STATE SPACE MODELS OF STOCHASTIC SYSTEMS

GIORGIO PICCI

Department of Information Engineering,
Università di Padova, Italy

Scuola di Dottorato March 2005

STATE SPACE MODELS OF DETERMINISTIC SYSTEMS

$\{\mathbf{y}(t)\}$ m-dimensional output $\{\mathbf{u}(t)\}$ r-dimensional input

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t) \end{cases} , \quad \begin{array}{l} \mathbf{x}(t_0) = \mathbf{x}_0 \\ t \geq t_0 \end{array}$$

$\{\mathbf{x}(t)\}$ n-dimensional **state** of the system. A, B, C, D COSTANT system parameters. May be time-varying (but known).

REACHABILITY $\quad \mathrm{rank}\left[BAB \ldots, A^{n-1}B\right] = n$

OBSERVABILITY $\quad \mathrm{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} = n$

THEOREM 1 *The model is MINIMAL (*$\dim \mathbf{x}(t)$ *as small as possible) iff the system is* **reachable and observable***.*

TRANSFER FUNCTION

$$F(z) = C\,[zI - A]^{-1}B + D$$

$m \times r$ matrix. Elements $F_{ij}(z)$ are **proper rational functions of** z

$$F(z) = C\frac{\mathrm{Adj}\,[zI - A]}{\det\,[zI - A]}B + D$$

Characteristic Polynomial:

$$\Delta(z) = \det\,[zI - A]$$

$$\Delta(p_k) = 0 \quad \Leftrightarrow \quad \text{POLES OF } W(z)$$

BIBO STABILITY (Bounded inputs $\Rightarrow$ Bounded outputs) $\quad \Leftrightarrow \quad |p_k| < 1$

Z- (Fourier) TRANSFORM

$$\hat{f}(z) := \sum_{t=-\infty}^{+\infty} f(t) z^{-1}, \qquad z \in \mathbb{C}$$

Frequency response $\theta = \omega T$, $\qquad F(e^{j\theta})$.

SIMILARITY TRANSFORMATION

Warning: use of incongruous units may lead to ill-conditioned models.

State can be transformed $\hat{x}(t) := Tx(t)$; T $n \times n$ nonsingular matrix

$$\hat{A} = T^{-1}AT, \quad \hat{B} = T^{-1}B \quad \hat{C} = CT$$

$$\begin{cases} \hat{\mathbf{x}}(t+1) &= \hat{A}\hat{\mathbf{x}}(t) + \hat{B}\hat{\mathbf{u}}(t) & \hat{\mathbf{x}}(t_0) = \hat{\mathbf{x}}_0 \\ \mathbf{y}(t) &= \hat{C}\hat{\mathbf{x}}(t) + D\mathbf{u}(t) \quad , & t \geq t_0 \end{cases}$$

has the same transfer function

$$F(z) = \hat{C}\left[zI - \hat{A}\right]^{-1}\hat{B} + D$$

SINGULAR VALUE DECOMPOSITION (SVD)

THEOREM 2 *Let $A \in \mathbb{R}^{m\times p}$ of rank $n \leq \min(m,p)$. Can find two orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{p\times p}$ and positive numbers $\{\sigma_1 \geq, \ldots, \geq \sigma_n\}$, the* **singular values** *of A, so that*

$$A = U\Delta V^\top \qquad \Delta = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}, \qquad \Sigma = \mathrm{diag}\{\sigma_1, \ldots, \sigma_n\}$$

Full-rank factorization of A

$$A = [u_1, \ldots, u_n]\,\Sigma\,[v_1, \ldots, v_n]^\top := U_n \Sigma V_n^\top$$

where U_n, V_n submatrices of U, V keeping only the first n columns

$$U_n^\top U_n = I_n = V_n^\top V_n$$

$$Ax = \sum_{k=1}^{n} u_k \, \sigma_k \langle v_k \, , x \rangle$$

$U = [u_1, \ldots, u_m]$ = normalized eigenvectors of $AA^\top$;

$V := \left[v_1, \ldots, v_p\right]$ normalized eigenvectors of $A^\top A$.

$\{\sigma_1^2 \geq, \ldots, \geq \sigma_n^2\}$ (non zero) eigenvalues of $AA^\top$ (or of $A^\top A$).

MATRIX NORMS

2- norm of $A \in \mathbb{R}^{m \times p}$ Let $\|x\|$ be the Euclidean norm.

$$\|A\|_2 := \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_1 \quad (\sigma_{MAX}(A))$$

The **Frobenius norm** $\|A\|_F$ is

$$\|A\|_F^2 = \sum_{i,j} a_{i,j}^2 = \sigma_1^2 + \ldots + \sigma_n^2$$

Condition number

$$\kappa(A) = \frac{\|A\|_2}{\|A^{-1}\|_2} = \frac{\sigma_{MAX}(A)}{\sigma_{MIN}(A)}$$

USEFUL FEATURES OF SVD

Range and Nullspace of A:

$$\mathrm{Im}\,(A) = \mathrm{Im}\,(U_n), \qquad Ax = 0 \Leftrightarrow x \in \mathrm{span}\left(\left[v_{n+1}, \ldots, v_p\right]\right) = \mathrm{Im}\,V_n^{\perp}$$

Approximation properties

$$A_k := \sum_{i=1}^{k} \sigma_i u_i v_i^{\top}, \qquad k \le n$$

is the best approximant of rank k of A

$$\begin{aligned} \min_{\mathrm{rank}(B)=k} \|A - B\|_2 &= \|A - A_k\|_2 = \sigma_{k+1} \\ \min_{\mathrm{rank}(B)=k} \|A - B\|_F^2 &= \|A - A_k\|_F^2 = \sigma_{k+1}^2 + \ldots + \sigma_w^2 \end{aligned}$$

BALANCING

Assume: Eigenvalues of A strictly less than 1: $|\lambda(A)| < 1$, (A, B) reachable + (C, A) observable.

$$\Pi := \sum_{0}^{+\infty} A^k B B^\top (A^\top)^k, \qquad \Omega := \sum_{0}^{+\infty} (A^\top)^k C^\top C A^k$$

Reachability and Observability Gramians, solutions of the dual LYAPUNOV EQUATIONS

$$\begin{aligned} \Pi &= A \Pi A^\top + B B^\top \\ \Omega &= A^\top \Omega A + C^\top C \end{aligned}$$

THEOREM 3 *Assume the eigenvalues of A are strictly less than 1.* **System is reachable** *if and only if $\Pi > 0$.* **System is observable** *if and only if $\Omega > 0$. If both hold the model is MINIMAL ($\dim \mathbf{x}(t)$ as small as possible).*

INTERPRETATION OF THE GRAMIANS

Assume we can use only finite energy controls:

$$\|u\|_2^2 := \sum_{k=0}^{+\infty} u(k)^\top u(k) \leq 1$$

Energy of the state $\mathbf{x}(0) = \sum_0^{+\infty} A^k B\mathbf{u}(-k) := \mathbb{R}\mathbf{u}$

$$\max_{\|\mathbf{u}\|\leq 1} \frac{\|\mathbf{x}(0)\|^2}{\|\mathbf{u}\|^2} = \max_{\|\mathbf{u}\|\leq 1} \frac{\langle \mathbf{u}, \mathbb{R}^*\mathbb{R}\mathbf{u}\rangle}{\|\mathbf{u}\|^2} = \|\mathbb{R}^*\mathbb{R}\|_2 = \|\Pi\|_2 = \lambda_{\max}^2(\Pi)$$

Diagonalize:

$$U_c^\top \Pi U_c \Rightarrow \text{diag}\{\lambda_{c,1}^2, \ldots, \lambda_{c,n}^2\} \qquad \lambda_{c,1}^2 \geq, \ldots, \geq \lambda_{c,n}^2 > 0$$

Change coordinates $\mathbf{x}_c(0) := U_c^\top \mathbf{x}(0)$

Along the k-th eigenvector the maximum energy gain is $\lambda_{c,k}^2$

Energy ratios of the (orthogonal) state components

$$\frac{\|x_{1c}(0)\|}{\|x_{nc}(0)\|} = \frac{\lambda_{1c}}{\lambda_{nc}} \geq \ldots \geq \frac{\|x_{n-,1c}(0)\|}{\|x_{nc}(0)\|} = \frac{\lambda_{n-1,c}}{\lambda_{nc}}$$

$\frac{\lambda_{1c}}{\lambda_{nc}}$ may be large: the effect of the input on certain directions in the state space nearly invisible $\Rightarrow$ BAD CONDITIONING!

INTERPRETATION OF THE GRAMIANS (Cont.)

Dual meaning of the observability Gramian: Maximal L^2-energy ($\|\mathbf{y}\|_2$) of the output $\mathbf{y}(t) = CA^t\mathbf{x}(0)$ for $\|\mathbf{x}(0)\| \leq 1$: maximum singular value of Ω.

Diagonalization:

$$U_o^\top \Omega U_o \Rightarrow \mathrm{diag}\{\lambda_{o,1}^2, \ldots, \lambda_{o,n}^2\} \qquad \lambda_{o,1}^2 \geq, \ldots, \geq \lambda_{o,n}^2 > 0$$

Change coordinates $\mathbf{x}_o(t) := U_o^\top \mathbf{x}(t)$

Energy of the (orthogonal) state components (for $t \to \infty$)

$$\frac{\|y_{1,o}(0)\|}{\|y_{n,o}(0)\|} = \frac{\lambda_{1,o}}{\lambda_{n,o}} \geq \ldots \geq \frac{\|y_{n-1,o}(0)\|}{\|y_{n,o}(0)\|} = \frac{\lambda_{n-1,o}}{\lambda_{no}}$$

$\frac{\lambda_{1,o}}{\lambda_{n,o}}$ may be large: the effect of some states nearly invisible $\Rightarrow$ BAD CONDITIONING

(INTERNALLY) BALANCED MODELS

Changing bases can make things better

$$\hat{\Pi} = T^{-1}\Pi T^{-T}, \qquad \hat{\Omega} = T^{\top}\Omega T$$

Definition: Linear system in **Balanced form** if both $\hat{\Pi}$ and $\hat{\Omega}$ are **diagonal and equal**.

THEOREM 4 *Every linear model with* $|\lambda(A)| < 1$*,* (A,B) *reachable +* (C,A) *observable can be transformed to balanced form.*

ALGORITHM :

1. Compute Π and Ω, solutions of the two dual Lyapunov equations.

2. Compute the SVD

$$\Omega = U \Lambda_o U^\top$$

where Λ_o is the diagonal matrix of eigenvalues of Ω

3. Change basis $T_1 := \Lambda_o^{-1/2} U^\top$ so that $\hat{\Omega} = I$; compute

$$\hat{\Pi} = U \Lambda_o^{1/2} \Pi \Lambda_o^{1/2} U^\top$$

4. Compute the SVD

$$\hat{\Pi} = V \Lambda^2 V^\top$$

where Λ^2 is diagonal matrix with the (ordered) eigenvalues of $\hat{\Pi}$

5. Second change of basis defined by $T_2 := V\Lambda^{1/2}$ so as to make $\bar{\Pi} := T_2^{-1}\hat{\Pi}T_2^{-T} = \Lambda$, diagonal.

   With this change of basis

$$\bar{\Omega} = T_2^{\top}\hat{\Omega}T_2 = \Lambda^{1/2}V^{\top}IV\Lambda^{1/2} = \Lambda$$

The Gramians are diagonal and equal $\bar{\Pi} = \bar{\Omega} = \Lambda$

MATLAB

```
BALREAL  Balanced state-space realization and model reductior
  [Ab,Bb,Cb] = BALREAL(A,B,C) returns a balanced state-space
  realization of the system (A,B,C).

  [Ab,Bb,Cb,G,T] = BALREAL(A,B,C) also returns a vector G cor
  the diagonal of the gramian of the balanced realization, ar
  matrix T, the similarity transformation used to convert (A,
  to (Ab,Bb,Cb).  If the system (A,B,C) is normalized properl
  small elements in gramian G indicate states that can be rem
  reduce the model to lower order.
```

SINGULAR VALUES OF A LINEAR SYSTEM

The diagonal matrix Λ is a **system invariant** (does not change if basis is changed). Input-output map (from zero initialconditions)

$$\mathbf{y}(t) = \sum_{k=0}^{t-1} CA^{k-1}B\,\mathbf{u}(t-k) + D\mathbf{u}(t)$$

In matrix form $y = \mathbb{H}u$, where

$$\mathbb{H} := \begin{bmatrix} CB & CAB & CA^2B & \dots \\ CAB & CA^2B & CA^3B & \dots \\ CA^2B & CA^3B & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

Λ is diag of the **singular values of the Hankel matrix** $\mathbb{H}$ of the system

$$SVD(\mathbb{H}) = U \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} V^\top$$

MODEL REDUCTION
BY BALANCED TRUNCATION

How to best approximate a “Large” model (assumed stable + reach + obs)

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t)+B\mathbf{u}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t)+D\mathbf{u}(t) \end{cases}$$

Bring it to balanced form. Let Λ be partitioned

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}$$

Λ_2 $n_2 \times n_2$ made of small singular values ($\Lambda_1 >> \Lambda_2$)

BALANCED TRUNCATION

Ideally: Best rank n_1 approximation of $\mathbb{H}$ ....

$$\left\{ \begin{array}{lll} \begin{bmatrix} \mathbf{x}_1(t+1) \\ \mathbf{x}_2(t+1) \end{bmatrix} & = & \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) & = & \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + D\mathbf{u}(t) \end{array} \right.$$

$$\simeq$$

$$\left\{ \begin{array}{lll} \mathbf{x}_1(t+1) & = & A_{11}\mathbf{x}_1(t) + B_1\mathbf{u}(t) \\ \mathbf{y}(t) & = & C_1\mathbf{x}_1(t) + D\mathbf{u}(t) \end{array} \right.$$

N.B. STABILITY, REACHABILITY AND OBSERVABILITY **ARE PRESERVED**.

STATE-SPACE MODELS OF RANDOM SIGNALS

$\mathbf{y} = \{\mathbf{y}(t,\omega)\}$ discrete-time m-dimensional **random signal** $t \in [t_0, +\infty)$.

Expected value: $\mathbb{E}\,\mathbf{y}(t) = 0 \quad \Leftrightarrow \quad \int_\Omega \mathbf{y}(t,\omega)\,dP = 0$ can be subtracted off. All random quantities **zero mean**.

STOCHASTIC STATE-SPACE MODEL

$$\begin{cases} \mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{w}(t) & \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{w}(t) \quad , & t \geq t_0 \end{cases}$$

A, B, C, D COSTANT matrices $\{\mathbf{w}(t)\}$ p-dimensional **white noise** process of variance

$$\mathbb{E}\,\mathbf{w}(t)\,\mathbf{w}(s)^\top = I_p\delta(t-s) \qquad \mathbb{E}\,\mathbf{x}_0\,\mathbf{w}(t)^\top = 0 \quad \forall t \geq t_0$$

Initial (random) data $\quad \mathbb{E}\,\mathbf{x}_0 = 0 \quad , \quad \mathrm{Var}\,\mathbf{x}_0 = \Sigma_0$

MINIMAL MODELS

REACHABILITY: $\quad \text{rank}\left[BAB\ldots,A^{n-1}B\right]=n$

OBSERVABILITY: $\quad \text{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \; = \; n$

are necessary but not enough for minimality.

$$\begin{cases} \mathbf{x}(t+1) & = & -a\mathbf{x}(t)+(1-a^2)\mathbf{w}(t) \\ \mathbf{y}(t) & = & \mathbf{x}(t)+a\mathbf{w}(t) \end{cases}$$

$\mathbf{y}(t)$ is white noise. Has a minimal representation of order $n=0$.

STATE SPACE MODELS (Cont.)

