DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

AN OVERVIEW OF SUBSPACE IDENTIFICATION

Alessandro Chiuso

University of Padova
chiuso@dei.unipd.it
www.dei.unipd.it/~chiuso

Stockholm - May 11th, 2007

Outline of the Talk

- Identification
  - Statement of the problem
  - Closed loop identification
- Subspace identification and stochastic realization
  - Motivations
  - Main ideas
- Main (classical) Algorithms (CCA,N4SID,MOESP)
- Closed Loop Subspace Identification (New Alg.)
- Recent contributions
  - Statistical Analysis (Asymptotic Variance)
  - Comparison of algorithms
  - Role of the Future Horizon
- Simulation results
- Conclusions
  - Open questions

What I Shall not discuss…

- Identification of cointegrated process *[Bauer-Wagner]*
- ARMA processes with conditionally heteroskedastic innovation

  (e.g. (G)ARCH processes) *[Bauer]*
- Order Estimation *[Sorelius, Peternell et al., Bauer, Camba-Mendez et al…..]*

Outline of the Talk

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

- **Identification**
  - Statement of the problem
  - Closed loop identification
- Subspace identification and stochastic realization
  - Motivations
  - Main ideas
- Main (classical) Algorithms (CCA,N4SID,MOESP)
- Closed Loop Subspace Identification (New Alg.)
- Recent contributions
  - Statistical Analysis (Asymptotic Variance)
  - Comparison of algorithms
  - Role of the Future Horizon
- Simulation results
- Conclusions
  - Open questions

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Identification of LTI systems

Given observed data $\{y_1,..,y_T,u_1,..,u_T\}$, $y_t \in \mathbb{R}^{p_y}$, $u_t \in \mathbb{R}^{m_u}$, construct (approximate) model

Model order will always be n

$$\begin{cases} \mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{e}(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{e}(t) \end{cases} \quad \textbf{(LTI)}$$

which "describes" the data ($\mathbf{e}(t)$ is "white noise")

"Innovation" = one step ahead prediction error

If you prefer transfer functions

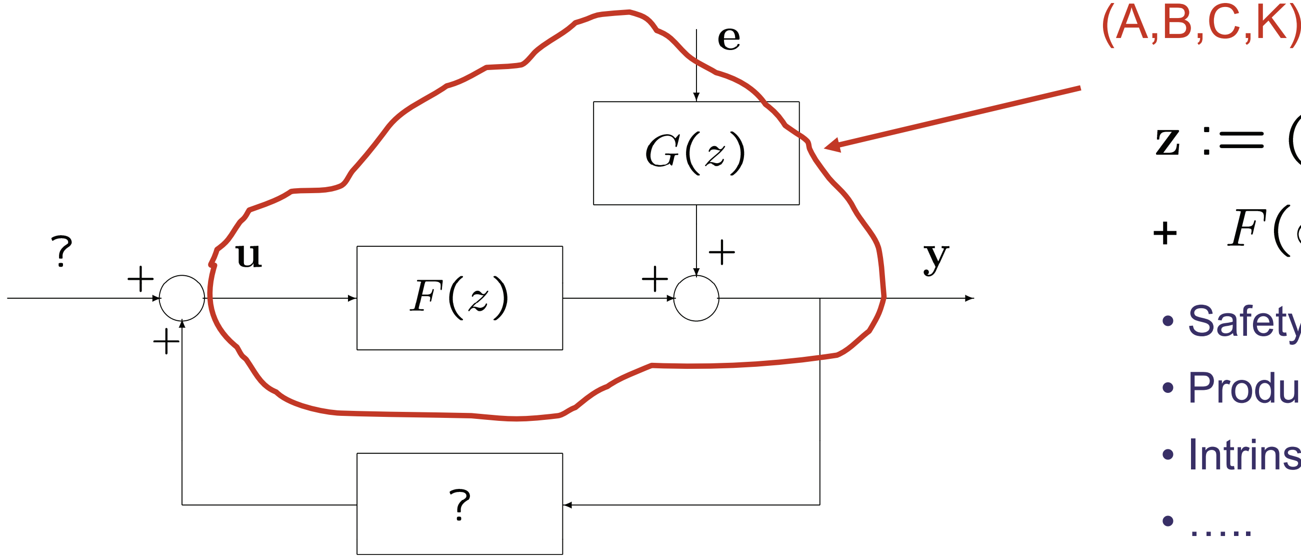
$$\mathbf{y}(t) = F(z)\mathbf{u}(t) + G(z)\mathbf{e}(t)$$

where

$$F(z) := C(zI-A)^{-1}B \qquad G(z) := C(zI-A)^{-1}K + I$$

"deterministic" "stochastic"

Identification (..closed loop)



$$\mathbf{z} := (\mathbf{y}^\top, \mathbf{u}^\top)^\top .$$

+ $F(\infty) = 0.$

- Safety,
- Production quality
- Intrinsic feedback
- .....

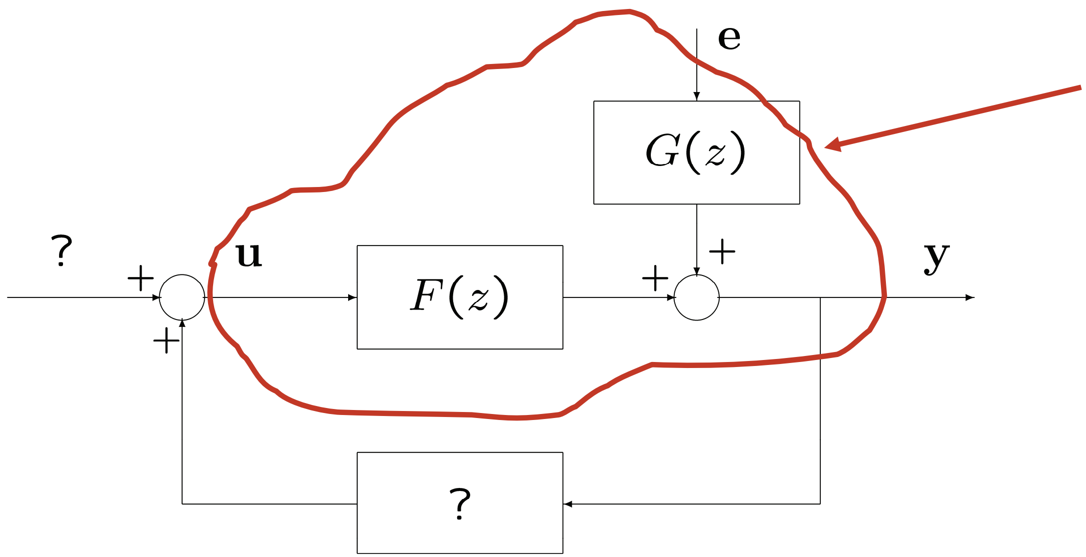
- Sometimes we shall need

$$mI \leq \Phi_{\mathbf{z}}(z) \leq MI \quad , \quad 0 < m \leq M < \infty$$

Joint Spectrum

Identification (..closed loop)

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA



(A,B,C,K)

$\mathbf{z} := (\mathbf{y}^\top, \mathbf{u}^\top)^\top.$

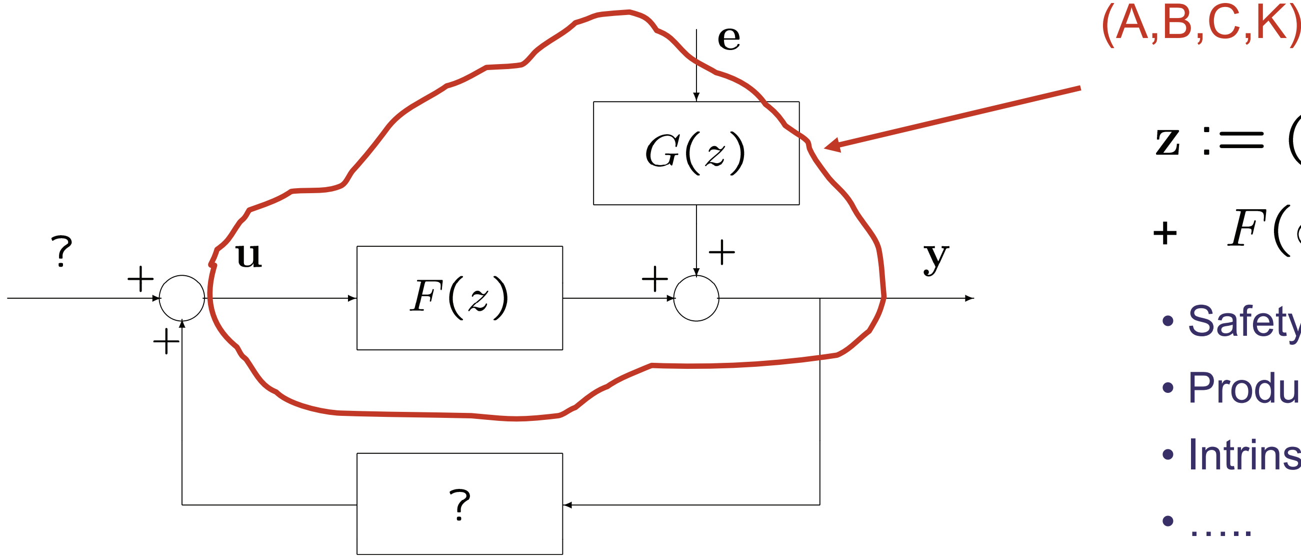
+ $F(\infty) = 0.$

- Safety,
- Production quality
- Intrinsic feedback
- .....

