MODELING, FILTERING AND IDENTIFICATION OF

MULTIVARIABLE STOCHASTIC SYSTEMS

GIORGIO PICCI

Department of Information Engineering,
Università di Padova, Italy

March 2012

Main Objectives of the course

To provide a working knowledge and a thorough understanding of modern state-space identification methods, called SUBSPACE IDENTIFICATION methods. These are the only working and easy-to-use methods for the identification of MULTIVARIABLE SYSTEMS

1. Multivariable linear stochastic models
2. Kalman filtering
3. Statistical methods: PCA and CCA
4. Numerical LinearAlgebra: SVD and LQ decompositions
5. Stochastic model reduction/approximation
6. Asymptotic statistical properties (consistency, efficiency)

PART ONE

1. Review of State Space models of stationary processes
   - General facts about stationary processes
   - State Space models and spectral factorization
   - The Linear matrix Inequality (LMI)
   - Predictor spaces and Kalman filtering
2. The Stochastic Realization Problem from covariance data
   - Problem Statement
   - Deterministic realization theory
   - Canonical correlation analysis
3. The Stochastic Realization Problem from sample covariance data
   - Statistical estimation by the method of moments
   - Properties of the solution
   - Early algorithms (Aoki)
   - The CCA algorithm

Wide-sense stationary random processes

$\mathbf{y} = \{\mathbf{y}(t, \omega)\}$ discrete-time m-dimensional random process $t \in [t_0, +\infty)$.
Expected value: $\mathbb{E}\,\mathbf{y}(t) = \int_\Omega \mathbf{y}(t, \omega)\, dP = \mu(t)$
can be subtracted off. All random quantities will be **zero mean**. Assume a finite Covariance function:

$$\mathbb{E}\,\mathbf{y}(t)\mathbf{y}(s)^\top = \Lambda(t, s)\,, \qquad m \times m \quad \text{matrix function}\,.$$

This is the basic mathematical description of the process. A *second order process* is the equivalence class of all stochastic process having (zero mean and) the same covariance function. Contains a Gaussian representative. Second order processes can be described by Linear models.

$\mathbf{y}$ is a (wide sense) stationary process if its covariance function depends on the difference $t-s$: $\Lambda(t, s) \equiv \Lambda(t-s)$.
We shall study stationary processes on the time line $\mathbb{Z}\,(t_0 = -\infty)$.

Hilbert space setting for second order processes

The closure in $L^2(\Omega, P)$ of all finite linear combinations of the random variables $\mathbf{y}_k(t), k = 1, 2, \ldots, m, t \in \mathbb{Z}$, is a Hilbert space

$$\mathbf{H}(\mathbf{y}) := \operatorname{span}\{\mathbf{y}_k(t)\,;\, k = 1, 2, \ldots, m\,;\, t \in \mathbb{Z}\} \equiv \operatorname{span}\{\mathbf{y}(t)\,;\, t \in \mathbb{Z}\}$$

with inner product $\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle = \mathbb{E}\{\boldsymbol{\xi}\bar{\boldsymbol{\eta}}\}$.
The shift operator $\mathbf{U} : \mathbf{H}(\mathbf{y}) \to \mathbf{H}(\mathbf{y})$ is the linear extension of

$$\mathbf{U}\mathbf{y}_k(t) := \mathbf{y}_k(t+1)\,, \qquad k = 1, 2, \ldots, m, t \in \mathbb{Z}$$

is Unitary (preserves inner product).

Purely non deterministic stationary random processes

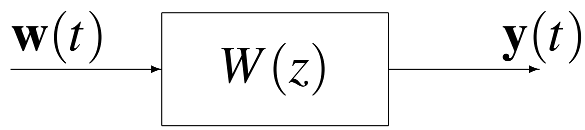
A stationary random process is purely non deterministic (p.n.d) if it can be represented as the output of a causal ℓ^2-stable linear system driven by a white noise

$$\mathbf{y}(t) = \sum_{k=-\infty}^{t} W(t-k)\mathbf{w}(k)$$

$\{\mathbf{w}(t)\}$ p-dimensional **white noise** process of variance $\mathbb{E}\,\mathbf{w}(t)\mathbf{w}(s)^\top = I_p\delta(t-s)$.
The $m \times p$ impulse response $W(t)$ is a causal function in ℓ^2: $W(t) = 0$ for $t < 0$.
The Fourier transform has an analytic extension $W(z)$ to $\{|z| > 1\}$ in H^2.
The representation is highly non unique. Input white noise is a *latent variable*; special white input is the innovation process = one step prediction error given the infinite past.

Spectrum from shaping filters

Shaping filter representation



Wiener-Kintchin formula gives the **spectral density matrix** $\Phi(e^{j\theta})$ of $\{\mathbf{y}(t)\}$

$$\Phi(e^{j\theta}) = \sum_{k=-\infty}^{+\infty} e^{-j\theta\tau}\Lambda(\tau) = W(e^{j\theta})W(e^{-j\theta})^{\top} \qquad \text{spectral factorization.}$$

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

The covariance function $\Lambda(\tau)$ of a p.n.d process admits Fourier transform.

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

Shaping filters and ARMA models

Assume $W(z)$ is a rational matrix function. Since $W(z)$ is **stable** i.e. analytic in $\{|z| > 1\}$, can be written as a ratio of polynomial matrices $W(z) = D(z)^{-1}\,N(z)$ with $\det D(z) \neq 0$ in $\{|z| > 1\}$;

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

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

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

There are *many* ARMA model representations !

Purely deterministic stationary random processes

$\mathbf{y}$ is a purely deterministic (p.d) process if it has zero innovation. Can be predicted exactly based on the infinite past.
Example (elementary)

$$\mathbf{y}(t) = \sum_{k=1}^{\nu} \mathbf{x}_k \cos \omega_k t + \mathbf{z}_k \sin \omega_k t\,, \qquad \mathbb{E}\,\mathbf{x}_k^2 = \mathbb{E}\,\mathbf{z}_k^2 = \sigma_k^2$$

all random variables $\{\mathbf{x}_k\,,\,\mathbf{z}_k\,;\,k = 1, 2, \ldots, \nu\}$ mutually uncorrelated.

$$\mathbf{H}(\mathbf{y}) := \operatorname{span}\{\mathbf{x}_k\,,\,\mathbf{z}_k\,;\,k = 1, 2, \ldots, \nu\} = \mathbf{H}_t^-(\mathbf{y}) = \mathbf{H}_t^+(\mathbf{y})$$

The spectral density does not exists. Formally is a sum of delta functions (spectral lines).

Wold decomposition

Theorem 1 (Wold decomposition) *Every stationary process can be decomposed uniquely as*

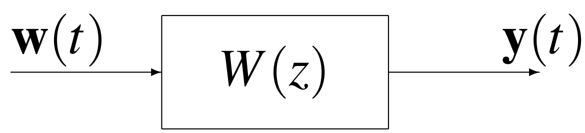
$$\mathbf{y}(t) = \mathbf{u}(t) + \mathbf{v}(t) \quad ,$$

where $\{\mathbf{u}(t)\}$ *is p.n.d.,* $\{\mathbf{v}(t)\}$ *is p.d. and* $\{\mathbf{u}(t)\}$ *and* $\{\mathbf{v}(t)\}$ *are uncorrelated; i.e.* $\langle \mathbf{u}_k(t), \mathbf{v}_j(s) \rangle = 0$*,* $\forall t, s \in \mathbb{Z}$*,* $\forall k, j = 1, \ldots, m$.

The spectrum of $\mathbf{y}$ is the sum of an absolutely continuous part (spectral density) plus a singular part (spectral lines + ..).

State space models of random processes

Any p.n.d. process $\{\mathbf{y}(t)\}$ is the response to some normalized white noise process $\{\mathbf{w}(t)\}$ of a linear *Shaping Filter* of transfer function $W(z)$



If $W(z)$ is rational; i.e. $W(z) = C(zI - A)^{-1}B + D$, can be realized as a state space system

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

$\{\mathbf{w}(t)\}$ p-dimensional **white noise** process of variance $\mathbb{E}\,\mathbf{w}(t)\mathbf{w}(s)^\top = I_p\delta(t-s)$
Initial (random) data

$$\mathbb{E}\,\mathbf{x}_0 = 0 \quad , \qquad \mathrm{Var}\,\mathbf{x}_0 = \Sigma_0 \quad , \qquad \mathbb{E}\,\mathbf{x}_0\,\mathbf{w}(t)^\top = 0 \quad \forall t \geq t_0 \,.$$

Minimality

REACHABILITY: $\quad \operatorname{rank}\begin{bmatrix} B & AB & \dots & A^{n-1}B \end{bmatrix} = n$

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

In any case if Reachability + Observability hold $W(z)$ analytic $\Leftrightarrow \; |\lambda(A)| < 1$
Reachability + Observability are necessary but not sufficient for minimality.
Example

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

this $\mathbf{y}(t)$ is white noise. Particular case $a = 0$; i.e. $W(z) = z^{-1}$. These models have a minimal representation of order $n = 0$.

Stochastic minimality holds iff there are no nontrivial right all-pass divisors of $W(z)$. This is the same as minimality of the spectral factor which will be defined later as minimality of the McMillan degree.

Unnormalized white noises

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t)+\mathbf{v}_1(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t)+\mathbf{v}_2(t) \end{cases}, \qquad \begin{matrix} \mathbf{x}(t_0)=\mathbf{x}_0 \\ t\geq t_0 \end{matrix}$$

with *unnormalized* white inputs:

$$\mathbb{E}\left\{ \begin{bmatrix} \mathbf{v}_1(t) \\ \mathbf{v}_2(t) \end{bmatrix} \begin{bmatrix} \mathbf{v}_1(t)^\top & \mathbf{v}_2(t)^\top \end{bmatrix} \right\} = \begin{bmatrix} Q & S \\ S^\top & R \end{bmatrix} \geq 0.$$

Factorize (full rank)

$$\begin{bmatrix} Q & S \\ S^\top & R \end{bmatrix} = \begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B^\top & D^\top \end{bmatrix}$$

Then $\mathbf{v}_1(t) := B\mathbf{w}(t)$, $\mathbf{v}_2(t) := D\mathbf{w}(t)$ with $\mathbf{w}$ the same normalized white noise

Note: $\mathbf{v}_1(t)$ and $\mathbf{v}_2(t)$ are in general *correlated* white noise processes unless

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

The State process

Since $\mathbb{E}\,\mathbf{x}_0\,\mathbf{w}(t)^\top = 0 \quad \forall t \geq t_0$,

$$\mathbf{x}(t) = A^{t-t_0}\mathbf{x}(t_0) + \sum_{k=t_0}^{t-1} A^{t-1-k}B\mathbf{w}(k) \qquad \text{Orthogonal sum} \qquad (*)$$

Then $\{\mathbf{x}(t)\}$ is a wide-sense **Markov process**; i.e.

$$\hat{\mathbb{E}}\left[\mathbf{x}(t) \mid \mathbf{x}(\tau),\, t_0 \leq \tau \leq s\right] = \hat{\mathbb{E}}\left[\mathbf{x}(t) \mid \mathbf{x}(s)\right] \quad , \qquad \forall t \geq s \quad ,$$

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

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

Satisfies a Lyapunov difference equation

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

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

Proof of the Markov property

Writing $(*)$ with s in place of t, you see that $\mathbf{w}(t) \perp \mathbf{x}(s)$ for all $t \geq s$. Hence if you project

$$\mathbf{x}(t) = A^{t-s}\mathbf{x}(s) + \sum_{k=s}^{t-1} A^{t-1-k}B\mathbf{w}(k)\,,$$

onto $\mathbf{H}(\mathbf{x}(t_0), \ldots, \mathbf{x}(s)) := \mathbf{H}_s^-(\mathbf{x}) \subset \mathbf{H}_s^-(\mathbf{w})$ the second term of the sum projects to zero and the (components of the) first belong to the space and stays unchanged. In fact

$$\hat{\mathbb{E}}\left[\mathbf{x}(t) \mid \mathbf{x}(s)\right] = A^{t-s}\mathbf{x}(s)\,.$$

Conditional orthogonality

Ricordiamo che due sottospazi di variabili aleatorie del secondo ordine $\mathbf{X}, \mathbf{Y}$, si definiscono *condizionatamente scorrelati* o, meglio, *condizionatamente ortogonali*, dato un terzo sottospazio di variabili del secondo ordine $\mathbf{Z}$, se risulta

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \hat{\mathbb{E}}\,(\mathbf{x} \mid \mathbf{Z}), \hat{\mathbb{E}}\,(\mathbf{y} \mid \mathbf{Z}) \rangle \tag{2}$$

per ogni $\mathbf{x} \in \mathbf{X}$, $\mathbf{y} \in \mathbf{Y}$. Notazione: $\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}$
La (2) è equivalente alla

$$\langle \mathbf{x} - \mathbb{E}\,(\mathbf{x} \mid \mathbf{Z}), \mathbf{y} - \mathbb{E}\,(\mathbf{y} \mid \mathbf{Z}) \rangle = \operatorname{cov}(\mathbf{x}, \mathbf{y} \mid \mathbf{Z}) = 0\,,$$

per cui la *covarianza incrociata condizionata di* $\mathbf{x}$ *e* $\mathbf{y}$*, date le variabili del sottospazio* $\mathbf{Z}$*, è zero*.
Che questa sia la naturale versione "debole" dell' indipendenza condizionata si può vedere supponendo che $\mathbf{X}$, $\mathbf{Y}$ e $\mathbf{Z}$ siano popolati da variabili congiuntamente Gaussiane. Allora la densità congiunta condizionata $p_{\mathbf{x}\mathbf{y}}(\cdot \mid \mathbf{Z})$ fattorizza nel prodotto delle due densità $p_{\mathbf{x}}(\cdot \mid \mathbf{Z})\, p_{\mathbf{y}}(\cdot \mid \mathbf{Z})$ (si può inizialmente supporre $\mathbf{Z}$ generato da un vettore z di dimensione finita e notare

che il ragionamento vale, indipendentemente dalla dimensione di $\mathbf{Z}$). La nozione intuitiva di incorrelazione condizionata è poi catturata in modo esplicito dal seguente enunciato (la cui dimostrazione si può ad esempio trovare in [**?**].

Lemma 1 *Si ha* $\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}$ *se e solo se vale una delle tre condizioni equivalenti*

(i) $\hat{\mathbb{E}}(\mathbf{x} \mid \mathbf{Y} \vee \mathbf{Z}) = \hat{\mathbb{E}}(\mathbf{x} \mid \mathbf{Z}) \qquad \mathbf{x} \in \mathbf{X}$

(ii) $\hat{\mathbb{E}}(\mathbf{y} \mid \mathbf{X} \vee \mathbf{Z}) = \hat{\mathbb{E}}(\mathbf{y} \mid \mathbf{Z}) \qquad \mathbf{y} \in \mathbf{Y}$

(iii) $\mathbf{X} \vee \mathbf{Z} \perp \mathbf{Y} \vee \mathbf{Z} \mid \mathbf{Z}$

dove $\mathbf{X} \vee \mathbf{Y}$ *denota lo spazio di Hilbert generato dalle variabili aleatorie dei sottospazi* $\mathbf{X}$ *e* $\mathbf{Y}$ *(ovvero la chiusura della somma vettoriale* $\mathbf{X} + \mathbf{Y}$*).*

The (wide-sense) Markov property

Coordinte-free notion of **Markov process**

Let $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ be a family of subspaces of some ambient Hilbert space of second order random variables. Let

$$\mathbf{X}_t^- := \overline{\text{span}}\{\mathbf{X}_s\,;\, s<t\}, \quad \mathbf{X}_t^+ := \overline{\text{span}}\{\mathbf{X}_s\,;\, s>t\} \qquad \text{strict past and future}$$
$$\bar{\mathbf{X}}_t^- := \overline{\text{span}}\{\mathbf{X}_s\,;\, s\le t\}, \quad \bar{\mathbf{X}}_t^+ := \overline{\text{span}}\{\mathbf{X}_s\,;\, s\ge t\} \qquad \text{past and future}\,.$$

Definition 1 *The family $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ is Markovian if past and future are conditionally uncorrelated given the present; i.e. $\mathbf{X}_t^- \perp \mathbf{X}_t^+ \mid \mathbf{X}_t$ for all t. Any n-dimensional process $\mathbf{x}(t) = [\mathbf{x}_1(t),\dots,\mathbf{x}_n(t)]^\top$ such that $\mathbf{X}_t = \text{span}\{\mathbf{x}_1(t),\dots,\mathbf{x}_n(t)\}$ is a (wide sense)* **Markov proces***.*

By lemma 1 this is equivalent to

$$\hat{\mathbb{E}}\left[\mathbf{x}(t+k)\mid \mathbf{X}_t^-\right] = \hat{\mathbb{E}}\left[\mathbf{x}(t+k)\mid \mathbf{X}_t\right], \quad \hat{\mathbb{E}}\left[\mathbf{x}(t-k)\mid \mathbf{X}_t^+\right] = \hat{\mathbb{E}}\left[\mathbf{x}(t-k)\mid \mathbf{X}_t\right] \quad \forall k\ge 0$$

for all $t\in\mathbb{Z}$. One can exchange $\mathbf{X}_t^{+/-}$ for $\bar{\mathbf{X}}_t^{+/-}$ everywhere.

Conventions on past and future

For a generic process $\mathbf{y}$ both the strict past and future do not contain the present. It is convenient to define past and future in such a way that at least one of them contains the present. A common convention is to *include the present in the future and not in the past*; i.e. to define

$$\mathbf{H}_t^-(\mathbf{y}) := \overline{\text{span}}\{\mathbf{y}(s)\,;\, s < t\} \qquad \mathbf{H}_t^+(\mathbf{y}) := \overline{\text{span}}\{\mathbf{y}(s)\,;\, s \geq t\}$$

For a Markov process this distinction is immaterial since strict past (future) and whole past (or future) play the same role but fo rgeneral processes this is not true.

Conditional orthogonality and state equations

Define the $n+m$ dimensional random vector

$$\mathbf{z}(t) := \begin{bmatrix} \mathbf{x}(t+1) \\ \mathbf{y}(t) \end{bmatrix}$$

Let

$$\mathbf{H}_t^+(\mathbf{z}) := \mathbf{H}_{t+1}^+(\mathbf{x}) \vee \mathbf{H}_t^+(\mathbf{y}), \qquad \mathbf{H}_t^-(\mathbf{z}) := \mathbf{H}_{t+1}^-(\mathbf{x}) \vee \mathbf{H}_t^-(\mathbf{y})$$

Proposition 1 *If* $(\mathbf{x},\ \mathbf{y})$ *satisfy a linear state equation, then one has the conditional orthogonality relation*

$$\mathbf{H}_t^-(\mathbf{z}) \perp \mathbf{H}_t^+(\mathbf{z}) \mid \mathbf{X}_t \qquad t_0 \leq t \tag{3}$$

where $\mathbf{X}_t := \operatorname{span}\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ *is the* **state space of the system at time** t *:.*

Proof:

TO BE ADDED

A geometric approach

Let $\mathbf{X}_t$ denote the state space of the system

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

The state vector is just a choice of basis in $\mathbf{X}_t$. Once a basis in $\mathbf{X}_t$ is fixed all the matrices $(ABCD)$ are uniquely determined. In fact, let $P = \mathbb{E}\,\mathbf{x}(t)\mathbf{x}(t)^\top$, then

$$\begin{cases} A = \mathbb{E}\,\mathbf{x}(t+1)\mathbf{x}(t)^\top P^{-1}\,, \\ C = \mathbb{E}\,\mathbf{y}(t)\mathbf{x}(t)^\top P^{-1}\,, \\ \bar{C} = \mathbb{E}\,\mathbf{y}(t-1)\mathbf{x}(t)^\top\,, \\ \begin{bmatrix} B\mathbf{w}(t) \\ D\mathbf{w}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{x}(t+1) \\ \mathbf{y}(t) \end{bmatrix} - \mathbb{E}\left(\begin{bmatrix} \mathbf{x}(t+1) \\ \mathbf{y}(t) \end{bmatrix} \mid \mathbf{x}(t) \right). \end{cases} \tag{4}$$

so in a sense, constructing a stochastic model is just a matter of finding the state space and choosing a suitable basis.

Conditional orthogonality and state equations

Proposition 2 *Conversely, if* (3) *holds and* $\mathbf{X}_t$ *is finite dimensional with basis* $\{\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)\}$*, then* $\mathbf{x}$, $\mathbf{y}$ *satisfy a state space equation of the form* (1)*.*

Proof : Since $\mathbf{z}(t) \in \mathbf{H}_t^+(\mathbf{z})$ and $\mathbf{x}(t)$ is a basis for $\mathbf{X}_t$

$$\hat{\mathbf{z}}(t \mid t-1) := \hat{\mathbb{E}}\left[\mathbf{z}(t) \mid \mathbf{H}_{t-1}^-(\mathbf{z})\right] = \hat{\mathbb{E}}\left[\mathbf{z}(t) \mid \mathbf{H}_t^-(\mathbf{z})\right] = \hat{E}\left[\mathbf{z}(t) \mid \mathbf{x}(t)\right] = \begin{bmatrix} A \\ C \end{bmatrix} \mathbf{x}(t).$$

Let $\mathbf{z}(t) = \hat{\mathbf{z}}(t \mid t-1) + \mathbf{v}(t)$, then $\mathbf{v}(t)$ is the innovation of $\mathbf{z}$ and hence is white noise. Normalize and define $\mathbf{w}(t)$ so that

$$\mathbf{v}(t) = \begin{bmatrix} B \\ D \end{bmatrix} \mathbf{w}(t)\,.$$

The p.n.d. and p.d. subsystems of a state space model

Proposition 3 *The n random variables $\{\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_n(t)\}$ form a basis in $\mathbf{X}_t$ if and only if $\Sigma(t) := \mathbb{E}\,\mathbf{x}(t)\mathbf{x}(t)^\top > 0$.*

Pick a basis (at time zero) $\mathbf{x} = [\mathbf{x}_s \; \mathbf{x}_d]^\top$ where $\mathbf{x}_s$ is the p.n.d. and $\mathbf{x}_d$ the p.d. component . They span two orthogonal complementary subspaces $\mathbf{X}_s$ and $\mathbf{X}_d$. In fact by Wold decomposition, $\mathbf{H}(\mathbf{x}_s)$ and $\mathbf{H}(\mathbf{x}_d)$ are orthogonal and hence the Markov property holds separately for the two state processes;

$$\begin{bmatrix} \mathbf{x}_s(t+1) \\ \mathbf{x}_d(t+1) \end{bmatrix} = \begin{bmatrix} A_s & 0 \\ 0 & A_d \end{bmatrix} \begin{bmatrix} \mathbf{x}_s(t) \\ \mathbf{x}_d(t) \end{bmatrix} + \begin{bmatrix} B_s \\ B_d \end{bmatrix} \mathbf{w}(t)$$

Note: B_d must be zero since otherwise $\mathbf{x}_d$ would have a non-zero innovation (Prove this!).

Second order description

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

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

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

$$\Lambda(t,s) \quad = \quad \begin{cases} CA^{t-s-1}\,\bar{C}(s)^\top & t > s \\ C\Sigma(t)C^\top + DD^\top & t = s \\ \bar{C}(t)\,(A^\top)^{s-t-1}C^\top & t < s \end{cases}$$

$$\bar{C}(s)^\top := A\Sigma(s)C^\top + BD^\top$$

Asymptotic stationarity of p.n.d. processes

Definition: $\{\mathbf{y}(t)\}$ is *asymptotically stationary* if for $t - t_0 \to \infty$, $\Lambda(t,s) \Rightarrow \Lambda(t-s)$, depends 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

$$
\begin{aligned}
\Sigma_{\mathbf{x}}(t-s) &= A^{t-s}\,\bar{\Sigma} \quad , \qquad t \geq s \quad , \\
\Lambda(t-s) &= \begin{cases} CA^{t-s-1}\,\bar{C}^\top & t > s \\ C\bar{\Sigma}C^\top + DD^\top & t = s \end{cases}
\end{aligned}
$$

where $\bar{C}^\top := A\bar{\Sigma}C^\top + BD^\top$ 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.

Homework

Particularize the formulas above for the case when $\mathbf{x}$ is (stationary) purely deterministic.

In this case $\bar{\Sigma} = \mathbb{E}\,\mathbf{x}_0\mathbf{x}_0^\top = \Sigma_0$ satisfies a homogeneous Lyapunov equation. Assume $\Sigma_0 > 0$ and show that in a suitable basis A must be an orthogonal matrix, namely

$$AA^\top = A^\top A = I$$

Describe the eigenvalues/eigenvectors of an orthogonal matrix.

The Lyapunov equation

Theorem 2 *The discrete Lyapunov equation*

$$X = AXA^{\top} + Q,$$

with arbitrary Q, has a solution (necessarily unique) if and only if the spectrum of A is **unmixing***; i.e. does not contain reciprocal elements, that is $\lambda_k \in \sigma(A) \Rightarrow 1/\lambda_k \notin \sigma(A)$. If $Q = Q^{\top}$ the solution is symmetric.*
Let A have unmixing spectrum and P be the solution. If $Q = BB^{\top}$ and (A, B) is reachable, the number of eigenvalues of A with modulus less [greater] than 1 is equal to the number of positive [negative] eigenvalues of P. In particular P is non singular.

Corollary 1 *Any of the two conditions below imply the remaining one*

i) (A, B) *is reachable*

ii) A *is asymptotically stable*

iii) The equation $X = AXA^{\top} + BB^{\top}$ *has a unique solution* $P = P^{\top} > 0$ *.*

Causal/Acausal models

The state space model

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{w}(t)\,, \\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{w}(t) \end{cases} \tag{5}$$

with $|\lambda(A)| < 1$ is a *causal* or *forward* representation since the equations can be solved to yield

$$\mathbf{x}(t) = \sum_{s=-\infty}^{t-1} A^{t-1-s}B\mathbf{w}(s)\,, \qquad \mathbf{y}(t) = \sum_{s=-\infty}^{t-1} CA^{t-1-s}B\mathbf{w}(s) + D\mathbf{w}(t)$$

which represent $\mathbf{x}(t)$ and $\mathbf{y}(t)$ as a causal function of the past $\{\mathbf{w}(s)\,;\, s \leq t\}$. However this is just a representation out of many possible others. A random process is just a *flow*; has no intrinsic causality. Shaping filter representation could be anti-causal or even of “mixed” causal-anticausal type. The only condition which is really needed is that the impulse response be in $\ell^2(\mathbb{Z})$. For rational processes $\sigma(W(z)) \cap \{|z| = 1\} = \emptyset$. Otherwise the spectrum of $W(z)$ could be arbitrary.

Introduction to rational spectral factorization

Inverse problem (stochastic realization): Given a p.n.d. stationary process $\mathbf{y}$ with a rational spectral density matrix $\Phi(z)$ want to find and classify the state space realizations of $\mathbf{y}$.

We want to classify all Minimal state space models where $\dim \mathbf{x}(t)$ is the smallest possible.

This involves computing the rational spectral factors parametrized in the form $W(z) = C(zI - A)^{-1}B + D$. Here we will not worry about constructing the noise process $\mathbf{w}$.

Will solve this problem obtaining a parametrization of all minimal analytic (i.e. causal) spectral factors.

Minimal spectral factors are of minimal Mc Millan degree. Will also study a more general problem.

Causal-Anticausal splitting of $\Phi(z)$

Covariance function of the process $\mathbf{y}$

$$\Lambda(\tau) = \begin{cases} CA^{\tau-1}\,\bar{C}^\top & \tau > 0 \\ C\bar{\Sigma}C^\top + DD^\top & \tau = 0 \end{cases}$$

where $\quad \bar{C}^\top := A\Sigma C^\top + BD^\top$ and Σ satisfies the Lyapunov equation

$$\Sigma = A\Sigma A^\top + BB^\top\,.$$

Since $\mathbf{y}$ is p.n.d. the Fourier transform of Λ exists. Get the Laurent expansion:

$$\begin{aligned}
\Phi(z) &= \sum_{\tau=-\infty}^{+\infty} \Lambda(\tau) z^{-\tau} \qquad \Lambda(-\tau) = \Lambda(\tau)^\top \\
&= \left[C(zI-A)^{-1}\bar{C}^\top + \Lambda_0/2\right] + \left[\Lambda_0/2 + \bar{C}(z^{-1}I - A^\top)^{-1}C^\top\right] \\
&:= \qquad\qquad \Phi_+(z) \qquad\quad + \qquad\quad \Phi_+(z^{-1})^\top
\end{aligned}$$

Since $|\lambda(A)| < 1$ this is an analytic + co-analytic decomposition.

The positive real part of $\Phi(z)$

A rational $m \times m$ matrix function $\Phi(z)$ is a spectral density iff it satisfies
∘ the para-Hermitian symmetry $\Phi(z) = \Phi(z^{-1})^\top$
∘ $\Phi(e^{j\theta})$ integrable on the unit circle $\Rightarrow$ no poles on the unit circle
∘ positive semidefinite on the unit circle $\Phi(e^{j\theta}) \geq 0$.

Hence $\Phi_+(z)$ is analytic on $\{|z| \geq 1\}$ and

$$\Phi(e^{j\theta}) = \Phi_+(e^{j\theta}) + \Phi_+(e^{-j\theta})^\top = 2\Re e\,\Phi_+(e^{j\theta}) \geq 0$$

therefore $\Phi_+(z)$ is a positive real function, the Positive Real part of $\Phi(z)$, see e.g. the book by Anderson and Vongpanitlerd for the definition and implications of this important property.

Generalization of the additive splitting

Call a rational $m \times m$ matrix function $\Phi(z)$ parahermitian if it satisfies only the first two conditions.

So far we have been showing a causal-anticausal decomposition of the spectrum with A an asymptotically stable matrix. One can have many other additive decompositions of a parahermitian $\Phi(z)$. The poles of $\Phi(z)$ have reciprocal symmetry; i.e. if $\Phi(z)$ has a pole in $z = p_k$ then $1/p_k$ must also be a pole (be it finite or not) of the same multiplicity.

The set of poles, $\sigma(\Phi)$, of a $\Phi(z)$ of degree $2n$ can then be split in two reciprocal subsets σ_1 and σ_2 each containing n complex numbers (repeated according to multiplicity), such that $\sigma_2 = 1/\sigma_1$. This decomposition of the spectrum yields, by partial fraction expansion, a rational additive decomposition of $\Phi(z)$ of the type

$$\Phi(z) = Z(z) + Z(z^{-1})^{\top}, \qquad (\ddagger)$$

where $Z(z)$ is a rational function with poles in σ_1 and those of $Z(z^{-1})^{\top}$ necessarily in $\sigma_2 = 1/\sigma_1$. Clearly, here A need not be asymptotically stable.

Unmixing

Recall that an $n \times n$ matrix A has **unmixed spectrum** if $\sigma(A)$, does not contain reciprocal pairs counting multiplicity. Assume minimality of $(C, A, \bar{C}^\top)$. Then A has unmixed spectrum if and only if the selected pole set $\sigma_1 \equiv \sigma(A)$ has no self-reciprocal elements. It is obvious that this happens if and only if $\sigma_1 \cap \sigma_2 = \emptyset$.

Example 1 *The parahermitian function* $\Phi(z) = \dfrac{K^2}{(z - \frac{1}{2})^2 (z^{-1} - \frac{1}{2})^2}$ *has pole set* $\{\frac{1}{2}, \frac{1}{2}, 2, 2\}$ *so the only non intersecting subsets of two elements are* $\{\frac{1}{2}, \frac{1}{2}\}$ *and* $\{2, 2\}$*. Hence in this case either* $Z(z)$ *is stable and coincides with* $\Phi(z)_+$ *or is totally antistable with poles in* $\{2, 2\}$*.*

Note that the unmixed spectrum condition is exactly the condition insuring that the Lyapunov equation $P - APA^\top = Q$ has a (unique) solution for arbitrary Q.

