Exact and approximate linear system identification

Ivan Markovsky

University of Southampton

Outline

- Exact identification: the most powerful unfalsified model
- Identifiability conditions and algorithms
- From exact to approximate identification: misfit vs latency
- Misfit computation and minimization

Exact identification: $w_{\mathrm{d}} \mapsto \mathscr{B}$, $w_{\mathrm{d}} \in \mathscr{B}$

An exact identification problem

Problem P1 (Exact identification)

Given two vector time series

$$u_{\mathrm{d}} = \big(u_{\mathrm{d}}(1),\dots,u_{\mathrm{d}}(T)\big) \in (\mathbb{R}^{\mathtt{m}})^T \qquad \text{“inputs”}$$
$$y_{\mathrm{d}} = \big(y_{\mathrm{d}}(1),\dots,y_{\mathrm{d}}(T)\big) \in (\mathbb{R}^{\mathtt{p}})^T \qquad \text{“outputs”}$$

find $\mathtt{n} \in \mathbb{N}$ and LTI system $\mathscr{B}$ of order n, with m inputs and p outputs, s.t.

$$w_{\mathrm{d}} := (u_{\mathrm{d}}, y_{\mathrm{d}}) \in \mathscr{B},$$

i.e., w_{d} is a trajectory of $\mathscr{B}$.

How can we check that “$w_{\mathrm{d}} \in \mathscr{B}$”?

Checking that $w_d \in \mathscr{B}$

Let $\mathscr{B}$ be defined by a minimal input/state/output representation

$$\mathscr{B} = \mathscr{B}_{\text{i/s/o}}(A,B,C,D) := \{\, (u,y) \mid \sigma x = Ax + Bu,\ y = Cx + Du \,\}$$

$(u_d, y_d) \in \mathscr{B}_{\text{i/s/o}}(A,B,C,D) \iff$ there exists $x_{\text{ini}} \in \mathbb{R}^{n}$, such that

$$y_d = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{bmatrix}}_{\mathscr{O}_T(A,C)} x_{\text{ini}} + \begin{bmatrix} D & & & & \\ CB & D & & & \\ CAB & CB & D & & \\ \vdots & \ddots & \ddots & \ddots & \\ CA^{T-1}B & \cdots & CAB & CB & D \end{bmatrix} u_d$$

(y_d is the response of $\mathscr{B}$ under input u_d and initial condition x_{ini})

Comments

- P1 is an exact fitting problem, a most basic SYSID problem
- easily generalizable to a set of N time series
  $u_{d,1}, \dots, u_{d,N} \in (\mathbb{R}^{\mathtt{m}})^T$ and $y_{d,1}, \dots, y_{d,N} \in (\mathbb{R}^{\mathtt{p}})^T$
- the realization problem

  $$\text{impulse response} \quad \mapsto \quad (A,B,C,D)$$

  is a special case of P1 for a set of $\mathtt{m}$ time series
- while $\mathtt{m}$ is given, finding $\mathtt{n}$ is part of the problem
  any observable system of order $\mathtt{n} \geq \mathtt{p}T$ is a (trivial) solution
- we are actually interested is a solution of a minimal order

Revised exact identification problem

Problem P1' (Exact identification)

Given two vector time series

$$u_d = \big(u_d(1), \dots, u_d(T)\big) \in (\mathbb{R}^{\mathtt{m}})^T \qquad \text{"inputs"}$$
$$y_d = \big(y_d(1), \dots, y_d(T)\big) \in (\mathbb{R}^{\mathtt{p}})^T \qquad \text{"outputs"}$$

find the smallest $\mathtt{n} \in \mathbb{N}$ and LTI system $\mathscr{B}$ of order $\mathtt{n}$, with $\mathtt{m}$ inputs and $\mathtt{p}$ outputs, such that

$$w_d = (u_d, y_d) \in \mathscr{B}.$$

Set of LTI systems with a bounded complexity

Notation: $\mathscr{L}^{\mathtt{w},\mathtt{n}}_{\mathtt{m},\ell}$ is the set of all LTI systems with

