Least squares optimal identification of LTI dynamical systems

Bart De Moor

KU Leuven
Dept.EE: ESAT - STADIUS

bart.demoor@kuleuven.be

KU LEUVEN

Outline

Outline

- Eigenvalues and vectors: For matrix $A \in \mathbb{R}^{n \times n}$:

$$Ax = x\lambda \, , x \in \mathbb{C}^n, \lambda \in \mathbb{C}, x \neq 0.$$

- Characteristic equation - fundamental theorem of algebra

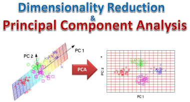
$$p(\lambda) = \det(\lambda I_n - A) = \lambda^n + \alpha_1 \lambda^{n-1} + \ldots + \alpha_{n-1}\lambda + \alpha_n = 0.$$

- Since Galois, for $n \geq 5$: no solution in radicals $\Longrightarrow$ iterative algorithms
- Eigenvalue decomposition - Jordan Canonical Form (JCF)

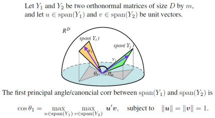
$$A = XJX^{-1}.$$

- Spectra of
  - Algebras
  - Operators: $d\, e^{(\alpha \pm j\beta t)}/dt = (\alpha \pm j\beta t)e^{(\alpha \pm j\beta t)}$
  - Geometrical shapes: moments inertia, eigenfrequencies, modal shapes, ...
  - ...

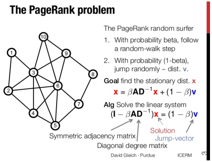
KU LEUVEN



PCA



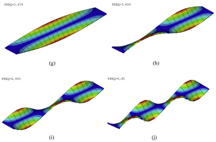
Can. Corr./Principal Angles



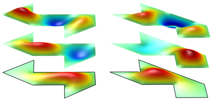
Graph spectral analysis

$$\frac{\partial^2 y}{\partial t^2} = \frac{1}{v_w^2}\frac{\partial^2 y}{dt^2}$$

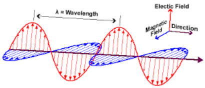
Wave equation



Modal shapes



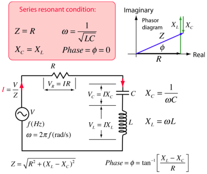
Hear the shape of a drum?



Maxwell's laws

1. $\nabla \cdot \mathbf{D} = \rho_V$
2. $\nabla \cdot \mathbf{B} = 0$
3. $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$
4. $\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$

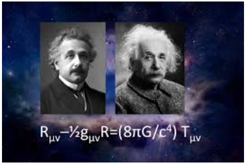
Maxwell's field equations



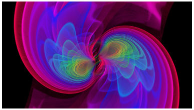
RLC circuits

$$H(t)|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle$$

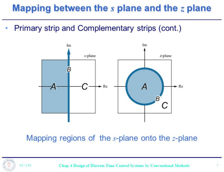
Schrodinger equation



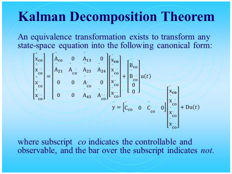
Matter curves spacetime moves matter



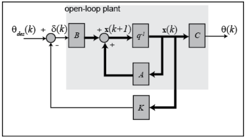
Gravitational waves



Stability



Controllability/observability

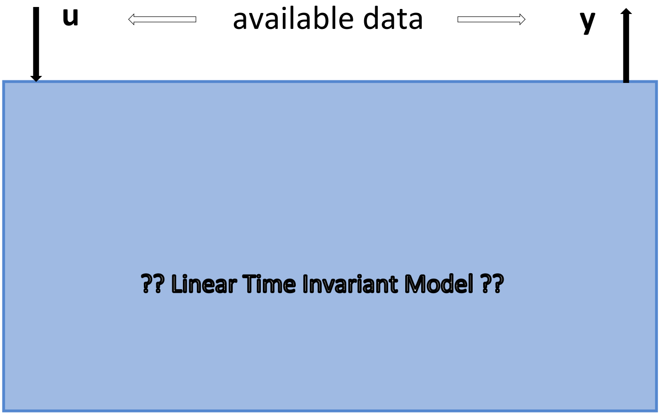


Pole placement

Observers	Kalman Filter Riccati Hamil. EVP	H_∞-filter Riccati Sympl. EVP
Control	LQR Riccati Hamil. EVP	H_∞-control Riccati Sympl. EVP

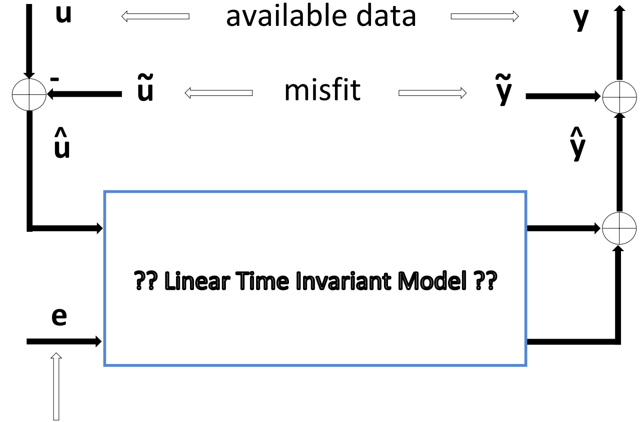
Kalman, Willems, bdm

Outline



Hypotheses non fingo. Newton.
Let the data speak for themselves. Kalman.

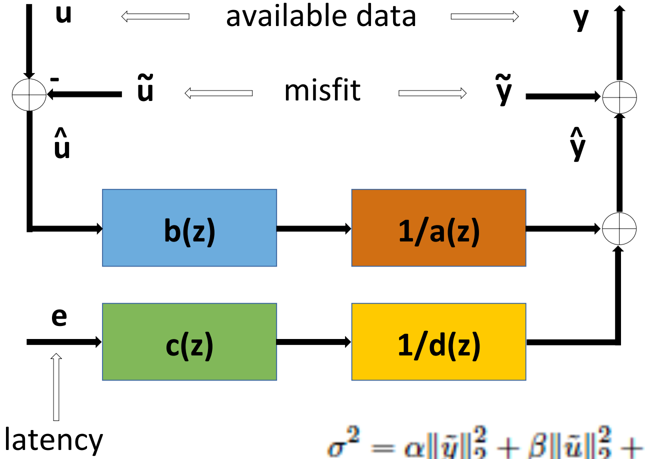
Misfit-Latency LTI Models



$$\sigma^2 = \alpha\|\tilde{y}\|_2^2 + \beta\|\tilde{u}\|_2^2 + \gamma\|\tilde{e}\|_2^2,$$

Models are a matter of deduction, not inspiration. Jan Willems.

Misfit-Latency LTI Models



$$\sigma^2 = \alpha\|\tilde{y}\|_2^2 + \beta\|\tilde{u}\|_2^2 + \gamma\|\tilde{e}\|_2^2,$$

Errors using inadequate data are much less
than those using no data at all.
Charles Babbage.

How nonlinear is least squares linear system identification ?

	Nonlinearity	'Heuristic' remedy
State space	$x_{k+1} = \mathbf{Ax_k} + Bu_k$ Unknown $A \times x_k$	Subspace: Oblique projection and SVD
PEM	Unknown parameters $\times$ latency input e	Nonlinear optimization
EIV	Unknown parameters $\times$ misfits $\tilde{u}, \tilde{y}$	Instrumental Variables

But:

All 'nonlinearities' are sums of products of unkowns.

Hence multivariate polynomial.

- All 'nonlinearities' are multivariate polynomial and occur in the model and data equations
- The objective function (sum-of-squares) is polynomial
- Hence, the problem is a multivariate polynomial optimization problem: multivariate polynomial objective function and constraints
- Taking derivatives of multivariate polynomials (first order optimality) results in a set of multivariate polynomials equal to zero
- The roots of this set are local and global minima and maxima, and saddle points
- We only need the one or several roots that correspond to the global minimum of the objective function.
- Evaluate a multivariate polynomial (the objective function - the critical polynomial) over the roots

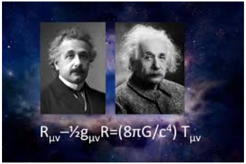
How to find the roots of a set of multivariate polynomials ?

What do we mean by 'solution' and 'to solve' ?