Rational spectral factorization

The rational spectral factorization problem is, given a parahermitian rational matrix $\Phi(z)$ find rational matrix functions $W(z)$ such that $\Phi(z) = W(z)W(z^{-1})^\top$ In the classical setting $\Phi(z)$ is actually a spectral density and one starts from the additive decomposition $\Phi(z) = \Phi_+(z) + \Phi_+(z^{-1})^\top$ with $\Phi_+(z)$ positive real and looks for *analytic* spectral factors.

We want to generalize this problem. Start from a parahermitian matrix given by a general additive decomposition $(\ddagger)$ where the function $Z(z) := C(zI - A)^{-1}\bar{C}^\top + \frac{1}{2}\Lambda_0$ is not necessarily positive real and look for rational spectral factors with the same poles of $Z(z)$. Note that the existence of spectral factors is not automatically guaranteed in this case and is in fact *equivalent to the the positivity of* $\Phi(e^{j\theta})$ since, irrespective of analiticity, if a spectral factor W exists, then

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

From additive decomposition to spectral factorization

Would like to factor the function given by the additive decomposition $(\ddagger)$, namely

$$\Phi(z) = \begin{bmatrix} C(zI-A)^{-1} & I \end{bmatrix} \begin{bmatrix} 0 & \bar{C}^\top \\ \bar{C} & \Lambda_0 \end{bmatrix} \begin{bmatrix} (z^{-1}I - A^\top)^{-1}C^\top \\ I \end{bmatrix} \qquad (*)$$

as $\Phi(z) = W(z)W(z^{-1})^\top$. Note that this seems impossible at first sight since the matrix in the middle is not positive semidefinite. But there are many constant matrices representing the same quadratic form.

Lemma 2 *Let*

$$M = \begin{bmatrix} P - APA^\top & -APC^\top \\ -CPA^\top & -CPC^\top \end{bmatrix}$$

for some $n \times n$ symmetric matrix P. Then we have

$$\begin{bmatrix} C(zI-A)^{-1} & I \end{bmatrix} M \begin{bmatrix} (z^{-1}I - A^\top)^{-1}C^\top \\ I \end{bmatrix} \equiv 0$$

identically.

Proof : Use the identity

$$P - APA^\top = AP(z^{-1}I - A^\top) + (zI - A)PA^\top + (zI - A)P(z^{-1}I - A^\top)\,. \tag{6}$$

Let $P - APA^\top := Q$ and use the shorthand $A(z) := (zI - A)$. Left multiply by $CA(z)^{-1}$ and right multiply by $A(z^{-1})^{-\top}C^\top$ to get

$$CA(z)^{-1}QA(z^{-1})^{-\top}C^\top = CA(z)^{-1}APC^\top + CPA^\top A(z^{-1})^{-\top}C^\top + CPC^\top$$

which is what we need to show.

The Linear matrix Inequality (LMI)

Lemma 3 *If there exists $P = P^\top$ such that the following Linear Matrix Inequality*

$$M(P) = \begin{bmatrix} P - APA^\top & \bar{C}^\top - APC^\top \\ \bar{C} - CPA^\top & \Lambda_0 - CPC^\top \end{bmatrix} \geq 0 \qquad \textbf{LMI}$$

is satisfied, the parahermitian matrix $\Phi(z)$ admits spectral factors. In fact, let $W(z) = C(zI - A)^{-1}B + D$ where B, D be defined by the factorization

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

then $W(z) = C(zI - A)^{-1}B + D$ is a spectral factor.

The lemma follows from the content of the previous slide. Hence if the LMI has a solution, $\Phi(z)$ is actually a spectral density. Note that no stability of A nor minimality are required.

Exercise: Show that if A is unmixing and $C(zI-A)^{-1}\bar{C}^{\top}$ is a minimal realization then $C(zI-A)^{-1}B$ is also a minimal realization.

Hint: A proof can be based on the fact that $\sigma_1 \cap \sigma_2 = \emptyset$ implies that the McMillan degree $\delta(Z) = \frac{1}{2}\delta(\Phi) = n$ so that the dimension of A is $n \times n$ which implies that (A, B) must be reachable otherwise the dimension of a minimal realization of W would be smaller than n and hence $\Phi(z) = W(z)W(z^{-1})^{\top}$ would have McMillan degree smaller than $2n$ which is in contrast with the minimality of the realization of $Z(z)$.

If A is not unmixing there may be ambiguities in forming the decomposition $\Phi(z) = Z(z) + Z(z^{-1})^{\top}$. Actually in some case the decomposition may not even exist.

The Linear matrix Inequality (LMI): Necessity

Conversely want to show that for any spectral factor there is a $P = P^\top$ satisfying the LMI constructed with the parameters $(A, C, \bar{C}, \Lambda_0)$ of some $Z(z)$. Note that whenever $W(z) = C(zI - A)^{-1}B + D$ is a spectral factor then we can write

$$\Phi(z) = \begin{bmatrix} C(zI-A)^{-1} & I \end{bmatrix} \begin{bmatrix} BB^\top & BD^\top \\ DB^\top & DD^\top \end{bmatrix} \begin{bmatrix} (z^{-1}I - A^\top)^{-1}C^\top \\ I \end{bmatrix} \qquad (\dagger)$$

The LMI is defined once we show that this $\Phi(z)$ has a parahermitian additive decomposition like $(\ddagger)$, or equivalently $(*)$. Now in force of Lemma 2, this amounts to showing that there are matrices $P, \tilde{C}$ and R such that

$$\begin{bmatrix} BB^\top & BD^\top \\ DB^\top & DD^\top \end{bmatrix} = \begin{bmatrix} 0 & \tilde{C}^\top \\ \tilde{C} & R \end{bmatrix} + \begin{bmatrix} P - APA^\top & -APC^\top \\ -CPA^\top & -CPC^\top \end{bmatrix}, \qquad (**)$$

since by multiplying this equation on the left by $[C\,(zI - A)^{-1} \;\; I]$ and on the right by $[C\,(Iz^{-1} - A)^{-1} \;\; I]^\top$ and by using Lemma 2 we would conclude that $\Phi(z)$ has a parahermitian additive decomposition of the same form as $(*)$ with $Z(z) = C\,(zI - A)^{-1}\tilde{C} + \frac{1}{2}R$.

The Linear matrix Inequality (LMI): Necessity

Lemma 4 *Let* $W(z) = C(zI - A)^{-1}B + D$ *be a rational spectral factor of* $\Phi(z)$ *with an unmixing* A *matrix. Then there is a corresponding additive decomposition* $\Phi(z) = Z(z) + Z(z^{-1})^{\top}$ *with* $= C(zI - A)^{-1}\tilde{C}^{\top} + \frac{1}{2}\Lambda_0$ *and a unique* $P = P^{\top}$ *satisfying the linear matrix inequality*

$$\begin{bmatrix} P - APA^{\top} & \tilde{C}^{\top} - APC^{\top} \\ \tilde{C} - CPA^{\top} & \Lambda_0 - CPC^{\top} \end{bmatrix} \geq 0\,.$$

If $|\lambda(A)| < 1$ *and* (A, B) *is reachable, then* $P > 0$*.*

Proof: For analytic spectral factors ($|\lambda(A)| < 1$) the lemma is nearly obvious. A short probabilistic proof goes as follows. Take a stationary realization with transfer function $W(z)$ and compute the variance of

$$\begin{bmatrix} B \\ D \end{bmatrix} \mathbf{w}(t) = \begin{bmatrix} \mathbf{x}(t+1) \\ \mathbf{y}(t) \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} \mathbf{x}(t) = \mathbf{z}(t) - \hat{\mathbb{E}}\left[\mathbf{z}(t) \mid \mathbf{x}(t)\right]$$

and notice that the variance of the quantity on the left must obviously be positive semidefinite. By Pithagora's theorem the variance of the second member is

$$\begin{bmatrix} P & \bar{C}^\top \\ \bar{C} & \Lambda_0 \end{bmatrix} - \begin{bmatrix} APA^\top & APC^\top \\ CPA^\top & CPC^\top \end{bmatrix}$$

since

$$\mathbb{E}\,\mathbf{y}(t)\mathbf{x}(t+1)^\top = \mathbb{E}\,(C\mathbf{x}(t)+D\mathbf{w}(t))\,(\mathbf{x}(t)^\top A^\top + \mathbf{w}(t)^\top B^\top) = CPA^\top + DB^\top = \bar{C}$$

(this follows since the model is causal). In this case the solution P of the LMI is the variance of $\mathbf{x}(t)$.

For arbitrary A we can give an algebraic proof as follows.

By the unmixing assumption the Lyapunov equation $P - APA^\top = BB^\top$ has a unique symmetric solution P. Therefore solving equation $(**)$ with this fixed P yields

$$\begin{aligned} \tilde{C}^\top &= APC^\top + BD^\top \\ \tilde{C} &= CPA^\top + DB^\top \\ R &= CPC^\top + DD^\top \end{aligned}$$

whereby $\Phi(z) = C(zI - A)^{-1}\tilde{C}^\top + R + \tilde{C}(z^{-1}I - A^\top)^{-1}C^\top := \tilde{Z}(z) + R + \tilde{Z}(z^{-1})^\top$. Now $\Phi(z)$ has a spectral factor and hence is a bona fide spectral density matrix which admits a unique Laurent expansion. Now in a suitably small neighborhood of the unit circle $\tilde{Z}(z)$ has an expansion in powers of z^{-1} without constant term while $\tilde{Z}(z^{-1})^\top$ has an expansion in positive powers of z also without constant term. It follows that $R = \Lambda_0$ and there is a unique $P = P^\top$ satisfying the LMI corresponding to $Z(z) = C\,(zI - A)^{-1}\tilde{C} + \frac{1}{2}\Lambda_0$. $\square$

The Positive Real Lemma

Corollary 2 (Positive Real Lemma) *Let $Z(z)$ be a $m \times m$ transfer function with a minimal realization $Z(z) = C(zI - A)^{-1}\bar{C}^\top + \frac{1}{2}\Lambda_0$ where A is an asymptotically stable $n \times n$ matrix. Then, the set of symmetric solutions P of the linear matrix inequality (LMI)*

$$\mathcal{P} := \{P = P^\top ; M(P) \geq 0\}$$

is nonempty if and only if $Z(z)$ is positive real. All $P \in \mathcal{P}$ are positive definite.

Positive definitness of P follows from the minimality of (C, A, B) which in turn follows from that of $(C, A, \bar{C})$ (see the Exercise).

Parametrization of all minimal analytic spectral factors

Theorem 3 (B.D.O. Anderson) *Let $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ be a minimal realization of Φ_+, the positive real part of Φ. Then there is a one-to-one correspondence between minimal analytic spectral factors of Φ and symmetric $n \times n$ matrices P solving the* Linear Matrix Inequality *LMI in the following sense: Corresponding to each solution $P = P^\top$ of the LMI, necessarily positive definite, there corresponds a minimal analytic spectral factor $W(z) = C(zI - A)^{-1}B + D$ where A and C are as above and $\begin{bmatrix} B \\ D \end{bmatrix}$ is the unique (modulo orthogonal transformations) full-rank factor of $M(P)$.*
Conversely, corresponding to each minimal analytic spectral factor having minimal realization $W(z) = C(zI - A)^{-1}B + D$ there is a unique positive definite $P \in \mathcal{P}$ solving the LMI with $\bar{C} = CPA^\top + DB^\top$.

Proof

That from the minimality of $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ follows that any spectral factor of the form $W(z) = C(zI - A)^{-1}B + D$ is a minimal spectral factor, is the content of the Exercise on slide 37. Hence to each solution $P = P^\top$ of the LMI (necessarily positive definite since $|\lambda(A)| < 1$), there corresponds a minimal analytic spectral factor.
Conversely, let $W(z) = C(zI - A)^{-1}B + D$ be a minimal realization of a minimal spectral factor, then by Lemma 4 there is a unique P with $P = P^\top > 0$ solution of the LMI constructed from $Z(z) = C(zI - A)^{-1}\bar{C}^\top + \frac{1}{2}\Lambda_0$ where $\bar{C} = CPA^\top + DB^\top$. $Z(z)$ is indeed the positive real part of the spectrum. $\square$

Note: The matrix $\bar{C} = CPA^\top + DB^\top$ must be the same for all minimal spectral factors $W(z) = C(zI - A)^{-1}B + D$ since it is the "B" parameter of a minimal realization of Φ_+; hence it does not depend on which spectral factor is chosen to form $\Phi(z)$. In other words, $\bar{C}$ is an *invariant over the family of all minimal stochastic realizations expressed with a fixed* (C, A) *pair.*
Recall that $\bar{C} = \mathbb{E}\,\mathbf{y}(t)\mathbf{x}(t+1)^\top$ so this quantity is also an invariant.

The set $\mathscr{P}$ of solutions of the LMI

Easy to see that $\mathscr{P}$ is a closed and convex set.

Need a condition, called *regularity*, that precludes the presence of zeros either at $z = 0$ or at $z = \infty$ in the spectral density matrix of the process.

$$\Delta(P) := \Lambda_0 - CPC^\top > 0 \qquad \text{for all } P \in \mathcal{P}.$$

Clearly, a regular process must be full rank.

The number of columns of the spectral factor $W(z)$ varies with $P \in \mathcal{P}$. In fact, if the rank of $\begin{bmatrix} B^\top & D^\top \end{bmatrix}^\top$ is full, $W(z)$ is $m \times p$, where $p := \operatorname{rank} M(P)$. Then, letting $T := -(\bar{C}^\top - APC^\top)\Delta(P)^{-1}$, one has a block-diagonalization of $M(P)$

$$\begin{bmatrix} I & T \\ 0 & I \end{bmatrix} M(P) \begin{bmatrix} I & 0 \\ T^\top & I \end{bmatrix} = \begin{bmatrix} R(P) & 0 \\ 0 & \Delta(P) \end{bmatrix},$$

where

$$R(P) = P - APA^\top - (\bar{C}^\top - APC^\top)\Delta(P)^{-1}(\bar{C}^\top - APC^\top)^\top.$$

Hence, $P \in \mathcal{P}$ if and only if it satisfies the **Algebraic Riccati inequality**

$$R(P) \geq 0. \qquad \text{(ARI)}$$

Moreover, $p = \operatorname{rank} M(P) = m + \operatorname{rank} R(P) \geq m$. If P satisfies the **Algebraic Riccati Equation** $R(P) = 0$, i.e.,

$$P = APA^\top + (\bar{C}^\top - APC^\top)\Delta(P)^{-1}(\bar{C}^\top - APC^\top)^\top, \qquad \text{(ARE)}$$

then $\operatorname{rank} M(P) = m$ and the spectral factor $W(z)$ is $m \times m$.
The family of P's solving the ARE; i.e. corresponding to square spectral factors, form a subfamily $\mathcal{P}_0$ of $\mathcal{P}$. If $P \notin \mathcal{P}_0$, $W(z)$ is rectangular.

The extreme points of $\mathscr{P}$

Theorem 4 *There are two* **special** *solutions of the* **ARE***: P_-, P_+, such that*
$P_- \leq P \leq P_+, \quad \text{for all } P \in \mathscr{P}\,.$

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

is the **minimum phase** *spectral factor whose zeros are all in $\{|z| < 1\}$ i.e. with a* **causal inverse**

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

the **maximum phase** *spectral factors whose zeros are all in $\{|z| > 1\}$ i.e. with an* **anticausal inverse***.*

Proof based on Kalman filter theory.

The Kalman filter

Problem: compute the minimum variance estimate of the state at time $t+1$ of the stationary linear model

$$\begin{cases} \mathbf{x}(t+1) & = & A\,\mathbf{x}(t) + B\mathbf{w}(t) \\ \mathbf{y}(t) & = & C\,\mathbf{x}(t) + D\mathbf{w}(t) \end{cases}, \qquad \mathrm{Cov}\left\{ \begin{bmatrix} B\mathbf{w}(t) \\ D\mathbf{w}(t) \end{bmatrix} \right\} = \begin{bmatrix} Q & S \\ S^\top & R \end{bmatrix} \delta(t-s)\,.$$

Assume $R > 0$ but no conditions on A for now. Given past measurements of $\{\mathbf{y}(t)\}$ up to time $t \geq t_0$ with $\mathbf{x}(t_0) = \mathbf{x}_0 \perp \mathbf{w}(t)\,, t \geq t_0$ and given $\mathbb{E}\,\mathbf{x}_0 = 0\,, \mathrm{Var}\,\{\mathbf{x}_0\} = \Sigma_0$,
want to compute $\hat{\mathbf{x}}(t+1 \mid t) := \hat{\mathbb{E}}\,[\mathbf{x}(t+1) \mid \mathbf{y}(s),\ t_0 \leq s \leq t\,]$.
This is given by the Kalman one-step ahead predictor:

$$\hat{\mathbf{x}}(t+1 \mid t) = A\hat{\mathbf{x}}(t \mid t-1) + K(t)\mathbf{e}(t)\,, \qquad t \geq t_0$$

where the **Innovation process** $\mathbf{e}(t) := \mathbf{y}(t) - C\hat{\mathbf{x}}(t \mid t-1)$, is **white noise !**.

The Kalman filter cont.d

The Kalman gain $K(t)$ and innovation Variance

$$K(t) := \left[A\tilde{P}(t)\,C^\top + S\right]\Lambda(t)^{-1} \qquad \Lambda(t) = \mathbb{E}\,\mathbf{e}(t)\mathbf{e}(t)^\top = C\tilde{P}(t)\,C^\top + R$$

where, let $\tilde{\mathbf{x}}(t \mid t-1) := \mathbf{x}(t) - \hat{\mathbf{x}}(t \mid t-1)\}$, the **error covariance matrix**, $\tilde{P}(t)$ normally denoted $P(t \mid t-1)$, is:

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

where $\Sigma(t) = \mathbb{E}\,\mathbf{x}(t)\mathbf{x}(t)^\top$, $\hat{P}(t) = \mathbb{E}\,\hat{\mathbf{x}}(t \mid t-1)\hat{\mathbf{x}}(t \mid t-1)^\top$ are the covariance matrices of the state and state estimate.

The error covariance matrix $\tilde{P}(t)$, satisfies the **Riccati Difference Equation**

$$\tilde{P}(t+1) = A\tilde{P}(t)A^\top - K(t)\Lambda(t)K(t)^\top + Q\,, \qquad \tilde{P}(t_0) = \Sigma_0\,.$$

The Riccati equation of Kalman filtering

Substitute $\tilde{P}(t) = \Sigma(t) - \hat{P}(t)$ in the Kalman gain:

$$K(t) = \left[A\Sigma(t)C^\top + S - A\hat{P}(t)C^\top\right]\Lambda(t)^{-1} = \left[\bar{C}^\top(t) - A\hat{P}(t)C^\top\right]\Lambda(t)^{-1}$$

where (recall $S = BD^\top$) define $\bar{C}^\top(t) := A\Sigma(t)C^\top + BD^\top$.
Now use the Lyapunov difference equation for the state variance $\Sigma(t)$

$$\Sigma(t+1) = A\Sigma(t)A^\top + Q$$

(N.B.: this holds if $\mathbf{x}(t) \perp \mathbf{w}(t)$ although A may be arbitrary) and get an equivalent Riccati equation in terms of $\hat{P}(t)$, the covariance of $\hat{\mathbf{x}}(t \mid t-1)$

$$\hat{P}(t+1) = A\hat{P}(t)A^\top + (\bar{C}^\top(t) - A\hat{P}(t)C^\top)(\Lambda_0(t) - C\hat{P}(t)C^\top)^{-1}(\bar{C}(t) - C\hat{P}(t)A^\top), \quad \hat{P}(t_0) = 0$$

Here $\Lambda_0(t) = C\Sigma(t)C^\top + R = \mathbb{E}\,\mathbf{y}(t)\mathbf{y}(t)^\top$ is the transient output covariance.

Asymptotics of the Riccati equation under stationarity

Theorem 5 *Let* $\mathbf{x}(t)$ *be an asymptotically stationary state process. Then for* $t - t_0 \to \infty$*, the predictor* $\hat{\mathbf{x}}(t \mid t-1)$ *converges to a stationary process and its covariance matrix converges to a constant* $\hat{P} = \hat{P}^\top \geq 0$*. If in addition,* $|\lambda(A)| < 1$*, so that* $\bar{\Sigma} := \lim_{t-t_0 \to \infty} \Sigma(t)$ *satisfies the algebraic Lyapunov equation, then* $\hat{P}$ *satisfies the ARE of Spectral factorization :*

$$\hat{P} = A\hat{P}A^\top + (\bar{C}^\top - A\hat{P}C^\top)(\Lambda_0 - C\hat{P}C^\top)^{-1}(\bar{C} - C\hat{P}A^\top)\,.$$

where $\bar{C}^\top := A\bar{\Sigma}C^\top + BD^\top$*. Under these conditions* $P \geq \hat{P}$ *for all solutions* P *of the LMI corresponding to the parameters* $(C, A, \bar{C}, \Lambda_0)$*.*

Proof: Proof that $\hat{P}(t)$ converges is in [Dispense] hence the limit error covariance $\bar{\Sigma} - \hat{P}$ exists and is positive semidefinite (error covariance solution of the ARE of Kalman filtering). It follows that $\bar{\Sigma} \geq \hat{P}$. *This is true for all state space models having an asymptotically stable* A *matrix*.
Take all models corresponding to solutions P of the LMI (Theorem 3); hence $P \geq \hat{P}$ for all solutions P of the LMI, since these P's are state variances of a (minimal) model. $\square$

Invariance properties

Fix a minimal realization $C(zI-A)^{-1}\bar{C}^\top + \Lambda_0/2$ of the positive real part of a spectral density $\Phi(z)$. Then according to Theorem 3 the whole family of minimal spectral factors of $\Phi(z)$ is parametrized $1:1$ by a solution of the LMI. Let

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

be a minimal stochastic realization of $\mathbf{y}$ in this family. Then

Proposition 4 *All these models have the same steady state Kalman filter, in fact, the state of the Kalman filter is the predictor, $\hat{\mathbf{x}}(t \mid t-1)$, of the state of any minimal state space model of $\mathbf{y}$ and is itself the model with minimal state covariance matrix $\hat{P} \equiv P_-$.*

In fact the Kalman gain $\bar{K}$ and the innovation variance $\bar{\Lambda}$ are independent of the parameters (B,D) of the model and the predictor satisfies the same-steady state Kalman filter equation irrespective of which model one starts from.

The steady state Kalman filter with asymptotically stable A

The steady-state Kalman filter corresponding to $\bar{K} := \lim_{t-t_0\to\infty} K(t)$; i.e.

$$\bar{K} = (\bar{C}^\top - A\hat{P}C^\top)(\Lambda_0 - C\hat{P}C^\top)^{-1}$$

where $\hat{P}$ is the limit of $\hat{P}(t)$, is a state space representation of $\mathbf{y}$:

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

where $\mathbf{e}$ is the *stationary innovation process* of $\mathbf{y}$ of covariance $\bar{\Lambda} = (\Lambda_0 - C\hat{P}C^\top)$. Hence the square transfer function $\bar{W}(z) = C(zI-A)^{-1}\bar{K} + I$ must be the **minimum phase (unnormalized) spectral factor of** $\Phi(z)$. Note that $\bar{W}(z)$ may have zeros on the unit circle so the minimum phase condition is not strict in general. For this to be true we would need strict positivity of the spectral density on the unit circle. Will discuss this later.

The incremental LMI

Assume A is unmixing and we have fixed a spectral factor say $W_0(z) = C(zI - A)^{-1}B_0 + D_0$ of $\Phi(z)$ corresponding to the solution P_0 of the LMI (Lemma 4). We consider the sum

$$\begin{aligned}\tilde{M}(X) &:= \begin{bmatrix} B_0B_0^\top & B_0D_0^\top \\ D_0B_0^\top & D_0D_0^\top \end{bmatrix} - \begin{bmatrix} X - AXA^\top & -AXC^\top \\ -CXA^\top & -CXC^\top \end{bmatrix} \\ &= \begin{bmatrix} -X + AXA^\top + Q_0 & AXC^\top + S_0 \\ CXA^\top + S_0^\top & CXC^\top + R_0 \end{bmatrix}\end{aligned}$$

with an obvious meaning of the symbols. In virtue of Lemma 2 we have

$$\Phi(z) = \begin{bmatrix} C(zI - A)^{-1} & I \end{bmatrix} \tilde{M}(X) \begin{bmatrix} (z^{-1}I - A^\top)^{-1}C^\top \\ I \end{bmatrix}$$

and all spectral factors of $\Phi(z)$ can be derived from solutions $X = X^\top$ of the LMI $\tilde{M}(X) \geq 0$ by the usual factorization trick. Now, by uniqueness of the parametrization of spectral factors (Lemma 4) to each such X there must correspond a unique P solving the usual $M(P) \geq 0$ and conversely. This happens just when $X = P_0 - P$ since in this case, setting $\bar{C}^\top = AP_0C^\top + S_0$, one LMI turns into the other. Now, still assuming that $CXC^\top + R_0 = \Lambda_0 - CPC^\top$ is non singular, we can derive the Riccati inequality equivalent to the LMI, by block-diagonalization:

$$-X + AXA^\top + Q_0 - (AXC^\top + S_0)(CXC^\top + R_0)^{-1}(CXA^\top + S_0^\top) \geq 0$$

This leads in particular to the ARE

$$X = AXA^\top - (AXC^\top + S_0)(CXC^\top + R_0)^{-1}(CXA^\top + S_0^\top) + Q_0 \qquad (AREKF)$$

which is the Riccati equation of Kalman Filtering. Note that all of this holds just provided A is unmixing (so that the Lyapunov equation $P_0 = AP_0A^\top + Q_0$ has a solution). When does this ARE have a unique positive semidefinite solution is a classical question in system theory.

The steady state Kalman filter with general A

Theorem 6 *The (AREKF) has a positive semidefinite solution $\bar{X} = \bar{X}^\top$ if and only if (A, C), is detectable. For any such such solution let*

$$\bar{K} := K(\bar{X}) := (A\bar{X}C^\top + S_0)(C\bar{X}C^\top + R_0)^{-1} \qquad \bar{\Gamma} := A - \bar{K}C,$$

then the eigenvalues of the matrix $\bar{\Gamma}$ are strictly inside the unit circle; i.e. $\bar{X}$ is a **stabilizing solution**, *if and only if the pair $(F, \tilde{Q}_0)$, defined as*

$$F := A - S_0 R_0^{-1} C, \qquad \tilde{Q}_0 := Q_0 - S_0 R_0^{-1} S_0^\top$$

is **stabilizable**. *In this case (and only in this case) $\bar{X}$ is the* **unique positive semidefinite solution**. *In addition $\bar{X} > 0$ if and only if $(F, \tilde{Q}_0)$ is a reachable pair.*

Under the above conditions $\bar{X}$ is the (unique) limit of the error covariance, $\lim_{t-t_0 \to \infty} \tilde{P}(t)$ irrespective of the initial condition Σ_0, and hence the steady-state Kalman filter is still a bona-fide state space representation of $\mathbf{y}$:

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

where $\mathbf{e}$ is the *stationary innovation process* of $\mathbf{y}$.

Stabilizability and Spectral density

The stabilizability of $(F, \tilde{Q}_0)$ seems to be a condition depending on the particular spectral factor $W_0(z) = C(zI - A)^{-1}B_0 + D_0$ of $\Phi(z)$. In fact this is not the case.

Proposition 5 *A rational spectral density $\Phi(z)$ does not have zeros on the unit circle; i.e*

$$\Phi(e^{j\theta}) > 0, \qquad \forall\, \theta \in [-\pi, \pi]$$

if and only if it admits a spectral factor $W(z) = C(zI - A)^{-1}B + D$ with (C, A) detectable and $(F, \tilde{Q})$ stabilizable where $F := A - SR^{-1}C$ and $\tilde{Q} := Q - SR^{-1}S^{\top}$. In particular, any minimal spectral factor of a rational spectral density which has no zeros on the unit circle, must have a stabilizable $(F, \tilde{Q})$ pair.

Spectral zeros on the unit circle; example

Wonder if the condition on $(F, \tilde{Q})$ has anything to do with the triplet (A, B, C). The answer is NO.

Let $\mathbf{w}(t)$ be scalar normalized white noise and consider the model

$$\mathbf{x}(t+1) = 1/2\,\mathbf{x}(t) - 1/2\mathbf{w}(t)$$
$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{w}(t)$$

Here $A = 1/2$, $S = -1/2$, $R = 1$, $Q = 1/4$, $C = 1$ so that

$$F = A - SR^{-1}C = 1/2 + 1/2 = 1\,, \qquad \tilde{Q} = 1/4 - 1/4 = 0$$

hence $(F, \tilde{Q})$ is **not stabilizable** although the triplet (A, B, C) is trivially minimal. In fact the transfer function of this model is $W(z) = \dfrac{z-1}{z-1/2}$ and is a spectral factor of a spectrum $\Phi(z)$ which has has a zero at $z = 1$.

What kind of a model is the s.s. Kalman filter ?

Since A need not be asymptotically stable in general $\Sigma(t)$ does not satisfy the Lyapunov equation and there is in general no limit of $\Sigma(t)$ neither of $\hat{P}(t)$. Nevertheless if $\Phi(z)$ does not have zeros on the unit circle the closed loop matrix $A - \bar{K}C$ of the **Inverse system** (whitening filter)

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

e is asymptotically stable; i.e. $A - \bar{K}C$ has eigenvalues (the zeros !) inside the unit circle. So the s.s. Kalman filter is still a minimum phase model !!.

Problem :

Is there still a minimal solution of the LMI (of spectral factorization) even for unstable A?

Answer : Yes provided (A, C), is detectable.

The structure of rational all-pass functions

Another application of Lemma 4 is to rational spectral factors of the spectral density $\Phi(z) \equiv I$; i.e. to rational matrix functions $Q(z) = C(zI - A)^{-1}B + D$ such that $Q(z)Q(z^{-1})^{\top} = I$. These are called **all-pass** functions.
Assume (A, C) is a fixed observable pair so that the pole structure of $Q(z)$ is fixed. Then $\bar{C} = 0$. We shall look for the "square" solutions of the spectral factorization LMI

$$\begin{bmatrix} P - APA^{\top} & -APC^{\top} \\ -CPA^{\top} & I - CPC^{\top} \end{bmatrix} \geq 0$$

assuming $I - CPC^{\top} := DD^{\top}$ non singular. This permits to solve for B to get $B = -APC^{\top}D^{-\top}$ and leads to the homogeneous Riccati equation

$$P - A\left[P + PC^{\top}(I - CPC^{\top})^{-1}CP\right]A^{\top} = 0 \qquad (HARE)$$

which has the trivial solution $P = 0$, corresponding to $Q(z) = D$, a constant unitary matrix. The other solutions parametrize the non-trivial square all-pass functions with the given denominator.

Rational all-pass functions (cont.d)

We shall now assume that A is invertible. Consider the zero-dynamics matrix

$$\Gamma := A - BD^{-1}C = A + APC^{\top}(I - CPC^{\top})^{-1}C$$

using the Riccati equation one derives the invariance relation $\Gamma P = PA^{-\top}$ so that, if P is an invertible solution

$$P^{-1}\Gamma P = A^{-\top}.$$

So far we don't know if there are any invertible solutions of the Riccati equation. By the matrix inversion lemma, they must satisfy

$$P - A\left[P^{-1} - C^{\top}C\right]^{-1}A^{\top} = 0$$

which, since A is invertible turns into $\quad P^{-1} = A^{-\top}P^{-1}A^{-1} - A^{-\top}C^{\top}CA^{-1}$ that is into the Lyapunov equation

$$P^{-1} = A^{\top}P^{-1}A + C^{\top}C, \qquad (\dagger\dagger)$$

which by observability and unmixing has a unique nonsingular solution (Theorem 2). In case A is asymptotically stable P is positive definite.