Unnormalized white inputs:

$$\mathbf{v}(t) := G\mathbf{w}(t) \quad , \qquad \mathbf{w}(t) := D\mathbf{w}(t) \quad ,$$

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t) + \mathbf{v}(t) & \qquad \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) &= C\mathbf{x}(t) + \mathbf{w}(t) \quad , & \qquad t \geq t_0 \end{cases}$$

$$Q := E\{\mathbf{v}(t)\mathbf{v}(t)^\top\} \qquad S := E\{\mathbf{v}(t)\mathbf{w}(t)^\top\} \qquad R := E\{\mathbf{w}(t)\mathbf{w}(t)^\top\}$$

$$E\left\{ \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix} \begin{bmatrix} \mathbf{v}(t)^\top \; \mathbf{w}(t)^\top \end{bmatrix} \right\} = \begin{bmatrix} Q & S \\ S^\top & R \end{bmatrix} \quad .$$

$\{\mathbf{v}(t)\}$ and $\{\mathbf{w}(t)\}$ in general *correlated* white noise processes

$$B = [\bar{B}\, 0], \qquad D = [0\, \bar{D}] \qquad S = BD^\top = 0$$

THE STATE PROCESS

$\{\mathbf{x}(t)\}$ is a wide-sense **Markov process**,

$$\hat{\mathbb{E}}\left[\mathbf{x}(t) \mid \mathbf{x}(\tau)\, \tau \leq s\right] = \hat{\mathbb{E}}\left[\mathbf{x}(t) \mid \mathbf{x}(s)\right] \quad , \qquad \forall t \geq s \quad ,$$

If $\{\mathbf{w}(t)\}$ and $\mathbf{x}_0$ jointly Gaussian, then $\{\mathbf{x}(t)\}$ is Gaussian and Markov in strict sense. **State Variance**

$$\Sigma(t) = \mathbb{E}\,\mathbf{x}(t)\,\mathbf{x}(t)^\top := \mathrm{Var}\left(\mathbf{x}(t)\right)$$

Satisfies a LYAPUNOV DIFFERENCE EQUATION

$$\Sigma(t+1) = A\Sigma(t)\,A^\top + BB^\top \quad , \qquad \Sigma(t_0) = \Sigma_0 \quad .$$

$$\Sigma_{\mathbf{x}}(t,s) = A^{t-s}\,\Sigma(s) \quad , \qquad t \geq s$$

SECOND-ORDER DESCRIPTION

Joint covariances of $\{\mathbf{y}(t)\}$ and $\{\mathbf{x}(t)\}$ are completely determined by the model!

Output Covariance $\Sigma_{\mathbf{y}}(t,s) = \mathbb{E}\,\mathbf{y}(t)\,\mathbf{y}(s)^\top$

$$\Sigma_{\mathbf{x}}(t,s) = \begin{cases} A^{t-s}\,\Sigma(s) & t \geq s \\ \Sigma(t)\,(A^\top)^{s-t} & t \leq s \end{cases}$$

$$\Sigma_{\mathbf{y}}(t,s) = \begin{cases} CA^{t-s-1}\,G(s) & t > s \\ C\,\Sigma(t)\,C^\top + DD^\top & t = s \\ G(t)^\top\,(A^\top)^{s-t-1}C^\top & t < s \end{cases}$$

$$G(s) := A\Sigma(s)\,C^\top + BD^\top$$

ASYMPTOTIC STATIONARITY

Definition: $\{\mathbf{y}(t)\}$ is *asymptotically stationary* if for $t - t_0 \to \infty$, $\Sigma_{\mathbf{y}}(t,s)$, $t,s \geq t_0$, tends to depend on the difference $t - s$.

If A **(as.) stable** $|\lambda(A)| < 1$ then $\{\mathbf{x}(t)\}$ and $\{\mathbf{y}(t)\}$ for $t - t_0 \to +\infty$, jointly asympt. stationary

$$\Sigma_{\mathbf{x}}(t-s) = A^{t-s}\,\bar{\Sigma} \quad , \qquad t \geq s \quad ,$$

$$\Sigma_{\mathbf{y}}(t-s) = \begin{cases} CA^{t-s-1}\,\bar{G} & t > s \\ C\bar{\Sigma}C^\top + DD^\top & t = s \end{cases}$$

where $\bar{G} := A\bar{\Sigma}C' + BD'$ and $\bar{\Sigma} := \lim_{t-t_0 \to +\infty} \Sigma(t)$ satisfies the LYAPUNOV EQUATION

$$\bar{\Sigma} = A\bar{\Sigma}A^\top + BB^\top \quad .$$

$\bar{\Sigma}$, asympt. state variance, does not depend on the initial condition Σ_0.

THE LYAPUNOV EQUATION

FACT: Any two conditions imply the remaining one

i) (A,B) is reachable

ii) A is asymptotically stable

iii) The Lyapunov equation

$$X = AXA^\top + BB^\top$$

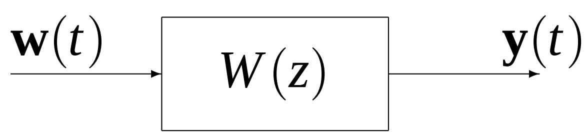
has a unique solution $P = P^\top > 0$

SHAPING FILTERS AND ARMA MODELS

Assume A **(as.) stable** i.e. $|\lambda(A)| < 1$ then $\{\mathbf{x}(t)\}$ and $\{\mathbf{y}(t)\}$ $t - t_0 \rightarrow +\infty$, **jointly asympt. stationary**. Effect of initial conditions disappears

$\{\mathbf{y}(t)\}$: response to normalized white noise process $\{\mathbf{w}(t)\}$ of a linear filter (*Shaping Filter*) with transfer function

$$W(z) = C(zI - A)^{-1}B + D$$



$W(z)$ is a rational matrix function. Can be written as a ratio of polynomial matrices

$$W(z) = D(z)^{-1}\, N(z);$$

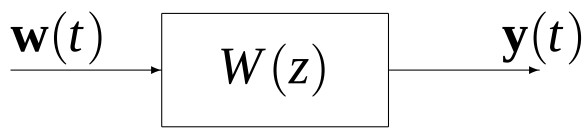
$$D(z) = Iz^{\nu} + \sum_{1}^{\nu} A_k z^{\nu-k} \qquad N(z) = N_0 z^{\nu} + \sum_{1}^{\nu} N_k z^{\nu-k}$$

$\{\mathbf{y}(t)\}$ may be described by a (multivariabile) **ARMA model**

$$\mathbf{y}(t) + \sum_{1}^{\nu} A_k \mathbf{y}(t-k) = N_0 \mathbf{w}(t) + \sum_{1}^{\nu} N_k \mathbf{w}(t-k) \quad .$$

WARNING: There are **many** ARMA model representations!

SHAPING FILTERS AND SPECTRUM



Wiener-Kintchine formula gives the **spectral density matrix** $\Phi(z)$ of $\{\mathbf{y}(t)\}$

$$\Phi(z) = W(z)\, W(z^{-1})^\top$$

Spectrum is a **rational function** of z.

Positivity: $$\Phi(e^{j\theta}) = W(e^{j\theta})\, W(e^{-j\theta})^\top \geq 0$$

SPECTRAL FACTORIZATION

FACT: The shaping filter $W(z)$ is a **spectral factor** of $\Phi(z)$

$$\Phi(z) = W(z)\, W(z^{-1})^{\top}$$

Conversely: modeling by SF is computing (rational) spectral factors from given (rational) spectrum.

Minimal spectral factors (minimal Mc Millan degree) $\Rightarrow$ Minimal state space models.

FREQUENCY-DOMAIN COMPUTATIONS

Computing the spectrum of $\mathbf{y}$:

$$\Phi(z) = \begin{bmatrix} C(zI-A)^{-1} & I \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B^\top D^\top \end{bmatrix} \begin{bmatrix} (z^{-1}I - A^\top)^{-1} C^\top \\ I \end{bmatrix} \quad .$$

$$\begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B^\top D^\top \end{bmatrix} = \begin{bmatrix} BB^\top & BD^\top \\ DB^\top & DD^\top \end{bmatrix} := \begin{bmatrix} Q & S \\ S^\top & R \end{bmatrix} = E\left\{ \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix} \begin{bmatrix} \mathbf{v}(t)^\top \ \mathbf{w}(t)^\top \end{bmatrix} \right\}$$

Scalar process: spectrum from ARMA :

$$\Phi(e^{j\theta}) = \frac{N(e^{j\theta})N(e^{j\theta})^*}{D(e^{j\theta})D(e^{j\theta})^*} = \left| \frac{N(e^{j\theta})}{D(e^{j\theta})} \right|^2$$

SUMMARY: THREE CLASSES OF MODELS

State Space Models: $\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{w}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t) \end{cases}$

Shaping filters/ARMA : $\mathbf{y}(t) + \sum_1^{\nu} A_k \mathbf{y}(t-k) = N_0 \mathbf{w}(t) + \sum_1^{\nu} N_k \mathbf{w}(t-k)$

Spectrum: $\Phi(z) = W(z)\, W(z^{-1})^{\top};$ $\qquad W(z) = C(zI - A)^{-1} B + D = D(z)^{-1} N(z)$

MODELS FROM COVARIANCE

Assume **given** the covariance function $\Sigma_{\mathbf{y}}(k) \;\; k = 1, 2, \ldots$

PROBLEM (stochastic realization): From $\{\Sigma_{\mathbf{y}}(k) \;\; k = 0, 1, 2, \ldots\}$ compute $\{A, B, C, D\}$ of a minimal state space model of $\mathbf{y}$.

NECESSARY CONDITIONS

Form the **Hankel Matrix** of $\Sigma_{\mathbf{y}}$

$$\mathbb{G} := \begin{bmatrix} \Sigma_{\mathbf{y}}(1) & \Sigma_{\mathbf{y}}(2) & \Sigma_{\mathbf{y}}(3) & \dots \\ \Sigma_{\mathbf{y}}(2) & \Sigma_{\mathbf{y}}(3) & \Sigma_{\mathbf{y}}(4) & \dots \\ \Sigma_{\mathbf{y}}(3) & \Sigma_{\mathbf{y}}(4) & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

If $\Sigma_{\mathbf{y}}$ is generated by a linear state space model

$$\Sigma_{\mathbf{y}}(k) \qquad = CA^{k-1}\,\bar{G} \qquad \qquad \begin{cases} \bar{G} & := & A\bar{\Sigma}C^{\top} + BD^{\top} \\ \bar{\Sigma} & = & A\bar{\Sigma}A^{\top} + BB^{\top} \end{cases}$$

$$\Sigma_{\mathbf{y}}(0) \quad = C\bar{\Sigma}C^{\top} + DD^{\top} \qquad \qquad \text{For}\, k = 0$$

Must admit a **factorization** of the type

$$\mathbb{G} = \begin{bmatrix} C\overline{G} & CA\overline{G} & CA^2\overline{G} & \dots \\ CA\overline{G} & CA^2\overline{G} & CA^3\overline{G} & \dots \\ CA^2\overline{G} & CA^3\overline{G} & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \begin{bmatrix} \overline{G} & A\overline{G} & A^2\overline{G} & \dots \end{bmatrix}$$

Necessary condition: $\operatorname{rank} \mathbb{G} = n$

Positivity of the function $k \to CA^{k-1}\,\overline{G} \quad (\simeq \Sigma_{\mathbf{y}}(k))$

TWO-STEPS SOLUTION

Step 1: From a *finite submatrix* $\mathbb{G}_N$ of $\mathbb{G}$, of $\mathrm{rank} = n$ compute $\{A, C, \bar{G}\}$

ALGORITHM (HO-KALMAN) :

1. Compute the SVD

$$\mathbb{G}_N = U \Delta V^\top \qquad \Delta = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

   where $\Sigma := \mathrm{diag}\{\sigma_1, \ldots, \sigma_n\}$ is the diagonal matrix of **nonzero singular values**.

2. Rank n factorization

$$\mathbb{G}_N = U_n \Sigma V_n^\top = U_n \Sigma^{1/2} \, \Sigma^{1/2} V_n^\top := \Omega \, \bar{\Omega}$$

3. Impose

$$\Omega = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{N-1} \end{bmatrix} \qquad \bar{\Omega} = \begin{bmatrix} \bar{G} & A\bar{G} & A^2\bar{G} & \ldots & A^{N-1}\bar{G} \end{bmatrix}$$

and solve for $C, A, \bar{G}$.

THE HO-KALMAN ALGORITHM (CONT'D)

Computing A:

$$\Omega = \begin{bmatrix} C \\ (\downarrow \Omega) \end{bmatrix} = \begin{bmatrix} (\uparrow \Omega) \\ CA^{N-1} \end{bmatrix}; \qquad (\downarrow \Omega) = (\uparrow \Omega)A \quad \Rightarrow \quad A = (\uparrow \Omega)^{-L}(\downarrow \Omega)$$

$$\bar{\Omega} = \begin{bmatrix} \bar{G} & \overrightarrow{\bar{\Omega}} \end{bmatrix} = \begin{bmatrix} \overleftarrow{\bar{\Omega}} & A^{N-1}\bar{G} \end{bmatrix}; \qquad A\overrightarrow{\bar{\Omega}} = \overleftarrow{\bar{\Omega}} \quad \Rightarrow \quad A = (\overleftarrow{\bar{\Omega}})\,(\overrightarrow{\bar{\Omega}})^{-R}.$$

Found a minimal state space model for the **Causal part** of the spectrum $\Phi(z) = \Phi_+(z) + \Phi_+(z^{-1})^\top$

$$\{A, C, \bar{G}, \Sigma_{\mathbf{y}}(0)\} \Rightarrow \Phi_+(z) = C\,[zI - A]^{-1}\,\bar{G} + 1/2\Sigma_{\mathbf{y}}(0)$$

STOCHASTIC REALIZATION ALGORITHM

Step 2: From the spectrum $(\Phi_+(z))$ to a state-space model

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t)+B\mathbf{w}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t)+D\mathbf{w}(t) \end{cases},$$

$A, C,$ can be taken the same! Just need to compute (B, D).

Recall: $W(z) := C(zI-A)^{-1}B+D$ is a **spectral factor** $\Phi(z) = W(z)W(1/z)^\top$

ALGORITHM:

Given $(A, C, \bar{G}, \frac{1}{2}\Sigma_{\mathbf{y}}(0))$ a minimal realization of $\Phi_+(z)$,

1. Find $n \times n$ matrices $P = P^\top$ solving the *Linear Matrix Inequality*

$$M(P) := \begin{bmatrix} P - APA^\top & \bar{G}^\top - APC^\top \\ \bar{G} - CPA^\top & \Sigma_{\mathbf{y}}(0) - CPC^\top \end{bmatrix} \geq 0$$

2. Compute full column rank matrix factors $\begin{bmatrix} B \\ D \end{bmatrix}$ of $M(P)$,

$$M(P) = \begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B^\top & D^\top \end{bmatrix},$$

3. $W(z) = C(zI - A)^{-1}B + D$. is a minimal spectral factor (a minimal shaping filter). And conversely....

All symmetric solutions P of the **LMI** are positive definite : State variance, solution of $P - APA^\top = BB^\top$

SOME SPECIAL STOCHASTIC MODELS

Minimal state-space models $\Leftrightarrow$ set of solutions $\mathcal{P}$ ofthe LMI.

If $\Sigma_{\mathbf{y}}(0) - CPC^\top > 0$, easy to see that $M(P) \geq 0$ iff P satisfies the **Algebraic Riccati Inequality**

$$P - APA^\top - (\bar{G}^\top - APC^\top)(\Sigma_{\mathbf{y}}(0) - CPC^\top)^{-1}(\bar{G} - CPA^\top) \geq 0.$$

In particular, if P satisfies the **Algebraic Riccati Equation (ARE)**

$$P = APA^\top + (\bar{G}^\top - APC^\top)(\Sigma_{\mathbf{y}}(0) - CPC^\top)^{-1}(\bar{G} - CPA^\top),$$

the corresponding $W(z)$ is **square** $m \times m$.

