Least squares optimal realization of autonomous LTI systems is an eigenvalue problem

Prof. Dr. Bart De Moor

KU Leuven
Dept.EE: ESAT - STADIUS

bart.demoor@esat.kuleuven.be

Back to the roots:

- Of linear algebra: Eigenvalue problem (=roots)
- Of algebraic geometry: Common roots of sets of multivariate polynomials
- To do something about the heuristics in modern day data science
- Honor my mentors (Kailath, Golub, Willems, ...)

I have deeply regretted
that I did not proceed far enough
at least to understand something
of the great leading principles of mathematics
for men thus endowed
seem to have an extra sense

Charles Darwin

Outline

Outline

1. Eigenvalues
2. Models and data
3. Rooting multivariate polynomials
4. Multivariate Optimization
5. Misfit realization
6. Conclusions

Eigenvalue problem

For matrix $A \in \mathbb{R}^{n\times n}$, find $x \neq 0 \in \mathbb{C}^n, \lambda \in \mathbb{C}$ so that

$$A\,x = x\,\lambda$$

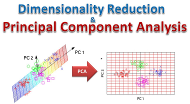
Eigenvectors: invariant directions for operator A in $\mathbb{R}^n$
Eigenvalues: Characteristic equation - fundamental theorem of algebra

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

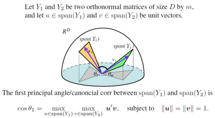
Since Galois, for $n \geq 5$: no solution in radicals $\Longrightarrow$ iterative algorithms

Eigenvalue decomposition: Jordan Canonical Form (JCF)

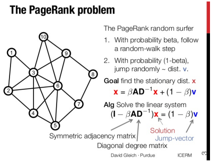
$$A = XJX^{-1}$$



PCA



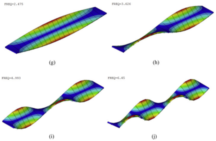
Can. Corr./Principal Angles



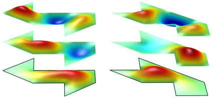
Graph spectral analysis

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

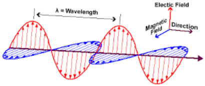
Wave equation



Modal shapes



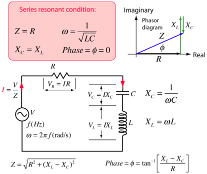
Hear the shape of a drum?



Maxwell's laws

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

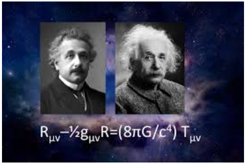
Maxwell's field equations



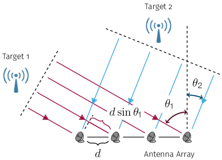
RLC circuits

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

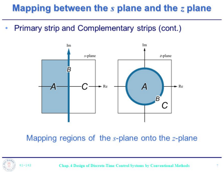
Schrodinger equation



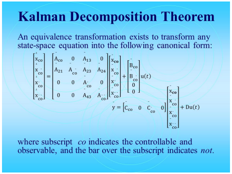
Matter curves spacetime moves matter



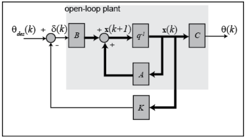
Direction of Arrival



Stability

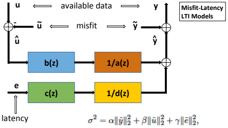


Controllability/observability



Pole placement

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



LS LTI System ID = EVP !

Multiparameter Eigenvalue Problem (MEVP)

Given $A_0, \ldots, A_m \in \mathbb{R}^{p \times q}$ with $p \geq q$, find $\lambda_i \in \mathbb{C}, i = 1, \ldots, m$ and $x \neq 0 \in \mathbb{C}^q$ so that

$$(A_0 + A_1\lambda_1 + \ldots + A_m\lambda_m)\, x = 0$$

Special cases:

- Ordinary EVP: $A_0 \in \mathbb{R}^{n \times n}$, $A_1 = -I_n$, $A_i = 0, i \geq 2$
- 'Generalized' EVP: $A_0, A_1 \in \mathbb{R}^{n \times n}$, $A_i = 0, i \geq 2$