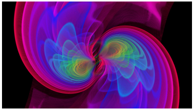
- When do we consider a mathematical problem to be solved ?
  - A conjecture is 're'-solved: e.g. Fermat's Last Theorem; A mathematical proof;
  - There is an analytical solution: e.g. linear ODEs
  - Reduction to a set of linear equations
  - Reduction to a convex optimization problem
  - Reduction to an eigenvalue problem
  - ....
- The computational complexity can still be deceiving (e.g. worst case behavior of the simplex method for LP).
- Set of linear equations and/or EVP: 50 years of spectacular progress in numerical linear algebra (Matlab, sparsity, iterative methods, large scale (HPC), ...)

$$H(t)|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle$$

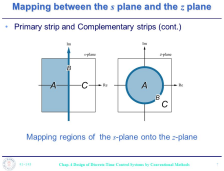
Schrodinger equation



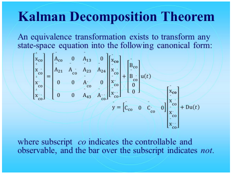
Matter curves spacetime moves matter



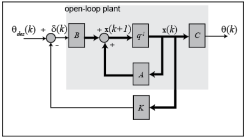
Gravitational waves



Stability

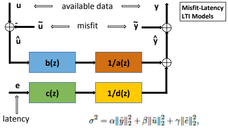


Controllability/observability



Pole placement

Observers	Kalman Filter Riccati Ham. EVP
Control	LQR Riccati Ham. EVP



LS LTI System ID = EVP !

Outline

Least squares optimal system identification of LTI models is an eigenvalue problem

- Realization theory in 1D and shift-invariant subspaces
- Realization theory in nD and multi-shift-invariant subspaces
- Roots in 1 variable: The null spaces of Toeplitz and Sylvester matrices are shift-invariant
- Roots in n variables: The null spaces of (quasi-Toeplitz) Macaulay and block Macaulay matrices are multi-shift-invariant
- Representative ID cases: MA, LS realization, dynamic TLS

Outline

1D realization theory

Singular autonomous system, states $x_k \in \mathbb{R}^n$, outputs $y_k \in \mathbb{R}^l$, singular E:

$$\begin{aligned} Ex_{k+1} &= Ax_k, \\ y_k &= Cx_k, \end{aligned}$$

Convert $(E, A) \to (PEQ, PAQ)$ to Weierstrass Canonical Form (WCF) with regular state $x_k^R \in \mathbb{R}^{n_1}$, singular state $x_k^S \in \mathbb{R}^{n_2}$, $n_2 = n - \text{rank}(E)$.
Rearrange in an a-causal autonomous system, with E_1 nilpotent with nilpotency index ν: $E^k = 0, k \geq \nu$:

$$\begin{aligned} x_{k+1}^R &= A_1 x_k^R & \to \text{causal}, \\ x_{k-1}^S &= E_1 x_k^S & \to \text{anti} - \text{causal}, \\ y_k &= C_R x_k^R + C_S x_k^S & \to \text{a} - \text{causal}. \end{aligned}$$

Characteristic polynomial with n_1 affine ('finite') and n_2 poles at infinity:

$$\det\left[\begin{pmatrix} I_{n_1} & 0 \\ 0 & E_1 \end{pmatrix} z - \begin{pmatrix} A_1 & 0 \\ 0 & I_{n_2} \end{pmatrix}\right] = \det(zI_{n_1} - A_1)\det(zE_1 - I_{n_2}) = 0.$$

Realization problem:

Given $y^T = (y_0 \; y_1 \; \ldots y_{N-1})$: find n, A_1, E_1, x_k^R and x_k^S.

Factorize $pl \times q$ (block) Hankel matrix $(N = p + q - 1)$ e.g. via SVD:

$$Y = \begin{pmatrix} y_0 & y_1 & y_2 & \cdots & y_{q-2} & y_{q-1} \\ y_1 & y_2 & y_3 & \cdots & y_{q-1} & y_q \\ y_2 & y_3 & \cdots & \cdots & y_q & y_{q+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{p-2} & y_{p-1} & \cdots & \cdots & y_{N-3} & y_{N-2} \\ y_{p-1} & y_p & \cdots & \cdots & y_{N-2} & y_{N-1} \end{pmatrix} = \Gamma\Delta$$

$$= \left(\begin{array}{cc} C_R & 0 \\ C_R A_1 & 0 \\ \vdots & \vdots \\ C_R A_1^{n_1-1} & 0 \\ \hline C_R A_1^{n_1} & 0 \\ \vdots & \vdots \\ C_R A_1^{p-\nu-1} & 0 \\ \hline C_R A_1^{p-\nu} & C_S E_1^{\nu-1} \\ \vdots & \vdots \\ C_R A_1^{p-3} & C_S E_1^2 \\ C_R A_1^{p-2} & C_S E_1 \\ C_R A_1^{p-1} & C_S \end{array}\right) \begin{pmatrix} x_0^R & A_1 x_0^R & \cdots & \cdots & \cdots & \cdots & A_1^{N-p} x_0^R \\ 0 & \cdots & 0 & E_1^{\nu-1} x_{N-1}^S & \cdots & E_1 x_{N-1}^S & x_{N-1}^S \end{pmatrix}$$

$$\Gamma\, T = \begin{pmatrix} \overset{n_1}{\Gamma_1} & \overset{n_2}{\Gamma_2} \end{pmatrix} = \left(\begin{array}{cc} C_R & 0 \\ C_R A_1 & 0 \\ \vdots & \vdots \\ C_R A_1^{n_1-1} & 0 \\ \hline C_R A_1^{n_1} & 0 \\ \vdots & \vdots \\ C_R A_1^{p-\nu-1} & 0 \\ \hline C_R A_1^{p-\nu} & C_S E_1^{\nu-1} \\ \vdots & \vdots \\ C_R A_1^{p-3} & C_S E_1^2 \\ C_R A_1^{p-2} & C_S E_1 \\ C_R A_1^{p-1} & C_S \end{array}\right)$$

- $\mathrm{rank}(Y) = n =$ total number of poles
- $Y = \Gamma\Delta$ (e.g. via SVD); $\Gamma \in \mathbb{R}^{pl\times n}$, only unique up to within non-singular $T \in \mathbb{R}^{n\times n}$.
- 3 **row** zones in Γ independent of T:

$\leftarrow$ **I. First block rows:** *'Affine-pole'-zone*:
Rank increases with at least 1 per block up to block n_1 = number of affine poles;

$\leftarrow$ **II. Middle block rows:** *'Mind-the-gap'-zone*:
Rank does not increase;

$\leftarrow$ **III. Last block rows:** *'A-bout-du-souffle'-zone*:
Rank increases per block.

- T is a column compression (e.g. SVD): reduces column space of **first zone** to n_1 linear independent columns = number of affine poles.

The 'affine-pole'-column space is a **shift-invariant subspace**:

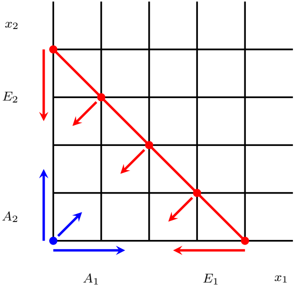
$$\underline{\Gamma}_1 \; A_1 \; = \overline{\Gamma}_1 = \begin{pmatrix} C_R \\ C_R A_1 \\ C_R A_1^2 \\ \vdots \\ C_R A_1^{p-3} \\ C_R A_1^{p-2} \end{pmatrix} A_1 = \begin{pmatrix} C_R A_1 \\ C_R A_1^2 \\ \vdots \\ C_R A_1^{p-3} \\ C_R A_1^{p-1} \end{pmatrix}$$

- Subspace is invariant after shifting up a block $\mathrm{Range}(\underline{\Gamma}_1) = \mathrm{Range}(\overline{\Gamma}_1)$ (if A_1 is nonsingular).
- Allows to find A_1 by solving set of linear equations, e.g. $A_1 = \underline{\Gamma}_1^{\dagger} \overline{\Gamma}_1$.
- Affine poles are eigenvalues of A_1
- A shift invariant subspace is determined by the eigenvalues of its shift A_1 (uniquely for $l = 1$, also by C_R for $l > 1$).

nD realization theory

nD singular multi-dimensional autonomous systems on discrete grids (here illustrated for $n = 2$, WCF already applied):

$$\begin{aligned} x^R_{k+1,l} &= A_1 x^R_{k,l}, \\ x^S_{k-1,l} &= E_1 x^S_{k,l}, \\ x^R_{k,l+1} &= A_2 x^R_{k,l}, \\ x^S_{k,l-1} &= E_2 x^S_{k,l}, \\ y_{k,l} &= C_R x^R_{k,l} + C_S x^S_{k,l}, \end{aligned}$$



with $A_1, A_2 \in \mathbb{R}^{n_1 \times n_1}$, $C_R \in \mathbb{R}^{l \times n_1}$, $C_S \in \mathbb{R}^{l \times n_2}$, $E_1, E_2 \in \mathbb{R}^{n_2 \times n_2}$, both nilpotent, $n = n_1 + n_2$. Commuting matrices (hence *Commutative Algebra*):

$$A_1 A_2 = A_2 A_1 \ , \ E_1 E_2 = E_2 E_1 .$$

Realization problem:

Given $y_{k,l}$. Find $n, n_1,\ A_1, A_2, C_R, C_S, E_1, E_2,\ x^R_{k,l},\ x^S_{k,l}$.

