Identification of Network Connected systems with "unknown" communication links

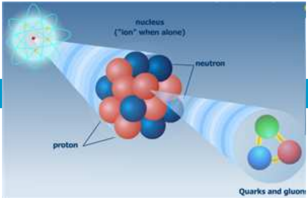
This lecture: Focus on Identifying KrARX (Quarks) models

Baptiste Sinquin and Michel Verhaegen [1]

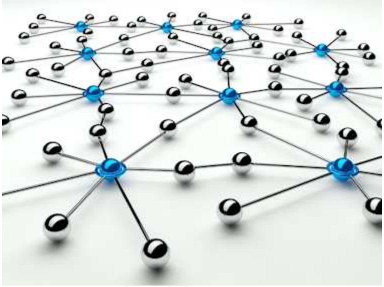
[1]Control for Scientific Imaging Systems, (CSI) Group
Delft Center for Systems and Control

[2]$\mathcal{L}$ecture at the ERNSI Workshop Cison di Valmarino 2016

26 September 2016



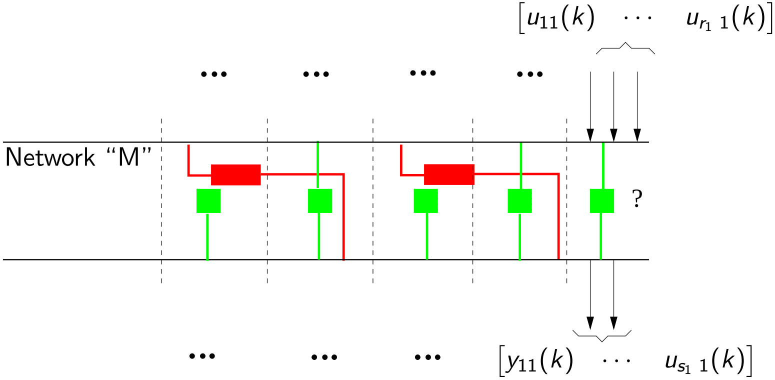
Structure in Sensor Networks



- *A (large-2D) network of (identical) sensors that are spatially and temporally coupled*
- Goal: Find spatial "topology" and temporal order PLUS the "parameters" of the "structured" spatial-temporal model that governs the sensor dynamics.

Static Interaction between two data vectors

$$y(k) = Mu(k)$$

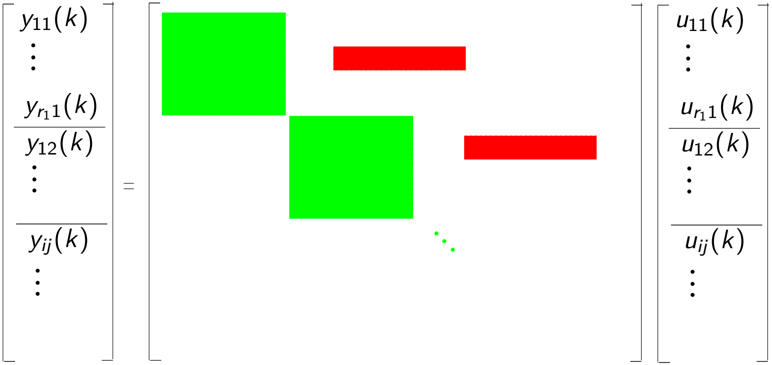


Static Interaction between two data vectors

$$y(k) = Mu(k)$$

Features:

1. Data is partitioned:

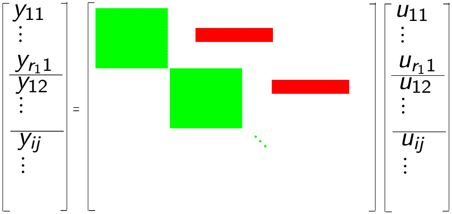


Static Interaction between two data vectors

$$y(k) = Mu(k)$$

Features:

1. Data is partitioned:



2. Structured matrices:
   - Toeplitz, Semi-Sequentially Separable, etc.
   - Low rank plus sparse (Mattia Zorzi, Alessandro Chiuso, 2016)
   - α-*decomposable matrices*
   - . . .

α-decomposable matrices

A "generalization" of the class of system matrices of decomposable systems [1][1] introduced in [2][2]

Definition

Let $\mathscr{P}$ be an $N \times N$ "pattern matrix". Define $\beta_j = \sum_{i=1}^{j} N_i$ (with $\beta_0 = 0$) and $I_{[a_1:a_2]}$ as an $N \times N$ diagonal matrix which contains 1 in the diagonal entries of indices from a_1 to a_2 (included) and 0 elsewhere, then an α-decomposible matrix (for a given α) is a matrix of the following kind:

$$\mathscr{M} = \sum_{i=1}^{\alpha} (I_{[\beta_{i-1}+1:\beta_i]} \otimes M_a^{(i)}) + \sum_{i=1}^{\alpha} (I_{[\beta_{i-1}+1:\beta_i]} \mathscr{P} \otimes M_b^{(i)})$$

The matrices $M_a^{(i)}$ are the diagonal blocks of $\mathscr{M}$, while the matrices $M_b^{(i)}$ constitute the off-diagonal blocks, according to the structure of $\mathscr{P}$.