Basic idea to solve an MEVP (illustrated for $m = 2$)

$$(A_0 + A_1\lambda_1 + A_2\lambda_2)\, x = 0$$

$$\begin{array}{c} \times 1 \\ \times\lambda_1 \\ \times\lambda_2 \\ \times\lambda_1^2 \\ \vdots \end{array} \quad \begin{pmatrix} A_0 & A_1 & A_2 & 0 & 0 & 0 & 0 & \dots \\ 0 & A_0 & 0 & A_1 & A_2 & 0 & 0 & \dots \\ 0 & 0 & A_0 & 0 & A_1 & A_2 & 0 & \dots \\ 0 & 0 & 0 & A_0 & 0 & 0 & A_1 & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \begin{pmatrix} x \\ \hline x\lambda_1 \\ x\lambda_2 \\ \hline x\lambda_1^2 \\ x\lambda_1\lambda_2 \\ x\lambda_2^2 \\ \hline x\lambda_1^3 \\ \vdots \end{pmatrix} = 0$$

Block 'quasi'-Toeplitz structure + 'generalized' Vandermonde structure

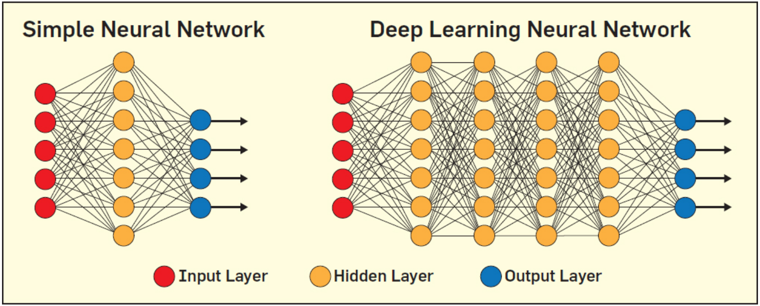
Outline

Basic modelling loop:

1. Collect data (preprocess, wrangle, clean, ...)
2. Select model class parametrized by unknown parameters
3. Select an approximation criterion
4. 'Solve' using nonlinear optimization
5. Validate the results
6. Re-iterate when necessary

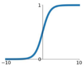
What do we mean by **'solved'** ?

- Result of nonlinear optimization ? Trouble with:
  1. Starting points (feasibility);
  2. Convergence (step sizes, rate, stopping criteria,...)
  3. Local minima
- Solved = *convex* or *set of linear equations* or *eigenvalue problem* !

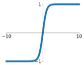


Activation Functions

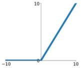
Sigmoid
$\sigma(x) = \frac{1}{1+e^{-x}}$



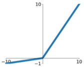
tanh
$\tanh(x)$



ReLU
$\max(0, x)$

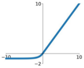


Leaky ReLU
$\max(0.1x, x)$

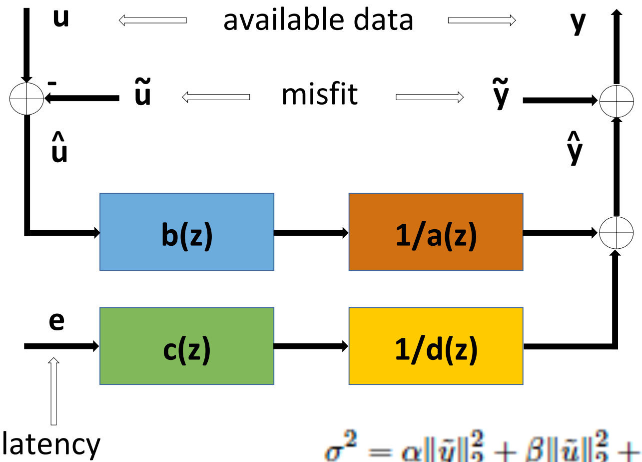


Maxout
$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU
$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$



Tackled by nonlinear optimization



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	$x_{k+1} = \mathbf{A}\mathbf{x_k} + Bu_k$ 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.

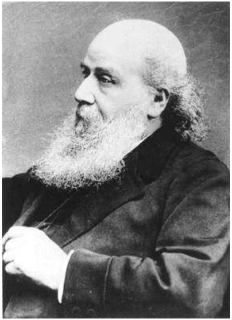
Hence multivariate polynomial !

- 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 common roots of this set are local and global minima and maxima, and saddle points
- We only need the root(s) 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 ?

