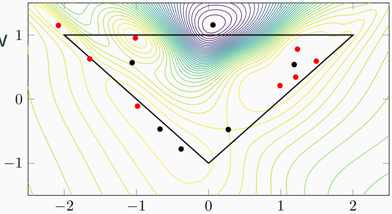
MAX PLANCK INSTITUTE
FOR MATHEMATICS
IN THE SCIENCES

The Algebraic Degree of System Identification and Model Order Reduction in LTI System Theory

CSBD Math Seminar

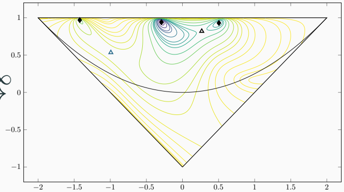
Leonie Kayser & Sibren Lagauw
leokayser.github.io

April 30, 2026



$5 \to \infty$

$\longrightarrow$



Today’s menu

Linear Time-Invariant Dynamical Systems

System Identification

Model Order Reduction

To Infinity And Beyond

Can a mathematician and an engineer be friends?

Back to the roots …

‘Back to the roots’ concentrates on time series, observations generated by systems, the signals of which evolve as a function of time. Examples include traffic systems, data from electrical power consumption, medical patient monitoring records or time-stamped data from complex industrial processes. Starting from such data, using advanced numerical algorithms, mathematical models are computed, which are then used to simulate and predict, to monitor the state of the system and whenever necessary adapt or control its behavior. Before a model is calculated, first one has to specify a model class, for instance neural networks or linear dynamical systems. From the data in the time-series, one then finds the best model in the model class, using large scale numerical optimization algorithms.

Every day, this methodology is massively deployed in thousands of applications within the AI community. Yet, there are still fundamental, open problems. Despite the fact the qualification of ‘best’ is well defined, paradoxically there is no guarantee that optimization algorithms actually also find the best model. Additionally, when calculations are repeated, e.g., with other initial values, most often other ‘solutions’ are found too, that are better or worse. Said in other words, current day optimization practices are heuristic, deliver results that are not necessarily reproducible and therefore difficult to interpret. In many applications, especially the critical ones (safety, health, industrial process control, etc.), the obtained results can not be fully trusted and are therefore often unreliable.

A mathematician and an engineer walk into a bar

Engineer	Mathematician
Uses calculus	Teaches calculus
Can build a bridge but doesn't know why it holds	Will count the number of possible bridges
Likes two columns	Hates two columns
Has funding from industry	
Likes the smell of whiteboard markers	Crazy for specific chalk from Japan
Cares about real solutjons	Invents imaginary numbers and points at ∞ just to be right
Wants solutions quickly	Wants correct solutions

Linear Time-Invariant Dynamical Systems

Realization of linear time-invariant difference equations

$$a_0 \hat{y}_i + a_1 \hat{y}_{i+1} + \cdots + a_r \hat{y}_{i+r} = 0, \qquad i = 0, \dots, N-1-r$$

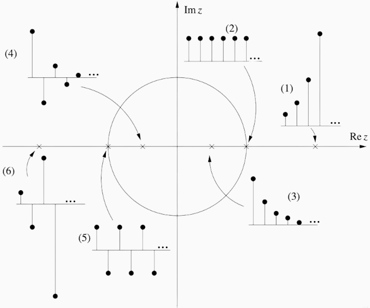
- Discrete-time (physical) system generating signals $y = (y_0, y_1, \dots, y_{N-1})^\top \in \mathbb{R}^N$
- Explain observed data with a mathematical model
- Impose a model class: *autonomous LTI models of finite order*
  - autonomous = no input signals, no influence from outside world
  - linear = linear relation between past outputs
  - time-invariant = coefficients $a = (a_0, \dots, a_r)^\top$ are independent of time
  - finite order r = the relation involves at most r past outputs
- $\hat{y}$ "model compliant" data

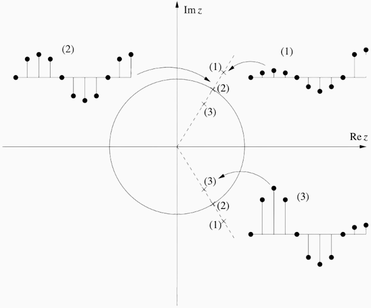
Roots of $a(z) = \sum_{i=0}^{r} a_i z^i$ determine dynamics of model

- **Simple roots:** Each root λ generates mode $\mathrm{vand}(\lambda) = (1, \lambda, \lambda^2, \ldots, \lambda^{N-1})^\mathsf{T}$

$$\hat{y} = \sum_{\lambda} c_\lambda \cdot \mathrm{vand}(\lambda) = \left[\sum_{\lambda} c_\lambda \cdot \lambda^k\right]_{k=0}^{N-1}$$

- **Multiple roots** introduce *confluent Vandermonde vectors* $\frac{\partial^j}{\partial \lambda^j} \mathrm{vand}(\lambda)$
- **Magnitude** of λ's determines growth or decay, **argument** determines phase





Exact realization = Linear Algebra

- Model population of rabbits $\hat{y} = (2, 3, 5, 8, 13)^{\mathsf{T}}$
- $T^{a}_{N-r}\hat{y} = 0$ is equivalent to $H^{\hat{y}}_{r} a = 0$
- $\hat{y}$ satisfies LTI difference equation iff $\operatorname{rank} H^{\hat{y}}_{r} \leq r$, all such $\hat{y}$ form a variety

$$X_{N-1,r} := \left\{ \hat{y} \in \mathbb{C}^{N} \;\middle|\; \operatorname{rank} H^{\hat{y}}_{r} \leq r \right\}$$

