AN IDENTIFICATION ALGORITHM FOR ARMAX SYSTEMS

First the X, then the AR, finally the MA

Jan C. Willems, K.U. Leuven

Joint work with Ivan Markovsky (K.U. Leuven)
and Paolo Rapisarda (Un. Maastricht)

General introduction

SYSTEM ID

Observed data $\longmapsto$ **System model**

Observed data $\mapsto$ System model

Case on interest: Data = a time-series record:

$$\tilde{w}(1), \tilde{w}(2), \ldots, \tilde{w}(T) \qquad w(t) \in \mathbb{R}^{\mathtt{w}}$$

Required: an algorithm to obtain a dynamical system that 'explains' this time-series.

SYSTEM ID

Observed data $\mapsto$ System model

Case on interest: <u>Data</u> = a time-series record:

$$\tilde{w}(1), \tilde{w}(2), \ldots, \tilde{w}(T) \qquad w(t) \in \mathbb{R}^{\mathtt{w}}$$

In the theory, the case $T \to \infty$ and (bi-)infinite data records

$$\tilde{w}(1), \tilde{w}(2), \ldots, \tilde{w}(t), \cdots$$

$$\cdots, \tilde{w}(-t), \cdots, \tilde{w}(0), \tilde{w}(1), \tilde{w}(2), \ldots, \tilde{w}(t), \cdots$$

play an important role.

SYSTEM ID

Observed data $\mapsto$ System model

Difficulties to cope with:

- **'blackbox' data**
- **unmeasured inputs 'latency'**
- **any element of the model class fits the data only approximately 'misfit'**
- **measurement 'errors'**
- **danger of 'overfitting'**

ARMAX SYSTEM ID

Usual approach: Data = an input/output record

$$\begin{bmatrix} \tilde{u}(1) \\ \tilde{y}(1) \end{bmatrix}, \begin{bmatrix} \tilde{u}(2) \\ \tilde{y}(2) \end{bmatrix}, \ldots, \begin{bmatrix} \tilde{u}(T) \\ \tilde{y}(T) \end{bmatrix}$$

$$\tilde{u}(t) \in \mathbb{R}^{\mathtt{m}}, \tilde{y}(t) \in \mathbb{R}^{\mathtt{p}}$$

ARMAX SYSTEM ID

Usual approach: Data = an input/output record

$$\begin{bmatrix} \tilde{u}(1) \\ \tilde{y}(1) \end{bmatrix}, \begin{bmatrix} \tilde{u}(2) \\ \tilde{y}(2) \end{bmatrix}, \ldots, \begin{bmatrix} \tilde{u}(T) \\ \tilde{y}(T) \end{bmatrix}$$

$\tilde{u}(t) \in \mathbb{R}^{\mathtt{m}}, \tilde{y}(t) \in \mathbb{R}^{\mathtt{p}}$

System model = an 'ARMAX' model

$\sigma =$ **'shift'**, $(\sigma f)(t) := f(t+1)$

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon$$

$\varepsilon =$ **'noise'**

$P, Q, M \in \mathbb{R}^{\bullet \times \bullet}[\xi]$, **suitably sized polynomial matrices.**

u, ε **stationary ergodic gaussian,** ε **white, independent of** u**.**

ARMAX SYSTEM ID

System model = an ‘ARMAX’ model

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon$$

$\varepsilon =$ ‘noise’

$P, Q, M \in \mathbb{R}^{\bullet \times \bullet}[\xi]$, suitably sized polynomial matrices.

u, ε stationary ergodic gaussian, ε white, independent of u.

$M(\sigma)\varepsilon$: lack of fit, prevents predictability of y from u, etc.

Note *subtle non-uniqueness* of the ARMAX representation.

ARMAX SYSTEM ID

System model = an 'ARMAX' model

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon$$

$\varepsilon =$ 'noise'

Well-known: ARMAX systems are those that allow finite dimensional state representations

$$\sigma x = Ax + Bu + G\varepsilon, \;\; y = Cx + Du + J\varepsilon$$

ARMAX SYSTEM ID

System model = an 'ARMAX' model

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon$$

$\varepsilon =$ 'noise'

ID Algorithm:

$$(\tilde{u}, \tilde{y}) \quad \mapsto \quad (\hat{P}, \hat{Q}, \hat{M})$$

(or another repr. of the ARMAX model)

Quality of the ID algorithm:

Assume that the data has been generated by an element of the model class; then require asymptotic convergence to the 'true system', for $T \to \infty$ (consistency, efficiency, etc.)

CENTRAL PARADIGM

Test the proposed algorithm assuming that the data has been generated by a model from a given (ARMAX) model class. The algorithm should perform well in this ‘test case’. In other words,

ID algorithms should perform well with simulated data

CENTRAL PARADIGM

Test the proposed algorithm assuming that the data has been generated by a model from a given (ARMAX) model class. The algorithm should perform well in this 'test case'. In other words,

ID algorithms should perform well with simulated data

There is no need to refer to the 'real' or 'true' system.

As a(n approximate) description of reality, the stochastic assumptions about u, ε are indeed rather tenuous!

CENTRAL PARADIGM

Test the proposed algorithm assuming that the data has been generated by a model from a given (ARMAX) model class. The algorithm should perform well in this 'test case'. In other words,

ID algorithms should perform well with simulated data

Same sort of justification for Kalman filtering, LQ-, LQG-, $\mathcal{H}_\infty$-control, adaptive control, etc.: We want that our algorithm works well under certain 'ideal' circumstances.

Stochasticity can thus in good conscience be interpreted as relative frequency.

Is the CENTRAL PARADIGM reasonable?

ID algorithms should perform well with simulated data

- **What does 'perform well' mean?**
- **What 'simulated data' should one test the algorithm for?**

Is the CENTRAL PARADIGM reasonable?

ID algorithms should perform well with simulated data

- **What does ‘perform well’ mean?**
- **What ‘simulated data’ should one test the algorithm for?**

The ARMAX model class, with stochastic inputs and disturbances is a very broad model class, but it puts ‘stochasticity’ very central.

Is the CENTRAL PARADIGM reasonable?

ID algorithms should perform well with simulated data

- What does 'perform well' mean?
- What 'simulated data' should one test the algorithm for?

Approximation deserves a much more central place in system ID. It (data produced by high order, nonlinear, time-varying system) seems much more the core problem in system ID than protection against unmeasured stochastic inputs or measurement 'errors'.