- $\mathtt{w}$ (external) variables
- at most $\mathtt{m}$ inputs
- minimal state dimension at most $\mathtt{n}$ and
- lag (= observability index) at most ℓ

For $t \geq \mathtt{n}$, the set $\mathscr{B}|_t$ of all t samples long traj. of $\mathscr{B}$ has dimension

$$\dim(\mathscr{B}|_t) \;\leq\; t\mathtt{m} + \mathtt{n} \;\leq\; t\mathtt{m} + \mathtt{p}\ell$$

(where $\mathtt{p}(\ell - 1) \leq \mathtt{n} \leq \mathtt{p}\ell$)

$\implies$ $(\mathtt{m},\mathtt{n})$ and $(\mathtt{m},\ell)$ specify the complexity of the model class $\mathscr{L}^{\mathtt{w},\mathtt{n}}_{\mathtt{m},\ell}$

Another exact identification problem

Problem P2 (Exact identification)

Given a vector time series

$$w_d = \big(w_d(1), \dots, w_d(T)\big) \in (\mathbb{R}^{\mathtt{w}})^T$$

find the smallest $\mathtt{m} \in \mathbb{N}$ and $\ell \in \mathbb{N}$ and LTI system $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell}$, s.t. $w_d \in \mathscr{B}$.

Comments:

- no separation between inputs and outputs
- the complexity is defined by $(\mathtt{m}, \ell)$

Most powerful unfalsified model

The most powerful unfalsified model in the model class $\mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell}$ of a time series $w_d \in (\mathbb{R}^{\mathtt{w}})^T$ is the system $\mathscr{B}_{\text{mpum}}$ that is

1. in the model class, *i.e.*, $\mathscr{B}_{\text{mpum}} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell}$,
2. unfalsified, *i.e.*, $w_d \in \mathscr{B}_{\text{mpum}}|_T$, and
3. most powerful among all LTI unfalsified systems, *i.e.*,

$$\mathscr{B}' \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell} \text{ and } w_d \in \mathscr{B}'|_T \quad \Longrightarrow \quad \mathscr{B}_{\text{mpum}}|_T \subseteq \mathscr{B}'|_T.$$

MPUM may not exist, but if it does, then it is unique

Identifiability

Identifiability question

P2 is the problem of computing the MPUM of w_d in $\mathscr{L}^{\mathtt{w}}$

The following related question is of interest:

Suppose that

$$w_d \in \overline{\mathscr{B}} \in \mathscr{L}^{\mathtt{w}}$$

and upper bounds $\mathtt{n}_{\max}$, $\ell_{\max}$ of the order $\mathtt{n}$ and lag ℓ of $\overline{\mathscr{B}}$ are given.

$$\boxed{\text{Under what conditions } \mathscr{B}_{\text{mpum}}(w_d) \text{ is equal to the system } \overline{\mathscr{B}}?}$$

the answer is given by the following lemma

Fundamental Lemma

Let $\bar{\mathscr{B}} \in \mathscr{L}^{\mathrm{w},\mathrm{n}}_{\mathrm{m}}$ be controllable and let $w_\mathrm{d} := (u_\mathrm{d}, y_\mathrm{d}) \in \bar{\mathscr{B}}|_T$.