FACT:

Two **special** solutions of the **ARE**: P_-, P_+, such that $P_- \leq P \leq P_+, \quad \text{for all } P \in \mathcal{P}$,

$$P_- \Rightarrow \begin{bmatrix} B_- \\ D_- \end{bmatrix} \Rightarrow W_-(z) = C(zI - A)^{-1} B_- + D_-$$

The **minimum phase** model: zeros in $\{|z| \leq 1\}$ i.e. **Causal inverse**

$$P_+ \Rightarrow \begin{bmatrix} B_+ \\ D_+ \end{bmatrix} \Rightarrow W_+(z) = C(zI - A)^{-1} B_+ + D_+$$

The **maximum phase** model: zeros in $\{|z| \geq 1\}$ i.e. **Anticausal inverse**

NB: $w(t) = W(z)^{-1} y(t)$ tells how to construct the white noise input!

WARNING: with real data the parameters $\{A, C, \bar{G}\}$ computed by Ho-Kalman may not satisfy the **positivity condition** that $\Phi_+(z)$ must be the causal part of a power spectrum

$$\Phi_+(e^{j\theta}) + \Phi_+(e^{-j\theta})^\top = W(e^{j\theta})\, W(e^{-j\theta})^\top \geq 0$$

This prevents solvability of the Riccati equation.

THE KALMAN FILTER

PROBLEM: Estimate the state of the linear model

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t) + \mathbf{v}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + \mathbf{w}(t) \end{cases} \quad , \qquad t \geq t_0 \quad ,$$

given past measurements of $\{\mathbf{y}(t)\}$ (m-dimensional) up to time t.

$$\mathbb{E}\left\{ \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix} \begin{bmatrix} \mathbf{v}(s)^\top, \mathbf{w}(s)^\top \end{bmatrix} \right\} = \begin{bmatrix} Q & S \\ S^\top & R \end{bmatrix} \delta(t-s) \qquad R > 0$$

$$\mathbb{E}\,\mathbf{x}_0 = \mu_0 \quad , \qquad \mathrm{Var}\{\mathbf{x}_0\} = P_0 \quad .$$

KALMAN FILTER (PREDICTOR):

$$\hat{\mathbf{x}}(t+1 \mid t) = A\hat{\mathbf{x}}(t \mid t-1) + G(t)\mathbf{e}(t)$$

The one-step output predictor: $\hat{\mathbf{y}}(t \mid t-1) = C\hat{\mathbf{x}}(t \mid t-1)$.
Innovation process $\mathbf{e}(t) := \mathbf{y}(t) - C\hat{\mathbf{x}}(t \mid t-1)$ is **white noise !**

THE KALMAN FILTER (CONTD)

The Kalman gain $K(t)$

$$K(t) := \left[AP(t \mid t-1)C^\top + S\right] \Lambda(t)^{-1}$$

Need **error covariance matrix**, $P(t \mid t-1) = \mathrm{Var}\{\tilde{\mathbf{x}}(t \mid t-1) := \mathbf{x}(t) - \hat{\mathbf{x}}(t \mid t-1)\}$

$$P(t \mid t-1) = \mathbb{E}\,\tilde{\mathbf{x}}(t \mid t-1)\,\tilde{\mathbf{x}}(t \mid t-1)^\top = P - \hat{P}(t)$$

Innovation covariance $\Lambda(t) = \mathbb{E}\,\mathbf{e}(t)\mathbf{e}(t)^\top$,

$$\Lambda(t) = CP(t \mid t-1)\,C^\top + R$$

Riccati Equation for the error covariance $P(t+1 \mid t)$,

$$P(t+1 \mid t) = AP(t \mid t-1)A^\top - K(t)\Lambda(t)K(t)^\top + Q$$

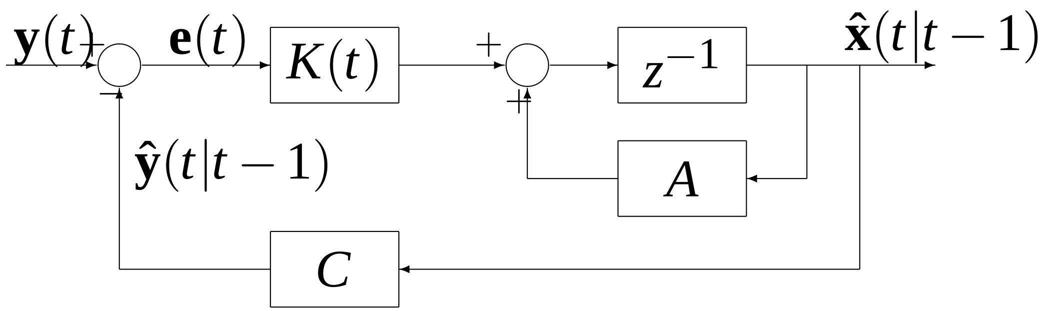
RICCATI EQUATION

Equivalent form in terms of covariance of $\hat{\mathbf{x}}(t \mid t-1)$

$$\hat{P}(t+1) = A\hat{P}(t)A^\top + (\overline{G}^\top - A\hat{P}(t)C^\top)(\Sigma_{\mathbf{y}}(0) - C\hat{P}(t)C^\top)^{-1}(\overline{G} - C\hat{P}(t)A^\top),$$

THE STEADY-STATE KALMAN FILTER

The Kalman filter is an asymptotically stable feedback system!!



Closed loop matrix $\Gamma(t) = A - K(t)C$, for $t - t_0 \rightarrow \infty$ asymptotically stable under very mild conditions

S.S. KALMAN FILTER IS ALSO A STATE MODEL FOR $\mathbf{y}$!

$$\begin{cases} \hat{\mathbf{x}}(t+1) & = & A\hat{\mathbf{x}}(t) + K_\infty \mathbf{e}(t) \\ \mathbf{y}(t) & = & C\hat{\mathbf{x}}(t) + \mathbf{e}(t) \end{cases}$$

Steady state solution of the RE, $\lim_{t-t_0\to\infty}\hat{P}(t)=P_\infty$. Solution of the ARE

$$P_\infty = AP_\infty A^\top + (\overline{G}^\top - AP_\infty C^\top)(\Sigma_{\mathbf{y}}(0) - CP_\infty C^\top)^{-1}(\overline{G} - CP_\infty A^\top),$$

SAME RICCATI EQUATION OF STOCHASTIC REALIZATION !!! $\Rightarrow$ S.S. KALMAN FILTER MODEL

$$\begin{cases} \hat{\mathbf{x}}(t+1) &= A\hat{\mathbf{x}}(t) + K_\infty \mathbf{e}(t) \\ \mathbf{y}(t) &= C\hat{\mathbf{x}}(t) + \mathbf{e}(t) \end{cases}$$

WHAT KIND OF MODEL IS THE STEADY-STATE KALMAN FILTER?

Closed loop matrix $\Gamma_\infty = A - K_\infty C$, of the steady state KF is asymptotically stable under very mild conditions.

Inverse system (whitening filter)

$$\begin{cases} \hat{\mathbf{x}}(t+1) &= [A - K_\infty C]\,\hat{\mathbf{x}}(t) + K_\infty \mathbf{y}(t) \\ \mathbf{e}(t) &= -C\hat{\mathbf{x}}(t) + \mathbf{y}(t) \end{cases}$$

Has eigenvalues inside the unit circle. So SSKF is the MINIMUM PHASE MODEL!!

$$P_\infty = P_-$$

THE BACKWARD KALMAN FILTER

PROBLEM: Estimate the state of the linear model

$$\begin{cases} \mathbf{x}(t+1) & = & A\mathbf{x}(t) + B\mathbf{w}(t) \\ \mathbf{y}(t) & = & C\mathbf{x}(t) + D\mathbf{w}(t) \end{cases} ,$$

given **future** measurements of $\{\mathbf{y}(t)\}$ (m-dimensional) from time t on.

Backward models......

(EARLY) SUBSPACE IDENTIFICATION FOR TIME SERIES [Aoki]

Given observed data (zero mean)

$$\{y_t \mid t = 0, 1, 2, \ldots, N\}$$

Algorithm:

1. Form covariance estimates

$$\Lambda_k = \frac{1}{N} \sum_{t=0}^{N-k} y_{t+k} y_t^\top \quad (\to \ \Sigma_{\mathbf{y}}(k))$$

2. Form the Hankel matrix

$$\mathbb{H}_\Lambda := \begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 & \dots & \Lambda_\nu \\ \Lambda_2 & \Lambda_3 & \Lambda_4 & \dots & \Lambda_{\nu-1} \\ \Lambda_3 & \Lambda_4 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \Lambda_{\nu+1} & \dots & \dots & \dots & \Lambda_{2\nu} \end{bmatrix}$$

Choose ν “large enough” $(\nu \geq n)$.

3. Compute the SVD

$$\mathbb{H}_\Lambda = U \Delta V^\top \qquad \Delta = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}$$

where $\Sigma := \mathrm{diag}\{\sigma_1, \dots, \sigma_n\}$ is the diagonal matrix of **dominant singular values**. $\Sigma_2 \simeq 0$ are neglected.

4. Rank n factorization

$$\mathbb{H}_\Lambda \simeq U_n \Sigma_1 V_n^\top = U_n \Sigma_1^{1/2}\, \Sigma_1^{1/2} V_n^\top := \Omega\, \bar{\Omega}$$

5. Impose

$$\Omega = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^\nu \end{bmatrix} \qquad \bar{\Omega} = \begin{bmatrix} \bar{G} & A\bar{G} & A^2\bar{G} & \dots & A^{\nu-1}\bar{G} \end{bmatrix}$$

and get $C, \bar{G}$ by inspection. Compute A by solving $\quad (\downarrow \Omega) = (\uparrow \Omega) A$

$$A = (\uparrow \Omega)^{-L} (\downarrow \Omega) = (\uparrow U_n \Sigma_1^{1/2})^{-L} (\downarrow U_n \Sigma_1^{1/2}) = \Sigma_1^{-1/2} (\uparrow U_n)^\top (\downarrow U_n) \Sigma_1^{1/2}$$

INNOVATION MODEL IDENTIFICATION

From previuos step: $(A, C, \bar{G})$. Want $K,\ \Lambda = \mathbb{E}\,\mathbf{e}(t)\mathbf{e}(t)^\top$ in

$$\begin{cases} \hat{\mathbf{x}}(t+1) &= A\hat{\mathbf{x}}(t) + K\mathbf{e}(t) \\ \mathbf{y}(t) &= C\hat{\mathbf{x}}(t) + \mathbf{e}(t) \end{cases}$$

Solve the ARE (minimal solution $P = P^\top > 0$)

$$P = APA^\top + (\bar{G}^\top - APC^\top)(\Lambda_0 - CPC^\top)^{-1}(\bar{G} - CPA^\top),$$

$$K = \left[\bar{G}^\top - APC^\top\right] R(P)^{-1} \qquad R(P) = \Lambda_0 - CPC^\top$$

WARNING: with real data the parameters $\{A, C, \bar{G}\}$ computed by Ho-Kalman may not satisfy the **positivity condition** that $\Phi_+(z)$ must be the causal part of a power spectrum

$$\Phi_+(e^{j\theta}) + \Phi_+(e^{-j\theta})^\top = W(e^{j\theta})\, W(e^{-j\theta})^\top \geq 0$$

This prevents solvability of the Riccati equation.

Main drawback of the method: the estimates $\Lambda(k)$ in general rather poor!

SUBSPACE IDENTIFICATION FROM INFINITE/FINITE INPUT-OUTPUT DATA

OUTLINE OF THE NEXT LECTURES:

1. Some Hilbert space background
2. State construction for stationary processes. Canonical Correlation Analysis (CCA). Stochastic Balancing
3. Realization of stationary processes (no input) with infinite/finite data

4. Subspace algorithms for time series. Relation with to Ho-Kalman algorithm

5. State construction for stationary stoch. systems. Conditional Canonical Correlation Analysis (CCCA)

6. Finite interval realization of stationary stochastic systems with inputs

7. Subspace identification algorithms: CCA, N4SID, MOESP.

8. Numerical aspects

BASIC IDEA OF SUBSPACE IDENTIFICATION FOR TIME SERIES

Assume we can observe also a **state trajectory** $\{x_0, x_1, x_2, \ldots, x_N\}$ of the model, corrresponding to the data

$$\{y_0, y_1, y_2, \ldots, y_N\}, \quad y_t \in \mathbb{R}^m$$

Form the “tail” matrices $\mathbf{Y}_t$, $\mathbf{X}_t$,

$$\mathbf{Y}_t := [\, y_t, y_{t+1}, y_{t+2}, \ldots]$$
$$\mathbf{X}_t := [\, x_t, x_{t+1}, x_{t+2}, \ldots]$$

Every sample trajectory $\{y_t\}$, $\{x_t\}$ of the system must satisfy the model equations, so there exist $\{e_t\}$ s.t.

$$\begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} K \\ I \end{bmatrix} \mathbf{E}_t$$

SUBSPACE IDENTIFICATION OF TIME SERIES (cont'd)

$$\begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} K \\ I \end{bmatrix} \mathbf{E}_t$$

Linear Regression ! Solve by Least Squares :

$$\min_{A,C} \| \begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} \mathbf{X}_t \|$$

getting

$$\begin{bmatrix} \hat{A} \\ C \end{bmatrix}_N := \frac{1}{N} \begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} \mathbf{X}_t^\top \left\{ \frac{1}{N} \mathbf{X}_t \mathbf{X}_t^\top \right\}^{-1}$$

BASIC IDEA OF SUBSPACE IDENTIFICATION (cont'd)

Theorem: If the data are **second order ergodic**, and the inverse exists:

$$\lim_{N \to \infty} \widehat{\begin{bmatrix} A \\ C \end{bmatrix}}_N = \begin{bmatrix} A \\ C \end{bmatrix} \qquad (\dagger)$$

consistent estimate of A, C.

Proof: HOMEWORK!

FINITE DATA

Meaning of Finite Data: finite string of observed data

$$\{y_0, y_1, y_2, \ldots, y_N\}$$

N sufficiently large so that

$$\frac{1}{N+1}\sum_{t=0}^{N} y_{t+k} y_t^\top \qquad k = 1, 2, \ldots, T$$

is a "good approximation" of a *finite* set of covariance lags,

$$\{\Lambda(0), \Lambda(1), \ldots, \Lambda(T)\},$$

Need to bound T so that $T << N$. Rule of thumb is $T \simeq (1/50)N$

Equivalent to $\forall a, b \in \mathbb{R}^m$

$$\frac{1}{N+1}\sum_{t=0}^{N} a^\top y_{t+k} y_{t+j}^\top b \simeq a^\top \mathbb{E}\{\mathbf{y}(k)\mathbf{y}(j)^\top\} b \qquad |k-j| \leq T$$

For $N \to \infty$ the **sample covariances** $\simeq$ **true covariances**.

Assuming N "very large" numerical TAIL sequences same as **random vectors** !

$$\mathbf{Y}_t \Leftrightarrow \mathbf{y}(t) \qquad \frac{1}{N}\mathbf{Y}_t \mathbf{Y}_s^\top \simeq \mathbb{E}\{\mathbf{y}(t)\,\mathbf{y}(s)^\top\}$$

EXACTLTY THE SAME AS if we had a **finite sequence** of TRULY RANDOM vectors

$$\{\mathbf{y}(0), \mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(T)\},$$

extracted from $\mathbf{y}$. CAN PRETEND had observations of $\mathbf{y}$ on the **finite interval** $[0, T]$. SAME FORMULAS!