The behavior of ARMAX systems

The behavior of an ARMAX system

When does the stochastic process

$$(u, y) : \mathbb{Z} \to \mathbb{R}^{\mathtt{m}} \times \mathbb{R}^{\mathtt{p}}$$

belong to the behavior of the ARMAX system

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon \textbf{ ?}$$

The behavior of an ARMAX system

When does the stochastic process

$$(u, y) : \mathbb{Z} \to \mathbb{R}^{\mathrm{m}} \times \mathbb{R}^{\mathrm{p}}$$

belong to the behavior of the ARMAX system

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon \ ?$$

:$\Leftrightarrow$ (u, y) is zero mean, stationary, gaussian, and there exist a stationary gaussian white noise process ε, independent of u such that

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon, \quad \text{a.s.}$$

Cfr. the work of Picci, Lindquist, (and co-workers), Deistler, Ljung, e.a.

The behavior of an ARMAX system

Deterministic language

When does the time-series

$$(u, y) : \mathbb{Z} \to \mathbb{R}^{\mathtt{m}} \times \mathbb{R}^{\mathtt{p}}$$

belong to the behavior of the ARMAX system

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon \text{ ?}$$

The behavior of an ARMAX system

There is an underlying shift-invariant Hilbert space $\mathfrak{H}$ of time-series $f : \mathbb{Z} \to \mathbb{R}$ to which the components of all the signals are assumed to belong. $< f, g > = < \sigma f, \sigma g >$

The behavior of an ARMAX system

There is an underlying shift-invariant Hilbert space $\mathfrak{H}$ of time-series $f : \mathbb{Z} \rightarrow \mathbb{R}$ to which the components of all the signals are assumed to belong. $< f, g > = < \sigma f, \sigma g >$

Examples:

- jointly stationary ergodic gaussian stochastic processes

$$||f||^2 = \mathfrak{E}||f(t)||^2 = \lim_{T \to \infty} \frac{1}{2T+1} \sum_{t=-T}^{T} ||f(t)||^2$$

- ℓ_2
- almost periodic sequences

The behavior of an ARMAX system

$$(u, y) : \mathbb{Z} \to \mathbb{R}^{\mathtt{m}} \times \mathbb{R}^{\mathtt{p}}$$

belongs to the behavior $\mathfrak{B}$ of the ARMAX system (P, Q, M), i.e.

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon \quad :\Leftrightarrow$$

1. its components $\in \mathfrak{H}$
2. there exists $\varepsilon : \mathbb{Z} \to \mathbb{R}^{\mathtt{e}}$ with
   (a) components $\in \mathfrak{H}$
   (b) the $\sigma^t\varepsilon$'s are orthonormal, $t \in \mathbb{Z}$
   (c) the $\sigma^t\varepsilon$'s are $\perp$ to the $\sigma^t u$'s, $t \in \mathbb{Z}$

   such that $P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon$

The notion of an input

$$P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon$$

Call u free if $\forall\ u \in \mathfrak{H}^{\mathtt{m}}, \exists\ y \in \mathfrak{H}^{\mathtt{p}}$ such that $(u, y) \in \mathfrak{B}$.

Maximally free :$\Leftrightarrow$ no further free components in y.

Maximally free = 'input'. Then y = 'output'.

u is input $\Leftrightarrow$ P square and $\det(P) \neq 0$.

The behavior of an ARMA system

An **ARMA system** is an ARMAX system that has only outputs. Analog of 'autonomous' system. $y : \mathbb{Z} \to \mathbb{R}^{\mathrm{p}}$ belongs to the behavior $\mathfrak{B}$ of the ARMA system (P, M), i.e.

$$P(\sigma)y = M(\sigma)\varepsilon, \quad \det(P) \neq 0 \qquad :\Leftrightarrow$$

1. its components $\in \mathfrak{H}$
2. there exists $\varepsilon : \mathbb{Z} \to \mathbb{R}^{\mathrm{e}}$ with
   (a) components $\in \mathfrak{H}$
   (b) the $\sigma^t e$'s are **orthonormal**, $t \in \mathbb{Z}$

   such that $P(\sigma)y = M(\sigma)\varepsilon$

To avoid difficulties yet to be dealt with in the proofs, assume throughout that $[PP^* \;\; MM^*]$ is left prime and that $\det(P)$ has no roots on the unit circle.

The behavior of an ARMA system

An **ARMA system** is an ARMAX system that has only outputs.

Analog of 'autonomous' system.

$$P(\sigma)y = M(\sigma)\varepsilon, \ \ \det(P) \neq 0$$

The behavior of an ARMA system consists of the $y \in \mathfrak{H}^{\mathrm{p}}$ that have the same **autocorrelation function** $\rho_{yy} : \mathbb{Z} \to \mathbb{R}^{\mathrm{p}\times\mathrm{p}}$ with

$$\rho_{yy}(t) := < \sigma^t y, y > .$$

etc., etc.

The behavior of an MA system

An **MA system** is a special ARMA system. $y : \mathbb{Z} \to \mathbb{R}^{\mathrm{p}}$ belongs to the behavior $\mathfrak{B}$ of the MA system M, i.e.

$$y = M(\sigma)\varepsilon, \qquad :\Leftrightarrow$$

1. its components $\in \mathfrak{H}$
2. there exists $\varepsilon : \mathbb{Z} \to \mathbb{R}^{\mathrm{e}}$ with
   (a) components $\in \mathfrak{H}$
   (b) the $\sigma^t e$'s are **orthonormal**, $t \in \mathbb{Z}$

   such that $y = M(\sigma)\varepsilon$

The behavior of an MA system

An **MA system** is a special ARMA system.

$$y = M(\sigma)\varepsilon,$$

The behavior of an MA system consists of the $y \in \mathfrak{H}^{\mathrm{p}}$ that have the same compact support autocorrelation function $\rho_y : \mathbb{Z} \to \mathbb{R}^{\mathrm{p}\times\mathrm{p}}$ with

$$\rho_y(t) := < \sigma^t y, y > .$$

etc., etc.

What subsets $\mathfrak{B}$ of $\mathfrak{H}^{\mathtt{m}+\mathtt{p}}$ are representable as a ARMAX, ARMA, MA system?

