The geometry behind eigenvalue methods for symmetric tensor decomposition

NoGAGS Spring 2024

Leonie Kayser (feat. Fulvio Gesmundo & Simon Telen)
leokayser.github.io

March 1, 2024

MAX PLANCK INSTITUTE
FOR MATHEMATICS
IN THE SCIENCES

The geometry behind eigenvalue methods for symmetric tensor decomposition

(symmetric) tensor decomposition

What is a tensor?

A tensor...

- ... is an object that transforms like a tensor
- ... is an element of a tensor product of vector spaces $U \otimes V \otimes W$
- ... is a multidimensional array of numbers $A = (A_{ijk})_{i,j,k}$
- ... in $V^{\otimes d}$ is symmetric if its entries are invariant under permutations $\sigma \in \mathfrak{S}_d$
- Symmetric tensors can be identified with homogeneous polynomials (in char. 0)

$$\mathrm{S}^d V \ni v_1 \cdots v_d \quad \mapsto \quad \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)} \in \mathrm{Sym}^d V \subseteq V^{\otimes d}$$

Tensor decomposition and rank

- A tensor of the form $(u_i v_j w_k)_{i,j,k} \hat{=} u \otimes v \otimes w$ is simple
- Every tensor is a sum of simple tensors

$$A = \sum_{i=1}^{r} \lambda_i u^{(i)} \otimes v^{(i)} \otimes w^{(i)}$$

- The smallest such r is the tensor rank of A
- Generalizes matrix rank: $A = S \cdot \mathrm{diag}(\underbrace{1, \ldots, 1}_{\mathrm{rank}\, A}, 0, \ldots, 0) \cdot T$
- If the simple tensors are unique up to scaling, then A is called identifiable
- Symmetric case: Simple tensor $v^{\otimes d} \hat{=} \ell^d$, $F = \sum_{i=1}^{r} \lambda_i \ell_i^d$, symmetric tensor rank, ...

Examples

We will identify symmetric tensors with homogeneous polynomials in $T = \mathbb{C}[X_0, \dots, X_n]$.

- Rank 1 = powers of linear forms ℓ^d = (cone over) Veronese variety

$$V_{d,n} = \nu_d(\mathbb{P}(T_1)) \subseteq \mathbb{P}(T_d), \qquad \nu_d([\ell]) = [\ell^d]$$

- Quadratic forms = sym. matrices: $F = x^{\mathsf{T}} A x$, then $\operatorname{rk} F = \operatorname{rank} A$
- Fun exercise: $\operatorname{rk}(X_1^d + \cdots + X_n^d) = n$
- $\operatorname{rk}(X_0 X_1) = 2$, as $X_0 X_1 = \frac{1}{4}(X_0 + X_1)^2 - \frac{1}{4}(X_0 - X_1)^2$
- $\operatorname{rk}(X_0 X_1^{d-1}) = d$, more generally for $\alpha_0 \leq \alpha_1 \leq \dots$

$$\operatorname{rk}(X_0^{\alpha_0} \cdots X_n^{\alpha_n}) = (\alpha_1 + 1) \cdots (\alpha_n + 1)$$

- But $d X_0 X_1^{d-1} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(\varepsilon X_0 + X_1)^d - \frac{1}{\varepsilon} X_1^d$, so rk is *not* lower semi-continuous

A general form walks into the door

Theorem (Alexander–Hirschowitz)

For $r(n+1) \leq \binom{n+d}{d}$ the affine cone

$$\widehat{\sigma_r V_{d,n}} = \overline{\{ F \in T_d \mid \operatorname{rk}(F) \leq k \}} \subseteq T_d = \mathbb{A}^{\binom{n+d}{n}}$$

has the expected dimension $r(n+1)$ except for

$$(d, n, r) = (2, \geq 2, \geq 2), (3, 4, 7), (4, 2, 5), (4, 3, 9), (4, 4, 14).$$

In particular, a general polynomial has rank $\left\lceil \frac{1}{n+1}\binom{n+d}{n} \right\rceil$.

Running example:

A general $F \in \mathbb{C}[X_0, X_1, X_2]_{10}$ has $\operatorname{rk} F = \frac{1}{3}\binom{2+10}{2} = 22$. The set of such forms of rank 18 has dimension 54 in $\mathbb{A}^{66}$

General forms of subgeneric rank are identifiable

Theorem (Ballico, Mella, Chiantini–Ottaviani–Vannieuwenhoven, . . .)