- Identify model a via kernel of Hankel matrix, $\operatorname{Ker}\begin{bmatrix} 2 & 3 & 5 \\ 3 & 5 & 8 \\ 5 & 8 & 13 \end{bmatrix} = \mathbb{R}\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$

$$\underbrace{\begin{bmatrix} a_0 & a_1 & \cdots & a_r & & & \\ & a_0 & a_1 & \cdots & a_r & & \\ & & \ddots & \ddots & \ddots & \ddots & \\ & & & a_0 & a_1 & \cdots & a_r \end{bmatrix}}_{=:\, T^{a}_{N-r} \text{ Toeplitz matrix } (N-r)\times N} \begin{pmatrix} \hat{y}_0 \\ \vdots \\ \hat{y}_{N-1} \end{pmatrix} = \underbrace{\begin{bmatrix} \hat{y}_0 & \hat{y}_1 & \cdots & \hat{y}_r \\ \hat{y}_1 & \hat{y}_2 & \cdots & \hat{y}_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{N-r-1} & \hat{y}_{N-r} & \cdots & y_{N-1} \end{bmatrix}}_{=:\, H^{\hat{y}}_{r} \text{ Hankel matrix } (N-r)\times(r+1)} \begin{pmatrix} a_0 \\ \vdots \\ a_r \end{pmatrix} \overset{!}{=} 0$$

Least squares realization

- Fix a 2-norm on $\mathbb{R}^N$, $Q(y) = \frac{1}{2}\|y\|^2 = \frac{1}{2} y^\mathsf{T} \Lambda y$, for example $Q(y) = \sum_{i=1}^N y_i^2$
- Real world scenario: Don't have access to $\hat{y}$, measure noisy $y = \hat{y} + \varepsilon$

$\rightsquigarrow$ y never satisfies a difference equation exactly, $\operatorname{rank} H_r^y = r + 1$ almost surely

- System Identification problem: Given data $y \in \mathbb{R}^N$, identify the system of order r whose response $\hat{y}$ minimizes the distance $\|y - \hat{y}\|$
- If ε is Gaussian white noise, then closest $\hat{y}$ is maximum likelihood estimator

$$\hat{y} = \underset{\hat{y} \in X_{N-1,r}(\mathbb{R})}{\operatorname{argmin}} \|y - \hat{y}\|^2 = \underset{\hat{y} \in X_{N-1,r}(\mathbb{R})}{\operatorname{argmin}} \mathcal{L}(\hat{y} \mid y = \hat{y} + \varepsilon)$$

Let's get some input: SISO LTI systems

- Extend the model class to allow for an input $(u_k)_{k\in\mathbb{N}}$ at each time step k
- Consider discrete-time single-input (u_k) single-output (y_k) LTI models
- State space description: states $x_0, x_1, \ldots \in \mathbb{R}^r$ and

$$x_{k+1} = Ax_k + Bu_k, \quad y_k = C^\top x_k + d \cdot u_k, \qquad A \in \mathbb{R}^{r\times r},\ B, C \in \mathbb{R}^r,\ d \in \mathbb{R}$$

- A is the state transition matrix; will always assume initial state $x_0 = 0$
- The impulse response $h := \mathrm{response}[\delta]$, $\delta = (1, 0, 0, \ldots)$ is

$$h_k = \begin{cases} d & k = 0 \\ C^\top A^{k-1} B & k \geq 1 \end{cases}$$

- Any response is convolution with impulse response: $\mathrm{response}[u] = h * u$
- Distance of systems is ℓ^2-norm of impulse responses $\|h - \hat{h}\|_{\ell^2}$

Transfer functions

- $\mathcal{Z}$-transform: If $y \in \ell^2$, then

$$\mathcal{Z}\{y\}(z) := \sum_{k \geq 0} y_k z^{-k}$$

- Multiplicative $\mathcal{Z}(y * y') = \mathcal{Z}(y) \cdot \mathcal{Z}(y')$, therefore

$$\mathcal{Z}\{y\}(z) = H(z) \cdot \mathcal{Z}\{u\}(z) \qquad H := \mathcal{Z}\{h\} = C^{\mathsf{T}}(Iz - A)^{-1}B + d$$

- Parseval identity:

$$\|h\|_{\ell^2}^2 = \|H\|_{\mathcal{H}_2}^2 := \frac{1}{2\pi \mathrm{i}} \oint_{\mathbb{S}^1} |H(z)|^2 \frac{\mathrm{d}z}{z}$$

  where $\mathbb{S}^1 \subset \mathbb{C}$ is the contour around the unit circle
- State-space models can be “faithfully” represented by the rational function H

Kronecker's theorem on linear difference equations

- $H = C^\mathsf{T}(Iz - A)^{-1}B + d$ is a rational function

Theorem (Kronecker)

The following are equivalent for $\hat{h} = (\hat{h}_1, \hat{h}_2, \dots) \in \prod_{k \geq 1} \mathbb{C}$*:*

1. $\hat{h}$ *satisfies a linear recurrence relation of minimal order* r.
2. *The formal power series* $\widehat{H}(z) = \sum_{k=1}^{\infty} \hat{h}_k z^{-k} \in \mathbb{C}[\![z^{-1}]\!]$ *is a rational function:* $\widehat{H}(z) = b(z)/a(z)$, $\gcd(a, b) = 1$, *and* $r = \deg a > \deg b$.

The polynomial $a(z)$ *in 2. is the char.polynomial of the minimal recurrence relation.*

- If the state space model is "minimal", then $a(z) = \mathrm{charpol}_A(z)$
- $a(z)$ governs the behavior/dynamics of the system

Stable polynomials and the main optimization problem

- A polynomial $a(z)$ is (Schur-)stable if all roots in open unit disc

Corollary

The following are equivalent for $\hat{h} = (\hat{h}_1, \hat{h}_2, \dots)$*:*

1. $\hat{h}$ *satisfies a linear recurrence relation of minimal order* r *and* $\hat{h} \in \ell^2$*;*
2. *The* $\mathcal{Z}$*-transform* $\widehat{H}(z)$ *is a rational function:* $\widehat{H}(z) = b(z)/a(z)$, $\gcd(a, b) = 1$, *and* $r = \deg a > \deg b$, *and* $a(z)$ *stable.*

- Model order reduction problem: Given model via transfer function $H = d/c$ of order R, minimize the distance $\|H - \widehat{H}\|_{\mathcal{H}_2}$ to models $\widehat{H} = b/a$ of order r
- Suffices to assume $H, \widehat{H}$ to be strictly proper (if not, match $d = \hat{d}$)
- Can be seen as "limit" of system identification as $N \to \infty$