Factorize the generalized block Hankel matrix

$$
\begin{aligned}
Y &= \left(\begin{array}{c|cc|ccc|cc}
y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} & y_{30} & \cdots \\
\hline
y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} & y_{40} & \cdots \\
y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} & y_{31} & \cdots \\
\hline
y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} & y_{50} & \cdots \\
y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} & y_{41} & \cdots \\
y_{02} & y_{12} & y_{13} & y_{22} & y_{13} & y_{04} & y_{32} & \cdots \\
\hline
y_{30} & y_{40} & y_{31} & y_{50} & \cdots & \cdots & \cdots & \cdots \\
y_{21} & y_{31} & y_{22} & y_{41} & \cdots & \cdots & \cdots & \cdots \\
y_{12} & y_{22} & y_{13} & y_{32} & \cdots & \cdots & \cdots & \cdots \\
y_{03} & y_{13} & y_{04} & y_{23} & \cdots & \cdots & \cdots & \cdots \\
\hline
y_{40} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}\right) \\
&= \Gamma \; \Delta .
\end{aligned}
$$

Y is a quasi-block-Hankel-block matrix.

$$\Gamma\, T = \begin{pmatrix} \overset{n_1}{\Gamma_1} & \overset{n_2}{\Gamma_2} \end{pmatrix}$$

$$= \left(\begin{array}{cc} C_R & 0 \\ \hline C_R A_1 & 0 \\ C_R A_2 & 0 \\ \hline C_R A_1^2 & 0 \\ C_R A_1 A_2 & 0 \\ C_R A_2^2 & 0 \\ \hline \vdots & \vdots \\ \hline C_R A_1^{n_1-1} & 0 \\ C_R A_1^{n_1-2} A_2 & 0 \\ \vdots & \vdots \\ C_R A_2^{n_1-1} & 0 \\ \hline\hline C_R A_1^{n_1} & 0 \\ \vdots & \vdots \\ \hline\hline \vdots & * \\ \vdots & * \\ \hline \vdots & * \\ \vdots & * \end{array}\right)$$

- $\operatorname{rank}(Y) = n =$ state space dimension
- $Y = \Gamma\Delta$ (e.g. via SVD); $\Gamma \in \mathbb{R}^{pl\times n}$, only unique up to within non-singular $T \in \mathbb{R}^{n\times n}$.

- 3 **row** zones in Γ independent of T:

  $\leftarrow$ **I. First block rows:** *'Regular'-zone*:
  Rank increases with at least 1 per block up to block $n_1 =$ dimension of regular state space;

  $\leftarrow$ **II. Middle block rows:** *'Mind-the-gap'-zone*:
  Rank does not increase;

  $\leftarrow$ **III. Last block rows:** *'A-bout-du-souffle'-zone*:
  Rank increases per block.

- T is a column compression (e.g. SVD)

KU LEUVEN

The 'regular'-column space is a **multi-shift-invariant subspace**:

$$\underline{\Gamma}_1 \; A_1 \;= S_1\Gamma = \begin{pmatrix} C_R \\ \hline C_RA_1 \\ C_RA_2 \\ \hline C_RA_1^2 \\ C_RA_1A_2 \\ C_RA_2^2 \\ \hline \vdots \\ \hline C_RA_1^{p-2} \\ C_RA_1^{p-3}A_2 \\ \vdots \\ C_RA_2^{p-2} \end{pmatrix} A_1 = \begin{pmatrix} C_RA_1 \\ \hline C_RA_1^2 \\ C_RA_1A_2 \\ \hline C_RA_1^3 \\ C_RA_1^2A_2 \\ C_RA_1A_2^2 \\ \hline \vdots \\ \hline C_RA_1^{p-1} \\ C_RA_1^{p-2}A_2 \\ \vdots \\ C_RA_1A_2^{p-2} \end{pmatrix} \quad \text{and} \quad \underline{\Gamma}_1 \; A_2 \;= S_2\Gamma$$

- Selector matrix S_1 selects the block rows $(2, 4, 5, 7, 8, 9, \ldots)$.
- Selector matrix S_2 selects the block rows $(3, 5, 6, 8, 9, 10, \ldots)$.
- Allows to find A_1, A_2 by solving set of linear equations

$$A_1 = \underline{\Gamma}_1^{\dagger} S_1 \Gamma_1 \;\; \text{and} \;\; A_2 = \underline{\Gamma}_1^{\dagger} S_2 \Gamma_1 \; .$$

- A multi-shift invariant subspace is determined by the eigenvalues of its shifts A_1 and A_2 (uniquely for $l = 1$, also by C_R for $l > 1$).

Outline

Univariate polynomial of degree 3:

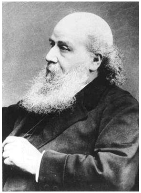
$$x^3 + a_1 x^2 + a_2 x + a_3 = 0,$$

having three distinct roots x_1, x_2 and x_3

$$\begin{bmatrix} a_3 & a_2 & a_1 & 1 & 0 & 0 \\ 0 & a_3 & a_2 & a_1 & 1 & 0 \\ 0 & 0 & a_3 & a_2 & a_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ x_1^2 & x_2^2 & x_3^2 \\ x_1^3 & x_2^3 & x_3^3 \\ x_1^4 & x_2^4 & x_3^4 \\ x_1^5 & x_2^5 & x_3^5 \end{bmatrix} = 0$$

- Banded Toeplitz; linear homogeneous equations
- Null space: (Confluent) Vandermonde structure
- Corank (nullity) = number of solutions
- Realization theory in null space: eigenvalue problem

Two univariate polynomials: common roots ?



James Joseph Sylvester

$$\begin{aligned} f(x) &= x^3 - 6x^2 + 11x - 6 = (x-1)(x-2)(x-3) \\ g(x) &= -x^2 + 5x - 6 = -(x-2)(x-3) \end{aligned}$$

$$\begin{array}{l} f(x) = 0 \\ x \cdot f(x) = 0 \\ g(x) = 0 \\ x \cdot g(x) = 0 \\ x^2 \cdot g(x) = 0 \end{array} \qquad \begin{array}{c} \begin{array}{ccccc} 1 & x & x^2 & x^3 & x^4 \end{array} \\ \begin{bmatrix} -6 & 11 & -6 & 1 & 0 \\ & -6 & 11 & -6 & 1 \\ -6 & 5 & -1 & & \\ & -6 & 5 & -1 & \\ & & -6 & 5 & -1 \end{bmatrix} \end{array} \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \\ x_1^2 & x_2^2 \\ x_1^3 & x_2^3 \\ x_1^4 & x_2^4 \end{bmatrix} = 0$$

where $x_1 = 2$ and $x_2 = 3$ are the common roots of f and g

- Nullity of Sylvester matrix = number of common zeros
- Null space = intersection of null spaces of two banded Toeplitz matrices = shift invariant subspace
- Common roots follow from realization theory in null space
- Notice 'double' Toeplitz-structure of Sylvester matrix

KU LEUVEN

The vectors in the Vandermonde kernel K obey a 'shift structure':

$$\begin{bmatrix} 1 & 1 \\ x_1 & x_2 \\ x_1^2 & x_2^2 \\ x_1^3 & x_2^3 \end{bmatrix} \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ x_1^2 & x_2^2 \\ x_1^3 & x_2^3 \\ x_1^4 & x_2^4 \end{bmatrix}$$

or

$$\underline{K}.D = S_1 K D = \overline{K} = S_2 K$$

The Vandermonde kernel K is not available directly, instead we compute Z, for which $ZV = K$. We now have

$$\begin{aligned} S_1 K D &= S_2 K \\ S_1 Z V D &= S_2 Z V \end{aligned}$$

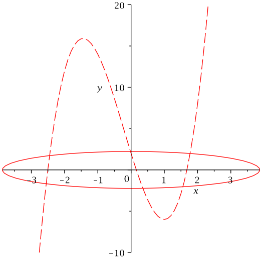
leading to the generalized eigenvalue problem

$$(S_2 Z) V = (S_1 Z) V D$$

Two polynomials in two variables

- Consider

$$\begin{cases} p(x,y) & = & x^2 + 3y^2 - 15 = 0 \\ q(x,y) & = & y - 3x^3 - 2x^2 + 13x - 2 = 0 \end{cases}$$



- Fix a monomial order, e.g., $1 < x < y < x^2 < xy < y^2 < x^3 < x^2y < \ldots$

- Construct quasi-Toeplitz Macaulay matrix M:

$$\begin{array}{c} \\ p(x,y) \\ q(x,y) \\ x \cdot p(x,y) \\ y \cdot p(x,y) \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & x & y & x^2 & xy & y^2 & x^3 & x^2y & xy^2 & y^3 \end{array} \\ \left[\begin{array}{cccccccccc} -15 & & & 1 & & 3 & & & & \\ -2 & 13 & 1 & -2 & & & -3 & & & \\ & -15 & & & & & 1 & & 3 & \\ & & -15 & & & & & 1 & & 3 \end{array}\right] \end{array} \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ \vdots \\ xy^2 \\ y^3 \end{pmatrix} = 0$$

KU LEUVEN

$$\begin{cases} p(x,y) &= x^2 + 3y^2 - 15 = 0 \\ q(x,y) &= y - 3x^3 - 2x^2 + 13x - 2 = 0 \end{cases}$$

Continue to enlarge M:

it #	form	1	x	y	x^2	xy	y^2	x^3	x^2y	xy^2	y^3	x^4	x^3y	x^2y^2	xy^3	y^4	x^5	x^4y	x^3y^2	x^2y^3	xy^4	y^5	→
$d=3$	p	−15			1		3																
	xp		−15					1		3													
	yp			−15					1		3												
	q	−2	13	1	−2			−3															
$d=4$	x^2p				−15							1		3									
	xyp					−15							1		3								
	y^2p						−15							1		3							
	xq		−2		13	1		−2				−3											
	yq			−2		13	1		−2				−3										
$d=5$	x^3p							−15									1		3				
	x^2yp								−15									1		3			
	xy^2p									−15									1		3		
	y^3p										−15									1		3	
	x^2q				−2			13	1				−2				−3						
	xyq					−2			13	1				−2				−3					
	y^2q						−2			13	1				−2				−3				
	↓							⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱	⋱

- \# rows grows faster than \# cols $\Rightarrow$ overdetermined system
- If solution exists: rank deficient by construction!

