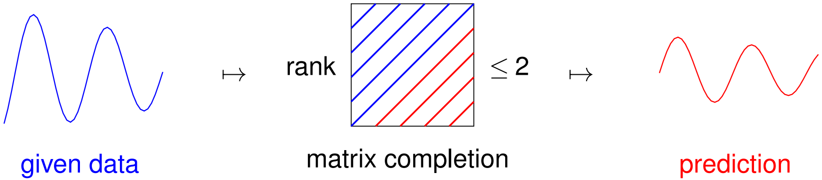
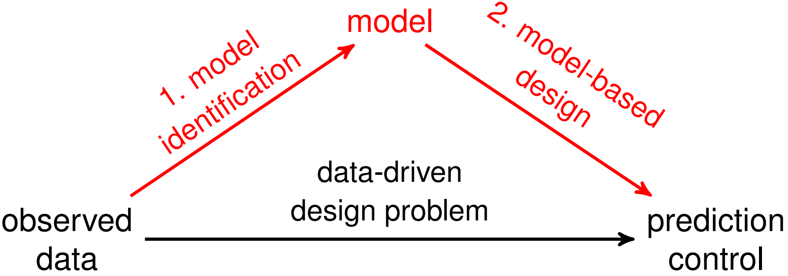
A missing data approach to data-driven filtering and control

Ivan Markovsky

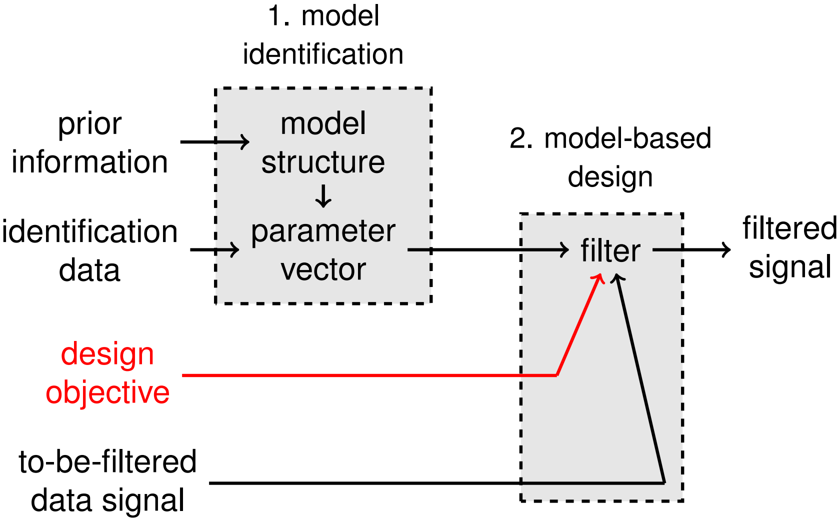
Vrije Universiteit Brussel



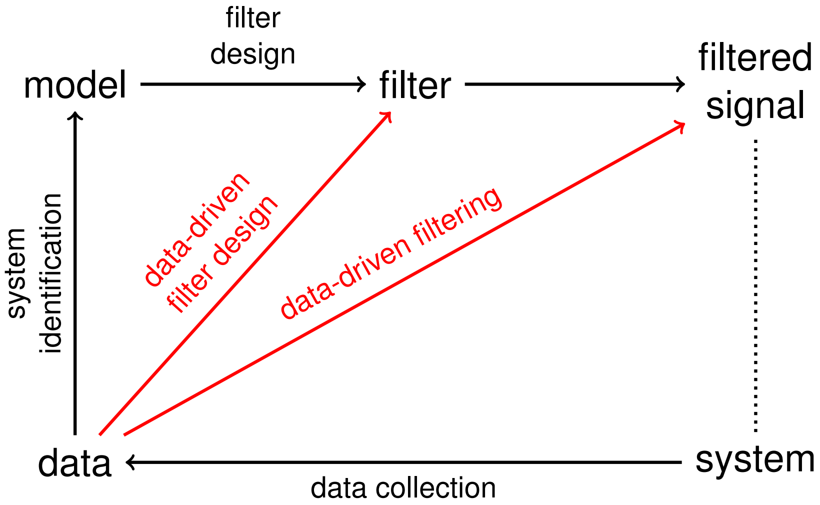
Modern filtering/control is model-based: the design problem is split into two steps