CONSTRUCTING THE STATE FROM FINITE DATA

Construct $\hat{\mathbf{x}}(t)$: **state of transient Kalman filter** on $[t_0, T]$:

$$\left\{ \begin{array}{rcl} \hat{\mathbf{x}}(t+1) & = & A\hat{\mathbf{x}}(t) + K(t)\hat{\mathbf{e}}(t) \\ \mathbf{y}(t) & = & C\hat{\mathbf{x}}(t) + \hat{\mathbf{e}}(t) \\ \hat{\mathbf{x}}(t_0) & = & 0 \end{array} \right.$$

Predictor of finite future based on finite past data :

$$\hat{\mathbf{y}}_t^+ := \mathbb{E}\left[\mathbf{y}_t^+ \mid \mathbf{y}_t^-\right] = \Gamma_k \hat{\mathbf{x}}(t) \qquad k = T - t$$

$$\hat{\mathfrak{X}}_t = \operatorname{span} \mathbb{E}\left[\mathbf{y}_t^+ \mid \mathbf{y}_t^-\right]$$

THE STATE BY CANONICAL CORRELATION ANALYSIS

Introduce **Finite past and future** at time t:

$$\mathbf{y}_t^- := \begin{bmatrix} \mathbf{y}(t_0) \\ \mathbf{y}(t_0+1) \\ \vdots \\ \mathbf{y}(t-1) \end{bmatrix} \simeq \mathbf{Y}_t^- := \begin{bmatrix} \mathbf{Y}_{t_0} \\ \mathbf{Y}_{t_0+1} \\ \vdots \\ \mathbf{Y}_{t-1} \end{bmatrix}$$

$$\mathbf{y}_t^+ := \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t+1) \\ \vdots \\ \mathbf{y}(T) \end{bmatrix} \simeq \mathbf{Y}_t^+ := \begin{bmatrix} \mathbf{Y}_{t} \\ \mathbf{Y}_{t+1} \\ \vdots \\ \mathbf{Y}_{T} \end{bmatrix}$$

CCA of finite future and past spaces $\simeq$ CCA of rowspaces of $\mathbf{Y}_t^-$ and $\mathbf{Y}_t^+$

CANONICAL CORRELATION ANALYSIS

CCA is an old concept in statistics. Given two finite-dimensional subspaces $\mathbf{A}, \mathbf{B}$ of zero-mean random variables of dimension n and m, one wants to find two special orthonormal bases say $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $\mathbf{A}$, and $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ for $\mathbf{B}$ such that

$$\mathbb{E}\{\mathbf{u}_k \mathbf{v}_h\} = \sigma_k \delta_{k,h}, \qquad k, h = 1, \dots, \min\{n, m\}$$

This is the same as asking that the correlation matrix of the two random vectors $\mathbf{u} := [\mathbf{u}_1, \dots, \mathbf{u}_n]'$ and $\mathbf{v} := [\mathbf{v}_1, \dots, \mathbf{v}_m]'$ made with the elements of the two bases, should be diagonal, i.e. assuming for example that $n \geq m$,

$$\mathbb{E}\{\mathbf{u}\mathbf{v}'\} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots \\ \vdots & & \ddots & \\ & & & \sigma_m \\ 0 \;\; 0 & \dots & 0 & 0 \end{bmatrix}$$

To make this choice of basis unique one further requires that all the σ_k's be nonnegative and ordered in decreasing magnitude.

That two orthonormal bases of this kind always exist follows by considering the singular value decomposition of the projection operator $\mathbb{E}^{\mathbf{A}}_{|\mathbf{B}}$.

Choosing as orthonormal basis in $\mathbf{A}$ and in $\mathbf{B}$ precisely the principal directions $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of $\mathbb{E}^{\mathbf{A}}_{|\mathbf{B}}$, one has

$$\mathbb{E}^{\mathbf{A}}_{|\mathbf{B}}\xi = \sum_{k=1}^{n} \sigma_k \langle \xi, \mathbf{v}_k \rangle \mathbf{u}_k$$

from which it is obvious that the two bases have the required properties. Uniqueness is guaranteed when and only when the singular values $\{\sigma_k\}$, which in this context are called *canonical correlation coefficients*, are all distinct.

CCA ALGORITHM

1. **Normalization:** Form $T_- := \frac{1}{N}\mathbf{Y}_t^-(\mathbf{Y}_t^-)^\top \qquad T_+ := \frac{1}{N}\mathbf{Y}_t^+(\mathbf{Y}_t^+)^\top$

   Compute (Cholesky) factors $T_- = L_- L_-^\top, \quad T_+ = L_+ L_+^\top$

$$\hat{\mathbf{Y}}_t^- := L_-^{-1}\mathbf{Y}_t^- \qquad \hat{\mathbf{Y}}_t^+ := L_+^{-1}\mathbf{Y}_t^+$$

2. **SVD :**

$$\frac{1}{N}\hat{\mathbf{Y}}_t^+(\hat{\mathbf{Y}}_t^-)^\top = \begin{bmatrix}\hat{U} & \tilde{U}\end{bmatrix} \begin{bmatrix}\hat{\Sigma} & 0 \\ 0 & \tilde{\Sigma}\end{bmatrix} \begin{bmatrix}\hat{V} & \tilde{V}\end{bmatrix}^\top$$

   Can be done (QSVD) without forming the Hankel matrix $\frac{1}{N}\hat{\mathbf{Y}}_t^+(\hat{\mathbf{Y}}_t^-)^\top$

   Order estimation: Choose n so that $\hat{\Sigma} \gg \tilde{\Sigma}$

3. **Canonical Variables**

$$\hat{\mathbf{X}}_t := \hat{U}^\top \hat{\mathbf{Y}}_t^- = \hat{U}^\top L_-^{-1} \mathbf{Y}_t^- \qquad \hat{\bar{\mathbf{X}}}_t := \hat{V}^\top \hat{\mathbf{Y}}_t^+ = \hat{V}^\top L_+^{-1} \mathbf{Y}_t^+$$

$\hat{\bar{\mathbf{X}}}_t$ basis for the *Backward Kalman filter*.

4. **Balancing of Canonical Variables**

$$\mathbf{Z}_t := \hat{\Sigma}^{1/2} \hat{\mathbf{X}}_t \qquad \bar{\mathbf{Z}}_t := \hat{\Sigma}^{1/2} \hat{\bar{\mathbf{X}}}_t \qquad \frac{1}{N} \mathbf{Z}_t \mathbf{Z}_t^\top = \hat{\Sigma} = \frac{1}{N} \bar{\mathbf{Z}}_t \bar{\mathbf{Z}}_t^\top$$

5. Repeat for $t = t + 1$ to get $\mathbf{Z}_{t+1}$ basis in $\hat{\mathcal{X}}_{t+1}$ and solve

$$\begin{bmatrix} \mathbf{Z}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \mathbf{Z}_t + \begin{bmatrix} K(t) \\ I \end{bmatrix} \hat{\mathbf{E}}_t$$

by Least-Squares.
N.B. $\mathbf{Z}_{t+1}$ must be a *coherent basis with* $\mathbf{Z}_t$.

THE BACKWARD K.F. AND THE $\bar{G} = \bar{C}$ PARAMETERS

$$\begin{cases} \hat{\bar{\mathbf{x}}}(t-1) = A^\top \hat{\bar{\mathbf{x}}}(t) + \bar{K}(t)\bar{\mathbf{e}}(t-1) \\ \mathbf{y}(t-1) = \bar{C}\hat{\bar{\mathbf{x}}}(t) + \bar{\mathbf{e}}(t-1) \\ \bar{\mathbf{x}}(T) = 0 \end{cases}$$

Predictor of past based on finite future data :

$$\hat{\mathbf{y}}_t^- := \mathbb{E}\left[\mathbf{y}_t^- \mid \mathbf{y}_t^+\right] = \bar{\Gamma}_k \hat{\bar{\mathbf{x}}}(t) \qquad k = t$$

The Backward sate space :

$$\hat{\bar{\mathfrak{X}}}_t = \text{span}\, \mathbb{E}\left[\mathbf{y}_t^- \mid \mathbf{y}_t^+\right]$$

Backward covariance $\Sigma_{\mathbf{y}}(-\tau) = \mathbb{E}\,\mathbf{y}(-\tau)\mathbf{y}(0)^\top = \bar{C}A^{\tau-1}C^\top$

ESTIMATING THE B, D PARAMETERS

We have the stationary parameters $(A, C, \bar{G})$ and $\Lambda_0 \simeq \Sigma_{\mathbf{y}}(0)$

Solve the Algebraic Riccati Equation

$$P = APA^\top + (\bar{G}^\top - APC^\top)(\Lambda_0 - CPC^\top)^{-1}(\bar{G} - CPA^\top) \qquad \text{(ARE)}$$

To get the minimal (stabilizing) solution P_-

$$K = \left[\bar{G}^\top - AP_-C^\top\right] R(P_-)^{-1} \qquad R(P_-) = \Lambda_0 - CP_-C^\top = D_-D_-^\top$$

Equivalently $B_- = KD_-$

The ARE has a solution iff $(A, C, \bar{G}, \Lambda_0)$ is positive real !

COMPARISON WITH THE "EARLY" ALGORITHM

Conceptually the algorithm is the same as HO-KALMAN applied to the finite **Normalized** Hankel matrix

$$\hat{\mathbb{H}}_\Lambda := L_+^{-1}\begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 & \dots & \Lambda_\nu \\ \Lambda_2 & \Lambda_3 & \Lambda_4 & \dots & \Lambda_{\nu-1} \\ \Lambda_3 & \Lambda_4 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \Lambda_{\nu+1} & \dots & \dots & \dots & \Lambda_{2\nu} \end{bmatrix} L_-^{-\top}$$

$$\hat{\mathbb{H}}_\Lambda = \simeq \hat{U}\hat{\Sigma}\hat{V}^\top \qquad \hat{\Sigma} = \mathrm{diag}\{\sigma_1, \dots, \sigma_n\}$$

Leads to exactly the same formulas as

$$\begin{bmatrix} \hat{A} \\ C \end{bmatrix}_N = \begin{bmatrix} \mathbf{Z}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} \mathbf{Z}_t \mathbf{Z}_t^\top$$

NUMERICAL ASPECTS

The LQ factorization a key step in subspace identification algorithms.

$$\begin{bmatrix} U \\ Y \end{bmatrix} = \begin{bmatrix} L_{uu} & 0 \\ L_{yu} & L_{yy} \end{bmatrix} \begin{bmatrix} Q_u^\top \\ Q_y^\top \end{bmatrix}$$

where $Q_u^\top Q_u = I,\ Q_y^\top Q_y = I,\ Q_u^\top Q_y = 0$ and L_{uu}, L_{yy} are lower triangular.

$$\begin{aligned} \mathbb{E}\left[Y \mid \mathcal{U}\right] &= Y Q_u \left[Q_u^\top Q_u\right]^{-1} Q_u^\top = L_{yu} Q_u^\top \\ \mathbb{E}\left[Y \mid \mathcal{U}^\perp\right] &= Y Q_y \left[Q_y^\top Q_y\right]^{-1} Q_y^\top = L_{yy} Q_y^\top \end{aligned}$$

$Q_y^\top$ an orthonormal basis for the orthogonal complement $\mathcal{U}^\perp$ in $\mathcal{U} \vee \mathcal{Y}$.

NUMERICAL ASPECTS (Cont'd)

SVD step: Compute from LQ factorization

$$\mathbb{E}\{\hat{\mathbf{Y}}_t^+ \mid \hat{\mathbf{Y}}_t^-\} = \begin{bmatrix}\hat{U} & \tilde{U}\end{bmatrix} \begin{bmatrix}\hat{\Sigma} & 0 \\ 0 & \tilde{\Sigma}\end{bmatrix} \begin{bmatrix}\hat{V} & \tilde{V}\end{bmatrix}^\top$$

Do **order estimation:** pick n such that $\tilde{U} \simeq 0$

Extended Observability matrix from

$$\mathbb{E}\{\hat{\mathbf{Y}}_t^+ \mid \hat{\mathbf{Y}}_t^-\} \simeq \hat{U}\hat{\Sigma}^{1/2} \; \hat{\Sigma}^{1/2}\hat{V}^\top := \Omega_t \; \hat{\mathbf{X}}(t)$$

Get A, C from **Shift-Invariance method :**

$$\hat{U}\hat{\Sigma}^{1/2} = \Omega = \begin{bmatrix} C \\ (\downarrow \Omega) \end{bmatrix} = \begin{bmatrix} (\uparrow \Omega) \\ CA^{N-1} \end{bmatrix}; \qquad (\downarrow \Omega) = (\uparrow \Omega)A \quad \Rightarrow \quad A = (\uparrow \Omega)^{-L}(\downarrow \Omega)$$

N.B. **no need** to compute $\mathbf{Z}_{t+1}$*coherent basis in* $\hat{\mathfrak{X}}_{t+1}$ with $\mathbf{Z}_t$ and solving

$$\begin{bmatrix} \mathbf{Z}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \mathbf{Z}_t + \begin{bmatrix} K(t) \\ I \end{bmatrix} \hat{\mathbf{E}}_t$$

by Least-Squares.

Get $A^\top, \bar{C}$ from the backward procedure

$$\mathbb{E}\{\hat{\mathbf{Y}}_t^- \mid \hat{\mathbf{Y}}_t^+\} \simeq \hat{V}\hat{\Sigma}^{1/2}\ \hat{\Sigma}^{1/2}\hat{U}^\top := \bar{\Omega}_t\ \hat{\bar{\mathbf{X}}}(t)$$

$$\bar{\Omega} = \begin{bmatrix} \bar{C} \\ (\downarrow \bar{\Omega}) \end{bmatrix} = \begin{bmatrix} (\uparrow \bar{\Omega}) \\ \bar{C}(A^\top)^{N-1} \end{bmatrix}; \qquad (\downarrow \bar{\Omega}) = (\uparrow \bar{\Omega})A^\top \quad \Rightarrow \quad A^\top = (\uparrow \bar{\Omega})^{-L}(\downarrow \bar{\Omega})$$

Only need to pick the first block of m rows to get $\bar{C}$.

ORDER SELECTION

Minimize Akaike-type criterion

$$NIC(n) := \sum_{k=n+1}^{n_{MAX}} \hat{\sigma}_k^2 - d(n)\frac{logN}{N}$$

where $d(n)$ = number of additional free parameters in a model of order $n_{MAX} > n$.

Consistency If data are generated by a true model of order n_0 and $N \to \infty$ the minimum NIC estimate of n is consistent:

$$\hat{n} \to n_0 \qquad \text{with probability one.}$$

STATISTICAL PROPERTIES

- **Consistency** If data are generated by a true model

- **Asymptotic Variance of** A, C

- **Efficiency**

STATE-SPACE MODELS WITH INPUT SIGNALS

$\mathbf{u} = \{\mathbf{u}(t,\omega)\}$ discrete-time p-dimensional zero-mean **random signal** in $t \in [t_0, +\infty)$.

STOCHASTIC STATE-SPACE MODEL WITH INPUTS

$$\begin{cases} \mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + G\mathbf{w}(t) & \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t) + J\mathbf{w}(t), & t \geq t_0 \end{cases}$$

A, B, C, D, G, J constant matrices, $\{\mathbf{x}(t)\}$ is the state process of dimension n, and $\{\mathbf{w}(t)\}$ is a normalized white noise process. Assume $|\lambda(A)| < 1$ (causality).

N.B: **We are not interested in modelling the input** $\{\mathbf{u}(t)\}$.