Then, if u_d is persistently exciting of order $L + \mathrm{n}$,

$$\text{image}\left(\begin{bmatrix} w_\mathrm{d}(1) & w_\mathrm{d}(2) & w_\mathrm{d}(3) & \cdots & w_\mathrm{d}(T-L+1) \\ w_\mathrm{d}(2) & w_\mathrm{d}(3) & w_\mathrm{d}(4) & \cdots & w_\mathrm{d}(T-L+2) \\ w_\mathrm{d}(3) & w_\mathrm{d}(4) & w_\mathrm{d}(5) & \cdots & w_\mathrm{d}(T-L+3) \\ \vdots & \vdots & \vdots & & \vdots \\ w_\mathrm{d}(L) & w_\mathrm{d}(L+1) & w_\mathrm{d}(L+2) & \cdots & w_\mathrm{d}(T) \end{bmatrix}\right) = \bar{\mathscr{B}}|_L$$

$\Longrightarrow$ under the conditions of the FL, any L samples long response y of $\mathscr{B}$ can be obtained as $y = \mathscr{H}_L(y_\mathrm{d})g$, for certain g $\rightsquigarrow$ algorithms

$\Longrightarrow$ with $L = \ell_{\max} + 1$, the FL gives conditions for identifiability

Persistency of excitation

$u_\mathrm{d} = \big(u_\mathrm{d}(1), \dots, u_\mathrm{d}(T)\big)$ is persistently exciting of order L if

$$\mathscr{H}_L(u_\mathrm{d}) := \begin{bmatrix} u_\mathrm{d}(1) & u_\mathrm{d}(2) & u_\mathrm{d}(3) & \cdots & u_\mathrm{d}(T-L+1) \\ u_\mathrm{d}(2) & u_\mathrm{d}(3) & u_\mathrm{d}(4) & \cdots & u_\mathrm{d}(T-L+2) \\ u_\mathrm{d}(3) & u_\mathrm{d}(4) & u_\mathrm{d}(5) & \cdots & u_\mathrm{d}(T-L+3) \\ \vdots & \vdots & \vdots & & \vdots \\ u_\mathrm{d}(L) & u_\mathrm{d}(L+1) & u_\mathrm{d}(L+2) & \cdots & u_\mathrm{d}(T) \end{bmatrix} \quad \text{is full row rank}$$

System theoretic interpretation:

u_d is persistently exciting of order L $\iff$ there is no LTI system with # of inputs $< \mathrm{m}$ and lag $< L$ for which u_d is a trajectory

Overview of algorithms

1. $w_\mathrm{d} \mapsto R(\xi)$
2. $w_\mathrm{d} \mapsto$ impulse response H
3. $w_\mathrm{d} \mapsto (A, B, C, D)$ (possibly balanced)

   3.1 $w_\mathrm{d} \mapsto R(\xi) \mapsto (A, B, C, D)$ or $w_\mathrm{d} \mapsto H \mapsto (A, B, C, D)$

   3.2 $w_\mathrm{d} \mapsto \mathscr{O}_{\ell_{\max}+1}(A, C) \mapsto (A, B, C, D)$

   3.3 $w_\mathrm{d} \mapsto \big(x_\mathrm{d}(1), \dots, x_\mathrm{d}(\mathrm{n}_{\max} + \mathrm{m} + 1)\big) \mapsto (A, B, C, D)$

MATLAB toolbox:

```
ftp.esat.kuleuven.be/pub/SISTA/markovsky/
                                abstracts/05-122.html
```


References

1. J. C. Willems.
   From time series to linear system—Part II. Exact modelling.
   Automatica, 22(6):675–694, 1986.
2. J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor.
   A note on persistency of excitation.
   Systems & Control Letters, 54(4):325–329, 2005.
3. I. Markovsky, J. C. Willems, P. Rapisarda, and B. De Moor.
   Algorithms for deterministic balanced subspace identification.
   Automatica, 41(5):755–766, 2005.
4. I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor.
   Exact and Approximate Modeling of Linear Systems
   SIAM, 2006

From exact to approximate identification

Identifiability

Provided that $w_d \in \overline{\mathscr{B}} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$, find conditions under which

$$\mathscr{B}_{\text{mpum}}(w_d) = \overline{\mathscr{B}}.$$

Main theoretical result in the exact identification setting:

$\overline{\mathscr{B}}$ is identifiable from $w_d = (u_d, y_d)$ in $\mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$ if