Rational all-pass functions (cont.d)

All other solutions of the homogeneous Riccati equation must be singular.

Theorem 7 *Let A be unmixing and nonsingular and (C, A) be observable. There is a 1:1 correspondence between square all pass rational matrix functions of the form $Q(z) = C(zI - A)^{-1}B + D$, defined modulo multiplication from the right by an arbitrary constant unitary matrix, and solutions $P = P^\top$ of the HARE.*
Consider the orthogonal direct sum decomposition

$$\mathbb{R}^n = \operatorname{Im} P \oplus \operatorname{Ker} P\,, \qquad (\perp)$$

then $\operatorname{Im} P$ is an invariant subspace for Γ and $\operatorname{Ker} P$ is a left-invariant subspace for A which is orthogonal to the reachable subspace of (A, B). The McMillan degree of $Q(z)$ is then equal to $\dim\{\operatorname{Im} P\}$. In a basis adapted to the direct sum decomposition $(\perp)$, $P = \operatorname{diag}\{\hat{P}_1,\, 0\}$ and the restrictions $\hat{A}_{11}^{-\top}$, $\hat{\Gamma}_{11}$ of $A^{-\top}$ and of Γ to $\operatorname{Im} P$ are similar; i.e.

$$\hat{P}_1^{-1}\hat{\Gamma}_{11}\hat{P}_1 = \hat{A}_{11}^{-\top}\,.$$

Proof

The first statement is just a particularization of Theorem **??**.
The orthogonal direct sum decomposition $(\perp)$ holds since P is symmetric. From the invariance relation $\Gamma P = PA^{-\top}$, it follows that for any $v \in \mathbb{R}^n$, $\Gamma P v \in \operatorname{Im} P$ and hence $\operatorname{Im} P$ is invariant for Γ. Next, for any $x \in \operatorname{Ker} P$ we have $PA^{-\top}x = \Gamma P x = 0$ and hence $\operatorname{Ker} P$ is an invariant subspace for $A^{-\top}$. In fact $\operatorname{Ker} P$ is orthogonal to the reachable subspace for (A, B) as $x^\top B = -x^\top APC^\top D^{-\top} = 0$ since $\operatorname{Ker} P$ is also an invariant subspace for $A^\top$ and hence $x^\top A$ belongs to the left nullspace of P.
Since P is symmetric there is an orthogonal basis of eigenvectors in which $P = \operatorname{diag}\{\hat{P}_1, 0\}$ with $\hat{P}_1$ non singular and the invariance relation can be written

$$\begin{bmatrix} \hat{P}_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{A}_{11}^{-\top} & \hat{A}_{12}^{-\top} \\ \hat{A}_{21}^{-\top} & \hat{A}_{22}^{-\top} \end{bmatrix} = \begin{bmatrix} \hat{\Gamma}_{11} & \hat{\Gamma}_{12} \\ \hat{\Gamma}_{21} & \hat{\Gamma}_{22} \end{bmatrix} \begin{bmatrix} \hat{P}_1 & 0 \\ 0 & 0 \end{bmatrix}$$

from which the similarity of $\hat{A}_{11}^{-\top}$ to $\hat{\Gamma}_{11}$ follows. In this basis $Q(z)$ has a realization $(\hat{C}_1, \hat{A}_{11}, \hat{B}_1, D)$ of dimension equal to $\dim\{\operatorname{Im} P\}$. Since $\hat{P}_1$ is non singular and satisfies the Lyapunov equation $\hat{P}_1 = \hat{A}_{11}\hat{P}_1\hat{A}_{11}^\top + \hat{B}_1\hat{B}_1^\top$ this realization must be reachable. $\square$

Comments

Some of the above can be obtained by a simple similarity argument based on the fact that $Q(z)^{-1} = Q(z^{-1})^\top$. Assume A is invertible and recall the formula

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

letting $Q(z) = J + H(zI - F)^{-1}G$, a minimal realization with F invertible, one has

$$Q(z)^{-1} = J^{-1} - J^{-1}H(zI - \Gamma)^{-1}GJ^{-1}, \qquad \Gamma = F - GJ^{-1}H$$
$$Q(z^{-1})^\top = J^\top - G^\top F^{-\top}H^\top - G^\top F^{-\top}(zI - F^{-\top})^{-1}F^{-\top}H^\top$$

and hence there must exist an invertible T such that

$$T^{-1}(F - GJ^{-1}H)T = F^{-\top}, \quad J^{-1}HT = G^\top F^{-\top}, \quad T^{-1}GJ^{-1} = F^{-\top}H^\top$$

These equations determine a unique T which must solve either of the two dual Lyapunov equations

$$T = FTF^\top - GG^\top, \qquad T^{-1} = F^\top T^{-1}F - H^\top H$$

Co-analytic spectral factorization

Let $\Phi_+(z) = C(zI - A)^{-1}\bar{C}^\top + \frac{1}{2}\Lambda_0$ be the positive real part of a spectral density. We want to describe the family of all minimal Anticausal (co-analytic) spectral factors:

$$\Phi(z) = \bar{W}(z)\bar{W}(z^{-1})^\top, \qquad \bar{W}(z) \text{ analytic inside the unit circle}.$$

Consider the family of all analytic spectral factors of the transpose,

$$\Phi^\top(z) = \tilde{W}(z)\tilde{W}(z^{-1})^\top, \qquad \text{so that} \qquad \Phi(z) = \tilde{W}(z^{-1})\tilde{W}(z)^\top$$

and hence there is a 1:1 correspondence between coanalytic spectral factors of $\Phi(z)$ and analytic factors of $\Phi^\top(z)$ given by the formula

$$\bar{W}(z) = \tilde{W}(z^{-1}).$$

Since we have the decomposition

$$\Phi^\top(z) = \bar{C}(zI - A^\top)^{-1}C^\top + \Lambda_0 + C(z^{-1}I - A)^{-1}\bar{C}^\top$$

we can set up a “coanalytic” version of Theorem 3.

The anticausal (backwards) LMI

Form the linear matrix inequality of $\Phi^\top(z)$

$$\bar{M}(\bar{P}) := \begin{bmatrix} \bar{P} - A^\top \bar{P} A & C^\top - A^\top \bar{P} \bar{C}^\top \\ C - \bar{C} \bar{P} A & \Lambda_0 - \bar{C} \bar{P} \bar{C}^\top \end{bmatrix} \geq 0. \tag{7}$$

Then $\bar{B}$ and $\bar{D}$ determined via full rank factorization

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

yield all coanalytic spectral factors

$$\bar{W}(z) = \bar{C}(z^{-1} I - A^\top)^{-1} \bar{B} + \bar{D}.$$

Under the same regularity assumption the dual LMI is equivalent to the dual (backward) Algebraic Riccati Inequality

$$\bar{P} - A^\top \bar{P} A - (C^\top - A^\top \bar{P} \bar{C}^\top) \Delta(\bar{P})^{-1} (C - \bar{C} \bar{P} A) \geq 0$$

where $\Delta(\bar{P}) = \Lambda_0 - \bar{C} \bar{P} \bar{C}^\top$.

Parametrization of all minimal co-analytic spectral factors

Theorem 8 *Let* $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ *be a minimal realization of* Φ_+*, the positive real part of* $\Phi(z)$*. Then there is a one-to-one correspondence between minimal coanalytic spectral factors of* $\Phi(z)$ *and symmetric* $n \times n$ *matrices* $\bar{P}$ *solving the* dual Linear Matrix Inequality *LMI*

$$\bar{M}(\bar{P}) := \begin{bmatrix} \bar{P} - A^\top \bar{P} A & C^\top - A^\top \bar{P} \bar{C}^\top \\ C - \bar{C}\bar{P}A & \Lambda_0 - \bar{C}\bar{P}\bar{C}^\top \end{bmatrix} \geq 0$$

Corresponding to each solution $\bar{P} = \bar{P}^\top$ *of the LMI, necessarily positive definite, there corresponds a minimal coanalytic spectral factor* $\bar{W}(z) = \bar{C}(z^{-1}I - A^\top)^{-1}\bar{B} + \bar{D}$ *where* $\begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}$ *is the unique (modulo orthogonal transformations) full-rank left factor of* $\bar{M}(\bar{P})$*.*
Conversely, corresponding to each minimal coanalytic spectral factor having minimal realization $\bar{W}(z) = \bar{C}(z^{-1}I - A^\top)^{-1}\bar{B} + \bar{D}$ *there is a unique positive definite* $\bar{P}$ *solving the LMI with* $C := \bar{C}\bar{P}A + \bar{D}\bar{B}^\top$*.*

Theorem 9 *Assume* $\mathbf{y}$ *has a forward representation* (5)*. Then it can also be represented by the backward system*

$$\begin{cases} \bar{\mathbf{x}}(t-1) = A^\top \bar{\mathbf{x}}(t) + \bar{B}\bar{\mathbf{w}}(t) \\ \mathbf{y}(t) \; = \bar{C}\bar{\mathbf{x}}(t) + \bar{D}\bar{\mathbf{w}}(t), \end{cases} \tag{8}$$

with state covariance

$$\bar{P} := \mathbb{E}\{\bar{\mathbf{x}}(t)\bar{\mathbf{x}}(t)^\top\} = P^{-1}.$$

Where $\bar{C} = CPA^\top + BD^\top$ *and* $\bar{B}$, $\bar{D}$ *are defined, uniquely modulo an orthogonal transformation, via a minimum-rank factorization*

$$\begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix} \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}^\top = \begin{bmatrix} \bar{P} - A^\top \bar{P} A & C^\top - A^\top \bar{P} \bar{C}^\top \\ C - \bar{C}\bar{P}A & \Lambda_0 - \bar{C}\bar{P}\bar{C}^\top \end{bmatrix},$$

and $\bar{\mathbf{w}}$ *is a normalized white noise with the property that*

$$\mathbf{H}_t^-(\bar{w}) \perp \left(\mathbf{H}_t^+(\bar{\mathbf{x}}) \vee \mathbf{H}_{t+1}^+(\mathbf{y})\right) \quad \textit{for all } t \in \mathbb{Z}.$$

Proof

Let $\bar{\mathbf{x}}(t) := P^{-1}x(t+1)$, $\quad \bar{\mathbf{z}}(t) := \begin{bmatrix} \bar{\mathbf{x}}(t-1) \\ \mathbf{y}(t) \end{bmatrix}$ and $\quad \bar{\mathbf{z}}(t) = \hat{\bar{\mathbf{z}}}(t \mid t+1) + \bar{\mathbf{v}}(t)$
where $\hat{\bar{\mathbf{z}}}(t \mid t+1)$ is the backward one-step predictor

$$\hat{\bar{\mathbf{z}}}(t \mid t+1) := \hat{\mathbb{E}}\left[\bar{\mathbf{z}}(t) \mid \mathbf{H}^{+}_{t+1}(\bar{\mathbf{z}})\right]$$

and $\bar{\mathbf{v}}(t) := \bar{\mathbf{z}}(t) - \hat{\bar{\mathbf{z}}}(t \mid t+1)$ is the backward innovation process, which must be a white noise, i.e., $\mathbb{E}\{\bar{\mathbf{v}}(t)\bar{\mathbf{v}}(s)'\} = \bar{V}\delta_{ts}$, with $\bar{V}$ of size $(n+m)\times(n+m)$

$$\begin{aligned} \hat{\bar{\mathbf{z}}}(t \mid t+1) &= \mathbb{E}\left[\bar{\mathbf{z}}(t) \mid \mathbf{H}^{+}_{t+1}(\bar{\mathbf{z}})\right] = \hat{\mathbb{E}}\left[\bar{\mathbf{z}}(t) \mid \mathbf{X}_{t+1}\right] \\ &= \mathbb{E}\{\bar{\mathbf{z}}(t)\mathbf{x}(t+1)^{\top}\}\mathbb{E}\{\mathbf{x}(t+1)\mathbf{x}(t+1)^{\top}\}^{-1}\mathbf{x}(t+1) \\ &= \begin{bmatrix} A^{\top} \\ CPA^{\top} + DB^{\top} \end{bmatrix} P^{-1}\mathbf{x}(t+1), \end{aligned}$$

since $\mathbb{E}\{\mathbf{x}(t)\mathbf{x}(t+1)^{\top}\} = PA^{\top}$ and $\mathbb{E}\{\mathbf{y}(t)\mathbf{x}(t+1)^{\top}\} = CPA^{\top} + DB^{\top}$. Consequently,

$$\hat{\bar{\mathbf{z}}}(t \mid t+1) = \begin{bmatrix} A^{\top} \\ \bar{C} \end{bmatrix} \bar{\mathbf{x}}(t), \tag{9}$$

Hence $\begin{bmatrix} \bar{\mathbf{x}}(t-1) \\ \mathbf{y}(t) \end{bmatrix} = \begin{bmatrix} A^\top \\ \bar{C} \end{bmatrix} \bar{\mathbf{x}}(t) + \bar{\mathbf{v}}(t)$ so that,

$$\mathbb{E}\{ \begin{bmatrix} \bar{\mathbf{x}}(t-1) \\ \mathbf{y}(t) \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}(t-1)^\top & \mathbf{y}(t) \end{bmatrix}^\top \} = \begin{bmatrix} A^\top \\ \bar{C} \end{bmatrix} \bar{P} \begin{bmatrix} A & \bar{C}^\top \end{bmatrix} + \mathbb{E}\{\bar{\mathbf{v}}(t)\bar{\mathbf{v}}(t)^\top\},$$

consequently $\bar{P}$ solves the LMI $M(\bar{P}) \geq 0$ and $\bar{V}$ is equal to

$$\mathbb{E}\{\bar{\mathbf{v}}(t)\bar{\mathbf{v}}(t)^\top\} = \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix} \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}^\top .$$

If $\begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}$ is full rank, can normalize $\bar{\mathbf{v}}(t)$ to $\bar{\mathbf{w}}(t) := \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}^{-L} \bar{\mathbf{v}}(t)$ and get the backward model in standard form.

$\bar{\mathbf{w}}(t)$ is the normalized backward innovation of $\bar{\mathbf{z}}(t)$ so that $\mathbf{H}_t^-(\bar{w}) \perp \mathbf{H}_{t+1}^+(\bar{\mathbf{z}})$.

$\square$

The anticausal (backwards) Kalman filter

Problem: compute the minimum variance estimate of the state at time $t-1$ of the linear backward model

$$\begin{cases} \bar{\mathbf{x}}(t-1) = A^\top \bar{\mathbf{x}}(t) + \bar{B}\,\bar{\mathbf{w}}(t) \\ \mathbf{y}(t) \;= \bar{C}\,\bar{\mathbf{x}}(t) + \bar{D}\,\bar{\mathbf{w}}(t) \end{cases} \quad , \qquad \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix} \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}^\top := \begin{bmatrix} \bar{Q} & \bar{S} \\ \bar{S}^\top & \bar{R} \end{bmatrix},$$

given **future** measurements of $\{\mathbf{y}(s)\,;\, s \geq t\}$ (m-dimensional) from time t on. Completely symmetric story for the steady-state estimate $\hat{\bar{\mathbf{x}}}(t-1 \mid t)$ which satisfies

$$\begin{cases} \hat{\bar{\mathbf{x}}}(t-1 \mid t) = A^\top \hat{\bar{\mathbf{x}}}(t \mid t+1) + \bar{K}\,\bar{\mathbf{e}}(t) \\ \qquad\quad \mathbf{y}(t) \;= \bar{C}\,\hat{\bar{\mathbf{x}}}(t \mid t+1) + \bar{\mathbf{e}}(t), \end{cases}$$

The backward error covariance $\bar{X} = \bar{P} - \hat{\bar{P}}$ where $\hat{\bar{P}} = \mathbb{E}\,\hat{\bar{\mathbf{x}}}(t \mid t+1)\hat{\bar{\mathbf{x}}}(t \mid t+1)^\top$, satisfies the **Backward Algebraic Riccati equation**

$$\bar{X} = A^\top \bar{X} A - (A^\top \bar{X} \bar{C}^\top + \bar{S})(\bar{R} + \bar{C}\bar{X}\bar{C}^\top)^{-1}(\bar{C}\bar{X}A + \bar{S}^\top) + \bar{Q} \qquad \overline{ARE}$$

which has a steady state error variance solution $\bar{X} = \mathbb{E}\,\tilde{\bar{\mathbf{x}}}(t)\tilde{\bar{\mathbf{x}}}(t)^\top \geq 0$ which by the minimum error variance property satisfies

$$\bar{X} = \bar{P} - \hat{\bar{P}} > 0\,, \qquad \forall\, \bar{P} = \mathbb{E}\,\bar{\mathbf{x}}(t)\bar{\mathbf{x}}(t)^\top \ \text{ solutions of the dual LMI}$$

so that $[\hat{\bar{P}}]^{-1} > \bar{P}^{-1}$ But every $\bar{P}^{-1} \equiv P$ is a solution of the (causal) LMI and hence $[\hat{\bar{P}}]^{-1}$ is the maximal solution $\equiv P_+$ of the (forward) LMI.

Backward models by coprime factorization

Let $W(z) = C(zI - A)^{-1}B + D$ be a minimal $m \times p$ *analytic* spectral factor. Intuitively it should be possible to choose a $p \times p$ all pass function $\bar{Q}(z)$ to cancel all the (stable) poles of $W(z)$ so that $\bar{W}(z) := W(z)\bar{Q}(z)$ is a coanalytic spectral factor with spectrum the reciprocal spectrum of $W(z)$.
If we write $Q(z) := \bar{Q}(z)^{-1}$ then $Q(z)$ is also all pass but with a stable denominator; i.e. is a **rational inner function** that is analytic and all-pass. The factorization $\bar{W}(z) = W(z)Q^{-1}(z)$ is a **coprime factorization** in $\mathbb{R}H^{\infty}$, the algebra of rational functions which are analytic outside of the unit disk. It can be shown that the representation of $\bar{W}(z)$ by a coprime factorization is unique.
Since $W(z)$ and $Q(z)$ must have the same (stable) poles, we can write

$$Q(z) = H(zI - A)^{-1}APH^{\top}J^{-\top} + J\,, \qquad JJ^{-\top} = I + HPH^{\top}$$

with $H \in \mathbb{R}^{p \times n}$ (assuming (H, A) observable) and P solving the homogeneous Riccati equation

$$P - A\left[P + PH^{\top}(I - HPH^{\top})^{-1}HP\right]A^{\top} = 0\,.$$

Backward models by coprime factorization (cont.d)

Let us assume A invertible and let $P = P^\top > 0$ be the (invertible) solution of the Lyapunov equation $(\dagger\dagger)$

$$P^{-1} = A^\top P^{-1} A + H^\top H$$

then the zero dynamics of $Q(z)$ is governed by the matrix $\Gamma = A + APH^\top (I - HPH^\top)^{-1} H$ which is similar to $A^{-\top}$, namely $P^{-1}\Gamma P = A^{-\top}$.

(TO BE COMPLETED)

Homework n. 1

Assume $W(z) = C(zI - A)^{-1}B + D$ is a minimal realization of a minimal degree spectral factor; i.e. (C, A, B) is a minimal triplet and $W(z)$ has no non-trivial right inner divisors. Find $Q(z)$.

Hint: check (for continuous time coprime factorizations) C.N. Nett, C.A. Jacobson and M.J. Balas *IEEE Trans A.C.* vol AC-29 September 1984, pp 831-832.

There should be explicit formulas also for discrete time but I don't know of a good precise reference.

Example

Consider a spectral density $\Phi(z)$ with the positive real part

$$\Phi_+(z) = \frac{\frac{5}{3}}{z - \frac{1}{2}} + \frac{7}{6}.$$

Then $A = \frac{1}{2}$, $\bar{C} = \frac{5}{3}$, $C = 1$, and $\Lambda_0 = \frac{7}{3}$, and the linear matrix inequality becomes

$$M(P) = \begin{bmatrix} \frac{3}{4}P & \frac{5}{3} - \frac{1}{2}P \\ \frac{5}{3} - \frac{1}{2}P & \frac{7}{3} - P \end{bmatrix} \geq 0,$$

which holds if and only if $P > 0$, $\frac{7}{3} - P > 0$ and

$$\det M(P) = -P^2 + \frac{41}{12}P - \frac{25}{9} = -(P - \frac{4}{3})(P - \frac{25}{12}) \geq 0.$$

These inequalities hold precisely for $P \in [\frac{4}{3}, \frac{25}{12}]$, and hence $\mathcal{P}$ is the interval $[P_-, P_+]$, where $P_- = \frac{4}{3}$ and $P_+ = \frac{25}{12}$.

Since

$$M(P_-) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

$P = P_-$ yields $B = 1$ and $D = 1$ and the minimal spectral factor

$$W_-(z) = \frac{1}{z - \frac{1}{2}} + 1 = \frac{z + \frac{1}{2}}{z - \frac{1}{2}},$$

which clearly is minimum phase. On the other hand,

$$M(P_+) = \begin{bmatrix} 25/16 & 5/8 \\ 5/8 & 1/4 \end{bmatrix}$$

yielding $B = \frac{5}{4}$ and $D = \frac{1}{2}$ and the maximum phase spectral factor

$$W_+(z) = \frac{\frac{5}{4}}{z - \frac{1}{2}} + \frac{1}{2} = \frac{1 + \frac{1}{2}z}{z - \frac{1}{2}}.$$

Finally, let us take a P in the interior of $\mathcal{P}$. With $P = 2 \in [\frac{4}{3}, \frac{25}{12}]$ we obtain

$$M(P) = \begin{bmatrix} 3/2 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}.$$

Without restriction we may take

$$\begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ d & 0 \end{bmatrix}$$

and then

$$\begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix}' = \begin{bmatrix} b_1^2 + b_2^2 & b_1 d \\ b_1 d & d^2 \end{bmatrix} = \begin{bmatrix} 3/2 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$$

which may be solved to yield $d = \frac{1}{\sqrt{3}}$, $b_1 = \frac{2}{\sqrt{3}}$ and, choosing one root, $b_2 = \frac{1}{\sqrt{6}}$, thus defining a rectangular spectral factor

$$W(z) = \left(\frac{\frac{2}{\sqrt{3}}}{z - \frac{1}{2}} + \frac{1}{\sqrt{3}}, \frac{\frac{1}{\sqrt{6}}}{z - \frac{1}{2}} \right).$$

In this example all minimal spectral factors, except W_- and W_+ which are scalar, are 1×2 matrix valued.

REALIZATION THEORY

1. The Stochastic Realization Problem from covariance data
   - Problem Statement
   - Review of deterministic realization theory
   - The shift-invariance algorithm
2. The Stochastic Realization Problem from sample covariance data
   - Early algorithms (Aoki) and critiques
   - Splitting subspaces and stochastic models
   - The state space construction by CCA

Motivations

We know how to compute (A, B, C, D, P) from $\Phi_+(z) = C(zI - A)^{-1}\bar{C} + \frac{1}{2}\Lambda_0$.

How to compute $(C, A, \bar{C})$? The available data are (estimates of) a finite covariance sequence

$$\{\Lambda(\tau), ; \tau = 0, 1, \ldots, \nu\} \stackrel{?}{\to} C, A, \bar{C}$$

This is a (deterministic, partial) *realization problem*.

Deterministic realization theory

Input-output map (free response form initial conditions built by past u's at time zero)

$$y(t) = \sum_{k=-\infty}^{-1} G_{t-k}\, u(k)\ , \qquad t \geq 0$$

Notations $y^+ := \begin{bmatrix} y(0) \\ y(1) \\ \vdots \end{bmatrix}$ $\quad u^- := \begin{bmatrix} u(-1) \\ u(-2) \\ \vdots \end{bmatrix}$ In matrix form $y^+ = \mathbb{H} u^-$, where $\mathbb{H}$ is an infinite Hankel matrix

$$\mathbb{H} := \begin{bmatrix} G_1 & G_2 & G_3 & \dots \\ G_2 & G_3 & G_4 & \dots \\ G_3 & G_4 & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

Problem A: When is it true that $G_k = CA^{k-1}B$ for some triplet (A, B, C) and how to compute them ?

Deterministic realization theory cont.d

necessary and sufficient condition: $\mathbb{H}$ should have finite rank n i.e. admit a rank n factorization $\mathbb{H} = \Omega\Gamma$:

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

then the Input-output map is realized by a minimal system of dimension n (where $D = G_0$ may be absent)

$$\begin{cases} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t)\,, \end{cases}$$

Sufficiency proof is constructive: Ho-Kalman Algorithm

The partial realization problem

The Ho-Kalman algorithm assumes knowledge of n and infinite data. One never has an infinite sequence $\{G_k\}$

Problem B:(the partial realization problem) Given a finite sequence $\{G_k\,; k = 1, 2, \ldots, \nu\}$ find all minimal triplets (A, B, C) such that

$$G_k = CA^{k-1}B\,; \qquad k = 1, 2, \ldots, \nu\,.$$

any such minimal triplet (A, B, C) provides a *minimal rational extension* of the finite sequence $\{G_k\,; k = 1, 2, \ldots, \nu\}$.

In general this problem has infinite solutions.

Partial realization of a finite covariance sequence

Consider a sequence of $m \times m$ (covariance) matrices

$$\Lambda := \{\Lambda_1, \ldots, \Lambda_{2k}\},$$

and the following block-Hankel structures

$$H_k := \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_k \\ \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_k & \Lambda_{k+1} & \cdots & \Lambda_{2k-1} \end{bmatrix} \tag{10a}$$

$$H_{k+1} := \begin{bmatrix} \Lambda_1 \quad \Lambda_2 \quad \cdots \quad \Lambda_k \\ \sigma H_k \end{bmatrix} \qquad \bar{H}_{k+1} := \begin{bmatrix} \begin{matrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_k \end{matrix} & \sigma H_k \end{bmatrix} \tag{10b}$$

where σH_k is the *shifted Hankel matrix*, of the same dimensions of H_k but with all entries shifted by one time unit i.e. with Λ_{i+1} replacing Λ_i everywhere.

Lemma 5 *Assume the following* equal rank condition *holds:*

$$\operatorname{rank} H_k = \operatorname{rank} H_{k+1} = \operatorname{rank} \bar{H}_{k+1} = n \tag{11}$$

where $n < mk$*, and let*

$$H_k = \Omega_k \bar{\Omega}_k^\top, \qquad \Omega_k,\ \bar{\Omega}_k \in \mathbb{R}^{mk \times n}, \tag{12}$$

be a factorization of H_k *where both factors* $\Omega_k, \bar{\Omega}_k$ *have* n *linearly independent columns. Then there are unique full rank left- and right factors,* $\Omega_{k+1}, \bar{\Omega}_{k+1}^\top$*, of* H_{k+1} *and* $\bar{H}_{k+1}$*; such that*

$$H_{k+1} = \Omega_{k+1} \bar{\Omega}_k^\top, \qquad \bar{H}_{k+1} = \Omega_k \bar{\Omega}_{k+1}^\top \tag{13}$$

and unique matrices $(C, A, \bar{C})$ *solving the shift-invariance equations:*

$$\Omega_{k+1} = \begin{bmatrix} C \\ \Omega_k A \end{bmatrix}, \qquad \bar{\Omega}_{k+1} = \begin{bmatrix} \bar{C} \\ \bar{\Omega}_k A^\top \end{bmatrix}. \tag{14}$$

Proof: By equal rank,

$$\operatorname{rowspan} H_k = \operatorname{rowspan} H_{k+1} \qquad \operatorname{columnspan} H_k = \operatorname{columnspan} \bar{H}_{k+1}$$

Hence there exist matrices $C, \Delta, \bar{C}, \bar{\Delta}$ such that

$$\begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_k \end{bmatrix} = C\bar{\Omega}_k^\top, \quad \sigma H_k = \Delta\bar{\Omega}_k^\top \qquad \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_k \end{bmatrix} = \Omega_k \bar{C}^\top, \quad \sigma H_k = \Omega_k \bar{\Delta}$$

Since the n rows of $\bar{\Omega}_k^\top$ form a basis for $\operatorname{rowspan} H_{k+1}$ and the n columns of Ω_k are a basis for $\operatorname{columnspan} \bar{H}_{k+1}$, $C, \Delta, \bar{C}, \bar{\Delta}$ are unique.

Likewise, from the last two equalities on the right and (10b), there must exist unique matrices $A, \bar{A}$ of dimension $n \times n$ such that $\Delta = \Omega_k A$ and $\bar{\Delta} = \bar{A}\bar{\Omega}_k^\top$ so that

$$\sigma H_k = \Omega_k A \bar{\Omega}_k^\top = \Omega_k \bar{A} \bar{\Omega}_k^\top .$$

That $A = \bar{A}$ follows since from this equation one gets $\Omega_k (A - \bar{A}) \bar{\Omega}_k^\top = 0$ which can only happen if $A = \bar{A}$. Hence H_{k+1} and $\bar{H}_{k+1}$ can be expressed as

$$H_{k+1} = \begin{bmatrix} C\bar{\Omega}_k^\top \\ \Omega_k A \bar{\Omega}_k^\top \end{bmatrix} = \begin{bmatrix} C \\ \Omega_k A \end{bmatrix} \bar{\Omega}_k^\top := \Omega_{k+1} \bar{\Omega}_k^\top \tag{15}$$

and

$$\begin{aligned}\bar{H}_{k+1} &= \left[\begin{array}{cc} \Omega_k \bar{C}^\top & \Omega_k A \bar{\Omega}_k^\top \end{array}\right] = \Omega_k \left[\begin{array}{cc} \bar{C}^\top & A \bar{\Omega}_k^\top \end{array}\right] \\ &:= \Omega_k \, \bar{\Omega}_{k+1}^\top \,. \end{aligned} \tag{16}$$

and since both the factors $\bar{\Omega}_k^\top$ and Ω_k are of full column rank, $(C, A, \bar{C})$ must satisfy the recursions (14).

Then we have the following important result on the uniqueness of partial realization.

Theorem 10 (Kalman) *If and only if the equal rank condition* (11) *holds, to each rank factorization* (12) *of H_k there corresponds a minimal triplet $(A, C, \bar{C}^\top)$, solution of the shift invariance equations* (14)*, such that*

$$\Lambda_i = CA^{i-1}\bar{C}^\top \qquad \text{for } i = 1, 2, \ldots, 2k$$

Such a triplet is unique in the following sense. If $(A_1, C_1, \bar{C}_1^\top)$ and $(A_2, C_2, \bar{C}_2^\top)$ are minimal partial realizations corresponding to different rank n factorizations (12)*, then there is a nonsingular $n \times n$ matrix T such that*

$$A_2 = T^{-1}A_1T, \quad C_2 = C_1T, \quad \bar{C}_2^\top = T^{-1}\bar{C}_1^\top.$$

Hence all minimal partial realizations corresponding to different rank n factorizations (12) are related by similarity.

The Singular Value Decomposition (SVD)

Theorem 11 *Let $A \in \mathbb{R}^{m \times p}$ of rank $n \le \min(m,p)$. Can find two orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{p \times p}$ and positive numbers $\{\sigma_1 \ge, \ldots, \ge \sigma_n\}$, the* **singular values** *of A, so that*

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

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

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

Proof is based on eigenvalue-eigenvector decomposition of $AA^\top$ and $A^\top A$.

$U = [u_1, \ldots, u_m]$ = 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 $A^\top A$).

