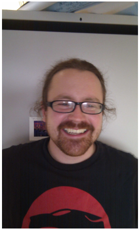
Back to the Roots: Solving Polynomial Systems with Numerical Linear Algebra Tools



Philippe Dreesen

Kim Batselier

Bart De Moor

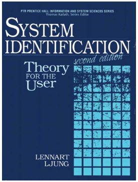
KU Leuven
Department of Electrical Engineering
ESAT-STADIUS

KU LEUVEN

Outline

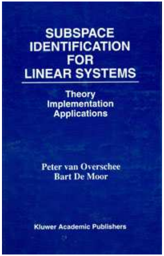
System Identification: PEM

- LTI models
- Non-convex optimization
- Considered 'solved' early nineties



Linear Algebra approach

$\Rightarrow$ **Subspace methods**



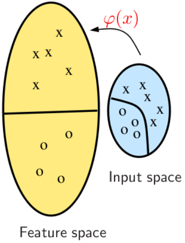
Nonlinear regression, modelling and clustering

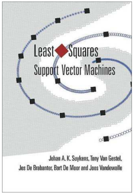
- Most regression, modelling and clustering problems are nonlinear when formulated in the input data space
- This requires nonlinear nonconvex optimization algorithms

Linear Algebra approach

$\Rightarrow$ **Least Squares Support Vector Machines**

- 'Kernel trick' = projection of input data to a high-dimensional feature space
- Regression, modelling, clustering problem becomes a large scale linear algebra problem (set of linear equations, eigenvalue problem)





Nonlinear Polynomial Optimization

- Polynomial object function + polynomial constraints
- Non-convex
- Computer Algebra, Homotopy methods, Numerical Optimization
- Considered 'solved' by mathematics community

Linear Algebra Approach

$\Rightarrow$ **Linear Polynomial Algebra**

Conceptual/Geometric Level

- Polynomial system solving is an eigenvalue problem!
- Row and Column Spaces: Ideal/Variety $\leftrightarrow$ Row space/Kernel of M, ranks and dimensions, nullspaces and orthogonality
- Geometrical: intersection of subspaces, angles between subspaces, Grassmann's theorem,...

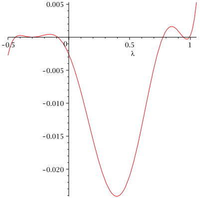
Numerical Linear Algebra Level

- Eigenvalue decompositions, SVDs,...
- Solving systems of equations (consistency, nb sols)
- QR decomposition and Gram-Schmidt algorithm

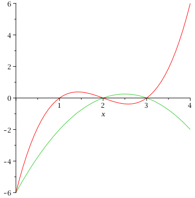
Numerical Algorithms Level

- Modified Gram-Schmidt (numerical stability), GS 'from back to front'
- Exploiting sparsity and Toeplitz structure (computational complexity $O(n^2)$ vs $O(n^3)$), FFT-like computations and convolutions,...
- Power method to find smallest eigenvalue (= minimizer of polynomial optimization problem)

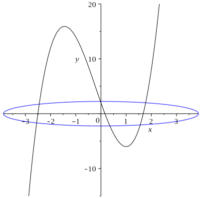
$$p(\lambda) = \det(A - \lambda I) = 0$$



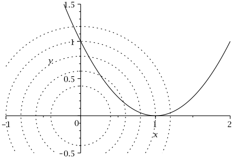
$$\begin{aligned} (x-1)(x-3)(x-2) &= 0 \\ -(x-2)(x-3) &= 0 \end{aligned}$$



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



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



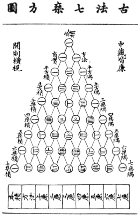
KU LEUVEN

Outline

Solving Polynomial Systems: a long and rich history. . .