1. w_d is exact, *i.e.*, $w_d \in \overline{\mathscr{B}}$,
2. the model class is correct, *i.e.*, $\overline{\mathscr{B}} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$,
3. $\overline{\mathscr{B}}$ is controllable, and
4. u_d is persistently exciting of order $\ell_{\max} + \mathtt{n}$

Notes

- the conditions are only sufficient
- conditions 1, 2, and 3 are not verifiable from the given data and should therefore be postulated
- a known input/output partitioning of the data is assumed
- from a practical point of view, conditions 1 and 2 are strong and limit the applicability of the exact SYSID methods

approaches for relaxing condition 1 are described next

MPUM in the case of "noisy" data

Q: What is the MPUM of a noisy trajectory

$$w_d = \overline{w} + \widetilde{w} \qquad \text{where} \quad \overline{w} \in \overline{\mathscr{B}} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}} \quad \text{and} \quad \widetilde{w} \text{ is random and zero mean?} \tag{EIV}$$

A: With probability 1,

$$\mathscr{B}_{\text{mpum}}(w_d) = (\mathbb{R}^{\mathtt{w}})^{\mathbb{Z}_+} \qquad \text{(all variables are inputs)}$$

This is a trivial model because it fits every trajectory.

Alternatively, $\mathscr{B}_{\text{mpum}}(w_d)$ does not exist in a model class $\mathscr{M} = \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$ of bounded ($\mathtt{m} < \mathtt{w}$, $\ell_{\max} \ll T$) complexity.

In what follows we assume a given bounded complexity model class $\mathscr{M} = \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$, so we refer to lack of existence rather than trivial MPUM.

In practice, w_{d} is often generated by a

nonlinear, infinite dimensional, time-varying system $\overline{\mathscr{B}}$

possibly with process and measurement noises, *i.e.*, $\overline{\mathscr{B}} \notin \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$

$\implies$ even without noise, identifying exact model is often not possible

It can be argued that in practice the approximation aspect ($\widehat{\mathscr{B}} \approx \overline{\mathscr{B}}$) is often more important than the stochastic estimation ($\widehat{\mathscr{B}} \to \overline{\mathscr{B}}$ as $T \to \infty$)

An approximate $\widehat{\mathscr{B}} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$ is what is anyway needed:

Many prediction and control methods are based on LTI models

$\implies$ even if it was possible to identify $\overline{\mathscr{B}}$, it would be necessary to approximate it by $\widehat{\mathscr{B}} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$

Misfit vs latency

Modifications of the MPUM concept

Unless the model class $\mathscr{M} = \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$ is enlarged, *i.e.*, $(\mathtt{m}, \ell_{\max})$ is increased until the MPUM exists in $\mathscr{M}$,

we have to accept falsified models in $\mathscr{M}$
$\rightsquigarrow$ approximate SYSID

The main question in approximate SYSID is:

Q: Which (falsified) model in $\mathscr{M}$ to choose?

A: In some sense the “least falsified” (“approximately unfalsified”) one.

Two major notions of “least falsified” are small misfit and small latency.

They quantify the discrepancy between the model and the data.

Misfit approach for approximate SYSID

Consider given data w_{d} and model class $\mathscr{M} = \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$.

If the MPUM does not exists in $\mathscr{M}$, *i.e.*,

$$\mathscr{B}_{\text{mpum}}(w_{\text{d}}) \notin \mathscr{M}$$

we aim to find an approximate model $\widehat{\mathscr{B}}$ for w_{d} in $\mathscr{M}$.

The misfit approach modifies w_{d} as little as possible, so that the modified data, say $\widehat{w}$, has MPUM in $\mathscr{M}$, *i.e.*,

$$\mathscr{B}_{\text{mpum}}(\widehat{w}) \in \mathscr{M}$$

The approximate model for w_{d} in $\mathscr{M}$ is defined as $\widehat{\mathscr{B}}_{\text{misfit}} := \mathscr{B}_{\text{mpum}}(\widehat{w})$.