Denote the family of these subsets as
$\mathbb{B}_{\textbf{ARMAX}}, \mathbb{B}_{\textbf{ARMA}}, \mathbb{B}_{\textbf{MA}}.$

What subsets $\mathfrak{B}$ of $\mathfrak{H}^{\mathtt{m}+\mathtt{p}}$ are representable as a ARMAX, ARMA, MA system?

Denote the family of these subsets as $\mathbb{B}_{\mathbf{ARMAX}}, \mathbb{B}_{\mathbf{ARMA}}, \mathbb{B}_{\mathbf{MA}}$.

If $\mathfrak{B} \in \mathbb{B}_{\mathbf{ARMAX}}, \mathbb{B}_{\mathbf{ARMA}}, \mathbb{B}_{\mathbf{MA}}$, what are all its representations $(P, Q, M), (P, M), M$?
$\rightsquigarrow$ an equivalence relation on tuples of polynomial matrices

Identifiability

Given $(u, y) \in \mathfrak{B}$, are there simple conditions (say on the u) such that there is **only one** element in $\mathbb{B}_{\mathrm{ARMAX}}$ that contains this (u, y)?

$\rightsquigarrow$ **persistency of excitation**

Identification problem

Give an algorithm

$$(\tilde{u}, \tilde{y}) \mapsto (\hat{P}, \hat{Q}, \hat{M}) \cong \hat{\mathfrak{B}}.$$

estimate, finite data records, consistency, etc.

The structure of ARMAX systems

The module of orthogonalisers

Consider

$(P, Q, M) \mapsto \mathfrak{B} \in \mathbb{B}_{\mathrm{ARMAX}}$, P **square,** $\det(P) \neq 0$.

Define the following set of polynomial vectors

$$\Pi_{\mathfrak{B}} := \{\pi \in \mathbb{R}^{\mathtt{m}+\mathtt{p}}[\xi, \xi^{-1}] \mid \pi^\top(\sigma) \left[\begin{smallmatrix} y \\ u \end{smallmatrix}\right] \perp \sigma^t u$$

for all $t \in \mathbb{Z}$ **and for all** $(u, y) \in \mathfrak{B}\}$

The module of orthogonalisers

Define the following set of polynomial vectors

$$\Pi_{\mathfrak{B}} := \{\pi \in \mathbb{R}^{\mathtt{m}+\mathtt{p}}[\xi, \xi^{-1}] \mid \pi^\top(\sigma) \left[\begin{smallmatrix} y \\ u \end{smallmatrix}\right] \perp \sigma^t u$$

$$\text{for all } t \in \mathbb{Z} \text{ and for all } (u, y) \in \mathfrak{B}\}$$

Easy: $\Pi_{\mathfrak{B}}$ submodule of $\mathbb{R}^{\mathtt{m}+\mathtt{p}}[\xi, \xi^{-1}]$ viewed as a module over $\mathbb{R}[\xi, \xi^{-1}]$. Hence finitely generated.

FAQ: What is $\Pi_{\mathfrak{B}}$ in terms of the ARMAX matrices (P, Q, M)?

The module of orthogonalisers

$$\Pi_{\mathfrak{B}} := \{\pi \in \mathbb{R}^{\mathtt{m}+\mathtt{p}}[\xi, \xi^{-1}] \mid \pi^\top(\sigma, \sigma^{-1}) \left[\begin{smallmatrix} y \\ u \end{smallmatrix}\right] \perp \sigma^t u\}$$

$\mathfrak{B} \cong P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon$ and $\sigma^t\varepsilon$'s $\perp$ $\sigma^{t'}u$'s

$\Rightarrow$ the transposes of the rows of $[P \;\; Q] \in \Pi_{\mathfrak{B}}$.

But, these do not always form a set of generators.

The module of orthogonalisers

$$\Pi_{\mathfrak{B}} := \{\pi \in \mathbb{R}^{\mathtt{m}+\mathtt{p}}[\xi, \xi^{-1}] \mid \pi^\top(\sigma, \sigma^{-1}) \left[\begin{smallmatrix} y \\ u \end{smallmatrix}\right] \perp \sigma^t u\}$$

$\mathfrak{B} \cong P(\sigma)y + Q(\sigma)u = M(\sigma)\varepsilon$ and $\sigma^t\varepsilon$'s $\perp$ $\sigma^{t'}u$'s

$\Rightarrow$ the transposes of the rows of $[P \;\; Q] \in \Pi_{\mathfrak{B}}$.

But, these do not always form a set of generators.

Proposition:

Let $[P \;\; Q] = A[P_c \;\; Q_c]$

with A square and non-singular, and $[P_c \;\; Q_c]$ left prime.

The transposes of the rows of $[P_c \;\; Q_c]$ form a (minimal) set of generators of the module $\Pi_{\mathfrak{B}}$.

The module of orthogonalisers

$$\Pi_{\mathfrak{B}} := \{\pi \in \mathbb{R}^{\mathtt{m}+\mathtt{p}}[\xi,\xi^{-1}] \mid \pi^\top(\sigma,\sigma^{-1})\begin{bmatrix} y \\ u \end{bmatrix} \perp \sigma^t u\}$$

Proposition:

Let $$[P\ \ Q] = A[P_c\ \ Q_c]$$

with A square and non-singular, and $[P_c\ \ Q_c]$ left prime.

The transposes of the rows of $[P_c\ \ Q_c]$ form a (minimal) set of generators of the module $\Pi_{\mathfrak{B}}$.

In behavioral language, $[P_c\ \ Q_c]$ defines a **'controllable'** kernel; $P_c^{-1}Q_c$ is the transfer function of the **'deterministic part'** of our ARMAX system.

The module of orthogonalisers

$$\Pi_{\mathfrak{B}} := \{\pi \in \mathbb{R}^{\mathtt{m}+\mathtt{p}}[\xi, \xi^{-1}] \mid \pi^\top(\sigma, \sigma^{-1}) \left[\begin{smallmatrix} y \\ u \end{smallmatrix}\right] \perp \sigma^t u\}$$

Proposition:

Let $[P \;\; Q] = A[P_c \;\; Q_c]$

with A square and non-singular, and $[P_c \;\; Q_c]$ left prime.

The transposes of the rows of $[P_c \;\; Q_c]$ form a (minimal) set of generators of the module $\Pi_{\mathfrak{B}}$.

