# THE MODULE STRUCTURE

# of

# ARMAX SYSTEMS

**Jan Willems, K.U. Leuven, Belgium**

**CDC 2006, San Diego** **December 13, 2006**

**Joint paper with**

**Ivan Markovsky & Bart De Moor (K.U. Leuven)**

# ARMAX systems

# Processes

**In this talk** **'process'** **means:**

**a vector of real stochastic processes on $\mathbb{Z}$,**
**gaussian, zero mean, stationary, and ergodic.**

**'White noise'**: **a process $\varepsilon$ with**

$$\text{the } \varepsilon(t)\text{'s i.i.d. and } \mathbb{E}\Big(\varepsilon(0)\varepsilon^\top(0)\Big) = I$$

$\perp$ **means: 'independence'.**

**A (stochastic) system**

**:= a collection of processes = the 'behavior'**

# ARMAX

**Consider the difference eq'ns**

$$W_0 w(t) + W_1 w(t+1) + \cdots + W_{\mathtt{L}} w(t+\mathtt{L})$$
$$= E_0 \varepsilon(t) + E_1 \varepsilon(t+1) + \cdots + E_{\mathtt{L}} w(t+\mathtt{L})$$

$\mathtt{L} =$ **the lag. In shorthand:**

$$\boxed{W(\sigma)w = E(\sigma)\varepsilon} \qquad \textbf{(ARMAX)}$$

**with** $W, E$ **real polynomial matrices;**

$\sigma =$ **the 'shift':** $\quad (\sigma f)(t) := f(t+1)$.

# ARMAX

**Consider the difference eq'ns**

$$W(\sigma)w = E(\sigma)\varepsilon \qquad \textbf{(ARMAX)}$$

**The stochastic system consisting of all processes** $w$ **satisfying (ARMAX) with** $\varepsilon$ **white noise**

**is called the ARMAX system** $(W, E)$**.**

# ARMAX

**Consider the difference eq'ns**

$$\boxed{W(\sigma)w = E(\sigma)\varepsilon} \qquad \textbf{(ARMAX)}$$

**WLOG** $W$ **of full row rank** $\rightsquigarrow$ **the difference eq'ns**

$$\boxed{Y(\sigma)y + U(\sigma)u = E(\sigma)\varepsilon}$$

**with** $Y, U, E$ **real polynomial matrices,** $Y$ **square,** $\det(Y) \neq 0$**;**

**Under 'generic' conditions,** $u$ **is free:** $\forall u \ \exists\, y$ **... ,**
$u$ **is an 'exogeneous' input;** $y$ **an 'endogenous' output.**

# AR-MA-X

**WLOG:** $W$ **full row rank in** $W(\sigma)w = E(\sigma)\varepsilon$**.**

**Refine the ARMAX notation, by factoring out** $A$**, to:**

$$A(\sigma)(R(\sigma)w) = M(\sigma)\varepsilon$$

$A, R, M$ **real polynomial matrices,**

$A$ **square,** $\det(A) \neq 0$**,**

$R$ **left prime (i.e.** $R(\lambda)$ **of full row rank** $\forall \lambda \in \mathbb{C}$**).**

# AR-MA-X

**Refine the ARMAX notation, by factoring out $A$, to:**

$$A(\sigma)(R(\sigma)w) = M(\sigma)\varepsilon$$

**Note that $v = R(\sigma)w$ is an ARMA process, satisfying**

$$A(\sigma)v = E(\sigma)\varepsilon.$$

**Call $A$ the AR-part, $M$ the MA-part, $R$ the X-part.**

$G = P^{-1}Q \quad (R = [\,P \;\; Q\,])$ **= tf f'n of the 'deterministic part'.**