The modification of the data is measured by the misfit $\|w_{\text{d}} - \widehat{w}\|$

Latency

The latency approach augments w_d by, as small as possible, variable e, so that the augmented data $w_\mathrm{ext} := \mathrm{col}(e, w_\mathrm{d})$ has MPUM in the augmented model class, *i.e.*,

$$\mathscr{B}_\mathrm{mpum}\big(\mathrm{col}(e, w_\mathrm{d})\big) \in \mathscr{M}_\mathrm{ext} := \mathscr{L}^{\mathtt{w}+\mathtt{e}}_{\mathtt{m}+\mathtt{e},\ell_\mathrm{max}}$$

Let Π_w be the projection of $\mathrm{col}(e, w)$ on w. The approximate model for w_d in $\mathscr{M}$ is defined as

$$\widehat{\mathscr{B}}_\mathrm{latency} := \Pi_w\, \mathscr{B}_\mathrm{mpum}\big(\mathrm{col}(e, w_\mathrm{d})\big)$$

The size of e is measured by the latency $\|e\|$

Notes

- Both the misfit and latency approaches reduce the approximate SYSID problem to (different) exact SYSID problems:
  - $\widehat{\mathscr{B}}_\mathrm{misfit}$ is exact for the modified data $\widehat{w}$
  - $\widehat{\mathscr{B}}_\mathrm{latency}$ is obtained from an exact model for $\mathrm{col}(e, w_\mathrm{d})$
- If $\mathscr{B}_\mathrm{mpum}(w_\mathrm{d}) \in \mathscr{M}$ (the data is exact),

  $$\widehat{\mathscr{B}}_\mathrm{misfit} = \widehat{\mathscr{B}}_\mathrm{latency} = \mathscr{B}_\mathrm{mpum}(w_\mathrm{d}) \qquad (\widehat{w} = w_\mathrm{d} \text{ and } e = 0)$$

  So, misfit and latency are indeed extensions of the MPUM.
- The misfit approach modifies w_d but does not change $\mathscr{M}$.
- The latency approach modifies $\mathscr{M}$ but does not change w_d.

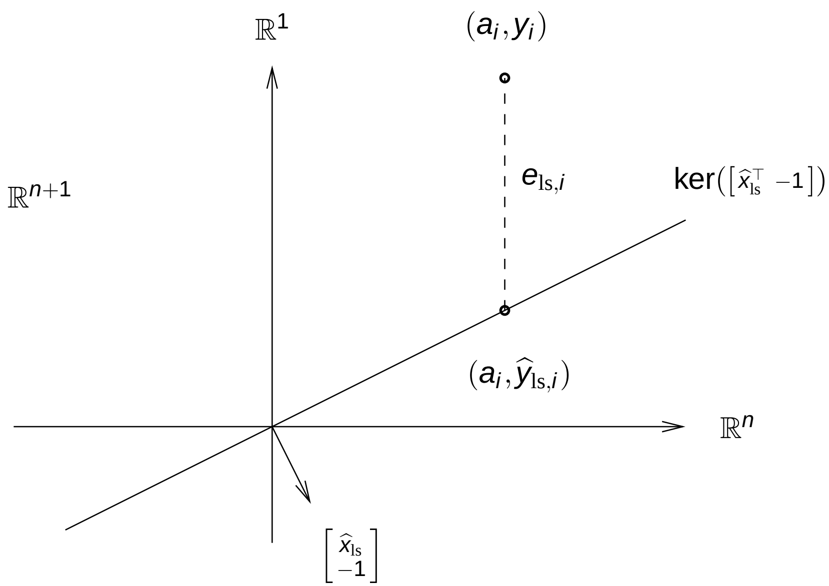
Static case: latency $\leftrightarrow$ LS misfit $\leftrightarrow$ TLS