$[P_c \;\; Q_c]$ unique up to $[P_c \;\; Q_c] \mapsto U[P_c \;\; Q_c]$, U unimodular

This submodule $\cong$ the X part of the ARMAX system

The AR module

Consider the ARMA system

$(P, M) \mapsto \mathfrak{B} \in \mathbb{B}_{\mathrm{ARMA}}$, P **square,** $\det(P) \neq 0$.

Define the following set of polynomial vectors

$$\Gamma_{\mathfrak{B}} := \{\gamma \in \mathbb{R}^{\mathtt{p}}[\xi] \mid \gamma^\top(\sigma) y \perp \sigma^t y \; \forall\, t \in \mathbb{Z}, t < 0\}$$

The AR module

Consider the ARMA system

$(P, M) \mapsto \mathfrak{B} \in \mathbb{B}_{\mathrm{ARMA}}$, P **square,** $\det(P) \neq 0$**.**

Define the following set of polynomial vectors

$$\Gamma_{\mathfrak{B}} := \{\gamma \in \mathbb{R}^{\mathrm{p}}[\xi] \mid \gamma^{\top}(\sigma) y \perp \sigma^{t} y \; \forall \, t \in \mathbb{Z}, t < 0\}$$

Easy: $\Gamma_{\mathfrak{B}}$ **submodule of** $\mathbb{R}^{\mathrm{p}}[\xi]$**. Hence finitely generated.**

FAQ: What is $\Gamma_{\mathfrak{B}}$ **in terms of the ARMAX matrices** (P, M)**?**

The AR module

Consider the ARMA system

$(P, M) \mapsto \mathfrak{B} \in \mathbb{B}_{\mathrm{ARMA}}$, P **square,** $\det(P) \neq 0$.

Define the following set of polynomial vectors

$$\Gamma_{\mathfrak{B}} := \{\gamma \in \mathbb{R}^{\mathrm{p}}[\xi] \mid \gamma^{\top}(\sigma)y \perp \sigma^{t}y \; \forall\, t \in \mathbb{Z}, t < 0\}$$

$\Gamma_{\mathfrak{B}}$ **is the module generated by the transp. of the rows of** H**, with**

$$H(\xi)H^{\top}(\xi^{-1}) = P(\xi)P^{\top}(\xi^{-1})$$

a unit circle 'spectral factorization'.

The AR module

Now, given an ARMAX system (P, Q, M), with behavior $\mathfrak{B}$, and $[P \;\; Q] = A[P_c \;\; Q_c]$ with $[P_c \;\; Q_c]$ generators of the module of orthogonalisers, it can be shown that

$$\mathfrak{B}' = [P_c(\sigma) \;\; Q_c(\sigma)]\mathfrak{B} = \{f : \mathbb{Z} \to \mathbb{R}^{\mathrm{p}} \mid \exists \begin{bmatrix} y \\ u \end{bmatrix} \in \mathfrak{B} \text{ such that } f = P_c(\sigma)y + Q_c(\sigma)u\}$$

is an ARMA system.

The AR module

$$\mathfrak{B}' = [P_c(\sigma) \;\; Q_c(\sigma)]\mathfrak{B} = \{f : \mathbb{Z} \to \mathbb{R}^{\mathrm{p}} \mid$$
$$\exists \begin{bmatrix} y \\ u \end{bmatrix} \in \mathfrak{B} \text{ such that } f = P_c(\sigma)y + Q_c(\sigma)u\}$$

is an ARMA system. Described by

$$A(\sigma)f = M(\sigma)\varepsilon.$$

The associated AR submodule $\cong$ the AR part of the ARMAX system. Assume generated by the polynomial matrix H.

The AR module

Described by

$$A(\sigma)f = M(\sigma)\varepsilon.$$

The associated AR submodule $\cong$ the **AR part** of the ARMAX system. Assume generated by the polynomial matrix H.

It follows finally that $H(\sigma)\mathfrak{B}'$ is an MA system, $m = M(\sigma)\varepsilon$.

The associated MA matrix $\cong$ the **MA part** of the ARMAX system.

Recapitulation

Time-series $u : \mathbb{Z} \to \mathbb{R}^{\mathtt{m}}$, $y : \mathbb{Z} \to \mathbb{R}^{\mathtt{p}}$.

Behavior $\mathfrak{B}$, an ARMAX behavior.

u input, y output.

1. The module of orthogonalisers $\rightsquigarrow$ $[P_c \ \ Q_c]$.
2. $f = P_c(\sigma)y + Q_c(\sigma)u$ is an ARMA system. The AR-module $\rightsquigarrow$ H.
3. $m = H(\sigma)f$ is an MA-system $\rightsquigarrow$ M
4. $\rightsquigarrow$ the ARMAX representation

$$H(\sigma)P_c(\sigma)y + H(\sigma)Q_c(\sigma)u = M(\sigma)\varepsilon$$

An identification algorithm

An identification algorithm

For notational simplicity, we only treat the case that an infinite time series

$$\ldots, \begin{bmatrix} \tilde{y}(-t) \\ \tilde{u}(-t) \end{bmatrix}, \ldots, \begin{bmatrix} \tilde{y}(0) \\ \tilde{u}(0) \end{bmatrix}, \begin{bmatrix} \tilde{y}(1) \\ \tilde{u}(1) \end{bmatrix}, \begin{bmatrix} \tilde{y}(2) \\ \tilde{u}(2) \end{bmatrix}, \ldots, \begin{bmatrix} \tilde{y}(t) \\ \tilde{u}(t) \end{bmatrix}, \ldots$$

is given, components in $\mathfrak{H}$.

Algorithm

Data:

$$\ldots, \begin{bmatrix} \tilde{y}(-t) \\ \tilde{u}(-t) \end{bmatrix}, \ldots, \begin{bmatrix} \tilde{y}(0) \\ \tilde{u}(0) \end{bmatrix}, \begin{bmatrix} \tilde{y}(1) \\ \tilde{u}(1) \end{bmatrix}, \begin{bmatrix} \tilde{y}(2) \\ \tilde{u}(2) \end{bmatrix}, \ldots, \begin{bmatrix} \tilde{y}(t) \\ \tilde{u}(t) \end{bmatrix}, \ldots$$

We assume that the data has been produced by an ARMA system, and we are looking for an algorithm that returns (P, Q, M), equivalently, $([P_c \; Q_c], H, M)$.