**$\exists$ a 'classification up to equivalence issue' for $(R, A, M)$.**

# System identification

# AR-MA-X SYSID

**Observe**

$$\tilde{w} = (\tilde{w}(1), \tilde{w}(2) \ldots, \tilde{w}(T))$$

**for simplicity, today,**

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

**SYSID pbm : Estimate one, all, a special, ARMAX system**

$$(W, E) \quad \textbf{or} \quad (R, A, M)$$

**to which the observed $\tilde{w}$ belongs.**

# AR-MA-X system ID

**Estimate** $(W,E)$, **or** $(R,A,M)$ **from observed**

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

**Today, we explain how to obtain the 'X-part': how to compute**

$$\tilde{w} \mapsto R$$

$$\boxed{\tilde{w} \longmapsto R}$$

## $\tilde{w} \mapsto R$

**Assume** $\tilde{w} = \begin{bmatrix} \tilde{u} \\ \tilde{y} \end{bmatrix}$ **and a representation**

$$\boxed{A(\sigma)\left(Y(\sigma)\tilde{y} + U(\sigma)\tilde{u}\right) = M(\sigma)\varepsilon}$$

**with** $\tilde{u} \perp \varepsilon$**.**

$$A(\sigma)\left(P(\sigma)\tilde{y} + Q(\sigma)\tilde{u}\right) = A(\sigma)\left(R(\sigma)\tilde{w}\right) = M(\sigma)\varepsilon$$

$$\Rightarrow \qquad R(\sigma)\tilde{w} = \sum_{t=-\infty}^{+\infty} H(t)\ \sigma^t \varepsilon$$

$\Rightarrow$ **(since** $\varepsilon \perp \tilde{u}$**)**

$$\boxed{R(\sigma)\tilde{w}\ \perp\ \tilde{u}.}$$

# $\tilde{w} \mapsto R$

**Idea:** **the linear combinations of the rows of the observed**

$$\begin{bmatrix} \tilde{w}(1) & \tilde{w}(2) & \tilde{w}(3) & \cdots & \tilde{w}(t) & \cdots \\ \tilde{w}(2) & \tilde{w}(3) & \tilde{w}(4) & \cdots & \tilde{w}(t+1) & \cdots \\ \tilde{w}(3) & \tilde{w}(4) & \tilde{w}(5) & \cdots & \tilde{w}(t+2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

**that are orthogonal to the rows of the observed**

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

**determine** $R$**.**

# $\tilde{w} \mapsto R$

$$\begin{bmatrix} \tilde{w}(1) & \tilde{w}(2) & \tilde{w}(3) & \cdots & \tilde{w}(t) & \cdots \\ \tilde{w}(2) & \tilde{w}(3) & \tilde{w}(4) & \cdots & \tilde{w}(t+1) & \cdots \\ \tilde{w}(3) & \tilde{w}(4) & \tilde{w}(5) & \cdots & \tilde{w}(t+2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

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

**$\exists$ an $\infty$ number of 'orthogonalizing' linear combinations.**

**What special structure do they have, so that they are determined by a finite number of them, $\cong$ $R$ ?**

**Is there a way to limit the number of rows?**

# Modules

**... see my website ...**

# Modules

**A module can be thought of as 'a vector space over a ring'.**

**A mathematician's favorite example:**

**the module** $\mathbb{Z} + \sqrt{2}\,\mathbb{Z} + \sqrt{5}\,\mathbb{Z}$ **over the ring** $\mathbb{Z}$

**A system theorist's favorite example:**

**the module** $\mathbb{R}^{\mathtt{n}}[\xi]$ **over the ring** $\mathbb{R}[\xi]$

$\mathbb{R}^{\mathtt{n}}[\xi]$**:= the** $\mathtt{n}$**-dimensional vectors of polynomials with real coefficients, in the indeterminate** $\xi$**.**

# Modules

A **submodule** of $\mathbb{R}^{\mathtt{n}}[\xi]$: subset that is also a module over $\mathbb{R}[\xi]$.

E.g., for given polynomial vectors $v_1, v_2, \ldots, v_{\mathtt{k}}$, all sums

$$p_1 v_1 + p_2 v_2 + \cdots + p_{\mathtt{k}} v_{\mathtt{k}}$$

$p$'s polynomials. The submodule **'generated by'** $v_1, v_2, \ldots, v_{\mathtt{k}}$.