- Estimate Parameters (A,B,C,K) [Point estimators in statistics]
- Attach “quality tag’’ to the estimators [e.g. Variance]

Identification (..closed loop)

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA



$$\mathbf{z} := (\mathbf{y}^\top, \mathbf{u}^\top)^\top.$$

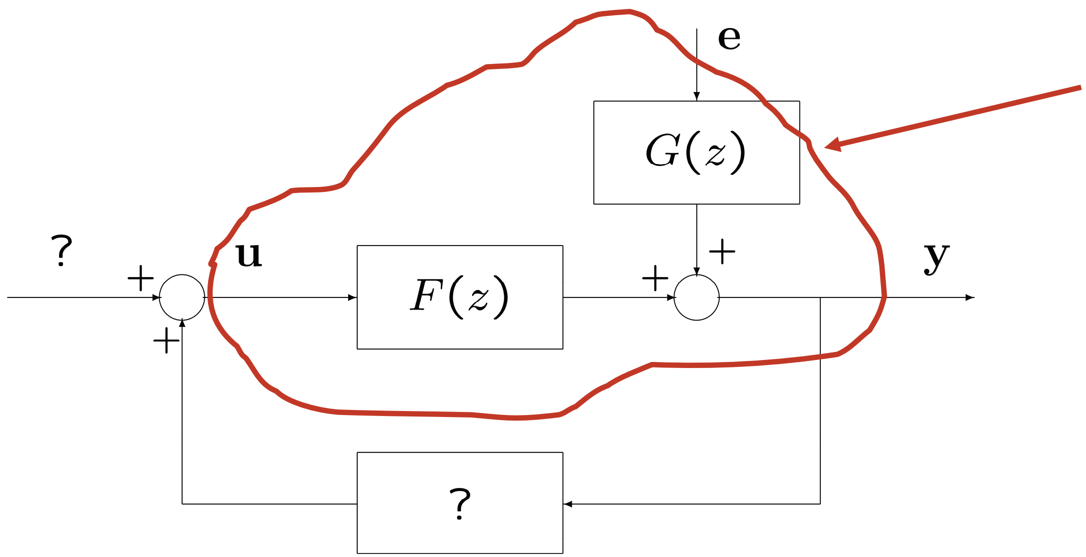
\+ $F(\infty) = 0.$

- Safety,
- Production quality
- Intrinsic feedback
- .....

- Estimate Parameters (A,B,C,K) [Point estimators in statistics]
- Attach "quality tag" to the estimators [e.g. Variance]

WELL-POSED ???

Identification (..closed loop)



$$\mathbf{z} := (\mathbf{y}^\top, \mathbf{u}^\top)^\top.$$

+ $F(\infty) = 0.$

"Essentially" $1:1$ correspondence between the covariance sequence of the joint process $(\mathbf{y}, \mathbf{u})$ and internally stable (linear) feedback models $\rightarrow$ UNIQUE $F(z)$!

[Ljung-Soderström-Gustavsson-Gevers-Anderson-Caines-Chan] [1974-1981]

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Outline of the Talk

- Identification
    - Statement of the problem
    - Closed loop identification
- **Subspace identification and stochastic realization**
    - Motivations (i.e. WHY?)
    - Main ideas (i.e. WHAT IS IT?)
- Main (classical) Algorithms (CCA,N4SID,MOESP)
- Closed Loop Subspace Identification (New Alg.)
- Recent contributions
    - Statistical Analysis (Asymptotic Variance)
    - Comparison of algorithms
    - Role of the Future Horizon
- Simulation results
- Conclusions
    - Open questions

WHY Subspace ID? (and why not!)

- *PROS:*
  1. No parametric optimization (as PEM), only linear algebra (QR, SVD,..) (+ Riccati)
  2. Implements Stochastic realization procedures *(understand what's going on!)*
  3. Identification with feedback (*Recent* !)

     → *MIMO identification (APPLICATIONS)*

- *CONS:*
  1. Statistical Analysis Difficult
  2. Many algorithms, relative efficiency not completely clear + finite data behavior

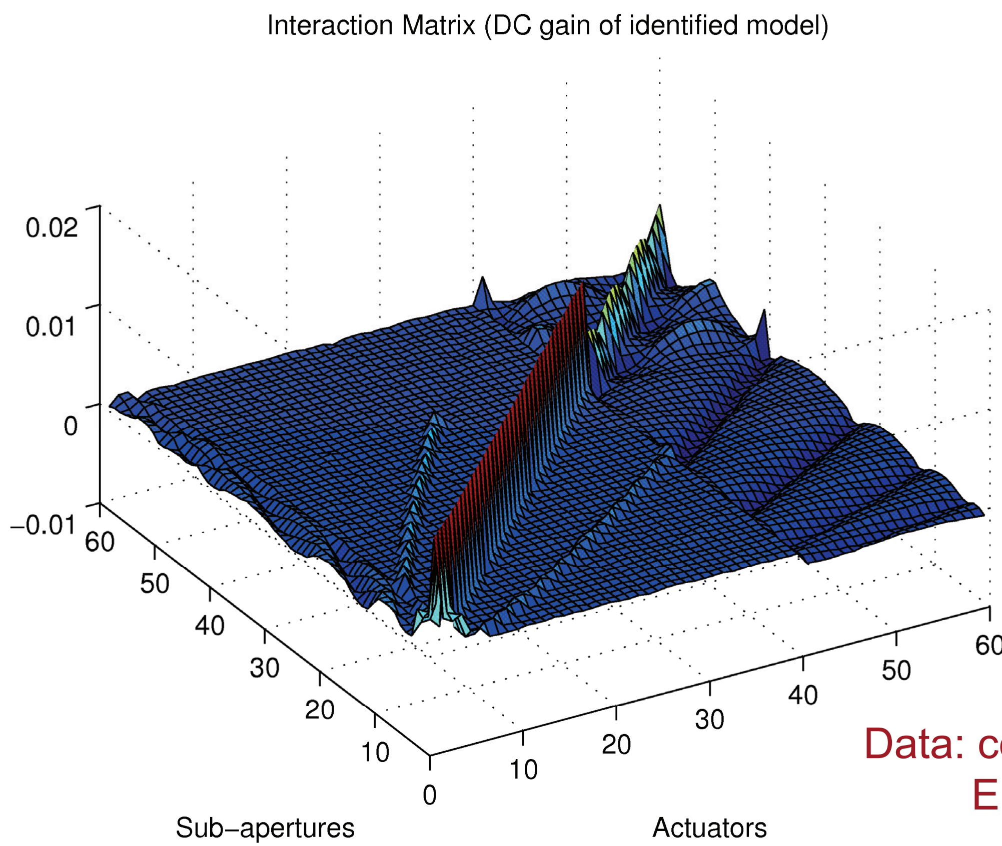
A 60x60 MIMO real system…

- Closed loop Identification of the Deformable Mirror in adaptive optics (DC Gain shown)

A 60x60 MIMO real system…

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

- Closed loop Identification of the Deformable Mirror in adaptive optics (DC Gain shown)



Identification of the “full” 60x60 system with PBSIDopt Using closed loop data (i.e. when the mirror curvature is controlled)

Matches very well with the DC gain which can be computed from a very accurate (non-linear) model of the plant and with physical insight

Data: courtesy of European Southern Observatory (ESO)

Models from covariances

Note: given $\{z_t\}_{t\in\mathbb{R}}$, $z_t = [y_t^\top u_t^\top]^\top \rightarrow \Sigma(\tau)$ (Ergodicity)

Problem: Given $\Sigma(\tau)$ find $A, B, C, K,\ \Lambda = \text{Var}\{\mathbf{e}(t)\}$ compute a (minimal) state space model (LTI).

STOCHASTIC REALIZATION PROBLEM

Huge liter.: $\underbrace{Akaike, Ruckebusch, Faurre, Lindquist\ and\ Picci}$, $\underbrace{Chiuso\ and\ Picci}$

(Just to cite a few!)

No measured "inputs" u(t) (time-series identification) (1969-1974-1985-1991)

Measured "inputs" u(t) (identification with inputs) (2002-2005)

Geometric Stochastic Realization

Given a stat. stoch. process $\{\mathbf{z}(t)\}_{t\in\mathbb{R}}$, $\mathbf{z}(t) = [\mathbf{y}^\top(t), \mathbf{u}^\top(t)]^\top$

Let $\mathcal{H} := \overline{\{\eta := \sum_i a_i^\top \mathbf{z}(t_i)\}} = \overline{\text{span}}\{\mathbf{z}(s), s \in \mathbb{Z}\}$

Geometric Stochastic Realization

Given a stat. stoch. process $\{\mathbf{z}(t)\}_{t\in\mathbb{R}}$, $\mathbf{z}(t) = [\mathbf{y}^\top(t), \mathbf{u}^\top(t)]^\top$

Let $\mathcal{H} := \overline{\{\eta := \sum_i a_i^\top \mathbf{z}(t_i)\}} = \overline{\text{span}}\{\mathbf{z}(s), s \in \mathbb{Z}\}$

Similar considerations in Willems (1986-2006) in a deterministic setting

Geometric Stoch. Realization:

1. construct $\mathcal{X}_t := \text{span}\{\mathbf{x}(t)\}$ via "geometric" operations on the space $\mathcal{H}$ (COVARIANCE SEQUENCE).

2. to each (many !!!) "state space" one can attach a realization (A, B, C, D) choosing a basis $\mathbf{x}(t)$ *(equivalence class as for deterministic realization)*

From Stochastic Realization to Subspace Algorithms

$$\mathbf{y}(t) \quad \underset{N\to\infty}{\overset{1:1}{\longleftrightarrow}} \quad \{y_t, y_{t+1}, \ldots, y_{t+N-1}\} = \mathbf{Y}_t$$

$$Stoch.\ \ Processes \quad \leftrightarrow \quad Lin.\ \ Algebra$$

From Stochastic Realization to Subspace Algorithms

$$\mathbf{y}(t) \quad \overset{1:1}{\underset{N\to\infty}{\longleftrightarrow}} \quad \{y_t, y_{t+1}, \ldots, y_{t+N-1}\} = \mathbf{Y}_t$$

$$Stoch.\ \ Processes \quad \leftrightarrow \quad Lin.\ \ Algebra$$

QR, SVD...

$$\mathcal{Y}_{[t,s]} = \mathrm{span}\{\mathbf{y}(t), .., \mathbf{y}(s)\} \quad \overset{1:1}{\underset{N\to\infty}{\longleftrightarrow}} \quad \mathbf{Y}_{[t,s]} := \begin{bmatrix} y_t & y_{t+1} & \cdots & y_{t+N-1} \\ y_{t+1} & y_{t+2} & \cdots & y_{t+N} \\ \vdots & \vdots & \ddots & \vdots \\ y_s & y_{s+1} & \cdots & y_{s+N-1} \end{bmatrix}$$

(abstract) Spaces of random variables ⟷ Row space of (Block) Hankel Data Matrices

Key Steps in Subspace ID

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

sometimes called past induced….