PERSISTENCY of EXCITATION

A key ingredient is **'persistency of excitation'**.

The vector time-series $\tilde{u}$ is said to be ***persistently exciting of order L*** if the Hankel matrix

$$\begin{bmatrix} \cdots & \tilde{u}(1) & \tilde{u}(2) & \tilde{u}(3) & \cdots & \tilde{u}(T-L+1) & \cdots \\ \cdots & \tilde{u}(2) & \tilde{u}(3) & \tilde{u}(4) & \cdots & \tilde{u}(T-L+2) & \cdots \\ \cdots & \tilde{u}(3) & \tilde{u}(4) & \tilde{u}(5) & \cdots & \tilde{u}(T-L+3) & \cdots \\ \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots \\ \cdots & \tilde{u}(L) & \tilde{u}(L+1) & \tilde{u}(L+2) & \cdots & \tilde{u}(T) & \cdots \end{bmatrix}$$

is of <u>full row rank</u>.

Persistency of excitation $\Leftrightarrow$ no linear relations of order L.

PERSISTENCY of EXCITATION

A key ingredient is ‘persistency of excitation’.

The vector time-series $\tilde{u}$ is said to be *persistently exciting of order L* if the Hankel matrix

$$\begin{bmatrix} \cdots & \tilde{u}(1) & \tilde{u}(2) & \tilde{u}(3) & \cdots & \tilde{u}(T-L+1) & \cdots \\ \cdots & \tilde{u}(2) & \tilde{u}(3) & \tilde{u}(4) & \cdots & \tilde{u}(T-L+2) & \cdots \\ \cdots & \tilde{u}(3) & \tilde{u}(4) & \tilde{u}(5) & \cdots & \tilde{u}(T-L+3) & \cdots \\ \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots \\ \cdots & \tilde{u}(L) & \tilde{u}(L+1) & \tilde{u}(L+2) & \cdots & \tilde{u}(T) & \cdots \end{bmatrix}$$

is of full row rank.

Assume persistency of excitation as needed.

Estimating the X part

Determine the orthogonalisers, for instance by computing the 'MPUM' type module generated by the left kernel of the Hankel matrix of the mixed correlation matrix

$$\begin{bmatrix} \rho_{uu}(1) & \rho_{uu}(2) & \cdots & \rho_{uu}(T) & \cdots \\ \rho_{yu}(1) & \rho_{yu}(2) & \cdots & \rho_{yu}(T) & \cdots \\ \rho_{uu}(2) & \rho_{uu}(3) & \cdots & \rho_{uu}(T+1) & \cdots \\ \rho_{yu}(2) & \rho_{yu}(3) & \cdots & \rho_{yu}(T+1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{uu}(\Delta) & \rho_{uu}(\Delta+1) & \cdots & \rho_{uu}(T+\Delta-1) & \cdots \\ \rho_{yu}(\Delta) & \rho_{yu}(\Delta+1) & \cdots & \rho_{yu}(T+\Delta-1) & \cdots \end{bmatrix}$$

with $\rho_{uu}(t) =< \sigma^t \tilde{u}, \tilde{u} >, \rho_{yu}(t) =< \sigma^t \tilde{y}, \tilde{u} >$

Estimating the X part

Determine the orthogonalisers, for instance by computing the 'MPUM' type module generated by the left kernel of the Hankel matrix of the mixed correlation matrix

$$\begin{bmatrix} \rho_{uu}(1) & \rho_{uu}(2) & \cdots & \rho_{uu}(T) & \cdots \\ \rho_{yu}(1) & \rho_{yu}(2) & \cdots & \rho_{yu}(T) & \cdots \\ \rho_{uu}(2) & \rho_{uu}(3) & \cdots & \rho_{uu}(T+1) & \cdots \\ \rho_{yu}(2) & \rho_{yu}(3) & \cdots & \rho_{yu}(T+1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{uu}(\Delta) & \rho_{uu}(\Delta+1) & \cdots & \rho_{uu}(T+\Delta-1) & \cdots \\ \rho_{yu}(\Delta) & \rho_{yu}(\Delta+1) & \cdots & \rho_{yu}(T+\Delta-1) & \cdots \end{bmatrix}$$

with $\rho_{uu}(t) =< \sigma^t\tilde{u}, \tilde{u} >, \rho_{yu}(t) =< \sigma^t\tilde{y}, \tilde{u} >$

Result: $\hat{P}_c, \hat{Q}_c$

AR part

Compute the ARMA signal

$$\tilde{f} = \hat{P}_c(\sigma)\tilde{y} + \hat{Q}_c(\sigma)\tilde{u}$$

AR part

Compute the ARMA signal

$$\tilde{f} = \hat{P}_c(\sigma)\tilde{y} + \hat{Q}_c(\sigma)\tilde{u}$$

Compute its AR module for instance by computing the MPUM type module generated by the left kernel of the Hankel matrix

$$\begin{bmatrix} \rho_{ff}(1) & \rho_{ff}(2) & \cdots & \rho_{ff}(T) & \cdots \\ \rho_{ff}(2) & \rho_{ff}(3) & \cdots & \rho_{ff}(T+1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{ff}(\Delta) & \rho_{ff}(\Delta+1) & \cdots & \rho_{ff}(T+\Delta-1) & \cdots \end{bmatrix}$$

with $\rho_{ff}(t) =< \sigma^t \tilde{f}, \tilde{f} >$.