Diophantus (c200-c284) *Arithmetica*

Al-Khwarizmi (c780-c850)



Zhu Shijie (c1260-c1320) *Jade Mirror of the Four Unknowns*

Pierre de Fermat (c1601-1665)

René Descartes (1596-1650)

Isaac Newton (1643-1727)

Gottfried Wilhelm Leibniz (1646-1716)

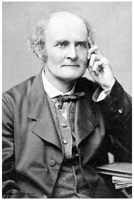
. . . leading to “Algebraic Geometry”

Etienne Bézout
(1730-1783)

Carl Friedrich Gauss
(1777-1755)

Jean-Victor Poncelet
(1788-1867)

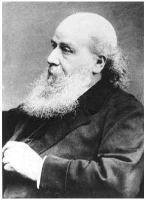
Evariste Galois
(1811-1832)



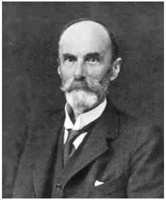
Arthur Cayley
(1821-1895)

Leopold Kronecker
(1823-1891)

Edmond Laguerre
(1834-1886)



James J. Sylvester
(1814-1897)

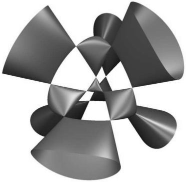


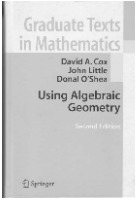
Francis S. Macaulay
(1862-1937)

David Hilbert
(1862-1943)

Computational Algebraic Geometry

- Emphasis on symbolic manipulations
- Computer algebra
- Huge body of literature in Algebraic Geometry
- Computational tools: Gröbner Bases (next slide)





Wolfgang Gröbner
(1899-1980)

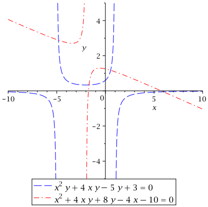
Bruno Buchberger

KU LEUVEN

Example: Gröbner basis

Input system:

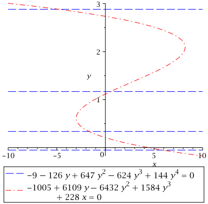
$$\begin{aligned} x^2y + 4xy - 5y + 3 &= 0 \\ x^2 + 4xy + 8y - 4x - 10 &= 0 \end{aligned}$$



- Generates simpler but equivalent system (same roots)
- Symbolic eliminations and reductions
- Monomial ordering (e.g., lexicographic)
- Exponential complexity
- Numerical issues! Coefficients become very large

Gröbner Basis:

$$\begin{aligned} -9 - 126y + 647y^2 - 624y^3 + 144y^4 &= 0 \\ -1005 + 6109y - 6432y^2 + 1584y^3 + 228x &= 0 \end{aligned}$$

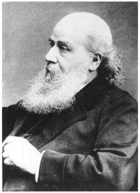


Outline

$$\underset{p\times q}{A}\ \underset{q\times(q-r)}{X} = \underset{p\times(q-r)}{0}$$

- $C(A^T) \perp C(X)$
- $\text{rank}(A) = r$
- $\dim N(A) = q - r = \text{rank}(X)$

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$
$$\Downarrow$$
$$X = V_2$$



James Joseph Sylvester

$$\underset{p\times q}{A}\quad \underset{q\times(q-r)}{X} = \underset{p\times(q-r)}{0}$$

Reorder columns of A and partition

$$\overset{p\times q}{\overline{A}} = \left[\overset{p\times(q-r)}{\overline{A_1}}\quad \overset{p\times r}{\overline{A_2}}\right] \qquad \text{rank}(\overline{A_2}) = r \quad (\overline{A_2} \text{ full column rank})$$

Reorder rows of X and partition accordingly

$$\left[\overline{A_1}\quad \overline{A_2}\right] \overset{q-r}{\begin{bmatrix} \overline{X_1} \\ \overline{X_2} \end{bmatrix}} \begin{matrix} q-r \\ r \end{matrix} = 0$$

$$\boxed{\begin{array}{ccc} \text{rank}(\overline{A_2}) & = & r \\ & \Updownarrow & \\ \text{rank}(\overline{X_1}) & = & q-r \end{array}}$$

$$\begin{bmatrix} \overline{A_1} & \overline{A_2} \end{bmatrix} \overset{q-r}{\begin{bmatrix} \overline{X_1} \\ \overline{X_2} \end{bmatrix}} \begin{matrix} q-r \\ r \end{matrix} = 0$$

- $\overline{X_1}$: independent variables
- $\overline{X_2}$: dependent variables

$$\begin{aligned} \overline{X_2} &= -\overline{A_2}^{\dagger}\, \overline{A_1}\, \overline{X_1} \\ \overline{A_1} &= -\overline{A_2}\, \overline{X_2}\, \overline{X_1}^{-1} \end{aligned}$$