KU LEUVEN

nD realization in the null space after column compression to deflate the zeros at ∞:

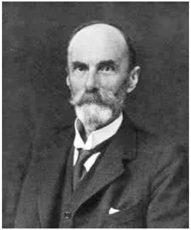
- Macaulay matrix M:

$$M = \begin{bmatrix} \times & \times & \times & \times & 0 & 0 & 0 \\ 0 & \times & \times & \times & \times & 0 & 0 \\ 0 & 0 & \times & \times & \times & \times & 0 \\ 0 & 0 & 0 & \times & \times & \times & \times \end{bmatrix}$$

- Solutions generate vectors in kernel of M:

$$MK = 0$$

- Number of solutions s follows from rank decisions 'mind-the-gap':



Francis Sowerby Macaulay

Vandermonde nullspace K built from s solutions (x_i, y_i):

$$\left[\begin{array}{c|c|c|c} 1 & 1 & \ldots & 1 \\ \hline x_1 & x_2 & \ldots & x_s \\ y_1 & y_2 & \ldots & y_s \\ \hline x_1^2 & x_2^2 & \ldots & x_s^2 \\ x_1y_1 & x_2y_2 & \ldots & x_sy_s \\ y_1^2 & y_2^2 & \ldots & y_s^2 \\ \hline x_1^3 & x_2^3 & \ldots & x_s^3 \\ x_1^2y_1 & x_2^2y_2 & \ldots & x_s^2y_s \\ x_1y_1^2 & x_2y_2^2 & \ldots & x_sy_s^2 \\ y_1^3 & y_2^3 & \ldots & y_s^3 \\ \hline x_1^4 & x_2^4 & \ldots & x_4^4 \\ x_1^3y_1 & x_2^3y_2 & \ldots & x_s^3y_s \\ x_1^2y_1^2 & x_2^2y_2^2 & \ldots & x_s^2y_s^2 \\ x_1y_1^3 & x_2y_2^3 & \ldots & x_sy_s^3 \\ y_1^4 & y_2^4 & \ldots & y_s^4 \\ \hline \vdots & \vdots & \vdots & \vdots \end{array}\right]$$

Setting up an eigenvalue problem in x

- Choose s linear independent rows in K

$$S_1 K$$

- This corresponds to finding linear dependent columns in M

$$
\left[\begin{array}{c|c|c|c}
1 & 1 & \dots & 1 \\
\hline
x_1 & x_2 & \dots & x_s \\
y_1 & y_2 & \dots & y_s \\
\hline
x_1^2 & x_2^2 & \dots & x_s^2 \\
x_1 y_1 & x_2 y_2 & \dots & x_s y_s \\
y_1^2 & y_2^2 & \dots & y_s^2 \\
\hline
x_1^3 & x_2^3 & \dots & x_s^3 \\
x_1^2 y_1 & x_2^2 y_2 & \dots & x_s^2 y_s \\
x_1 y_1^2 & x_2 y_2^2 & \dots & x_s y_s^2 \\
y_1^3 & y_2^3 & \dots & y_s^3 \\
\hline
x_1^4 & x_2^4 & \dots & x_4^4 \\
x_1^3 y_1 & x_2^3 y_2 & \dots & x_s^3 y_s \\
x_1^2 y_1^2 & x_2^2 y_2^2 & \dots & x_s^2 y_s^2 \\
x_1 y_1^3 & x_2 y_2^3 & \dots & x_s y_s^3 \\
y_1^4 & y_2^4 & \dots & y_s^4 \\
\hline
\vdots & \vdots & \vdots & \vdots
\end{array}\right]
$$

KU LEUVEN

Shifting the selected rows gives (shown for 3 columns)

$$
\left[\begin{array}{c|c|c}
1 & 1 & 1 \\ \hline
x_1 & x_2 & x_3 \\
y_1 & y_2 & y_3 \\ \hline
x_1^2 & x_2^2 & x_3^2 \\
x_1 y_1 & x_2 y_2 & x_3 y_3 \\
y_1^2 & y_2^2 & y_3^2 \\ \hline
x_1^3 & x_2^3 & x_3^3 \\
x_1^2 y_1 & x_2^2 y_2 & x_3^2 y_3 \\
x_1 y_1^2 & x_2 y_2^2 & x_3 y_3^2 \\
y_1^3 & y_2^3 & y_3^3 \\ \hline
x_1^4 & x_2^4 & x_4^4 \\
x_1^3 y_1 & x_2^3 y_2 & x_3^3 y_3 \\
x_1^2 y_1^2 & x_2^2 y_2^2 & x_3^2 y_3^2 \\
x_1 y_1^3 & x_2 y_2^3 & x_3 y_3^3 \\
y_1^4 & y_2^4 & y_3^4 \\ \hline
\vdots & \vdots & \vdots
\end{array}\right]
\longrightarrow \text{“shift with } x\text{”} \longrightarrow
\left[\begin{array}{c|c|c}
1 & 1 & 1 \\ \hline
x_1 & x_2 & x_3 \\
y_1 & y_2 & y_3 \\ \hline
x_1^2 & x_2^2 & x_3^2 \\
x_1 y_1 & x_2 y_2 & x_3 y_3 \\
y_1^2 & y_2^2 & y_3^2 \\ \hline
x_1^3 & x_2^3 & x_3^3 \\
x_1^2 y_1 & x_2^2 y_2 & x_3^2 y_3 \\
x_1 y_1^2 & x_2 y_2^2 & x_3 y_3^2 \\
y_1^3 & y_2^3 & y_3^3 \\ \hline
x_1^4 & x_2^4 & x_4^4 \\
x_1^3 y_1 & x_2^3 y_2 & x_3^3 y_3 \\
x_1^2 y_1^2 & x_2^2 y_2^2 & x_3^2 y_3^2 \\
x_1 y_1^3 & x_2 y_2^3 & x_3 y_3^3 \\
y_1^4 & y_2^4 & y_3^4 \\ \hline
\vdots & \vdots & \vdots
\end{array}\right]
$$

simplified:

$$
\left[\begin{array}{c|c|c}
1 & 1 & 1 \\
x_1 & x_2 & x_3 \\
y_1 & y_2 & y_3 \\
x_1 y_1 & x_2 y_2 & x_3 y_3 \\
x_1^3 & x_2^3 & x_3^3 \\
x_1^2 y_1 & x_2^2 y_2 & x_3^2 y_3
\end{array}\right]
\begin{bmatrix} x_1 & & \\ & x_2 & \\ & & x_3 \end{bmatrix}
=
\left[\begin{array}{c|c|c}
x_1 & x_2 & x_3 \\
x_1^2 & x_2^2 & x_3^2 \\
x_1 y_1 & x_2 y_2 & x_3 y_3 \\
x_1^2 y_1 & x_2^2 y_2 & x_3^2 y_3 \\
x_1^4 & x_2^4 & x_4^4 \\
x_1^3 y_1 & x_2^3 y_2 & x_3^3 y_3
\end{array}\right]
$$

Finding the x-roots

Let $D_x = \text{diag}(x_1, x_2, \ldots, x_s)$, then

$$S_1 K D_x = S_x K,$$

where S_1 and S_x select rows from K w.r.t. shift property We have

$$S_1 K D_x = S_x K$$

Generalized Vandermonde K is not known as such, instead a null space basis Z is calculated, which is a linear transformation of K:

$$ZV = K$$

which leads to

$$(S_x Z)V = (S_1 Z)V D_x$$

Here, V is the matrix with eigenvectors, D_x contains the roots x as eigenvalues.

KU LEUVEN

Setting up an eigenvalue problem in y

It is possible to shift with y as well...