For $r(n+1) < \binom{n+d}{d}$ *a general form of rank* r *is identifiable except in the cases*

$$(d,n,r) = (2, \geq 2, \geq 2), (6,2,9), (4,3,8), (3,5,9).$$

- For applications tensors are often of subgeneric rank $\rightsquigarrow$ generic identifiability
- A general $F \in \mathbb{C}[X_0, X_1, X_2]_{10}$ of rank 18 has an essentially *unique* representation

$$F = \sum_{i=1}^{18} \lambda_i \ell_i^{10}, \qquad \ell_i \in \mathbb{C}[X_0, X_1, X_2]_1$$

- Given F, how do we find the ℓ_i algorithmically?

Apolarity and eigenvalue methods

The fundamental theorem of tensor decomposition

- There is a the natural apolar action $\mathrm{S}^d V^* \times \mathrm{Sym}^D V \to \mathrm{Sym}^{D-d} V$
- Let $S = \mathrm{S}^d V^* = \mathbb{C}[\partial_0, \dots, \partial_n]$ then S acts on $T = \mathbb{C}[X_0, \dots, X_n]$ by differentiation

$$\partial^\alpha \bullet x^\beta = \frac{\beta!}{(\beta-\alpha)!} x^{\beta-\alpha} \text{ if } \beta \geq \alpha, \text{ else } 0$$

- In this way S is the homogeneous coordinate ring of $\mathbb{P}(T_1)$: $g([\ell]) = g \bullet \ell^{\deg g}$
- For $F \in T$ let $F^\perp = \mathrm{Ann}_S(F) = \{\, g \in S \mid g \bullet F = 0 \,\}$

Theorem (Apolarity lemma)

For $F \in T_D$ and $\ell_1, \dots, \ell_r \in T_1$ the following are equivalent:

1. $F = \lambda_1 \ell_1^D + \cdots + \lambda_r \ell_r^D$ *for some* $\lambda_i \in \mathbb{C}$;
2. $I([\ell_1], \dots, [\ell_r]) \subseteq F^\perp$ *in* S.

The Catalecticant method

- If $F = \sum_{i=1}^{r} \lambda_r \ell_1^D + \cdots + \lambda_r \ell_r^D$, then $F^\perp$ *contains* equations vanishing on $[\ell_i]$
- For $d \leq \frac{D}{2}$, $r < \binom{d+n}{n} - n$ and $F \in T_D$ general of rank r, then actually

$$(F^\perp)_d = I([\ell_1], \dots, [\ell_r])_d$$

- By definition $(F^\perp)_d = \operatorname{Ker} \operatorname{Cat}_{d,D-d}(F)$ where

$$\operatorname{Cat}_{d,D-d}(F) \colon S_d \to T_{D-d}, \qquad g \mapsto g \bullet F$$

- Algorithmic approach:
  - Compute basis $\mathcal{F}$ of kernel
  - Solve polynomial system $\{\mathcal{F} = 0\}$ to get $Z = \{[\ell_1], \dots, [\ell_r]\}$,
  - Solve *linear* equations to get λ_i

⇝ When is $\mathrm{V}(F_d^\perp) = Z$? Equivalently $\mathrm{V}(I(Z)_d) = Z$?

Eigenvalue methods for polynomial system solving

Task: Given 0-dim'l system $J \subseteq S$, compute $Z = \{z_1, \ldots, z_r\} = \mathrm{V}(J) \subseteq \mathbb{P}^n$

- ▷ For t large enough, $h_{S/J}(t) := \dim_{\mathbb{C}}(S/J)_t = r$ and $J_t = I(Z)_t$
- ▷ Multiplication map for $g \in S_e$:

$$M_g \colon (S/J)_d \xrightarrow{\cdot g} (S/J)_{d+e}$$

- ▷ Under "suitable conditions" $M_h^{-1} M_g \colon (S/J)_d \to (S/J)_d$ has left eigenpairs

$$\{\, (\mathrm{ev}_{z_i}, \tfrac{g}{h}(z_i)) \mid i = 1, \ldots, r \,\}, \qquad \mathrm{ev}_{z_i}(f) = f(z_i)/h(z_i)$$

⇝ Translate problem into large eigenvalue problem, solve numerically

- ▷ For this need $h_{S/J}(d+e) = h_{S/J}(d) = r$, want $d, d+e$ as small as possible

Example: J saturated

If $J = I(Z)$ and Z is a general set of points, then $h_{S/I(Z)} = \min\{h_S(t), r\}$.
Hence $d = \min\{\, t \mid h_S(t) \geq r \,\}$ and $e = 1$ work.