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

In particular on the singular vectors A acts like multiplication

$$A v_j = \sum_{k=1}^{n} u_k \, \sigma_k \langle v_k , v_j \rangle = \sigma_j u_j$$

Useful Features of the SVD

Range and Nullspace of A:

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

Approximation properties

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

is the best approximant of rank k of A:

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

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

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

Stochastic Models from a covariance sequence

Given a covariance sequence $\Lambda_k \;\; k = 1, 2, \ldots, \nu$

Problem (Partial) stochastic realization: From $\{\Lambda_k \;\; k = 0, 1, 2, \ldots, \nu\}$ want to compute a minimal realization $\{A, B, C, D\}$ with $|\lambda(A)| < 1$, such that

$$\Lambda_k = CA^{k-1}\bar{C}^\top, \; k = 0, 1, 2, \ldots, \nu,$$

where

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

N.B.: The model covariance sequence $\{\frac{1}{2}\Lambda_0, \; CA^{k-1}\bar{C}^\top; k = 1, 2, \ldots, \}$ must be a sequence of positive type

Problem of rational positive extension of $\Lambda_k \;\; k = 1, 2, \ldots, \nu$, of minimal degree.

Rational positive (covariance) extension is a very old problem (Caratheodory, Schur, Levinson..)

The so-called *maximum entropy solution* is a model of the AR type of very high degree ($n \geq \nu$).

Early subspace identification for time series [Aoki]

Given observed data (zero mean)

$$\{\mathbf{y}_t\,;\,t=0,1,2,\ldots,N\}$$

Algorithm:

1. Form sample covariance estimates

$$\hat{\Lambda}_k = \frac{1}{N}\sum_{t=0}^{N-k} \mathbf{y}_{t+k}\mathbf{y}_t^\top, \qquad k=0,1,\ldots,2\nu \ll N$$

2. Form the sample Hankel matrix

$$\mathbb{H}_{\hat{\Lambda}} := \begin{bmatrix} \hat{\Lambda}_1 & \hat{\Lambda}_2 & \hat{\Lambda}_3 & \ldots & \hat{\Lambda}_\nu \\ \hat{\Lambda}_2 & \hat{\Lambda}_3 & \hat{\Lambda}_4 & \ldots & \hat{\Lambda}_{\nu-1} \\ \hat{\Lambda}_3 & \hat{\Lambda}_4 & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \hat{\Lambda}_{\nu+1} & \ldots & \ldots & \ldots & \hat{\Lambda}_{2\nu} \end{bmatrix}$$

Assume ν “large enough”, ideally $(\nu > n)$.

3. Compute the SVD

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

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

4. Rank n factorization

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

5. Impose the structure (and get $C, \bar{C}$ by inspection)

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

6. Let

$$\downarrow\Omega := \begin{bmatrix} CA \\ CA^2 \\ \vdots \\ CA^{\nu} \end{bmatrix} \qquad \uparrow\Omega := \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{\nu-1} \end{bmatrix}$$

Compute A by solving the shift invariance equation

$$(\downarrow\Omega) = (\uparrow\Omega)A$$

Solve by least squares (no exact solution in general)

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

This provides estimates of $C, \bar{C}, A$.

Second step

Step 2: Assume that $\{A, C, \bar{C}, \frac{1}{2}\Lambda_0\}$ $\Rightarrow \Phi_+(z) = C\,[zI - A]^{-1}\,\bar{C}^\top + 1/2\Lambda_0$ is Positive Real.

From the half-spectrum $(\Phi_+(z))$ to get a state-space model

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

solve the LMI. Just need to compute (B, D).
Recall: $W(z) := C(zI - A)^{-1}B + D$ is a **spectral factor** $\Phi(z) = W(z)W(1/z)^\top$

Innovation model identification

To get

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

Solve the ARE and find the minimal solution $P_- = P_-^\top > 0$. This is the unique stabilizing solution.

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

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

Critiques

Main drawback: with real data the estimates $\hat{\Lambda}_k$ are in general rather poor ! The parameters $\{A, C, \bar{C}\}$ computed by the realization algorithm may not satisfy the **positivity condition** that $\Phi_+(z)$ must be the causal part of a power spectrum: i.e. the condition that

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

may not hold.

This prevents solvability of the Riccati equation.

May try to increase N and ν... but there may be no “true” finite-dim. system generating the data.

If the truncated matrix $\uparrow \Omega$ is not of full rank n and /or $\downarrow \Omega$ does not belong to its column space, then A is not uniquely determined. $A = (\uparrow \Omega)^\dagger (\downarrow \Omega)$ involves a pseudo inverse and is not unique.

Finally use of the Hankel matrix $\mathbb{H}_{\hat{\Lambda}}$ is not optimal. (Will see this later on)

Markovian splitting subspaces

The state space $\mathbf{X}$ (say at $t = 0$) of any linear stochastic system with output $\mathbf{y}$ is a Markovian splitting subspace for $\mathbf{y}$:

$$(\mathbf{H}(\mathbf{y})^- \vee \mathbf{X}^-) \perp (\mathbf{H}(\mathbf{y})^+ \vee \mathbf{X}^+) \mid \mathbf{X}. \tag{17}$$

This implies that $\{\mathbf{X}_t\}$ is a Markovian family and that the past and the future spaces of the process $\mathbf{y}$ are conditionally orthogonal to the state space $\mathbf{X}$ (at $t = 0$), i.e.,

$$\mathbf{H}(\mathbf{y})^- \perp \mathbf{H}(\mathbf{y})^+ \mid \mathbf{X}\,. \tag{18}$$

Theorem 12 *If $\mathbf{X}$ is a Markovian splitting subspace then*

$$\mathbb{E}\,[\lambda \mid \mathbf{H}(\mathbf{y})^- \vee \mathbf{X}^-\,] = \mathbb{E}\,[\lambda \mid \mathbf{X}] \qquad \textit{for all } \lambda \in \mathbf{H}(\mathbf{y})^+ \tag{19}$$

or, equivalently,

$$\mathbb{E}\,[\lambda \mid \mathbf{H}(\mathbf{y})^+ \vee \mathbf{X}^+] = \mathbb{E}\,[\lambda \mid \mathbf{X}] \qquad \textit{for all } \lambda \in \mathbf{H}(\mathbf{y})^-. \tag{20}$$

Consequently, $\mathbf{X}$ serves as a “memory” or “sufficient statistics” which contains everything from the past which is needed in predicting the future and everything from the future which is needed in predicting the past.

Theorem 13 *Any choice of basis in a Markovian splitting subspace provides the state vector of a linear stochastic system with output* $\mathbf{y}$*.*

Example 2 *The predictor spaces*

$$\mathbf{X}^{+/-} := \mathbb{E}\left[\mathbf{H}(\mathbf{y})^{+} \mid \mathbf{H}(\mathbf{y})^{-}\right] \quad \textit{and} \quad \mathbf{X}^{-/+} := \mathbb{E}\left[\mathbf{H}(\mathbf{y})^{-} \mid \mathbf{H}(\mathbf{y})^{+}\right]$$

are minimal splitting subspaces.

The predictor spaces

$$\mathbf{X}^{+/-} := \mathrm{span}\,\{\,\mathbb{E}\,[\mathbf{y}(t)\mid \mathbf{H}(\mathbf{y})^{-}]\,;\,t=1,2,3\ldots,\,\}$$

is spanned by the forward predictors of the output given the past (at time zero). Take any basis vector say $\mathbf{x}_{-}(0)$ in $\mathbf{X}^{+/-}$.
Then $\mathbf{x}_{-}(t)$ is the state of an innovation representation (i.e. a steady state Kalman filter) of $\mathbf{y}$.
Dually take any basis vector say $\mathbf{x}_{+}(0)$ in $\mathbf{X}^{-/+}$. Then $\mathbf{x}_{+}(t)$ is the state of a backward innovation representation (i.e. a backward steady state Kalman filter) of $\mathbf{y}$.

Special bases in $\mathbf{X}^{+/-}$ and $\mathbf{X}^{-/+}$ can be constructed by **Canonical Correlation Analysis** of the past and future spaces of $\mathbf{y}$.

Canonical correlation analysis

Given two finite-dimensional subspaces $\mathbf{H}_1, \mathbf{H}_2$ of second order zero-mean random variables of dimension n and m, one wants to find two orthonormal bases, $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $\mathbf{H}_1$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ for $\mathbf{H}_2$, such that

$$\mathbb{E}\{\mathbf{u}_k \mathbf{v}_j\} = \sigma_k \delta_{kj}, \quad k = 1, \dots, n, \; j = 1, \dots, m.$$

This is the same as requiring that the covariance matrix of the two random vectors $\mathbf{u} := (\mathbf{u}_1, \dots, \mathbf{u}_n)^\top$ and $\mathbf{v} := (\mathbf{v}_1, \dots, \mathbf{v}_m)^\top$ be diagonal; i.e.,

$$\mathbb{E}\{\mathbf{u}\mathbf{v}^\top\} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & \dots & 0 \\ \vdots & & \ddots & & \dots & \vdots \\ 0 & & & \sigma_r & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \end{bmatrix}, \qquad r \le \min(n, m)$$

where we want $\sigma_1, \sigma_2, \dots, \sigma_r$ nonnegative and ordered in decreasing magnitude. Note that all $\sigma_k \le 1$ since the random variables $\mathbf{u}_k, \mathbf{v}_j$ have unit variance (norm). In fact say, $\sigma_1 = 1$ if and only if $\mathbf{u}_1$ and $\mathbf{v}_1$ are parallel and hence coincide.

The variables $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ and $(\mathbf{v}_1, \dots, \mathbf{v}_m)$ are called **canonical variables**.

Canonical angles

Since $\sigma_k \leq 1\,,;\, k = 1, \ldots, r$ we can define canonical (or principal) angles, θ_k, between the subspaces $\mathbf{H}_1$ and $\mathbf{H}_2$ by setting

$$\cos\theta_k := \sigma_k\,,; \qquad k = 1, \ldots, r$$

We have $\sigma_1 < 1$ if and and only if the (first canonical) angle between $\mathbf{H}_1$ and $\mathbf{H}_2$ is positive, which is equivalent to $\mathbf{H}_1 \cap \mathbf{H}_2 = \{0\}$

Construction of the canonical variables

Choose $\mathbf{u}$ and $\mathbf{v}$ as orthonormal bases in $\mathbf{H}_1$ and $\mathbf{H}_2$. By definition

$$\mathbb{E}\left[\mathbf{u}_k \mid \mathbf{H}_2\right] = \mathbb{E}\left[\mathbf{u}_k \mid \mathbf{v}\right] = \sigma_k \mathbf{v}_k, \quad k = 1, 2, \ldots \min(n, m),$$

This is like doing SVD to the Orthogonal Projection operator from $\mathbf{H}_1$ onto $\mathbf{H}_2$

$$\hat{\mathbb{E}}\left[\cdot \mid \mathbf{H}_2\right] : \mathbf{H}_1 \to \mathbf{H}_2$$

Want a matrix representation of this operator. Choose arbitrary bases $\mathbf{x}$, $\mathbf{y}$ in $\mathbf{H}_1$ and $\mathbf{H}_2$. Then for arbitrary $\boldsymbol{\xi} = a^\top \mathbf{x} \in \mathbf{H}_1$,

$$\hat{\mathbb{E}}\left[\boldsymbol{\xi} \mid \mathbf{H}_2\right] = a^\top \mathbb{E}\{\mathbf{x}\,\mathbf{y}^\top\}\Sigma_{\mathbf{y}}^{-1}\mathbf{y}, \qquad \Sigma_{\mathbf{y}} := \mathbb{E}\{\mathbf{y}\mathbf{y}^\top\}.$$

So in the chosen bases the representation of $\hat{\mathbb{E}}\left[\cdot \mid \mathbf{H}_2\right]$ is matrix multiplication (from the left)

$$a^\top \to a^\top \mathbb{E}\{\mathbf{x}\mathbf{y}^\top\}\Sigma_{\mathbf{y}}^{-1} = a^\top \Sigma_{\mathbf{x}\mathbf{y}}\Sigma_{\mathbf{y}}^{-1}$$

A Warning

Note that in order to express the inner product of random elements in $\mathbf{H}_1, \mathbf{H}_2$ in terms of their coordinates, we must introduce appropriate weights to form the inner products in the coordinate spaces. In fact, the inner product of two elements $\xi_i = a_i^\top \mathbf{x} \in \mathbf{H}_1$, $i = 1, 2$, induces in $\mathbb{R}^n$ the inner product

$$\langle a_1, a_2 \rangle_{\Sigma_\mathbf{x}} := a_1^\top \Sigma_\mathbf{x} a_2, \qquad \Sigma_\mathbf{x} := \mathbb{E}\{\mathbf{x}\mathbf{x}^\top\}.$$

Similarily, there is an inner product $\langle b_1, b_2 \rangle_{\Sigma_\mathbf{y}} := b_1^\top \Sigma_\mathbf{y} b_2$ corresponding to the basis $\mathbf{y}$ for $\mathbf{H}_2$.

In particular, the SVD needs to be done in the weighted inner product spaces !

To obtain the usual Euclidean inner product in $\mathbb{R}^n$, **the bases need to be orthonormal !**. Only in this case the matrix representation of the adjoint of $\hat{\mathbb{E}}[\cdot \mid \mathbf{H}_2]$ is the transpose of its matrix representation.

Computing the CCA in coordinates

Let L_x and L_y be the lower triangular Cholesky factors of the covariance matrices $\Sigma_{\mathbf{x}}$ and $\Sigma_{\mathbf{y}}$, respectively; i.e., $L_x L_x^\top = \Sigma_{\mathbf{x}}, \qquad L_y L_y^\top = \Sigma_{\mathbf{y}}$ and introduce orthonormal bases in $\mathbf{H}_1$ and $\mathbf{H}_2$:

$$\boldsymbol{\nu}_{\mathbf{x}} := L_x^{-1}\mathbf{x}, \qquad \boldsymbol{\nu}_y := L_y^{-1}\mathbf{y}$$

Then, in this orthonormal basis

$$\hat{\mathbb{E}}\,[a^\top \boldsymbol{\nu}_x \mid \mathbf{H}_2] = \hat{\mathbb{E}}\,[a^\top \boldsymbol{\nu}_x \mid \boldsymbol{\nu}_y] = a^\top H\,\boldsymbol{\nu}_y$$

where H is the $n \times m$ matrix

$$H := \mathbb{E}\{\boldsymbol{\nu}_x \boldsymbol{\nu}_y^\top\} = L_x^{-1}\mathbb{E}\{\mathbf{x}\mathbf{y}^\top\}(L_y^\top)^{-1}$$

Compute the **singular value decomposition** of H

$$H = U\Sigma V^\top, \qquad UU^\top = I_m, \quad VV^\top = I_n$$

Then the canonical variables are

$$\mathbf{u} := U^\top \boldsymbol{\nu}_x, \qquad \mathbf{v} := V^\top \boldsymbol{\nu}_y.$$

and the canonical correlation coefficients of the subspaces $\mathbf{H}_1$ and $\mathbf{H}_2$ are the (nonzero) singular values of H.

CCA of infinite past and future subspaces

First represent past and future $\mathbf{H}(\mathbf{y})^-$ and $\mathbf{H}(\mathbf{y})^+$ as spanned by infinite vectors

$$\mathbf{y}_- = \begin{bmatrix} \mathbf{y}(-1) \\ \mathbf{y}(-2) \\ \mathbf{y}(-3) \\ \vdots \end{bmatrix}, \qquad \mathbf{y}_+ = \begin{bmatrix} \mathbf{y}(0) \\ \mathbf{y}(1) \\ \mathbf{y}(2) \\ \vdots \end{bmatrix}, \qquad H_\infty := \mathbb{E}\{\mathbf{y}_+\mathbf{y}_-^\top\} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 & \dots \\ \Lambda_2 & \Lambda_3 & \Lambda_4 & \dots \\ \Lambda_3 & \Lambda_4 & \Lambda_5 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

Let L_- and L_+ be the lower triangular Cholesky factors of the infinite block Toeplitz matrices

$$T_- := \mathbb{E}\{\mathbf{y}_-\mathbf{y}_-^\top\} = L_-L_-^\top \qquad T_+ := \mathbb{E}\{\mathbf{y}_+\mathbf{y}_+^\top\} = L_+L_+^\top$$

and let orthonormal bases in $\mathbf{H}(\mathbf{y})^-$ and $\mathbf{H}(\mathbf{y})^+$ be

$$\boldsymbol{\nu} := L_-^{-1}\mathbf{y}_- \qquad \bar{\boldsymbol{\nu}} := L_+^{-1}\mathbf{y}_+$$

Then, in this orthonormal bases, $\mathcal{H} := \hat{\mathbb{E}}\,[\,\cdot \mid \mathbf{H}(\mathbf{y})^-] : \mathbf{H}(\mathbf{y})^+ \to \mathbf{H}(\mathbf{y})^-$ has the matrix representation

$$\hat{H}_\infty := \mathbb{E}\{\bar{\boldsymbol{\nu}}\boldsymbol{\nu}^\top\} = L_+^{-1}H_\infty(L_-)^{-1} \neq H_\infty.$$

Note that the normalization of the block Hankel matrix H_∞ is necessary in order for the singular values to become the canonical correlation coefficients; i.e., the singular values of $\mathcal{H}$. In fact, if we were to use the unnormalized Hankel matrix representation H_∞, instead, as may seem simpler and more natural, the transpose of H_∞ would not be the matrix representation of $\mathcal{H}^*$ in the same bases, a property which is crucial in singular value decomposition above. This is because H_∞ corresponds to bases which are not orthonormal.

A Digression on

Principal Components Analysis

of Deterministic Systems

We shall discuss special bases for deterministic state space systems (A, B, C, D).

Gramians

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

$$\Gamma_\infty := \begin{bmatrix} B & AB & A^2B & \dots \end{bmatrix}, \qquad \Omega_\infty^\top := \begin{bmatrix} C^\top & C^\top A^\top & C^\top (A^\top)^2 & \dots \end{bmatrix}$$

The **Reachability and Observability Gramians**,

$$\Pi := \sum_0^{+\infty} A^k BB^\top (A^\top)^k = \Gamma_\infty \Gamma_\infty^\top \qquad \Delta := \sum_0^{+\infty} (A^\top)^k C^\top C A^k = \Omega_\infty^\top \Omega_\infty,$$

are solutions of the dual Lyapunov equations

$$\Pi = A\Pi A^\top + BB^\top, \qquad \Delta = A^\top \Delta A + C^\top C.$$

Theorem 14 *The eigenvalues of $\Delta\Pi$ are all positive and invariant under similarity.*

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

The singular values of a linear system

Consider the Hankel matrix of a deterministic linear minimal system (A, B, C) of dimension n

$$\mathbf{H}_\infty := \begin{bmatrix} CB & CAB & CA^2B & \dots \\ CAB & CA^2B & CA^3B & \dots \\ CA^2B & CA^3B & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} = \Omega_\infty \Gamma_\infty$$

Consider eigenvalues-eigenvectors of $\Delta\Pi$

$$\Delta\Pi\, v_k = \lambda_k^2\, v_k\,; \qquad k = 1, 2, \dots, n\,.$$

Since $\mathbf{H}_\infty^\top \mathbf{H}_\infty = \Gamma_\infty^\top \Omega_\infty^\top \Omega_\infty \Gamma_\infty$,

$$\mathbf{H}_\infty^\top \mathbf{H}_\infty (\Gamma_\infty^\top v_k) = \Gamma_\infty^\top \Delta\Pi\, v_k = \Gamma_\infty^\top \lambda_k^2\, v_k = \lambda_k^2\, (\Gamma_\infty^\top v_k)$$

Theorem: The squares of the n nonzero singular values of the Hankel matrix $\mathbf{H}_\infty$ are the eigenvalues of $\Delta\Pi$.

Interpretation of the Gramians

Assume we can only use finite energy controls:

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

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

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

Diagonalize:

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

Change coordinates $\mathbf{x}_c(0) := U_c^\top \mathbf{x}(0)$. Along the k-th eigenvector the energy gain is $\lambda_{c,k}$

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

may be very large: the effect of the input on certain directions in the state space nearly invisible $\Rightarrow$ bad conditioning !

Interpretation of the Gramians cont.d

Dual meaning of the observability Gramian: By stability the output $y(t) = CA^t\mathbf{x}(0)\,;t = 0,1,\dots$ is in ℓ^2, then :

$$\mathbf{y} = \Omega_\infty\,\mathbf{x}(0)\,, \qquad \|\mathbf{y}\|^2 = \mathbf{x}(0)^\top\Omega_\infty^\top\Omega_\infty\,\mathbf{x}(0)$$

Then

$$\max_{\|\mathbf{x}(0)\|\leq 1}\frac{\langle\mathbf{x}(0),\,\Delta\mathbf{x}(0)\rangle}{\|\mathbf{x}(0)\|^2} = \|\Delta\|_2 = \lambda_{\max}(\Delta)$$

Diagonalization:

$$U_o^\top\Delta U_o \Rightarrow \mathrm{diag}\{\lambda_{o,1},\dots,\lambda_{o,n}\} \qquad \lambda_{o,1}\geq,\dots,\geq\lambda_{o,n}>0$$

Change coordinates $\mathbf{x}_o(t) := U_o^\top\mathbf{x}(t)$. Energy of the outputs corresponding to (orthogonal) state eigenvectors

$$\frac{\|\mathbf{y}_{1,o}\|}{\|\mathbf{y}_{n,o}\|} = \frac{\lambda_{1,o}}{\lambda_{n,o}}$$

may be very large: the effect of some state direction nearly invisible $\Rightarrow$ bad conditioning !

Deterministic balancing

Changing bases can make things better

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

Definition 2 *A stable linear system is in* Balanced form *if both* $\hat{\Pi}$ *and* $\hat{\Delta}$ *are* **diagonal and equal**.

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

Algorithm :

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

2. Compute the SVD

$$\Delta = U \Lambda_o U^\top$$

where Λ_o is the diagonal matrix of eigenvalues of Δ

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

6. With this change of basis

$$\bar{\Delta} = T_2^{\top} \hat{\Delta} T_2 = \Lambda^{1/2} V^{\top} I V \Lambda^{1/2} = \Lambda$$

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

MATLAB command

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

Model reduction by balanced truncation

How to approximate a "Large" model (assumed as. stable + reach + obs) by a lower dimensional one

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

Bring it to balanced form: Gramians $\Pi = \Delta = \Lambda$ diagonal. 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 ($\Lambda_1 >> \Lambda_2$) to be neglected.

Balanced truncation

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

Approximate by n_1-dimensional "principal subsystem"

$$\begin{cases} \mathbf{x}_1(t+1) & = & A_{11}\mathbf{x}_1(t) + B_1\mathbf{u}(t) \\ \mathbf{y}(t) & = & C_1\mathbf{x}_1(t) + D\mathbf{u}(t) \end{cases}$$

Theorem 15 *System* (A_{11}, B_1, C_1) *is as. stable and minimal. If* $\lambda_{n_1} > \lambda_{n_1+1}$,

$$\|G(e^{j\theta}) - G_1(e^{j\theta})\|_\infty \leq 2 \sum_{k=n_1+1}^{n} \lambda_k$$

References on Deterministic Model reduction and Balanced truncation

1. K.Glover, All optimal Hankel norm approximations of linear multivariable systems and their L^∞ error bounds, *Intern. J. Control* , **39** (1984), 1115–1193.

2. B.C. Moore, Principal components analysis in linear systems, *IEEE Trans Aut. Control*, **AC-26**, 1981.

3. L. Pernebo and L. M. Silverman, Model reduction via balanced state space representations, *IEEE Trans. Automatic Control* **AC-27** (1982), 382–387.

4. K. Zhou, J. Doyle and K. Glover, *Robust and Optimal Control*, Prentice Hall, 1995, Section 21.8.

CCA of a finite dimensional process

Now, assume that $\mathbf{y}$ can be realized by a finite dimensional system of dimension n. Then the infinite block Hankel matrix

$$H_\infty := \mathbb{E}\{\mathbf{y}_+\mathbf{y}_-^\top\} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 & \dots \\ \Lambda_2 & \Lambda_3 & \Lambda_4 & \dots \\ \Lambda_3 & \Lambda_4 & \Lambda_5 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where $\Lambda_k = \mathbb{E}\{\mathbf{y}(t+k)\mathbf{y}(t)^\top\} = CA^{k-1}\bar{C}^\top$, has rank n and admits a factorization

$$H_\infty = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \begin{bmatrix} \bar{C} \\ \bar{C}A^\top \\ \bar{C}(A^\top)^2 \\ \vdots \end{bmatrix}^\top := \Omega_\infty \bar{\Omega}_\infty^\top, \qquad \bar{C} = CPA^\top + DB^\top$$

The normalized $\hat{H}_\infty$ has a singular-value decomposition $\hat{H}_\infty = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^\top$

CCA of a finite dimensional process cont.d

Since $\operatorname{rank} \hat{H}_\infty = \operatorname{rank} H_\infty$, $\hat{H}_\infty$ has exactly n nonzero singular values, the **canonical correlation coefficients**

$$\Sigma = \operatorname{diag}\{\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n\}$$

which are arranged in decreasing order $1 \geq \sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots \geq \sigma_n > 0$.
σ_k's are the cosines of the principal angles between the past $\mathbf{H}(\mathbf{y})^-$ and the future $\mathbf{H}(\mathbf{y})^+$. We have $\sigma_1 < 1$ if and and only if $\mathbf{H}(\mathbf{y})^- \cap \mathbf{H}(\mathbf{y})^+ = 0$.
Can show that this holds if and only if the spectral density of $\mathbf{y}$ is coercive.
$U \in \mathbb{R}^{\infty \times \infty}$, $V \in \mathbb{R}^{\infty \times \infty}$ have orthonormal columns

$$U^\top U = I = V^\top V \,.$$

(can be made rigorous as operators in ℓ^2). Rotate the orthonormal basis $\bar{\boldsymbol{\nu}}$ in $\mathbf{H}(\mathbf{y})^+$ and $\boldsymbol{\nu}$ in $\mathbf{H}(\mathbf{y})^-$ to recover the **canonical vectors**

$$\mathbf{u} := V^\top \boldsymbol{\nu}, \qquad \mathbf{v} := U^\top \bar{\boldsymbol{\nu}}\,, \qquad \mathbb{E}\,\mathbf{v}\,\mathbf{u}^\top = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}.$$

Canonical bases in the predictor spaces

Convenient bases in $\mathbf{X}^{+/-}$ and $\mathbf{X}^{-/+}$ are obtained by C.C.A. of the infinite past and future subspaces.

Theorem 16 *Let* $\mathbf{y}$ *be realized by a finite dimensional model.Then*

$$\mathbf{X}^{+/-} = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}, \qquad \mathbf{X}^{-/+} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}.$$

i.e. the first n *left* canonical vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ *of* $\mathcal{H}$ *constitute an orthonormal basis of the predictor space* $\mathbf{X}^{+/-}$*, while* $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ *form an orthonormal basis of* $\mathbf{X}^{-/+}$*.*

Proof :

$$\mathbf{H}^- = \hat{\mathbb{E}}[\mathbf{H}^+ \mid \mathbf{H}^-] \oplus [\mathbf{H}^- \cap (\mathbf{H}^+)^\perp] = \mathbf{X}^{+/-} \oplus \mathbf{N}^- \qquad \text{State Space plus past Junk}$$
$$\mathbf{H}^+ = \hat{\mathbb{E}}[\mathbf{H}^- \mid \mathbf{H}^+] \oplus [\mathbf{H}^+ \cap (\mathbf{H}^-)^\perp] = \mathbf{X}^{-/+} \oplus \mathbf{N}^+ \qquad \text{State Space plus future Junk}$$

Past Junk $\mathbf{N}^-$ is the subspace of past variables which are orthogonal to the future. The projection of the future onto this subspace is zero.
$\mathbf{X}^{+/-}$ is precisely the subspace of random variables in $\mathbf{H}^-$ having nonzero correlation with the future $\mathbf{H}^+$.

$\mathbf{X}^{+/-}$ is the range space of the operator $\hat{\mathbb{E}}[\cdot \mid \mathbf{H}^-]$ restricted to $\mathbf{H}^+$
$\mathbf{X}^{-/+}$ is the range space of the operator $\hat{\mathbb{E}}[\cdot \mid \mathbf{H}^+]$ restricted to $\mathbf{H}^-$

$$\mathbf{N}^- = \mathbf{H}^- \cap (\mathbf{H}^+)^\perp = \operatorname{span}\{\mathbf{u}_{n+1}, \mathbf{u}_{n+2}, \mathbf{u}_{n+3}, \ldots\} = \operatorname{Ker} \hat{\mathbb{E}}[\cdot \mid \mathbf{H}^+]$$
$$\mathbf{N}^+ = \mathbf{H}^+ \cap (\mathbf{H}^-)^\perp = \operatorname{span}\{\mathbf{v}_{n+1}, \mathbf{v}_{n+2}, \mathbf{v}_{n+3}, \ldots\} = \operatorname{Ker} \hat{\mathbb{E}}[\cdot \mid \mathbf{H}^+]$$

are the nullspaces (kernels) of the restricted projections $\hat{\mathbb{E}}[\cdot \mid \mathbf{H}^+]$ and $\hat{\mathbb{E}}[\cdot \mid \mathbf{H}^-]$.

$$\mathbf{H}(\mathbf{y}) = \mathbf{H}^- \vee \mathbf{H}^+ = \mathbf{N}^- \oplus \mathbf{H}^\square \oplus \mathbf{N}^+,$$

where $\mathbf{H}^\square := \mathbf{X}^{+/-} \vee \mathbf{X}^{-/+}$ (has dimension $2n$).

Balanced canonical bases

Balanced canonical bases

$$\mathbf{z} = \Sigma^{1/2}\mathbf{u} = \begin{bmatrix} \sigma_1^{1/2}\mathbf{u}_1 \\ \sigma_2^{1/2}\mathbf{u}_2 \\ \vdots \\ \sigma_n^{1/2}\mathbf{u}_n \end{bmatrix}, \qquad \bar{\mathbf{z}} = \Sigma^{1/2}\mathbf{v} = \begin{bmatrix} \sigma_1^{1/2}\mathbf{v}_1 \\ \sigma_2^{1/2}\mathbf{v}_2 \\ \vdots \\ \sigma_n^{1/2}\mathbf{v}_n \end{bmatrix} \tag{21}$$

Clearly $\mathbf{z}$ is a basis in $\mathbf{X}^{+/-}$ and $\bar{\mathbf{z}}$ is a basis in $\mathbf{X}^{-/+}$, and they have the property that

$$\mathbb{E}\{\mathbf{z}\mathbf{z}^\top\} = \Sigma = \mathbb{E}\{\bar{\mathbf{z}}\bar{\mathbf{z}}^\top\},$$

Will show that this is a sort of “balancing” property for stochastic systems (same Gramian). As for deterministic systems this will be an Important property for doing model reduction. In fact important for order estimation.

Homework

May seem more natural to keep the normalized vectors $\mathbf{u}$ and $\mathbf{v}$ as state variables. Show that this choice leads to **different realization parameters** $(A, C, \bar{C})$ of $\Phi(z)_+$.

Instead $\mathbf{z}$ and $\bar{\mathbf{z}}$ are particular bases in $\mathbf{X}^{+/-}$ and $\mathbf{X}^{-/+}$ which lead to the same realization parameters $(A, C, \bar{C})$.

How do we compute the canonical correlation coefficients and the canonical vectors ?