We find

$$S_1 K D_y = S_y K$$

with D_y diagonal matrix of y-components of roots, leading to

$$(S_y Z)V = (S_1 Z)V D_y$$

Some interesting observations:

- same eigenvectors V!
- $(S_x Z)^{-1}(S_1 Z)$ and $(S_y Z)^{-1}(S_1 Z)$ commute
  $\Longrightarrow$ 'commutative algebra'

Rank, nullity and null space: SVD-ize the Macaulay matrix

Basic Algorithm outline

Find a basis for the nullspace of M using an SVD:

$$M = \begin{bmatrix} \times & \times & \times & \times & 0 & 0 & 0 \\ 0 & \times & \times & \times & \times & 0 & 0 \\ 0 & 0 & \times & \times & \times & \times & 0 \\ 0 & 0 & 0 & \times & \times & \times & \times \end{bmatrix} = \begin{bmatrix} X & Y \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} W^T \\ Z^T \end{bmatrix}$$

Hence,

$$MZ = 0$$

Deflate roots at ∞ by detecting 'mind-the-gap' and column compression:

$$ZT = \begin{pmatrix} Z_{11} & 0 \\ Z_{21} & Z_{22} \end{pmatrix}$$

We have

$$S_1 K D = S_{\text{shift}} K$$

with K generalized Vandermonde, not known as such. Instead a basis Z_{11} is computed as

$$Z_{11} V = K$$

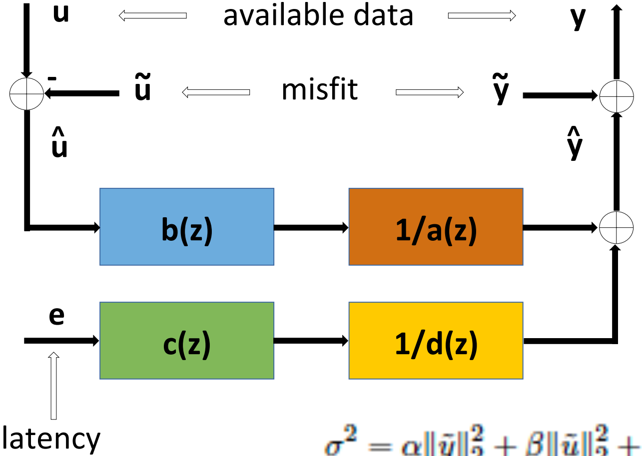
which leads to

$$(S_{\text{shift}} Z_{11}) V = (S_1 Z_{11}) V D$$

S_1 selects linear independent rows.
S_{shift} selects rows 'hit' by the shift.

Outline

Misfit-Latency LTI Models



$$\sigma^2 = \alpha\|\bar{y}\|_2^2 + \beta\|\bar{u}\|_2^2 + \gamma\|\bar{e}\|_2^2,$$

SISO transfer function (with $\mathcal{Z}\{x_k\} = x(z)$), e.g. ARMAX:

$$y(z) = \frac{b(z)}{a(z)}u(z) + \frac{c(z)}{a(z)}e(z),$$

with polynomial $a(z)$ (monic), $b(z)$, $c(z)$ (monic) of degree n_a, n_b, n_c.

Corresponding difference equation with $\alpha_i, \beta_i, \gamma_i \in \mathbb{R}$:

$$\begin{aligned} y_{k+n_a} + \alpha_1 y_{k+n_a-1} + \ldots + \alpha_{n_a} y_k \quad &= \quad \beta_0 u_{k+n_b} + \beta_1 y_{k+n_b-1} + \ldots + \alpha_{n_b} u_k \\ &\quad\;\; + e_{k+n_c} + \gamma_1 e_{k+n_c-1} + \ldots + \gamma_{n_c} e_k \end{aligned}$$

KU LEUVEN

Algebraic representation, e.g. ARMAX.

$$T_a y = T_b u + T_c e$$

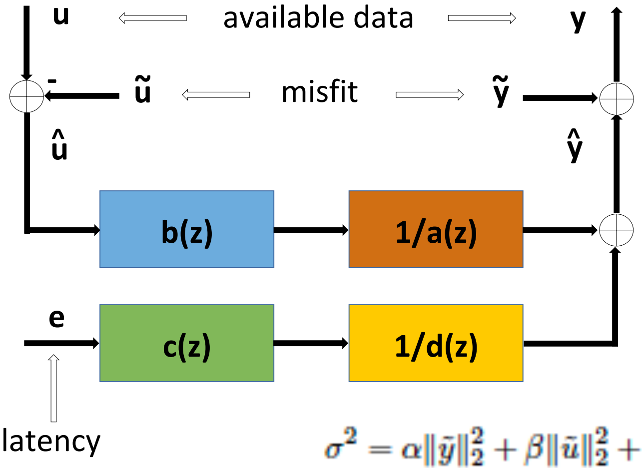
where $y^T = (y_0 \; y_1 \; \dots \; y_N)$ and e, u alike.

T_a, T_b, T_c are banded Toeplitz convolution operators, e.g. T_c:

$$\begin{pmatrix} \gamma_{n_c} & \gamma_{n_c-1} & \dots & \dots & \gamma_1 & 1 & 0 & 0 & \dots & 0 \\ 0 & \gamma_{n_c} & \gamma_{n_c-1} & \dots & \gamma_2 & \gamma_1 & 1 & 0 & \dots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \dots & \dots & \dots & \dots & \dots & \gamma_{n_c} & \gamma_{n_c-1} & \dots & \dots & 1 \end{pmatrix}$$

What we will not do here

- Experiment design, preprocessing data, validation, ...
- Which LTI model class to choose.
- Assessing optimal degrees n_a, n_b, n_c
- Bringing in a priori information in the objective function (weighting, a priori known coloring, missing variables)
- Identifiability conditions: persistancy of excitation, sufficient richness, ...
- Second order optimality conditions
- Interpretations/assumptions ('Hypotheses non fingo') such as maximum likelihood, statistical efficiency ...
- Error-covariance matrices, sensitivity, condition numbers
- ....

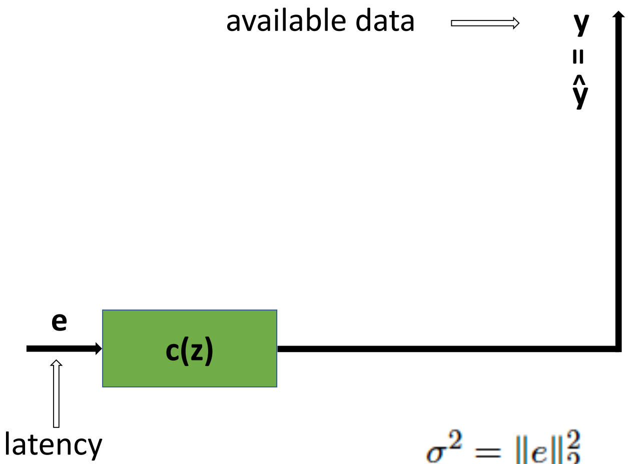


Misfit-Latency LTI Models

$$\sigma^2 = \alpha\|\tilde{y}\|_2^2 + \beta\|\tilde{u}\|_2^2 + \gamma\|\tilde{e}\|_2^2,$$

KU LEUVEN

Latency: MA
Moving Average



$$y = T_c e$$

$$\sigma^2 = \|e\|_2^2$$

Latency case: Moving average: Given $y \in \mathbb{R}^N$.

$$\min_{e \in \mathbb{R}^{N+n_c}} \sigma^2 = \|e\|_2^2 \text{ subject to } y = T_c e.$$

$T_c \in \mathbb{R}^{N \times (N+n_c)}$ = banded Toeplitz of full row rank (monic:$\gamma_0 = 1$). $e \in \mathbb{R}^{N+n_c}$ because of n_c initial conditions.
Underdetermined set of linear equations: minimum norm solution

$$e = T_c^{\dagger} y = T_c^T (T_c T_c^T)^{-1} y,$$

so that

$$\sigma^2 = \|e\|_2^2 = e^T e = y^T (T_c T_c^T)^{-1} y = y^T D_c^{-1} y \ ,$$

where D_c is symm. pos. def. banded Toeplitz, quadratic in the γ_i.

Interpretation: We look for a metric D_c^{-1} in which the weighted norm of y is minimal. $T_c^{\dagger}$ is a 'whitening' filter.

First order optimality conditions from $\sigma^2 = y^T D_c^{-1} y$:

$$\frac{\partial \sigma^2}{\partial \gamma_i} = y^T \frac{\partial D_c^{-1}}{\partial \gamma_i} y = y^T D_c^{-1} \frac{\partial D_c}{\partial \gamma_i} D_c^{-1} y = 0 \ , \ i = 1, \ldots, n_c. \quad (1)$$

These are n_c ‘nonlinear’ equations in the n_c unknowns γ_i.
Since

$$D_c^{-1} = \operatorname{adj}(D_c) / \det(D_c),$$

where the adjugate matrix $\operatorname{adj}(D_c)$ is multivariate polynomial in the γ_i, equations (1) constitute n_c multivariate polynomials in n_c variables γ_i:

$$\frac{\partial \sigma^2}{\partial \gamma_i} = 0 = y^T \operatorname{adj}(D_c) \frac{\partial D_c}{\partial \gamma_i} \operatorname{adj}(D_c) y \ , i = 1, \ldots, n_c.$$

The γ_i are the roots of a set of n_c multivariate polynomials in n_c unkowns.