1. Construct a basis for the (predictor) state space $\mathbf{X}_t$ (a matrix!!)

2. $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{K}$ (+ Riccati...) FROM

$$\begin{cases} \mathbf{X}_{t+1} = A\mathbf{X}_t + B\mathbf{U}_t + K\hat{\mathbf{E}}_t \\ \mathbf{Y}_t = C\mathbf{X}_t \overset{\perp}{+} \mathbf{E}_t \end{cases}$$

where $\hat{\mathbf{E}}_t := \mathbf{Y}_t - \hat{C}\mathbf{X}_t$

$$\frac{\mathbf{E}_t\mathbf{X}_t^\top}{N} \simeq \mathbb{E}[\mathbf{e}(t)\mathbf{x}^\top(t)] = 0$$

How do we get the state?

1. The state $\mathcal{X}_p$ is the "interface" between "past" $(-\infty, p)$ and "future" $[p, +\infty)$ **(=SPLITTING)**

2. *(no $\mathbf{u}(t)$)* Every future output can be decomposed in a part that lies in the past (and hence in the state) and one that is "uncorrelated" with it (minimum variance estimator)

$$\mathbf{y}(p+k) \quad = \quad \hat{\mathbf{y}}(p+k|p) \overset{\perp}{+} \tilde{\mathbf{y}}(p+k|p)$$

AKAIKE 1975…

where

$$\hat{\mathbf{y}}(p+k|p) := E[\mathbf{y}(p+k)|\mathcal{Y}_p^-] = CA^k\mathbf{x}(p) \in \mathcal{X}_p$$

Orthogonal projection (minimum variance predictor)

The state space (no inputs)

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Orthogonal projection

1. A little of work will show that

$$\mathcal{X}_p := E\left[\mathcal{Y}_p^+ | \mathcal{Y}_p^-\right]$$

   is a "valid state space" **(Markovian Splitting)**

2. How do we construct it in practice?

   Canonical correlation analysis (CCA) comes into play: allows to discover **subsets** of variables in $\mathcal{Y}_p^-$ that are "maximally correlated" with $\mathcal{Y}_p^+$.
   (**model reduction/approximation..**)

Hotelling, 1936: *Relations between two sets of variables*

CCA (no inputs)

1. Compute

p=length of ``past''
f+1=length of ``future''

$$\begin{aligned}\hat{\Sigma}_{fp} &:= \frac{1}{N}\mathbf{Y}_{[p,p+f]}\mathbf{Y}^{\top}_{[0,p)}\\ \hat{\Sigma}_{pp} &:= \frac{1}{N}\mathbf{Y}_{[0,p)}\mathbf{Y}^{\top}_{[0,p)}\\ \hat{\Sigma}_{ff} &:= \frac{1}{N}\mathbf{Y}_{[p,p+f]}\mathbf{Y}^{\top}_{[p,p+f]}\end{aligned}$$

CCA (no inputs)

1. Compute

p=length of ``past''
f+1=length of ``future''

$$\hat{\Sigma}_{fp} := \frac{1}{N}\mathbf{Y}_{[p,p+f]}\mathbf{Y}_{[0,p)}^{\top}$$
$$\hat{\Sigma}_{pp} := \frac{1}{N}\mathbf{Y}_{[0,p)}\mathbf{Y}_{[0,p)}^{\top}$$
$$\hat{\Sigma}_{ff} := \frac{1}{N}\mathbf{Y}_{[p,p+f]}\mathbf{Y}_{[p,p+f]}^{\top}$$

2. Then SVD decompose

$$U_n S_n V_n^{\top} \simeq \hat{\Sigma}_{ff}^{-1/2}\hat{\Sigma}_{fp}\hat{\Sigma}_{pp}^{-1/2}$$

CCA (no inputs)

1. Compute

p=length of ``past''
f+1=length of ``future''

$$
\begin{aligned}
\hat{\Sigma}_{fp} &:= \frac{1}{N}\mathbf{Y}_{[p,p+f]}\mathbf{Y}_{[0,p)}^\top \\
\hat{\Sigma}_{pp} &:= \frac{1}{N}\mathbf{Y}_{[0,p)}\mathbf{Y}_{[0,p)}^\top \\
\hat{\Sigma}_{ff} &:= \frac{1}{N}\mathbf{Y}_{[p,p+f]}\mathbf{Y}_{[p,p+f]}^\top
\end{aligned}
$$

2. Then SVD decompose

$$U_n S_n V_n^\top \simeq \hat{\Sigma}_{ff}^{-1/2}\hat{\Sigma}_{fp}\hat{\Sigma}_{pp}^{-1/2}$$

3. Set

$$\hat{\mathbf{X}}_p := S_n^{-1/2}U_n^\top\hat{\Sigma}_{ff}^{-1/2}\hat{\Sigma}_{fp}\hat{\Sigma}_{pp}^{-1}\mathbf{Y}_{[0,p)}$$

With Inputs/Feedback

- Similar ideas/more complicated formulas
- Involves “oblique projections” and conditional CCA
- Feedback? Needs new point of view:
  the predictor model
  (rather simple idea to be honest!)

Outline of the Talk

- Identification
  - Statement of the problem
  - Closed loop identification
- Subspace identification and stochastic realization
  - Motivations
  - Main ideas
- **Main (classical) Algorithms (CCA,N4SID,MOESP)**
- Closed Loop Subspace Identification (New Alg.)
- Recent contributions
  - Statistical Analysis (Asymptotic Variance)
  - Comparison of algorithms
  - Role of the Future Horizon
- Simulation results
- Conclusions
  - Open questions

Milestones in Subspace ID

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

- Akaike (1975-76) Canonical Variables as “state”.
- Moonen et al. (1988) Deterministic ID: “*Intersection algorithms*” (in the spirit of Willems 1986)
- Aoki (1990): Realization applied to estimated covariance (*a la* Ho-Kalman)
- Larimore (1983-1990): *first “**CCA**” method*
- Verhaegen et al. (1992-1994): ***MOESP*** *“class” (work on observability matrix rather than state – “**shift invariance**”)*
- *van Overschee - De Moor (1993): first time ``finite-interval stochastic realization translated right!” (time-series id.)[CCA + Riccati to compute K]*
- *van Overschee - De Moor (1994):* ***N4SID*** *algorithm*
- *Picci-Katayama, Chiuso-Picci (1996-1999) – ORT DEC Algorithm*

Box-Jenkins sort of model

NOTATION

Observability matrix

$$\Gamma := \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^f \end{bmatrix}$$

``deterministic'' (**u** as input)

$$H_d = \begin{bmatrix} 0 & 0 & \dots & 0 \\ CB & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-1}B & CA^{f-2}B & \dots & O \end{bmatrix}$$

Toeplitz matrices with Markov Parameters

$$H_s = \begin{bmatrix} I & 0 & \dots & 0 \\ CK & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-1}K & CA^{f-2}K & \dots & I \end{bmatrix}$$

``stochastic'' (**e** as input)

CCA, N4SID, MOESP

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

1. Using

State belongs to “past”

$$\mathbf{Y}_{[p,p+f]} = \Gamma \mathbf{X}_p + H_d \mathbf{U}_{[p,p+f]} + H_s \mathbf{E}_{[p,p+f]}$$

CCA, N4SID, MOESP

1. Using

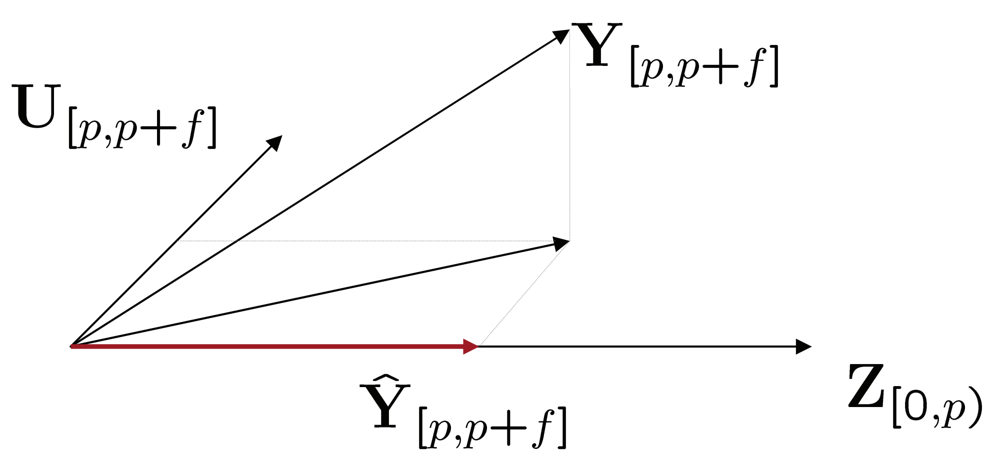
State belongs to "past"

$$\mathbf{Y}_{[p,p+f]} = \Gamma \mathbf{X}_p + H_d \mathbf{U}_{[p,p+f]} + H_s \mathbf{E}_{[p,p+f]}$$

2.