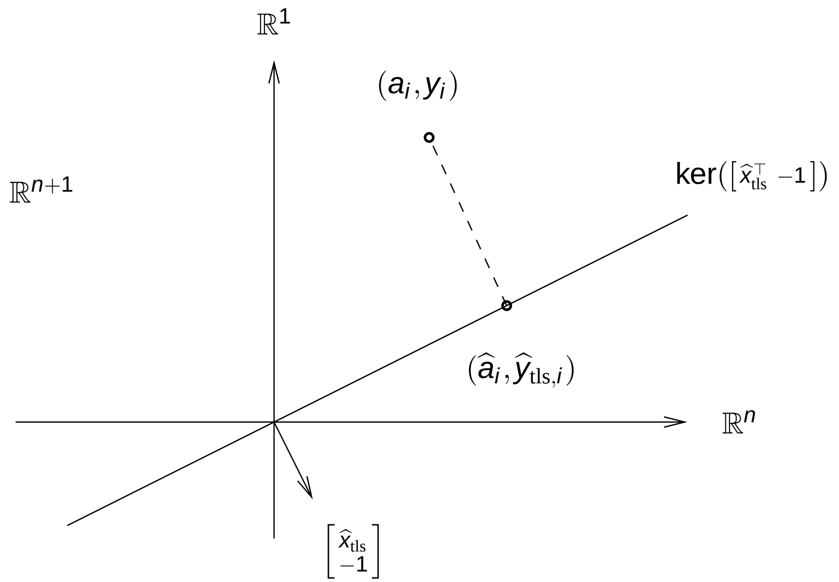
LS: $\quad \text{minimize}_{e,x}\ \|e\|_2 \quad \text{subject to} \quad Ax = y + e$ $\qquad$ e latent variable

$$\mathrm{latency}\big((A,y),x\big) := \Big(\min_e \|e\|_2^2 \text{ s.t. } Ax = y + e\Big) = \|Ax - y\|_2^2$$

TLS: $\quad \text{minimize}_{\Delta A, \Delta y, x}\ \big\|\begin{bmatrix}\Delta A & \Delta y\end{bmatrix}\big\|_\mathrm{F}$ subject to $(A + \Delta A)x = y + \Delta y$ $\qquad$ ΔA, Δy data corrections

$$\begin{aligned}\mathrm{misfit}\big((A,b),x\big) &:= \min_{\Delta A, \Delta b} \big\|\begin{bmatrix}\Delta A & \Delta b\end{bmatrix}\big\|_\mathrm{F} \text{ s.t. } (A+\Delta A)x = b + \Delta b \\ &= \frac{\|Ax - b\|_2}{\sqrt{1 + \|x\|_2^2}}\end{aligned}$$

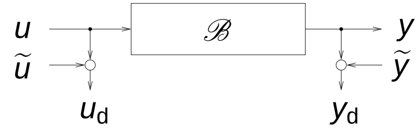




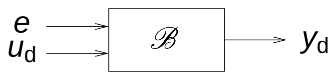
Statistical interpretation of misfit and latency

misfit	$\leftrightarrow$	errors-in-variables (EIV) model
latency	$\leftrightarrow$	ARMAX model

EIV model: $\widetilde{w} = (\widetilde{u}, \widetilde{y})$ — measurement errors



ARMAX model: e — process noise



Assumptions: $\widetilde{w}$, e — zero mean, stationary, white, ergodic, Gaussian

Misfit minimization

Identification problems

Misfit minimization (GTLS): given $w_{\text{d}} \in (\mathbb{R}^{\mathtt{w}})^T$ and $\ell_{\max} \in \mathbb{N}$, find

$$\widehat{\mathscr{B}}^*_{\text{gtls}} := \arg\min_{\widehat{\mathscr{B}}, \widehat{w}} \; \|w_{\text{d}} - \widehat{w}\| \quad \text{subject to} \quad \widehat{w} \in \widehat{\mathscr{B}} \in \mathscr{L}_{\mathtt{m},\ell_{\max}}$$