Assume we are given the covariance sequence $\{\Lambda_k\}$. Need to compute eigenvalues/eigenvectors of $\hat{H}_\infty^\top \hat{H}_\infty$ or $\hat{H}_\infty \hat{H}_\infty^\top$. Recall

$$T_- := \mathbb{E}\{\mathbf{y}_- \mathbf{y}_-^\top\} = L_- L_-^\top \qquad T_+ := \mathbb{E}\{\mathbf{y}_+ \mathbf{y}_+^\top\} = L_+ L_+^\top$$

Then

$$\begin{aligned}
\hat{H}_\infty^\top \hat{H}_\infty &= (L_-)^{-1} H_\infty^\top L_+^{-\top} L_+^{-1} H_\infty (L_-)^{-\top} \\
&= (L_-)^{-1} \bar{\Omega}_\infty \, \Omega_\infty^\top (T_+)^{-1} \Omega_\infty \, \bar{\Omega}_\infty^\top (L_-)^{-\top} \\
\hat{H}_\infty \hat{H}_\infty^\top &= (L_+)^{-1} H_\infty (L_-)^{-\top} (L_-)^{-1} H_\infty^\top L_+^{-\top} \\
&= (L_+)^{-1} \Omega_\infty \, \bar{\Omega}_\infty^\top (T_-)^{-1} \bar{\Omega}_\infty \, \Omega_\infty^\top L_+^{-\top}
\end{aligned}$$

C.C.A. and Riccati equations

Theorem 17 *Let P_- and P_+ be the minimal (stabilizing) solutions of the two dual Riccati inequalities (in fact of the dual LMI's)*

$$
\begin{aligned}
P = APA^\top + (\bar{C}^\top - APC^\top)\Delta(P)^{-1}(\bar{C}^\top - APC^\top)^\top \geq 0 \\
\bar{P} = A^\top \bar{P} A + (C^\top - A^\top \bar{P}\bar{C}^\top)\Delta(\bar{P})^{-1}(C - \bar{C}\bar{P}A) \geq 0
\end{aligned}
$$

Then

$$
P_- = \bar{\Omega}_\infty^\top (T_-)^{-1} \bar{\Omega}_\infty , \qquad \bar{P}_+ = \Omega_\infty^\top (T_+)^{-1} \Omega_\infty .
$$

Theorem 18 *The eigenvalues of $P_-\bar{P}_+$ are the squares of the canonical correlation coefficients of the finite dimensional process $\mathbf{y}$.*

Proof : Let $\hat{H}_\infty^\top \hat{H}_\infty \, v = \sigma^2 v$. Since

$$
\begin{aligned}
\bar{\Omega}_\infty^\top (L_-)^{-\top} \hat{H}_\infty^\top \hat{H}_\infty &= \bar{\Omega}_\infty^\top (L_-)^{-\top} (L_-)^{-1} H_\infty^\top L_+^{-\top} L_+^{-1} H_\infty (L_-)^{-\top} \\
&= \bar{\Omega}_\infty^\top (T_-)^{-1} \bar{\Omega}_\infty \, \Omega_\infty^\top (T_+)^{-1} \Omega_\infty \, \bar{\Omega}_\infty^\top (L_-)^{-\top} \\
&= P_- \bar{P}_+ \; \bar{\Omega}_\infty^\top (L_-)^{-\top}
\end{aligned}
$$

it follows that $\bar{\Omega}_\infty^\top (L_-)^{-\top} v$ is an eigenvector of $P_-\bar{P}_+$ with eigenvalue σ^2.

Proof: Hankel matrix and Riccati equations

A forward-backward pair of stochastic realizations of the same process $\mathbf{y}$

$$\begin{cases} \mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{w}(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{w}(t) \end{cases}, \qquad \begin{cases} \bar{\mathbf{x}}(t-1) = A^\top\bar{\mathbf{x}}(t) + \bar{B}\bar{\mathbf{w}}(t) \\ \mathbf{y}(t) = \bar{C}\bar{\mathbf{x}}(t) + \bar{D}\bar{\mathbf{w}}(t), \end{cases}$$

where $\bar{\mathbf{x}}(t) := P^{-1}\mathbf{x}(t+1)$, $\bar{P} := \mathbb{E}\,\bar{\mathbf{x}}(t)\bar{\mathbf{x}}(t)^\top = P^{-1}$, $\bar{C} = CPA^\top + DB^\top$ etc.
Then

$$\mathbf{y}_+ = \begin{bmatrix} \mathbf{y}(0) \\ \mathbf{y}(1) \\ \mathbf{y}(2) \\ \vdots \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \mathbf{x}(0) \overset{\perp}{+} \begin{bmatrix} D & 0 & \dots & & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \dots & & & \ddots & 0 \end{bmatrix} \mathbf{w}_+ := \Omega_\infty \mathbf{x}(0) \overset{\perp}{+} \Psi_\infty \mathbf{w}_+,$$

$$\mathbf{y}_- = \begin{bmatrix} \mathbf{y}(-1) \\ \mathbf{y}(-2) \\ \mathbf{y}(-3) \\ \vdots \end{bmatrix} = \begin{bmatrix} \bar{C} \\ \bar{C}A^\top \\ \bar{C}(A^\top)^2 \\ \vdots \end{bmatrix} \bar{\mathbf{x}}(-1) \overset{\perp}{+} \begin{bmatrix} \bar{D} & 0 & \dots & & 0 \\ \bar{C}\bar{B} & \bar{D} & 0 & \dots & 0 \\ \bar{C}A^\top\bar{B} & \bar{C}\bar{B} & \bar{D} & \dots & 0 \\ \dots & & & \ddots & 0 \end{bmatrix} \bar{\mathbf{w}}_- := \bar{\Omega}_\infty \bar{\mathbf{x}}(-1) \overset{\perp}{+} \bar{\Psi}_\infty \bar{\mathbf{w}}_-$$

Note that $\mathbf{w}_+ \perp \mathbf{H}^-(\mathbf{x}, \mathbf{y}) \subset \mathbf{H}^-(\bar{\mathbf{w}}_-)$ so that

$$H_\infty := \mathbb{E}\{\mathbf{y}_+ \mathbf{y}_-^\top\} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \begin{bmatrix} \bar{C} \\ \bar{C}A^\top \\ \bar{C}(A^\top)^2 \\ \vdots \end{bmatrix}^\top := \Omega_\infty \bar{\Omega}_\infty^\top$$

Since $\bar{\mathbf{x}}(-1) := P^{-1}\mathbf{x}(0)$ so $\mathbb{E}\,\mathbf{x}(0)\bar{\mathbf{x}}(-1)^\top = \mathbb{E}\,\mathbf{x}(0)\mathbf{x}(0)^\top P^{-1} = I$.

Hankel matrix and Riccati equations cont.d

Rewrite

$$\mathbf{y}_- = \bar{\Omega}_\infty P^{-1}\mathbf{x}(0) \overset{\perp}{+} \Psi_\infty \bar{\mathbf{w}}_- \,, \qquad \mathbf{y}_+ = \Omega_\infty P \bar{\mathbf{x}}(-1) \overset{\perp}{+} \bar{\Psi}_\infty \mathbf{w}_+$$

Classical estimation formulas

$$\mathbb{E}\,[\,\mathbf{x}(0)\mid \mathbf{H}^-(\mathbf{y})\,] = \mathbb{E}\,[\,\mathbf{x}(0)\mid \mathbf{y}_-\,] = \mathbb{E}\,\{\mathbf{x}(0)(\mathbf{y}_-)^\top\}(T_-)^{-1}\,\mathbf{y}_-$$

But $\mathbb{E}\,\{\mathbf{y}_-\,\mathbf{x}(0)^\top\} = \bar{\Omega}_\infty$ and so

$$\mathbb{E}\,[\,\mathbf{x}(0)\mid \mathbf{H}^-(\mathbf{y})\,] = \bar{\Omega}_\infty^\top (T_-)^{-1}\,\mathbf{y}_- \tag{22}$$

This is the s.s. Kalman filter estimate ! Hence

$$P_- = \mathrm{Var}\,\{\,\mathbb{E}\,[\,\mathbf{x}(0)\mid \mathbf{H}^-(\mathbf{y})\,]\} = \bar{\Omega}_\infty^\top (T_-)^{-1}\,\bar{\Omega}_\infty\,.$$

Dually,

$$\bar{P}_+ = \mathrm{Var}\,\{\,\mathbb{E}\,[\,\bar{\mathbf{x}}(-1)\mid \mathbf{H}^+(\mathbf{y})\,]\} = \Omega_\infty^\top (T_+)^{-1}\,\Omega_\infty\,.$$

The Stochastic Gramians

Consider

$$P_- = \mathrm{Var}\{\mathbb{E}[\mathbf{x}(0) \mid \mathbf{H}^-(\mathbf{y})]\} = \bar{\Omega}_\infty^\top (T_-)^{-1} \bar{\Omega}_\infty \,.$$

$$\bar{P}_+ = \mathrm{Var}\{\mathbb{E}[\bar{\mathbf{x}}(-1) \mid \mathbf{H}^+(\mathbf{y})]\} = \Omega_\infty^\top (T_+)^{-1} \Omega_\infty \,.$$

These are the stochastic analogs of the **recostructibility and observability Gramians**. Note that they depend only on the parameters of $\Omega_\infty, \bar{\Omega}_\infty$ and T_-, T_+. In they fact depend only on $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ since the two dual Riccati equations depend only on these parameters.

Proposition 6 *The stochastic reconstructibilityand observability Gramians, P_- and $\bar{P}_+$ don't depend on the particular state space model (realization) but depend only on the realization $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ of $\Phi_+(z)$.*

Stationary stochastic (Positive-Real) Balancing

Let $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ be a minimal realization of $\Phi(z)_+$ and look at the dual Riccati equations

$$
\begin{aligned}
P &= APA^\top + (\bar{C}^\top - APC^\top)(\Lambda_0 - CPC^\top)^{-1}(\bar{C} - CPA^\top) \\
\bar{P} &= A^\top \bar{P} A + (C^\top - A^\top \bar{P} \bar{C}^\top)(\Lambda_0 - \bar{C}\bar{P}\bar{C}^\top)^{-1}(C - \bar{C}\bar{P}A)
\end{aligned}
$$

Definition 3 *One says that the system $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ is in* **stochastic or Positive-Real balanced form** *if the two minimal solutions P_-, $\bar{P}_+$, are diagonal and equal.*

Idea: balancing observability with *recostructability* i.e. reconstructing he state at time zero from the future evolution of $\mathbf{y}$. Riccati equations now play the role of the Lyapunov equations

$$
\Pi = A\Pi A^\top + \bar{C}^\top \bar{C}, \qquad \Delta = A^\top \Delta A + C^\top C,
$$

Here P_-, $\bar{P}_+$ have interpretation as (stochastic) observability and recostructibility Gramians.

Stationary stochastic (Positive-Real) Balancing

Recall that the forward and backward steady state Kalman filter realizations

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

depend only on $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ and are the same for all minimal (forward and backward) realizations. The state covariances are related by $\mathbb{E}\{\bar{\mathbf{x}}(t)\bar{\mathbf{x}}(t)^\top\} = \bar{P} = P^{-1}$.

A change of basis $(A, C, \bar{C}^\top,) \Rightarrow (T^{-1}AT, CT, T^{-1}\bar{C}^\top,)$ induces a change of basis in all realizations, in particular on the Kalman filters. The state covariances transform according to $\hat{P}_- = TP_-T^\top, \qquad \hat{\bar{P}}_+ = T^{-\top}\bar{P}_+T^{-1}.$

Theorem 19 *There is an essentially unique T which brings the system $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ into stochastically balanced form. In this basis*

$$\hat{P}_- = TP_-T^\top = \Sigma = T^{-\top}\bar{P}_+T^{-1} = \hat{\bar{P}}_+$$

where Σ is the diagonal matrix of canonical correlation coefficients. Hence the CCA procedure produces automatically stochastic balanced realizations.

Stochastic model reduction from balancing

Assume we have a stationary process of very high order n. Do canonical correlation analysis. Pick n-dimensional canonical bases $\mathbf{z}$, $\bar{\mathbf{z}}$ in the predictor spaces so that stationary canonical correlation coefficients

$$\mathrm{Var}\{\mathbf{z}\} = \mathrm{Var}\{\bar{\mathbf{z}}\} = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_n\} = P_- = \bar{P}_+$$

Pick the first $k < n$ canonical correlation coefficients and define k-dimensional subvectors

$$\mathbf{z}_1 = \begin{bmatrix} \sigma_1^{1/2}\mathbf{u}_1 \\ \sigma_2^{1/2}\mathbf{u}_2 \\ \vdots \\ \sigma_k^{1/2}\mathbf{u}_k \end{bmatrix}, \qquad \bar{\mathbf{z}}_1 = \begin{bmatrix} \sigma_1^{1/2}\mathbf{v}_1 \\ \sigma_2^{1/2}\mathbf{v}_2 \\ \vdots \\ \sigma_k^{1/2}\mathbf{v}_k \end{bmatrix}$$

by which define a k-dimensional reduced system with $\Sigma_1 = \mathrm{diag}\{\sigma_1, \ldots, \sigma_k\}$ and

$$\begin{cases} A_1 = \mathbb{E}\,\mathbf{z}_1(t+1)\mathbf{z}_1(t)^\top \Sigma_1^{-1}, \\ C_1 = \mathbb{E}\,\mathbf{y}(t)\mathbf{z}_1^\top \Sigma_1^{-1}, \\ \bar{C}_1 = \mathbb{E}\,\mathbf{y}(t-1)\mathbf{z}_1(t)^\top. \end{cases}$$

Stationary stochastic model reduction 2

Get a reduced-degree half-spectrum $\Phi_1(z) = C_1(zI - A_1)^{-1}\bar{C}_1^\top + \frac{1}{2}\Lambda_0$. Clearly we get "principal submatrices" of the original $(A, C, \bar{C})$

$$A = \begin{bmatrix} A_1 & A_{12} \\ A_{21} & A_2 \end{bmatrix}, \qquad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \qquad \bar{C} = \begin{bmatrix} \bar{C}_1 & C_2 \end{bmatrix}$$

Questions:

1. Is A_1 also stable ?

2. Is $(A_1, C_1, \bar{C}_1)$ still a minimal triplet ?

3. **Is $\Phi_1(z)$ still positive real ?**

4. Is $(A_1, C_1, \bar{C}_1)$ still in (stochastic) balanced form ?

Third condition is crucial: only if $\Phi_1(z)$ is positive real we can compute a stochastic model of reduced the complexity.

Stationary stochastic model reduction: Proof

Theorem 20 *If* $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ *is positive real and in stochastic balanced form, then the reduced degree function* $\Phi_1(z)$ *is positive real. In particular* A_1 *is stable. In general* $(A_1, C_1, \bar{C}_1)$ *is not in balanced form.*

Proof : Since $\Phi_+(z)$ is positive real, and $\Sigma = \mathrm{diag}\{\Sigma_1,\ \Sigma_2\}$ solves the LMI,

$$M(\Sigma) = \begin{bmatrix} \Sigma_1 - A_1\Sigma_1 A_1^\top - A_{12}\Sigma_2 A_{12}^\top & * & \bar{C}_1^\top - A_1\Sigma_1 C_1^\top - A_{12}\Sigma_2 C_2' \\ * & * & * \\ \bar{C}_1 - C_1\Sigma_1 A_1^\top - C_2\Sigma_2 A_{12}^\top & * & \Lambda_0 - C_1\Sigma_1 C_1^\top - C_2\Sigma_2 C_2^\top \end{bmatrix} \geq 0,$$

where the blocks which do not enter the analysis are marked with an asterisk. Consequently,

$$M_1(\Sigma_1) - \begin{bmatrix} A_{12} \\ C_2 \end{bmatrix} \Sigma_2 \begin{bmatrix} A_{12} \\ C_2 \end{bmatrix}^\top \geq 0, \quad \text{where} \quad M_1(\Sigma_1) = \begin{bmatrix} \Sigma_1 - A_1\Sigma_1 A_1^\top & \bar{C}_1^\top - A_1\Sigma_1 C_1^\top \\ \bar{C}_1 - C_1\Sigma_1 A_1^\top & \Lambda_0 - C_1\Sigma_1 C_1^\top \end{bmatrix}$$

Therefore, $M(\Sigma_1) \geq 0$ is the LMI corresponding to the reduced triplet $(A_1, C_1, \bar{C}_1)$.

BASIC IDEA OF SUBSPACE IDENTIFICATION

We are given a "long" trajectory of a finite dimensional process (zero mean second order stationary)

$$\{y_0, y_1, y_2, \dots, y_{N'}\}, \qquad y_t \in \mathbb{R}^m$$

Assume we can also observe a **state trajectory** $\{x_0, x_1, x_2, \dots, x_{N'}\}$ of an underlying model, generating the data

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

$$\begin{cases} x_{t+1} = Ax_t + Bw_t \\ y_t \;\; = Cx_t + Dw_t\,, \qquad t = 0, 1, 2, \dots, N' \end{cases}$$

Form the "tail" matrices Y_t^N, X_t^N,

$$\begin{aligned} Y_t^N &:= [\, y_t,\, y_{t+1},\, y_{t+2}, \dots,\, y_{t+N}] \\ X_t^N &:= [\, x_t,\, x_{t+1},\, x_{t+2}, \dots,\, x_{t+N}] \end{aligned}$$

We assume N' large enough so that we can form tail matrices of the same length $N+1$.

The Idea of subspace identification

Then can write

$$\begin{bmatrix} X_{t+1}^N \\ Y_t^N \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} X_t^N + \begin{bmatrix} B \\ D \end{bmatrix} W_t^N$$

Interpret as **Linear Regression !** Solve by Least Squares :

$$\min_{A,C} \| \begin{bmatrix} X_{t+1}^N \\ Y_t^N \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} X_t^N \|$$

getting

$$\begin{bmatrix} \hat{A} \\ C \end{bmatrix}_N := \frac{1}{N} \begin{bmatrix} X_{t+1}^N \\ Y_t^N \end{bmatrix} (X_t^N)^\top \left\{ \frac{1}{N} X_t^N (X_t^N)^\top \right\}^{-1}$$

Also for the corresponding backward model get: $\hat{\bar{C}}_N := \frac{1}{N} Y_{t-1}^N (X_t^N)^\top$

Asymptotics

If the generating processes are **second order ergodic**,

$$\lim_{N\to\infty} \frac{1}{N} Y_t^N (X_s^N)^\top = \mathbb{E}\,\mathbf{y}(t)\mathbf{x}(s)^\top$$

Theorem 21 *If the underlying processes* $\mathbf{y}, \mathbf{x}$ *are second order ergodic, and the limit* $\lim_{N\to\infty} \dfrac{1}{N} X_t^N (X_t^N)^\top$ *is invertible,*

$$\lim_{N\to\infty} \hat{\begin{bmatrix} A \\ C \end{bmatrix}}_N = \begin{bmatrix} A \\ C \end{bmatrix}, \qquad \lim_{N\to\infty} \hat{\bar{C}}_N = \bar{C}$$

i.e. we get **consistent estimates** *of* $A, C, \bar{C}$*.*

Proof: see equation (4).

Basic assumption on the data: the limit

$$\lim_{N\to\infty} \frac{1}{N}\, Y^N_{t+\tau}\,(Y^N_t)^\top := \Lambda_\tau$$

exists for all $\tau \geq 0$ and is independent of t. If the data are compatible with second order ergodicity (no periodic trends etc.) Λ_τ is the true covariance. Then as $N \to \infty$ the family $\{\mathbf{Y}_t\,;\, t \geq 0\}$ of **infinite tail matrices** where

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

behaves (to second order) like a w-stationary stochastic process !
For infinite tail matrices can use the “stochastic” notation:

$$\mathbb{E}\,\mathbf{Y}_{t+\tau}\mathbf{Y}_t^\top \equiv \lim_{N\to\infty} \frac{1}{N}\, Y^N_{t+\tau}\,(Y^N_t)^\top = \Lambda_\tau$$

Can set up a isometric isomorphism between $\mathbf{H}(\mathbf{y})$ and the inner product space spanned by the rows of $\{Y_t\,;\, t \geq 0\}$.

$$\{\mathbf{y}(t)\,;\, t \geq 0\} \iff \{\mathbf{Y}_t\,;\, t \geq 0\}$$

Interpretation of the Least squares estimates $(N \to \infty)$:

L.S. formulas with semi-infinite tail matrices

$$\begin{bmatrix} \hat{A} \\ C \end{bmatrix} := \mathbb{E} \begin{bmatrix} X_{t+1} \\ Y_t \end{bmatrix} X_t^{\top} \left\{ \mathbb{E} X_t X_t^{\top} \right\}^{-1}$$

$$\hat{\bar{C}} := \mathbb{E} Y_{t-1} X_t^{\top}$$

are exactly the same formulas as (4).

How to get a state sequence ?

CONSTRUCT THE STATE FROM THE OBSERVED OUTPUT SAMPLES

Can construct a sample estimate of $\mathbf{X}^{+/-}$ and a canonical basis by sample canonical correlation analysis. Use exactly the same procedure as that done for random variables. Just substitute random variables $\mathbf{y}(t)$ by long (ideally semi-infinite) tail sequences Y_t^N. Below for simplicity we shall omit the superscript ^{N}.

Sample CCA Algorithm: Given observed data $\{y_t\,;\, t = 0, 1, 2, \ldots, N'\}$,

1. **Form sample covariance estimates**

$$\hat{\Lambda}_\tau = \frac{1}{N}\sum_{t=0}^{N-\tau} y_{t+\tau} y_t^\top\,, = \frac{1}{N} Y_\tau [Y_0]^\top \qquad \tau = 0, 1, \ldots, 2\nu + 1 \ll N$$

2. **Form the sample Hankel matrix**

$$\mathbb{H}_{\nu+1,\nu+1} := \begin{bmatrix} \hat{\Lambda}_1 & \hat{\Lambda}_2 & \hat{\Lambda}_3 & \dots & \hat{\Lambda}_{\nu+1} \\ \hat{\Lambda}_2 & \hat{\Lambda}_3 & \hat{\Lambda}_4 & \dots & \hat{\Lambda}_{\nu+2} \\ \hat{\Lambda}_3 & \hat{\Lambda}_4 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \hat{\Lambda}_{\nu+1} & \dots & \dots & \dots & \hat{\Lambda}_{2\nu+1} \end{bmatrix}$$

Assume ν "large enough", ideally should be that $(\nu > n)$ but here we assume ν very large. Define past and future strings at "time" $\nu+1$

$$\mathbf{Y}^{-}_{\nu+1} := \begin{bmatrix} Y_\nu \\ Y_{\nu-1} \\ \dots \\ Y_0 \end{bmatrix}, \qquad \mathbf{Y}^{+}_{\nu+1} := \begin{bmatrix} Y_{\nu+1} \\ Y_{\nu+2} \\ \dots \\ Y_{2\nu+1} \end{bmatrix} \quad \text{both of dimension } m(\nu+1)\times N$$

Formally: $$\mathbb{H}_{\nu+1,\nu+1} = \frac{1}{N}\,\mathbf{Y}^{+}_{\nu+1}(\mathbf{Y}^{-}_{\nu+1})^{\top}.$$

3. **Form the Toeplitz matrices**

$$T^{-}_{\nu+1} := \begin{bmatrix} \hat{\Lambda}_0 & \hat{\Lambda}_1 & \hat{\Lambda}_2 & \dots & \hat{\Lambda}_\nu \\ \hat{\Lambda}_1^\top & \hat{\Lambda}_0 & \hat{\Lambda}_1 & \dots & \hat{\Lambda}_{\nu-1} \\ \hat{\Lambda}_2^\top & \hat{\Lambda}_1^\top & \ddots & \dots & \dots \\ \dots & \dots & \dots & \ddots & \dots \\ \hat{\Lambda}_\nu^\top & \dots & \dots & \dots & \hat{\Lambda}_0 \end{bmatrix} \qquad T^{+}_{\nu+1} := \begin{bmatrix} \hat{\Lambda}_0 & \hat{\Lambda}_1^\top & \hat{\Lambda}_2^\top & \dots & \hat{\Lambda}_\nu^\top \\ \hat{\Lambda}_1 & \hat{\Lambda}_0 & \hat{\Lambda}_1^\top & \dots & \hat{\Lambda}_{\nu-1}^\top \\ \hat{\Lambda}_2 & \hat{\Lambda}_1 & \ddots & \dots & \dots \\ \dots & \dots & \dots & \ddots & \dots \\ \hat{\Lambda}_\nu^\top & \dots & \dots & \dots & \hat{\Lambda}_0 \end{bmatrix}$$

4. **Compute (Cholesky) factors**

$$T^{-}_{\nu+1} = L^{-}_{\nu+1}(L^{-}_{\nu+1})^\top, \quad T^{+}_{\nu+1} = L^{+}_{\nu+1}(L^{+}_{\nu+1})^\top$$

5. **Normalization:**

$$\hat{\mathbb{H}}_{\nu+1,\nu+1} = (L^{+}_{\nu+1})^{-1}\,\mathbb{H}_{\nu+1,\nu+1}\,(L^{-}_{\nu+1})^{-\top}$$

6. **SVD :**

$$\hat{\mathbb{H}}_{\nu+1,\nu+1} \;=\; \begin{bmatrix} \hat{\mathbb{U}}_{\nu+1} & \tilde{\mathbb{U}}_{\nu+1} \end{bmatrix} \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & \tilde{\Sigma} \end{bmatrix} \begin{bmatrix} \hat{\mathbb{V}}_{\nu+1} & \tilde{\mathbb{V}}_{\nu+1} \end{bmatrix}^\top$$

7. **Order estimation:** Choose n so that $\hat{\Sigma} \gg \tilde{\Sigma}$. Then we have a rank factorization

$$\mathbb{H}_{\nu+1,\nu+1} = \Omega_{\nu+1}\,\bar{\Omega}^{\top}_{\nu+1}\,, \qquad \Omega_{\nu+1} = L^{+}_{\nu+1}\hat{\mathbb{U}}_{\nu+1}\hat{\Sigma}^{1/2}, \quad \bar{\Omega}^{\top}_{\nu+1} = \hat{\Sigma}^{1/2}\hat{\mathbb{V}}^{\top}_{\nu+1}L^{-}_{\nu+1}$$

8. **Canonical Variables**

$$\begin{aligned} \hat{\mathbf{Y}}^{-}_{\nu+1} &= (L^{-}_{\nu+1})^{-1}\mathbf{Y}^{-}_{\nu+1}\,, \qquad & \hat{\mathbf{Y}}^{+}_{\nu+1} &= (L^{+}_{\nu+1})^{-1}\mathbf{Y}^{+}_{\nu+1} \\ \mathbf{U}_{\nu+1} &= \hat{\mathbb{V}}^{\top}_{\nu+1}\hat{\mathbf{Y}}^{-}_{\nu+1} & \mathbf{V}_{\nu+1} &= \hat{\mathbb{U}}^{\top}\hat{\mathbf{Y}}^{+}_{\nu+1}\,, \end{aligned}$$

9. **Balancing**

$$\mathbf{Z}_{\nu+1} := \hat{\Sigma}^{1/2}\,\hat{\mathbb{V}}^{\top}_{\nu+1}\,\hat{\mathbf{Y}}^{-}_{\nu+1} = \hat{\Sigma}^{1/2}\,\hat{\mathbb{V}}^{\top}(L^{-}_{\nu+1})^{-1}\mathbf{Y}^{-}_{\nu+1} = \bar{\Omega}^{\top}_{\nu+1}\,(T^{-}_{\nu+1})^{-1}\mathbf{Y}^{-}_{\nu+1},$$

The last formula is the sample analog of (22). Note that in principle we would need infinite past data.

10. **Compute $\mathbf{Z}_\nu$ the state at time ν** If Y_{-1} were available we would have

$$\mathbf{Z}_\nu = \bar{\Omega}_{\nu+1}^\top (T_{\nu+1}^-)^{-1} \sigma(\mathbf{Y}_{\nu+1}^-) \qquad \text{where} \quad \sigma(\mathbf{Y}_{\nu+1}^-) = \begin{bmatrix} Y_{\nu-1} \\ Y_{\nu-2} \\ \dots \\ Y_0 \\ Y_{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_\nu \\ Y_{-1} \end{bmatrix}$$

approximate by deleting Y_{-1}:

11. **Compute** $\bar{\Omega}_\nu := \uparrow \bar{\Omega}_{\nu+1}$: delete the last block of m rows from $\bar{\Omega}_{\nu+1}$. Let T_ν be the principal $m\nu \times m\nu$ submatrix of $T_{\nu+1}$ then

$$\mathbf{Z}_\nu := \bar{\Omega}_\nu^\top (T_\nu^-)^{-1} \mathbf{Y}_\nu^-$$

is the **Canonical Variable at time** ν. Still this has finite memory. In theory should involve the infinite past.

12. **Solve by Least-Squares to get** A, C

$$\begin{bmatrix} \mathbf{Z}_{\nu+1} \\ Y_\nu \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \mathbf{Z}_\nu + \begin{bmatrix} K \\ I \end{bmatrix} \hat{\mathbf{E}}_\nu$$

13. **Compute** $\bar{C}$

$$\bar{C} = \frac{1}{N} Y_\nu \, \mathbf{Z}_{\nu+1}^\top \, .$$

This provides estimates of $C, \bar{C}, A$.

Need to deal with data on a finite interval

The CCA procedure as if infinite Y_k were available, induces some errors.
We must do the **finite data** setting right: constructing stationary models given **finite data** (Partial realization).
Will first do CCA and state construction based on *finite histories* of the process; after we shall translate the procedure in terms of time series data.

Motivation of the finite data setting

Meaning of Finite Data: we have a finite string of observed data $\{y_0, y_1, y_2, \ldots, y_N\}$ assume N very large so that

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

are a "good approximation" of the *true* covariance lags, $\{\Lambda(0), \Lambda(1), \ldots, \Lambda(T)\}$
However the sample estimates are poor if k is large. Need to bound T so that $T << N$. Rule of thumb is $T \leq (1/50)N$.
Key point: having only a *finite* set of (estimated) covariance lags,

$$\{\Lambda_0, \Lambda_1, \ldots, \Lambda_T\},$$

is exactly equivalent to having a **finite chunk** of random variables extracted from $\mathbf{y}$:

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

Can pretend we have observations of $\mathbf{y}$ on the **finite interval** $[0, T]$.
Will initially pretend these are true covariances of some p.n.d. process. In particular they form a block-Toeplitz matrix which is positive definite.

Constructing the state from finite history

We want to do a geometric construction: construct a state space realization using only the finite chunk of available random variables of the process.

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

Let t be a "present time" between time 0 and T (to simplify notations $T = 2t$). Start from finite past and future vectors and a Hankel matrix of covariance data

$$\mathbf{y}_t^- := \begin{bmatrix} \mathbf{y}(t-1) \\ \mathbf{y}(t-2) \\ \vdots \\ \mathbf{y}(0) \end{bmatrix}, \quad \mathbf{y}_t^+ := \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t+1) \\ \vdots \\ \mathbf{y}(T) \end{bmatrix}, \quad H = \mathbb{E}\{\mathbf{y}_+ \mathbf{y}_-^\top\} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \ldots & \Lambda_t \\ \Lambda_2 & \Lambda_3 & \ldots & \Lambda_{t+1} \\ \vdots & \vdots & \ddots & \ldots \\ \Lambda_{t+1} & \Lambda_{t+2} & \ldots & \Lambda_T \end{bmatrix},$$

Models describing finite data must be non-stationary !

What to expect: Non-stationary state space models of the type

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

defined on the finite interval $[0, T]$. The state must be Markov on a finite interval. The noise process in the Kalman Filter realization must be the normalized one-step prediction error given **data on a finite interval.**

Both turn out to be non-stationary processes. Note however that $(A, C, \bar{C})$ must be constant since $\mathbf{y}$ is **stationary** so its covariance still must admit stationary realizations

$$\Lambda_k = CA^{k-1}\bar{C}$$

with constant parameters $(C, A, \bar{C})$.