Assumption: **there is no feedback from $\mathbf{y}$ to $\mathbf{u}$**. This is the same as: **the processes $\{\mathbf{u}(t)\}$ and $\{\mathbf{w}(t)\}$ are completely uncorrelated**.

DETERMINISTIC + STOCHASTIC DECOMPOSITION

State Space Model for $\mathbf{y}$: *parallel* of models (in general **Non Minimal!**)

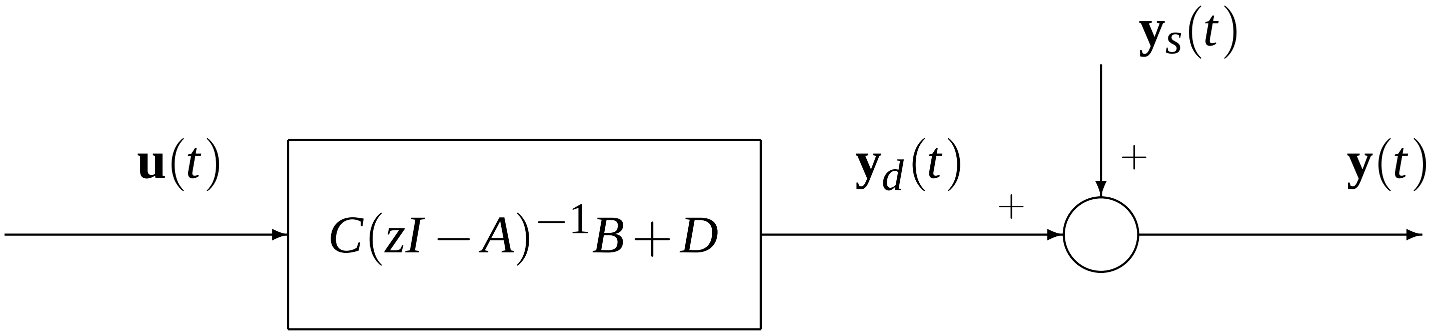
$$\text{Stochastic Model} \quad \begin{cases} \mathbf{x}_S(t+1) &= A\mathbf{x}_S(t) + G\mathbf{w}(t) \\ \mathbf{y}_S(t) &= C\mathbf{x}_S(t) + J\mathbf{w}(t) \end{cases}$$

$$\text{Deterministic Model} \quad \begin{cases} \mathbf{x}_d(t+1) &= A\mathbf{x}_d(t) + B\mathbf{u}(t) \\ \mathbf{y}_d(t) &= C\mathbf{x}_d(t) + D\mathbf{u}(t) \end{cases}$$

$$\mathbf{y}(t) = \mathbf{y}_S(t) + \mathbf{y}_d(t) = C\left[\mathbf{x}_S(t) + \mathbf{x}_d(t)\right] + D\mathbf{u}(t) + J\mathbf{w}(t)$$

NB. $\mathbf{x}_S(t)$ **uncorrelated with** $\mathbf{u}$ $\quad\Rightarrow\quad$ $\mathbf{x}_S(t)$ **uncorrelated with** $\mathbf{x}_d$!

FROM STATE-SPACE TO ARMAX



Deterministic system + "stochastic error" decomposition :

$$\begin{aligned}\mathbf{y}(t) &= \left[C(zI-A)^{-1}B+D\right]\mathbf{u}(t) + \left[C(zI-A)^{-1}G+J\right]\mathbf{w}(t)\\ &:= F(z)\mathbf{u}(t) + G(z)\mathbf{w}(t)\end{aligned}$$

NB: $F(z)$ and $G(z)$ realized with the same (A,C) pair. In general non-minimal realizations

$F(z)$ $G(z)$ rational. Can be written as a ratio of polynomial matrices with **same denominator**

$$F(z) = A(z)^{-1} B(z); \quad G(z) = A(z)^{-1} C(z)$$

$$A(z) = Iz^{\nu} + \sum_{1}^{\nu} A_k z^{\nu-k} \qquad B(z) = \sum_{1}^{\nu} B_k z^{\nu-k} \qquad C(z) = C_0 z^{\nu} + \sum_{1}^{\nu} C_k z^{\nu-k}$$

$\{\mathbf{y}(t)\}$ may be described also by the **ARMAX model**

$$\mathbf{y}(t) + \sum_{1}^{\nu} A_k \mathbf{y}(t-k) = \sum_{1}^{\nu} B_k \mathbf{u}(t-k) + C_0 \mathbf{w}(t) + \sum_{1}^{\nu} C_k \mathbf{w}(t-k) \quad .$$

IDENTIFICATION OF SYSTEMS WITH INPUTS (NO FEEDBACK)

Could be done in two ways. Either identify the **joint model** or first compute $\mathbf{y}_d(t) = \mathbb{E}\{\mathbf{y}(t) \mid H(\mathbf{u})\}$ and identify a **Deterministic Model**

$$\begin{cases} \mathbf{x}_d(t+1) & = & A\mathbf{x}_d(t) + B\mathbf{u}(t) \\ \mathbf{y}_d(t) & = & C\mathbf{x}_d(t) + D\mathbf{u}(t) \end{cases}$$

Then identify a stochastic model for the **disturbance**

$$\mathbf{y}_s(t) = \mathbf{y}(t) - \mathbf{y}_d(t) = \mathbb{E}\{\mathbf{y}(t) \mid H(\mathbf{u})^{\perp}\}$$

Shall do the JOINT model only.

JOINT INNOVATION MODEL WITH INPUTS

Steady state Kalman filter: $\hat{\mathbf{x}}(t+1) = \mathbb{E}\{\mathbf{x}(t+1) \mid \mathbf{y}(s), \mathbf{u}(s)\, s \leq t\}$

Innovation process (of $\mathbf{y}$) : $\hat{\mathbf{e}}(t) = \mathbf{y}(t) - C\hat{\mathbf{x}}(t) - D\mathbf{u}(t)$ **white noise !**

$$\begin{bmatrix} \hat{\mathbf{x}}(t+1) \\ \mathbf{y}(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}(t) \\ \mathbf{u}(t) \end{bmatrix} + \begin{bmatrix} K \\ I \end{bmatrix} \mathbf{e}(t)$$

No feedback from $\mathbf{y}$ to $\mathbf{u}$: $\mathbf{e}(t) \perp \mathbf{u}(\tau) \quad \forall\, \tau,\, t$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \mathbb{E}\{\begin{bmatrix} \hat{\mathbf{x}}(t+1) \\ \mathbf{y}(t) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}(t) \\ \mathbf{u}(t) \end{bmatrix}^\top\} \left(\mathbb{E}\{\begin{bmatrix} \hat{\mathbf{x}}(t) \\ \mathbf{u}(t) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}(t) \\ \mathbf{u}(t) \end{bmatrix}^\top\} \right)^{-1}$$

Parameters are uniquely determined by the basis $\mathbf{x}(t)$!

IDENTIFICATION OF THE DETERMINISTIC SUBSYSTEM

Problem : Assume the data are generated by a true stochastic system of order n. From observed input-output time series

$$\{y_0, y_1, y_2, \ldots, y_N\}, \quad y_t \in \mathbb{R}^m \qquad \{u_0, u_1, u_2, \ldots, u_N\}, \quad u_t \in \mathbb{R}^p$$

find estimates (in a certain basis) $\widehat{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_N$

such that (**consistency**)

$$\lim_{N \to \infty} \widehat{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_N = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

BASIC IDEA OF SUBSPACE IDENTIFICATION

Assume we can observe also a **state trajectory** $\{x_0, x_1, x_2, \dots, x_N\}$, corrre-sponding to the I/O data

$$\{y_0, y_1, y_2, \dots, y_N\}, \quad y_t \in \mathbb{R}^m \qquad \{u_0, u_1, u_2, \dots, u_N\}, \quad u_t \in \mathbb{R}^p$$

Form the “tail” matrices $\mathbf{Y}_t$, $\mathbf{X}_t$, $\mathbf{U}_t$

$$\begin{aligned}\mathbf{Y}_t &:= [\, y_t, y_{t+1}, y_{t+2}, \dots] \\ \mathbf{X}_t &:= [\, x_t, x_{t+1}, x_{t+2}, \dots] \\ \mathbf{U}_t &:= [\, u_t, u_{t+1}, u_{t+2}, \dots]\end{aligned}$$

Every sample trajectory $\{y_t\}$, $\{x_t\}$, $\{u_t\}$ of the system must satisfy the model equations, so

$$\begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix} + \begin{bmatrix} K \\ I \end{bmatrix} \mathbf{E}_t$$

BASIC IDEA OF SUBSPACE IDENTIFICATION (cont'd)

$$\begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix} + \begin{bmatrix} K \\ I \end{bmatrix} \mathbf{E}_t$$

Linear Regression ! Solve by Least Squares :

$$\min_{A,C,B,D} \| \begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix} \|$$

getting

$$\widehat{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_N := \frac{1}{N} \begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} \begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix}^\top \left\{ \frac{1}{N} \begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix} \begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix}^\top \right\}^{-1}$$

BASIC IDEA OF SUBSPACE IDENTIFICATION (cont'd)

$$\widehat{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_N := \frac{1}{N}\begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix}\begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix}^\top \left\{ \frac{1}{N}\begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix}\begin{bmatrix} \mathbf{X}_t \\ \mathbf{U}_t \end{bmatrix}^\top \right\}^{-1}$$

Theorem: If the data are **second order ergodic**, there is no feedback and the inverse exists:

$$\lim_{N\to\infty} \widehat{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_N = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad (\dagger)$$

consistent estimate of A, B, C, D.

Proof: HOMEWORK

SECOND ORDER ERGODICITY

For $N \to \infty$ sample covariances converge to true covariances, say

$$\frac{1}{N}\sum_{k=t}^{t+N}\{y_k u_k^\top\} = \frac{1}{N}\mathbf{Y}_t\mathbf{U}_s^\top \to \mathbb{E}\{\mathbf{y}(t)\,\mathbf{u}(s)^\top\} \qquad N \to \infty$$

For $N \to \infty$ the **sample covariances can be substituted by the true ones**.

Assuming N “very large” can use **random variables** instead of numerical sequences!

$$\mathbf{y}(t) \Leftrightarrow \mathbf{Y}_t, \quad \mathbf{u}(t) \Leftrightarrow \mathbf{U}_t, \qquad \text{etc.}$$

COMMENTS

STATE SEQUENCE IS NOT AVAILABLE: NEED TO CONSTRUCT THE STATE FROM INPUT-OUTPUT DATA!

Easy to do if **infinite past data** were available at time t: want to construct the

Steady state Kalman filter: $\hat{\mathbf{x}}(t) = \mathbb{E}\{\mathbf{x}(t) \mid \mathbf{y}(s), \mathbf{u}(s)\, s < t\}$

Innovation process (of $\mathbf{y}$) : $\hat{\mathbf{e}}(t) = \mathbf{y}(t) - C\hat{\mathbf{x}}(t) - D\mathbf{u}(t)$ **white noise !**

$$\begin{bmatrix} \hat{\mathbf{x}}(t+1) \\ \mathbf{y}(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}(t) \\ \mathbf{u}(t) \end{bmatrix} + \begin{bmatrix} K \\ I \end{bmatrix} \mathbf{e}(t)$$

Pick basis vector in the state space of this model : Generalize Akaike procedure by **conditional CCA**

CONSTRUCTING THE STATE SPACE OF A JOINT STATIONARY MODEL

$\mathbf{X}_t := \text{span}\,\{\hat{\mathbf{x}}_1(t), \hat{\mathbf{x}}_2(t) \ldots, \hat{\mathbf{x}}_n(t)\}$ State space of a joint innovation model.
Assume we have data starting from $t = -\infty$

$$\begin{aligned} \mathcal{P}_t \quad &:= \quad \mathcal{Y}_t \vee \mathcal{U}_t = \text{span}\{\mathbf{y}(s), s < t, \mathbf{u}(s), s < t\} \\ \mathcal{U}_t^+ \quad &:= \quad \text{span}\{\mathbf{u}(s), s \geq t\} \end{aligned}$$

Theorem If the data are generated by a finite-dimensional stationary model and there is no feedback,

$$\mathbf{X}_t = \text{span}\{\mathbb{E}_{\|\mathcal{U}_t^+}\{\mathbf{y}(t+h) \mid \mathcal{P}_t\};\, h = 0, 1, \ldots, n\}$$

State-Space = Oblique Predictor Space= Oblique projection of future outputs onto joint past along future inputs

Proof:

$$\begin{aligned}\mathbf{y}(t+h) &= CA^h\hat{\mathbf{x}}(t) + \textstyle\sum_{k=0}^{h-1} CA^{h-1-k}B\mathbf{u}(t+k) + D\mathbf{u}(t+h) \\ &+ \textstyle\sum_{k=0}^{h-1} CA^{h-1-k}K\mathbf{e}(t+k) + J\mathbf{e}(t+h)\end{aligned}$$

since $\mathbf{e}(t+k) \quad \perp \; \mathcal{P}_t$:

$$\begin{aligned}\mathbb{E}\{\mathbf{y}(t+h) \mid \mathcal{P}_t \vee \mathcal{U}_t^+\} &= \mathbb{E}\{\mathbf{y}(t+h) \mid \mathcal{P}_t \vee \mathcal{U}_{t\,|t+h}\} \\ &= CA^h\hat{\mathbf{x}}(t) \quad + \quad \textstyle\sum_{k=0}^{h-1} CA^{h-1-k}B\mathbf{u}(t+k) + D\mathbf{u}(t+h) \\ &= \mathbb{E}_{\|\mathcal{U}_t^+}\{\mathbf{y}(t+h) \mid \mathcal{P}_t\} \quad + \quad \mathbb{E}_{\|\mathcal{P}_t}\{\mathbf{y}(t+h) \mid \mathcal{U}_{[t\,t+h]}\}\end{aligned}$$

HENCE: $\mathbb{E}_{\|\mathcal{U}_t^+}\{\mathbf{y}(t+h) \mid \mathcal{P}_t\} = CA^h\hat{\mathbf{x}}(t) \; h = 0, 1, \ldots$. QED

OBLIQUE PROJECTIONS

Let $\mathcal{A} = \text{span}\{\mathbf{a}\}, \quad \mathcal{B} = \text{span}\{\mathbf{b}\}$.
The **oblique projection of $\mathbf{v}$ onto $\mathcal{A}$ along $\mathcal{B}$** is

$$\mathbb{E}_{\,||\mathcal{B}}\{\mathbf{v}\,|\,\mathcal{A}\} = \left[\mathbb{E}\left\{\mathbf{v}\mathbf{a}^\top\right\}\; E\left\{\mathbf{v}\mathbf{b}^\top\right\}\right]\left[\begin{array}{cc}\mathbb{E}\left\{\mathbf{a}\mathbf{a}^\top\right\} & \mathbb{E}\left\{\mathbf{a}\mathbf{b}^\top\right\}\\ \mathbb{E}\left\{\mathbf{b}\mathbf{a}^\top\right\} & \mathbb{E}\left\{\mathbf{b}\mathbf{b}^\top\right\}\end{array}\right]^\dagger\left[\begin{array}{c}\mathbf{a}\\ 0\end{array}\right]$$

If $\mathcal{A} \perp \mathcal{B}$ i.e. $\mathcal{A}$ and $\mathcal{B}$ are **orthogonal**

$$\mathbb{E}_{\,||\mathcal{B}}\{\mathbf{v}\,|\,\mathcal{A}\} = \mathbb{E}\,\{\mathbf{v}\,|\,\mathcal{A}\}$$

If $\mathcal{A} \cap \mathcal{B} = \{0\}$ i.e. $\mathcal{A}$ and $\mathcal{B}$ are in **direct sum** unique decomposition

$$\mathbb{E}\,\{\mathbf{v}\,|\,\mathcal{A}+\mathcal{B}\} = \mathbb{E}_{\,||\mathcal{B}}\{\mathbf{v}\,|\,\mathcal{A}\} + \mathbb{E}_{\,||\mathcal{A}}\{\mathbf{v}\,|\,\mathcal{B}\}$$