System identification does not take into account the design objective



Data-driven methods avoid modeling



Combined modeling+design has benefits

identification ignores the design objective

the two-step approach is suboptimal

objective: define and solve a direct problem

$$\begin{matrix}\text{observed}\\ \text{data}\end{matrix} + \begin{matrix}\text{filtering}\\ \text{objective}\end{matrix} \mapsto \begin{matrix}\text{filtered}\\ \text{signal}\end{matrix}$$

Plan

Example: data-driven Kalman smoothing

Generalization: missing data estimation

Solution approach: matrix completion

Plan

Example: data-driven Kalman smoothing

Generalization: missing data estimation

Solution approach: matrix completion

A dynamical system $\mathscr{B}$ is a set of signals w

$w \in \mathscr{B} \quad \iff$

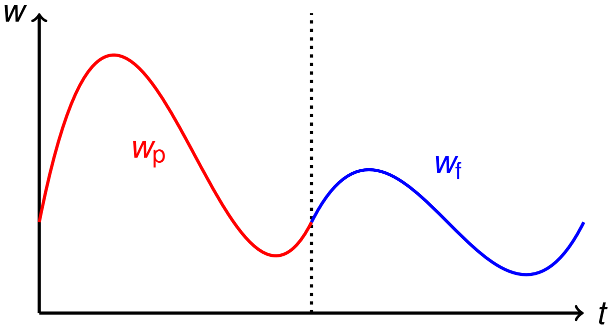
- the signal w is trajectory of the system $\mathscr{B}$
- $\mathscr{B}$ is an exact model for w
- $\mathscr{B}$ is unfalsified by w

we consider linear time-invariant systems $\quad \left(w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]\right)$

$\mathscr{L}$ — linear time-invariant model class

Initial conditions are specified by "past" traj.

$$w = w_{\mathrm{p}} \wedge w_{\mathrm{f}}$$



Representation free definition of smoothing

observer: given model $\mathscr{B}$ and exact trajectory w_f

$$\text{find } w_p, \text{ such that } w_p \wedge w_f \in \mathscr{B}$$

smoother: given model $\mathscr{B}$ and noisy trajectory w_f

$$\text{minimize} \quad \|w_f - \widehat{w}_f\| \quad \text{subject to} \quad \widehat{w}_p \wedge \widehat{w}_f \in \mathscr{B} \qquad \text{(MBS)}$$

When does a trajectory $w_d \in \mathscr{B}$ specify $\mathscr{B}$?

identifiability conditions

1. u_d is persistently exciting of "sufficiently high order"
2. $\mathscr{B}$ is controllable

how to obtain $\mathscr{B}$ back from w_d?

$w_d \mapsto \mathscr{B}$ by choosing the simplest exact model for w_d

The most powerful unfalsified model of w_d, $\mathscr{B}_{\mathrm{mpum}}(w_\mathrm{d})$ is the data generating system

complexity $\leftrightarrow$ # inputs $\mathtt{m}$ and # states $\mathtt{n}$

$$\mathrm{c}(\mathscr{B}) = (\mathtt{m}, \mathtt{n})$$

the most powerful unfalsified model

$$\mathscr{B}_{\mathrm{mpum}}(w_\mathrm{d}) := \underbrace{\arg \min_{\widehat{\mathscr{B}} \in \mathscr{L}} \mathrm{c}(\widehat{\mathscr{B}})}_{\text{most powerful}} \quad \text{subject to} \quad \underbrace{w_\mathrm{d} \in \widehat{\mathscr{B}}}_{\text{unfalsified model}}$$

$\mathscr{L}_{\mathtt{m},\mathtt{n}}$ — set of models with complexity bounded by $(\mathtt{m}, \mathtt{n})$

Data-driven smoothing replaces the model $\mathscr{B}$ by trajectory $w_\mathrm{d} \in \mathscr{B}$

observer: given trajectories w_d and w_f of $\mathscr{B}$

$$\text{find } w_\mathrm{p}, \text{ such that } w_\mathrm{p} \wedge w_\mathrm{f} \in \mathscr{B}_\mathrm{mpum}(w_\mathrm{d})$$

smoother: given noisy traj. w_d and w_f of $\mathscr{B}$ and $(\mathtt{m}, \ell)$

$$\begin{array}{ll} \text{minimize} & \underbrace{\|w_\mathrm{f} - \widehat{w}_\mathrm{f}\|_2^2}_{\text{estimation error}} + \underbrace{\|w_\mathrm{d} - \widehat{w}_\mathrm{d}\|_2^2}_{\text{identification error}} \\ \text{subject to} & \widehat{w}_\mathrm{p} \wedge \widehat{w}_\mathrm{f} \in \mathscr{B}_\mathrm{mpum}(\widehat{w}_\mathrm{d}) \in \mathscr{L}_{\mathtt{m},\ell} \end{array} \qquad \text{(DDS)}$$