Recap

We are lead to the following setup:

- Given a general form $F = \sum_{i=1}^{r} \lambda_i \ell_i^r \in \mathbb{C}[X_1, \ldots, X_n]_D$ of rank $r < \binom{n+\lfloor D/2 \rfloor}{n} - n$
- Decomposition is unique, want to find $Z = \{[\ell_1], \ldots, [\ell_r]\} \in \mathbb{P}^n$
- Have access to $\mathcal{F} = I(Z)_d$ only for $d \leq \frac{D}{2}$
- Want to solve polynomial system $\mathcal{F}$ using the eigenvalue method
- Is $\mathrm{V}(\mathcal{F}) = Z$ (scheme-theoretically)?
- What is the Hilbert function of the subideal $\langle \mathcal{F} \rangle_S \subseteq I(Z)$? When $= r$?

Running example

$n = 2$, $D = 10$, $r = 18$. $F = \sum_{i=1}^{18} \lambda_i \ell_i^{10} \in \mathbb{C}[X_0, X_1, X_2]_{10}$.
Only interesting: $d = D/2 = 5$, since for $d \leq 4$ we have $I(Z)_d = 0$!

Some nice geometry behind this!

Mathematics > Commutative Algebra

[Submitted on 5 Jul 2023]

Hilbert Functions of Chopped Ideals

Fulvio Gesmundo, Leonie Kayser, Simon Telen

A chopped ideal is obtained from a homogeneous ideal by considering only the generators of a fixed degree. We investigate cases in which the chopped ideal defines the same finite set of points as the original one-dimensional ideal. The complexity of computing these points from the chopped ideal is governed by the Hilbert function and regularity. We conjecture values for these invariants and prove them in many cases. We show that our conjecture is of practical relevance for symmetric tensor

Rediscovering a notion introduced by [Ahmed–Fröberg–Rafiq]

Definition (Chopped ideal)

The *chopped ideal* of a homogeneous ideal $I \subseteq S$ in degree d is $I_{\langle d \rangle} := \langle I_d \rangle_S$.

From now on $Z \subseteq \mathbb{P}^n$ is a general set of r points, $I = I(Z)$, $d = \min\{\, t \mid h_S(t) \geq r \,\}$.

- The minimal generators of I live in degrees $\{d, d+1\}$
- Can we recover Z from $I(Z)_{\langle d \rangle}$?
- When does $(I(Z)_{\langle d \rangle})_{d+e} = I(Z)_{d+e}$?
- What is the Hilbert function $h_{I(Z)_{\langle d \rangle}}(t)$?

Example: Z = 18 points in the plane

t	$\dots$	3	4	5	6	7
$h_S(t)$	$\dots$	10	15	21	28	36
$h_I(t)$	$\dots$	0	0	3	10	18
$h_{I_{\langle 5\rangle}}(t)$	$\dots$	0	0	3	9	18

t	0	1	2	3	4	5	6	7
$h_S(t)$	1	3	6	10	15	21	28	36
$h_{S/I}(t)$	1	3	6	10	15	18	18	18
$h_{S/I_{\langle 5\rangle}}(t)$	1	3	6	10	15	18	19	18

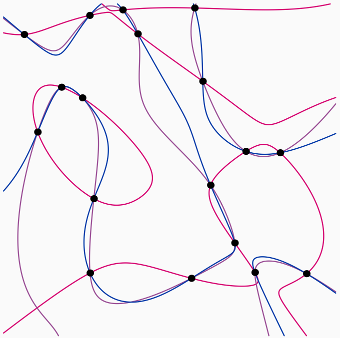
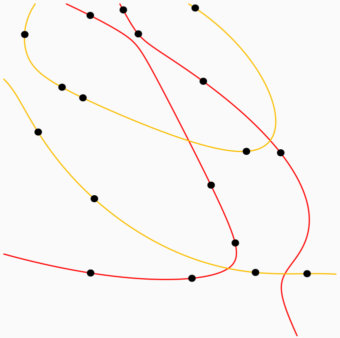


Figure 1: Three quintics $\langle q_1, q_2, q_3\rangle_{\mathbb{C}} = I_5$ passing through 18 general points (left) and the missing split sextic $cc' \in I_6$ (right).