[1] [1] Massioni, Verhaegen, "Distributed Control for Identical Dynamically Coupled Systems: A Decomposition Approach" IEEE-TAC 54(1):124-135, 2009

[2] [2] Massioni, "Distributed control for alpha-heterogeneous dynamically coupled systems" Syst & Con. Let. 72:30-35, 2014

Sum of Kronecker Products [Van Loan, LA 2004]

Definition operator $\mathscr{R}(.)$

Let a matrix A be partitioned as:

$$\left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right] = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$$

then the "reshuffle" operator $\mathscr{R}(.)$ applied on A yields:

$$\mathscr{R}(A) = \left[\begin{array}{cccc} a_{11} & a_{21} & a_{12} & a_{22} \\ \hline a_{31} & a_{41} & a_{32} & a_{42} \\ \hline a_{13} & a_{23} & a_{14} & a_{24} \\ \hline a_{33} & a_{43} & a_{34} & a_{44} \end{array}\right] = \left[\begin{array}{c} \mathrm{vec}(A_{11})^T \\ \hline \mathrm{vec}(A_{21})^T \\ \hline \mathrm{vec}(A_{12})^T \\ \hline \mathrm{vec}(A_{22})^T \end{array}\right]$$

Sum of Kronecker Products [Van Loan, LA 2004]

Theorem: The Kronecker Product SVD

Let

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1,c_2} \\ \vdots & \ddots & \vdots \\ A_{r_2,1} & \cdots & A_{r_2,c_2} \end{bmatrix} \quad A_{i_2,j_2} \in \mathbb{R}^{r_1 \times c_1}$$

then there exists $U_1, \cdots, U_{r_{KP}} \in \mathbb{R}^{r_2 \times c_2}$, $V_1, \cdots, V_{r_{KP}} \in \mathbb{R}^{r_1 \times c_1}$, and scalars $\sigma_1 \geq \cdots \geq \sigma_{r_{KP}} > 0$, such that,

$$A = \sum_{\ell=1}^{r_{KP}} \sigma_\ell U_\ell \otimes V_\ell$$

With the sets $\{\mathrm{vec}(U_\ell)\}$ and $\{\mathrm{vec}(V_\ell)\}$ orthonormal and r_{KP} is the Kronecker rank of A w.r.t. block-partioning of A.

Kronecker (product) ARX models (KrARX = “quarks”)

Definition

Let the data of a 2D “square” sensor grid be stored into the matrix:

$$S_k = \begin{bmatrix} s_{1,1}(k) & s_{1,2}(k) & \cdots & s_{1,N}(k) \\ s_{2,1}(k) & s_{2,2}(k) & & s_{2,N}(k) \\ \vdots & & \ddots & \vdots \\ s_{N,1}(k) & s_{N,2}(k) & \cdots & s_{N,N}(k) \end{bmatrix} \in \mathbb{R}^{pN \times N}$$

then the KrARX model of temporal order n and “spatial” order r is given as:

$$\mathrm{vec}\Big(S_k\Big) = \sum_{i=1}^{n} A_i \mathrm{vec}\Big(S_{k-i}\Big) + v(k)$$

with $v(k)$ a zero-mean white noise sequence and $A_i \in \mathbb{R}^{pN^2 \times pN^2}$ given as $\sum_{j=1}^{r} M(b_i^{(j)})^T \otimes M(a_i^{(j)})$. For affine parametrizations $M(b_i^{(j)})$ and $M(a_i^{(j)})$.

Matrix Format of KrARX

Lemma

The KrARX model of order (n, r) can also be written into the following Matrix Format:

$$S_k = \sum_{i=1}^{n} \left(\sum_{j=1}^{r} M(a_i^{(j)}) S_{k-i} M(b_i^{(j)}) \right) + V_k$$

with $\mathrm{vec}\left(V_k\right) = v(k)$. (Proof follows from $\mathrm{vec}\left(XYZ\right) = \left(Z^T \otimes X\right)\mathrm{vec}\left(Y\right)$)

This can also be written explicitly as,

$$S_k = \sum_{i=1}^{n} \begin{bmatrix} M(a_i^{(1)}) & \cdots & M(a_i^{(r)}) \end{bmatrix} \begin{bmatrix} S_{k-i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S_{k-i} \end{bmatrix} \begin{bmatrix} M(b_i^{(1)}) \\ \vdots \\ M(b_i^{(r)}) \end{bmatrix} + V_k$$

Example: KrARX of order (1,1) is given as:

$$S_k = B^T S_{k-1} A + V_k$$

Estimation Method 1

For a given choice of the temporal order n, the estimation of the KrARX coefficient matrix A_i can be done via the following *multi-criteria optimization problem*:

$$\min_{A_i} \sum_{k=n+1}^{N_{\mathrm{dat}}} \|\mathrm{vec}\Big(S_k\Big) - \sum_{i=1}^{n} A_i \mathrm{vec}\Big(S_{k-i}\Big)\|_F^2 + \lambda \mathrm{rank}\Big(\mathscr{R}(A_i)\Big)$$

for the "trade-off" parameter λ.