Classical approach: divide and conquer

1. identification: given w_d and $(\mathtt{m}, \ell)$

$$\text{minimize} \quad \|w_\text{d} - \widehat{w}_\text{d}\| \quad \text{subject to} \quad \mathscr{B}_\text{mpum}(\widehat{w}_\text{d}) \in \mathscr{L}_{\mathtt{m},\ell}$$

2. model-based filtering: given w_f and $\widehat{\mathscr{B}} := \mathscr{B}_\text{mpum}(\widehat{w}_\text{d})$

$$\text{minimize} \quad \|w_\text{f} - \widehat{w}_\text{f}\| \quad \text{subject to} \quad \widehat{w}_\text{p} \wedge \widehat{w}_\text{f} \in \widehat{\mathscr{B}}$$

Summary

model-based smoothing

given model $\mathscr{B}$ and trajectory w_{f}

$$\text{minimize} \quad \|w_{\mathrm{f}} - \widehat{w}_{\mathrm{f}}\| \quad \text{subject to} \quad \widehat{w}_{\mathrm{p}} \wedge \widehat{w}_{\mathrm{f}} \in \mathscr{B} \qquad \text{(MBS)}$$

data-driven smoothing

given trajectories w_{d} and w_{f} and complexity $(\mathtt{m}, \ell)$

$$\begin{array}{ll} \text{minimize} & \|w_{\mathrm{f}} - \widehat{w}_{\mathrm{f}}\|_2^2 + \|w_{\mathrm{d}} - \widehat{w}_{\mathrm{d}}\|_2^2 \\ \text{subject to} & \widehat{w}_{\mathrm{p}} \wedge \widehat{w}_{\mathrm{f}} \in \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_{\mathrm{d}}) \in \mathscr{L}_{\mathtt{m},\ell} \end{array} \qquad \text{(DDS)}$$

Plan

Example: data-driven Kalman smoothing

Generalization: missing data estimation

Solution approach: matrix completion

We aim to find missing part of trajectory

missing data — interpolated from $w \in \mathscr{B}$

exact data— kept fixed

inexact / "noisy" data — approximated by $\min \|\text{error}\|_2$

Other examples fit in the same setting

? — missing, E — exact, N — noisy

$w = \Pi \begin{bmatrix} u \\ y \end{bmatrix}$, u — input, y — output

example	w_p	u_f	y_f
state estimation	?	E	E
EIV Kalman smoothing	?	N	N
classical Kalman smoothing	?	E	N
simulation	E	E	?
partial realization	E	E	E/?
noisy realization	E	E	N/?
output tracking	E	?	N

classical Kalman filter

	past	future
input	?	u
output	?	y

$$\text{minimize} \quad \|y - \widehat{y}\|$$
$$\text{subject to} \quad w_{\text{p}} \wedge (u, \widehat{y}) \in \mathscr{B}$$

output tracking control

	past	future
input	u_{p}	?
output	y_{p}	y_{ref}

$$\text{minimize} \quad \underbrace{\|y_{\text{ref}} - \widehat{y}\|}_{\text{tracking error}}$$
$$\text{subject to} \quad w_{\text{p}} \wedge (\widehat{u}, \widehat{y}) \in \mathscr{B}$$

Weighted approximation criterion accounts for exact, missing, and noisy data

error vector: $e := w - \widehat{w}$

$$\|e\|_v := \sqrt{\sum_t \sum_i \; v_i(t) e_i^2(t)}$$

weight	used for	to	by
$v_i(t) = \infty$	$w_i(t)$ exact	interpolate $w_i(t)$	$e_i(t) = 0$
$v_i(t) \in (0, \infty)$	$w_i(t)$ noisy	approx. $w_i(t)$	$\min \; \|e_i(t)\|$
$v_i(t) = 0$	$w_i(t)$ missing	fill in $w_i(t)$	$\widehat{w} \in \widehat{\mathscr{B}}$