OBLIQUE PROJECTIONS

"Conditional quantities"

$$\mathbb{E}\left\{\mathbf{v} \mid \mathbf{b}^{\perp}\right\} := \mathbf{v} - \mathbb{E}\left\{\mathbf{v} \mid \mathbf{b}\right\}, \qquad \mathbb{E}\left\{\mathbf{a} \mid \mathbf{b}^{\perp}\right\} := \mathbf{a} - \mathbb{E}\left\{\mathbf{a} \mid \mathbf{b}\right\}$$

$$\Sigma_{\mathbf{va}|\mathbf{b}} := \mathrm{Cov}\left[\mathbb{E}\left\{\mathbf{v} \mid \mathbf{b}^{\perp}\right\}, \mathbb{E}\left\{\mathbf{a} \mid \mathbf{b}^{\perp}\right\}\right], \qquad \Sigma_{\mathbf{aa}|\mathbf{b}} := \mathrm{Var}\left[\mathbb{E}\left\{\mathbf{a} \mid \mathbf{b}^{\perp}\right\}\right]$$

Fact: Assume $\mathcal{A} \cap \mathcal{B} = \{0\}$ and $\mathbf{a}$ and $\mathbf{b}$ are bases:

$$\mathbb{E}_{\,||\mathcal{B}}\left\{\mathbf{v} \mid \mathcal{A}\right\} = \Sigma_{\mathbf{va}|\mathbf{b}}\Sigma_{\mathbf{aa}|\mathbf{b}}^{-1}\mathbf{a} \qquad \mathbb{E}_{\,||\mathcal{A}}\left\{\mathbf{v} \mid \mathcal{B}\right\} = \Sigma_{\mathbf{vb}|\mathbf{a}}\Sigma_{\mathbf{bb}|\mathbf{a}}^{-1}\mathbf{b}$$

THE STATE SPACE
OF A STATIONARY MODEL

Any choice of basis in the **oblique predictor space**

$$\mathbf{X}_t = \mathbb{E}_{\|\mathcal{U}_{[t,t+n]}}\{\mathcal{Y}_{[t,t+n]} \mid \mathcal{Y}_t^- \vee \mathcal{U}_t^-\} = \operatorname{span}\{\mathbb{E}_{\|\mathcal{U}_t^+}[\mathbf{y}(t+h) \mid \mathcal{P}_t]\,;\, h=0,1,2'\ldots\}$$

provides a minimal innovation model (Steady state Kalman filter)

$$\begin{cases} \hat{\mathbf{x}}(t+1) &= A\hat{\mathbf{x}}(t)+B\mathbf{u}(t)+K\mathbf{e}(t) \\ \mathbf{y}(t) &= C\hat{\mathbf{x}}(t)+D\mathbf{u}(t)+\mathbf{e}(t) \end{cases}$$

Can compute a basis in $\mathbf{X}_t$ by **conditional CCA**: SVD of the normalized conditional covariance of future outputs $\mathbf{y}_t^+$ and (joint!) past

$$\mathbf{p}(t) := \begin{bmatrix} \mathbf{u}_t^- \\ \mathbf{y}_t^- \end{bmatrix} \qquad [\infty \times 1 \text{ past observations}]$$

given future inputs $\mathbf{u}_t^+$.

STATIONARY CONDITIONAL CCA

If data are described by a true n-dimensional model, the **Conditional Hankel Matrix**

$$H_{\mathbf{y}^+\mathbf{p}|\mathbf{u}^+} := \mathrm{Cov}\left[\mathbb{E}\left\{\mathbf{y}_t^+ \mid (\mathbf{u}^+)^\perp\right\}, \mathbb{E}\left\{\mathbf{p}(t) \mid (\mathbf{u}_t^+)^\perp\right\}\right]$$

has finite rank $n \Rightarrow \mathbf{y}_t^+$ and $\mathbf{u}_t^+$ can be taken to be finite dimensional vectors.

Cholesky factors

$$H_{\mathbf{y}^+\mathbf{y}^+|\mathbf{u}^+} := \mathrm{Var}\left[\mathbb{E}\left\{\mathbf{y}_t^+ \mid (\mathbf{u}^+)^\perp\right\}\right] = L_{\mathbf{y}^+|\mathbf{u}^+} L_{\mathbf{y}^+|\mathbf{u}^+}^\top,$$

$$H_{\mathbf{pp}|\mathbf{u}^+} := \mathrm{Var}\left[\mathbb{E}\left\{\mathbf{p}(t) \mid (\mathbf{u}^+)^\perp\right\}\right] = L_{\mathbf{p}|\mathbf{u}^+} L_{\mathbf{p}|\mathbf{u}^+}^\top$$

Do SVD of the *normalized conditional Hankel matrix*

$$\hat{H}_{\mathbf{y}^+\mathbf{p}|\mathbf{u}^+} := L_{\mathbf{y}^+|\mathbf{u}^+}^{-1} H_{\mathbf{y}^+\mathbf{p}|\mathbf{u}^+} L_{\mathbf{p}|\mathbf{u}^+}^{-\top}$$

Order estimation

$$\hat{H}_{\mathbf{y}^+\mathbf{p}|\mathbf{u}^+} := \begin{bmatrix}\hat{U} & \tilde{U}\end{bmatrix} \begin{bmatrix}\hat{\Sigma} & 0 \\ 0 & \tilde{\Sigma}\end{bmatrix} \begin{bmatrix}\hat{V} & \tilde{V}\end{bmatrix}^\top \simeq \hat{U}\hat{\Sigma}\hat{V}^\top$$

Canonical state

$$\mathbf{z}(t) = \hat{\Sigma}^{1/2}\hat{V}^\top L^{-1}_{\mathbf{p}|\mathbf{u}^+}\mathbf{p}(t)$$

MAIN DIFFICULTY: The **The infinite past** $\mathbf{p}(t)$ spanning $\mathcal{Y}_{-\infty|t} \vee \mathcal{U}_{-\infty|t}$ is not available !! Approximation with available **finite past** yields biased estimates. Bias may be large if the zeros of the true system are far from the unit circle.

ONLY FINITE DATA ARE AVAILABLE!

Infinite past approximation leads to **errors (bias)** in the estimate which do not $\rightarrow 0$ as $N \rightarrow \infty$.

Bias can be made arbitrarily large taking the zeros of the stochastic subsystem arbitrarily close to the unit circle.

For consistency with finite regression data: NEED FINITE-INTERVAL (NON–STATIONARY) STOCHASTIC REALIZATION

FINITE-INTERVAL INNOVATION MODEL

Modeling using "data" on a finite-interval $[t_0, T]$. The estimate

$$\hat{\mathbf{x}}(t) := \mathbb{E}\left[\mathbf{x}(t) \mid \mathcal{P}_{[t_0,t)} \vee \mathcal{U}_{[t,T]}\right]$$

satisfies the *transient conditional Kalman filter equation*

$$\begin{cases} \hat{\mathbf{x}}(t+1) &= A\hat{\mathbf{x}}(t) + B\mathbf{u}(t) + K(t)\hat{\mathbf{e}}(t) \\ \mathbf{y}(t) &= C\hat{\mathbf{x}}(t) + D\mathbf{u}(t) + \hat{\mathbf{e}}(t) \\ \hat{\mathbf{x}}(t_0) &= \mathbb{E}\left[\mathbf{x}(t_0) \mid \mathcal{U}_{[t_0 T]}\right] \end{cases}$$

How to construct $\hat{\mathbf{x}}(t)$?

Is $\hat{\mathbf{x}}(t)$ a basis in some predictor space? e.g $\mathbb{E}_{\|\mathcal{U}_{[t,T]}}\left[\mathcal{Y}_{[t,T]} \mid \mathcal{P}_{[t_0,t)}\right]$?

Cannot be $\mathbb{E}\left[\mathcal{Y}_{[t,T]} \mid \mathcal{P}_{[t_0,t)}\right]$; would introduce innovation of $\mathbf{u}$!!

FINITE-INTERVAL REALIZATION THEORY

$$H_d := \begin{bmatrix} D & 0 & \dots & 0 & 0 \\ CB & D & \dots & 0 & 0 \\ \vdots & & & \ddots & \vdots \\ CA^{\nu-1}B & CA^{\nu-2}B & \dots & CB & D \end{bmatrix},$$

$$\begin{aligned} &\mathbb{E}\left[\mathbf{y}(t+h) \mid \mathcal{P}_{[t_0,t)} \vee \mathcal{U}_{[t,T]}\right] = \\ &= \mathbb{E}_{\|\mathcal{U}_{[t,T]}}\left[\mathbf{y}(t+h) \mid \mathcal{P}_{[t_0,t)}\right] + \mathbb{E}_{\|\mathcal{P}_{[t_0,t)}}\left[\mathbf{y}(t+h) \mid \mathcal{U}_{[t,T]}\right] \\ &= CA^h \mathbb{E}_{\|\mathcal{U}_{[t,T]}}\left[\hat{\mathbf{x}}(t) \mid \mathcal{P}_{[t_0,t)}\right] + CA^h \mathbb{E}_{\|\mathcal{P}_{[t_0,t)}}\left[\hat{\mathbf{x}}(t) \mid \mathcal{U}_{[t,T]}\right] + H_{d,h}\mathbf{u}_t^+ \end{aligned}$$

Projecting along $\mathcal{U}_{[t,T]}$ kills one piece of

$$\hat{\mathbf{x}}(t) = \mathbb{E}_{\|\mathcal{U}_{[t,T]}}\left[\hat{\mathbf{x}}(t) \mid \mathcal{P}_{[t_0,t)}\right] + \mathbb{E}_{\|\mathcal{P}_{[t_0,t)}}\left[\hat{\mathbf{x}}(t) \mid \mathcal{U}_{[t,T]}\right]$$

PATCHING UP

Causal component of the state: $\boldsymbol{\xi}(t) := \mathbb{E}_{\|\mathcal{U}_{[t,T]}}\left[\hat{\mathbf{x}}(t) \mid \mathcal{P}_{[t_0,t)}\right]$

$$\mathbb{E}_{\|\mathcal{U}_{[t,T]}}\left[\mathbf{y}(t+h) \mid \mathcal{P}_{[t_0,t)} \vee \mathcal{U}_{[t,T]}\right] := CA^h\,\boldsymbol{\xi}(t)$$

From an oblique projection can recover the **Observability Matrix** Γ_k:

$$\hat{\mathbf{y}}_t^+ = \mathbb{E}\left\{\begin{bmatrix}\mathbf{y}(t)\\ \mathbf{y}(t+1)\\ \vdots \\ \mathbf{y}(t+k)\end{bmatrix} \mid \mathcal{P}_{[t_0,t)} \vee \mathcal{U}_{[t,T]}\right\} = \Gamma_k\,\boldsymbol{\xi}(t) + \text{part in } \mathcal{U}_{[t,T]}$$

Need

$$\mathbb{E}\{\boldsymbol{\xi}(t)\,\boldsymbol{\xi}(t)^\top\} > 0 \qquad\qquad \text{(consistency condition)}$$

THE VAN OVERSCHEE-DE MOOR MODEL

Pseudostate: $\quad \bar{\mathbf{x}}(t) := \Gamma_k^{-L}\hat{\mathbf{y}}_t^+ \;=\; \hat{\mathbf{x}}(t) + \Gamma_k^{-L} H_k \mathbf{u}_t^+$

$$\begin{bmatrix} \bar{\mathbf{x}}(t+1) \\ \mathbf{y}(t) \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \bar{\mathbf{x}}(t) + \begin{bmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{bmatrix} \mathbf{u}_t^+ + \mathbf{w}_t^\perp \qquad (*)$$

$\mathcal{K}_1\ \mathcal{K}_2$ known linear functions of (B,D).

$$\begin{aligned} \begin{bmatrix} A \\ C \end{bmatrix} \Sigma_{\overline{\mathbf{x}\mathbf{x}}|\mathbf{u}^+} \quad &= \quad \begin{bmatrix} \Sigma_{\bar{\mathbf{x}}_1\bar{\mathbf{x}}|\mathbf{u}^+} \\ \Sigma_{\overline{\mathbf{y}\mathbf{x}}|\mathbf{u}^+} \end{bmatrix} \\ \begin{bmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{bmatrix} \Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}} \quad &= \quad \begin{bmatrix} \Sigma_{\bar{\mathbf{x}}_1\mathbf{u}^+|\bar{\mathbf{x}}} \\ \Sigma_{\mathbf{y}\mathbf{u}^+|\bar{\mathbf{x}}} \end{bmatrix} \end{aligned}$$

Solve in terms of **Conditional Covariances**:

$$\Sigma_{\bar{\mathbf{x}}\bar{\mathbf{x}}|\mathbf{u}^+} = E\{\left[\bar{\mathbf{x}}(t) - E\left(\bar{\mathbf{x}}(t) \mid \mathbf{u}_t^+\right)\right]\left[\bar{\mathbf{x}}(t) - E\left(\bar{\mathbf{x}}(t) \mid \mathbf{u}_t^+\right)\right]^\top\} \quad = \quad \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$$

$$\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}} = E\{\left[\mathbf{u}_t^+ - E(\mathbf{u}_t^+ \mid \bar{\mathbf{x}}(t))\right]\left[\mathbf{u}_t^+ - E(\mathbf{u}_t^+ \mid \bar{\mathbf{x}}(t))\right]^\top\}$$

THE N4SID ALGORITHM

[vanOverschee-DeMoor94]

1. Predictor matrix based on joint input-output data

$$\hat{Y}_{[t,T-1]} := \mathbb{E}\left[Y_{[t,T-1]} \mid Y_{[t_0,t)} \vee U_{[t_0,T]}\right]$$

(projection onto the joint rowspace).

2. Compute the oblique projection along $U_{[t,T]}$

$$Z(t) := \mathbb{E}_{\|\mathcal{U}_{[t,T]}}\left[\hat{Y}_{[t,T-1]} \mid Y_{[t_0,t)} \vee U_{[t_0,t)}\right]$$

to get an estimate of $\Gamma_k \Xi_t$

3 Estimate the order and the observability matrix Γ_k by SVD factorization.

4. The "Pseudostate" $\bar{X}_t := \Gamma_k^{-L} \hat{Y}_{[t,T-1]}$ obeys the recursion

$$\begin{bmatrix} \bar{X}_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \bar{X}_t + \begin{bmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{bmatrix} U_{[t,T]} + W^{\perp}$$

5. Compute a coherent pseudostate at time $t+1$: $\quad \bar{X}_{t+1} := \Gamma_k^{-L} \hat{Y}_{[t+1,T]}$

6. Solve by LS for the unknown parameters (A, C) and $(\mathcal{K}_1, \mathcal{K}_2)$

7. Estimate (B, D) from $(\mathcal{K}_1, \mathcal{K}_2)$.

“ MOESP”

Start from the stationary innovation model (state vector $\mathbf{x}(t)$) future horizon: $k = T - t$,

$$
\begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t+1) \\ \vdots \\ \mathbf{y}(T) \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^k \end{bmatrix} \mathbf{x}(t) \quad + \begin{bmatrix} D & & & 0 & 0 \\ CB & D & & & 0 \\ \vdots & \ddots & \ddots & & \\ CA^{k-1}B & \cdots & CB & D & \end{bmatrix} \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{u}(t+1) \\ \vdots \\ \mathbf{u}(T) \end{bmatrix}
$$

$$
+ \begin{bmatrix} I & & & 0 & 0 \\ CK & I & & & 0 \\ \vdots & \ddots & \ddots & & \\ CA^{k-1}K & \cdots & CK & I & \end{bmatrix} \begin{bmatrix} \mathbf{e}(t) \\ \mathbf{e}(t+1) \\ \vdots \\ \mathbf{e}(T) \end{bmatrix}
$$

$$
\mathbf{y}_t^+ = \Gamma_k \mathbf{x}(t) + H_d \, \mathbf{u}_t^+ + H_e \, \mathbf{e}_t^+
$$

Want to kill the last two pieces.