Oblique projection (QR decomp.)

$$\Gamma \mathbf{X}_p \simeq \hat{\mathbf{Y}}_{[p,p+f]} := E_{||\mathbf{U}_{[p,p+f]}} \left[\mathbf{Y}_{[p,p+f]} \mid \mathbf{Z}_{[0,p)} \right]$$



DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

CCA, N4SID, MOESP

1. Using

State belongs to "past"

$$\mathbf{Y}_{[p,p+f]} = \Gamma \mathbf{X}_p + H_d \mathbf{U}_{[p,p+f]} + H_s \mathbf{E}_{[p,p+f]}$$

2. Oblique projection (QR decomp.)

$$\Gamma \mathbf{X}_p \simeq \hat{\mathbf{Y}}_{[p,p+f]} := E_{||\mathbf{U}_{[p,p+f]}} \left[\mathbf{Y}_{[p,p+f]} \mid \mathbf{Z}_{[0,p)} \right]$$

3. Use "weighted SVD" to estimate Γ and/or $\mathbf{X}_p$

$$U_n S_n V_n^\top \simeq W_r \hat{\mathbf{Y}}_{[p,p+f]} W_c$$

van Overschee – De Moor 1995

CCA, N4SID, MOESP

1. Using

State belongs to "past"

$$\mathbf{Y}_{[p,p+f]} = \Gamma \mathbf{X}_p + H_d \mathbf{U}_{[p,p+f]} + H_s \mathbf{E}_{[p,p+f]}$$

2.

Oblique projection (QR decomp.)

$$\Gamma \mathbf{X}_p \simeq \hat{\mathbf{Y}}_{[p,p+f]} := E_{||\mathbf{U}_{[p,p+f]}} \left[\mathbf{Y}_{[p,p+f]} \mid \mathbf{Z}_{[0,p)} \right]$$

3. Use "weighted SVD" to estimate Γ and/or $\mathbf{X}_p$

$$U_n S_n V_n^\top \simeq W_r \hat{\mathbf{Y}}_{[p,p+f]} W_c$$

van Overschee – De Moor 1995

4.

$$\hat{\Gamma} := W_r^{-1} U_n S_n^{1/2} \qquad \hat{\mathbf{X}}_p := \hat{\Gamma}_r^{-L} \hat{\mathbf{Y}}_{[p,p+f]}$$

PROBLEM: CLOSED-LOOP DATA

- CCA,MOESP,N4SID are *not consistent* if there is feedback [see e.g. Ljung and McKelvey 1996]
- What does go wrong?

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Where do things go wrong?

RECALL:

“ORTHOGONAL”

$$\mathbf{Y}_{[p,p+f]} = \Gamma \mathbf{X}_p + H_d \mathbf{U}_{[p,p+f]} + H_s \mathbf{E}_{[p,p+f]}$$

NOT “ORTHOGONAL”
WITH FEEDBACK
(future inputs depend upon past outputs/noise!)

Where do things go wrong?

RECALL:

"ORTHOGONAL"

$$\mathbf{Y}_{[p,p+f]} = \Gamma \mathbf{X}_p + H_d \mathbf{U}_{[p,p+f]} + H_s \mathbf{E}_{[p,p+f]}$$

NOT "ORTHOGONAL"
WITH FEEDBACK
(future inputs depend upon past outputs/noise!)

- Construction of the state! Need a NEW LOOK at stochastic realization theory when there are input (and possibly feedback) [see Chiuso and Picci 2002,Chiuso and Picci 2005]

Outline of the Talk

- Identification
  - Statement of the problem
  - Closed loop identification
- Subspace identification and stochastic realization
  - Motivations
  - Main ideas
- Main (classical) Algorithms (CCA,N4SID,MOESP)
- **Closed Loop Subspace Identification (New Alg.)**
- Recent contributions
  - Statistical Analysis (Asymptotic Variance)
  - Comparison of algorithms
  - Role of the Future Horizon
- Simulation results
- Conclusions
  - Open questions

Closed-loop (Subspace) ID

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

- Early contributions (Verhaegen '93/Ljung-McKelvey'96/ van Overschee-De Moor '97/Chou-Verhaegen '99/Katayama et al.)
- Chiuso-Picci 2002/2003: construction of the state with feedback-stochastic realization
- Qin-Ljung 2003 Innovation Estimation Algorithm (IEM)
- Jansson 2003 Subspace Algorithm with pre-estimation (SS-ARX) [similar to Peternell 1996-Shi and McGregor 2001] [equivalent to Larimore 2004 (ADAPTx) (Chiuso ACC '07 to appear.)]
- Chiuso-Picci 2004/2005 "Geometric version of SSARX" (PBSID) (more in the geometric spirit of subspace ID)

Predictor Based Subspace IDentification

Assumptions

Most algorithms need the past horizon p to go to infinity with N while satisfying ($\rho = max|\lambda(\bar{A})|$):

$$p \geq \frac{logN^{-d/2}}{log|\rho|} \qquad 1 < d < \infty$$

$$p = o((logN)^{\alpha}) \qquad \alpha < \infty$$

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Assumptions

Most algorithms need the past horizon p to go to infinity with N while satisfying ($\rho = max|\lambda(\bar{A})|$):

$$p \geq \frac{logN^{-d/2}}{log|\rho|} \qquad 1 < d < \infty$$
$$p = o((logN)^{\alpha}) \qquad \alpha < \infty$$

Further Motivation:

Subspace methods are "covariance based": it is well known that there is no Efficient estimator for ARMA(X) processes based on a finite number of covariances [Walker '61, Porat-Friedlander '85] (unless the process is AR(X))

Innovation Estimation

[Qin-Ljung 2003 (Chiuso-Picci 2002)]

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

RECALL:

LIES IN THE "PAST" → $\Gamma \mathbf{X}_p$

``DETERMINISTIC INPUT TERM'' → $H_d \mathbf{U}_{[p,p+f)} + H_s \mathbf{E}_{[p,p+f)}$

$$\mathbf{Y}_{[p,p+f)} = \Gamma \mathbf{X}_p + H_d \mathbf{U}_{[p,p+f)} + H_s \mathbf{E}_{[p,p+f)}$$

IF WE KNEW $\mathbf{E}_{[p,p+f)}$

$$\Gamma \mathbf{X}_p = E_{||\mathbf{U}_{[p,p+f)} \vee \mathbf{E}_{[p,p+f)}} \left[\mathbf{Y}_{[p,p+f)} \mid \mathbf{Z}_{[0,p)} \right]$$

SIMILAR (BUT NOT THE SAME) Construction as Moonen et al (1988), (Willems 1986)

INNOVATION ESTIMATION: estimate innovation $\hat{\mathbf{E}}_{[p,p+f]}$ (OLD CIRCLE OF IDEAS)

"Future-Corrected" CCA

(ADAPTx) [Shi-McGregor '01 – Larimore '04]

RECALL:

$$\mathbf{Y}_{[p,p+f]} = \Gamma \mathbf{X}_p + H_d \mathbf{U}_{[p,p+f]} + H_s \mathbf{E}_{[p,p+f]}$$

IF WE KNEW $\mathbf{H}_d$

ESTIMATE VIA ARX MODELING

$$\Gamma \mathbf{X}_p = E\left[\underbrace{\mathbf{Y}_{[p,p+f]} - \widehat{H}_d \mathbf{U}_{[p,p+f]}}_{\text{"Corrected" Future}} \mid \underbrace{\mathbf{Z}_{[0,p)}}_{\text{Past}}\right]$$

Do **CCA** between corrected future and past! (closely related to **uCCA** in Peternell et al. '96)

Predictor-Based SID

[Jansson '03, Chiuso-Picci '04, Chiuso '05-'06]

- Based on "predictor model" as PEM idenfication.

$$\begin{cases} \mathbf{X}_{t+1} = (A - KC)\mathbf{X}_t + B\mathbf{U}_t + K\mathbf{Y}_t \\ \mathbf{Y}_t = C\mathbf{X}_t + \mathbf{E}_t \end{cases} \quad (1)$$

Output of a system (known inputs) measured in white noise.
<u>CAREFUL: noise CORRELATED with future inputs (outputs)</u>

DISCUSSED IN DETAIL in CHIUSO-PICCI Automatica '05- IFAC '05

NOTATION

(PREDICTOR MODEL $\bar{A} := A - KC$)

Observability matrix

$$\bar{\Gamma} := \begin{bmatrix} C \\ C\bar{A} \\ C\bar{A}^2 \\ \vdots \\ C\bar{A}^f \end{bmatrix}$$

``deterministic'' (**u** as input)

$$\bar{H}_d = \begin{bmatrix} 0 & 0 & \dots & 0 \\ CB & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C\bar{A}^{f-1}B & C\bar{A}^{f-2}B & \dots & O \end{bmatrix}$$

Toeplitz matrices with Markov Parameters

$$\bar{H}_s = \begin{bmatrix} I & 0 & \dots & 0 \\ CK & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C\bar{A}^{f-1}K & C\bar{A}^{f-2}K & \dots & I \end{bmatrix}$$

``stochastic'' (**y** as input)

Predictor-Based SID

[Jansson '03, Chiuso-Picci '04, Chiuso '05-'06]

$$\mathbf{Y}_{[p,p+f]} = \bar{\Gamma}\mathbf{X}_p + \bar{H}_d\mathbf{U}_{[p,p+f]} + \bar{H}_s\mathbf{Y}_{[p,p+f]} + \mathbf{E}_{[p,p+f]}$$

1. Jansson 2003 (SSARX): Pre-estimates $\hat{\bar{H}}_d, \hat{\bar{H}}_s$ using ARX and form

$$\mathbf{Y}_{[p,p+f]} - \hat{\bar{H}}_d\mathbf{U}_{[p,p+f]} - \hat{\bar{H}}_s\mathbf{Y}_{[p,p+f]} \simeq \bar{\Gamma}\mathbf{X}_p + \mathbf{E}_{[p,p+f]}$$