System Identification

Our constained optimization problem

- ▷ Norm $Q(y) = \frac{1}{2}\|y\|^2 = \frac{1}{2}y^\mathsf{T}\Lambda y$
- ▷ System Identification problem:
  - Given data: $y \in \mathbb{R}^N$
  - Unknowns: $\hat{y} \in \mathbb{R}^N$, $a \in \mathbb{R}[z]_{\leq r}$

  Identify the system of order r whose response $\hat{y}$ minimizes the distance $\|y - \hat{y}\|$
- ▷ Constraint optimization problem: Impose rank condition on $\hat{y}$

$$
\begin{aligned}
& \underset{\hat{y}\in\mathbb{R}^N}{\text{minimize}} \; Q(y-\hat{y}) && \text{subject to} && \operatorname{rank} H_r^{\hat{y}} \leq r \\
\iff \quad & \underset{\hat{y}\in\mathbb{R}^N,\, a\in\mathbb{R}^r\setminus 0}{\text{minimize}} \; Q(y-\hat{y}) && \text{subject to} && H_r^{\hat{y}} \cdot a = 0
\end{aligned}
$$

Euclidean Distance Degree

- Given a variety $X \subseteq \mathbb{C}^N$ and a point $y \in \mathbb{R}^N$, find closest point on $X(\mathbb{R})$
- Distance measured using non-degenerate quadric $Q(y) = y^\mathsf{T} \Lambda y$

Definition (Euclidean distance degree, $\mathrm{EDD}_Q(X)$)

The number $\mathrm{EDD}_Q(X)$ of complex critical points of $\hat{y} \mapsto Q(\hat{y} - y)$ on X_{reg} for general $y \in \mathbb{R}^N$ is the Euclidean Distance degree of X (with respect to Q).

- For generic quadric obtain generic EDD; upper bound on specific $\mathrm{EDD}_Q(X)$
- Here $X = X_{N-1,r}$ is the r-th secant variety of the rational normal curve $\nu_{N-1}(\mathbb{P}^1)$
- $\mathrm{EDD}_Q(\mathcal{X}_{N-1,r})$ is algebraic degree of system identification

Heuristic approaches

- First idea goes back to Prony [dP95]
- Cadzow's method [Cad88] (assume standard norm on $\mathbb{R}^N$)
  1. Compute SVD of $H_r^y = U\Sigma V^\top$, singular values $\sigma_1 \geq \cdots \geq \sigma_{r+1} > 0$
  2. Setting $\sigma_{r+1} \rightsquigarrow 0$ yields rank-deficient matrix H', but lose Hankel structure
  3. Approximate H' by Hankel matrix $H_r^{y'}$, lose rank-deficiency
  4. Iterate 1.-3. until convergence to rank-deficient Hankel matrix
- Eckart–Young theorem: SVD gives optimal low rank approximation of a matrix
- Other heuristic approaches: iterative quadratic maximum likelihood (IQML), Steiglitz–McBride, for a comparison see [LVVHDM01]
- What if we care about *global* minima?

Let's get FONCy!

- ▷ $X_{N-1,r}$ is not smooth, $\{\, (y, a) \in \mathbb{C}^N \times \mathbb{P}^r \mid H_r^y \cdot a = 0 \,\}$ is desingularization
- ⇝ Prefer this formulation of the optimization problem

$$\underset{\hat{y} \in \mathbb{R}^N,\, a \in \mathbb{R}^r \setminus 0}{\text{minimize}} \; Q(y - \hat{y}) \qquad \text{subject to} \qquad H_r^{\hat{y}} \cdot a = 0 = T_{N-r}^{a} \cdot \hat{y}$$

- ▷ Introduce Lagrange multipliers $\ell \in \mathbb{R}^{N-r}$ to make unconstrained problem

$$\mathcal{L}_y(\hat{y}, a, \ell) = Q(\hat{y} - y) + \ell^{\mathsf{T}} \cdot H_r^{\hat{y}} \cdot a$$

- ▷ First order necessary conditions for optimality:

$$0 \overset{!}{=} \frac{\partial \mathcal{L}_y}{\partial \hat{y}} = \Lambda(\hat{y} - y) + (T_{N-r}^{a})^{\mathsf{T}} \ell$$

$$0 \overset{!}{=} \frac{\partial \mathcal{L}_y}{\partial a} = (H_r^{\hat{y}})^{\mathsf{T}} \ell = T_{N-2r}^{\ell} \hat{y}, \qquad 0 \overset{!}{=} \frac{\partial \mathcal{L}_y}{\partial \ell} = H_r^{\hat{y}} a = T_{N-r}^{a} \hat{y}$$

Lower-rank solutions are never optimal

Lemma

*If $(\hat{y}, a, \ell)$ is a solution to the FONC with $\operatorname{rank} H_r^{\hat{y}} \leq r - 1$, then $\hat{y}$ is **not** a local minimum of $Q(\hat{y} - y)$ on X_r.*

Idea: Can use additional degrees of freedom $\hat{y} + c \cdot \operatorname{vand}(\lambda)$ to decrease norm

Theorem (Characterization of rank r solutions)

Consider a solution $(\hat{y}, a, \ell)$, interpret $a \in S_{\leq r} := \mathbb{R}[z]_{\leq r}$, $\ell \in \mathbb{R}^{N-r} = S_{\leq N-r-1}$.