Data-driven signal processing can be posed as missing data estimation problem

$$\begin{array}{ll} \text{minimize} & \|w_{\mathrm{d}} - \widehat{w}_{\mathrm{d}}\|_2^2 + \|w - \widehat{w}\|_v^2 \\ \text{subject to} & \widehat{w} \in \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_{\mathrm{d}}) \in \mathscr{L}_{\mathrm{m},\ell} \end{array} \qquad \text{(DD-SP)}$$

the recovered missing values of $\widehat{w}$ are the desired result

Plan

Example: data-driven Kalman smoothing

Generalization: missing data estimation

Solution approach: matrix completion

$w \in \mathscr{B} \iff$ Hankel matrix is low-rank

exact trajectory $w \in \mathscr{B} \in \mathscr{L}_{\mathrm{m},\ell}$

$$\Updownarrow$$

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_\ell w(t+\ell) = 0$$

$$\Updownarrow$$

rank deficient

$$\mathscr{H}(w) := \begin{bmatrix} w(1) & w(2) & \cdots & w(T-\ell) \\ w(2) & w(3) & \cdots & w(T-\ell+1) \\ w(3) & w(4) & \cdots & w(T-\ell+2) \\ \vdots & \vdots & & \vdots \\ w(\ell+1) & w(\ell+2) & \cdots & w(T) \end{bmatrix}$$

relation at time $t = 1$

$$R_0 w(1) + R_1 w(2) + \cdots + R_\ell w(\ell+1) = 0$$

in matrix form:

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_\ell \end{bmatrix} \begin{bmatrix} w(1) \\ w(2) \\ \vdots \\ w(\ell+1) \end{bmatrix} = 0$$

relation at time $t = 2$

$$R_0 w(2) + R_1 w(3) + \cdots + R_\ell w(\ell + 2) = 0$$

in matrix form:

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_\ell \end{bmatrix} \begin{bmatrix} w(2) \\ w(3) \\ \vdots \\ w(\ell+2) \end{bmatrix} = 0$$

relation at time $t = T - \ell$

$$R_0 w(T-\ell) + R_1 w(T-\ell+1) + \cdots + R_\ell w(T) = 0$$

in matrix form:

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_\ell \end{bmatrix} \begin{bmatrix} w(T-\ell) \\ w(T-\ell+1) \\ w(T-\ell+2) \\ \vdots \\ w(T) \end{bmatrix} = 0$$

relation for $t = 1, \ldots, T - \ell$

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_\ell w(t+\ell) = 0$$

in matrix form:

$$\underbrace{\begin{bmatrix} R_0 & R_1 & \cdots & R_\ell \end{bmatrix}}_{R} \underbrace{\begin{bmatrix} w(1) & w(2) & \cdots & w(T-\ell) \\ w(2) & w(3) & \cdots & w(T-\ell+1) \\ w(3) & w(4) & \cdots & w(T-\ell+2) \\ \vdots & \vdots & & \vdots \\ w(\ell+1) & w(\ell+2) & \cdots & w(T) \end{bmatrix}}_{\mathscr{H}(w)} = 0$$

$$w \in \mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell}$$

$$\Updownarrow$$

there is $R \in \mathbb{R}^{(q-\mathtt{m})\times q(\ell+1)}$ full row rank,

such that $\quad R\mathscr{H}(w) = 0$

$$\Updownarrow$$

$$\operatorname{rank}\big(\mathscr{H}(w)\big) \leq q\ell + \mathtt{m}$$

q — # of variables

$\widehat{w} \in \mathscr{B}_{\text{mpum}}(\widehat{w}_{\text{d}})$ is equivalent to rank constraint on a mosaic-Hankel matrix

$$\widehat{w} \in \mathscr{B}_{\text{mpum}}(\widehat{w}_{\text{d}}) \in \mathscr{L}_{\mathtt{m},\ell}$$

$$\Downarrow$$

$$\widehat{w}_{\text{d}} \in \widehat{\mathscr{B}} \in \mathscr{L}_{\mathtt{m},\ell} \quad \text{and} \quad \widehat{w} \in \widehat{\mathscr{B}}$$

$$\Updownarrow$$

$$\text{rank}\Big(\underbrace{\begin{bmatrix} \mathscr{H}(\widehat{w}_{\text{d}}) & \mathscr{H}(\widehat{w}) \end{bmatrix}}_{\mathscr{H}(\widehat{w}_{\text{d}},\widehat{w})}\Big) \leq q\ell + \mathtt{m}$$