2. Chiuso-Picci 2004 (PBSID): Use a "geometric approach" (oblique projections)[equivalent to estimating f ARX models]

"Unconstrained" least squares (PBSID)

$$\mathbf{Y}_{[p,p+f]} = \bar{\Gamma}\mathbf{X}_p + \bar{H}_d\mathbf{U}_{[p,p+f]} + \bar{H}_s\mathbf{Y}_{[p,p+f]} + \mathbf{E}_{[p,p+f]}$$

$$\begin{bmatrix} \mathbf{Y}_p \\ \mathbf{Y}_{p+1} \\ \vdots \\ \mathbf{Y}_{p+f} \end{bmatrix} \doteq \begin{bmatrix} \Xi_0 \\ \Xi_1 \\ \vdots \\ \Xi_f \end{bmatrix} \mathbf{Z}_{[0,p)} + \begin{bmatrix} 0 & 0 & \dots & 0 \\ \Psi_{11} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{ff} & \dots & \Psi_{f1} & 0 \end{bmatrix} \mathbf{Z}_{[p,p+f]} + \begin{bmatrix} \mathbf{E}_p \\ \mathbf{E}_{p+1} \\ \vdots \\ \mathbf{E}_{p+f} \end{bmatrix}$$

$$\mathbf{E}_{[p,p+f]} := \begin{bmatrix} e_p & e_{p+1} & \dots & e_{p+N-1} \\ e_{p+1} & e_{p+2} & \dots & e_{p+N} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p+f} & e_{p+f+1} & \dots & e_{p+f+N-1} \end{bmatrix}$$

The vectorization has Singular Variance

"Optimal" Predictor-Based SID

[PBSIDopt: Chiuso '05-'06]

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Transform (quite naturally) the set of oblique projections (PBSID) in a weighted least-squares problem corresponding to minimizing $\sum_{t=p}^{p+f+N-1} \|e_t\|^2$

“Optimal” Predictor-Based SID

[PBSIDopt: Chiuso ‘05-’06]

Transform (quite naturally) the set of oblique projections (PBSID) in a weighted least-squares problem corresponding to minimizing $\sum_{t=p}^{p+f+N-1} \|e_t\|^2$

- Boils down to solving a (HUGE) equality constrained least squares problem (Computationally **very** intensive)

“Optimal” Predictor-Based SID

[PBSIDopt: Chiuso ‘05-’06]

Transform (quite naturally) the set of oblique projections (PBSID) in a weighted least-squares problem corresponding to minimizing $\sum_{t=p}^{p+f+N-1} \|e_t\|^2$

- Boils down to solving a (HUGE) equality constrained least squares problem (Computationally **very** intensive)

- The “optimized” PBSID automatically takes into account/enforces the
  - *block Hankel structure* of the data
  - *block Toeplitz structure* of the estimated parameters

"Optimal" Predictor-Based SID

[PBSIDopt: Chiuso '05-'06]

- PBSIDopt

  (exactly!) equivalent to estimating a long VARX *(of order p)* model

$$\mathbf{Y}_p = \begin{bmatrix} \hat{\Phi}_p & \dots & \hat{\Phi}_f & \dots & \hat{\Phi}_1 \end{bmatrix} \mathbf{Z}_{[0,p)} + \mathbf{E}_p$$

followed by state estim. via SVD (= "model reduction")

“Optimal” Predictor-Based SID

[PBSIDopt: Chiuso ’05-’06]

- PBSIDopt

  (exactly!) equivalent to estimating a long VARX *(of order p)* model

$$\mathbf{Y}_p = \left[\ \hat{\Phi}_p \quad \dots \quad \hat{\Phi}_f \quad \dots \quad \hat{\Phi}_1\ \right] \mathbf{Z}_{[0,p)} + \mathbf{E}_p$$

followed by state estim. via SVD (= “model reduction”)

  - Advantages for computational complexity (much easier)
  - Take advantage of results from Stochastic Realization

- *Dahlen-Scherrer (2002)* proved that CCA is asymptotically equivalent to AR modeling + balanced model reduction

- Related approaches in [*Qin-Liung 2006*, *Onodera-Emoto-Qin 2006*]

“Optimal” Predictor-Based SID

[PBSIDopt: Chiuso ‘05-’06]

$$\Phi_i := C\bar{A}^{i-1}[K \ \ B]$$

$$\hat{\mathbf{Y}}_{[p,p+f]} = \begin{bmatrix} \hat{\Phi}_p & \dots & \hat{\Phi}_f & \dots & \hat{\Phi}_1 \\ 0 & \hat{\Phi}_p & \dots & \ddots & \hat{\Phi}_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \hat{\Phi}_p & \dots \end{bmatrix} \mathbf{Z}_{[0,p)}$$

Similar Ideas (Toeplitz structure) in *Peternell et al. 1996* and perhaps relation with cut and shift (*Rapisarda-Willems*) **BUT** in our case the structure follows from an “optimally weighted” procedure, not just from imposing “ad-hoc” constraints

Outline of the Talk

- Identification
  - Statement of the problem
  - Closed loop identification
- Subspace identification and stochastic realization
  - Motivations
  - Main ideas
- Main (classical) Algorithms (CCA,N4SID,MOESP)
- Closed Loop Subspace Identification (New Alg.)
- **Recent contributions**
  - Statistical Analysis (Asymptotic Variance)
  - Comparison of algorithms
  - Role of the Future Horizon
- **Simulation results**
- Conclusions
  - Open questions

Classical vs. New methods

- **CCA**,N4SID,MOESP:
  - **Consistent** *(Open Loop) [..many..]*
  - ***Known variance formulas*** *[Bauer et al. 00, Jansson 00, Bauer-Ljung 02, Chiuso-Picci 04, Bauer 05]*
  - **CCA efficient** *(time-series, [Bauer05])* + **optimal** *(white input, [Bauer-Ljung 02], state joint model = state model of y given u [Chiuso 07 TAC sub.])*

- **SSARX** [Jansson 03], IEM. [Qin-Ljung 03], **PBSID** [Chiuso-Picci 04] **ADAPTx** [Larimore 04, ADAPTx]
  - **Consistent** *(Closed Loop + Infinite ``past'') [Chiuso-Picci,Automatica '05]*
  - **Known Variance formulas** *[Chiuso TAC'06]*
  - **Similarities of PBSID with PEM** *[Chiuso-Picci IFAC'05] and relation* ***SSARX/*****PBSID/CCA/FC-CCA/PBSIDopt** *[TAC 07, Chiuso SYSID'06-Chiuso CDC'06-Chiuso AUTO 07, Chiuso ACC 07]*
  - *Relation with other methods (e.g. IEM)???*

Asymptotic Analysis

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

- $\mathrm{e}(t)$ is a martingale difference sequence + constant conditional variance + bound on fourth order moments

Give two estimators $\widehat{\theta}_N$, $\widehat{\psi}_N$ we shall use the notation

$$\widehat{\theta}_N \doteq \widehat{\psi}_N$$

if $\widehat{\theta}_N - \widehat{\psi}_N = o_P(1/\sqrt{N})$ and hence $\sqrt{N}\widehat{\theta}_N$ and $\sqrt{N}\widehat{\psi}_N$ have the *same* asymptoic distribution

Asymptotic Variance

- Open Loop (N4SID,CCA,MOESP) many papers:
  - *Bauer-Deistler-Scherrer (99), Bauer-Jansson (00), Jansson (00), Bauer-Ljung (02), Bauer (05), Chiuso-Picci (04) [+ Lovera-Bittanti using Bootstrap (00)]*
- Closed Loop: IEM,SSARX,PBSID (WFA)
  - *Chiuso (TAC 2006)*

Asymptotic Variance (PBSID)
(Chiuso, *TAC06*)

Let $\tilde{A}_N := \hat{A}_N - A_N$ ($\tilde{B}_N$ etc.), then $\sqrt{N}\mathrm{vec}\left(\tilde{A}_N\right)$, $\sqrt{N}\mathrm{vec}\left(\tilde{B}_N\right)$ and $\sqrt{N}\mathrm{vec}\left(\tilde{C}_N\right)$ are jointly asymptotically normal with as. variance matrix

$$\begin{pmatrix} M_{A1}P & M_{A2}\bar{P} \\ M_{B1}P & M_{B2}\bar{P} \\ M_C P & 0 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_{22} \end{pmatrix} \begin{pmatrix} M_{A1}P & M_{A2}\bar{P} \\ M_{B1}P & M_{B2}\bar{P} \\ M_C P & 0 \end{pmatrix}^\top$$

1. P and $\bar{P}$ permutation matrices
2. Σ_{ij} input-output conditional covariance matrices
3. M_{A1}, M_{A2}, depends on the "true" system parameters and some cond. cov. matrices (state-input-output)