Call $f = D_c^{-1} y$, then, with $\sigma^2 = y^T D_c^{-1} y$:

$$\begin{pmatrix} D_c & y \\ y^T & \sigma^2 \end{pmatrix} \begin{pmatrix} f \\ -1 \end{pmatrix} = 0. \tag{2}$$

First order optimality conditions: Chain rule with $D_c^{\gamma_i} = \partial D_c / \partial \gamma_i$, $f^{\gamma_i} = \partial f / \partial \gamma_i$ and $\partial \sigma^2 / \partial \gamma_i = 0$:

$$\begin{pmatrix} D_c^{\gamma_i} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} f \\ -1 \end{pmatrix} + \begin{pmatrix} D_c & y \\ y^T & \sigma^2 \end{pmatrix} \begin{pmatrix} f^{\gamma_i} \\ 0 \end{pmatrix} = 0. \tag{3}$$

$(N+1)(n_c+1)$ **equations:** $N+1$ in (2) and $n_c.(N+1)$ in (3).
$(N+1)(n_c+1)$ **unknowns:** N (f) + $n_c.N$ (f^{γ_i}) + n_c (γ_i) + 1 (σ^2).

The last row of (2) defines σ^2.

The last row of (3) defines n_c orthogonality relations $y^T f^{\gamma_i} = 0, i = 1, \ldots, n_c$.

Orthogonality

$$\begin{aligned} y^T f^{\gamma_i} &= 0 \\ &= y^T D_c^{-1} D_c^{\gamma_i} f \\ &= y^T D_c^{-1} T_c^{\gamma_i} T_c^T f \\ &= y^T D_c^{-1} T_c^{\gamma_i} e \ , \ i = 1, \ldots, n_c. \end{aligned}$$

$$y^T D_c^{-1} \begin{pmatrix} e_{-n_c} & e_{-n_c+1} & \cdots & e_{-1} \\ e_{-n_c+1} & e_{-n_c+2} & \cdots & e_0 \\ e_{-n_c+2} & e_{-n_c+3} & \cdots & e_1 \\ \vdots & \vdots & \vdots & \vdots \\ e_{N-n_c} & e_{N-n+c+1} & \cdots & e_{N-1} \end{pmatrix} = 0.$$

The data vector y is orthogonal to the column space of the $N \times n_c$ Hankel matrix with the latency estimates, in the metric given by D_c^{-1}.

Latency case: MA ($n_c = 1$)

$$\begin{pmatrix} D_c^\gamma & D_c & 0 \\ D_c & 0 & y \\ 0 & y^T & 0 \end{pmatrix} \begin{pmatrix} f \\ f^\gamma \\ -1 \end{pmatrix} = 0.$$

For $N = 4$:

$$\left(\begin{array}{cccc|cccc|c} 2\gamma & 1 & 0 & 0 & 1+\gamma^2 & \gamma & 0 & 0 & 0 \\ 1 & 2\gamma & 1 & 0 & \gamma & 1+\gamma^2 & \gamma & 0 & 0 \\ 0 & 1 & 2\gamma & 1 & 0 & \gamma & 1+\gamma^2 & \gamma & 0 \\ 0 & 0 & 1 & 2\gamma & 0 & 0 & \gamma & 1+\gamma^2 & 0 \\ \hline 1+\gamma^2 & \gamma & 0 & 0 & 0 & 0 & 0 & 0 & y_0 \\ \gamma & 1+\gamma^2 & \gamma & 0 & 0 & 0 & 0 & 0 & y_1 \\ 0 & \gamma & 1+\gamma^2 & \gamma & 0 & 0 & 0 & 0 & y_2 \\ 0 & 0 & \gamma & 1+\gamma^2 & 0 & 0 & 0 & 0 & y_3 \\ \hline 0 & 0 & 0 & 0 & y_0 & y_1 & y_2 & y_3 & 0 \end{array}\right) \left(\begin{array}{c} f_0 \\ f_1 \\ f_2 \\ f_3 \\ \hline f_0^\gamma \\ f_1^\gamma \\ f_2^\gamma \\ f_3^\gamma \\ \hline -1 \end{array}\right) = 0.$$

Regroup as **quadratic eigenvalueproblem** and ‘linearize’ :

$$(A_2\gamma^2 + A_1\gamma + A_0)z = 0 \text{ with } z = \begin{pmatrix} -1 \\ f \\ f^\gamma \end{pmatrix} \Longrightarrow \begin{pmatrix} 0 & I \\ A_0 & A_1 \end{pmatrix} \begin{pmatrix} z \\ z\gamma \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & -A_2 \end{pmatrix} \begin{pmatrix} z \\ z\gamma \end{pmatrix} \gamma.$$

Only need eigenvalue that minimizes objective function !
The latency $e = T_c^T f$, f is part of corresponding eigenvector.

Latency case MA ($n_c = 2$)

$$\begin{pmatrix} D_c^{\gamma_i} & D_c & 0 \\ D_c & 0 & y \\ 0 & y^T & 0 \end{pmatrix} \begin{pmatrix} f \\ f^{\gamma_i} \\ -1 \end{pmatrix} = 0 \ , \ i = 1, 2.$$

Regroup in a multi-parameter eigenvalueproblem with $z^T = (-1 \ f^T \ (f^{\gamma_1})^T \ (f^{\gamma_2})^T \ )$:

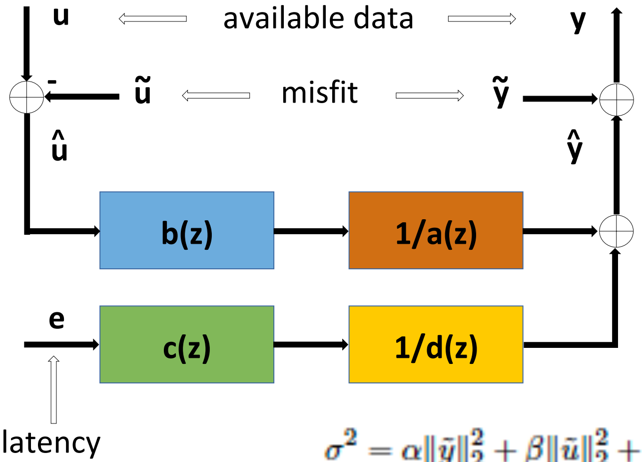
$$(A_{00} + A_{10}\gamma_1 + A_{01}\gamma_2 + A_{20}\gamma_1^2 + A_{11}\gamma_1\gamma_2 + A_{02}\gamma_2^2) \left(\begin{array}{c} z \\ \hline z\gamma_1 \\ z\gamma_2 \\ \hline z\gamma_1^2 \\ z\gamma_1\gamma_2 \\ z\gamma_1^2 \end{array}\right) = 0.$$

and build up block Macaulay recursively (quasi-Toeplitz-ify) until 'mind-the-gap' starts in the null space, which is **multi-shift invariant**:

$$\begin{array}{l} \\ 1 \\ \times\gamma_1 \\ \times\gamma_2 \\ \times\gamma_1^2 \\ \times\gamma_1\gamma_2 \\ \times\gamma_2^2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccccccccc} 1 & \gamma_1 & \gamma_2 & \gamma_1^2 & \gamma_1\gamma_2 & \gamma_2^2 & \gamma_1^3 & \gamma_1^2\gamma_2 & \gamma_1\gamma_2^2 & \gamma_2^3 & \gamma_1^4 & \dots \end{array} \\ \begin{pmatrix} A_{00} & A_{10} & A_{01} & A_{20} & A_{11} & A_{02} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & A_{00} & 0 & A_{10} & A_{01} & 0 & A_{20} & A_{11} & A_{02} & 0 & 0 & \dots \\ 0 & 0 & A_{00} & 0 & A_{10} & A_{01} & 0 & A_{20} & A_{11} & A_{02} & 0 & \dots \\ 0 & 0 & 0 & A_{00} & 0 & 0 & A_{10} & A_{01} & 0 & 0 & A_{20} & \dots \\ 0 & 0 & 0 & 0 & A_{00} & 0 & 0 & A_{10} & A_{01} & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & A_{00} & 0 & 0 & A_{10} & A_{01} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \end{array} \left(\begin{array}{c} z \\ \hline z\gamma_1 \\ z\gamma_2 \\ \hline z\gamma_1^2 \\ z\gamma_1\gamma_2 \\ z\gamma^2 \\ \hline z\gamma_1^3 \\ \vdots \end{array}\right) = 0$$

Next do 2D realization theory in the **multi-shift invariant** null space !