Splitting subspaces for finite data

Present and past Hilbert spaces of random variables at time t,

$$\mathbf{Y}_t^- := \text{span}\{\mathbf{y}(k)\ \ 0 \le k < t\}, \qquad \mathbf{Y}_t^+ := \text{span}\{\mathbf{y}(k)\ \ t \le k \le T\}$$

Theorem 22 *The* finite interval *forward and backward predictor spaces,*

$\hat{\mathbf{X}}^{+/-}(t) := \text{span}_{t \le k \le T}\{\mathbb{E}[\mathbf{y}(k) \mid \mathbf{Y}_t^-]\}$ *and* $\hat{\mathbf{X}}^{-/+}(t) = \text{span}_{0 \le k < t}\{\mathbb{E}[\mathbf{y}(k) \mid \mathbf{Y}_t^+]\}$
are minimal state spaces for $\mathbf{y}$ *on* $[0, T]$*; i.e. letting* $\hat{\mathbf{X}}$ *be either* $\hat{\mathbf{X}}^{+/-}$ *or* $\hat{\mathbf{X}}^{-/+}$*, and*

$$\hat{\mathbf{X}}_t^- := \text{span}\{\hat{\mathbf{X}}(s)\,;\, 0 \le s < t\}, \qquad \hat{\mathbf{X}}_t^+ := \text{span}\{\hat{\mathbf{X}}(s)\,;\, t \le s \le T\}$$

the Markovian splitting property holds

$$\mathbf{Y}_t^- \vee \hat{\mathbf{X}}_t^- \perp \mathbf{Y}_t^+ \vee \hat{\mathbf{X}}_t^+ \mid \hat{\mathbf{X}}(t)\,.$$

Therefore any choice of basis in either $\hat{\mathbf{X}}^{+/-}$ or $\hat{\mathbf{X}}^{-/+}$ will define a state space model for $\mathbf{y}$ on $[0, T]$. Dimension will vary with t.

Constructing the state from finite data

Assume $\mathbf{y}$ has a minimal stationary realization with state vector $\mathbf{x}(t)$ (or $\bar{\mathbf{x}}(t)$) of dimension n.

Theorem 23 (Invariance Theorem) *Suppose that $n \le t \le T - n$. Then*

$$\hat{\mathbf{x}}(t) := \mathbb{E}[\,\mathbf{x}(t) \mid \mathbf{Y}_t^-\,], \qquad \hat{\bar{\mathbf{x}}}(t) := \mathbb{E}[\,\bar{\mathbf{x}}(t) \mid \mathbf{Y}_t^+\,]$$

are bases for $\hat{\mathbf{X}}^{+/-}(t)$ and $\hat{\mathbf{X}}^{-/+}(t)$. All $\mathbf{x}$'s which project to the same $\hat{\mathbf{x}}$ describe models with the same minimal triplet $(A, C, \bar{C})$. Dually all $\bar{x}$ which project to the same $\hat{\bar{\mathbf{x}}}$ describe models with the same minimal triplet $(A^\top, \bar{C}, C)$.

Transient Kalman filter as a state space model

Take the state of any minimal stationary realization

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

Then $\hat{\mathbf{x}}(t) = \mathbb{E}\,[\,\mathbf{x}(t) \mid \mathbf{Y}_t^-\,]$ is a basis $\hat{\mathbf{x}}(t)$ in $\hat{\mathbf{X}}^{+/-}(t)$; i.e. is the state of a transient Kalman filter state-space model of $\mathbf{y}$ on $[0, T]$:

$$\begin{cases} \hat{\mathbf{x}}(t+1) &= A\,\hat{\mathbf{x}}(t) + K(t)\,\hat{\mathbf{e}}(t)\,, \qquad \hat{\mathbf{x}}(0) = 0 \\ \mathbf{y}(t) &= C\,\hat{\mathbf{x}}(t) + \hat{\mathbf{e}}(t) \end{cases}$$

To get a basis in $\hat{\mathbf{X}}^{+/-}(t)$ don't need to know the stationary model. Predictor of finite future of $\mathbf{y}$ based on finite past of $\mathbf{y}$:

$$\hat{\mathbf{X}}^{+/-}(t) = \mathrm{span}\,\{\,\mathbb{E}\,\left[\mathbf{y}(k) \mid \mathbf{y}_t^-\right]\,; t \le k \le T\}$$

Predicted finite future vector

$$\hat{\mathbf{y}}_t^+ := \mathbb{E}\,\left[\mathbf{y}_t^+ \mid \mathbf{y}_t^-\right] = \Omega_t\,\hat{\mathbf{x}}(t)\,, \qquad \Omega_t^\top = \left[C^\top \quad A^\top C^\top \quad \dots \quad (A^\top)^{T-t} C^\top\right]$$

Transient Kalman filter as a state space model (cont'd)

State covariance $\mathbb{E}\hat{\mathbf{x}}(t)\hat{\mathbf{x}}(t)^\top = P(t)$ can be computed by solving a Riccati Difference Equation (RDE)

$$P(k+1) = AP(k)A^\top + (\bar{C}^\top - AP(k)C^\top)\Delta(P(k))^{-1}(\bar{C}^\top - AP(k)C^\top)^\top, \qquad 0 \le k \le t\,,$$

where $\Delta(P(k)) = \Lambda_0 - CP(k)C^\top$, with initial condition $P(0) = 0$.
$A, C, \bar{C}$ are parameters of the stationary model; $\Lambda_k = CA^{k-1}\bar{C}^\top$.

Later we will see how to compute a transient K.F. realization by finite interval CCA of $\mathbf{y}$, without having (or assuming) a stationary model.

Transient Backward Kalman filter as a state space model

Take the state of any minimal backward stationary realization

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

Then $\hat{\bar{\mathbf{x}}}(t) = \mathbb{E}\,[\bar{\mathbf{x}}(t) \mid \mathbf{Y}_t^+]$ is a basis in $\hat{\mathbf{X}}^{-/+}(t)$; i.e. is the state of a backward transient Kalman filter state-space model of $\mathbf{y}$ on $[0,T]$:

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

To get a basis in $\hat{\mathbf{X}}^{-/+}(t)$ don't need to have a stationary model. Backward predictor of finite past of $\mathbf{y}$ based on finite future of $\mathbf{y}$:

$$\hat{\mathbf{X}}^{-/+}(t) = \operatorname{span}\{\mathbb{E}\left[\mathbf{y}(k) \mid \mathbf{y}_t^+\right]; 0 \le k < t\}$$

Back-predicted finite past vector

$$\hat{\mathbf{y}}_t^- := \mathbb{E}\left[\mathbf{y}_t^- \mid \mathbf{y}_t^+\right] = \bar{\Omega}_t\,\hat{\bar{\mathbf{x}}}(t)\,, \qquad \bar{\Omega}_t^\top := \begin{bmatrix}\bar{C}^\top & A\bar{C}^\top & \dots & A^{t-1}\bar{C}^\top\end{bmatrix}$$

Backward Kalman filtering cont.d

The backward Kalman gain is given by

$$\bar{K}(t) = (C^\top - A^\top \bar{P}(t)\bar{C}^\top)(\Lambda_0 - \bar{C}P(t)\bar{C}^\top)^{-1}$$

where $\bar{P}(t) = \mathbb{E}\hat{\bar{\mathbf{x}}}(t)\hat{\bar{\mathbf{x}}}(t)^\top$ is the solution at time t of the Backward Riccati Difference Equation (BRDE)

$$\bar{P}(k-1) = A^\top \bar{P}(k)A + (C^\top - A^\top \bar{P}(k)\bar{C}^\top)(\Lambda_0 - \bar{C}\bar{P}(k)\bar{C}^\top)^{-1}(C^\top - A^\top \bar{P}(k)\bar{C}^\top)^\top,$$

solved backwards with end condition $\bar{P}(T) = 0$.
Later we will see how to compute a transient backward K.F. realization by finite interval CCA of $\mathbf{y}$, without having (or assuming) a stationary model.

$A, C, \bar{C}$ from the state

The state $\hat{\mathbf{x}}(t)$ of the transient Kalman filter model contains all the information needed for reconstructing the minimal triplet $(A, C, \bar{C})$, and the same holds for the backward Kalman filter state $\hat{\bar{\mathbf{x}}}(t)$. This is of importantance in subspace identification, as we shall see.

Proposition 7 *Let $\hat{\mathbf{x}}(t)$ be the state of the forward transient Kalman filter realization, then*

$$\begin{aligned}
A &= \mathbb{E}\{\hat{\mathbf{x}}(t+1)\hat{\mathbf{x}}(t)^\top\}P(t)^{-1} && \text{(23a)}\\
C &= \mathbb{E}\{\mathbf{y}(t)\,\hat{\mathbf{x}}(t)^\top\}P(t)^{-1} && \text{(23b)}\\
\bar{C} &= \mathbb{E}\{\mathbf{y}(t)\,\hat{\mathbf{x}}(t+1)^\top\}, && \text{(23c)}
\end{aligned}$$

Likewise, let $\hat{\bar{x}}(t)$ be the state of the backward transient Kalman filter realization, then

$$\begin{aligned}
A^\top &= \mathbb{E}\{\hat{\bar{\mathbf{x}}}(t-1)\hat{\bar{\mathbf{x}}}(t)^\top\}\bar{P}(t)^{-1} && \text{(24a)}\\
C &= \mathbb{E}\{\mathbf{y}(t)\,\hat{\bar{\mathbf{x}}}(t-1)^\top\} && \text{(24b)}\\
\bar{C} &= \mathbb{E}\{\mathbf{y}(t)\,\hat{\bar{\mathbf{x}}}(t)^\top\}\bar{P}(t)^{-1}, && \text{(24c)}
\end{aligned}$$

$A, C, \bar{C}$ from the state; Proof

Same proof as for stationary models. Multiply from the right by $\mathbf{x}(t)^\top$ and take expectation

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

Since $\hat{\mathbf{e}}(t) = \mathbf{y}(t) - \mathbb{E}\,[\,\mathbf{y}(t) \mid \mathbf{Y}_t^-\,] \perp \mathbf{Y}_t^-$ and $\hat{\mathbf{x}}(t) = \mathbb{E}\,[\,\mathbf{x}(t) \mid \mathbf{Y}_t^-\,]$,

$$\mathbb{E}\,\hat{\mathbf{e}}(t)\hat{\mathbf{x}}(t)^\top = 0$$

Doing C.C.A. of FINITE past and future

The canonical correlation coefficients of a stationary process $\mathbf{y}$ defined on $\mathbb{Z}$ are for the **infinite** past and future histories of $\mathbf{y}$; in fact are the singular values of an **infinite** Hankel matrix. In practice we only have a finite sequence of covariances and can only form a finite Hankel matrix.
Assume we have finite past and future data at some generic time t,

$$\mathbf{y}_t^- = \begin{bmatrix} \mathbf{y}(t-1) \\ \mathbf{y}(t-2) \\ \vdots \\ \mathbf{y}(0) \end{bmatrix}, \qquad \mathbf{y}_t^+ = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t+1) \\ \vdots \\ \mathbf{y}(T) \end{bmatrix},$$

Let say $T = 2t$; CCA involves SVD (after normalization) of the finite Hankel matrix

$$H_{t+1,t} = \mathbb{E}\{\mathbf{y}_+ \mathbf{y}_-^\top\} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \dots & \Lambda_t \\ \Lambda_2 & \Lambda_3 & \dots & \Lambda_{t+1} \\ \vdots & \vdots & \ddots & \dots \\ \Lambda_{t+1} & \Lambda_{t+2} & \dots & \Lambda_T \end{bmatrix},$$

In general with different t we have past and future vectors of different dimension. Will get rectangular Hankel matrices of different dimensions.

CCA of a finite dimensional process cont.d

Let T_t^- and T_{t+1}^+ be the block Toeplitz matrices

$$T_t^- = \mathbb{E}\{\mathbf{y}_t^-(\mathbf{y}_t^-)^\top\}, \qquad T_{t+1}^+ = \mathbb{E}\{\mathbf{y}_t^+(\mathbf{y}_t^+)^\top\}.$$

and let $L_t^-(L_t^-)^\top = T_t^-$, $L_{t+1}^+(L_{t+1}^+)^\top = T_{t+1}^+$. Then $\hat{H}_{t+1,t} = (L_{t+1}^+)^{-1} H_{t+1,t} (L_t^-)^{-\top}$ has a singular-value decomposition

$$\hat{H}_{t+1,t} = U_{t+1} \begin{bmatrix} \Sigma_t & 0 \\ 0 & 0 \end{bmatrix} V_t^\top,$$

where we assume there are n nonzero singular values: the canonical correlation coefficients at time t, arranged in decreasing order

$$1 \geq \sigma_1(t) \geq \sigma_2(t) \geq \sigma_3(t) \ldots \geq \sigma_n(t) > 0\,.$$

U_{t+1}, V_t are orthonormal matrices $U_{t+1}^\top U_{t+1} = I = V_t^\top V_t$.
Rotate the orthonormal basis vector $\bar{\boldsymbol{\nu}}(t)$ in $\mathbf{Y}_t^+$ and $\boldsymbol{\nu}(t)$ in $\mathbf{Y}_t^-$ to recover the **canonical vectors**

$$\mathbf{u}(t) := V_t^\top \boldsymbol{\nu}(t), \qquad \mathbf{v}(t) := U_{t+1}^\top \bar{\boldsymbol{\nu}}(t)\,, \qquad \mathbb{E}\,\mathbf{v}(t)\mathbf{u}(t)^\top = \begin{bmatrix} \Sigma_t & 0 \\ 0 & 0 \end{bmatrix}.$$

CCA and finite interval balancing

Compute canonical correlation coefficients of finite past and future spaces at time t and construct Balanced canonical bases

$$\hat{\mathbf{z}}(t) = \begin{bmatrix} \sigma_1(t)^{1/2}\mathbf{u}_1(t) \\ \sigma_2(t)^{1/2}\mathbf{u}_2(t) \\ \vdots \\ \sigma_n(t)^{1/2}\mathbf{u}_n(t) \end{bmatrix}, \qquad \hat{\bar{\mathbf{z}}}(t) = \begin{bmatrix} \sigma_1(t)^{1/2}\mathbf{v}_1(t) \\ \sigma_2(t)^{1/2}\mathbf{v}_2(t) \\ \vdots \\ \sigma_n(t)^{1/2}\mathbf{v}_n(t) \end{bmatrix}$$

Fact: $\hat{\mathbf{z}}(t)$ is a basis in $\hat{\mathbf{X}}_t^{+/-}$ and $\hat{\bar{\mathbf{z}}}(t)$ is a basis in $\hat{\mathbf{X}}_t^{-/+}$

These bases have the **Finite interval balancing** property

$$\mathbb{E}\{\hat{\mathbf{z}}(t)\hat{\mathbf{z}}(t)^\top\} = \Sigma_t = \mathbb{E}\{\hat{\bar{\mathbf{z}}}(t)\hat{\bar{\mathbf{z}}}(t)^\top\},$$

So Σ_t is the state covariance of both forward and backward Kalman filter realizations. What are the parameters of these realizations?

Stationary $(A, C, \bar{C})$ parameters from finite interval CCA

Use the formulas (23) and (24)

$$
\begin{aligned}
A &= \mathbb{E}\{\hat{\mathbf{z}}(t+1)\hat{\mathbf{z}}(t)^{\top}\}\Sigma_t^{-1} && \text{(25a)}\\
C &= \mathbb{E}\{\mathbf{y}(t)\,\hat{\mathbf{z}}(t)^{\top}\}\Sigma_t^{-1} && \text{(25b)}\\
\bar{C} &= \mathbb{E}\{\mathbf{y}(t)\,\hat{\mathbf{z}}(t+1)^{\top}\}, && \text{(25c)}
\end{aligned}
$$

$$
\begin{aligned}
A^{\top} &= \mathbb{E}\{\hat{\bar{\mathbf{z}}}(t-1)\hat{\bar{\mathbf{z}}}(t)^{\top}\}\Sigma_t^{-1} && \text{(25d)}\\
C &= \mathbb{E}\{\mathbf{y}(t)\,\hat{\bar{\mathbf{z}}}(t-1)^{\top}\} && \text{(25e)}\\
\bar{C} &= \mathbb{E}\{\mathbf{y}(t)\,\hat{\bar{\mathbf{z}}}(t)^{\top}\}\Sigma_t^{-1}, && \text{(25f)}
\end{aligned}
$$

No need to solve Riccati Equation! the solution $\Sigma(t)$ is obtained by CCA. But $\Sigma_t > 0$ is only the **Transient solution** of the RDE at time t !!

Coherent bases

Hence we need to compute $\hat{\mathbf{z}}(t+1)$. In principle should re-do CCA at time $t+1$. Doing CCA for **shifted** finite past and future data

$$\mathbf{y}_{t+1}^{-} = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t-1) \\ \vdots \\ \mathbf{y}(0) \end{bmatrix}, \qquad \mathbf{y}_{t+1}^{+} = \begin{bmatrix} \mathbf{y}(t+1) \\ \mathbf{y}(t+2) \\ \vdots \\ \mathbf{y}(T) \end{bmatrix}, \qquad (T = 2t)$$

Involves normalized SVD of the Hankel matrix of dimension $mt \times m(t+1)$,

$$H_{t,t+1} = \mathbb{E}\{\mathbf{y}_{t+1}^{+}(\,\mathbf{y}_{t+1}^{-})^{\top}\} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \dots & \Lambda_t & \Lambda_{t+1} \\ \Lambda_2 & \Lambda_3 & \dots & \Lambda_{t+1} & \Lambda_{t+2} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \Lambda_t & \Lambda_{t+1} & \dots & \Lambda_{T-1} & \Lambda_T \end{bmatrix}$$

By CCA we get a basis at time $t+1$, say $\hat{\mathbf{z}}(t+1)$.
Warning: need the two bases $\hat{\mathbf{z}}(t+1)$ and $\hat{\mathbf{z}}(t)$ to be **coherent**: they should yield **time-invariant parameters** $(A, C, \bar{C})$ independent of time t.
Problem: doing SVD at two different times we may get time-varying realizations.

Bases and Hankel factors

Proposition 8 *Assume the Hankel matrix at some time t has rank n and let*

$$H_t = \Omega_t \, \bar{\Omega}_t^\top$$

be a rank factorization. There is a one-to-one correspondence between rank factorizations of H_t and choices of bases in the finite-interval predictor spaces $\hat{\mathbf{X}}^{+/-}(t)$ and $\hat{\mathbf{X}}^{-/+}(t)$. Given a rank factorization the stochastic n-vectors

$$\hat{\mathbf{x}}(t) := \bar{\Omega}_t^\top (T_t^-)^{-1} \mathbf{y}_t^- , \qquad \hat{\bar{\mathbf{x}}}(t) := \Omega_t^\top (T_t^+)^{-1} \mathbf{y}_t^+ \tag{26}$$

are bases in $\hat{\mathbf{X}}^{+/-}(t)$ and $\hat{\mathbf{X}}^{-/+}(t)$ yielding dual realizations with the same triplet $(A, C, \bar{C})$ uniquely determined by the factorization.
Conversely, given two such dual bases $\hat{\mathbf{x}}(t)$ and $\hat{\bar{\mathbf{x}}}(t)$, there are matrices Ω_t and $\bar{\Omega}_t$ such that

$$\mathbb{E}\,[\mathbf{y}_t^+ \mid \mathbf{Y}_t^-] = \Omega_t \, \hat{\mathbf{x}}(t), \qquad \mathbb{E}\,[\mathbf{y}_t^- \mid \mathbf{Y}_t^+] = \bar{\Omega}_t \, \hat{\bar{\mathbf{x}}}(t), \tag{27}$$

and $H_t = \Omega_t \bar{\Omega}_t^\top$ is a rank n factorization of H_t. The factors Ω_t and $\bar{\Omega}_t$ are the observability and constructibility matrices corresponding to the triplet $(A, C, \bar{C})$ of the two bases.

Coherent Hankel factors

Start with a large Hankel matrix $H_{t+1,t+1}$ of rank $= n$ and let $H_{t+1,t+1} = \Omega_{t+1}\bar{\Omega}_{t+1}^{\top}$ be a rank n factorization. Assume that the matrices Ω_t, $\bar{\Omega}_t$ determined by chopping off the last block of m rows of $\Omega_{t+1}, \bar{\Omega}_{t+1}$:

$$\Omega_t = \uparrow \Omega_{t+1}\,, \qquad \bar{\Omega}_t = \uparrow \bar{\Omega}_{t+1}$$

still have rank n. Then the (shifted) Hankel sub-matrices $H_{t,t+1}$ and $H_{t+1,t}$ obtained by deleting the last block row or the last block column in $H_{t+1,t+1}$ admit unique *coherent* rank n factorizations

$$H_{t+1,t} = \Omega_{t+1}\bar{\Omega}_t^{\top}\,, \qquad H_{t,t+1} = \Omega_t\bar{\Omega}_{t+1}^{\top}$$

where one large factor stays the same. In fact, this follows from

$$\begin{aligned} H_{t+1,t} &= H_{t+1,t+1}\begin{bmatrix} I_{mt} \\ 0 \end{bmatrix} = \Omega_{t+1}\,\bar{\Omega}_{t+1}^{\top}\begin{bmatrix} I_{mt} \\ 0 \end{bmatrix}, \\ H_{t,t+1} &= \begin{bmatrix} I_{mt} & 0 \end{bmatrix} H_{t+1,t+1} = \begin{bmatrix} I_{mt} & 0 \end{bmatrix} \Omega_{t+1}\;\bar{\Omega}_{t+1}^{\top}\,. \end{aligned}$$

Coherent bases from Hankel factors

Theorem 24 *Assume the rank of Ω_t and of $\bar{\Omega}_t$ is still equal to n, then*

$$\hat{\mathbf{x}}(t+1) := \bar{\Omega}_{t+1}^{\top}(T_{t+1}^{-})^{-1}\mathbf{y}_{t+1}^{-}, \qquad \hat{\mathbf{x}}(t) := \bar{\Omega}_t^{\top}(T_t^{-})^{-1}\mathbf{y}_t^{-} \tag{28}$$

are coherent bases in $\hat{\mathbf{X}}^{+/-}(t+1)$ and $\hat{\mathbf{X}}^{+/-}(t)$ and, similarly

$$\hat{\bar{\mathbf{x}}}(t-1) := \Omega_{t-1}^{\top}(T_{t-1}^{+})^{-1}\mathbf{y}_{t-1}^{+}, \qquad \hat{\bar{\mathbf{x}}}(t) := \Omega_t^{\top}(T_t^{+})^{-1}\mathbf{y}_t^{+} \tag{29}$$

are coherent bases in $\hat{\mathbf{X}}^{-/+}(t-1)$ and $\hat{\mathbf{X}}^{-/+}(t)$. Both of which yield the same coefficients $(A,C,\bar{C})$ as (28)*.*

The proof follows from the fact that Ω_{t+1}, $\bar{\Omega}_{t+1}$ have the standard structure of reconstructibility and observability matrices with the same triplet $(A,C,\bar{C})$. Naturally, $\operatorname{rank}\Omega_t = \operatorname{rank}\bar{\Omega}_t = n$ if t is large enough.

The finite-data CCA Algorithm

1. With the data form the stacked tail matrices (we use past and future strings of equal length but the procedure works in general)

$$Y^{-}_{t+1} := \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_0 \end{bmatrix}, \qquad Y^{+}_{t+1} := \begin{bmatrix} Y_{t+1} \\ Y_{t+1} \\ \vdots \\ Y_{2t+1} \end{bmatrix}$$

2. **Form the (approximately) Toeplitz matrices**

$$T^{-}_{t+1} := \frac{1}{N} Y^{-}_{t+1} (Y^{-}_{t+1})^{\top} \qquad T^{+}_{t+1} := \frac{1}{N} Y^{+}_{t+1} (Y^{+}_{t+1})^{\top}$$

3. **Compute (Cholesky) factors**

$$T^{-}_{t+1} = L^{-}_{t+1} (L^{-}_{t+1})^{\top}, \quad T^{+}_{t+1} = L^{+}_{t+1} (L^{+}_{t+1})^{\top}$$

4. **Normalization:**

$$\hat{Y}_{t+1}^{-} := (L_{t+1}^{-})^{-1} Y_{t+1}^{-} \qquad \hat{Y}_{t+1}^{+} := (L_{t+1}^{+})^{-1} Y_{t+1}^{+}$$

5. **SVD + Order estimation:**

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

Choose n so that $\hat{\Sigma}_{t+1} \gg \tilde{\Sigma}_{t+1}$

6. **Rank Factorization**

$$H_{t+1,t+1} = \Omega_{t+1} \bar{\Omega}_{t+1}^{\top} = L_{t+1}^{+} \hat{U}_{t+1} \hat{\Sigma}_{t+1}^{1/2} \left[L_{t+1}^{-} \hat{V}_{t+1} \hat{\Sigma}_{t+1}^{1/2} \right]^{\top}$$

7. **Canonical Variables at time $t+1$**

$$Z_{t+1} := \hat{\Sigma}_{t+1}^{1/2} \hat{V}_{t+1}^{\top} \hat{Y}_{t+1}^{-} = \hat{\Sigma}_{t+1}^{1/2} \hat{V}_{t+1}^{\top} (L_{t+1}^{-})^{\top} Y_{t+1}^{-} := \bar{\Omega}_{t+1}^{\top} (T_{t+1}^{-})^{-1} Y_{t+1}^{-},$$

8. **Compute** $\bar{\Omega}_t$: chop off the last block of $\bar{\Omega}_{t+1}$

$$\bar{\Omega}_t = \uparrow \left[L_{t+1}^{-} \hat{V}_{t+1} \hat{\Sigma}_{t+1}^{1/2} \right] ,$$

9. **Canonical Variables at time** t

$$Z_t := \bar{\Omega}_t^{\top} (T_t^{-})^{-1} Y_t^{-}$$

10. **Solve by Least-Squares to get** A, C

$$\begin{bmatrix} Z_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} Z_t + \begin{bmatrix} K(t) \\ I \end{bmatrix} \hat{E}_t$$

11. **Compute** $\bar{C}$

$$\bar{C} = \frac{1}{N} Y_t Z_{t+1}^{\top} .$$

The partial realization algorithm

Go through the same first 6 steps of the previous CCA algorithm

1. **Form** Ω_{t+1}, $\bar{\Omega}_{t+1}$ in particular let $\Omega_{t+1} = L_{t+1}^{+} \hat{U}_{t+1} \hat{\Sigma}_{t+1}^{1/2}$ and compute Ω_t by chopping off the last block of m rows of Ω_{t+1} so that

$$H_{t,t+1} = \Omega_t \, \bar{\Omega}_{t+1}^{\top}$$

2. **Compute** C, A **by shift-invariance**:

$$C = \begin{bmatrix} I_m & 0 & \dots & 0 \end{bmatrix} \Omega_{t+1} \,, \qquad \downarrow \Omega_{t+1} = \Omega_t \, A$$

3. **Compute** $\bar{C}$

$$\bar{C} = \begin{bmatrix} I_m & 0 & \dots & 0 \end{bmatrix} \bar{\Omega}_{t+1} \,.$$

The two procedures lead to the same formulas for the estimates of $(A, C, \bar{C})$. However numerically are not exactly equivalent.

Estimate the B, D parameters

We have the stationary parameters $(A, C, \bar{C})$ and $\hat{\Lambda}_0 \simeq \Lambda_0$

Solve the Algebraic Riccati Equation

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

To get the minimal (stabilizing) solution P_-

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

Equivalently $B_- = KD_-$

The ARE has a solution iff $(A, C, \bar{C}, \Lambda_0)$ is positive real !. To be addressed later.

Numerical aspects

The LQ factorization in subspace identification

In the algorithms above, many steps involve computations with data matrices with a very large number of columns!

Proposition 9 *Assume that* $U \in \mathbb{R}^{n_1 \times N}$ *and* $Y \in \mathbb{R}^{n_2 \times N}$*. Then there are matrices* Q_1*,* Q_2 *with* $Q_1^\top Q_1 = I_{n_1}$*,* $Q_2^\top Q_2 = I_{n_2}$ *and* $Q_1^\top Q_2 = 0$ *such that*

$$\begin{bmatrix} U \\ Y \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^\top \\ Q_2^\top \end{bmatrix},$$

where and L_{11}*,* L_{22} *are lower triangular.*

The rows of $Q_1^\top$ form an orthonormal basis for the rowspace $\mathbf{U}$; hence

$$\begin{aligned} \hat{\mathbb{E}}\left[Y \mid \mathbf{U}\right] &= YQ_1\left[Q_1^\top Q_1\right]^{-1} Q_1^\top = L_{21}Q_1^\top \\ \hat{\mathbb{E}}\left[Y \mid \mathbf{U}^\perp\right] &= YQ_2\left[Q_2^\top Q_2\right]^{-1} Q_2^\top = L_{22}Q_2^\top . \end{aligned}$$

Using the LQ factorization

In particular, taking $U = Y^{-}_{t+1}$ and $Y = Y^{+}_{t+1}$, one sees that $\hat{Y}^{-}_{t+1} = \sqrt{N}\,Q_1^{\top}$ and

$$T^{-}_{t+1} = \frac{1}{N} L_{11} L_{11}^{\top}, \qquad T^{+}_{t+1} = \frac{1}{N}\{L_{12}L_{12}^{\top} + L_{22}L_{22}^{\top}\},$$

which greatly facilitates the computation of the Cholesky factors. Moreover we have

$$H(t+1,t+1) = \mathbb{E}\, Y^{+}_{t+1}(Y^{-}_{t+1})^{\top} = \mathbb{E}\, Y^{+}_{t+1}(L^{-}_{t+1}\hat{Y}^{-}_{t+1})^{\top} = L_{21}(L^{-}_{t+1})^{\top},$$

which with a minimum amount of computation leads to the normalized Hankel matrix $\hat{H}(t+1,t+1)$ so that the SVD decomposition can be done directly on $(L^{+}_{t+1})^{-1}L_{21}$.

The question of positivity

For generic data there is no guarantee that $(A, C, \bar{C})$ obtained by the above procedures is positive real. The question has nothing to do with sample variability. We shall examine it in a probabilistic setting. Consider two situations

1. The underlying process $\mathbf{y}$ has no finite dimensional realization.

2. The underlying process $\mathbf{y}$ has a finite dimensional realization of very large dimension.

If $\mathbf{y}$ has a realization of dimension n small enough so that $\operatorname{rank} H_t = n$ then the realization procedure recovers a positive real triplet $(A, C, \bar{C})$. This means that the statistical algorithm in this case recovers a a positive real estimated triplet $(A, C, \bar{C})$ asymptotically for $N \to \infty$.

Finite-interval Positivity

Definition 4 *The system* $(A, C, \bar{C}, \Lambda_0)$ *is positive real on the interval* $[0, T]$ *if the Toeplitz matrix*

$$\mathbb{T}_T := \begin{bmatrix} \Lambda_0 & \Lambda_1^\top & \Lambda_2^\top & \dots & \Lambda_T^\top \\ \Lambda_1 & \Lambda_0 & \Lambda_1^\top & \dots & \Lambda_{T-1}^\top \\ \dots & \dots & \ddots & \dots & \dots \\ \Lambda_T & \Lambda_{T-1} & \dots & \Lambda_1 & \Lambda_0 \end{bmatrix}, \qquad \Lambda_k = CA^{k-1}\bar{C}$$

is positive definite.