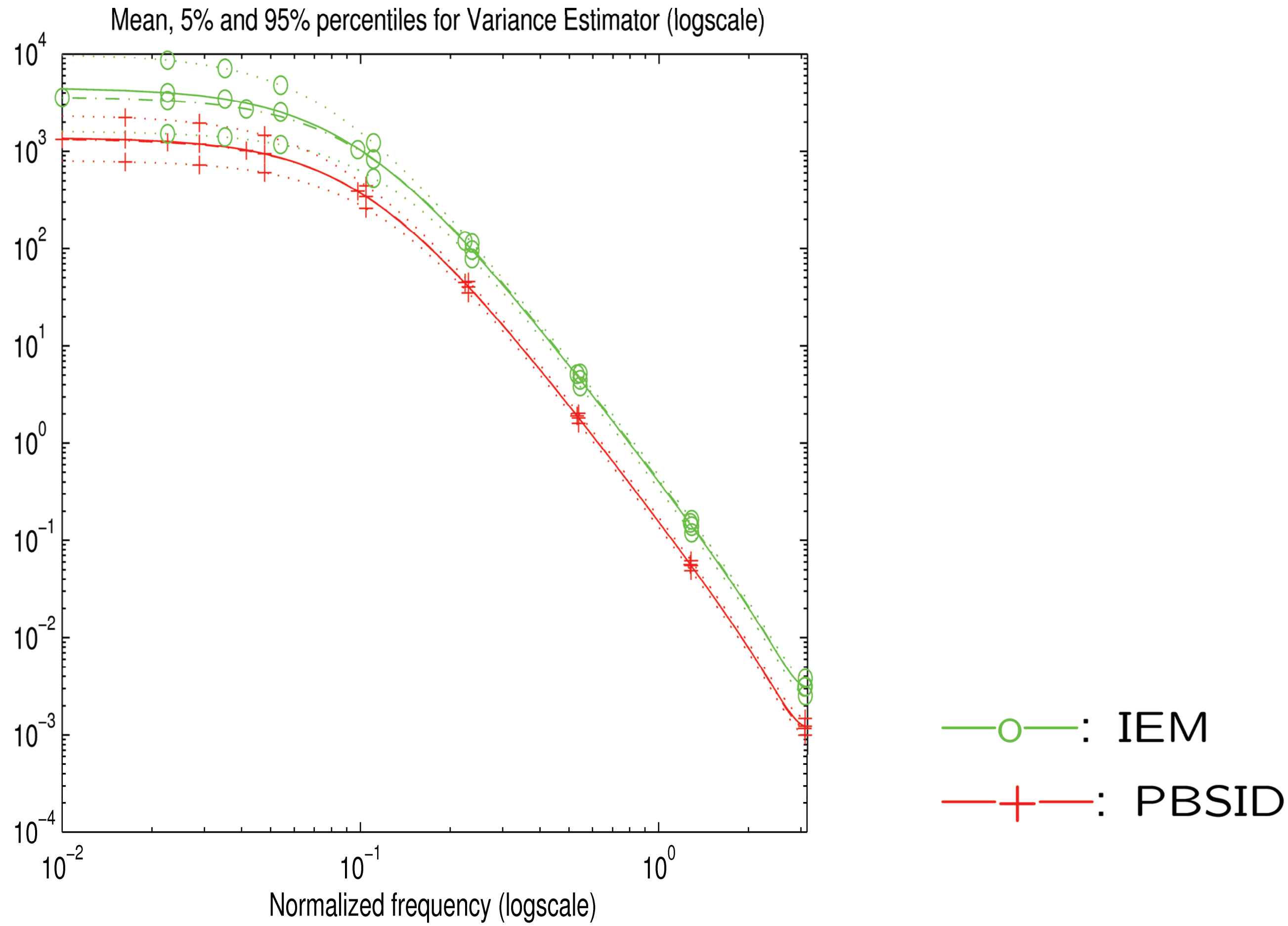
Practical Problem

- Want to assess model accuracy starting from data alone.
- Do the ASVAR formulas help us to this purpose?
  - Compute variance from data:
    - Difficult/Computationally expensive?
    - Reliable?

SIMPLE (O(N)) ALGORITHM TO APPROXIMATE VARIANCE FROM DATA

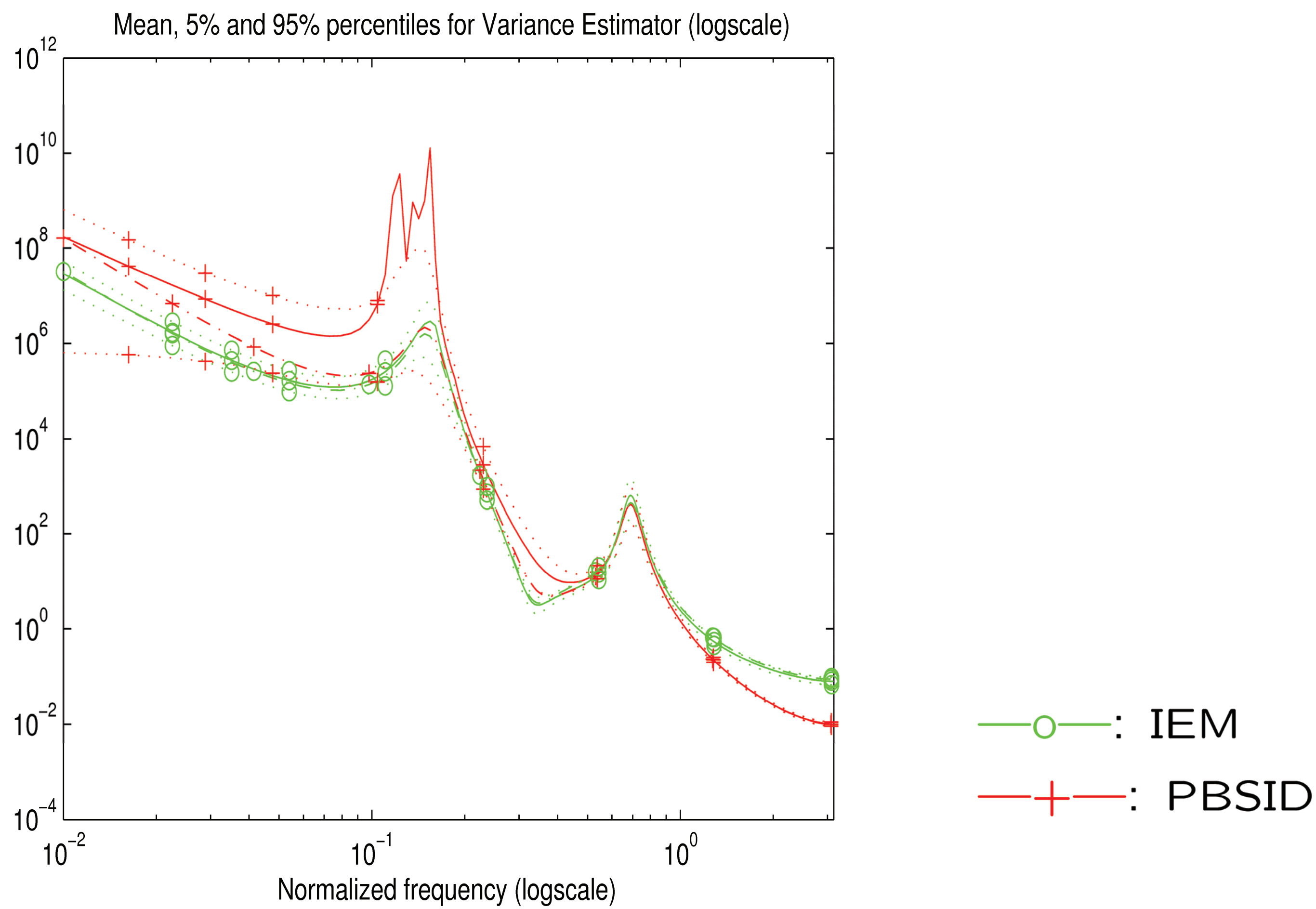
Estimated vs. Asympt. Variance

Order 1



Estimated vs. Asympt. Variance

Order 5



Outline of the Talk

- Identification
  - Statement of the problem
  - Closed loop identification
- Subspace identification and stochastic realization
  - Motivations
  - Main ideas
- Main (classical) Algorithms (CCA,N4SID,MOESP)
- Closed Loop Subspace Identification (New Alg.)
- **Recent contributions**
  - Statistical Analysis (Asymptotic Variance)
  - **Comparison of algorithms**
  - Role of the Future Horizon
- **Simulation results**
- Conclusions
  - Open questions

PBSID/PBSIDopt vs. CCA
(open loop)

- PBSID asympt. Eq. to CCA with white inputs
- PBSIDopt no worse than CCA

  [Chiuso TAC 07, to appear]

Theorem. *Denote with* $\hat{\Theta}^{CCA}$ *and* $\hat{\Theta}^{P_{opt}}$ *the estimators of any system invariant using respectively CCA and PBSID*$_{opt}$*; then,*

$$\text{AsVar}\{\sqrt{N}\hat{\Theta}^{P_{opt}}\} \leq \text{AsVar}\{\sqrt{N}\hat{\Theta}^{CCA}\}.$$

SSARX vs. PBSID vs. ADAPTx + Ljung-McKelvey '96 (closed loop)