Recovering the points from their chopped ideal

- Generally $I_{\langle d \rangle} \subsetneq I$, but maybe

$$I \overset{?}{=} (I_{\langle d \rangle})^{\mathrm{sat}} := \bigcup_{k \geq 0} (I_{\langle d \rangle} : \mathfrak{m}^k) \qquad \Longleftrightarrow \qquad \mathrm{V}(I) \underset{\text{schemes}}{\overset{?}{=}} \mathrm{V}(I_{\langle d \rangle}) \subseteq \mathbb{P}^n$$

Theorem

Let $Z \subseteq \mathbb{P}^n$ be a general set of r points and $d \in \mathbb{N}$.

1. *If $r > \binom{n+d}{n} - n$, then $\mathrm{V}(I_{\langle d \rangle})$ is a positive-dimensional complete intersection.*
2. *If $r = \binom{n+d}{n} - n$, then $\mathrm{V}(I_{\langle d \rangle})$ is a complete intersection of d^n points.*
3. *If $r < \binom{n+d}{n} - n$, then $I_{\langle d \rangle}$ cuts out Z scheme-theoretically.*

In particular, $I = (I_{\langle d \rangle})^{\mathrm{sat}}$ if and only if $r < \binom{n+d}{n} - n$ or $r = 1$ or $(n, r) = (2, 4)$.

Towards the expected Hilbert function – naively

- Graded components of $I_{\langle d\rangle}$ are images of multiplication map

$$\mu_e \colon S_e \otimes_{\mathbb{C}} I_d \to I_{d+e}, \qquad g \otimes f \mapsto g \cdot f$$

- One may expect μ_e to have *maximal rank*, i.e. to be injective or surjective:

$$h_{I_{\langle d\rangle}}(t) \overset{?}{=} \min\{h_I(t),\, h_S(t-d) \cdot h_I(d)\}$$

⇝ $e = 1$: Ideal generation conjecture (IGC) predicting number of minimal generators of I

- This turns out to be too optimistic; μ_e has elements in its kernel, for example

$$f_1 \otimes f_2 - f_2 \otimes f_1 \in \operatorname{Ker} \mu_d, \qquad f_1, f_2 \in I_d$$

- This *does* happen, e.g. $r = 52$ points in $\mathbb{P}^3$, then μ_5 does not have maximal rank

Towards the expected Hilbert function – more carefully

- The kernel of μ_e contains the Koszul syzygies Ksz_e generated by

$$g f_i \otimes f_j - g f_j \otimes f_i, \qquad g \in S_{e-d},\ f_i, f_j \in I_d$$

- Expecting $\operatorname{Ker} \mu_e = \mathrm{Ksz}_e$, a first estimate of $\dim_{\mathbb{C}} \operatorname{Ker} \mu_e$ is $h_S(e-d) \cdot \binom{h_I(d)}{2}$
- Expect the syzygies to also only have Koszul syzygies, correct by $h_S(e-2d) \cdot \binom{h_I(d)}{3}$
- And these also only have Koszul syzygies and ...
- This leads to the following estimate for $h_{S/I_{\langle d \rangle}}(t)$:

$$h_S(t) - \underbrace{h_S(t-d)h_I(d)}_{\text{gen's of } I_d} + \underbrace{h_S(t-2d)\binom{h_I(d)}{2}}_{\text{Koszul syzygies}} - \underbrace{h_S(t-3d)\binom{h_I(d)}{3}}_{\text{Koszul syzygy syzygies}} \pm \ldots$$

- On the other hand, as soon as $h_{I_{\langle d \rangle}}(t_0) \geq h_I(t_0)$, then $I_t = (I_{\langle d \rangle})_t$ for $t \geq t_0$

The main conjecture

Expected syzygy conjecture (ESC)

$$h_{S/I_{\langle d\rangle}}(t) = \begin{cases} \sum_{k\geq 0} (-1)^k \cdot h_S(t-kd) \cdot \binom{h_I(d)}{k} & t < t_0, \\ r & t \geq t_0, \end{cases}$$

where t_0 is the least integer $> d$ such that the sum is at most r.