AR part

Compute the ARMA signal

$$\tilde{f} = \hat{P}_c(\sigma)\tilde{y} + \hat{Q}_c(\sigma)\tilde{u}$$

Compute its AR module for instance by computing the MPUM type module generated by the left kernel of the Hankel matrix

$$\begin{bmatrix} \rho_{ff}(1) & \rho_{ff}(2) & \cdots & \rho_{ff}(T) & \cdots \\ \rho_{ff}(2) & \rho_{ff}(3) & \cdots & \rho_{ff}(T+1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{ff}(\Delta) & \rho_{ff}(\Delta+1) & \cdots & \rho_{ff}(T+\Delta-1) & \cdots \end{bmatrix}$$

with $\rho_{ff}(t) = < \sigma^t \tilde{f}, \tilde{f} > .$

Result: $\hat{H}$

MA part

Compute the signal

$$\tilde{m} = \hat{H}(\sigma)\tilde{f}$$

MA part

Compute the signal

$$\tilde{m} = \hat{H}(\sigma)\tilde{f}$$

Compute its MA representation for instance by computing the correlations $\rho_{mm}(t) =< \sigma^t \tilde{m}, \tilde{m} >$, and factoring as

$$\sum_{t=-\mathtt{n}}^{\mathtt{n}} \rho_{mm}(t)\xi^t = \hat{M}(\xi)\hat{M}^\top(\xi^{-1})$$

MA part

Compute the signal

$$\tilde{m} = \hat{H}(\sigma)\tilde{f}$$

Compute its MA representation for instance by computing the correlations $\rho_{mm}(t) = < \sigma^t \tilde{m}, \tilde{m} >$**, and factoring as**

$$\sum_{t=-\mathrm{n}}^{\mathrm{n}} \rho_{mm}(t)\xi^t = \hat{M}(\xi)\hat{M}^\top(\xi^{-1})$$

Result: $\hat{M}$

MA part

Real computations, finite time series, noise:

$\rightsquigarrow$ all sorts of approximations.

How can we guarantee that H is Schur,

how do we guarantee that the autocorrelation of $\tilde{m}$ has the necessary positivity and compact support properties?

We take a look at the second problem, in the scalar case.

MA part

Compute estimates $\hat{\rho}_{mm}(t) =< \sigma^t \tilde{m}, \tilde{m} >$.

We can approximate $\hat{\rho}$ with a 'legal' MA ρ by solving the LMI:

minimize $\sum_{t=-n}^{n} |\hat{\rho}(t) - \rho(t)|^2$ subject to

$$\exists\, \Gamma = \Gamma^T \geq 0 \text{ such that } \sum_{t=-n}^{n} \rho(t)\xi^t = \begin{bmatrix} 1 \\ \xi^{-1} \\ \vdots \\ \xi^{-n} \end{bmatrix}^\top \Gamma \begin{bmatrix} 1 \\ \xi \\ \vdots \\ \xi^{n} \end{bmatrix}$$

MA part

Compute estimates $\hat{\rho}_{mm}(t) =< \sigma^t \tilde{m}, \tilde{m} >$.

We can approximate $\hat{\rho}$ with a ‘legal’ MA ρ by solving the LMI:

minimize $\sum_{t=-\mathrm{n}}^{\mathrm{n}} |\hat{\rho}(t) - \rho(t)|^2$ subject to

$$\exists\, \Gamma = \Gamma^T \geq 0 \text{ such that } \sum_{t=-\mathrm{n}}^{\mathrm{n}} \rho(t)\xi^t = \begin{bmatrix} 1 \\ \xi^{-1} \\ \vdots \\ \xi^{-\mathrm{n}} \end{bmatrix}^\top \Gamma \begin{bmatrix} 1 \\ \xi \\ \vdots \\ \xi^{\mathrm{n}} \end{bmatrix}$$

The dyadic expansion of $\Gamma = m_1 m_1^\top + \cdots + m_{\mathrm{k}} m_{\mathrm{k}}^\top$ then

yields, with $M(\xi) = \begin{bmatrix} m_1 \\ \vdots \\ m_{\mathrm{k}} \end{bmatrix} \begin{bmatrix} 1 \\ \xi \\ \vdots \\ \xi^{\mathrm{n}} \end{bmatrix}$, an LS MA approximation. Can be

reduced, via storage functions, and another LMI, to a scalar M.

A simulation

The system is

$$\underbrace{H(\sigma)P_c(\sigma)}_{P(\sigma)}y = \underbrace{H(\sigma)Q_c(\sigma)}_{Q(\sigma)}u + M(\sigma)\varepsilon,$$

where the polynomials H, P_c, Q_c, and M are selected as follows:

$$H(\xi) = 1+\xi+0.5\xi^2, \quad Q_c(\xi) = 1-1.2\xi+0.6\xi^2-0.7\xi^3, \quad M(\xi) = 1+0.5\xi,$$

$$P_c(\xi) = 1 - 0.8713\xi - 1.539\xi^2 + 1.371\xi^3 + 0.6451\xi^4 - 0.5827\xi^5.$$

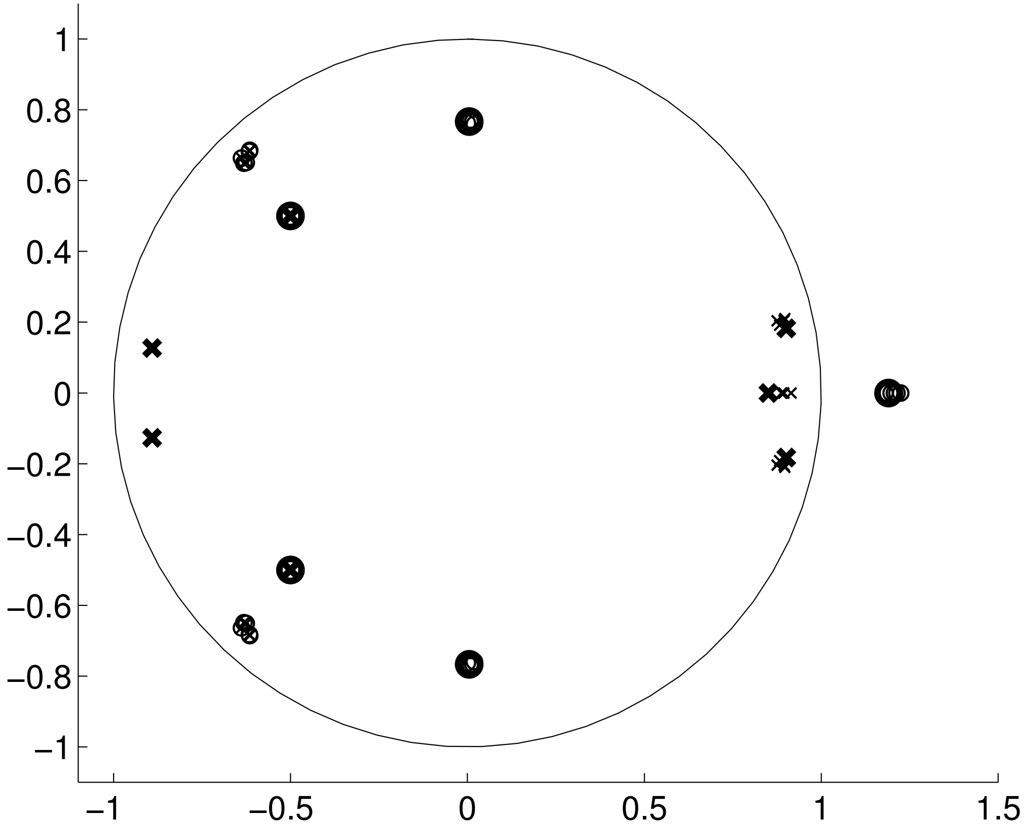
A simulation

The inputs u and ε are zero mean, gaussian, white, with variances 1 and 0.2, respectively. The initial condition, under which y is obtained from u and ε is a random vector.