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

- SSARX, ADAPTx and PBSID are asymptotically equivalent

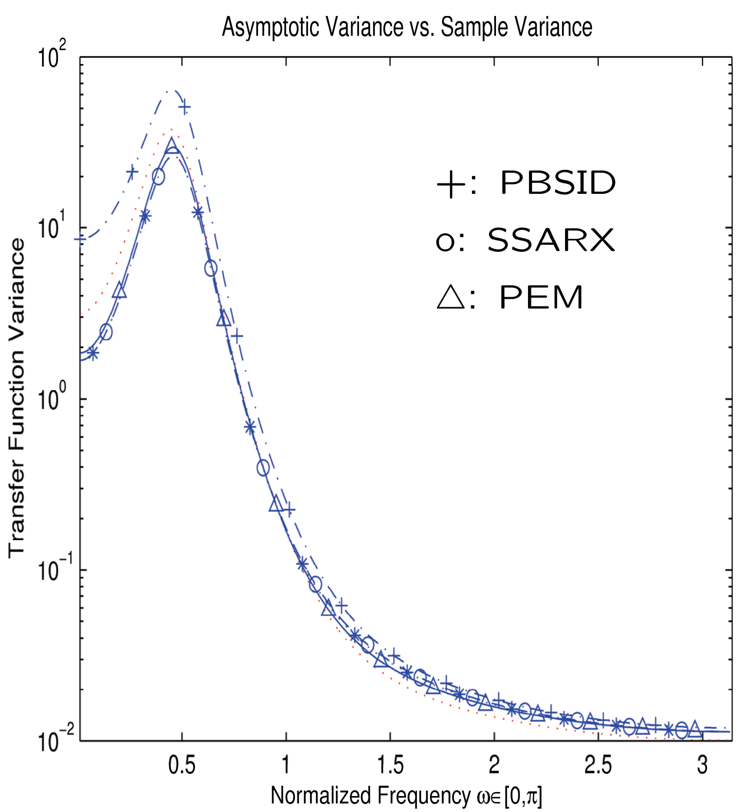
In the sense of VanOverschee-DeMoor '95

- PBSIDopt is a "weighted version" of Ljung-McKelvey 1996

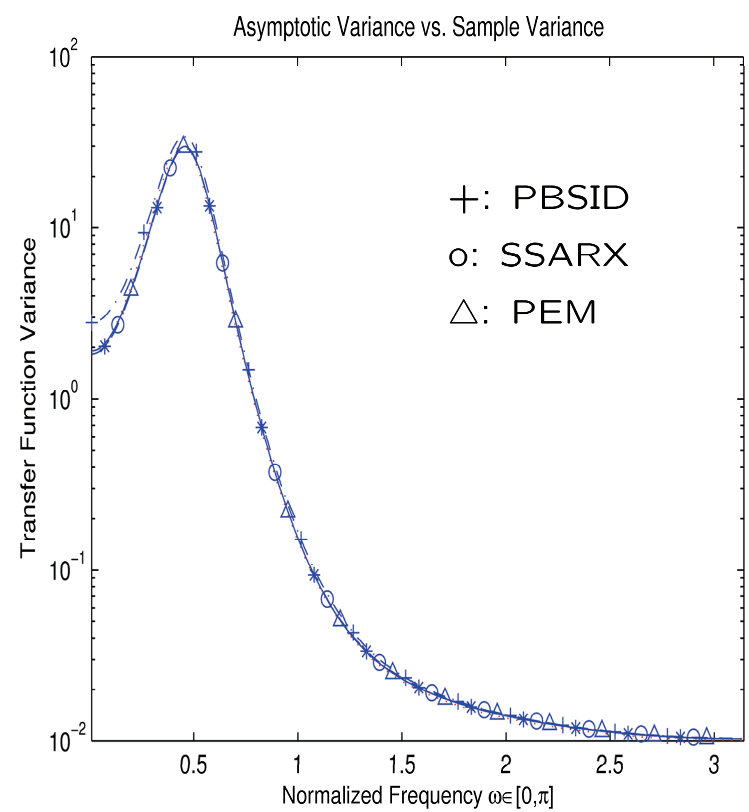
[Chiuso, Automatica 2007 + ACC 07 (to appear)]

SIMULATIONS
(PBSID vs. SSARX: II order):

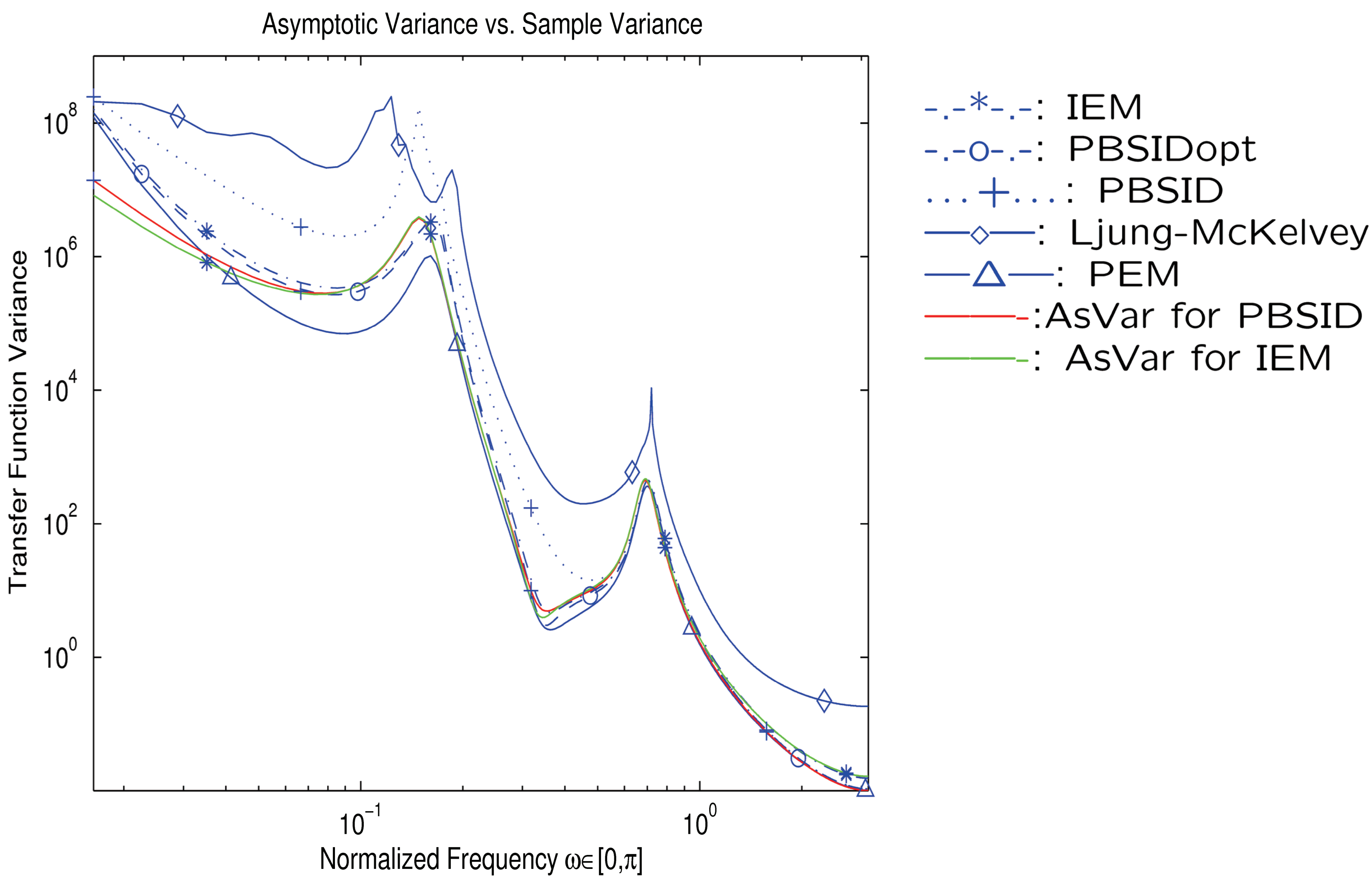
$f = 10,\ p = 10,\ N = 1000$



$f = 10,\ p = 30,\ N = 3000$



Fitfh-order example (CL)



Outline of the Talk

- Identification
  - Statement of the problem
  - Closed loop identification
- Subspace identification and stochastic realization
  - Motivations
  - Main ideas
- Main (classical) Algorithms (CCA,N4SID,MOESP)
- Closed Loop Subspace Identification (New Alg.)
- **Recent contributions**
  - Statistical Analysis (Asymptotic Variance)
  - Comparison of algorithms
  - **Role of the Future Horizon**
- **Simulation results**
- Conclusions
  - Open questions

Role of f (example)

- MODEL

$$\mathbf{y}(t) - 0.5\mathbf{y}(t-1) = \mathbf{u}(t-1) + \mathbf{e}(t) + 0.5\mathbf{e}(t-1)$$

- INPUT $\mathbf{u}(t)$

$$\mathbf{u}(t) - 0.9\mathbf{u}(t) = \mathbf{n}(t) - \alpha\mathbf{n}(t-1)$$

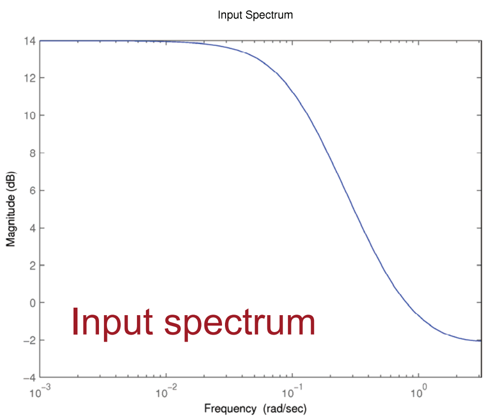
$\mathbf{n}(t)$ white noise

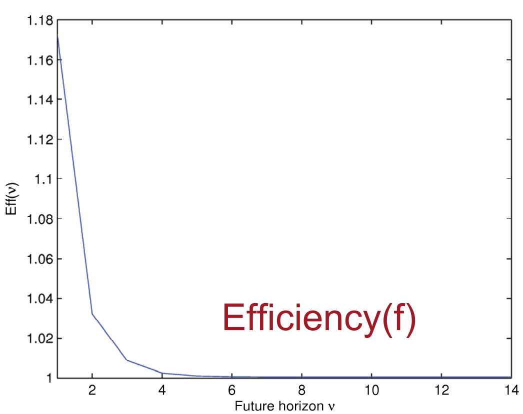
- "Efficiency" index

$$Eff(f) := \frac{\int_0^{2\pi} \text{AsVar}\left\{\hat{F}_f(e^{j\omega})\right\} d\omega}{\int_0^{2\pi} CRLB_{\hat{F}}(j\omega)\, d\omega} \qquad (1)$$