- This is always a (lexicographic) lower bound due to Fröberg (and recently Nenashev)
- Alternative expression for the ideal:

$$h_{I_{\langle d\rangle}}(t) = \begin{cases} \sum_{k\geq 1} (-1)^{k-1} \cdot h_S(t-kd) \cdot \binom{h_I(d)}{k} & t < t_0, \\ h_I(t) & t \geq t_0, \end{cases}$$

Is the complicated alternating sum really needed?

- For $\mathbb{P}^2$ the (ESC) "actually" says $h_{I_{\langle d \rangle}}(t) = \min\{h_I(d) \cdot h_S(t-d),\, h_I(t)\}$
- This is no longer true in higher dimension – in general n summands are required
- Smallest example: 52 points in $\mathbb{P}^3$

$$h_{S/I_{\langle 5 \rangle}}(t) = \begin{cases} h_S(t) - 4h_S(t-5) + 6h_S(t-10) & t < 11, \\ 52 & t \geq 11 \end{cases}$$

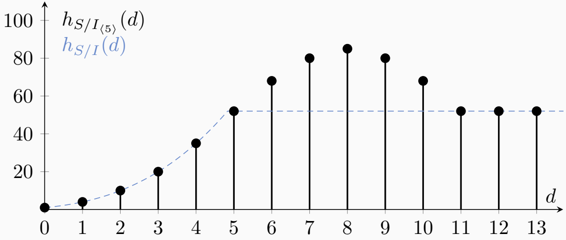


Figure 2: The Hilbert function of the chopped ideal of 52 general points in $\mathbb{P}^3$.

Main results

Theorem

Conjecture (ESC) is true in the following cases:

- $r_{\max} := h_S(d) - (n+1)$ *for all* d *in all dimensions* n.
- *In the plane for* $r_{\min} = \frac{1}{2}(d+1)^2$ *when* d *is odd.*
- $r \leq \frac{1}{n}\big((n+1)h_S(d) - h_S(d+1)\big)$ *and [*$n \leq 4$ *or generally whenever (IGC) holds].*
- *In a large number of individual cases in low dimension (next slide).*

The length of the saturation gap is bounded above by

$$\min\{\, e > 0 \mid (I_{\langle d \rangle})_{d+e} = I_{d+e} \,\} \leq (n-1)d - (n+1).$$

Whenever $I_{\langle d \rangle}$ *is non-saturated, one has* $\operatorname{reg}_{\mathrm{CM}} S/I_{\langle d \rangle} = \operatorname{reg}_{\mathrm{H}} S/I_{\langle d \rangle} - 1 = d + e - 1$.

Verification using computer algebra

- Testing the conjecture for particular values of (n, r):
  - Sample r random points from $\mathbb{P}^n(\mathbb{Q})$
  - Calculate $h_{S/I(Z)_{\langle d \rangle}}$ using a computer algebra system
  - If the sample satisfies (ESC), then the conjecture is true for general such Z

Theorem

The map $Z \mapsto h_{S/I(Z)_{\langle d \rangle}}(t)$ is upper semicontinuous on the set $U \subseteq (\mathbb{P}^n)^r$ of points with generic Hilbert function.

- To speed up computation, perform calculations over a finite field $\mathbb{F}_p$
- Using `Macaulay2` we verified the conjecture in the following cases

n	2	3	4	5	6	7	8	9	10
r	≤ 1825	≤ 1534	≤ 991	≤ 600	≤ 447	≤ 316	≤ 333	≤ 204	≤ 259
d	≤ 58	≤ 18	≤ 9	≤ 6	≤ 4	≤ 3	≤ 3	≤ 2	≤ 2

Visualization of the saturation gaps in $\mathbb{P}^2$

- ESC predicts exactly how large the difference between I and $I_{\langle d \rangle}$ is