Consider a stationary process sequence $\{\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(t), \dots, \mathbf{y}(T)\}$ with covariances $\Lambda_k = CA^{k-1}\bar{C}$; $k = 0, 1, \dots, T$. Then $\mathbb{T}_T$ is positive definite.
We have shown that $\mathbf{y}$ can be described on the finite interval $[0, T]$ by a non-stationary state space model

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

For example the transient Kalman filter model.

Finite-interval Positivity cont.d

For any such model

$$\mathbb{E}\begin{bmatrix}\mathbf{x}(t+1)-A\mathbf{x}(t)\\ \mathbf{y}(t)-C\mathbf{x}(t)\end{bmatrix}\begin{bmatrix}\mathbf{x}(t+1)-A\mathbf{x}(t)\\ \mathbf{y}(t)-C\mathbf{x}(t)\end{bmatrix}^\top=\begin{bmatrix}B(t)\\ D(t)\end{bmatrix}\begin{bmatrix}B(t)\\ D(t)\end{bmatrix}^\top\geq 0$$

Then, using the Markov property and $\mathbb{E}\,\mathbf{y}(t)\,\mathbf{x}(t+1)^\top=\bar{C}$ (constant) we see that $P(t):=\mathbb{E}\,\mathbf{x}(t)\mathbf{x}(t)^\top$ satisfies

$$\begin{bmatrix}P(t+1)-AP(t)A^\top & \bar{C}^\top-AP(t)C^\top\\ \bar{C}^\top-CP(t)A^\top & \Lambda_0-CP(t)C^\top\end{bmatrix}=\begin{bmatrix}B(t)\\ D(t)\end{bmatrix}\begin{bmatrix}B(t)^\top & D(t)^\top\end{bmatrix}\geq 0\,.$$

Hence there is a matrix function $P(t)$ (symmetric and positive definite) which satisfies the time varying LMI

$$M(P(t))\geq 0$$

on the interval $[0,\,T]$.

Finite-interval Positive-Realness

Theorem 25 *The system* $(A, C, \bar{C}, \Lambda_0)$ *is positive real on the interval* $[0, T]$*, or equivalently, the finite sequence* $\{\Lambda_0,\ \Lambda_k = CA^{k-1}\bar{C}\,;\, k = 1, \ldots, T\}$ *is positive real, if and only if there exist a symmetric and positive definite matrix function* $P(t)$ *satisfying the time varying Linear Matrix Inequality*

$$M(P(t)) := \begin{bmatrix} P(t+1) - AP(t)A^\top & \bar{C}^\top - AP(t)C^\top \\ \bar{C}^\top - CP(t)A^\top & \Lambda_0 - CP(t)C^\top \end{bmatrix} \geq 0\,.$$

on the interval $[0, T]$*.*

Note: The sequence $\Lambda_k = CA^{k-1}\bar{C}^\top\,;\, k = 1, \ldots$ can be extended as an infinite sequence. However there is no guarantee that if it is positive real on the interval $[0, T]$ it will be still positive real on a larger interval. In particular there is no guarantee that it will be positive real *as an infinite sequence*; i.e. that $C(zI - A)^{-1}\bar{C} + \frac{1}{2}\Lambda_0$ will be a positive real function.

Finite-interval stochastic model reduction

In practice we only operate with finite data. Assume the data are generated by a true system of high dimension n. We pick only the first $k < n$ canonical correlation coefficients at time t and define k-dimensional subvectors

$$\hat{\mathbf{z}}_1(t) = \begin{bmatrix} \sigma_1(t)^{1/2}\mathbf{u}_1(t) \\ \sigma_2(t)^{1/2}\mathbf{u}_2(t) \\ \vdots \\ \sigma_k(t)^{1/2}\mathbf{u}_k(t) \end{bmatrix}, \qquad \hat{\bar{\mathbf{z}}}_1(t) = \begin{bmatrix} \sigma_1(t)^{1/2}\mathbf{v}_1(t) \\ \sigma_2(t)^{1/2}\mathbf{v}_2(t) \\ \vdots \\ \sigma_k(t)^{1/2}\mathbf{v}_k(t) \end{bmatrix}$$

Get a reduced-degree system anyway. half-spectrum . Questions:

1. Is A_1 also stable ?
2. Is $(A_1, C_1, \bar{C}_1)$ a minimal triplet ?
3. Is $(A_1, C_1, \bar{C}_1)$ still positive real on $[0, T]$?
4. Is $(A_1, C_1, \bar{C}_1)$ still in (stochastic) balanced form ?

Third condition is crucial: only if $(A_1, C_1, \bar{C}_1)$ still positive real on $[0, T]$ we have a stochastic model of reduced complexity.

Theorem 26 *If $(A, C, \bar{C}, \frac{1}{2}\Lambda_0)$ is positive real on $[0, T]$ and in stochastic balanced form, then the reduced degree system $(A_1, C_1, \bar{C}_1)$ is still positive real on $[0, T]$.*

in general $(A_1, C_1, \bar{C}_1)$ is not in balanced form. Don't know if A_1 is always stable.

Homework 2

Read Katayama's book pages 227-229. Comment on Remark 8.1 and Remark 8.2.

Order estimation

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

Subspace identification $\simeq$ Estimation by the method of moments.

- **Consistency** If data are generated by a true model.
- **Asymptotic distribution/Variance of** A, C will see general formulas later on.
- **Asymptotic Efficiency** ?? don't know. Depends on the estimates of Λ_k; if they are M.L. then the estimates are also M.L.

Open problems if the generating process has a p.d. component.

Estimation of the Covariance matrix

The covariance matrix

$$\mathbb{T}_T := \begin{bmatrix} \Lambda_0 & \Lambda_1^\top & \Lambda_2^\top & \dots & \Lambda_T^\top \\ \Lambda_1 & \Lambda_0 & \Lambda_1^\top & \dots & \Lambda_{T-1}^\top \\ \dots & \dots & \ddots & \dots & \dots \\ \Lambda_T & \Lambda_{T-1} & \dots & \Lambda_1 & \Lambda_0 \end{bmatrix}, \qquad \Lambda_k = CA^{k-1}\bar{C}$$

must be a block-Toeplitz matrix. The sample estimates like

$$T_t^- := \frac{1}{N} Y_t^- (Y_t^-)^\top \qquad T_t^+ := \frac{1}{N} Y_t^+ (Y_t^+)^\top$$

are only asymptotically Toeplitz. Not efficient. Should impose the structure from the beginning. Sample estimate the Λ_k's and then form the Toeplitz matrices

Generalization to processes with a p.d. component

Will see this at the end.

SUBSPACE IDENTIFICATION
FROM INFINITE/FINITE INPUT-OUTPUT DATA

Outline of the next three lectures:

1. State space models with inputs, decomposition, innovations etc
2. Basic idea of susbspace identification with infinite data
3. State construction for stationary processes. Oblique projections.
4. Infinite data Conditional Canonical Correlation Analysis (CCA). Stochastic Balancing
5. Realization of deterministic systems with infinite/finite data
6. Models with finite data
7. State construction of stochastic systems with inputs from finite data
8. Finite interval realization of stationary stochastic systems with inputs
9. Subspace identification algorithms: CCA, N4SID, MOESP.
10. Consistency and ill-conditioning
11. The asymptotic variance

State space models with input signals

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

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

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. Standing assumptions: $|\lambda(A)| < 1$ (causality), $(A, [B, G])$ reachable and (A, C) observable.

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**.

The Deterministic-Stochastic decomposition

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

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

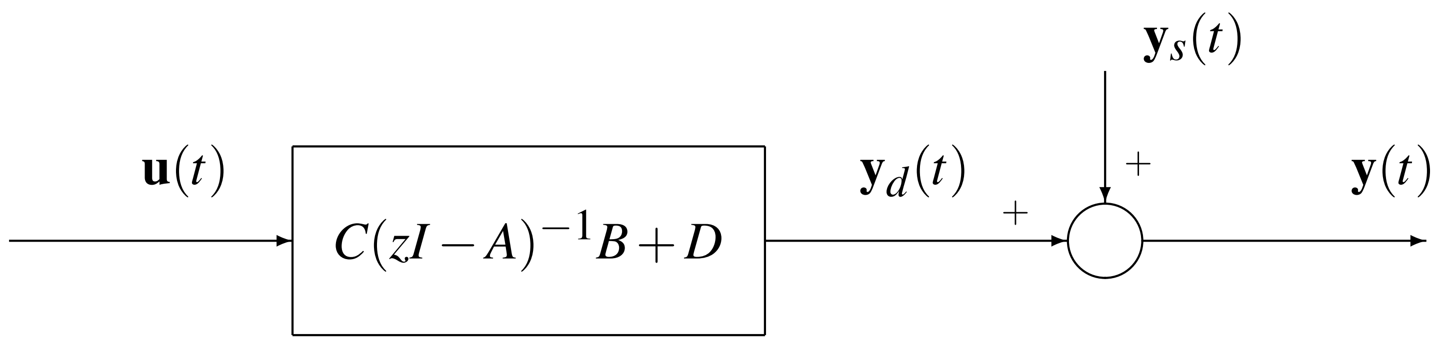
$$\text{Deterministic} \quad \text{Model} \quad \begin{cases} \mathbf{x}_d(t+1) & = & A\mathbf{x}_d(t) + B\mathbf{u}(t) \\ \mathbf{y}_d(t) & = & C\mathbf{x}_d(t) + D\mathbf{u}(t) \end{cases}$$

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

Both models have the same dimension; not true in general.

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

Relation with 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 these are non-minimal realizations.

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

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

where

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

Hence the joint state space model of $\{\mathbf{y}(t)\}$ has an I/O description by the **ARMAX model**

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

This may also be a redundant parametrization.

Deterministic/Stochastic identification

(No feedack!) With infinite data can project the **joint model** on $\mathbf{H}(\mathbf{u})$ and get a **Deterministic Model** for $\mathbf{y}_d(t) = \mathbb{E}\{\mathbf{y}(t) \mid H(\mathbf{u})\}$,

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

Identification of this deterministic system is a **Realization problem** starting from i/o time series data. See Katayama book.

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

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

Nice idea with infinite data but does not work very well with finite data.

The joint innovation model

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 !**

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

Even if there is feedback from $\mathbf{y}$ to $\mathbf{u}$: $\mathbf{e}(t) \perp \hat{\mathbf{x}}(t)\ \ \mathbf{u}(\tau), \quad \forall\, \tau \leq t$

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

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

Consistent subspace Identification problem

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

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

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

such that (**consistency**)

$$\lim_{N\to\infty} \widehat{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_N = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \text{ modulo change of basis}$$

Basic idea of subspace identification 1

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

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

Form the "tail" matrices Y_t, X_t, U_t

$$\begin{aligned} Y_t &:= [\, y_t, y_{t+1}, y_{t+2}, \ldots] \\ X_t &:= [\, x_t, x_{t+1}, x_{t+2}, \ldots] \\ U_t &:= [\, u_t, u_{t+1}, u_{t+2}, \ldots] \end{aligned}$$

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

$$\begin{bmatrix} X_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_t \\ U_t \end{bmatrix} + \begin{bmatrix} K \\ I \end{bmatrix} \mathbf{E}_t$$

Basic idea of subspace identification 2

$$\begin{bmatrix} X_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_t \\ U_t \end{bmatrix} + \begin{bmatrix} K \\ I \end{bmatrix} \mathbf{E}_t$$

Linear Regression ! Solve by Least Squares :

$$\min_{A,C,B,D} \| \begin{bmatrix} X_{t+1} \\ Y_t \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_t \\ U_t \end{bmatrix} \|$$

getting

$$\widehat{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_N := \frac{1}{N} \begin{bmatrix} X_{t+1} \\ Y_t \end{bmatrix} \begin{bmatrix} X_t \\ U_t \end{bmatrix}^\top \left\{ \frac{1}{N} \begin{bmatrix} X_t \\ U_t \end{bmatrix} \begin{bmatrix} X_t \\ U_t \end{bmatrix}^\top \right\}^{-1}$$

Basic idea of subspace identification 3

$$\widehat{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_N := \frac{1}{N} \begin{bmatrix} X_{t+1} \\ Y_t \end{bmatrix} \begin{bmatrix} X_t \\ U_t \end{bmatrix}^\top \left\{ \frac{1}{N} \begin{bmatrix} X_t \\ U_t \end{bmatrix} \begin{bmatrix} X_t \\ U_t \end{bmatrix}^\top \right\}^{-1}$$

Theorem 27 *If the data are* **second order ergodic***, there is no feedback and the inverse exists:*

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

the method provides **consistent estimates** *of* A, B, C, D*.*

Proof: Same as for time-series seen before. Note that the inverse must exist at lest for $N \to \infty$.

Second order ergodicity

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

$$\frac{1}{N}\sum_{k=0}^{N}\{y_{t+k}u_{s+k}^{\top}\} = \frac{1}{N}Y_tU_s^{\top} \to \mathbb{E}\{\mathbf{y}(t)\mathbf{u}(s)^{\top}\} \qquad N \to \infty$$

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

Assuming N "very large" numerical sequences behave like **random variables**: just take sample averages instead of expectations!

$$\mathbf{y}(t) \Leftrightarrow Y_t, \quad \mathbf{u}(t) \Leftrightarrow U_t, \qquad \text{etc.}$$

State sequence construction

Need to **construct the state from input-output data.**

First do when **infinite past data** are (theoretically) available at time t: want just to construct the state of 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 previous procedure by **conditional CCA** .

Geometry: Oblique projections

Consider a static situation. Let $\mathbf{X} = \mathrm{span}\{\mathbf{x}\}, \quad \mathbf{U} = \mathrm{span}\{\mathbf{u}\}$, say finite dimensional random vectors in some Hibert space $\mathbf{H}$.
Assume $\mathbf{X} \cap \mathbf{U} = \{0\}$. Then any $\mathbf{v}$ in $\mathbf{X} + \mathbf{U}$ (direct sum) has a **unique decomposition** into **Oblique projections**:

$$\mathbf{v} = A\mathbf{x} + B\mathbf{u} = \mathbb{E}_{\,||\mathbf{U}}\{\mathbf{v} \mid \mathbf{X}\} + \mathbb{E}_{\,||\mathbf{X}}\{\mathbf{v} \mid \mathbf{U}\}$$

We know how to compute A and B when the sum is orthogonal;
e.g. $A\mathbf{x} = \mathbb{E}\,[\mathbf{v} \mid \mathbf{X}] = \mathbb{E}\left\{\mathbf{v}\mathbf{x}^\top\right\}\mathbb{E}\left\{\mathbf{x}\mathbf{x}^\top\right\}^{-1}\mathbf{x}$.

Theorem 28 *Assume* $\mathbf{X} \cap \mathbf{U} = \{0\}$ *and* $\mathbf{x}$ *and* $\mathbf{u}$ *are bases. Then*

$$A\,\mathbf{x} = \mathbb{E}_{\,||\mathbf{U}}\{\mathbf{v} \mid \mathbf{X}\} = \Sigma_{\mathbf{vx}|\mathbf{u}}\Sigma_{\mathbf{xx}|\mathbf{u}}^{-1}\,\mathbf{x} \qquad B\,\mathbf{u} = \mathbb{E}_{\,||\mathbf{X}}\{\mathbf{v} \mid \mathbf{U}\} = \Sigma_{\mathbf{vu}|\mathbf{x}}\Sigma_{\mathbf{uu}|\mathbf{x}}^{-1}\,\mathbf{u}\,,$$

where $\Sigma_{\mathbf{vx}|\mathbf{u}}, \Sigma_{\mathbf{xx}|\mathbf{u}}, \Sigma_{\mathbf{uu}|\mathbf{x}}$ *are conditional covariances.*

Let $\mathbf{v} \mid \mathbf{u}^\perp := \mathbf{v} - \mathbb{E}\{\mathbf{v} \mid \mathbf{u}\}, \quad \mathbf{x} \mid \mathbf{u}^\perp := \mathbf{x} - \mathbb{E}\{\mathbf{x} \mid \mathbf{u}\}, \qquad \mathbf{u} \mid \mathbf{x}^\perp := \mathbf{u} - \mathbb{E}\{\mathbf{u} \mid \mathbf{x}\}$

$$\Sigma_{\mathbf{vx}|\mathbf{u}} = \mathbb{E}\,(\mathbf{v} \mid \mathbf{u}^\perp)(\mathbf{x} \mid \mathbf{u}^\perp)^\top, \qquad \Sigma_{\mathbf{xx}|\mathbf{u}} = \mathbb{E}\,(\mathbf{x} \mid \mathbf{u}^\perp)(\mathbf{x} \mid \mathbf{u}^\perp)^\top$$

Notations:

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

$$\Sigma_{\mathbf{vx}|\mathbf{u}} := \mathbb{E}\left[\mathbb{E}\left\{\mathbf{v} \mid \mathbf{u}^{\perp}\right\} \mathbb{E}\left\{\mathbf{x} \mid \mathbf{u}^{\perp}\right\}^{\top}\right], \qquad \Sigma_{\mathbf{xx}|\mathbf{u}} := \mathbb{E}\left[\mathbb{E}\left\{\mathbf{x} \mid \mathbf{u}^{\perp}\right\} \mathbb{E}\left\{\mathbf{x} \mid \mathbf{u}^{\perp}\right\}^{\top}\right]$$

Recall: in the Gaussian case both $\mathbf{v} \mid \mathbf{u}^{\perp}$, $\mathbf{x} \mid \mathbf{u}^{\perp}$ *and their product* are independent of $\mathbf{U}$ so the *conditional* covariance, given $\mathbf{u}$ is the same as their (unconditional) covariance.

Proof of the Oblique projection formula

To get $A\mathbf{x}$ do orthogonal projection of $\mathbf{v}$ onto $\mathbf{X}+\mathbf{U}$ and then set $\mathbf{u}=0$

$$\mathbb{E}_{\,||\mathbf{U}}\{\mathbf{v}\,|\,\mathbf{X}\} = \left[\mathbb{E}\left\{\mathbf{v}\mathbf{x}^\top\right\}\ E\left\{\mathbf{v}\mathbf{u}^\top\right\}\right] \begin{bmatrix} \mathbb{E}\left\{\mathbf{x}\mathbf{x}^\top\right\} & \mathbb{E}\left\{\mathbf{x}\mathbf{u}^\top\right\} \\ \mathbb{E}\left\{\mathbf{u}\mathbf{x}^\top\right\} & \mathbb{E}\left\{\mathbf{u}\mathbf{u}^\top\right\} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix}$$

Use matrix inversion lemma

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \Delta^{-1} & -\Delta^{-1}BD^{-1} \\ -D^{-1}C\Delta^{-1} & D^{-1}+D^{-1}C\Delta^{-1}\Delta^{-1}BD^{-1} \end{bmatrix}$$

where $\Delta = A - BD^{-1}C$. See Katayama p. 274. There is also a geometric proof.

RULE: Oblique projection $\mathbb{E}_{\,||\mathbf{U}}\{\mathbf{v}\,|\,\mathbf{X}\}$ is same as orthogonal projection of $\mathbb{E}\left\{\mathbf{v}\,|\,\mathbf{U}^\perp\right\}$ onto $\mathbb{E}\left\{\mathbf{X}\,|\,\mathbf{U}^\perp\right\}$ but with argument $\mathbf{x}$ in place of $\mathbf{x}\,|\,\mathbf{U}^\perp$.

Geometry: Conditionally Markovian subspaces

Standing assumption on the input process. Let

$$\mathbf{U}_t^+ := \text{span}\,\{\,\mathbf{u}(s)\,;\, s \geq t\,\} \qquad\qquad \mathbf{U}_t^- := \text{span}\,\{\,\mathbf{u}(s)\,;\, s < t\,\}$$

Assume:

$$\mathbf{U}_t^+ \cap \mathbf{U}_t^- = \{0\}\,, \quad \forall t \qquad \textbf{sufficient richness}$$

Means that the random variables $\{\mathbf{u}(t)\,; t \in \mathbb{Z}\}$ form a basis for $\mathbf{H}(\mathbf{u})$; $\mathbf{u}$ must be p.n.d. with a coercive spectral density.

Definition 5 *Let* $\mathbf{w}$ *be a white noise with* $\mathbf{w} \perp \mathbf{u}$ *(no feedback). Say that* $\mathbf{x}$ *is a* **Conditionally Markov process** *if*

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + G\mathbf{w}(t)\,, \qquad |\lambda(A)| < 1$$

Let $\mathbf{X}_t := \text{span}\,\{\mathbf{x}_1(t), \mathbf{x}_2(t) \ldots, \mathbf{x}_n(t)\}$ then

$$\mathbb{E}_{\|\mathbf{U}_t^+}\left\{\mathbf{X}_t^+ \mid \mathbf{X}_t^- \vee \mathbf{X}_t\right\} = \mathbb{E}_{\|\mathbf{U}_t^+}\left\{\mathbf{X}_t^+ \mid \mathbf{X}_t\right\}.$$

$\mathbf{X}_t$ is a **Conditionally (or Oblique) Markovian subspace**.

Proof

By causality $\mathbf{x}(t) \in \operatorname{span}\{\mathbf{u}(s), \mathbf{w}(s), ; s < t\}$ so that $\mathbf{X}_t \subset \mathbf{U}_t^- \oplus \mathbf{W}_t^-$. Since $\mathbf{X}_{t+1}^- = \mathbf{X}_t^- \vee \mathbf{X}_t \subset \mathbf{U}_t^- \oplus \mathbf{W}_t^-$

$$\mathbf{X}_{t+1}^- \cap \mathbf{U}_t^+ = \{0\}$$

(because $\mathbf{W} \perp \mathbf{U} \Rightarrow \mathbf{W} \cap \mathbf{U} = \{0\}$). Projecting

$$\mathbf{x}(t+k) = A^k \mathbf{x}(t) + T_B \mathbf{u}_t^+ + T_G \mathbf{w}_t^+ \qquad k > 0$$

onto $\mathbf{X}_{t+1}^- + \mathbf{U}_t^+$ you get $A^k \mathbf{x}(t) + T_B \mathbf{u}_t^+$. Hence

$$\mathbb{E}_{||\mathbf{U}_t^+}\{\mathbf{X}_t^+ \mid \mathbf{X}_t^- \vee \mathbf{X}_t\} = \mathbb{E}_{||\mathbf{U}_t^+}\{\mathbf{X}_t^+ \mid \mathbf{X}_t\}$$

$\square$

Backward notion involving $\mathbb{E}_{||\mathbf{U}_t^-}$ and $\bar{\mathbf{x}}(t-1)$ etc. left as an exercise.

Geometry: Conditionally Markovian subspaces cont.d

Proposition 10 *The oblique Markovian property is equivalent to*

$$\mathbb{E}\{\mathbf{X}_{t+1} \mid \mathbf{X}_{t+1}^{-} \vee \mathbf{U}_{t+1}^{-}\} = \mathbb{E}\{\mathbf{X}_{t+1} \mid \mathbf{X}_t + \mathbf{U}_t\}$$

that is

$$\mathbf{X}_{t+1} \perp \mathbf{X}_{t+1}^{-} \vee \mathbf{U}_{t+1}^{-} \mid \mathbf{X}_t + \mathbf{U}_t$$

Let $\mathbf{x}(t)$, $\mathbf{x}(t+1)$ be stationary bases and let

$$G\mathbf{w}(t) := \mathbf{x}(t+1) - \mathbb{E}\{\mathbf{x}(t+1) \mid \mathbf{X}_{t+1}^{-} \vee \mathbf{U}_{t+1}^{-}\}$$

($\mathbf{w}$ is white noise: normalized joint innovation process of $\mathbf{x}$). Then by the proposition above,

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + G\mathbf{w}(t)\,.$$

Geometry: Oblique Markovian splitting subspaces

Definition 6 . *The subspace* $\mathbf{X}_t$ *is an Oblique Markovian splitting subspace if*

$$\mathbb{E}_{||\mathbf{U}_t^+}\left\{\mathbf{Y}_t^+ \vee \mathbf{X}_t^+ \mid \mathbf{Y}_t^- \vee \mathbf{X}_t^- \vee \mathbf{X}_t\right\} = \mathbb{E}_{||\mathbf{U}_t^+}\left\{\mathbf{Y}_t^+ \vee \mathbf{X}_t^+ \mid \mathbf{X}_t\right\}$$

Theorem 29 *The subspace* $\mathbf{X}_t$ *is a finite dimensional Oblique Markovian splitting subspace iff for any basis* $\mathbf{x}(t)$ *we have a representation*

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

where $|\lambda(A)| < 1$*.*

Same proof as for the Markov case with $\tilde{\mathbf{x}}(t+1) := \begin{bmatrix}\mathbf{x}(t+1) & \mathbf{y}(t)\end{bmatrix}^\top$ in place of $\mathbf{x}(t+1)$. Define the joint innovation as

$$\begin{bmatrix} G \\ J \end{bmatrix} \mathbf{w}(t) := \begin{bmatrix} \mathbf{x}(t+1) \\ \mathbf{y}(t) \end{bmatrix} - \mathbb{E}\left\{ \begin{bmatrix} \mathbf{x}(t+1) \\ \mathbf{y}(t) \end{bmatrix} \mid \mathbf{X}_{t+1}^- \vee \mathbf{Y}_t^- \vee \mathbf{U}_{t+1}^- \right\}$$

$\mathbf{X}_t$ is a “Conditional state space given $\mathbf{U}_t^+$”.

The (forward) oblique predictor space of a stationary process

Assume we have histories starting from $t = -\infty$. Consider state spaces in the joint past

$$\mathbf{X}_t \subset \mathbf{P}_t := \mathbf{Y}_t^- \vee \mathbf{U}_t^- = \operatorname{span}\{\mathbf{y}(s), s < t, \mathbf{u}(s), s < t\}$$

Definition 7 *The (forward) oblique predictor space*

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

Theorem 30 *The (forward) oblique predictor space is the minimal oblique splitting subspace contained in the past. The state space model corresponding to a basis $\hat{\mathbf{x}}(t)$ is the* **stationary innovation model**

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

where $\mathbf{e}(t) = \mathbf{y}(t) - \mathbb{E}\{\mathbf{y}(t) \mid \mathbf{P}_t \vee \mathbf{U}_t\}$.

Verification

Take an observable innovation model

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

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

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

Hence $\quad \mathbb{E}_{\|\mathbf{U}_t^+}\{\mathbf{y}(t+h) \mid \mathbf{P}_t\} = CA^h\hat{\mathbf{x}}(t), \quad h = 0, 1, \ldots \quad$ so that

$$
\mathbb{E}_{\|\mathbf{U}_t^+}\{\mathbf{Y}_t^+ \mid \mathbf{P}_t\} = \mathrm{span}\,\{CA^h\hat{\mathbf{x}}(t)\,;\, h = 0, 1, \ldots\} = \mathbf{X}_t^{+/-}.
$$

QED

Constructing the state of a stationary model

Choose a basis in the **oblique predictor space**. If $\mathbf{y}$ admits a finite dimensional realization, can use finite future spaces with n large enough,

$$\mathbf{X}_t^{+/-} = \text{span}\{\mathbb{E}_{\|\mathbf{U}_t^+}[\mathbf{y}(t+h) \mid \mathbf{P}_t]\,;h=0,1,2,\dots\} = \mathbb{E}_{\|\mathbf{U}_{[t,t+n]}}\{\mathbf{Y}_{[t,t+n]} \mid \mathbf{Y}_t^- \vee \mathbf{U}_t^-\}$$

but for stationarity the past spaces need to be infinite !
Each basis yields a minimal innovation model (Steady state Kalman filter)

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

Once a basis is given, can solve for A, B, C, D and then from the residual vector compute $\Lambda = \text{Var}\{\mathbf{e}(t)\}$ and K. This (in theory !) under certain nonsingularity conditions which will be discussed later.

Stationary conditional CCA

Can compute a basis in $\mathbf{X}_t^{+/-}$ by **conditional CCA**. $\mathbf{X}_t^{+/-}$ is the range space of the oblique projection operator $\mathbb{E}_{\|\mathbf{U}_t^+}[\cdot \mid \mathbf{P}_t]$ restricted to $\mathbf{Y}_t^+$.
Pick a basis in $\mathbf{P}_t$:

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

and future inputs $\mathbf{u}_t^+$ (this may be finite of same length as $\mathbf{y}_t^+$). Recall:
oblique projection is like orthogonal projection of $\mathbf{y}_t^+ \mid (\mathbf{u}_t^+)^\perp$ onto $\mathbf{p}_t \mid (\mathbf{u}_t^+)^\perp$.

Cross covariance is the **Conditional Hankel Matrix**

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

Assume this has finite rank $n \Rightarrow \mathbf{y}_t^+$ and $\mathbf{u}_t^+$ can be taken to be finite dimensional vectors.
Do SVD of the normalized conditional covariance matrix.

Conditional CCA Algorithm

Cholesky factors

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

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

Do SVD of the *normalized conditional Hankel matrix*

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

Rank n SVD (or 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 vector

$$\mathbf{z}(t) = \hat{\Sigma}^{1/2} \hat{V}^\top L_-^{-1}\, \mathbf{p}(t)$$

From unnormalized Hankel factorization get the extended observability matrix

$$\Omega_t = L_+ \hat{U} \hat{\Sigma}^{1/2}$$

Then from (27)

$$\mathbb{E}_{\,\|\mathbf{U}_t^+} [\, \mathbf{y}_t^+ \mid \mathbf{P}_t \,] = \Omega_t \, \mathbf{z}(t) \,,$$

and

$$\mathbb{E} \, [\, \mathbf{y}_t^+ \mid \mathbf{P}_t \vee \mathbf{U}_t^+ \,] = \mathbb{E}_{\,\|\mathbf{U}_t^+} [\, \mathbf{y}_t^+ \mid \mathbf{P}_t \,] + \mathbb{E}_{\,\|\mathbf{P}_t} [\, \mathbf{y}_t^+ \mid \mathbf{U}_t^+ \,] = \Omega_t \, \mathbf{z}(t) + T_B \mathbf{u}_t^+ \,.$$

From $\Omega_{t\pm1}$ can update the state to get $\mathbf{z}(t \pm 1)$.

ONLY FINITE DATA ARE AVAILABLE!

The **The infinite past** $\mathbf{p}(t)$ spanning $\mathbf{Y}_{(-\infty,t)} \vee \mathbf{U}_{(-\infty,t)}$ is not available !! Will see later that approximate conditional CCA using available **finite past** data yields biased estimates. Bias may be large if the zeros of the true system are close to the unit circle. With real data the approximation leads to **errors (bias)** in the estimate which do not $\to 0$ as $N \to \infty$.

For consistency with finite regression data: Need **finite-interval (non-stationary) stochastic realization**.

Finite-interval innovation models

Must use (random) "data" on a finite-interval $[t_0, T]$ only. Project a stationary model onto $\mathbf{Y}_{[t_0,t)} \vee \mathbf{U}_{[t_0,T]} \equiv \mathbf{P}_{[t_0,t)} \vee \mathbf{U}_{[t,T]}$.
The estimate

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

satisfies the *transient conditional Kalman filter equation*

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

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

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

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

The non-causal component of the transient K.F. state

Project the state of a stationary model (by feedback-free $\mathbf{Y}_{[t_0,t)} \cap \mathbf{U}_{[t,T]} = \{0\}$)

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{x}(t) \mid \mathbf{P}_{[t_0,t)} \vee \mathbf{U}_{[t,T]}\right] = \\
&= \mathbb{E}_{\|\mathbf{U}_{[t,T]}}\left[\mathbf{x}(t) \mid \mathbf{P}_{[t_0,t)}\right] \;+\; \mathbb{E}_{\|\mathbf{P}_{[t_0,t)}}\left[\mathbf{x}(t) \mid \mathbf{U}_{[t,T]}\right] \\
&= \hat{\mathbf{x}}_{causal}(t) \;+\; \mathbf{u}_t^+ - \text{dependent part}
\end{aligned}
$$

If you keep only the oblique projection along $\mathbf{U}_{[t,T]}$ you miss one piece of the transient state!