1. *If $\operatorname{rank} H^{\hat{y}} = r$, then $\ell = g \cdot a$ (as polynomials) for some $g \in S_{\leq N-2r-1}$*
2. *If y is sufficiently random, then $\ell = g \cdot a$ also implies $\operatorname{rank} H^{\hat{y}} = r$.*

Idea: 1. Linear algebra (apolarity) 2. Dimension argument

Putting it all together

$$0 \overset{!}{=} \frac{\partial \mathcal{L}_y}{\partial \hat{y}} = \Lambda(\hat{y} - y) + (T^a_{N-r})^\mathsf{T} \ell \qquad \ell \overset{!}{=} g \cdot a$$

$$0 \overset{!}{=} \frac{\partial \mathcal{L}_y}{\partial a} = T^\ell_{N-2r} \hat{y} \qquad 0 \overset{!}{=} \frac{\partial \mathcal{L}_y}{\partial \ell} = T^a_{N-r} \hat{y}$$

- First equation allows to eliminate $\hat{y}$: $\hat{y} := y - \Lambda^{-1}(a \cdot \ell)$
- Assuming y is general, we can substitute $\ell := g \cdot a$ and simplify

Theorem

For general y, the FONC solutions $(\hat{y}, a, \ell)$ correspond to solutions (a, g) to

$$T^a_{N-r} y \quad = \quad T^a_{N-r} \Lambda^{-1} (T^a_{N-r})^\mathsf{T} (T^a_{N-2r})^\mathsf{T} g \quad = \quad T^a_{N-r} \Lambda^{-1} (a^2 \cdot g).$$

The isomorphism is given by $\ell = a \cdot g$, $\hat{y} = y - \Lambda^{-1}(a^2 \cdot g)$.

The bad locus

- Reduced to system of $N-r$ polynomial equations in $(a,g) \in (\mathbb{C}^{r+1} \setminus 0) \times \mathbb{C}^{N-2r}$

$$T^a_{N-r} y = B_\Lambda(a) g, \qquad B_\Lambda(a) := T^a_{N-r} \Lambda^{-1} (T^a_{N-r})^\mathsf{T} (T^a_{N-2r})^\mathsf{T}$$

- Almost linear in g, homogenize by g_{-1}

$$YAG := \{\, (y, a, (g_{-1} : g)) \mid T^a_{N-r} y \cdot g_{-1} = B_\Lambda(a) g \,\} \subseteq \mathbb{C}^N \times \mathcal{G} \times \mathbb{P}^{N-2r}$$

- g_{-1} can vanish if and only if $B_\Lambda(a)$ becomes rank-deficient for some $a \neq 0$
- Good locus $\mathcal{G} := \{\, a \in \mathbb{C}^{r+1} \mid \operatorname{rank} B_\Lambda(a) = N - 2r \,\}$, bad locus $\mathcal{B} := \mathbb{C}^{r+1} \setminus \mathcal{G}$

Lemma

YAG is a smooth irreducible global complete intersection of dimension $N+1$ and codimension $N-r$ in $\mathbb{C}^N \times \mathcal{G} \times \mathbb{P}^{N-2r}$

Assumption: The set $\mathbb{P}(\mathcal{B})$ should be finite. General Λ: $\mathbb{P}(\mathcal{B}) = \emptyset$

The multi-parameter eigenvalue problem

- Rearrange polynomial system to reveal MEP structure

$$T^a_{N-r} y \cdot g_{-1} = B_\Lambda(a) \cdot g \qquad \Longleftrightarrow \qquad \underbrace{\left[T^a_{N-r} y \mid B_\Lambda(a)\right]}_{=:\,M(a,y)} \cdot \begin{pmatrix} -g_{-1} \\ g \end{pmatrix} = 0$$

- This is almost homogeneous in y, after projecting onto (a, y) we have

$$AY \coloneqq \{\, (a, y) \mid \operatorname{rank} M(a, y) \leq N - 2r \,\} \subseteq \mathbb{P}(\mathcal{G}) \times \mathbb{P}^{N-1}$$

- AY has the structure of a projective subbundle $\mathbb{P}(\mathcal{F}) \subseteq \mathbb{P}(\mathcal{O}^N_{\mathbb{P}(\mathcal{G})})$

Theorem

AY is a smooth irreducible variety of dimension $N-1$ and codimension r in $\mathbb{P}^{N-1} \times \mathbb{P}\mathcal{G}$.

- Restricting to a (general) $y \in \mathbb{P}^{N-1}$, we obtain a finite reduced set of solutions!

Intersection theory saves the day

$$AY := \{ (a,y) \mid \operatorname{rank} M(a,y) \le k \} \subseteq \mathbb{P}^r \times \mathbb{P}^{N-1}, \qquad k := N - 2r$$

- Assume $\mathcal{B} = \emptyset$, satisfies for general Λ
- AY has the *expected dimension* 0, hence Porteous formula applies
- $M(a,y) = [T^a_{N-r} y \mid B_\Lambda(a)]$ has entries of degree $(1,1)$ and $(3,0)$ (k columns)

Theorem (A formula for $\mathrm{EDD}_{\mathrm{gen}}(X_r)$)

In the Chow ring $\mathrm{A}^\bullet(\mathbb{P}^r \times \mathbb{P}^{N-1}) = \mathbb{Z}[\alpha,\beta]/\langle \alpha^{r+1}, \beta^N \rangle$ *we have*

$$[AY] = \left\{ \frac{1}{(1-(\alpha+\beta))(1-3\alpha)^k} \right\}^r = \sum_{j=0}^{r} \sum_{i=0}^{j} \binom{k+r}{j-i} \binom{k-1+i}{i} 2^i \alpha^j \beta^{r-j}.$$

For general y, *the number of solutions is* $\sum_{i=0}^r \binom{k+r}{r-i}\binom{k-1+i}{i} 2^i = \sum_{j=0}^r \binom{k-1+j}{j} 3^j$.

What if the bad locus is non-empty?