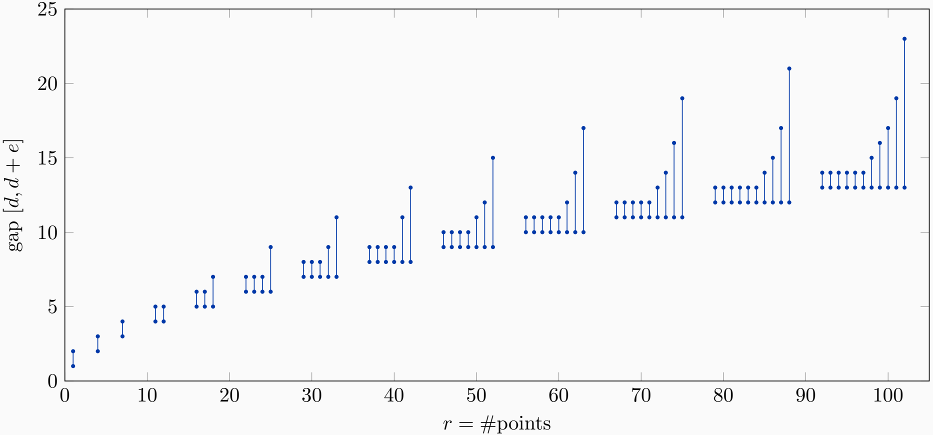


Figure 3: The saturation gaps for all values of $r \leq 102$ in $\mathbb{P}^2$.

Visualization of the saturation gaps in $\mathbb{P}^3$

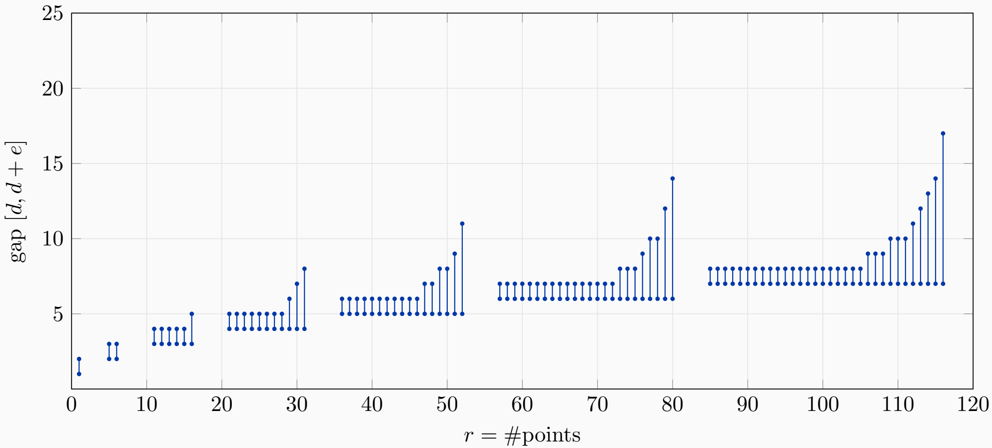


Figure 4: The saturation gaps for all values of $r \leq 116$ in $\mathbb{P}^3$.

Outlook

- Characteristic $p > 0$?

  Result expected to carry over, but char. 0 methods used (generic smoothness)

- Proving the conjecture in $\mathbb{P}^2$?

  Monomial degenerations seem to *not* work (despite resolving (MRC) & Fröberg)

- Generalizations multi-graded setting, e.g. points in $\mathbb{P}^n \times \mathbb{P}^m$ (original motivation)

  $\rightsquigarrow$ non-symmetric tensor decomposition problems

- State a conjecture for the minimal free resolution of $I(Z)_{\langle d \rangle}$

Thank you! Questions?

arXiv:2307.02560

References i

Chwas Ahmed, Ralf Fröberg, and Mohammed Rafiq Namiq.
The graded betti numbers of truncation of ideals in polynomial rings, 2022.

A. Bernardi, E. Carlini, M. V. Catalisano, A. Gimigliano, and A. Oneto.
The hitchhiker guide to: Secant varieties and tensor decomposition.
Mathematics, 6(12):314, 2018.

Ralf Fröberg.
An inequality for hilbert series of graded algebras.
Mathematica Scandinavica, 56(2), 1985.

Fulvio Gesmundo, Leonie Kayser, and Simon Telen.
Hilbert functions of chopped ideals, 2023.

Anna Lorenzini.
The minimal resolution conjecture.
Journal of Algebra, 156(1), 1993.

Simon Telen.
Solving Systems of Polynomial Equations.
PhD thesis, KU Leuven, Leuven, Belgium, 2020.

Simon Telen and Nick Vannieuwenhoven.
A normal form algorithm for tensor rank decomposition.
ACM Trans. on Math. Soft., 48(4):1–35, 2022.

Image credit

- Slide 10: “Hacken Holz” by vitranc on iStock
  `https://www.istockphoto.com/de/foto/hacken-holz-gm504268819-44840794`
- Slide 11: Created using Asymptote
  `https://asymptote.sourceforge.io/`