Advantages:

1. Convex relaxation via the nuclear norm.
2. "Automatic approximation" of the spatial order r

Disadvantages:

1. "Curse of dimensionality" ($A_i \in \mathbb{R}^{pN^2 \times pN^2}$)

Estimation Method 2: ALS

Illustrated for KrARX of order (1,1):

$$S_k = B^T S_{k-1} A + V_k \quad S_k, A, B \in \mathbb{R}^{N \times N}$$

Hereby we denote the ℓ-th column of A and B resp. as a_ℓ and b_ℓ, then the *alternating iteration*, consists (per iteration) of the following three steps:

1. *Step 1:* For a given $\hat{B}$ (estimate):

$$\forall \ell = 1, \cdots, N : \min_{a_\ell} \sum_{i=1}^{N_{\mathrm{dat}}} \|S_i(:, \ell) - \hat{B}^T S_{i-1} a_\ell\|_2^2 \quad \rightarrow \hat{\alpha}_\ell$$

2. *Step 2:* Normalization $\hat{a}_\ell = \frac{\hat{\alpha}_\ell}{\|\hat{\alpha}_\ell\|_2}$
3. *Step 3:* For a given $\hat{A}$ (estimate):

$$\forall \ell = 1, \cdots, N : \min_{b_\ell} \sum_{i=1}^{N_{\mathrm{dat}}} \|S_i(\ell, :) - b_\ell^T S_{i-1} \hat{A}\|_2^2 \quad \rightarrow \hat{b}_{\ell+1}$$

Convergence Analysis ALS

Consider the estimation of the *ij*-the entry in the KrARX (Matrix form) of order $(1, 1)$:

$$s_{ij}(k) = b_i^T S_{k-1} a_j + v_{ij}(k)$$

Then with a value of $\hat{b}_i(s)$ at iteration s the 3-step procedure can be summarized as:

$$\hat{b}_i(s+1) = \mathscr{F}_3(\mathscr{F}_2(\mathscr{F}_1(\hat{b}_i(s)) := \mathscr{F}(\hat{b}_i(s))$$

And the question of convergence boils down to the existance of a unique fixed point of the composite map $\mathscr{F}(.)$. [3]

[3] Based on the work of Guoqi Li, et. al, "Fixed point iteration in identifying bilinear models," System & Control Let 83:28-37, 2015.

Convergence Analysis ALS

Theorem [Guoqi Li, et. al]

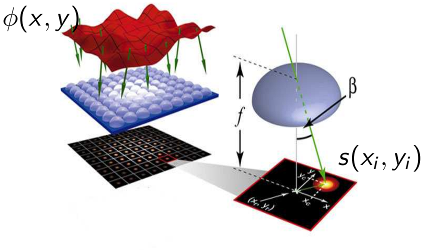
Under the assumptions that,

1. The noise $v_{ij}(k)$ is i.i.d. with mean zero and finite variance.
2. The least squares problem in Step 1 and 3 have a unique solution.
3. Either $\|b_i\|_2$ or $\|a_j\|_2$ is known and the first nonzero entry of b_i or a_j is positive.

The mapping $\mathscr{F}(.) : X_b \to X_b$ is a contraction mapping on X_b when $N_{\mathrm{dat}} \to \infty$ and *asymptotically* has a unique fixed point on X_b which corresponds to the true parameter b.

Wavefront sensor Prediction

Schematic Shack-Hartmann Sensor



Let the time-dependency of $s(x_i, y_i)$ be denoted as $s_{i,j}(k)$, then this 2D-sensor provides the data in matrix form:

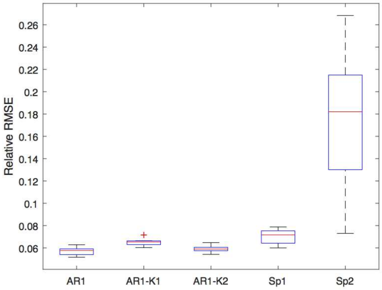
$$S(k)$$

Simulation [Beghi, et. al, 2008]

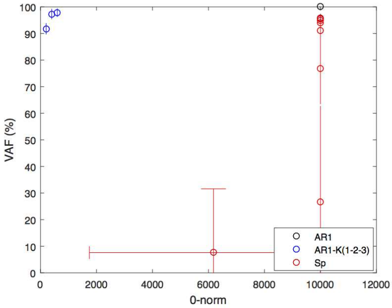
- Kolmogorov model for spatial correlation
- Frozen flow temporal dynamics
- $N \times N$ lenslet array
- $N_t = 5000$.

Reconstruction Performance versus model complexity

n = 1 AR models



VAF versus model complexity (0-norm)



Delft Center for Systems and Control

European Research Council
Funding for top researchers from anywhere in the world

Control
Scientific
Instruments

Just the beginning ...

- Non-rectangular measurement grids
- AR(MA)X model identification
- Subspace model identification
- etc.

B. Sinquin and M. Verhaegen, “Kronecker-based modeling of networks with unknown communication links,” http://arxiv.org/abs/1609.07518