The time horizon for the simulation is $T = 1000$ and the simulated time series (u, y) is used for estimation.

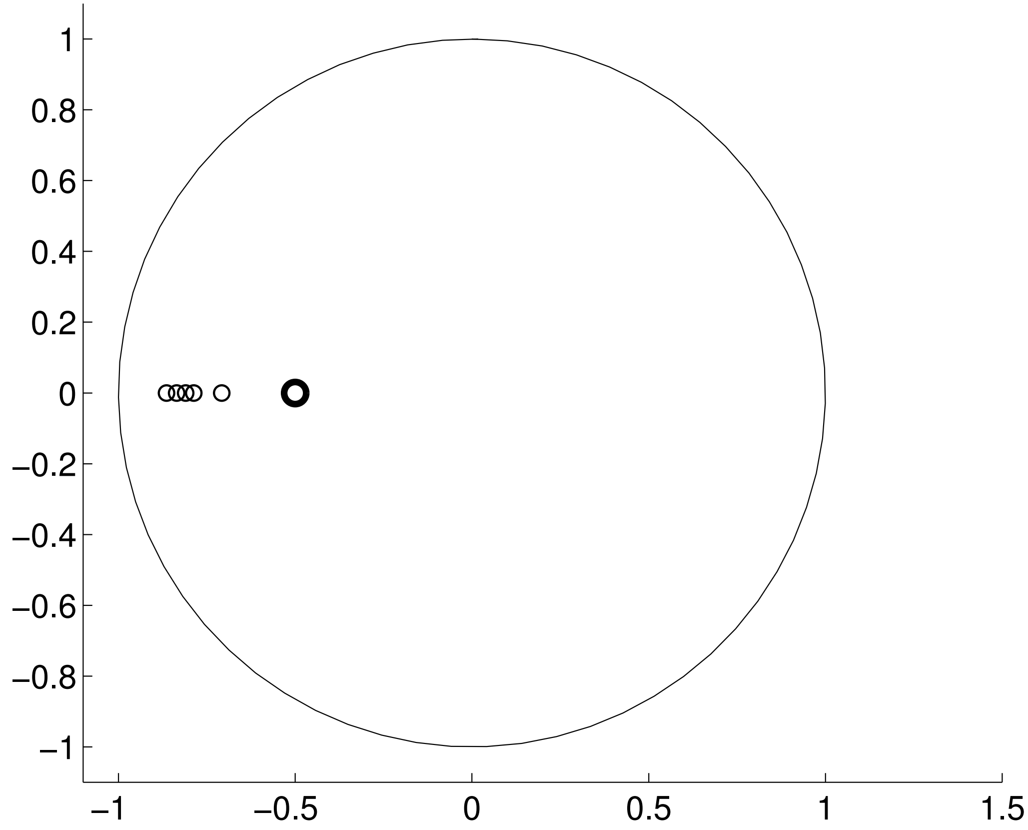
The experiment is repeated $N = 5$ times with different realizations of u and ε in each run.

A simulation



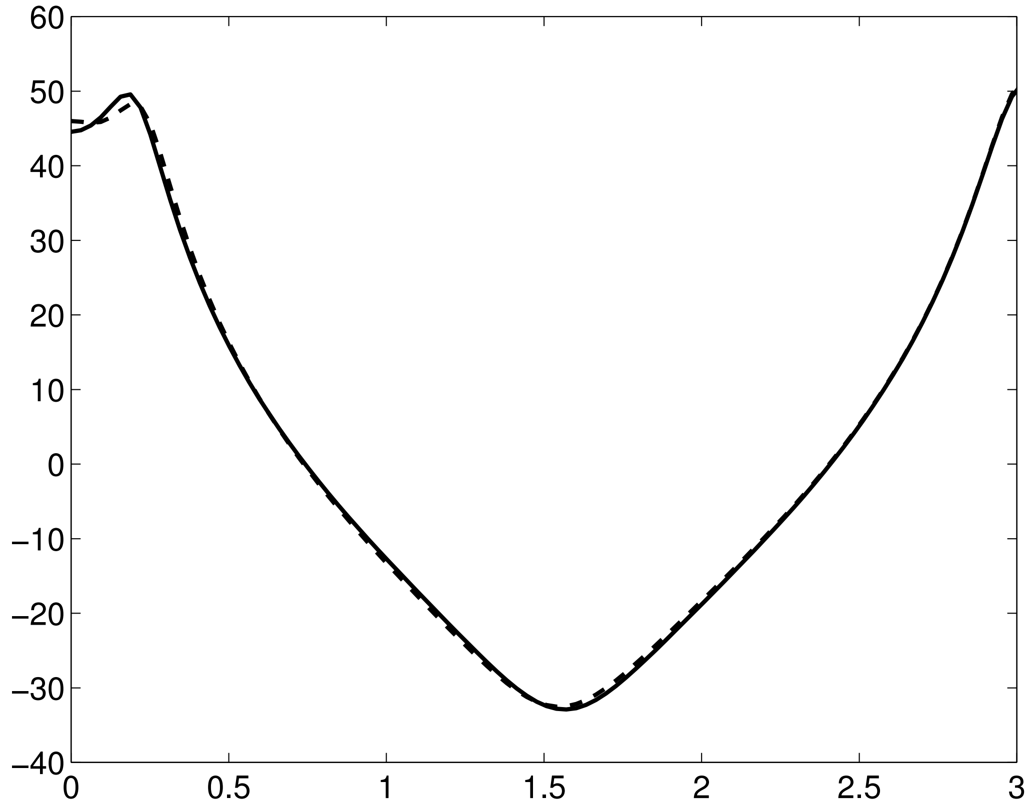
Roots of P, Q, $\hat{P}^{(k)}$, and $\hat{Q}^{(k)}$, for $k = 1, \ldots, N$.

