrm{d}}$$

## assuming

1. $e_1 = \cdots = e_N = 0$ (no random noise)
2. $\{ s_1, \ldots, s_N \}$ persistently exciting of order $n_\mathrm{s}$
3. $\mathscr{B}_{\mathrm{d},1}, \ldots, \mathscr{B}_{\mathrm{d},N}$ have no common poles

## we have that $\mathscr{B}_\mathrm{s} = \mathscr{B}_1 \cap \cdots \cap \mathscr{B}_N$

# Subspace method for common dynamics detection and estimation

assuming persistency of excitation

$$\mathscr{B}_i|_L = \text{image}\, \mathscr{H}_L(\bar{y}_i)$$

common dynamics $\rightsquigarrow$ subspace intersection

$$\mathscr{B}|_L = \mathscr{B}_1|_L \cap \cdots \cap \mathscr{B}_N|_L$$

with random noise:

1. preprocessing: estimate $\mathscr{B}_1|_L, \ldots, \mathscr{B}_N|_L$
2. find approximate subspace intersection

# Estimation of $\mathscr{B}_i|_L$ via low-rank approximation

## maximum likelihood estimation problem

$$\text{minimize} \quad \text{over } \widehat{y}_i \quad \|y_i - \widehat{y}_i\|$$
$$\text{subject to} \quad \text{rank } \mathscr{H}_{n+1}(\widehat{y}_i) \leq n$$

## suboptimal heuristic

$$\text{SVD truncation} \qquad \mathscr{H}_L(y_i) \approx \mathscr{O}_i \mathscr{C}_i$$

- ignore the Hankel structure
- impose the rank-$n$ structure

## Kung's method

- impose shift-invariance structure on $\mathscr{O}_i$

# Approximate subspace intersection

given: $\widehat{\mathscr{B}}_i|_L = \text{image}\, \mathscr{O}_i$, $i = 1, \ldots, N$, find:

$$\widehat{\mathscr{B}}|_L \approx \widehat{\mathscr{B}}_1|_L \cap \cdots \cap \widehat{\mathscr{B}}_N|_L$$

algorithm:

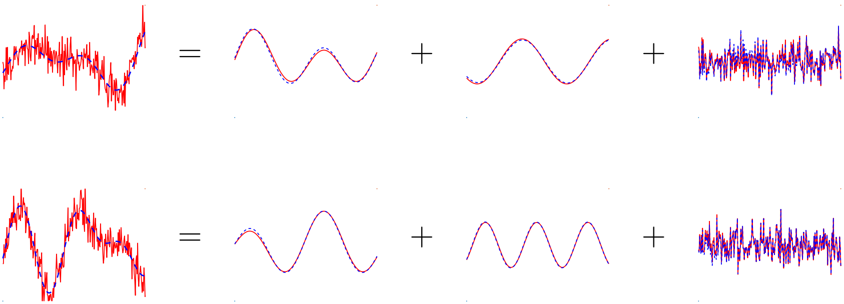
1. find $R_i \mathscr{O}_i = 0$, so that $\widehat{\mathscr{B}}_i|_L = \ker R_i$
2. define $R' = \text{col}(R_1, \ldots, R_N)$
3. find nonsingular $U$, such that $UR' \approx \left[\begin{smallmatrix} R \\ 0 \end{smallmatrix}\right]$, with $R$ f.r.r.

approximate intersection

$$\widehat{\mathscr{B}}|_L := \ker R' \approx \ker R$$

# Gives good results on the example comparing true and estimated signals

| observed signal $y$ | true signal $s$ | structured noise $d$ | random noise $e$ |
|---|---|---|---|



# Data generating model: $y_i = s_i + d_i + e_i$

$s_i \in \mathscr{B}_\mathrm{s} \in \mathscr{L}_{n_\mathrm{s}}$ — true signal

$d_i \in \mathscr{B}_{\mathrm{d},i} \in \mathscr{L}_{n_\mathrm{d}}$ — structured noise

$e_i \sim \mathrm{N}(0, s^2 I)$ — unstructured noise

$$n := n_\mathrm{s} + n_\mathrm{d}$$

# Maximum likelihood estimation problem

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{N} \|y_i - \widehat{s}_i - \widehat{d}_i\|^2 \\ \text{subject to} \quad & \widehat{s}_i \in \widehat{\mathscr{B}}_{\text{s}} \in \mathscr{L}_{n_{\text{s}}}, \quad \text{for } i = 1, \dots, N \\ & \widehat{d}_i \in \widehat{\mathscr{B}}_{\text{d},i} \in \mathscr{L}_{n_{\text{d}}}, \quad \text{for } i = 1, \dots, N \end{aligned} \tag{ML}$$

## identifiability conditions:

1. $\{ s_1, \dots, s_N \}$ — persistently exciting of order $n_{\text{s}}$
2. $\mathscr{B}_{\text{d},1}, \dots, \mathscr{B}_{\text{d},N}$ have no common poles

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

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{N} \|y_i - \widehat{s}_i - \widehat{d}_i\|^2 \\
\text{subject to} \quad & \operatorname{rank} \begin{bmatrix} \mathscr{H}_{n_\mathrm{s}+1}(\widehat{s}_1) & \cdots & \mathscr{H}_{n_\mathrm{s}+1}(\widehat{s}_N) \end{bmatrix} \leq n_\mathrm{s} \\
& \operatorname{rank} \mathscr{H}_{n_\mathrm{d}+1}(\widehat{d}_i) \leq n_\mathrm{d}, \quad \text{for } i = 1, \ldots, N
\end{aligned}$$

$N+1$ rank constraints

# Current/future work

optimization with multiple rank constraints

generalization to systems with inputs

(real-life) applications