Outline

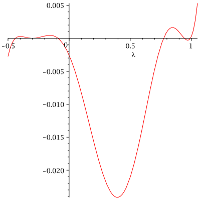
Rooting a set of multivariate polynomials is an eigenvalue problem !



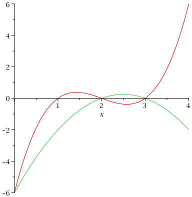
James Joseph Sylvester

- Algebra (fundamental theorem)
- Numerical linear algebra (power method and derivates, multiparameter eig. problem (MEVP), SVD)
- (Commutative) Algebraic geometry (ideals and varieties)
- Optimization theory (Lagrangean)
- System theory (state space, realization theory)
- nD system theory (nD realization)
- Operator theory (shift-invariant spaces)
- Interpolation theory (moment problems)

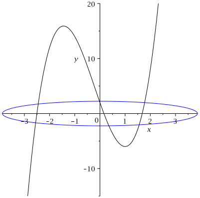
$$p(\lambda) = \det(A - \lambda I) = 0$$



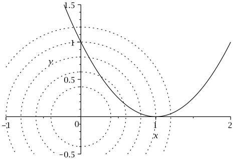
$$\begin{aligned} (x-1)(x-3)(x-2) &= 0 \\ -(x-2)(x-3) &= 0 \end{aligned}$$



$$\begin{aligned} x^2 + 3y^2 - 15 &= 0 \\ y - 3x^3 - 2x^2 + 13x - 2 &= 0 \end{aligned}$$



$$\begin{aligned} \min_{x,y} \quad & x^2 + y^2 \\ \text{s.t.} \quad & y - x^2 + 2x - 1 = 0 \end{aligned}$$



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

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

- Banded Toeplitz; linear homogeneous equations
- Null space: (Confluent) Vandermonde structure
- Corank (nullity) = number of solutions

Two univariate polynomials: common roots ?

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

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

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

- Sylvester matrix = double banded Toeplitz
- Null space = (confluent) Vandermonde structure
- Null space = intersection of null spaces of two banded Toeplitz matrices
- Nullity = number of common zeros

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

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

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

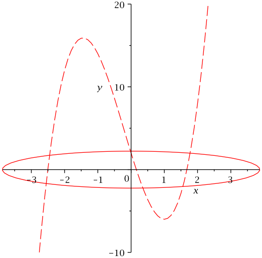
$$\begin{aligned} \underline{K}.D &= \overline{K} \\ \underline{Z}.V.D &= \overline{Z}.V \end{aligned}$$

- Generalized EVP with eigenvalues in D and eigenvectors the columns of V.
- Null space $\mathbf{R}(K) = \mathbf{R}(Z)$ is shift-invariant.

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:

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

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

Continue to enlarge M ('quasi-Toeplitzification'):

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 ⇒ overdetermined system
- If solution exists: rank deficient by construction!

KU LEUVEN

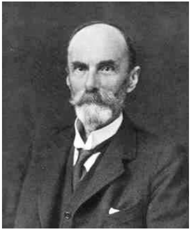
- Macaulay matrix M:

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

- Solutions generate vectors in kernel of M:

$$MK = 0$$

- Number of solutions s follows from rank decisions



Francis Sowerby Macaulay

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

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

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

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_s \\ y_1 & y_2 & \dots & y_s \\ x_1^2 & x_2^2 & \dots & x_s^2 \\ x_1 y_1 & x_2 y_2 & \dots & x_s y_s \\ y_1^2 & y_2^2 & \dots & y_s^2 \\ x_1^3 & x_2^3 & \dots & x_s^3 \\ x_1^2 y_1 & x_2^2 y_2 & \dots & x_s^2 y_s \\ x_1 y_1^2 & x_2 y_2^2 & \dots & x_s y_s^2 \\ y_1^3 & y_2^3 & \dots & y_s^3 \\ x_1^4 & x_2^4 & \dots & x_4^4 \\ x_1^3 y_1 & x_2^3 y_2 & \dots & x_s^3 y_s \\ x_1^2 y_1^2 & x_2^2 y_2^2 & \dots & x_s^2 y_s^2 \\ x_1 y_1^3 & x_2 y_2^3 & \dots & x_s y_s^3 \\ y_1^4 & y_2^4 & \dots & y_s^4 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Shifting the selected rows gives (shown for 3 columns)