The stationary conditional CCA applied to **finite past data** gives a causal basis, approximation of the transient Kalman state $\hat{\mathbf{x}}(t)$ with error depending on the initial condition.

On $[t_0, t]$ the *steady state* Kalman filter realization obeys

$$
\mathbf{x}(t+1) = (A-KC)\mathbf{x}(t) + (B-KD)\mathbf{u}(t) + K\mathbf{y}(t)\,, \qquad \mathbf{x}(t_0) = \mathbf{x}_0 \in \mathbf{P}_{t_0}
$$

so that

$$
\mathbb{E}_{\|\mathbf{P}_{[t_0,t)}}\{\mathbf{x}(t) \mid \mathbf{U}_{[t,T)}\} = (A-KC)^{t-t_0}\,\mathbb{E}_{\|\mathbf{P}_{[t_0,t)}}\{\mathbf{x}(t_0) \mid \mathbf{U}_{[t,T)}\}\,.
$$

Error analysis of conditional CCA with finite past data

The stationary conditional CCA procedure applied to **finite past data** provides an approximation of the transient Kalman state vector $\hat{\mathbf{x}}(t)$ with error depending on the initial condition.
The term

$$\mathbb{E}_{\|\mathbf{P}_{[t_0,t)}}\{\mathbf{x}(t) \mid \mathbf{U}_{[t,T)}\} = (A-KC)^{t-t_0}\,\mathbb{E}_{\|\mathbf{P}_{[t_0,t)}}\{\mathbf{x}(t_0) \mid \mathbf{U}_{[t,T)}\},$$

tends to zero when $t-t_0 \to \infty$, if $|\lambda(A-KC)| < 1$; i.e. the (true) system has no zeros on the unit circle.
In our case $t-t_0$ is **fixed** and often quite small. If there are zeros close to the unit circle, the stationary conditional CCA procedure applied to finite data, $\mathbf{y}_t^+$ and $\mathbf{p}_t \in \mathbf{P}_{[t_0,t)}$ may produce a state differing considerably (for $t-t_0$ finite) from the transient K.F. state $\hat{\mathbf{x}}(t)$.

A basic step of finite-interval subspace methods

Don't know how to construct transient Kalman filter state $\hat{\mathbf{x}}(t)$ from finite data. Conditional CCA may yield biased estimates. Need a different route. Project future outputs of a stationary model (state vector $\mathbf{x}(t)$) on the finite data space, (future horizon: $k = T - t$)

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

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

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

Want to project on finite data spaces so as to kill the last two pieces. In this way get a term which looks like $\Omega_k \mathbf{x}(t)$. From this can recover the column space of Ω_k.

Recovering Ω_k

Compute the output predictor based on all available data at time t

$$\hat{\mathbf{y}}_t^+ := \mathbb{E}\left\{ \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t+1) \\ \vdots \\ \mathbf{y}(t+k) \end{bmatrix} \mid \mathbf{P}_{[t_0,t)} + \mathbf{U}_{[t,T]} \right\} = \Omega_k\, \hat{\mathbf{x}}(t) + H_d\, \mathbf{u}_t^+$$

To kill the $\mathbf{u}_t^+$-dependent term, take oblique projection along $\mathbf{U}_{[t,T]}$:

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

where $\tilde{\mathbf{x}}(t)$ is the **causal component** of the K.F. state : $\tilde{\mathbf{x}}(t) := \mathbb{E}_{\|\mathbf{U}_{[t,T]}}\left[\hat{\mathbf{x}}(t) \mid \mathbf{P}_{[t_0,t)} \right]$
Hence if

$$\mathbb{E}\{\tilde{\mathbf{x}}(t)\,\tilde{\mathbf{x}}(t)^\top\} > 0 \qquad \text{(consistency condition)},$$

we can recover the column space of Ω_k from the oblique projection along $\mathbf{U}_{[t,T]}$ of the finite data predictors $\hat{\mathbf{y}}_t^+$.
This idea is used in the N4SID algorithm.

THE VAN OVERSCHEE-DE MOOR MODEL

The **Pseudostate** : $\bar{\mathbf{x}}(t) := \Omega_k^{-L}\hat{\mathbf{y}}_t^+ = \hat{\mathbf{x}}(t) + \Omega_k^{-L} H_d \mathbf{u}_t^+$ substitute $\hat{\mathbf{x}}(t)$ into the K.F. equation to get the linear recursion

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

$\mathcal{K}_1\ \mathcal{K}_2$ are known (complicated) linear functions of (A, C) and (B, D). Solve (*) for the unknown parameters in terms of the data $\bar{\mathbf{x}}(t)$ and $\mathbf{u}_t^+$ using the oblique projection formulas

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

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

Conditional Covariances (Notation: $\bar{\mathbf{x}}_1 \equiv \bar{\mathbf{x}}(t+1)$) :

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

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

The N4SID algorithm

[vanOverschee-DeMoor-94]

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 $U_{[t,T]}$

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

to get an estimate of $\Omega_k \tilde{X}_t$ (causal part of $\hat{X}_t$)

3. Estimate the order and the observability matrix Ω_k by SVD factorization keeping in Σ only the "relevant" singular values

$$\hat{Z}_{[t,T-1]} = U\Sigma V^\top \qquad \Omega_k := U\Sigma^{1/2}, \qquad \tilde{X}_t = \Sigma^{1/2}V^\top$$

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

$$\begin{bmatrix} \bar{X}_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \bar{X}_t + \begin{bmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{bmatrix} U_{[t,T]} \overset{\perp}{+} \begin{bmatrix} K(t) \\ I \end{bmatrix} \hat{E}_t$$

5. Need to compute a coherent pseudostate at time $t+1$:

$$\bar{X}_{t+1} := \Omega_k^{-L}\hat{Y}_{[t+1,T]}$$

(slight variation: use all future data $\hat{Y}_{[t,T]}$ to compute $\bar{X}_t$ and $\hat{Y}_{[t+1,T]}$ to compute $\bar{X}_{t+1}$)

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

7. Estimate (B,D) inverting the functions $(\mathcal{K}_1,\mathcal{K}_2)$ (see the book by van Ovrschee and De Moor).

The “Robust” N4SID algorithm

The crucial step in N4SID is the computation of $\hat{Z}_{[t,T-1]}$ by oblique projection along $\mathbf{U}_{[t,T]}$. Oblique projection may be ill-conditioned and poor estimates of Ω may result. Try to use only orthogonal projections !
Orthogonal decomposition of the data space at time t

$$\mathbf{P}_{[t_0,t)} + \mathbf{U}_{[t,T]} = \mathbf{P}_{[t_0,t)} \oplus \mathbb{E}\left\{\mathbf{P}_{[t_0,t)} \mid \mathbf{U}^{\perp}_{[t,T]}\right\}$$

where

$$\mathbb{E}\left\{\mathbf{P}_{[t_0,t)} \mid \mathbf{U}^{\perp}_{[t,T]}\right\} = \operatorname{span}\left\{\mathbf{p}(t) - \mathbb{E}\left\{\mathbf{p}(t) \mid \mathbf{u}_t^+\right\};\, \mathbf{p}(t) \in \mathbf{P}_{[t_0,t)}\right\}$$

(same general idea used earlier to compute oblique projections by orthogonal projections). Notation:

$$\mathbf{U}^{\perp}_{[t,T]} \equiv \mathbb{E}\left\{\mathbf{P}_{[t_0,t)} \mid \mathbf{U}^{\perp}_{[t,T]}\right\}$$

orthogonal complement of $\mathbf{U}_{[t,T]}$ in the ambient space $\mathbf{P}_{[t_0,t)} + \mathbf{U}_{[t,T]}$ (which varies with t). Compute instead

$$\hat{\mathbf{z}}_t^+ := \mathbb{E}\left\{\hat{\mathbf{y}}_t^+ \mid \mathbf{U}^{\perp}_{[t,T]}\right\} = \Omega_k\, \mathbb{E}\left\{\tilde{\mathbf{x}}(t) \mid \mathbf{U}^{\perp}_{[t,T]}\right\} := \Omega_k\, \hat{\mathbf{x}}_c(t).$$

The Ω_k is in a different basis (keep the same symbol for simplicity). See next slide for comments on $\hat{\mathbf{x}}_c(t)$.

Remark about Robust N4SID algorithm

Note: from the oblique decomposition of $\hat{\mathbf{x}}(t)$:

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

define

$$\hat{\mathbf{x}}_c(t) := \mathbb{E}\left\{\hat{\mathbf{x}}(t) \mid \mathbf{U}^{\perp}_{[t,T]}\right\} = \mathbb{E}\left\{\tilde{\mathbf{x}}(t) \mid \mathbf{U}^{\perp}_{[t,T]}\right\}.$$

Note that $\hat{\mathbf{x}}_c(t)$ is not equal to $\tilde{\mathbf{x}}(t)$ which $\in \mathbf{P}_{[t_0,t)}$, being an **oblique** projection.

While $\tilde{\mathbf{x}}(t)$ satisfies a transient Kalman Filter recursion with zero initial conditions there is no simple recursion for $\hat{\mathbf{x}}_c(t)$.

" MOESP" [Verhaegen 93-94]

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

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

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

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

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

3. Factorize $\hat{\mathbf{z}}_t^+$, i.e. the matrix $Z_{[t,T]}^c := \mathbb{E}\left[\hat{Y}_{[t,T]} \mid \mathbf{U}_{[t,T]}^{\perp}\right]$ by SVD

$$Z_{[t,T]}^c = \begin{bmatrix}\hat{U} & \tilde{U}\end{bmatrix} \begin{bmatrix}\hat{\Sigma} & 0 \\ 0 & \tilde{\Sigma}\end{bmatrix} \begin{bmatrix}\hat{V}^\top \\ \tilde{V}^\top\end{bmatrix} \simeq \hat{U}\hat{\Sigma}^{1/2}\,\hat{\Sigma}^{1/2}\hat{V}^\top$$

to get an estimate of the order n and of Ω_k e.g. $\hat{\Omega}_k = \hat{U}\hat{\Sigma}^{1/2}$.

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

5. Construct a matrix $\hat{\Omega}_k^{\perp}$ such that $\hat{\Omega}_k^{\perp}\hat{\Omega}_k = 0$. For example $\tilde{U}^{\top}$.

6. Compute

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

7. Once (A,C) are estimated, H_d is a linear function of (B,D) : can rewrite this as

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

Estimate (B,D) by linear regression.

Efficient implementation via L-Q factorization. See e.g. Katayama's book p. 157-159.

MOESP $\equiv$ "robust" N4SID with orthogonalized regressors

"Robust" N4SID computes $\hat{\Omega}_k$ as in MOESP by SVD of the orthogonal projection $Z^c_{[t,T]} = \mathbb{E}\left[Y_{[t,T]} \mid U^{\perp}_{[t,T]}\right]$. Then use the same definition of $\bar{X}_t$ so that after redefining $\mathcal{K}_1, \mathcal{K}_2$, get the same recursion

$$\begin{bmatrix} \bar{X}_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \bar{X}_t + \begin{bmatrix} \mathcal{K}^c_1 \\ \mathcal{K}^c_2 \end{bmatrix} U_{[t,T]} \stackrel{\perp}{+} \begin{bmatrix} K(t) \\ I \end{bmatrix} \hat{E}_t$$

Solve by least squares: the parameters can be gotten by the same oblique projection formula. Note

$$\Sigma_{\bar{\mathbf{x}}\bar{\mathbf{x}}|\mathbf{u}^+} = \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} = \Sigma_{\hat{\mathbf{x}}_c,\hat{\mathbf{x}}_c} \leftarrow \mathbb{E}\,\hat{X}^c_t(\hat{X}^c_t)^\top .$$

Can show that you get exactly the same estimates of (A, C). (see Chiuso-Picci 2004).

Consistency and conditioning

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

$$\begin{bmatrix} A \\ C \end{bmatrix} \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} = \begin{bmatrix} \Sigma_{\bar{\mathbf{x}}_1\bar{\mathbf{x}}|\mathbf{u}^+} \\ \Sigma_{\bar{\mathbf{y}}\bar{\mathbf{x}}|\mathbf{u}^+} \end{bmatrix} \qquad \begin{bmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{bmatrix} \Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}} = \begin{bmatrix} \Sigma_{\bar{\mathbf{x}}_1\mathbf{u}^+|\bar{\mathbf{x}}} \\ \Sigma_{\mathbf{y}\mathbf{u}^+|\bar{\mathbf{x}}} \end{bmatrix}$$

The **Conditional Covariances**:

$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ and $\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}$ must be non-singular !

(First is the Jansson-Wahlberg consistency condition). More generally: $\Sigma_{\hat{\mathbf{x}}_c\hat{\mathbf{x}}_c}$ $(= \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+})$ may be ill-conditioned $\Rightarrow$ computation of the parameters (A, C) of the regression will be **ill-conditioned**: **random fluctuation errors in the sample covariance data are amplified**.

$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ ill-conditioned $\Leftrightarrow$ subspaces $\hat{\mathbf{X}}_t$ and $\mathbf{U}_{[t,T]}$ are “nearly parallel”.

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

Conditioning of subspace identification

The conditioning of the problem (*) is determined by the singular values of the conditional covariances $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ and $\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}$. Let

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

Normalized Cross-Covariance Matrices

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

The singular values of $\hat{\Pi}$ are cosines of the **canonical angles** between $\hat{\mathbf{X}}_t$ and $\mathbf{U}_{[t,T]}$. From well-known expressions of the error covariances follows that

$$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} = \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}} - \Sigma_{\hat{\mathbf{x}}\mathbf{u}^+}\Sigma_{\mathbf{u}^+\mathbf{u}^+}^{-1}\Sigma_{\hat{\mathbf{x}}\mathbf{u}^+}^\top = L_{\hat{\mathbf{x}}}\left[I - \hat{\Pi}^\top\hat{\Pi}\right]L_{\hat{\mathbf{x}}}^\top$$

Condition numbers

Recall $\kappa(A) := \sigma_{max}(A)/\sigma_{min}(A)$ is the relative error amplification coefficient in the solution of $Ax = b$.

Theorem 31

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

$$\kappa\left(\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}\right) \le \kappa\left(\Lambda_{\mathbf{u}}\right) \frac{1}{1-\sigma^2_{max}(\hat{\hat{\Pi}})}$$

The bounds are sharp.

If $\sigma_{\max}\{\hat{\mathbf{X}}_t, \mathbf{U}_{[t,T]}\} \simeq 1 \Rightarrow \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ **very ill conditioned !**
$\kappa(\Lambda_{\mathbf{u}}) \simeq \max_{\omega}\Phi_{\mathbf{u}}(\omega)/\min_{\omega}\Phi_{\mathbf{u}}(\omega)$. If $\mathbf{u}$ is white noise $\kappa(\Lambda_u) = 1$.
If $\Phi_{\mathbf{u}}(\omega) \simeq 0$, locally (poor excitation), then $\kappa(\Lambda_{\mathbf{u}})$ large.

Conditioning of subspace identification cont.d

In the stationary setting $\hat{\mathbf{x}}_d(t)$ is uncorrelated with $\hat{\mathbf{x}}_s(t)$ (by feedback-free). Hence

$$\Pi := E\left[\mathbf{u}_t^+ \hat{\mathbf{x}}(t)^\top\right] = E\left[\mathbf{u}_t^+ \hat{\mathbf{x}}_d(t)^\top\right] := \Pi_d$$

- 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 $\bar{\Pi} \equiv E\left[\mathbf{u}_t^+ \bar{\hat{\mathbf{x}}}(t)^\top\right]$ 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

Conditioning in the stationary setting

Asymptotics for $T - t_0 \to \infty$, everything stationary.
Given an input with assigned spectrum $\Phi_{\mathbf{u}}$. Which deterministic systems $F(z) = C(zI - A)^{-1}B + D$ have the SMALLEST canonical angles of (the spaces spanned by) $\mathbf{u}_t^+$ and $\mathbf{x}_d(t)$ (worst conditioning of the identification problem) ??

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

$$\mathbf{U}^+ \qquad \text{and} \qquad \mathbf{X}_d := \text{span}\,\{\mathbf{x}_d(0)\} \subset \mathbf{U}^-$$

Theorem 32

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

Maximal when

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

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

Probing inputs

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.

Estimation of the stochastic parameters 1

Recall the Pseudostate in the (robust) Van Overschee-De Moor model $\bar{\mathbf{x}}(t) := \Omega_k^{-L}\hat{\mathbf{y}}_t^+ = \hat{\mathbf{x}}(t) + \Omega_k^{-L} H_d \mathbf{u}_t^+)$ substitute the expression of $\hat{\mathbf{x}}(t)$ into the transient Kalman Filter to get the linear recursion

$$\begin{bmatrix} \bar{\mathbf{x}}(t+1) \\ \mathbf{y}(t) \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \bar{\mathbf{x}}(t) + \begin{bmatrix} \mathcal{K}_1^c \\ \mathcal{K}_2^c \end{bmatrix} \mathbf{u}_t^+ + \begin{bmatrix} K(t) \\ I \end{bmatrix} \hat{\mathbf{e}}(t)$$

The residues of the Least Squares estimates

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

say

$$\widehat{\begin{bmatrix} \hat{K}(t) \\ I \end{bmatrix}}_N \hat{E}_t := \begin{bmatrix} \bar{X}_{t+1} \\ Y_t \end{bmatrix} - \widehat{\begin{bmatrix} A & \mathcal{K}_1^c \\ C & \mathcal{K}_2^c \end{bmatrix}}_N \begin{bmatrix} \bar{X}_t \\ U_{[t,T]} \end{bmatrix}$$

for $N \to \infty$

$$\frac{1}{N} \widehat{\begin{bmatrix} \hat{K}(t) \\ I \end{bmatrix}}_N \hat{E}_t \hat{E}_t^\top \widehat{\begin{bmatrix} \hat{K}(t) \\ I \end{bmatrix}}_N^\top \to \begin{bmatrix} \hat{K}(t) \\ I \end{bmatrix} \hat{\Lambda}(t) \begin{bmatrix} \hat{K}(t) \\ I \end{bmatrix}^\top = \begin{bmatrix} Q(t) & S(t) \\ S(t)^\top & \Lambda(t) \end{bmatrix}$$

A warning

In the literature people often say that since they are obtained as residual variance, i.e. by construction

$$\begin{bmatrix} Q(t) & S(t) \\ S(t)^\top & \Lambda(t) \end{bmatrix} \geq 0 \qquad (\dagger)$$

the matrices $Q(t), S(t), \Lambda(t)$ can be taken as estimates of the steady state noise parameters Q, S, R.

However we can prove that the matrix $(\dagger)$ will converge to a positive definite limit when $t \to \infty$, only if we assume that the data are generated by a true system of order n.

For generic data, although $(\dagger)$ holds for a finite t, there is no guarantee that the matrix will converge to anything nor that the limit will be positive definite.

Estimating $P(t)$

Recall: the causal component of the K.F. state $\tilde{\mathbf{x}}(t) := \Omega_k^{-L}\hat{\mathbf{z}}_t^+$ satisfies the Kalman recursion with zero initial conditions.
Recall $\hat{Z}_{[t,T]}$ is the *oblique projection* of $\hat{Y}_{[t,T]}$ onto the joint past space. From SVD

$$\hat{Z}_{[t,T]} = U\Sigma^{1/2}\,\Sigma^{1/2}V^\top = \Omega_k\,\tilde{X}_t$$

get

$$\tilde{X}_t = \Sigma^{-1/2}V^\top$$

After having estimated B, D can subtract from $\tilde{\mathbf{x}}(t)$ the $\mathbf{u}$-dependent mean, $\mu(t) = \sum_{i=0}^{t-1} A^{k-1-i}Bu(t_0 + i)$. Compute:

$$\tilde{X}_t - \Gamma_k U_{[t_0,t)}\,, \qquad \Gamma_k = \begin{bmatrix} A^{k-1}B & \dots & B \end{bmatrix}$$

from which estimate the state covariance $P(t) = \mathbb{E}\,(\tilde{\mathbf{x}}(t)-\mu(t))(\tilde{\mathbf{x}}(t)-\mu(t))^\top$ as

$$\hat{P}(t) = \frac{1}{N}\left[\tilde{X}_t - \Gamma_k U_{[t_0,t)}\right]\left[\tilde{X}_t - \Gamma_k U_{[t_0,t)}\right]^\top.$$

Estimating $P(t)$; Alternative route

From SVD of $\hat{Z}^c_{[t,T]}$ (orth. projection) obtain $\Omega_k^{-L} = \Sigma^{-1/2} U^\top$; can compute

$$\hat{X}_t = \Omega_k^{-L} \left(\hat{Y}_{[t,T]} - H_d U_{[t,T]} \right)$$

(this follows from $\hat{\mathbf{x}}(t) = \Omega_k^{-L} \left(\hat{\mathbf{y}}_t^+ - H_d \mathbf{u}_t^+ \right)$)
After having estimated B, D can subtract from $\hat{\mathbf{x}}(t)$ the $\mathbf{u}$-dependent mean, $\mu(t) = \sum_{i=0}^{t-1} A^{k-1-i} B u(t_0 + i)$. To this end compute:

$$\hat{X}_t - \Gamma_k U_{[t_0,t)}, \qquad \Gamma_k = \begin{bmatrix} A^{k-1}B & \dots & B \end{bmatrix}$$

This state covariance $P(t) = \mathbb{E}\,(\hat{\mathbf{x}}(t) - \mu(t))(\hat{\mathbf{x}}(t) - \mu(t))^\top$ satisfies the same Riccati difference equation but with nonzero initial condition:

$$P(t_0) = \mathbb{E}\left\{ \mathbb{E}\left[\mathbf{x}(t_0) \mid \mathbf{U}_{[t_0,T]}\right] \mathbb{E}\left[\mathbf{x}(t_0) \mid \mathbf{U}_{[t_0,T]}\right]^\top \right\}$$

can also be estimated as

$$\hat{P}(t) = \frac{1}{N} \left[\hat{X}_t - \Gamma_k U_{[t_0,t)} \right] \left[\hat{X}_t - \Gamma_k U_{[t_0,t)} \right]^\top .$$

Estimation of the stochastic parameters 2

Problem 1 *Assume $\mathbf{y}$ is described by a (minimal) stationary model of order n, (may take in innovation form)*

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

and that we know $(A, C B, D)$ and, for some fixed $t > t_0$, the transient Kalman filter parameters

$$\begin{bmatrix} K(t) \\ I \end{bmatrix} \Lambda(t) \begin{bmatrix} K(t) \\ I \end{bmatrix}^\top = \begin{bmatrix} Q(t) & S(t) \\ S(t)^\top & \Lambda(t) \end{bmatrix},$$

and the state covariance $P(t)$.
Can we compute the steady state Kalman gain and steady state innovation variance of the model?

Getting the stochastic steady state parameters

Solution: Since $K(t) = (\bar{C}^\top - AP(t)C^\top)\Lambda(t)^{-1}$ we can get

$$\bar{C}^\top = K(t)\Lambda(t) + AP(t)C^\top$$

and consider the Riccati Difference equation on $s \geq 0$,

$$P(s+1) = AP(s)A^\top + (\bar{C}^\top - AP(s)C^\top)\left(\Lambda_0 - CP(s)C^\top\right)^{-1}(\bar{C}^\top - AP(s)C^\top)^\top$$

with initial condition $P(0) = P(t)$.
If $(A, C, \bar{C}, \Lambda_0)$ is positive real and $P(t)$ is a (true) transient state covariance then

$$\lim_{s\to+\infty} P(s) = P_- .$$

From this compute K, Λ. Naturally this applies only for $N \to \infty$.

Homework n. 3

The key step in estimation of the stochastic parameters is to get the $\bar{C}$ matrix (in the right basis). Once you have estimates of $A, C, \bar{C}$ (and of Λ_0 of course) it is immediate to find the steady state innovation model parameters K, Λ by solving the ARE (in case it is solvable).

Describe how you would setup a “backward” subspace algorithm to identify the $A^\top, \bar{C}$ parameters of the dual backward transient Kalman filter model, based on finite data $\mathbf{U}_{[t_0,T]} \vee \mathbf{Y}_{[t,T]}$.

Numerical aspects

Cannot be discussed here. Based on extensive use of L-Q factorization. See the reference list.

Katayama (2011) shows how to get the stochastic parameters in the MOESP algorithm via L-Q factorization.

Software: there is a version of N4SID in Ljung's system identification toolbox. Don't know the actual algorithm used in here.
Also there is a MATLAB MOESP-based toolbox.
Better software written by A. Chiuso. Ask him if you need it.

Asymptotic variance of A, C

Theorem 33 *Assume consistency. 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,*

$$\begin{aligned}
\lim_{N\to\infty} N\,\mathbb{E}\left\{\mathrm{vec}\left(\tilde{A}_N\right)\mathrm{vec}\left(\tilde{A}_N\right)^\top\right\} &= \left\{\Sigma_{\hat{\mathbf{x}}_c\hat{\mathbf{x}}_c}^{-1}\otimes[MH_s]\right\}\cdot \\
&\cdot \sum_{|\tau|\le k}\Sigma_{\hat{\mathbf{x}}_c\hat{\mathbf{x}}_c}(\tau)\otimes\Sigma_{\bar{\mathbf{e}}^+\bar{\mathbf{e}}^+}(\tau)\cdot\left\{\Sigma_{\hat{\mathbf{x}}_c\hat{\mathbf{x}}_c}^{-1}\otimes[MH_s]\right\}^\top \\
\lim_{N\to\infty} N\,\mathbb{E}\left\{\mathrm{vec}\left(\tilde{C}_N\right)\mathrm{vec}\left(\tilde{C}_N\right)^\top\right\} &= \left\{\Sigma_{\hat{\mathbf{x}}_c\hat{\mathbf{x}}_c}^{-1}\otimes[RH_s]\right\}\cdot \\
&\cdot \sum_{|\tau|<k}\Sigma_{\hat{\mathbf{x}}_c\hat{\mathbf{x}}_c}(\tau)\otimes\Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau)\cdot\left\{\Sigma_{\hat{\mathbf{x}}_c\hat{\mathbf{x}}_c}^{-1}\otimes[RH_s]\right\}^\top
\end{aligned}$$

NOTATIONS

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

Ω the observability matrix in a certain basis.

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

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

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

Formula 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.

Application: asymptotic variance of eigenvalues

Assume for simplicity that A has simple eigenvalues.

There 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

$$\hat{\lambda}_N^i - \lambda^i \simeq \frac{\mathbf{v}_i^\top \tilde{A}_N \mathbf{u}_i}{\mathbf{v}_i^\top \mathbf{u}_i} + O(\| \tilde{A}_N \|^2)$$

where $\mathbf{v}_i$ and $\mathbf{u}_i$ are the normalized left and right eigenvectors of A correspoding to λ^i.

$$N \mathbb{E} (\hat{\lambda}_N^i - \lambda^i)^2 = \frac{1}{(\mathbf{v}_i^\top \mathbf{u}_i)^2} (\mathbf{u}_i^\top \otimes \mathbf{v}_i^\top) N \mathbb{E} \left\{ \mathrm{vec}\left(\tilde{A}_N\right) \mathrm{vec}\left(\tilde{A}_N\right)^\top \right\} (\mathbf{u}_i \otimes \mathbf{v}_i)$$

Note that $(\mathbf{v}_i^\top \mathbf{u}_i)^2$ 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 $\mathbf{v}_i = \mathbf{u}_i$).

Asymptotioc 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

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

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

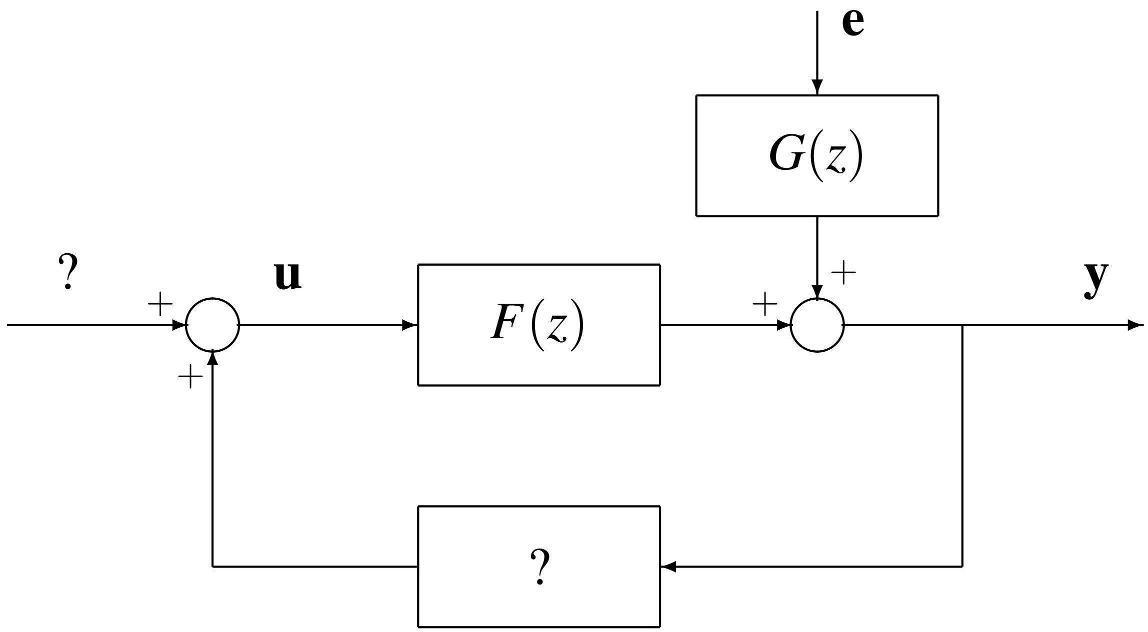
R and M being as before, and,

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

$\tilde{\mathbf{u}}^+_{t+\tau}$ 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

Preliminary material on feedback between stationary processes in Chap 7 of my book in Italian.



+ $F(\infty) = 0$.

PROBLEMS WITH STATE CONSTRUCTION

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

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

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

Needs $\mathbf{E}_t^+ \perp U_t^+$ 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**

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

$$\hat{\mathbf{y}}(t+h \mid t) = C(A-KC)^h\mathbf{x}(t) + \text{``}terms \;\; in \;\; \mathbf{U}_t^+ \vee Y_t^+\text{''}$$

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

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

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

2. Compute $\hat{X}_t^{+/-}$ 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 $([\mathbf{y}_t, \mathbf{y}_{t+1}, .., \mathbf{y}_{t+N}])$ $N \to \infty$, 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. A. Chiuso, G. Picci, "Geometry of Oblique Splitting, Minimality and Hankel Operators", in *Directions in Mathematical Systems Theory and Optimization*, A. Rantzer and C. Byrnes eds. Springer Lect. Notes in Control and information Sciences, **286**, pp. 85-124 (2003).

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

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

8. 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.

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

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

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

12. T. Katayama, *Subspace methods for system identification*, Springer Communication and Control Engineering series, 2005.

13. T. Katayama, A note on LQ decomposition in stochastic subspace identification, in *Perspectives in Mathematical System Theory, Control, and Signal Processing*, Willems, J.C.; Hara, S.; Ohta, Y.; Fujioka, H. (Eds.); Springer Lecture Notes in Control and Information Sciences, Vol. 398, 2010.

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

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

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

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

18. P. van Overschee and B. De Moor, *Subspace Identification for Linear Systems*, (1996), free download from `http://ebookee.org/dl/` .

19. 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.

20. M. Verhaegen, Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, **30**: pp. 61-74, 1994.

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

22. 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.