" MOESP" (Cont'd)

1. Project orthogonally onto $\mathcal{Y}_{[t_0,t)} \vee \mathcal{U}_{[t_0,T]}$

$$\hat{\mathbf{y}}_t^+ := \mathbb{E}\left[\mathbf{y}_t^+ \mid \mathcal{P}_{[t_0,t)} \vee \mathcal{U}_{[t,T]}\right] = \Gamma_k \hat{\mathbf{x}}(t) + H_d\,\mathbf{u}_t^+$$

2. Project onto the orthogonal complement $\mathcal{U}_{[t,T]}^\perp$

$$\hat{\mathbf{z}}_t^+ := \hat{\mathbf{y}}_t^+ - \mathbb{E}\left[\hat{\mathbf{y}}_t^+ \mid \mathcal{U}_{[t,T]}\right] = \Gamma_k\,\hat{\mathbf{x}}^c(t)$$

$$\hat{\mathbf{x}}^c(t) = \hat{\mathbf{x}}(t) - \mathbb{E}\left[\hat{\mathbf{x}}(t) \mid \mathcal{U}_{[t,T]}\right]$$

3. Factorize $\hat{\mathbf{z}}_t^+$, i.e. the matrix

$$Z^c(t) := \mathbb{E}\left[\hat{Y}_{[t,T]} \mid \mathcal{U}_{[t,T]}^\perp\right]$$

by SVD to get an estimate of the order n and of Γ_k e.g. $\hat{\Gamma}_k$.

4. Estimate (A, C) from the estimated observability matrix by the “shift-Invariance method”.

5. Construct a (projection) matrix $\hat{\Gamma}_k^{\perp}$ such that $\hat{\Gamma}_k^{\perp}\hat{\Gamma}_k = 0$

6. Compute

$$\hat{\Gamma}_k^{\perp}\hat{\mathbf{y}}_t^{+} = \Gamma_k^{\perp} H_d \, \mathbf{u}_t^{+} + \text{noise}$$

7. A linear function of (B, D) : estimate (B, D) by linear regression

$$\hat{\Gamma}_k^{\perp}\hat{\mathbf{y}}_t^{+} = L(A, C)\,\text{vec}\,(B, D) + \text{noise}$$

MOESP $\equiv$ ORTHOGONALIZING REGRESSORS

$$\begin{bmatrix} \bar{X}_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \bar{X}_t + \begin{bmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{bmatrix} U_{[t,T]} + W^{\perp}$$

Introduce orthogonal regressors $\bar{X}_t^c := \bar{X}_t - \mathbb{E}\{\bar{X}_t \mid U_{[t,T]}\}$

$$\begin{bmatrix} \bar{X}_{t+1}^c \\ Y_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \bar{X}_t^c + \begin{bmatrix} \mathcal{K}_1^c \\ \mathcal{K}_2^c \end{bmatrix} U_{[t,T]} + W^{\perp}$$

Least squares: right-multiply by $(\bar{X}_t^c)^\top$, $\quad \mathbb{E}\, U_{[t,T]}(\bar{X}_t^c)^\top \to 0$

$$\mathbb{E}\, \bar{X}_t^c (\bar{X}_t^c)^\top \to \Sigma_{\mathbf{x}^c,\mathbf{x}^c} = \Sigma_{\bar{\mathbf{x}}\bar{\mathbf{x}}|\mathbf{u}^+} \quad = \quad \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$$

WEIGHTING

Both N4SID ("Robust") and MOESP use the preliminary orthogonalization

basic step is the SVD of

$$Z^c(t) := \mathbb{E}\left[\hat{Y}_{[t,T]} \mid \mathcal{U}^{\perp}_{[t,T]}\right] \simeq \mathbb{E}\left[\hat{\mathbf{y}}_t^+ \mid (\mathbf{u}_t^+)^{\perp}\right] = \mathbb{E}\left[\mathbf{y}_t^+ \mid (\mathbf{p}(t) - \mathbb{E}\{\mathbf{p}(t) \mid \mathbf{u}_t^+\})\right]$$

$$\mathbf{z}^c(t) = \mathbb{E}\left[\mathbf{y}_t^+ (\mathbb{E}\{\mathbf{p}(t) \mid (\mathbf{u}_t^+)^{\perp}\})^{\top}\right] \mathrm{Var}\{\mathbb{E}\{\mathbf{p}(t) \mid (\mathbf{u}_t^+)^{\perp}\}^{-1} \mathbb{E}\{\mathbf{p}(t) \mid (\mathbf{u}_t^+)^{\perp}\}$$

Conditional Hankel Matrix

$$H_{\mathbf{y}^+\mathbf{p}|\mathbf{u}^+} := \mathrm{Cov}\left[\mathbb{E}\left\{\mathbf{y}_t^+ \mid (\mathbf{u}^+)^{\perp}\right\}, \mathbb{E}\left\{\mathbf{p}(t) \mid (\mathbf{u}_t^+)^{\perp}\right\}\right]$$

has finite rank $n \Rightarrow \mathbf{y}_t^+$ and $\mathbf{u}_t^+$ are finite dimensional vectors.

CCA : introduce Cholesky factors

$$H_{\mathbf{y}^+\mathbf{y}^+|\mathbf{u}^+} := \mathrm{Var}\left[\mathbb{E}\left\{\mathbf{y}_t^+ \mid (\mathbf{u}^+)^\perp\right\}\right] = L_{\mathbf{y}^+|\mathbf{u}^+} L_{\mathbf{y}^+|\mathbf{u}^+}^\top,$$

$$H_{\mathbf{pp}|\mathbf{u}^+} := \mathrm{Var}\left[\mathbb{E}\left\{\mathbf{p}(t) \mid (\mathbf{u}^+)^\perp\right\}\right] = L_{\mathbf{p}|\mathbf{u}^+} L_{\mathbf{p}|\mathbf{u}^+}^\top$$

CCA: Doing SVD of the *normalized conditional Hankel matrix*

$$\hat{H}_{\mathbf{y}^+\mathbf{p}|\mathbf{u}^+} := L_{\mathbf{y}^+|\mathbf{u}^+}^{-1} H_{\mathbf{y}^+\mathbf{p}|\mathbf{u}^+} L_{\mathbf{p}|\mathbf{u}^+}^{-\top}$$

Same as introducing a **weighting matrix on the left**

$$\mathbf{z}^c(t) \to L_{\mathbf{y}^+|\mathbf{u}^+}^{-1} \mathbf{z}^c(t)$$

Weighting on the right side does not make sense....

ASYMPTOTIC SOLUTION (CONSISTENCY)

For $N \to \infty$ the estimates tend to satisfy

$$\begin{bmatrix} A \\ C \end{bmatrix} \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} = \begin{bmatrix} \Sigma_{\overline{\mathbf{x}}_1 \overline{\mathbf{x}}|\mathbf{u}^+} \\ \Sigma_{\overline{\mathbf{y}\mathbf{x}}|\mathbf{u}^+} \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{bmatrix} \Sigma_{\mathbf{u}^+\mathbf{u}^+|\overline{\mathbf{x}}} = \begin{bmatrix} \Sigma_{\overline{\mathbf{x}}_1 \mathbf{u}^+|\overline{\mathbf{x}}} \\ \Sigma_{\mathbf{y}\mathbf{u}^+|\overline{\mathbf{x}}} \end{bmatrix}$$

Solve in terms of **Conditional Covariances**:

$$\Sigma_{\hat{\mathbf{x}}\,\hat{\mathbf{x}}|\mathbf{u}^+} = \Sigma_{\hat{\mathbf{x}}^c\,\hat{\mathbf{x}}^c}$$

$$\Sigma_{\mathbf{u}^+\mathbf{u}^+|\overline{\mathbf{x}}} = E\{\left[\mathbf{u}_t^+ - E(\mathbf{u}_t^+ \mid \overline{\mathbf{x}}(t))\right]\left[\mathbf{u}_t^+ - E(\mathbf{u}_t^+ \mid \overline{\mathbf{x}}(t))\right]^\top\}$$

CONSISTENCY CONDITION AND ILL-CONDITIONING

Jansson-Wahlberg consistency condition:

$$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} = \Sigma_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c} \quad \text{MUST BE NON SINGULAR!}$$

$\Sigma_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c} \; (= \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+})$ may be ILL– CONDITIONED! $\Rightarrow$

The computation of the parameters (A, C) of the regression will be **ill-conditioned**: **random fluctuation errors in the data will be amplified**. $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ ILL-CONDITIONED $\Leftrightarrow$ Rowspaces of $\hat{\mathbf{X}}_t$ and $\mathbf{U}_{[t,T]}$ are “NEARLY PARALLEL”

Similar analysis holds for $(\mathcal{K}_1, \mathcal{K}_2)$ and $\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}$.

Crucial question: how "parallel" are the rowspaces of

$$U_{[t,T]} \qquad \text{and} \qquad \bar{X}_t = \hat{X}_t + \Gamma_k^{-L} H_k U_{[t,T]}$$

If some (canonical) angles are nearly zero $\Rightarrow$ the computation of the parameters (A, C) and $(\mathcal{K}_1, \mathcal{K}_2)$ of the regression will be **ill-conditioned** (large errors).

CONDITIONING OF SUBSPACE IDENTIFICATION

The conditioning of the problem (*) is determined by the singular values of the conditional covariances

$$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} \qquad \text{and} \qquad \Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}$$

$$\Pi := E\left[\mathbf{u}_t^+ \, \hat{\mathbf{x}}(t)^\top\right] \qquad \bar{\Pi} := E\left[\mathbf{u}_t^+ \, \hat{\bar{\mathbf{x}}}(\mathbf{t})^\top\right] \qquad \Lambda_u = \mathrm{Cov}\left[\mathbf{u}_t^+\right]$$

$$\hat{\Pi} := L_{\mathbf{u}^+}^{-1} \Pi L_{\hat{\mathbf{x}}}^{-\top} \qquad \hat{\bar{\Pi}} := L_{\mathbf{u}^+}^{-1} \bar{\Pi} L_{\bar{\mathbf{x}}}^{-\top}$$

Singular values of $\hat{\Pi}$ are cosines of the **canonical angles** between $\hat{\mathcal{X}}_t$ and $\mathcal{U}_{[t,T]}$. **Condition numbers:**

$$\kappa\left(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}\right) \leq \kappa\left(\Sigma_{\hat{\mathbf{x}}}\right) \frac{1-\sigma_{min}^2(\hat{\Pi})}{1-\sigma_{max}^2(\hat{\Pi})}$$

$$\kappa\left(\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}\right) \leq \kappa\left(\Lambda_u\right) \frac{1}{1-\sigma_{max}^2(\hat{\bar{\Pi}})}$$

CONDITIONING OF SUBSPACE IDENTIFICATION (back to)

- singular values of $\hat{\Pi}$ = singular values of $\hat{\Pi}_d \equiv E\left[\mathbf{u}_t^+ \hat{\mathbf{x}}_d(t)^\top\right]$ cosines of the canonical angles of the spaces spanned by $\mathbf{u}_t^+$ and the **deterministic state** $\hat{\mathbf{x}}_d(t)$

- singular values of $\bar{\Pi} \equiv E\left[\mathbf{u}_t^+ \hat{\bar{\mathbf{x}}}(\mathbf{t})^\top\right]$ cosines of the canonical angles of the spaces spanned by $\mathbf{u}_t^+$ and $\bar{\mathbf{x}}(t)$

- conditioning of the input $\kappa\left(\Lambda_{\mathbf{u}^+}\right)$ large when the amplitude of the spectrum of $\mathbf{u}$ varies widely

PRINCIPAL ANGLES (Canonical correlations)

Introduce Cholesky factors:

$$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = L_{\hat{\mathbf{x}}} L_{\hat{\mathbf{x}}}^{\top} \qquad \Sigma_{\mathbf{u}^+,\mathbf{u}^+} = L_{\mathbf{u}^+} L_{\mathbf{u}^+}^{\top}$$

Normalized Cross-Covariance (Correlation Matrix)

$$\Pi := L_{\mathbf{u}^+}^{-1} \mathrm{Cov}\,\{\mathbf{u}_t^+\, \hat{\mathbf{x}}(t)\}\, L_{\hat{\mathbf{x}}}^{-\top}$$

Singular values of Π are cosines of the **canonical angles** between $\hat{\mathbf{X}}_t$ and $\mathbf{U}_{[t,T]}$.

$$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} = L_{\hat{\mathbf{x}}} \left[I - \Pi^{\top}\Pi \right] L_{\hat{\mathbf{x}}}^{\top}$$

$\sigma_{\mathrm{MAX}}\{\hat{\mathbf{X}}_t, \mathbf{U}_{[t,T]}\} \simeq 1 \;\Leftrightarrow\; \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ **Nearly Singular !**

CCA OF FUTURE INPUTS AND STATE SPACE

Given an input with assigned spectrum $\Phi_{\mathbf{u}}$. Which systems $F(z)$ have the SMALLEST canonical angles of (the spaces spanned by) $\mathbf{u}_t^+$ and $\mathbf{x}_d(t)$ (worst conditioning of the identification problem) ??

$\sigma_k(\mathcal{X}_d, \mathcal{U}^+)$ cosines of **Canonical Angles** btw. the subspaces

$$\mathcal{U}^+ \qquad \text{and} \qquad \mathcal{X}_d := \text{span}\{\mathbf{x}_d(t)\} \subset \mathcal{U}^-$$

Fact:

$$\sigma_k(\mathcal{X}_d, \mathcal{U}^+) \leq \sigma_k(\mathcal{U}^-, \mathcal{U}^+) \quad , \quad k = 1, 2, \ldots$$

Maximal when

$$\sigma_k(\mathcal{X}_d, \mathcal{U}^+) = \sigma_k(\mathcal{U}^-, \mathcal{U}^+) \quad k = 1, 2, \ldots, n_d$$

if and only if the first n_d principal directions of $\mathcal{U}^-$ for the pair of subspaces $(\mathcal{U}^-, \mathcal{U}^+)$ span $\mathcal{X}_d$.

PROBING INPUTS (ASYMPTOTICS FOR $N, T \to \infty$)

Theorem *Assume* $\mathbf{u}$ *has given rational spectral density matrix* Φ_u*. The maximal canonical correlation coefficients* $\sigma_k(\mathbf{X}, \mathbf{U}^+)$ *are obtained when, and only when there are* n_d *principal zeros of the spectral density matrix* Φ_u *of* $\mathbf{u}$ *cancelling all the poles of the deterministic transfer function* $F(z) = C(zI - A)^{-1}B + D$*.*

How to deal with ill-conditioning? Sometimes Decoupling + Orthogonalization helps.