Latency minimization (PEM): given $w_{\text{d}} \in (\mathbb{R}^{\mathtt{w}})^T$ and $\ell_{\max} \in \mathbb{N}$, find

$$\widehat{\mathscr{B}}^*_{\text{pem}} := \arg\min_{\widehat{\mathscr{B}}_{\text{ext}}, e} \; \|e\| \quad \text{subject to} \quad (e, \widehat{w}) \in \widehat{\mathscr{B}}_{\text{ext}} \in \mathscr{L}_{\mathtt{m}+\mathtt{e},\ell_{\max}}$$

Notes:

- nonconvex optimization problems
- solution methods based on local optimization
- initial approximation obtained from subspace methods

Misfit minimization

Define the misfit between w_d and $\mathscr{B}$ as follows

$$\text{misfit}(w_\text{d},\mathscr{B}) := \text{minimize}_{\widehat{w}} \quad \|w_\text{d} - \widehat{w}\|_{\ell_2} \quad \text{subject to} \quad \widehat{w} \in \mathscr{B}$$

the minimizer $\widehat{w}^*$ is projection of w_d on $\mathscr{B}$ (best ℓ_2 approx. of w_d in $\mathscr{B}$)

alternatively, $\widehat{w}^*$ is the smoothed estimate of w_d, given $\mathscr{B}$

our goal is to find the model $\widehat{\mathscr{B}}$ that minimizes $\text{misfit}(w_\text{d},\mathscr{B})$, *i.e.*,

$$\widehat{\mathscr{B}} := \arg\min_{\widehat{w}} \ \text{misfit}(w_\text{d},\mathscr{B}) \quad \text{subject to} \quad \mathscr{B} \in \mathscr{M}$$

a double minimization problem: inner minimization is projection on a subspace (easy), outer minimization is a nonconvex problem (difficult)

Maximum likelihood estimator in the EIV setup

Assuming that the data is generated according to the model

$$w_\text{d} = \overline{w} + \widetilde{w}, \qquad \text{where} \quad \overline{w} \in \overline{\mathscr{B}} \in \mathscr{M} \quad \text{and} \quad \widetilde{w} \sim \text{N}(0, s^2 I)$$

$\widehat{\mathscr{B}}$ is the maximum likelihood estimator of the true model $\overline{\mathscr{B}}$

$\widehat{\mathscr{B}}$ is a consistent estimator of the true model $\overline{\mathscr{B}}$, *i.e.*, $\widehat{\mathscr{B}} \to \overline{\mathscr{B}}$ as $T \to \infty$

The log-likelihood function is ("const" does not depend on $\widehat{w}$ and $\widehat{\mathscr{B}}$)

$$L(\widehat{\mathscr{B}},\widehat{w}) = \begin{cases} \text{const} - \frac{1}{2s^2}\|w_\text{d} - \widehat{w}\|_{\ell_2}^2, & \text{if } \widehat{w} \in \widehat{\mathscr{B}} \in \mathscr{M} \\ -\infty, & \text{otherwise}, \end{cases}$$

likelihood evaluation $\iff$ misfit computation

Computation of the misfit

Given w_d and $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\ell_{\max}}$, find $\text{misfit}(w_\text{d},\mathscr{B}) := \min_{\widehat{w} \in \mathscr{B}} \|w_\text{d} - \widehat{w}\|_{\ell_2}$

$$\begin{array}{lcl} \mathscr{B} \text{ is subspace} & \Longrightarrow & \text{the constraint is linear} \\ & \Longrightarrow & \text{ordinary LS problem} \end{array}$$

Using general purpose LS solvers, the comput. complexity is $O(T^3)$.

Time-invariance of $\mathscr{B}$, however, implies Toeplitz structure of the LS prob. In addition, $\ell_{\max} \ll T$, implies banded structure with bandwidth $\ell_{\max}$.

Structure exploiting misfit computation methods have complexity $O(T)$.