- Number of different ways of choosing r linearly independent columns out of q columns (upper bound):

$$\binom{q}{q-r} = \frac{q!}{(q-r)!\, r!}$$

$$\underset{p\times q}{A}\ \underset{q\times(q-r_A)}{X} = \underset{p\times(q-r_A)}{0} \quad \text{and} \quad \underset{p\times t}{B}\ \underset{t\times(t-r_B)}{Y} = \underset{p\times(t-r_B)}{0}$$

What is the nullspace of $[\,A \quad B\,]$?

$$[\,A \quad B\,]\begin{bmatrix} \overset{q-r_A}{X} & \overset{t-r_B}{0} & \overset{?}{?} \\ 0 & Y & ? \end{bmatrix} = 0$$

Let $\operatorname{rank}([\,A \quad B\,]) = r_{AB}$

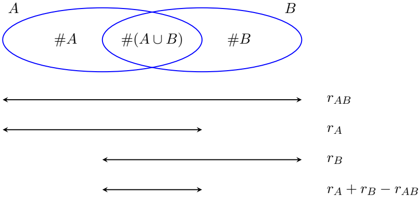
$$(q-r_A)+(t-r_B)+? = (q+t)-r_{AB} \quad \Rightarrow \quad ? = r_A + r_B - r_{AB}$$

KU LEUVEN

$$[\,A \quad B\,]\begin{array}{c} \begin{matrix} q-r_A & t-r_B & r_A+r_B-r_{AB} \end{matrix} \\ \begin{bmatrix} X & 0 & Z_1 \\ 0 & Y & Z_2 \end{bmatrix} \end{array} = 0$$

Intersection between column space of A and B:

$$AZ_1 = -BZ_2$$



Hermann Grassmann

$$\#(A \cup B) = \#A + \#B - \#(A \cap B)$$

KU LEUVEN

- **Characteristic Polynomial**
  The eigenvalues of A are the roots of

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

- **Companion Matrix**
  Solving

$$q(x) = 7x^3 - 2x^2 - 5x + 1 = 0$$

leads to

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1/7 & 5/7 & 2/7 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} = x \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}$$

Consider the univariate equation

$$x^3 + a_1x^2 + a_2x + a_3 = 0,$$

having three distinct roots x_1, x_2 and x_3

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

- Homogeneous linear system
- Rectangular Vandermonde
- $\text{corank} = 3$
- Observability matrix-like
- Realization theory!

Consider

$$\begin{aligned} x^3 + a_1 x^2 + a_2 x + a_3 &= 0 \\ x^2 + b_1 x + b_2 &= 0 \end{aligned}$$

Build the Sylvester Matrix:

$$\left[\begin{array}{ccccc} 1 & a_1 & a_2 & a_3 & 0 \\ 0 & 1 & a_1 & a_2 & a_3 \\ \hline 1 & b_1 & b_2 & 0 & 0 \\ 0 & 1 & b_1 & b_2 & 0 \\ 0 & 0 & 1 & b_1 & b_2 \end{array}\right] \left[\begin{array}{c} 1 \\ x \\ x^2 \\ x^3 \\ x^4 \end{array}\right] = 0$$

Row Space	Null Space
Ideal =union of ideals =multiply rows with powers of x	**Variety** =intersection of null spaces

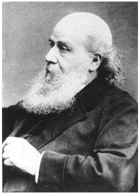
- Corank of Sylvester matrix = number of common zeros
- null space = intersection of null spaces of two Sylvester matrices
- common roots follow from realization theory in null space
- notice ‘double’ Toeplitz-structure of Sylvester matrix

- **Sylvester Resultant**
  Consider two polynomials $f(x)$ and $g(x)$:

$$f(x) = x^3 - 6x^2 + 11x - 6 = (x-1)(x-2)(x-3)$$
$$g(x) = -x^2 + 5x - 6 = -(x-2)(x-3)$$

Common roots iff $S(f,g) = 0$

$$S(f,g) = \det \left[\begin{array}{ccccc} -6 & 11 & -6 & 1 & 0 \\ 0 & -6 & 11 & -6 & 1 \\ \hline -6 & 5 & -1 & 0 & 0 \\ 0 & -6 & 5 & -1 & 0 \\ 0 & 0 & -6 & 5 & -1 \end{array}\right]$$



James Joseph Sylvester

The corank of the Sylvester matrix is 2!