**Fact:** Every submodule of $\mathbb{R}^{\mathtt{n}}[\xi]$ is like this: **'finitely generated'**.

**Fact:** Number of generators $\leq \mathtt{n}$. (mimimum =: the dimension)

# Modules

**A submodule of $\mathbb{R}^{\mathtt{n}}[\xi]$: subset that is also a module over $\mathbb{R}[\xi]$.**

**A submodule of $\mathbb{R}^{\mathtt{n}}[\xi]$ is said to be slim if it does not contain other submodules of the same dimension**

$$\Leftrightarrow V = \begin{bmatrix} v_1 & v_2 & \cdots & v_{\mathtt{k}} \end{bmatrix} \text{ right prime.}$$

**slim:** $\begin{bmatrix} * \\ \vdots \\ * \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ **not slim:** $p(\xi)\,\mathbb{R}[\xi]$.

# Modules

**Submodules of $\mathbb{R}^{\mathtt{n}}[\xi]$ are of great importance in system theory:**

**linear time-inv. diff. systems** $\overset{1:1}{\longleftrightarrow}$ **submodules of $\mathbb{R}^{\mathtt{n}}[\xi]$**

**controllable LTIS** $\overset{1:1}{\longleftrightarrow}$ **slim submodules**

# Modules

**Submodules of $\mathbb{R}^{\mathtt{n}}[\xi]$ are of great importance in system theory:**

**The 'left' or 'right' kernel of any Hankel matrix**

$$\begin{bmatrix} H(1) & H(2) & H(3) & \cdots & H(t'') & \cdots \\ H(2) & H(3) & H(4) & \cdots & H(t''+1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ H(t') & H(t'+1) & H(t'+2) & \cdots & H(t'+t''-1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$\cong$ **a submodule of** $\mathbb{R}^{\text{coldim}(H)}[\xi]$ **or** $\mathbb{R}^{\text{rowdim}(H)}[\xi]$**:**

$\rightsquigarrow$ **effectively not $\infty$-dimensional, but**

$\leq \text{rowdim}(H)$**- or** $\text{coldim}(H)$**-dimensional!**

# The orthogonalizers

# The orthogonalizers

**Let** $w = \begin{bmatrix} u \\ y \end{bmatrix}$ **be a process. Say,** $\mathtt{w}$**-dimensional.**

$n \in \mathbb{R}^{\mathtt{w}}[\xi]$ **is an orthogonalizer (for** $w$ **w.r.t** $u$**) if**

$$n^\top(\sigma)w \perp u.$$

**i.e. a linear combination of the components of the process** $w$ **and its shifts which becomes independent of the process** $u$**.**

# The orthogonalizers

**Let** $w = \begin{bmatrix} u \\ y \end{bmatrix}$ **be a process. Say,** $\mathtt{w}$**-dimensional.**

$n \in \mathbb{R}^{\mathtt{w}}[\xi]$ **is an orthogonalizer (for** $w$ **w.r.t** $u$**) if**

$$n^\top(\sigma)w \perp u.$$

**<u>Ex.</u>: the transpose of the rows of** $R$**, since**

$$A(\sigma)\big(R(\sigma)w\big) = M(\sigma)\varepsilon \;\Rightarrow R(\sigma)w \;\perp\; u.$$

$\Rightarrow$ **every element of the module generated by these.**
**Is this all? Are there no other orthogonalizers?**

# The orthogonalizers

**Let $w = \begin{bmatrix} u \\ y \end{bmatrix}$ be a process. Say, $\mathtt{w}$-dimensional.**

**Theorem:**

1. **The orthogonalizers for $w$ w.r.t. $u$ form a submodule of $\mathbb{R}^{\mathtt{w}}[\xi]$.**
2. **In fact, a slim one.**
3. **If $w = \begin{bmatrix} u \\ y \end{bmatrix}$ and $u$ is ‘ persistently exciting’, then it is precisely the submodule generated by the transposes of the rows of $R$.**

# The orthogonalizers

Let $w = \begin{bmatrix} u \\ y \end{bmatrix}$ be a process. Say, $\mathtt{w}$-dimensional.