A simulation



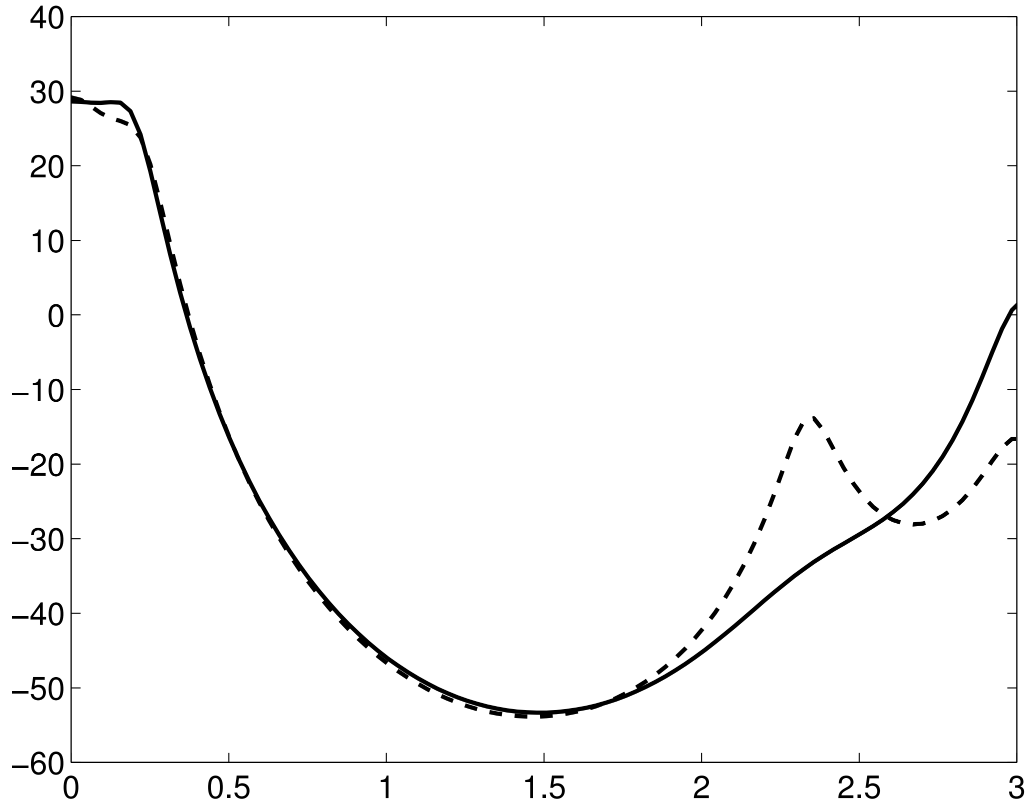
Roots of M and $\hat{M}^{(k)}$, for $k = 1, \ldots, N$.

A simulation



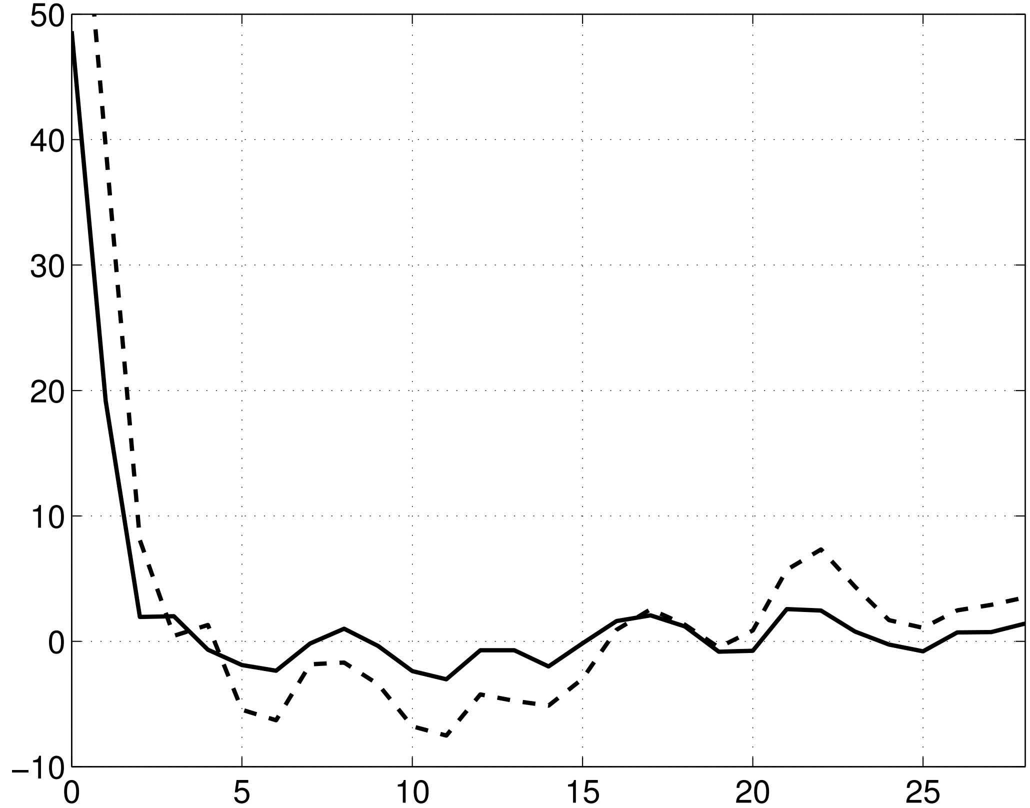
Bode plots of Q/P (solid line) and $\hat{Q}^{(1)}/\hat{P}^{(1)}$ (dashed line).

A simulation



Bode plots of M/P (solid line) and $\hat{M}^{(1)}/\hat{P}^{(1)}$ (dashed line).

A simulation



Autocorrelation of $P(\sigma)y - Q(\sigma)u$ (solid line) and ˆ (dashed line).

Remarks

- Note dramatic simplification to orthogonalisers/MA if deterministic part in controllable.
- The orthogonality suffices for a finite number of shifts.

Thank you

Thank you

Thank you

Thank you

Thank you

Thank you

Thank you

Thank you