Sylvester's construction can be understood from

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

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

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

$$\begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix} x = \begin{bmatrix} x \\ x^2 \\ x^3 \\ x^4 \end{bmatrix}$$

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

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

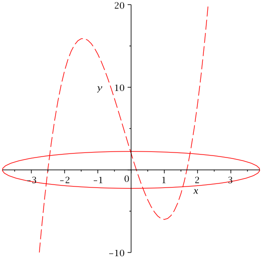
leading to the generalized eigenvalue problem

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

Outline

- Consider

$$\left\{ \begin{array}{lcl} p(x,y) & = & x^2 + 3y^2 - 15 = 0 \\ q(x,y) & = & y - 3x^3 - 2x^2 + 13x - 2 = 0 \end{array} \right.$$



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

- Construct M: write the system in matrix-vector notation:

	1	x	y	x^2	xy	y^2	x^3	x^2y	xy^2	y^3
$p(x,y)$	-15			1		3				
$q(x,y)$	-2	13	1	-2			-3			
$x \cdot p(x,y)$		-15					1		3	
$y \cdot p(x,y)$			-15					1		3

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

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 ⇒ overdetermined system
- rank deficient by construction!

- Coefficient 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 corank

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

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

- Choose s linear independent rows in K

$$S_1 K$$

- This corresponds to finding linear dependent columns in M

$$\begin{bmatrix}
1 & 1 & \ldots & 1 \\
x_1 & x_2 & \ldots & x_s \\
y_1 & y_2 & \ldots & y_s \\
x_1^2 & x_2^2 & \ldots & x_s^2 \\
x_1y_1 & x_2y_2 & \ldots & x_sy_s \\
y_1^2 & y_2^2 & \ldots & y_s^2 \\
x_1^3 & x_2^3 & \ldots & x_s^3 \\
x_1^2y_1 & x_2^2y_2 & \ldots & x_s^2y_s \\
x_1y_1^2 & x_2y_2^2 & \ldots & x_sy_s^2 \\
y_1^3 & y_2^3 & \ldots & y_s^3 \\
x_1^4 & x_2^4 & \ldots & x_4^4 \\
x_1^3y_1 & x_2^3y_2 & \ldots & x_s^3y_s \\
x_1^2y_1^2 & x_2^2y_2^2 & \ldots & x_s^2y_s^2 \\
x_1y_1^3 & x_2y_2^3 & \ldots & x_sy_s^3 \\
y_1^4 & y_2^4 & \ldots & y_s^4 \\
\vdots & \vdots & \vdots & \vdots
\end{bmatrix}$$

Shifting the selected rows gives (shown for 3 columns)

$$\begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ x_1^2 & x_2^2 & x_3^2 \\ x_1y_1 & x_2y_2 & x_3y_3 \\ y_1^2 & y_2^2 & y_3^2 \\ x_1^3 & x_2^3 & x_3^3 \\ x_1^2y_1 & x_2^2y_2 & x_3^2y_3 \\ x_1y_1^2 & x_2y_2^2 & x_3y_3^2 \\ y_1^3 & y_2^3 & y_3^3 \\ x_1^4 & x_2^4 & x_4^4 \\ x_1^3y_1 & x_2^3y_2 & x_3^3y_3 \\ x_1^2y_1^2 & x_2^2y_2^2 & x_3^2y_3^2 \\ x_1y_1^3 & x_2y_2^3 & x_3y_3^3 \\ y_1^4 & y_2^4 & y_3^4 \\ \vdots & \vdots & \vdots \end{bmatrix} \longrightarrow \text{“shift with } x\text{”} \longrightarrow \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ x_1^2 & x_2^2 & x_3^2 \\ x_1y_1 & x_2y_2 & x_3y_3 \\ y_1^2 & y_2^2 & y_3^2 \\ x_1^3 & x_2^3 & x_3^3 \\ x_1^2y_1 & x_2^2y_2 & x_3^2y_3 \\ x_1y_1^2 & x_2y_2^2 & x_3y_3^2 \\ y_1^3 & y_2^3 & y_3^3 \\ x_1^4 & x_2^4 & x_4^4 \\ x_1^3y_1 & x_2^3y_2 & x_3^3y_3 \\ x_1^2y_1^2 & x_2^2y_2^2 & x_3^2y_3^2 \\ x_1y_1^3 & x_2y_2^3 & x_3y_3^3 \\ y_1^4 & y_2^4 & y_3^4 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