$$
\begin{bmatrix}
1 & 1 & 1 \\
x_1 & x_2 & x_3 \\
y_1 & y_2 & y_3 \\
x_1^2 & x_2^2 & x_3^2 \\
x_1 y_1 & x_2 y_2 & x_3 y_3 \\
y_1^2 & y_2^2 & y_3^2 \\
x_1^3 & x_2^3 & x_3^3 \\
x_1^2 y_1 & x_2^2 y_2 & x_3^2 y_3 \\
x_1 y_1^2 & x_2 y_2^2 & x_3 y_3^2 \\
y_1^3 & y_2^3 & y_3^3 \\
x_1^4 & x_2^4 & x_4^4 \\
x_1^3 y_1 & x_2^3 y_2 & x_3^3 y_3 \\
x_1^2 y_1^2 & x_2^2 y_2^2 & x_3^2 y_3^2 \\
x_1 y_1^3 & x_2 y_2^3 & x_3 y_3^3 \\
y_1^4 & y_2^4 & y_3^4 \\
\vdots & \vdots & \vdots
\end{bmatrix}
\longrightarrow \text{“shift with } x\text{”} \longrightarrow
\begin{bmatrix}
1 & 1 & 1 \\
x_1 & x_2 & x_3 \\
y_1 & y_2 & y_3 \\
x_1^2 & x_2^2 & x_3^2 \\
x_1 y_1 & x_2 y_2 & x_3 y_3 \\
y_1^2 & y_2^2 & y_3^2 \\
x_1^3 & x_2^3 & x_3^3 \\
x_1^2 y_1 & x_2^2 y_2 & x_3^2 y_3 \\
x_1 y_1^2 & x_2 y_2^2 & x_3 y_3^2 \\
y_1^3 & y_2^3 & y_3^3 \\
x_1^4 & x_2^4 & x_4^4 \\
x_1^3 y_1 & x_2^3 y_2 & x_3^3 y_3 \\
x_1^2 y_1^2 & x_2^2 y_2^2 & x_3^2 y_3^2 \\
x_1 y_1^3 & x_2 y_2^3 & x_3 y_3^3 \\
y_1^4 & y_2^4 & y_3^4 \\
\vdots & \vdots & \vdots
\end{bmatrix}
$$

so that:

$$
\begin{bmatrix}
1 & 1 & 1 \\
x_1 & x_2 & x_3 \\
y_1 & y_2 & y_3 \\
x_1 y_1 & x_2 y_2 & x_3 y_3 \\
x_1^3 & x_2^3 & x_3^3 \\
x_1^2 y_1 & x_2^2 y_2 & x_3^2 y_3
\end{bmatrix}
\begin{bmatrix}
x_1 & & \\
 & x_2 & \\
 & & x_3
\end{bmatrix}
=
\begin{bmatrix}
x_1 & x_2 & x_3 \\
x_1^2 & x_2^2 & x_3^2 \\
x_1 y_1 & x_2 y_2 & x_3 y_3 \\
x_1^2 y_1 & x_2^2 y_2 & x_3^2 y_3 \\
x_1^4 & x_2^4 & x_4^4 \\
x_1^3 y_1 & x_2^3 y_2 & x_3^3 y_3
\end{bmatrix}
$$

Finding the x-roots

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

$$S_1 K D_x = S_x K,$$

where S_1 selects linear independent rows of K and S_x the ones 'hit' by the shift x.

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 the generalized eigenvalue problem

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

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

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'

Single shift invariance

$$\Gamma = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{p-2} \\ CA^{p-1} \end{pmatrix} \Longrightarrow \underline{\Gamma} A = \overline{\Gamma}$$

- Null space of Toeplitz or Sylvester
- Single shift (only one A)
- Cayley-Hamilton
- Shift-invariant $\mathbf{R}(\Gamma)$ fixed by $\lambda(A)$
- 1D observability
- 1D realization theory
- 1D Beurling-Lax
- 'Block' when $C =$ matrix

Multi-shift invariance $(n = 2)$

$$\Gamma = \left(\begin{array}{c} C \\ \hline CA_1 \\ CA_2 \\ \hline CA_1^2 \\ CA_1A_2 \\ CA_2^2 \\ \hline \vdots \\ \hline CA_1^{p-1} \\ CA_1^{p-2}A_2 \\ \vdots \\ CA_2^{p-1} \end{array}\right) \Longrightarrow \begin{array}{ll} \underline{\Gamma}\, A_1 & = S_1\Gamma \\ \underline{\Gamma}\, A_2 & = S_2\Gamma \end{array}$$