- $\mathbb{P}(\mathcal{B}) = \emptyset$ iff $B_\Lambda(a) = T^a_{N-r}\Lambda^{-1}(T^a_{N-r})^{\mathsf{T}}(T^a_{N-2r})^{\mathsf{T}}$ has full rank for all $a \neq 0$
- Recovers formula for $\mathrm{EDD}_{\mathrm{gen}}(X_r)$ from [OSS14, Theorem 3.7]
- If $\mathbb{P}(\mathcal{B})$ is non-empty but finite, then the determinantal formula still applies:

$$\mathrm{EDD}_\Lambda(X_r) = \sum_{j=0}^{r} \binom{k-1+j}{j} 3^j - (\text{multiplicity of } \mathcal{B} \text{ in ideal of minors of } M(a,y))$$

Theorem

Assume that $\mathbb{P}(\mathcal{B})$ is finite. One has

$$\mathrm{EDD}_{\mathrm{gen}}(X_r) - \deg \mathcal{B}^{\mathit{red}} \geq \mathrm{EDD}_\Lambda(X_r) \geq \mathrm{EDD}_{\mathrm{gen}}(X_r) - \deg(\textit{minors of } B_\Lambda(a)).$$

The latter inequality is strict if and only if the multiplicity structure of $\mathcal{B}$ in the ideal of minors of $M(a,y)$ does *depend on y. This can be verified explicitly.*

$\mathrm{EDD}_\Lambda(\mathcal{X}_{N-1,r})$ for special weights

$N \setminus r$	$\Lambda = 1$ Unit				$\Lambda = F$ Frobenius				$\Lambda = \Theta$ Bombieri			
	1	2	3	4	1	2	3	4	1	2	3	4
2	1				1				1			
3	4				2				2			
4	7	1			7	1			3	1		
5	10	13			6	9			4	7		
6	13	34	1		13	34	1		5	16	1	
7	16	64	40		10	38	34		6	28	20	
8	19	103	142	1	19	103	142	1	7	$\underline{43}..45$	$\underline{62}..64$	1
9	22	151	334	121	14	103	246	113	8	$\underline{61}..65$	$\underline{134}..142$	53
10	25	208	643	547	25	208	643	543	9	$\underline{82}..88$	$\underline{243}..263$	229

- Bombieri weights for $\mathcal{X}_{N-1,r}$ gives the first case where previous inequality is strict
- Efficient implementation in `Macaulay2` for extensive experimentation

Model Order Reduction

Checkpoint: The MOR problem, formally

- Given $H(z) = \frac{d(z)}{c(z)}$, $\deg d < \deg c = R$, $c(z)$ stable,

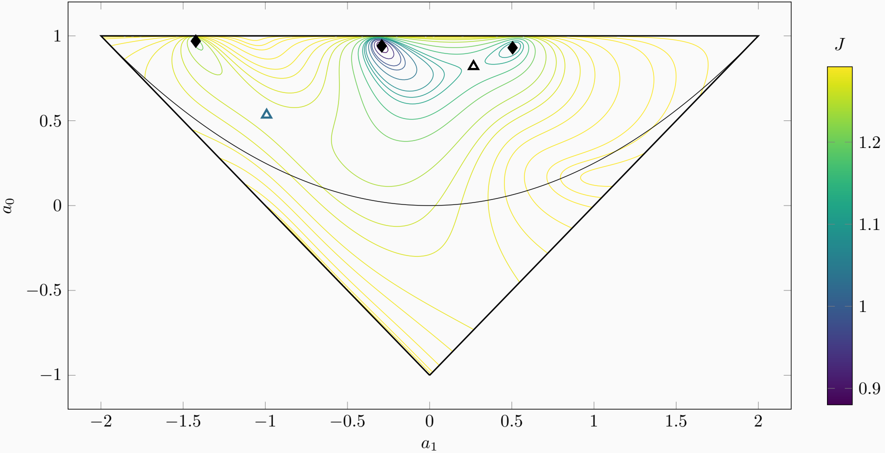
$$\underset{\widehat{H}=\frac{b(z)}{a(z)}}{\text{minimize}}\ J(\widehat{H}) \qquad \text{subject to } \deg b < \deg a \leq r,\ a(z) \text{ stable}$$

$$J(\widehat{H})^2 = \|\widehat{H} - H\|_{\mathcal{H}_2}^2 = \frac{1}{2\pi \mathrm{i}} \oint_{\mathbb{S}^1} |\widehat{H}(z) - H(z)|^2 \frac{\mathrm{d}z}{z}$$

- Given data: $(c, d) \in \mathbb{R}[z]_{<R} \times \mathbb{R}[z]_{\leq R}$
- Unknowns: $(a, b) \in \mathbb{R}[z]_{<r} \times \mathbb{R}[z]_{\leq r}$, $2r + 1$ coefficients (-1 up to scaling)
- Questions:
  1. Describe critical equations in the coefficients of a, b
  2. Is the critical locus finite, non-degenerate for *generic* c, d?
  3. What is the number of (complex) critical points in terms of R, r?

Example: Model order reduction from order 6 to 2

$$H(z) = \frac{0.0448z^5 + 0.2368z^4 + 0.0013z^3 + 0.0211z^2 + 0.2250z + 0.0219}{z^6 - 1.2024z^5 + 2.3675z^4 - 2.0039z^3 + 2.2337z^2 - 1.0420z + 0.8513}$$



Walsh's result on optimality

- If $\widehat{H}$ is a proper rational function $\frac{b}{a}$ in reduced form, then $\operatorname{rank}\widehat{H} := \deg a$

Lemma

If $\widehat{H}$ is a local minimum of J, then $\widehat{H}$ has rank r: $\deg a = r$ and $\gcd(a, b) = 1$.

Theorem (Walsh 1932)

Let $\widehat{H}$ be of rank exactly r, then $\widehat{H}$ is a critical point to $J(\widehat{H})$ if and only if

$$(zH(z))^{(j)}(\omega^{-1}) \overset{!}{=} (z\widehat{H}(z))^{(j)}(\omega^{-1}) \qquad \text{for all poles } \omega \text{ of } \widehat{H},\ j = 0, \dots, \operatorname{ord}_\omega a(z),$$

where $(-)^{(j)}$ denotes the j-th derivative.