simplified:

$$\begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ x_1y_1 & x_2y_2 & x_3y_3 \\ x_1^3 & x_2^3 & x_3^3 \\ x_1^2y_1 & x_2^2y_2 & x_3^2y_3 \end{bmatrix} \begin{bmatrix} x_1 & & \\ & x_2 & \\ & & x_3 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_1^2 & x_2^2 & x_3^2 \\ x_1y_1 & x_2y_2 & x_3y_3 \\ x_1^2y_1 & x_2^2y_2 & x_3^2y_3 \\ x_1^4 & x_2^4 & x_4^4 \\ x_1^3y_1 & x_2^3y_2 & x_3^3y_3 \end{bmatrix}$$

KU LEUVEN

- 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 wrt. shift property

- reminiscent of **Realization Theory**

We have

$$S_1 K D_x = S_x K$$

However, K is not known, instead a basis Z is computed that satisfies

$$ZV = K$$

Which leads to

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

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 results:

- same eigenvectors V!
- $(S_3 Z)^{-1}(S_1 Z)$ and $(S_2 Z)^{-1}(S_1 Z)$ commute

Nullspace of M

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,

$$MZ = 0$$

We have

$$S_1 K D = S_2 K$$

However, K is not known, instead a basis Z is computed as

$$ZV = K$$

Which leads to

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

Realization Theory and Polynomial System Solving

- Attasi model

$$v(k_1, \ldots, k_{i-1}, \mathbf{k_i} + \mathbf{1}, k_{i+1}, \ldots, k_n) = \mathbf{A_i} v(k_1, \ldots, k_n)$$

- Null space of Macaulay matrix: nD state sequence

$$\left(\begin{array}{c|cc|ccc|cccc} | & | & | & | & | & | & | & | & | & | \\ v_{00} & v_{10} & v_{01} & v_{20} & v_{11} & v_{02} & v_{30} & v_{21} & v_{12} & v_{03} \\ | & | & | & | & | & | & | & | & | & | \end{array} \right) =$$

$$\left(\begin{array}{c|cc|c|cccc} | & | & | & & | & | & | & | \\ v_{00} & A_1 v_{00} & A_2 v_{00} & \cdots & A_1^3 v_{00} & A_1^2 A_2 v_{00} & A_1 A_2^2 v_{00} & A_2^3 v_{00} \\ | & | & | & & | & | & | & | \end{array} \right)$$

- shift-invariance property, e.g., for x_2:

$$\begin{pmatrix} -v_{00}- \\ \hline -v_{10}- \\ -v_{01}- \\ \hline -v_{20}- \\ -v_{11}- \\ -v_{02}- \end{pmatrix} A_2^T = \begin{pmatrix} -v_{01}- \\ \hline -v_{11}- \\ -v_{02}- \\ \hline -v_{21}- \\ -v_{12}- \\ -v_{03}- \end{pmatrix},$$

- corresponding nD system realization

$$\begin{aligned} v(k+1,l) &= A_1 v(k,l) \\ v(k,l+1) &= A_2 v(k,l) \\ v(0,0) &= v_{00} \end{aligned}$$

- choice of basis null space leads to different system realizations
- eigenvalues of A_1 and A_2 invariant: x_1 and x_2 components

There are 3 kinds of roots:

1. Roots in zero
2. Finite nonzero roots
3. Roots at infinity

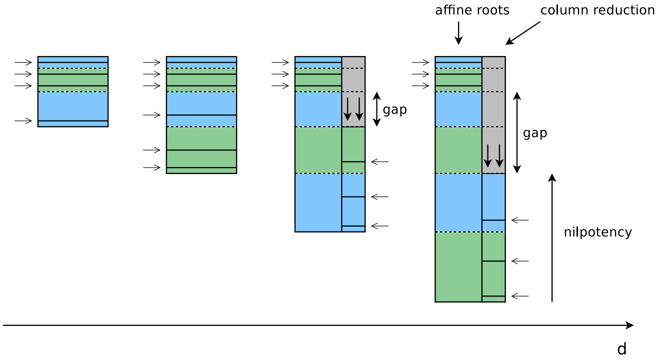
Applying Grassmann's Dimension theorem on the kernel allows to write the following partitioning