- Null space of Macaulay
- n shifts A_1, A_2: $A_1A_2 = A_2A_1$
- nD Cayley-Hamilton (new)
- Multi-shift invariant $\mathbf{R}(\Gamma)$ fixed by $\lambda(A_1)$ and $\lambda(A_2)$
- nD observability
- nD realization theory
- nD Beurling-Lax
- 'Block' when $C =$ matrix

Not treated here:

- Deflate roots at infinity
- Algorithms: kernel-driven versus data-driven (QR), SVD for rank decisions, ...
- Cayley-Hamilton (in 1D and nD)
- 1D and nD system theoretic interpretations of the null space (1D and nD observability matrices) based on 1D/nD state space models (possibly singular (roots at infinity))
- ...

Outline

Finding the minimum of univariate polynomial

$$p(x) = \alpha_0 x^n + \alpha_1 x^{n-1} + \ldots + \alpha_n$$

$$\min_{\sigma} \sigma = p(x) \text{ subject to } p'(x) = 0$$

Construct Sylvester matrix M with $\sigma = p(x)$ and $p'(x) = 0$:

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} - \sigma I & M_{22} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0$$

$$(M_{21} - M_{22} M_{12}^{-1} M_{11}) u = u\sigma$$

Generalizes to multivariate polynomial optimization problems:

$$\min_{x \in \mathbb{R}^n} \sigma = p_0(x) \text{ subject to } p_i(x) = 0, i = 1, \ldots, q \ .$$

Lagrangean:

$$\mathcal{L}(x, l) = p_0(x) + \sum_{i=1}^{q} \lambda_i \, p_i(x)$$

Taking derivatives:

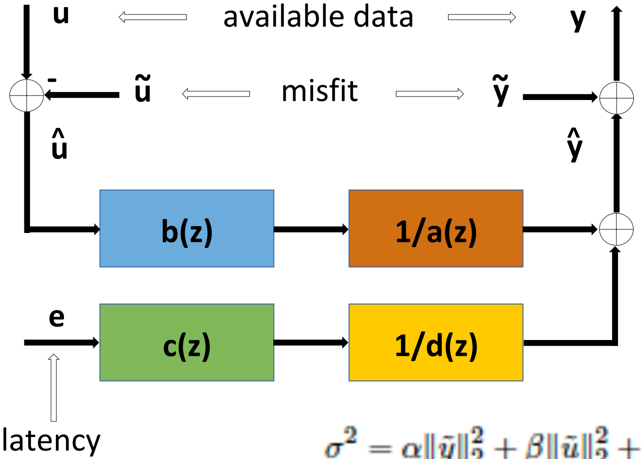
$$\begin{array}{ll} n \text{ equations} & \dfrac{\partial p_0}{\partial x} + \displaystyle\sum_{i=1}^{q} \lambda_i \, \dfrac{\partial p_i}{\partial x} = 0 \ , \\ q \text{ equations} & p_i(x) = 0 \ . \end{array}$$

Hence $q + n$ multivariate polynomial equations in $q + n$ unknowns
$=$ **Eigenvalue problem** !

Remark: only need roots that globally minimize !

Outline

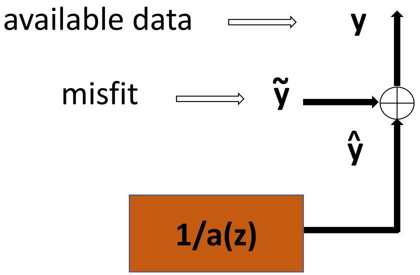
Misfit case: Least squares realization (n_a)



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)



Misfit:
LS realization

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

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

KU LEUVEN

Corresponding difference equation for 'exact' data $\hat{y}_k$ with $\alpha_i \in \mathbb{R}$:

$$\hat{y}_{k+n_a} + \alpha_1 \hat{y}_{k+n_a-1} + \ldots + \alpha_{n_a} \hat{y}_k = 0 \ , \ k \geq 0$$

