Can a mathematician and an engineer be friends?

Critical points to low-rank Hankel matrix approximation

Leonie Kayser & Sibren Lagauw

leokayser.github.io

May 8, 2025

MAX PLANCK INSTITUTE
FOR MATHEMATICS
IN THE SCIENCES

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

Realization of linear time-invariant difference equations

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

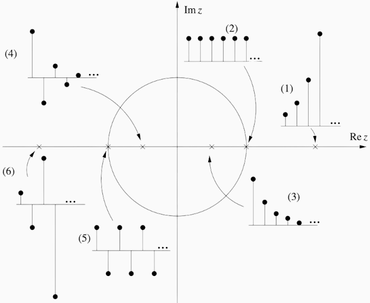
- Discrete-time (physical) system generating signals $y = (y_0, y_1, \ldots, y_{N-1})^\top \in \mathbb{R}^N$
- Explain observed data with a mathematical model
- Impose a model class: *autonomous LTI models of finite order*
  - autonomous = no input signals, no influence from outside world
  - linear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

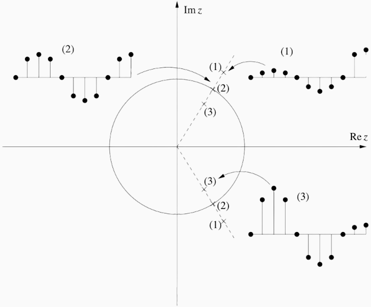
Roots of $a(z) = \sum_{i=0}^{r} a_i z^i$ determine dynamics of model

- 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

$$X_r := \left\{ \hat{y} \in \mathbb{C}^N \;\middle|\; \operatorname{rank} H^{\hat{y}}_r \leq r \right\}$$

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

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

Least squares realization

- Fix a 2-norm on $\mathbb{R}^N$, $Q(y) = \frac{1}{2}\|y\|^2 = \frac{1}{2}y^\mathsf{T}\Lambda y$
- 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

- If ε is Gaussian white noise, then closest $\hat{y}$ is maximum likelihood estimator

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

- Constraint optimization problem: Impose rank condition on $\hat{y}$

$$\underset{\hat{y}\in\mathbb{R}^N}{\text{minimize}} \;\; Q(y - \hat{y}) \qquad \text{subject to} \qquad \operatorname{rank} H_r^{\hat{y}} \leq r$$

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

Heuristic approaches

- First idea goes back to Prony [PGDB95]
- Cadzow's method [Cad88] (assume standard norm on $\mathbb{R}^N$)
  1. Compute SVD of $H_r^y = U\Sigma V^\mathsf{T}$, 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!

- $\mathbb{P}(X_r)$ is not smooth, $\{\,(y,a) \in \mathbb{P}^{N-1} \times \mathbb{P}^r \mid H_r^y \cdot a = 0\,\}$ is desingularization

$\rightsquigarrow$ 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:

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

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

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

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

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}

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

- 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

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

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

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

- 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!

AY is a determinantal variety

Lemma

Let M be a "tall" $m \times (n+1)$-matrix with polynomial entries over a variety X and

$$\mathcal{K} = \{\, (x, [v]) \mid M(x) \cdot v = 0 \,\} \subseteq X \times \mathbb{P}^n.$$

Let Z be the projection of $\mathcal{K}$ onto X. If $\mathcal{K}$ is reduced and for all $x \in X$ one has $\operatorname{rank} M(x) \in \{n, n+1\}$, then the ideal of Z is given by the $(n+1)$-minors of M.

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

Corollary

1. *The prime ideal of AY is locally given by the $(N - 2r + 1)$-minors of $M(a, y)$.*
2. *Restricting to a general $y \in \mathbb{P}^{N-1}$, the system of minors of $M(a, y)$ defines a finite set of reduced points in $\mathbb{P}(\mathcal{G})$, and hence in $\mathbb{P}^r$ (assuming $\mathbb{P}(\mathcal{B})$ finite).*

Intersection theory saves the day

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

- 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

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

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.*

Some special weights

$N \diagdown r$	1	2	3
3	4		
4	7		
5	10	13	
6	13	34	
7	16	64	40
8	19	103	142

$N \diagdown r$	1	2	3
3	2		
4	3		
5	4	7	
6	5	16	
7	6	28	20
8	7	43	134

- EDD's for standard norm (left) and Bombieri–Weyl norm (right)
- Bombieri–Weyl & $(N, r) = (8, 2)$ is the first case where the inequality is strict
- Efficient implementation in `Macaulay2`

The weight discriminant

- The *weight discriminant* is the set of norms giving subgeneric ED degree

$$\nabla_{\mathsf{EDw}}(X_r) = \overline{\{\, \Lambda \mid \operatorname{rank}\Lambda = N,\ \mathrm{EDD}_\Lambda(X_r) < \mathrm{EDD}_{\mathrm{gen}}(X_r)\,\}} \subseteq \mathbb{P}(\mathrm{Sym}(N))$$

- This is an irreducible variety, expected to be a hypersurface
- Using diagonal weights Λ in the definition gives $\nabla_{\mathsf{EDw,diag}}(X_r) \subseteq \mathbb{P}(\mathrm{Diag}(N))$

Theorem (The case $r = 1$)

For $r = 1, N \geq 3$ the discriminants are irreducible hypersurfaces

$$\deg \nabla_{\mathsf{EDw}}(X_1) = 4N - 6, \qquad \deg \nabla_{\mathsf{EDw,diag}}(X_1) = 2N - 4.$$

$\nabla_{\mathsf{EDw,diag}}$ *is the discriminant of a degree $N-1$ polynomial, the two are related by*

$$\Delta_{\mathsf{EDw}}(X_1)|_{\mathrm{Diag}(N)} = \lambda_0 \cdot \lambda_{N-1} \cdot \Delta_{\mathsf{EDw,diag}}(X_1)^2.$$

The ED discriminant

- Fixing Λ, our computation still relied on genericity of y
- The ED discriminant consists of $y \in \mathbb{C}^N$ such that the system has a multiple solution a.

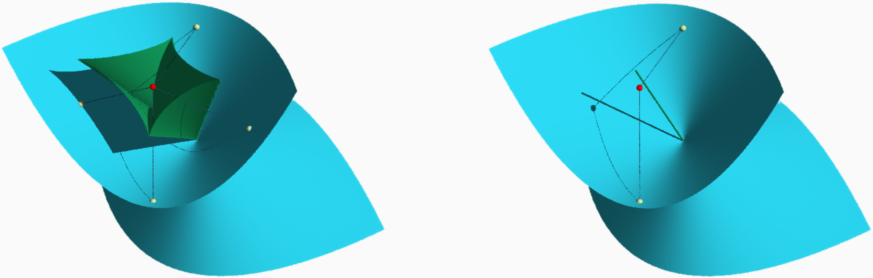


Figure 1: General and special (Bombieri–Weyl) weights

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.

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.

References ii

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.

C. Prony, F. M. Gaspard, and R. De Baron.
Essai experimental et analytique, ect.
Journal de l'Ecole Polytechnique de Paris, 1:24–76, 1795.

Image credit

- GeoGebra
- Slide 3: “With permission” from Sibren’s lecture on systems theory
- Slide 17: Thanks to Luca Sodomaco for letting me use his graphics!