$$[M_1 \ M_2] \begin{bmatrix} X_1 & 0 & X_2 \\ 0 & Y_1 & Y_2 \end{bmatrix} = 0$$

- X_1 corresponds with the roots in zero (multiplicities included!)
- Y_1 corresponds with the roots at infinity (multiplicities included!)
- $[X_2; Y_2]$ corresponds with the finite nonzero roots (multiplicities included!)

Mind the Gap!

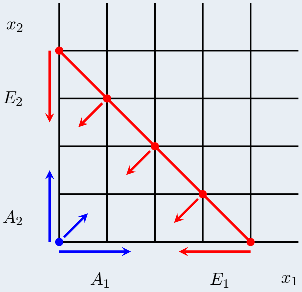
- dynamics in the null space of $M(d)$ for increasing degree d
- nilpotency gives rise to a 'gap'
- mechanism to count and separate affine from infinity



- Kronecker Canonical Form decoupling affine and infinity roots

$$\left(\frac{v(k+1)}{w(k-1)}\right) = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & E \end{array}\right)\left(\frac{v(k)}{w(k)}\right),$$

- Action of A_i and E_i represented in grid of monomials



Roots at Infinity: nD Descriptor Systems

Weierstrass Canonical Form decouples affine/infinity

$$\begin{bmatrix} v(k+1) \\ \hline w(k-1) \end{bmatrix} = \left[\begin{array}{c|c} A & 0 \\ \hline 0 & E \end{array}\right] \begin{bmatrix} v(k) \\ \hline w(k) \end{bmatrix}$$

Singular nD Attasi model (for $n = 2$)

$$\begin{array}{rcl} v(k+1,l) & = & A_x v(k,l) \\ v(k,l+1) & = & A_y v(k,l) \end{array}$$

$$\begin{array}{rcl} w(k-1,l) & = & E_x w(k,l) \\ w(k,l-1) & = & E_y w(k,l) \end{array}$$

with E_x and E_y nilpotent matrices.

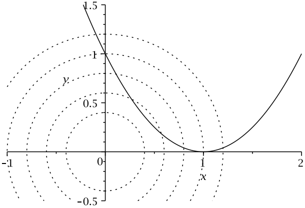
Summary

- solving multivariate polynomials
  - question in linear algebra
  - realization theory in null space of Macaulay matrix
  - nD autonomous (descriptor) Attasi model
- decisions made based upon (numerical) rank
  - $\#$ roots (nullity)
  - $\#$ affine roots (column reduction)
- mind the gap phenomenon: affine vs. infinity roots
- not discussed
  - multiplicity of roots
  - column-space based method
  - over-constrained systems

Outline

Polynomial Optimization Problems

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



Lagrange multipliers give conditions for optimality:

$$L(x, y, z) = x^2 + y^2 + z(y - x^2 + 2x - 1)$$

we find

$$\begin{aligned} \partial L/\partial x = 0 &\quad \rightarrow \quad 2x - 2xz + 2z = 0 \\ \partial L/\partial y = 0 &\quad \rightarrow \quad 2y + z = 0 \\ \partial L/\partial z = 0 &\quad \rightarrow \quad y - x^2 + 2x - 1 = 0 \end{aligned}$$

KU LEUVEN

Observations:

- everything remains polynomial
- system of polynomial equations
- shift with objective function to find minimum/maximum

Let

$$A_x V = xV$$

and

$$A_y V = yV$$

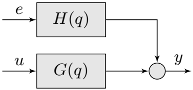
then find min/max eigenvalue of

$$(A_x^2 + A_y^2)V = (x^2 + y^2)V$$

Outline

- PEM System identification
- Measured data $\{u_k, y_k\}_{k=1}^{N}$
- Model structure

$$y_k = G(q)u_k + H(q)e_k$$



- Output prediction

$$\hat{y}_k = H^{-1}(q)G(q)u_k + (1 - H^{-1})y_k$$

- Model classes: ARX, ARMAX, OE, BJ

$$A(q)y_k = B(q)/F(q)u_k + C(q)/D(q)e_k$$

Class	Polynomials
ARX	$A(q), B(q)$
ARMAX	$A(q)$, $B(q)$, $C(q)$
OE	$B(q)$, $F(q)$
BJ	$B(q)$, $C(q)$, $D(q)$, $F(q)$