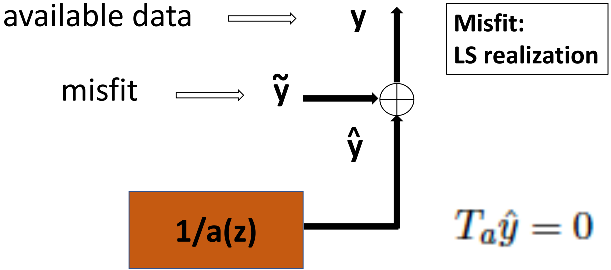
Misfit case: Least squares realization (n_a)



Misfit-Latency LTI Models

$$\sigma^2 = \alpha\|\tilde{y}\|_2^2 + \beta\|\tilde{u}\|_2^2 + \gamma\|\tilde{e}\|_2^2,$$

Misfit case: Least squares realization (n_a)



$$T_a \hat{y} = 0$$

$$\sigma^2 = \|\tilde{y}\|_2^2$$

Misfit case: Least squares realization

$$\min \|\tilde{y}\|_2^2 \ \text{subject to} \ \begin{array}{l} y = \hat{y} + \tilde{y}, \\ T_a \hat{y} = 0. \end{array}$$

Obviously

$$T_a y = T_a \tilde{y}.$$

Minimum norm solution using pseudo-inverse and T_a full row rank:

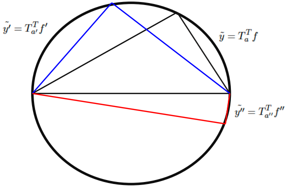
$$\tilde{y} = T_a^\dagger T_a y = T_a^T (T_a T_a^T)^{-1} T_a y = \Pi_a y.$$

Π_a = orthogonal projector onto row space of T_a. Define $D_a = T_a T_a^T$ and $f = D_a^{-1} T_a y$:

$$y = \hat{y} + \tilde{y} = \hat{y} + T_a^T f \Longrightarrow \hat{y} \perp \tilde{y} = T_a^T f.$$

$$\tilde{y} = T_a^T f.$$

The least squares residual
= f through FIR filter determined by a
= Finite dimensional form of
Beurling - Lax - Halmos theorem



Let

$$\sigma^2 = \|\tilde{y}\|_2^2 = y^T T_a^T (T_a T_a^T)^{-1} T_a y.$$

With $f = D_a^{-1} T_a y$:

$$\begin{pmatrix} D_a & T_a y \\ y^T T_a^T & \sigma^2 \end{pmatrix} \begin{pmatrix} f \\ -1 \end{pmatrix} = 0. \tag{4}$$

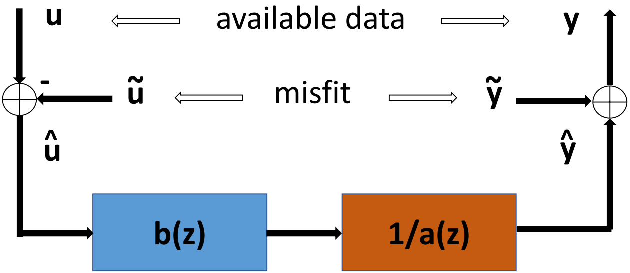
First order optimality conditions and chain rule:

$$\begin{pmatrix} D_a^{\alpha_i} & T_a^{\alpha_i} y \\ y^T (T_a^{\alpha_i})^T & 0 \end{pmatrix} \begin{pmatrix} f \\ -1 \end{pmatrix} + \begin{pmatrix} D_a & T_a y \\ y^T T_a^T & \sigma^2 \end{pmatrix} \begin{pmatrix} f^{\alpha_i} \\ -1 \end{pmatrix} = 0 \, , \; i = 1, \ldots, n_a. \tag{5}$$

Then:

- Define $z^T = (\, -1 \;\; f^T \;\; (f^{\alpha_1})^T \;\; \ldots (f^{\alpha_{n_a}})^T \,)$.
- Quasi-Toeplitz-ify eqs. (4) - (5) in block Macaulay with blocks in $1, \alpha_1, \ldots, \alpha_{n_a}, \alpha_1^2, \alpha_1 \alpha_2, \ldots$.
- Null space will be multi-shift invariant.
- Do nD realization theory in the null space.
- The last row of (4) allows to evaluate σ^2 in the roots
- The last row of (5) delivers interesting orthogonality properties (not derived here)
- Misfit vector $\tilde{y} = T_a^T f$ follows from eigenvector

Misfit case: Dynamic Total Least Squares (n_a, n_b)



Misfit:
Dynamic Total LS

$$T_a\hat{y} + T_b\hat{u} = 0$$

$$\sigma^2 = \|\tilde{u}\|_2^2 + \|\tilde{y}\|_2^2$$

Misfit case: dynamic total least squares

$$\min \|\tilde{u}\|_2^2 + \|\tilde{y}\|_2^2 \text{ subject to } \begin{array}{l} u = \hat{u} + \tilde{u} \\ y = \hat{y} + \tilde{y} \\ T_a\hat{y} + T_b\hat{u} = 0 \end{array} .$$

Then:

$$T_a y + T_b u = \begin{pmatrix} T_a & T_b \end{pmatrix} \begin{pmatrix} \tilde{y} \\ \tilde{u} \end{pmatrix} .$$

Pseudo-inverse minimum norm solution:

$$\begin{pmatrix} \tilde{y} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} T_a^T \\ T_b^T \end{pmatrix} (T_a T_a^T + T_b T_b^T)^{-1} \begin{pmatrix} T_a & T_b \end{pmatrix} \begin{pmatrix} y \\ u \end{pmatrix} = \Pi_{ab} \begin{pmatrix} y \\ u \end{pmatrix} .$$

Again 'Thales orthogonal decomposition' and 'Beurling-Lax-Halmos':

$$y = \hat{y} + \tilde{y} \Longrightarrow \begin{array}{l} \begin{pmatrix} T_a & T_b \end{pmatrix} \begin{pmatrix} \hat{y} \\ \hat{u} \end{pmatrix} = 0 \\ \begin{pmatrix} \tilde{y} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} T_a^T \\ T_b^T \end{pmatrix} f \end{array}$$

with

$$D_{ab} = (T_a T_a^T + T_b T_b^T) \text{ and } f = D_{ab}^{-1} \begin{pmatrix} T_a & T_b \end{pmatrix} \begin{pmatrix} y \\ u \end{pmatrix} .$$

Then

$$\sigma^2 = \|\tilde{u}\|_2^2 + \|\tilde{y}\|_2^2 = \begin{pmatrix} y^T & u^T \end{pmatrix} \begin{pmatrix} T_a^T \\ T_b^T \end{pmatrix} D_{ab}^{-1} \begin{pmatrix} T_a & T_b \end{pmatrix} \begin{pmatrix} y \\ u \end{pmatrix},$$

so that

$$\begin{pmatrix} D_{ab} & T_a y + T_b u \\ y^T T_a^T + u^T T_b^T & \sigma^2 \end{pmatrix} \begin{pmatrix} f \\ -1 \end{pmatrix} = 0. \quad (6)$$

First order optimality and chain rule:

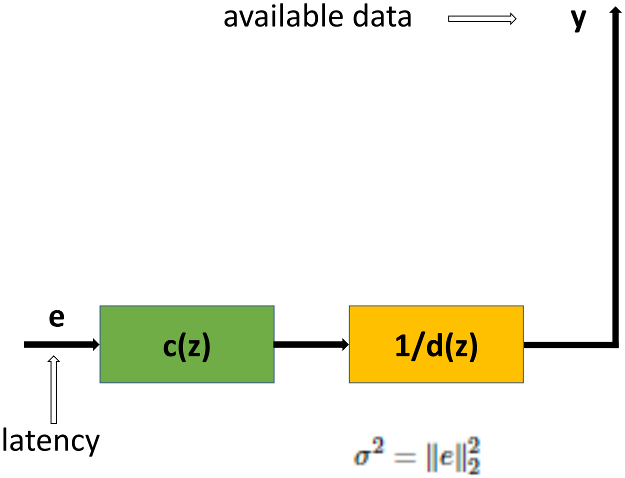
$$\begin{pmatrix} D_{ab}^{\alpha_i} & T_a^{\alpha_i} y \\ y^T (T_a^{\alpha_i})^T & 0 \end{pmatrix} \begin{pmatrix} f \\ -1 \end{pmatrix} + \begin{pmatrix} D_{ab} & T_a y + T_b u \\ y^T T_a^T + u^T T_b^T & \sigma^2 \end{pmatrix} \begin{pmatrix} f^{\alpha_i} \\ 0 \end{pmatrix}, \; i = 1, \ldots, n_a. \quad (7)$$

$$\begin{pmatrix} D_{ab}^{\beta_i} & T_b^{\beta_i} y \\ u^T (T_b^{\beta_i})^T & 0 \end{pmatrix} \begin{pmatrix} f \\ -1 \end{pmatrix} + \begin{pmatrix} D_{ab} & T_a y + T_b u \\ y^T T_a^T + u^T T_b^T & \sigma^2 \end{pmatrix} \begin{pmatrix} f^{\beta_i} \\ 0 \end{pmatrix}, \; i = 0, \ldots, n_b. \quad (8)$$