- $\widehat{H}$ simple poles $\omega_1, \dots, \omega_r \neq 0$: $H({\omega_i}^{-1}) \overset{!}{=} \widehat{H}({\omega_i}^{-1})$ and $H'({\omega_i}^{-1}) \overset{!}{=} \widehat{H}'({\omega_i}^{-1})$

Walsh polynomial system

Corollary

Let $\widehat{H}(z) = b(z)/a(z)$ *have simple poles, then* $\widehat{H}$ *is a critical point if and only if there exists a polynomial* $g \in \mathbb{R}[z]_{\leq R-r-1}$ *such that*

$$a \cdot d - b \cdot c = \tilde{a}^2 \cdot g, \qquad \tilde{a}(z) := z^r a(1/z) \in \mathbb{R}[z]_{\leq r}.$$

- *Proof:* By Walsh's theorem,

$$F(z) := \frac{zb(z)}{a(z)} - \frac{zd(z)}{c(z)}$$

  has zeros of order ≥ 2 at ω^{-1} for roots ω of $a(z)$
- Equivalently, $\tilde{a}^2 \mid F$, meaning $F = \tilde{a}^2 G$, where G has no poles outside unit disc
- Clear denominators to obtain equations ($g = acG/z$) □

The Walsh variety

- From now on, make restriction to $\widehat{H}$ having simple poles
- Complexify and give everybody a name

$$a \in A := \mathbb{C}[z]_{\leq r}, \quad b \in B := \mathbb{C}[z]_{<r},$$
$$c \in C := \mathbb{C}[z]_{\leq R}, \quad d \in D := \mathbb{C}[z]_{<R}, \quad g \in G := \mathbb{C}[z]_{<R-r}$$

- Shorthand notation $A^\circ := A \setminus 0$, $A^\circ BCDG := A^\circ \times B \times C \times D \times G$

Definition (Walsh variety)

The *Walsh variety* is the variety

$$\mathcal{W} := \left\{ (a, b, c, d, g) \;\middle|\; ad - bc = \tilde{a}^2 g \right\} \subseteq A^\circ BCDG.$$

- $R + r$ equations in $3R + r + 2$ variables $\rightsquigarrow$ dimension $2R + 2$?

Irreducibility

Theorem (K.–Lagauw 2025+)

The Walsh variety $\mathcal{W} = \{ad - bc = \tilde{a}^2 g\}$ is a reduced and irreducible ideal-theoretic complete intersection in $A^\circ BCDG$ of dimension $2R + 2$.

- Main idea: Stratify $\mathcal{W} = \bigcup_{k=0}^{r-1} \mathcal{W}_k$, $\mathcal{W}_k = \{\, (a, b, \dots) \in \mathcal{W} \mid \deg\gcd(a, b) = k \,\}$
- Each $\mathcal{W}_k$ is an affine bundle of rank $(R - r + 1 + k)$ over

$$\mathcal{T}_k := \left\{\, (a, b, g) \;\middle|\; \deg\gcd(a, b) = k,\ \gcd(a, b) \mid \tilde{a}^2 g \,\right\} \subseteq A^\circ BG$$

- $\mathcal{T}_k$ is a subset of dimension $\begin{cases} = R + r + 1, & k = 0 \\ < R + r + 1 - k, & k \geq 1 \end{cases}$
- Deduce that $\mathcal{W}$ is a set-theoretic complete intersection by dimension count
- $\mathcal{W}_0$ is smooth, irreducible and dense in $\mathcal{W}$; finally apply unmixedness theorem $\square$

Finiteness of the critical locus

- Consider the projection $\tau \colon \underset{\dim 2R+2}{\mathcal{W}} \to \underset{\dim 2R+1}{CD}$
- $\tau^{-1}(c, d)$ is the solution set to the Walsh polynomial system

Corollary

If $(c, d) \in CD \cong \mathbb{C}^{2R+1}$ is general, then every component of $\tau^{-1}(c, d)$ is

1. *reduced, of dimension 1 (invariant under $\lambda \cdot (a, b, g) = (\lambda a, \lambda b, \lambda^{-1} g)$) and*
2. *consists of tuples (a, b, g) such that $a(z)$ has r distinct roots and $\gcd(a, b) = 1$.*

- Either τ is not dominant, then vacously true (not the case, will see later),
- or the general fiber is reduced of dimension $\dim \mathcal{W} - \dim CD = 1$
- Gives alternative, conceptual proof of finiteness of the critical locus

This MEP has no bad locus!

- Let $M(f)\colon \mathbb{C}[z]_{<R+r-\deg(f)} \to \mathbb{C}[z]_{<R+r}$ be the linear map $q \mapsto q \cdot f$
- Walsh polynomial system is affine-linear in b, g:

$$ad - bc = \tilde{a}^2 g \iff \begin{bmatrix} M(ad) & M(c) & M(\tilde{a}^2) \end{bmatrix} \cdot \begin{pmatrix} -1 \\ b \\ g \end{pmatrix} = 0$$

- To eliminate b, g, we need to ensure that $\begin{bmatrix} M(c) & M(\tilde{a}^2) \end{bmatrix}$ never drops rank

Lemma (Emptyness of the bad locus)

$$\dim_{\mathbb{C}} \operatorname{Ker} \begin{bmatrix} M(c) & M(\tilde{a}^2) \end{bmatrix} = \max\{0, \deg \gcd(c, \tilde{a}^2) - r\}.$$

The matrix has full rank if either c has distinct roots or both a and c are stable.

Walsh meets Porteous

$$M(a,c,d) := \begin{bmatrix} M(ad) & M(c) & M(\tilde{a}^2) \end{bmatrix} : \mathcal{O}(-1) \oplus \mathcal{O}^r \oplus \mathcal{O}(-2)^{R-r} \to \mathcal{O}^{R+r}$$

Theorem (K.–Lagauw 2025+, main result)

For given (c,d) with c distinct roots, the projection

$$\{\, (a,b,g) \mid (a,b,c,d,g) \in \mathcal{W} \,\} \to A^\circ$$

is an isomorphism onto the degeneracy locus of $M(a,c,d) \in \mathbb{C}[\underline{a}]^{(R+r)\times(R+1)}$.
If (c,d) is general, then the degeneracy locus $D_R(M(a,c,d)) \subseteq \mathbb{P}A$ is a finite set of points of degree

$$\sum_{j=0}^{r} \binom{R-r-1+j}{j} \cdot 2^j.$$

Degree of the MOR problem for various $R > r$

$R \setminus r$	1	2	3	4	5	6	7	8
2	3							
3	5	7						
4	7	17	15					
5	9	31	49	31				
6	11	49	111	129	63			
7	13	71	209	351	321	127		
8	15	97	351	769	1 023	769	255	
9	17	127	545	1 471	2 561	2 815	1 793	511

- In example $(R, r) = (6, 2)$, the authors report 49 complex solutions, 11 real, 5 real stable.