KU LEUVEN

- Minimize the prediction errors $y - \hat{y}$, where

$$\hat{y}_k = H^{-1}(q)G(q)u_k + (1 - H^{-1})y_k,$$

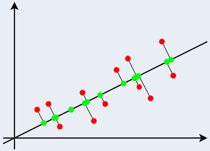
subject to the model equations

- Example

ARMAX identification: $G(q) = B(q)/A(q)$ and $H(q) = C(q)/A(q)$, where $A(q) = 1 + aq^{-1}$, $B(q) = bq^{-1}$, $C(q) = 1 + cq^{-1}$, $N = 5$

$$\begin{aligned} \min_{\hat{y},a,b,c} \quad & (y_1 - \hat{y}_1)^2 + \ldots + (y_5 - \hat{y}_5)^2 \\ \text{s.t.} \quad & \hat{y}_5 - c\hat{y}_4 - bu_4 - (c-a)y_4 = 0, \\ & \hat{y}_4 - c\hat{y}_3 - bu_3 - (c-a)y_3 = 0, \\ & \hat{y}_3 - c\hat{y}_2 - bu_2 - (c-a)y_2 = 0, \\ & \hat{y}_2 - c\hat{y}_1 - bu_1 - (c-a)y_1 = 0, \end{aligned}$$

Static Linear Modeling



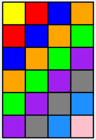
- Rank deficiency
- minimization problem:

$$\begin{array}{ll} \min & \| [\,\Delta A \quad \Delta b\,] \|_F^2 \,, \\ \text{s. t.} & (A + \Delta A) v = b + \Delta b, \\ & v^T v = 1 \end{array}$$

- Singular Value Decomposition:
  find (u, σ, v) which minimizes σ^2
  Let $M = [\,A \quad b\,]$

$$\left\{ \begin{array}{rcl} Mv & = & u\sigma \\ M^T u & = & v\sigma \\ v^T v & = & 1 \\ u^T u & = & 1 \end{array} \right.$$

Dynamical Linear Modeling



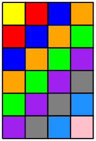
- Rank deficiency
- minimization problem:

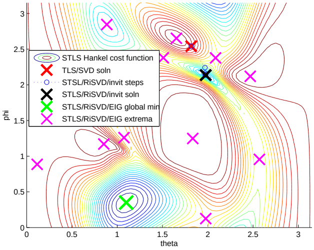
$$\begin{array}{ll} \min & \| [\Delta A \quad \Delta b] \|_F^2 \,, \\ \text{s. t.} & (A + \Delta A) v = b + \Delta b, \\ & v^T v = 1 \\ & [\Delta A \quad \Delta b] \ \text{structured} \end{array}$$

- Riemannian SVD:
  find (u, τ, v) which minimizes τ^2

$$\left\{ \begin{array}{rcl} Mv & = & D_v u \tau \\ M^T u & = & D_u v \tau \\ v^T v & = & 1 \\ u^T D_v u & = & 1 \ (= v^T D_u v) \end{array} \right.$$

$$\begin{array}{ll} \min\limits_{v} & \tau^2 = v^T M^T D_v^{-1} M v \\ \text{s.t.} & v^T v = 1. \end{array}$$





method	TLS/SVD	STLS inv. it.	STLS eig
v_1	.8003	.4922	**.8372**
v_2	-.5479	-.7757	**.3053**
v_3	.2434	.3948	**.4535**
τ^2	4.8438	3.0518	**2.3822**
global solution?	no	no	**yes**

CpG Islands

- genomic regions that contain a high frequency of sites where a cytosine (C) base is followed by a guanine (G)
- rare because of methylation of the C base
- hence CpG islands indicate functionality

Given observed sequence of DNA:

CTCACGTGATGAGAGCATTCTCAGA
CCGTGACGCGTGTAGCAGCGGCTCA

Problem

Decide whether the observed sequence came from a CpG island

The model

- 4-dimensional state space $[m] = \{\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}\}$
- Mixture model of 3 distributions on $[m]$
  1. : CG rich DNA
  2. : CG poor DNA
  3. : CG neutral DNA
- Each distribution is characterised by probabilities of observing base A,C,G or T