Equivalently

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

or

$$\begin{pmatrix} \alpha_{n_a} & \alpha_{n_a-1} & \ldots & \ldots & \alpha_1 & 1 & 0 & 0 & \ldots & 0 \\ 0 & \alpha_{n_a} & \alpha_{n_a-1} & \ldots & \alpha_2 & \alpha_1 & 1 & 0 & \ldots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \alpha_{n_a} & \alpha_{n_a-1} & \ldots & \ldots & 1 \end{pmatrix} \begin{pmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \end{pmatrix} = 0$$

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:

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

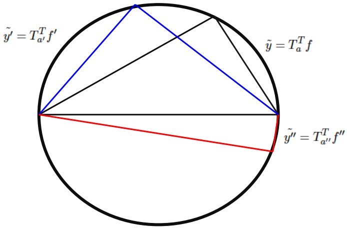
Π_a = orthogonal projector onto row space of T_a. Define $D_a = T_a T_a^T$ and $f = D_a^{-1} T_a y$:

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

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

The least squares residual $\tilde{y}$ is generated by filtering signal f through FIR filter determined by a

$=$ Finite dimensional form of Beurling/Lax/Halmos theorem



Thales theorem

Let

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

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

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

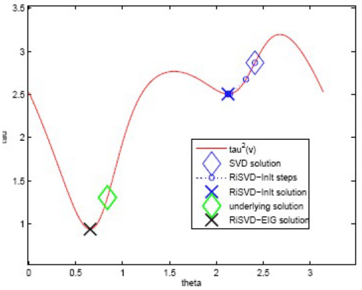
First order optimality conditions and chain rule $\forall\ \alpha_i$:

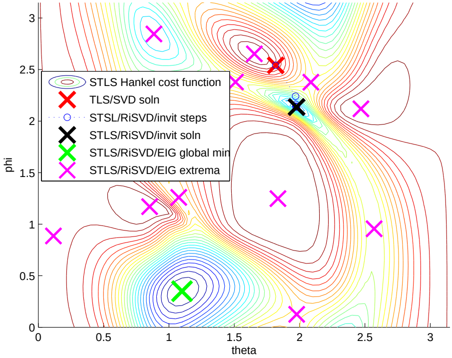
$$\begin{pmatrix} D_a^{\alpha_i} & T_a^{\alpha_i} y \\ y^T (T_a^{\alpha_i})^T & 0 \end{pmatrix} \begin{pmatrix} f \\ -1 \end{pmatrix} + \begin{pmatrix} D_a & T_a y \\ y^T T_a^T & \sigma^2 \end{pmatrix} \begin{pmatrix} f^{\alpha_i} \\ -1 \end{pmatrix} = 0 \qquad (2)$$

This is an MEVP !

- Define $z^T = (\ -1\ f^T\ (f^{\alpha_1})^T\ \ldots (f^{\alpha_{n_a}})^T\ )$.
- Quasi-Toeplitz-ify eqs. (1) - (2) 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 ('generalized' Vandermonde)

$n_a = 1$ $n_a = 2$


3.5
3
2.5
2
1.5
1
tau
0
0.5
1
1.5
2
2.5
3
theta
tau²(v)
SVD solution
RiSVD−InIt steps
RiSVD−InIt solution
underlying solution
RiSVD−EIG solution


3
2.5
2
1.5
1
0.5
0
phi
0
0.5
1
1.5
2
2.5
3
theta
STLS Hankel cost function
TLS/SVD soln
STSL/RiSVD/invit steps
STLS/RiSVD/invit soln
STLS/RiSVD/EIG global min
STLS/RiSVD/EIG extrema

Outline

Conclusions

Least squares identification of LTI models with 'latency' (e.g. Box-Jenkins, ARMAX, ...) or 'misfit' (e.g. errors-in-variables, dynamic total least squares,...), or 'both', requires multivariate polynomial optimization.

Multivariate polynomial optimization is a(n) (multiparameter) eigenvalue problem

Many results and connections not tackled in this presentation (numerical, conceptual, links with algebraic geometry and operator theory, etc.)

Reference (among others, Festschrift Kailath 85): *De Moor B., "Least squares optimal realisation of autonomous LTI systems is an eigenvalue problem", Communications in Information and systems, vol. 20, no. 2, 2020, pp. 163-207.*

Back to the roots !

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