**Theorem:**

1. The orthogonalizers for $w$ w.r.t. $u$ form a submodule of $\mathbb{R}^{\mathtt{w}}[\xi]$.
2. In fact, a slim one.
3. If $w = \begin{bmatrix} u \\ y \end{bmatrix}$ and $u$ is ' persistently exciting', then it is precisely the submodule generated by the transposes of the rows of $R$.

**Proof:** 1. is easy. 2. uses ergodicity! 3. a bit of module theory.

$$\boxed{\tilde{w} \longmapsto R}$$

# $\tilde{w} \mapsto R$

**Find (module basis for) the linear combinations of the rows of**

$$\begin{bmatrix} \tilde{w}(1) & \tilde{w}(2) & \tilde{w}(3) & \cdots & \tilde{w}(t) & \cdots \\ \tilde{w}(2) & \tilde{w}(3) & \tilde{w}(4) & \cdots & \tilde{w}(t+1) & \cdots \\ \tilde{w}(3) & \tilde{w}(3) & \tilde{w}(5) & \cdots & \tilde{w}(t+2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

**that are orthogonal to the rows of**

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

# $\tilde{w} \mapsto R$

**Find the linear combinations of the rows of**

$$W := \begin{bmatrix} \tilde{w}(1) & \tilde{w}(2) & \tilde{w}(3) & \cdots & \tilde{w}(t) & \cdots \\ \tilde{w}(2) & \tilde{w}(3) & \tilde{w}(4) & \cdots & \tilde{w}(t+1) & \cdots \\ \tilde{w}(3) & \tilde{w}(3) & \tilde{w}(5) & \cdots & \tilde{w}(t+2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

**that are orthogonal to the rows of**

$$U := \begin{bmatrix} \tilde{u}(1) & \tilde{u}(2) & \tilde{u}(3) & \cdots & \tilde{u}(t) & \cdots \\ \tilde{u}(2) & \tilde{u}(3) & \tilde{u}(4) & \cdots & \tilde{u}(t+1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

**!! Compute (a module basis for) the left kernel of $WU^\top$.**

# $\tilde{w} \mapsto R$

**Find the linear combinations of the rows of**

$$W := \begin{bmatrix} \tilde{w}(1) & \tilde{w}(2) & \tilde{w}(3) & \cdots & \tilde{w}(t) & \cdots \\ \tilde{w}(2) & \tilde{w}(3) & \tilde{w}(4) & \cdots & \tilde{w}(t+1) & \cdots \\ \tilde{w}(3) & \tilde{w}(3) & \tilde{w}(5) & \cdots & \tilde{w}(t+2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

**that are orthogonal to the rows of**

$$U := \begin{bmatrix} \tilde{u}(1) & \tilde{u}(2) & \tilde{u}(3) & \cdots & \tilde{u}(t) & \cdots \\ \tilde{u}(2) & \tilde{u}(3) & \tilde{u}(4) & \cdots & \tilde{u}(t+1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

**¿¿ Can we limit the number of rows that are needed ??**
**Yes, provided we assume a known bound on the lags.**

## $\tilde{w} \mapsto R$

**It is easy to prove that this, applied to a finite time-series, yields a consistent algorithm.**

**Note: no stability needed** $P$ **for** $R = [\, P \;\; Q \,]$**.**

## $\tilde{w} \mapsto A, M$

**Once we have $R$, we can compute the process**

$$\boxed{\tilde{a} = R(\sigma)\tilde{w}}$$

**This is an ARMA process.**

**Thinking in modules proceeds to $A$ and $M$.**

# Recapitulation

# Conclusion

**Linear algebra on the Hankel matrix of the data** $\rightsquigarrow$

**the module of orthogonalizers**

**with a limited depth (**$\mathtt{L}$**), contains the laws of the deterministic part of the system.**

***R* can be readily indentified**

**Reduces ARMAX SYSID to ARMA SYSID.**

**Thank you**

**Thank you**

**Thank you**

**Thank you**

**Thank you**

**Thank you**

**Thank you**

**Thank you**