Towards multiple poles

- Previous results also apply for specific c with distinct roots and general d
- If c has exactly one double root but the solution set of the Walsh system is still finite, then the number of solutions drops to

$$\#\text{solutions} \leq \sum_{j=0}^{r} \binom{R-r-1+j}{j} 2^j - \binom{R-2}{r-1}$$

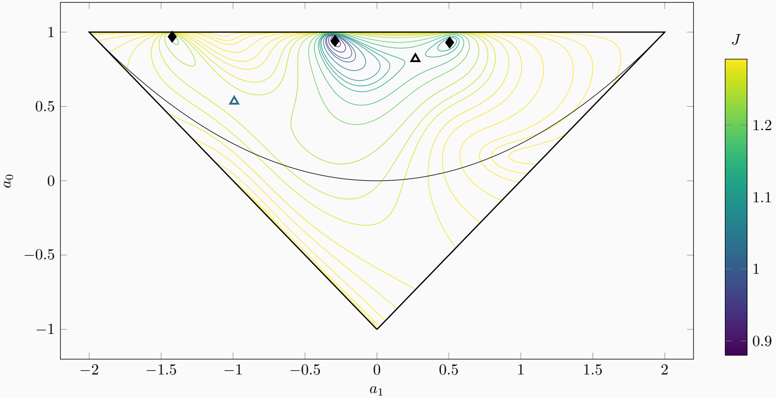
- This bound is sometimes sharp (e.g. $R=3, r=1$)
- But sometimes not; the points at infinity have multiplicities (e.g. $R=3, r=2$)

To Infinity And Beyond

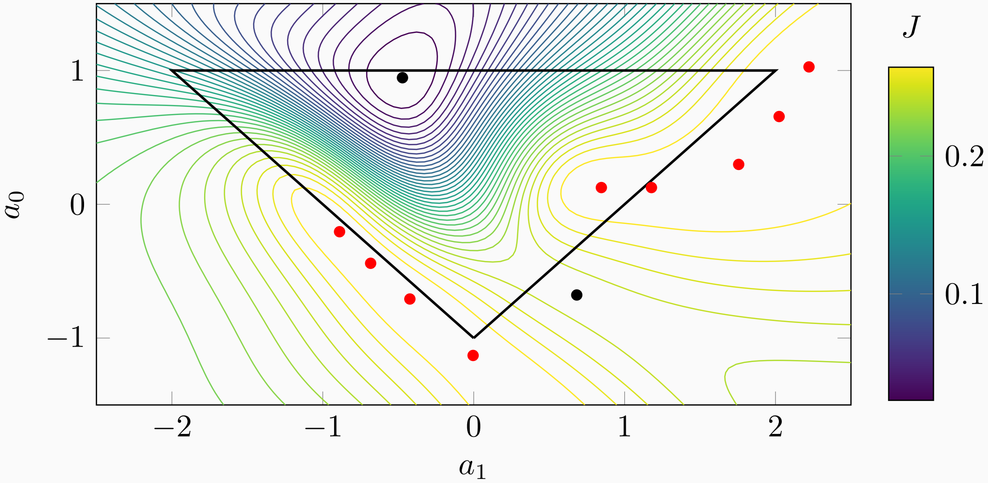
Example: Model order reduction from order 6 to 2

$$Y(z) = \frac{0.0448z^5 + 0.2368z^4 + 0.0013z^3 + 0.0211z^2 + 0.2250z + 0.0219}{z^6 - 1.2024z^5 + 2.3675z^4 - 2.0039z^3 + 2.2337z^2 - 1.0420z + 0.8513}$$

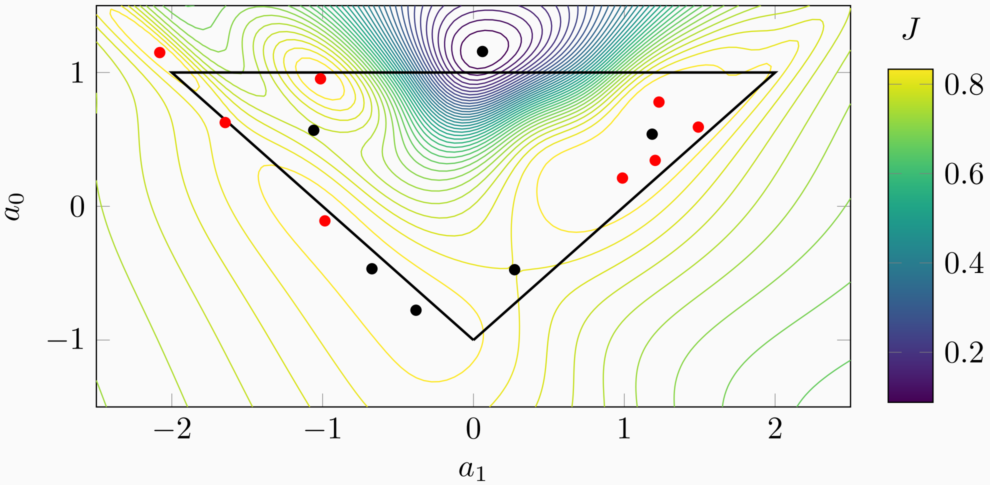
$$y = (0.0448, 0.2907, 0.2447, 0.2830, 0.2123, 0.3245, 0.0438, 0.5013, 0.4671, \dots)$$