Table : Probabilities for each of the distributions (Durbin; Pachter & Sturmfels)

DNA Type	A	C	G	T
CG rich	0.15	0.33	0.36	0.16
CG poor	0.27	0.24	0.23	0.26
CG neutral	0.25	0.25	0.25	0.25

- The probabilities of observing each of the bases A to T are given by

$$\begin{array}{rcl} p(A) & = & -0.10\,\theta_1 + 0.02\,\theta_2 + 0.25 \\ p(C) & = & +0.08\,\theta_1 - 0.01\,\theta_2 + 0.25 \\ p(G) & = & +0.11\,\theta_1 - 0.02\,\theta_2 + 0.25 \\ p(T) & = & -0.09\,\theta_1 + 0.01\,\theta_2 + 0.25 \end{array}$$

- θ_i is probability to sample from distribution i $(\theta_1 + \theta_2 + \theta_3 = 1)$
- Maximum Likelihood Estimate:

$$(\hat{\theta_1}, \hat{\theta_2}, \hat{\theta_3}) = \arg\max_{\theta} \; l(\theta)$$

where the log-likelihood $l(\theta)$ is given by

$$l(\theta) = 11\log p(A) + 14\log p(C) + 15\log p(G) + 10\log p(T)$$

- Need to solve the following polynomial system

$$\left\{ \begin{array}{rcccl} \frac{\partial l(\theta)}{\partial \theta_1} & = & \sum_{i=1}^{4} \frac{u_i}{p(i)} \frac{\partial p(i)}{\partial \theta_1} & = & 0 \\ \frac{\partial l(\theta)}{\partial \theta_2} & = & \sum_{i=1}^{4} \frac{u_i}{p(i)} \frac{\partial p(i)}{\partial \theta_2} & = & 0 \end{array} \right.$$

Solving the Polynomial System

- $\text{corank}(M) = 9$
- Reconstructed Kernel

$$K = \begin{bmatrix} 1 & 1 & 1 & 1 & \ldots \\ 0.52 & 3.12 & -5.00 & 10.72 & \ldots \\ 0.22 & 3.12 & -15.01 & 71.51 & \ldots \\ 0.27 & 9.76 & 25.02 & 115.03 & \ldots \\ 0.11 & 9.76 & 75.08 & 766.98 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{matrix} 1 \\ \theta_1 \\ \theta_2 \\ \theta_1^2 \\ \theta_1\theta_2 \\ \vdots \end{matrix} .$$

- θ_i's are probabilities: $0 \leq \theta_i \leq 1$
- Could have introduced slack variables to impose this constraint!
- Only solution that satisfies this constraint is $\hat{\theta} = (0.52, 0.22, 0.26)$

Applications are found in

- Polynomial Optimization Problems
- Structured Total Least Squares
- Model order reduction
- Analyzing identifiability nonlinear model structures
- Robotics: kinematic problems
- Computational Biology: conformation of molecules
- Algebraic Statistics
- Signal Processing
- ...

Outline

Conclusions

- Finding roots: linear algebra and realization theory!
- Polynomial optimization: extremal eigenvalue problems
- (Numerical) linear algebra/systems theory translation of algebraic geometry/symbolic algebra
- These relations 'convexify' (linearize) many problems
  - Algebraic geometry
  - System identification (PEM)
  - Numerical linear algebra (STLS, affine EVP $Ax = x\lambda + a$, etc.)
  - Multilinear algebra (tensor least squares approximation)
  - Algebraic statistics (HMM, Bayesian networks, discrete probabilities)
  - Differential algebra (Glad/Ljung)
- Convexification: projecting up to higher dimensional space (difficult in low number of dimensions; 'easy' in high number of dimensions)

Open Problems

Many challenges remain!

- Efficient construction of the eigenvalue problem - exploiting sparseness and structure
- Algorithms to find the minimizing solution directly (inverse power method?)
- Unraveling structure at infinity (realization theory)
- Positive dimensional solution set: parametrization eigenvalue problem
- nD version of Cayley-Hamilton theorem
- ...

Questions?

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