Data-driven signal processing $\iff$ structured low-rank approximation

$$\begin{array}{ll} \text{minimize} & \|w_{\mathrm{d}} - \widehat{w}_{\mathrm{d}}\|_2^2 + \|w - \widehat{w}\|_v^2 \\ \text{subject to} & \widehat{w} \in \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_{\mathrm{d}}) \in \mathscr{L}_{\mathrm{m},\ell} \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \text{minimize} & \|w' - \widehat{w}'\|_{v'} \\ \text{subject to} & \operatorname{rank}\left(\mathscr{H}(\widehat{w}')\right) \leq r \end{array}$$

Three main classes of solution methods

local optimization

nuclear norm relaxation

subspace methods

considerations

- generality
- user defined hyper parameters
- availability of efficient algorithms/software

Local optimization using variable projections: analytical elimination of $\widehat{w}$

kernel representation

$$\min_{R \text{ f.r.r.}} \left(\min_{\widehat{w}} \|w - \widehat{w}\| \text{ subject to } R\mathscr{H}(\widehat{w}) = 0 \right)$$

variable projection (VARPRO): elimination of $\widehat{w}$ leads to

$$\text{minimize} \quad f(R) \quad \text{subject to} \quad R \text{ full row rank}$$

Dealing with the "*R* full row rank" constraint

1. impose a quadratic equality constraint $RR^\top = I$
2. using specialized methods for optimization on a manifold
3. R full row rank $\iff$ $R\Pi = \begin{bmatrix} X & I \end{bmatrix}$ with Π permutation
   - Π fixed $\rightsquigarrow$ total least-squares
   - Π can be changed during the optimization

Summary of the VARPRO approach

kernel representation $\rightsquigarrow$ parameter opt. problem

$$\min_{\widehat{w}, R \text{ f.r.r.}} \|w - \widehat{w}\| \text{ subject to } R\mathscr{H}(\widehat{w}) = 0$$

elimination of $\widehat{w}$ $\rightsquigarrow$ optimization on a manifold

$$\min_{R \text{ f.r.r.}} f(R)$$

in case of mosaic-Hankel $\mathscr{H}$, f can be evaluated fast

Numerical example with Kalman smoothing

simulation setup

- $\overline{\mathscr{B}} \in \mathscr{L}_{1,2}$ — 2nd order LTI system
- $w_\text{f} = \overline{w}_\text{f} + \text{noise}, \quad \overline{w}_\text{f} \in \mathscr{B}$ — step response
- $w_\text{d} = \overline{w}_\text{d} + \text{noise}, \quad \overline{w}_\text{d} \in \mathscr{B}$

smoothing with known model

- state space solution
- solution of (MBS)

smoothing with unknown model

- identification + model-based design
- solution of (DDS)

Known model: the missing data approach (MBS) recovers the state space solution

state space solution

$$\text{minimize} \quad \left\| \begin{bmatrix} u_\text{f} \\ y_\text{f} \end{bmatrix} - \begin{bmatrix} 0 & I \\ \mathscr{O}_T(A,C) & \mathscr{T}_T(H) \end{bmatrix} \begin{bmatrix} \widehat{x}_\text{ini} \\ \widehat{u}_\text{f} \end{bmatrix} \right\| \qquad \text{(SSS)}$$

representation free solution

(MBS) is a generalized least squares

approximation error $\quad e := (\|\overline{w}_\text{f} - \widehat{w}_\text{f}\|)/\|\overline{w}_\text{f}\|$

method	(MBS)	(SSS)
error e	0.083653	0.083653

Unknown model: (DDS) gives better results than the model-based approach

classical approach

identification + (SSS)

data-driven approach

solution of (DDS) with local optimization

simulation result

method	(MBS)	(DDS)	classical
error e	0.083653	0.087705	0.091948

Conclusion

motivation: combine the modeling and design problems

we aim to find the missing part of a trajectory $w \in \mathscr{B}$

reformulation as weighted structured low-rank approx.

Future work

statistical analysis

computational efficiency / recursive computation

other methods: subspace, convex relaxation, ...