Then

- Define $z^T = \begin{pmatrix} -1 & f^T & (f^{\alpha_1})^T & \ldots (f^{\alpha_{n_a}})^T & (f^{\beta_1})^T & \ldots (f^{\beta_{n_b}})^T \end{pmatrix}$.
- Quasi-Toeplitz-ify in block Macaulay with blocks in $1, \alpha_1, \ldots, \alpha_{n_a}, \beta_0, \beta_1, \ldots, \alpha_1^2, \alpha_1\alpha_2, \ldots$.
- Null space will be multi-shift invariant.
- Do nD realization theory in the null space.
- The last row of (6) allows to evaluate σ^2 in the roots
- The last rows of (7) and (8) deliver interesting orthogonality properties (not derived here)
- Misfit vectors $\tilde{y}$ and $\tilde{u}$ follow from eigenvector

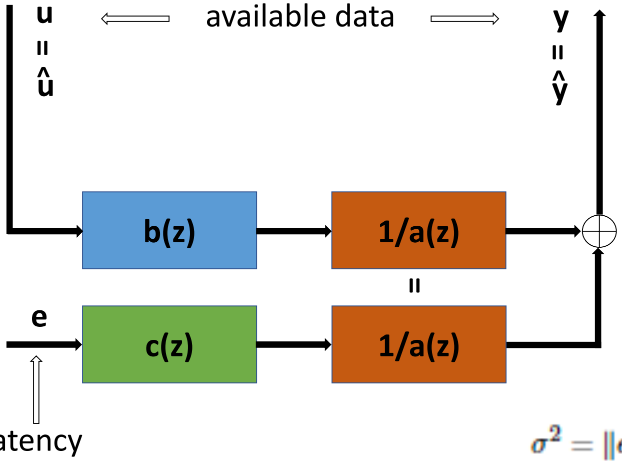
Latency case: ARMA (n_a, n_c)



Latency: ARMA Autoregressive Moving Average

$$T_a y = T_c e$$

Latency case: ARMAX (n_a, n_b, n_c)

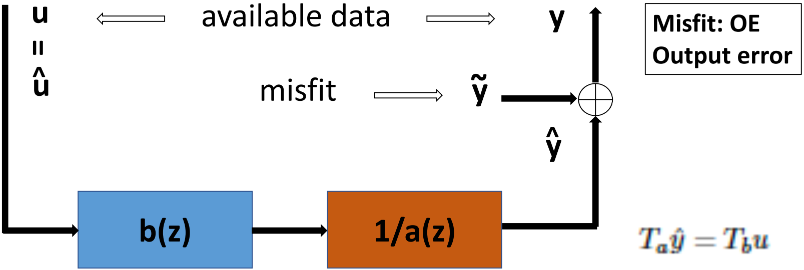


Latency: ARMAX ARMA + eXogeneous inputs

$$T_a y = T_b u + T_c e$$

$$\sigma^2 = \|e\|_2^2$$

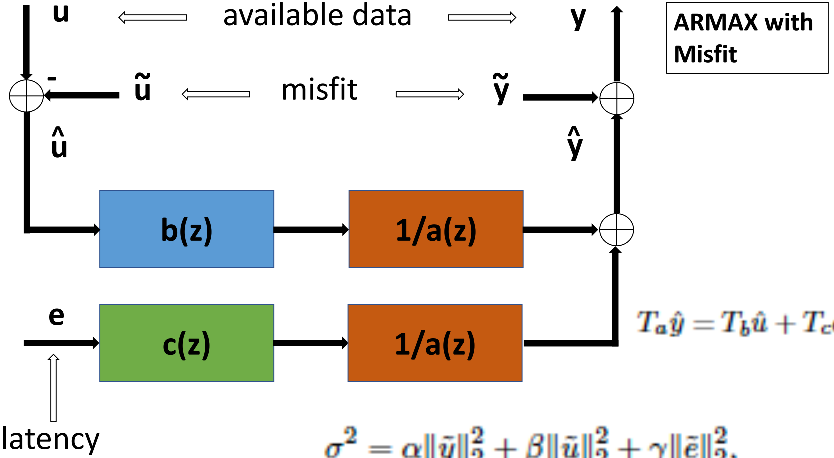
Misfit case: Output Error (n_a, n_b)



$$T_a \hat{y} = T_b u$$

$$\sigma^2 = \|\tilde{y}\|_2^2$$

Misfit+Latency case: ARMAX with I/O Misfit(n_a, n_b, n_c)



$$T_a \hat{y} = T_b \hat{u} + T_c e$$

$$\sigma^2 = \alpha \|\tilde{y}\|_2^2 + \beta \|\tilde{u}\|_2^2 + \gamma \|\tilde{e}\|_2^2,$$

Name	u	e	α	β	γ	a	b	c	d
Exact data									
Autonomous system	0	0	∞	∞	∞	a	1	1	1
Exact FIR	u	0	∞	∞	∞	1	b	1	1
Diff. eq.	u	0	∞	∞	∞	a	b	1	1
$\vdots$									
Latency									
MA	0	e	∞	∞	1	1	1	c	1
AR	0	e	∞	∞	1	1	1	1	d
ARMA	0	e	∞	∞	1	1	1	c	d
ARMAX	u	e	∞	∞	1	a	b	c	a
$\vdots$									
Misfit									
LS Realization	0	0	1	∞	∞	a	1	1	1
OE FIR	u	0	1	∞	∞	1	b	1	1
IE FIR	u	0	∞	1	∞	1	b	1	1
IE+OE FIR	u	0	α	β	∞	1	b	1	1
OE	u	0	1	∞	∞	a	b	1	1
IE	u	0	∞	1	∞	a	b	1	1
Dynamic TLS	u	0	α	β	∞	a	b	1	1
$\vdots$									
Misfit + Latency									
ARMAX with M+L	u	e	α	β	γ	a	b	c	a
$\vdots$									

Outline

What have we done ?

- System identification of LTI dynamical system least squares minimizing misfit and/or latency is solved!
- It is an eigenvalue problem, because
  - It is a multivariate polynomial optimization problem.
  - The first order optimality conditions generate a set of multivariate polynomials.
  - The optimal parameters belong to the roots of this set.
  - To find them, we recursively quasi-Toeplitz-ify the first order optimality conditions into 'growing' (block) Macaulay matrices.
  - The null spaces of these quasi-Toeplitz matrices are multi-shift invariant subspace, with 3 zones:
    - A 'regular' zone, recovered by rank tests and a column compression, that 'contains' the affine roots
    - A 'mind-the-gap'-zone that seperates the affine roots from those at infinity;
    - An 'a-bout-du-souffle'-zone that 'contains' the roots at infinity.
  - We apply nD realization theory in these multi-shift invariant subspaces
  - The roots are eigenvalues of the n shift matrices.
- We only need the minimizing affine roots (not covered here)

KU LEUVEN

What did we use ?

System and control theory: (Singular) observability matrices, parametrizations, ...

Optimization theory: Optimality conditions, Lagrange multipliers, ...

Advanced linear algebra: Cayley-Hamilton, SVD, JCF, WCF, ...

Algebraic geometry: *'queen of mathematics'*:

- Hilbert's theorem (nullstellensatz, basis thm, syzygies), ...
- 'Intersection' of fundamental theorem of algebra and linear algebra (null spaces and multi-shift invariance)
- Multi-parameter eigenvalue problems
- Translate (symbolic algebraic geometry: Grobner bases) into numerical linear algebra (floating point arithmetic)

Operator theory: shift-invariant subspaces, Beurling-Lax, ...

What are we to do in the (near) future ?

- Algorithms:
  - Numerical linear algebra: Large scale HPC implementation (exploiting structure (quasi-Toeplitz and multi-shift invariance), sparsity,....)
  - Compute only eigenvalues for minimum: power method and extensions (Lanczos, Krylov,....)
  - Recursiveness in the degrees n_a, n_b, n_c and in the number of data N: 'root loci' and 'stabilization diagrams'
  - Analyse all existing 'heuristic' approaches: PEM, VAPRO, IQML, Cadzow's iteration, ... (e.g. local versus global minima)
- Least squares and orthogonality: many interesting structured orthogonality results to be uncovered.
- Sensitivity, condition numbers, persistancy of excitation, sufficiently rich, ...
- Second-order optimality conditions, error covariance matrices, ...
- Extension for MIMO (find approach in state space so that 'non-uniqueness of parametrization does not matter, i.e. modulo non-uniqueness
- H_2 model reduction is solved: it's an eigenvalue problem. Bring in more operator theory (e.g. Commutant Lifting Theorem)
- Theory of multi-shift invariant spaces
- Least squares system id for linear partial difference equations
- ...

"What is difficult to solve in a low dimensional space,
is easier to solve in a high dimensional space."

Ex.: least squares realization is solved exactly by nD realization

"What is difficult to solve in a low dimensional space,
is easier to solve in a high dimensional space."
Ex.: least squares realization is solved exactly by nD realization

"Generalize to solve"

"What is difficult to solve in a low dimensional space,
is easier to solve in a high dimensional space."
Ex.: least squares realization is solved exactly by nD realization

"Generalize to solve"

"At the end of the day,
the only thing we really understand,
is linear algebra."