Role of f

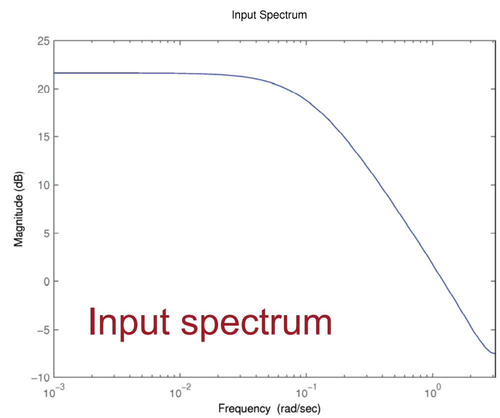
$\alpha = 0.5$

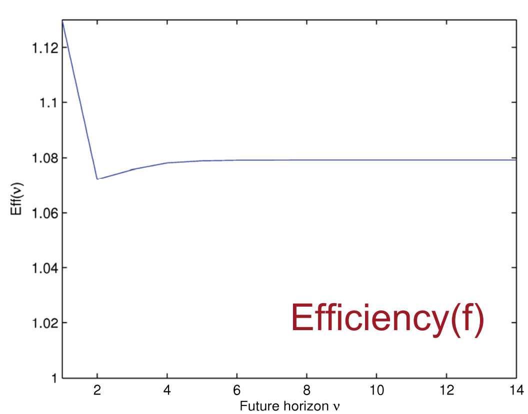




f

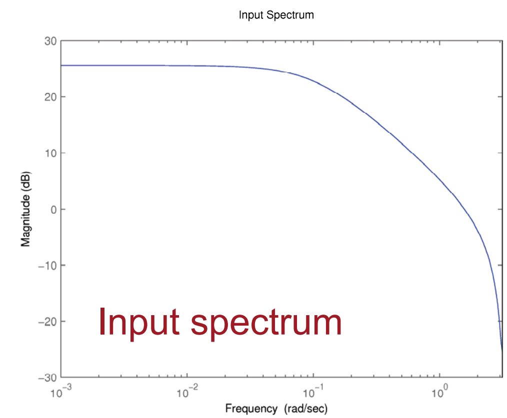
$\alpha = -0.2$

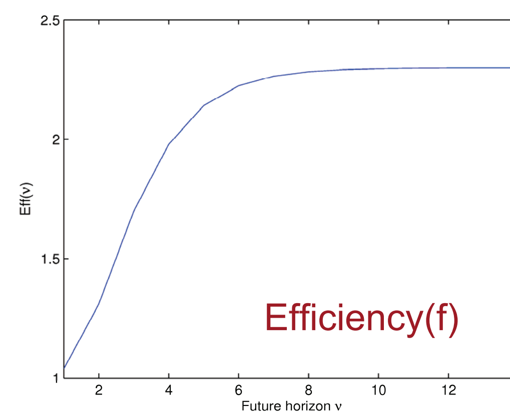




f

$\alpha = -0.9$





f

Role of f (future horizon)

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Can show that the estimated state via SSARX-ADAPTx-PBSID is of the form

$$
\begin{aligned}
\hat{X}_t - X_t &= S_n^{-1}(f)\hat{E}\left[V_t(f) \mid Z_{[t_0,t)}\right] + \\
&\quad -S_n^{-1}(f)U_n^\top(f)W_L^{-1}(f)\tilde{H}_f^d\hat{E}\left[U_{[t,T]} \mid Z_{[t_0,t)}\right]
\end{aligned}
$$

Role of f (future horizon)

Fixed basis

Can show that the estimated state via SSARX-ADAPTx-PBSID is of the form

$$\hat{X}_t - X_t = S_n^{-1}(f)\hat{E}\left[V_t(f) \mid Z_{[t_0,t)}\right] + \\ -S_n^{-1}(f)U_n^{\top}(f)W_L^{-1}(f)\tilde{H}_f^d\hat{E}\left[U_{[t,T]} \mid Z_{[t_0,t)}\right]$$

Goes to zero with N, Variance independent of f for u white

Error in estimation of the Markov Parameters

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Role of f (future horizon)

Can show that the estimated state via SSARX-ADAPTx-PBSID is of the form

$$\hat{X}_t - X_t = S_n^{-1}(f)\hat{E}\left[V_t(f) \mid Z_{[t_0,t)}\right] + \\ -S_n^{-1}(f)U_n^{\top}(f)W_L^{-1}(f)\tilde{H}_f^d\hat{E}\left[U_{[t,T]} \mid Z_{[t_0,t)}\right]$$

Fixed basis: X_t

Decreasing with f: $S_n^{-1}(f)$

Goes to zero with N, Variance independent of f: $\hat{E}\left[V_t(f) \mid Z_{[t_0,t)}\right]$

Error in estimation of the Markov Parameters: $\tilde{H}_f^d$

Increases with f, unless u(t) white, in which case vanishes with N: $\hat{E}\left[U_{[t,T]} \mid Z_{[t_0,t)}\right]$

- For white inputs variance is monotonically decreasing with f *[Bauer-Ljung]*

(the second term on the right hand side can be neglected while the first is monotonically decreasing)

Role of f (future horizon)

Theorem [Chiuso, TAC 2007, Submitted]

Assume the input process is such that the rows of X_t^u are, up to $o_P(1/\sqrt{N})$ terms, linear combinations of the rows of X_t (or equivalently span$\{\mathbf{x}^u(t)\} =: \mathcal{X}_t^u \subseteq \mathcal{X}_t :=$ span$\{\mathbf{x}(t)\}$); then the asymptotic variance of any system invariant estimated using the FC-CCA algorithm is a monotonically non increasing function of ν.

Role of f (future horizon)

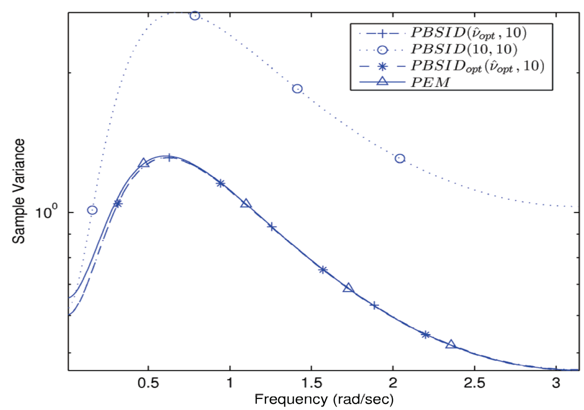
Theorem [Chiuso, TAC 2007, Submitted]

Assume the input process is such that the rows of X_t^u are, up to $o_P(1/\sqrt{N})$ terms, linear combinations of the rows of X_t (or equivalently span$\{\mathbf{x}^u(t)\} =: \mathcal{X}_t^u \subseteq \mathcal{X}_t :=$ span$\{\mathbf{x}(t)\}$); then the asymptotic variance of any system invariant estimated using the FC-CCA algorithm is a monotonically non increasing function of ν.

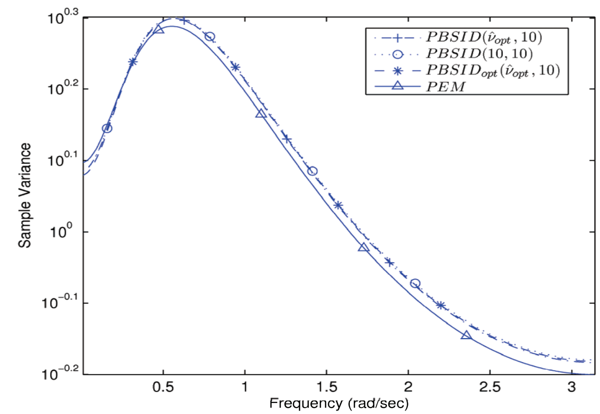
REMARK: FC-CCA as. eq. to PBSID,SSARX

Estimation of the "optimal" f using asymptotic variance formulas

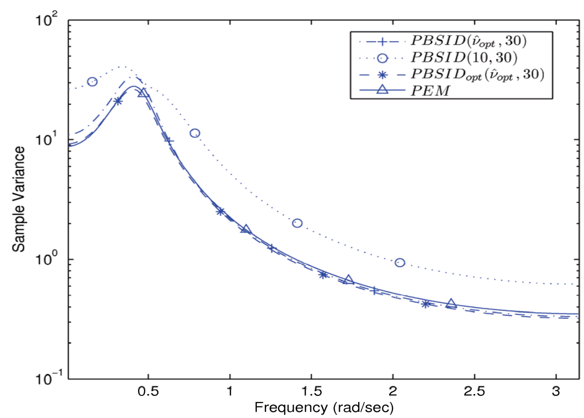
$\alpha = -0.9$



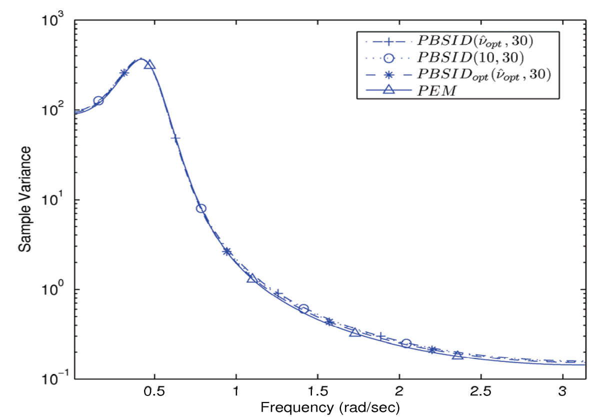
$\alpha = -0.2$



Sensitive to ν



Not sensitive to ν



Remark

- It was claimed in *[Larimore 04]* that the ADAPTx software is efficient as p and f go to infinity with the number of data.
- The plots in the previous slides show that indeed this is not true.
- Hence the quest for an efficient subspace procedure (if any) is still open.

CONCLUSIONS

- SUBSPACE ID is (ALMOST) a jungle of algorithms

  BUT

  Stochastic realization theory provides a unifying framework which has proven useful in deriving/comparing algorithms.

- Statistical analysis has been developed a lot since the late 90's. (Variance computation is less intensive than algorithm itself).

  Variance formulas provide also a natural tool to compare algorithms directly on data

  (we have used it to select the ``optimal'' f)

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

OPEN QUESTIONS!

- Statistically efficient procedures *(WHERE DOES STOCH. REALIZATION FAIL?)*
  - *Choice of f ?*
  - *Undermodeling?*
- Performance with finite data (comparison, when available, holds only asymptotically)
  - *Quality of the asymptotic approximation as a function of location in parameter space*
- Relation with ARX + model reduction
- Enforcing structure *(important for MIMO application when number of inputs and output is very large (50-100!))…*
- ………

Main References

CHECK OUT:

www.dei.unipd.it/~chiuso

THANK YOU