ASYMPTOTIC VARIANCE OF A, C

THEOREM 5 *Under standard assumptions on the true innovation noise, the estimation errors* $\tilde{A}_N := \hat{A}_N - A$, $\tilde{C}_N := \hat{C}_N - C$ *are asymptotically Normal,*

$$\lim_{N\to\infty} N\,\mathbb{E}\left\{\text{vec}\left(\tilde{A}_N\right)\,\text{vec}\left(\tilde{A}_N\right)^\top\right\} = \left\{\Sigma^{-1}_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}\otimes[MH_s]\right\}\cdot \cdot \sum_{|\tau|\leq k}\Sigma_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}(\tau)\otimes\Sigma_{\bar{\mathbf{e}}^+\bar{\mathbf{e}}^+}(\tau)\cdot\left\{\Sigma^{-1}_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}\otimes[MH_s]\right\}^\top$$

$$\lim_{N\to\infty} N\,\mathbb{E}\left\{\text{vec}\left(\tilde{C}_N\right)\,\text{vec}\left(\tilde{C}_N\right)^\top\right\} = \left\{\Sigma^{-1}_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}\otimes[RH_s]\right\}\cdot \cdot \sum_{|\tau|<k}\Sigma_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}(\tau)\otimes\Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau)\cdot\left\{\Sigma^{-1}_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}\otimes[RH_s]\right\}^\top$$

NOTATIONS

$$M := \begin{bmatrix} K & \Gamma^{\dagger} \end{bmatrix} - A \begin{bmatrix} \Gamma^{\dagger} & 0_{n\times m} \end{bmatrix} \qquad R := \begin{bmatrix} I_m & 0_{m\times m(k-1)} \end{bmatrix} - C\Gamma^{\dagger}$$

Γ the observability matrix in a certain basis.

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

$$\mathbf{e}_t^+ \quad := \quad \begin{bmatrix} \mathbf{e}(t) \\ \mathbf{e}(t+1) \\ \vdots \\ \mathbf{e}(T-1) \end{bmatrix} \qquad \bar{\mathbf{e}}_t^+ := \begin{bmatrix} \mathbf{e}_t^+ \\ \mathbf{e}(T) \end{bmatrix}$$

$$\Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau) \quad := \quad \mathbb{E}\{\mathbf{e}_{t+\tau}^+ (\mathbf{e}_t^+)^{\top}\} \quad \Sigma_{\bar{\mathbf{e}}^+\bar{\mathbf{e}}^+}(\tau) = \mathbb{E}\{\bar{\mathbf{e}}_{t+\tau}^+ (\bar{\mathbf{e}}_t^+)^{\top}\}$$

Valid for N4SID, MOESP, and also CCA.

- $\Sigma^{-1}_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c} = \Sigma^{-1}_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ Very “large” for ill-conditioned problems, the variance of the estimation errors will also be large.

- No (or white) input: $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} \equiv \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$

PREVIOUS AVAILABLE RESULTS

[Bauer, Bauer-Ljung, Bauer-Jansson]: asymptotic formulas valid for $N \to \infty$ AND $p := t - t_0$ (past data horizon), tending to infinity with N at a certain rate

Estimates neglect transient due to FINITE-INTERVAL DATA. Consistency only for $p \to \infty$

Different asymptotic formulas for different methods, CCA, MOESP, N4SID etc. Complicated and difficult to use.

Aymptotic formulas should be valid for FINITE p and "transient" estimates (in practice can only regress on **finite past**). Stationary approxim's are biased for finite p.

APPLICATIONS

Assume for simplicity that A has simple eigenvalues.

here is an eigenvalue λ^i of A such that the difference between the i-the eigenvalue of $\hat{A}_N$, $\hat{\lambda}_N^i$, and λ^i, satisfies

$$\hat{\lambda}_N^i - \lambda^i \simeq \frac{v_i^\top \tilde{A}_N u_i}{v_i^\top u_i} + O(\| \tilde{A}_N \|^2)$$

where v_i and u_i are the normalized left and right eigenvectors of A corre-spoding to λ^i.

$$N\,\mathbb{E}\,(\hat{\lambda}_N^i - \lambda^i)^2 = \frac{1}{(v_i^\top u_i)^2}(u_i^\top \otimes v_i^\top)\,N\,\mathbb{E}\left\{\mathrm{vec}\left(\tilde{A}_N\right)\,\mathrm{vec}\left(\tilde{A}_N\right)^\top\right\}(u_i \otimes v_i)$$

Note that $(v_i^\top u_i)^2$ is the square of the cosine of the angle between the two eigenvectors and is equal to one if the matrix A is symmetric (in which case $v_i = u_i$).

ASYMPTOTIC VARIANCE OF (B,D)

The vectorized parameter estimates $\mathrm{vec}(\hat{\mathcal{K}}_{1,N})$ $\mathrm{vec}(\mathcal{K}_{2,N})$ form an asymptotically Gaussian sequence

$$\begin{aligned}
\mathrm{AsVar}\left(\sqrt{N}\mathrm{vec}(\hat{\mathcal{K}}_{1,N})\right) &= \bar{G}\left\{\sum_{|\tau|\le k}\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}(\tau)\otimes\Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau)\right\}\bar{G}^\top \\
\mathrm{AsVar}\left(\sqrt{N}\mathrm{vec}(\hat{\mathcal{K}}_{2,N})\right) &= G\left\{\sum_{|\tau|<k}\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}(\tau)\otimes\Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau)\right\}G^\top
\end{aligned}$$

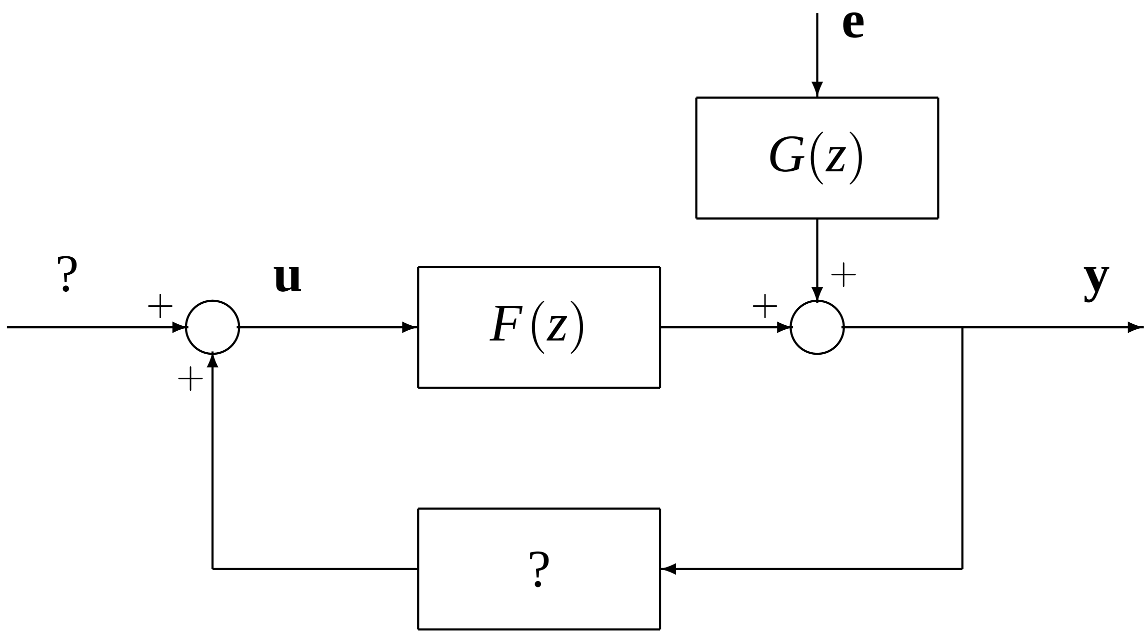
$$G := \Sigma^{-1}_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}\otimes[RH_s], \qquad \bar{G} := \Sigma^{-1}_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}\otimes[M\bar{H}_s]$$

R and M being as before, and,

$$\Sigma_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+|\bar{\mathbf{x}}}(\tau) := \mathbb{E}\{\tilde{\mathbf{u}}^+_{t+\tau}(\tilde{\mathbf{u}}^+_t)^\top\}, \qquad \Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau) = E\{\mathbf{e}^+_{t+\tau}(\mathbf{e}^+_t)^\top\}$$

$\tilde{\mathbf{u}}^+_{t+\tau}$ the τ-steps ahead stationary shift of the random vector $\tilde{\mathbf{u}}^+_t := \mathbf{u}^+_t - \mathbb{E}\left[\mathbf{u}^+_t \mid \bar{\mathbf{x}}(t)\right]$.

SUBSPACE IDENTIFICATION WITH FEEDBACK



+ $F(\infty) = 0$.

PROBLEMS WITH STATE CONSTRUCTION

$$\mathbf{y}(t+h) = CA^h\mathbf{x}(t) + \text{“}terms\ in\ \mathbf{U}_t^+\text{”} + \text{“}terms\ in\ \mathbf{E}_t^+\text{”} \qquad h = 0, 1, \ldots, k$$

Classical (N4SID, CVA, MOESP) construct the state space via the oblique projection

$$E_{\|\mathbf{U}_t^+}\left[\mathbf{Y}_t^+ \mid \mathbf{Y}_t^- \vee \mathbf{U}_t^-\right]$$

Needs $\mathbf{E}_t^+ \perp \mathbf{U}_t^+$ which is equivalent to *Absence of Feedback* from $\mathbf{y}$ to $\mathbf{u}$. (Granger)

Need an **alternative way to construct the state space**, see the discussion in *Ljung-McKelvey 1996*

REMEDY (Jansson 2003/Chiuso-Picci 2004)

FACT: $\mathbf{x}(t)$ is also the state space of the **predictor model**

$$\begin{cases} \mathbf{x}(t+1) &= (A-KC)\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{y}(t) \\ \hat{\mathbf{y}}(t \mid t-1) &= C\mathbf{x}(t) \end{cases}$$

$$\hat{\mathbf{y}}(t+h \mid t) = C(A-KC)^h \mathbf{x}(t) + \text{“}terms\ \ in\ \ \mathbf{U}_t^+ \vee \mathbf{Y}_t^+\text{”}$$

$$\mathbf{X}_t^{+/-} = E_{\|\mathbf{U}_t^+ \vee \mathbf{Y}_t^+}\left[\hat{\mathbf{Y}}_t^+ \mid \mathbf{U}_t^- \vee \mathbf{Y}_t^-\right]$$

Jansson 2003 Compute predictor space removing the effect of undesired terms pre-estimating Markov parameters of predictor using an ARX model.

“PREDICTOR IDENTIFICATION ALGORITHM:

1. Compute the oblique predictors

$$\hat{\mathbf{y}}(t+h \mid t) := E_{\|\mathbf{U}_{[t,t+h)} \vee \mathbf{Y}_{[t,t+h)}} \left[\mathbf{y}(t+h) \mid \mathbf{Y}_{[t_0,t)} \vee \mathbf{U}_{[t_0,t)} \right]$$

2. Compute $\hat{\mathbf{X}}_t^{+/-}$ as “best” n-dimensional approximation of the space spanned by $\hat{\mathbf{y}}(t+h \mid t)$, $h = 0, .., k$, repeat for $\hat{\mathbf{X}}_{t+1}^{+/-}$

3. Solve regression in the least squares sense to get $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{K}$.

COMMENTS:

- The classical subspace procedure to construct the state space turns out to be WRONG if data are collected in closed-loop.

- Subspace methods based on the **predictor model** work also with feedback !

- Predictor is always stable (joint spectrum bounded away from zero $\Rightarrow \; |\lambda(A-KC)| < 1$.)

- Ideally predictor space can be constructed without any assumption on feedback channel.

REMARKS

1. Predictor identification “ideally” yields consistent estimators

2. Practically need to work with **finite past** starting from a certain time t_0.

3. If number of data points ($[y_t, y_{t+1}, .., y_{t+N}]$) $N \longrightarrow \infty$, but $t - t_0$ **fixed and finite** Consistency not guaranteed.

4. “Transient” predictors (transient Kalman filter) involve also the dynamics of $\mathbf{u}$!

REFERENCES

1. B. D. O. Anderson, The inverse problem of stationary covariance generation, *J. Statistical Physics* 1:133–147, 1969.

2. B. D. O. Anderson, A System Theory Criterion for Positive–Real Matrices, *SIAM Journal on Control*, 5, 2:171–182, 1967.

3. K.S. Arun and S.Y. Kung, Balanced approximation of stochastic systems, *SIAM Journal on Matrix Analysis and Applications*, **11**: 42–68, 1990.

4. M. Aoki, *State Space Modeling of Time Series*, 2nd edition, Springer-Verlag, 1991.

5. Chiuso, A. and G. Picci (2004). The asymptotic variance of subspace estimates. *Journal of Econometrics* **118**(1-2), pp. 257–291.

6. Chiuso, A. and G. Picci (2004). On the ill-conditioning of subspace identification with inputs. *Automatica*, **40**, 575-589.

7. Dahlen, A. Lindquist A. and Mari, J, Experimental evidence showing that stochastic subspace identification methods may fail. *Systems Control Lett.* **34** (1998), no. 5, 303–312.

8. U.B. Desai, D. Pal and R.D. Kikpatrick, A realization approach to stochastic model reduction *International Journal of Control*, 42: 821–838 1985.

9. P. Faurre, M. Clerget, F. Germain, *Opérateurs Rationnels Positifs*, Dunod, 1979.

10. G. H. Golub and C. R. Van Loan, *Matrix Computations* (2nd ed.). The Johns Hopkins Univ. Press (1989).

11. A. Lindquist and G. Picci, Canonical Correlation Analysis, Approximate Covariance Extension and Identification of Stationary Time Series, *Automatica*, **32** (1996), 709-733.

12. A.Tether, Constructio of minimal state-variable models from input-output data, *IEEE Trans. Automatic Control* **AC-15** (1971), pp. 427-436.

13. P. van Overschee and B. De Moor, *Subspace algorithms for stochastic identification problem*, Automatica **3** (1993), 649-660.

14. Overschee, P. Van and B. De Moor (1994). N4SID: Subspace algorithms for the identification of combined deterministic– stochastic systems. *Automatica* **30**, 75–93.

15. Verhaegen, M. and P. Dewilde (1992). Subspace model identification, part 1. the output-error state-space model identification class of algorithms; part 2. analysis of the elementary output-error state-space model identification algorithm. *Int. J. Control* **56**, 1187–1210 & 1211–1241.

16. N. Wiener, Generalized Harmonic Analysis, in *The Fourier Integral and Certain of its Applications*, Cambridge U.P. 1933.

17. J. C. Willems, *Least squares stationary optimal control and the algebraic Riccati equation,* IEEE Trans. Automatic Control **AC-16** (1971), pp. 621–634.

18. H. P. Zeiger and A. J. McEwen, *Approximate linear realization of given dimension via Ho's algorithm*, IEEE Trans. Automatic Control **AC-19** (1974), p. 153.

References on Closed Loop

[1]. A. Chiuso, G. Picci (2004), "Consistency Analysis of some Closed-Loop Subspace Identification Methods" *Proc. of the 43rd IEEE Conf. on Decision and Control, 2004*, Nassau, Paradise Island.

[2]. A. Chiuso and G. Picci (2005), "Consistency Analysis of some Closed-loop Subspace Identification Methods", *Automatica: special issue on System Identification*, **41** pp. 377-391.

[3]. A. Chiuso, G. Picci (2005), "Prediction Error vs. Subspace Methods in Closed Loop Identification". *Proc of the 16th IFAC World Congress*, Prague.

[4]. M. Jansson (2003) “Subspace identification and ARX modelling”, in *Proc. of the IFAC Int. Symposium on System Identification (SYSID)*, Rotterdam, 2003.

[5]. W. Lin, S.J. Qin and L. Ljung, (2004)“A Framework for Closed-Loop Subspace Identification with Innovation Estimation”, *Texas-Wisconsin Modeling and Control Consortium Technical Report 07-2004*, available at www.che.wisc.edu/jbr-group/tech-reports/twmcc-2004-07.pdf.

[6]. L. Ljung and T. McKelvey (1996). Subspace identification from closed loop data. *Signal Processing* **52**(2), 209–216.

[7]. J. Qin, L. Ljung (2003) “Closed-Loop Subspace Identification with Innovation Estimation”, in *Proc. of the IFAC Int. Symposium on System Identification (SYSID)*, Rotterdam, 2003.