They are based on:

1. structured matrix computations (*e.g.*, generalized Schur alg.)
2. Riccati recursions (Kalman smoother)

Misfit computation using I/S/O representation

$$\text{minimize}_{\widehat{w}} \quad \|w_\text{d} - \widehat{w}\|_{\ell_2} \quad \text{subject to} \quad \widehat{w} \in \mathscr{B} := \mathscr{B}_{\text{i/s/o}}(A,B,C,D) \quad \text{(SS)}$$

Recall from Lecture 5 that

$w = (u,y) \in \mathscr{B}_{\text{i/s/o}}(A,B,C,D) \iff$ there exists $x_{\text{ini}} \in \mathbb{R}^{\mathtt{n}}$, such that

$$y = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{bmatrix}}_{\mathscr{O}} x_{\text{ini}} + \underbrace{\begin{bmatrix} H(0) & & & & \\ H(1) & H(0) & & & \\ H(2) & H(1) & H(0) & & \\ \vdots & \ddots & \ddots & \ddots & \\ H(T-1) & \cdots & H(2) & H(1) & H(0) \end{bmatrix}}_{\mathscr{T}_H} u$$

where $H(0) = D$ and $H(t) = CA^{t-1}B$, for $t = 1, 2, \ldots$

Then (SS) is equivalent to the ordinary LS problem:

$$\text{minimize}_{x_{\text{ini}},\widehat{u}} \quad \left\| \begin{bmatrix} u_{\text{d}} \\ y_{\text{d}} \end{bmatrix} - \begin{bmatrix} 0 & I \\ \mathscr{O} & \mathscr{T}_H \end{bmatrix} \begin{bmatrix} x_{\text{ini}} \\ \widehat{u} \end{bmatrix} \right\| \qquad \text{(SS')}$$

Efficient solution via Riccati recursion $\rightsquigarrow$ Kalman smoother

Theorem The solution of (SS) with $D = 0$ and given x_{ini} is

$$\widehat{u}(t) = -\big(B^\top P(t+1)B + I\big)^{-1}\big(B^\top P(t+1)A\widehat{x}(t) + B^\top s(t+1) - u_{\text{d}}(t)\big)$$

where $\widehat{x}$ is given by the forward recursion

$$\widehat{x}(t+1) = A\widehat{x}(t) + B\widehat{u}(t), \qquad \widehat{x}(0) = x_{\text{ini}}$$
$$\widehat{y}(t) = C\widehat{x}(t)$$

and P, s are given by the backward recursion

$$P(t) = -A^\top P(t+1)B\big(B^\top P(t+1)B + I\big)^{-1}\times B^\top P(t+1)A + A^\top P(t+1)A + C^\top C, \qquad P(T) = 0$$

$$s(t) = -A^\top P(t+1)B\big(B^\top P(t+1)B + I\big)^{-1}\times \big(B^\top s(t+1) - u_{\text{d}}(t)\big) + A^\top s(t+1) - C^\top y_{\text{d}}(t), \qquad s(T) = 0$$

The proof uses dynamic programming

I. Markovsky and B. De Moor, Linear dynamic filtering with noisy input and output, Automatica, *41(1):167–171, 2005*

Complete solution in the continuous-time case

I. Markovsky and J. C. Willems and B. De Moor, Continuous-time errors-in-variables filtering CDC 2002, pages 2576–2581

A Matlab toolbox for misfit identification:

`ftp.esat.kuleuven.be/pub/SISTA/markovsky/abstracts/04-221a.html`

References

1. B. Roorda and C. Heij,
   Global total least squares modeling of multivariate time series,
   IEEE-AC, 40(1):50–63, 1995
2. I. Markovsky et al.,
   Application of structured total least squares for system identification and model reduction,
   IEEE-AC, 50(10):1490–1500, 2005
3. P. Lemmerling and B. De Moor,
   Misfit versus latency,
   Automatica, 37:2057–2067, 2001.
4. I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor.
   Exact and Approximate Modeling of Linear Systems
   SIAM, 2006