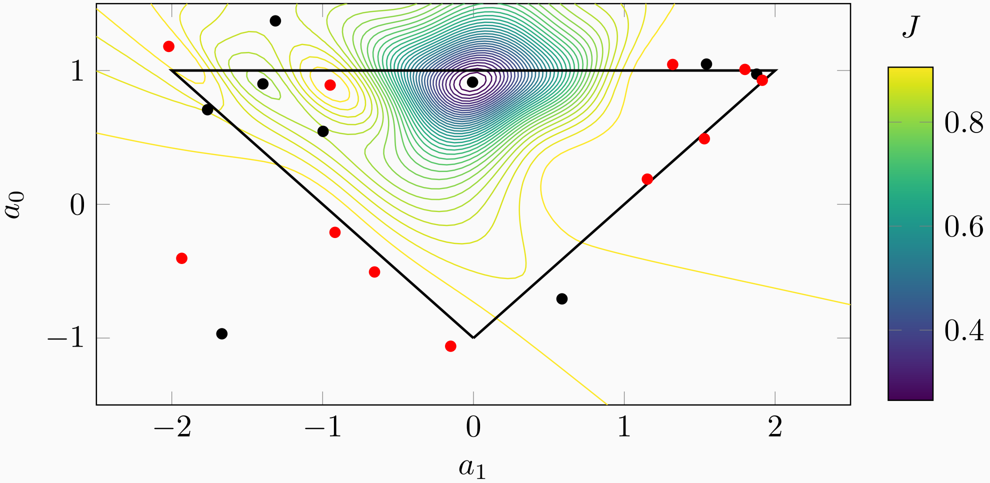
ED minimization of $(y_0, \ldots, y_4)$ to $\mathcal{X}_{4,2}$ $(N = 5)$



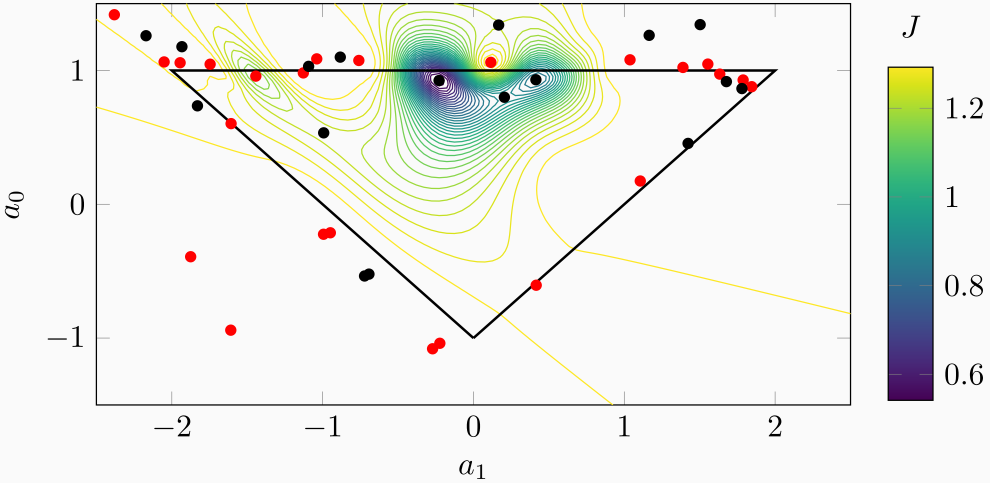
ED minimization of $(y_0, \ldots, y_9)$ to $\mathcal{X}_{9,2}$ ($N = 10$)



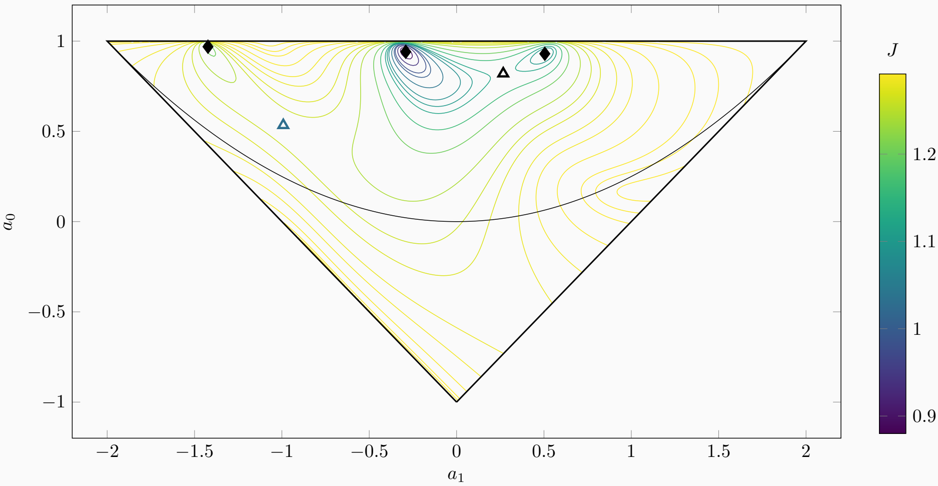
ED minimization of $(y_0, \dots, y_{14})$ to $\mathcal{X}_{14,2}$ $(N = 15)$



ED minimization of $(y_0, \ldots, y_{29})$ to $\mathcal{X}_{29,2}$ ($N = 30$)



ℓ_2-distance minimization of y to $\mathcal{X}_{\infty,2}$



Many future questions

- Why is the standard norm on $\mathbb{R}^N$ EDD-general?
- Prove more values $\mathrm{EDD}_\Theta(X_{N-1,r})$ for the Frobenius norm
- What is the bidegree of the critical incidence?
- Why is the standard norm the "correct one" and not the Bombieri norm?
- Can you formalize that critical points for finite N converge to critical points of the MOR problem as $N \to \infty$? What happens to the unstable critial points?
- Are the MEP descriptions of the two problems related?

Thank you! Questions?

Thank you! Questions?

References i

J. Cadzow.
Singal enhancement: a composite property mapping algorithm.
IEEE Transactions on Acoustics, Speech, and Signal Processing, 36(2):49–62, 1988.

Gaspard de Prony.
Essai experimental et analytique sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'alkool à différentes températures.
Journal de l'École Polytechnique de Paris, 1:24–76, 1795.

References ii

P. Lemmerling, L. Vanhamme, S. Van Huffel, and B. De Moor.
IQML-like algorithms for solving structured total least squares problems: a unified view.
Signal Processing, 81:1935–1945, 2001.

G. Ottaviani, P.-J. Spaenlehauer, and B. Sturmfels.
Exact solutions in structured low-rank approximation.
SIAM Journal on Matrix Analysis and Applications, 35(4):1521–1542, 2014.

Image credit

- Slide 3: "With permission" from Sibren's lecture on systems theory
- Slide 8: Created by Sibren using MATLAB and `pgfplot`