Exact Algebraic Computation of Learning Coefficients for Two-Dimensional Singular Statistical Models

Grégoire Sergeant-Perthuis [1] Elias Tsigaridas [12] Jules Tsukahara [12]

[1]Sorbonne Université

[2]INRIA

International Conference on Symbolic Computation and Machine Learning, 2026

Table of Contents

In This Talk

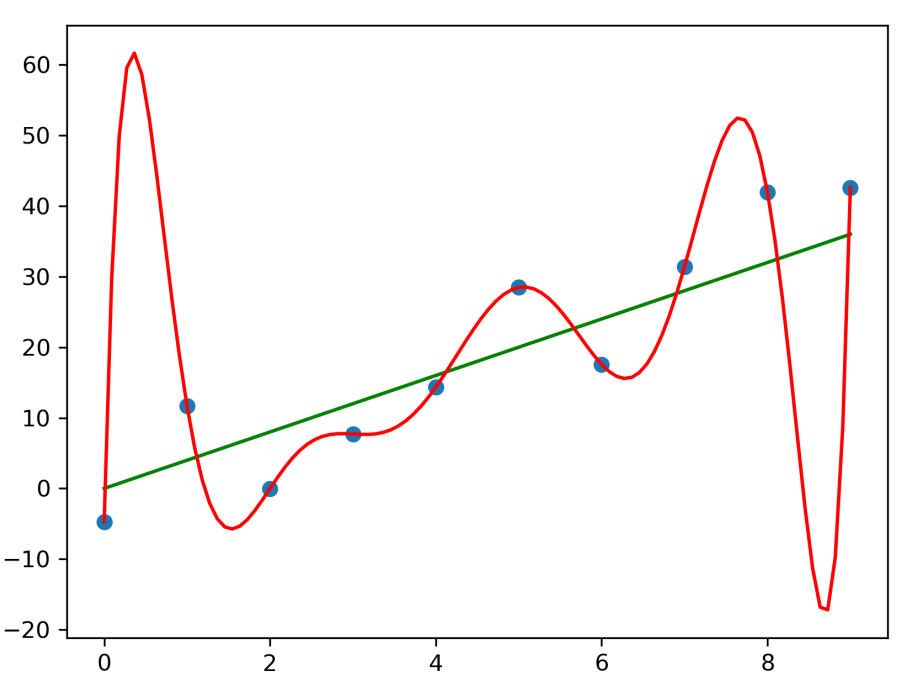
$$\mathrm{SC} \rightarrow \mathrm{ML}$$

In This Talk

$$\mathrm{SC} \longrightarrow \mathrm{ML}$$

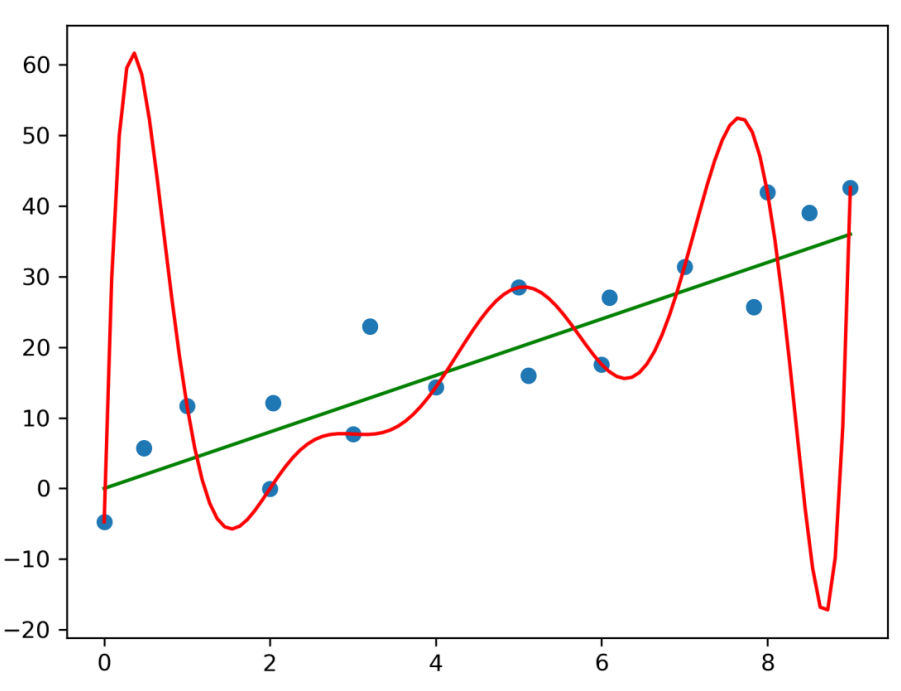
We give an application of Symbolic Computation to the analysis of Machine Learning models.

The Generalization Paradox



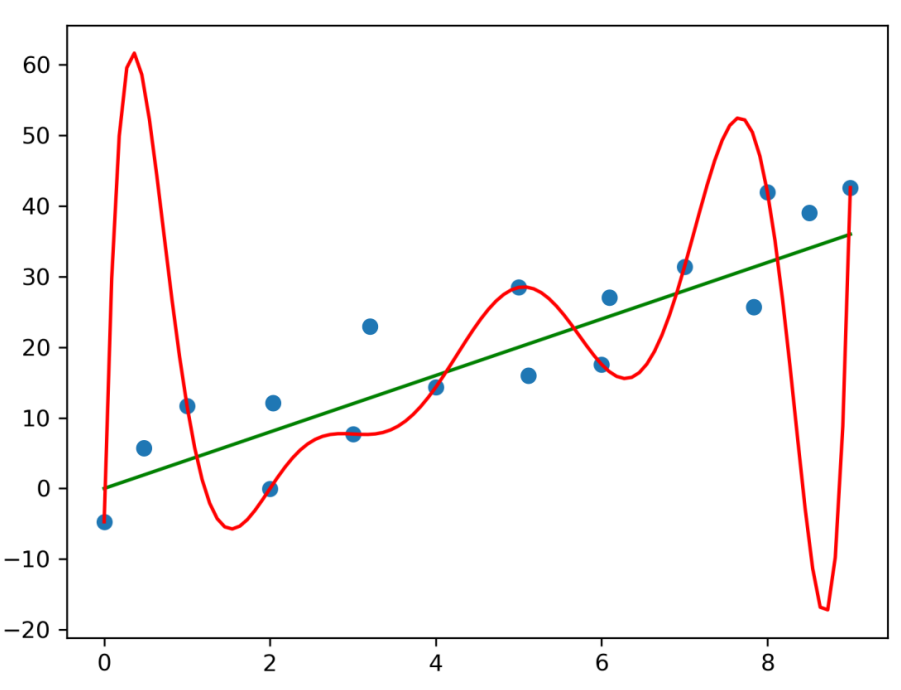
Model selection: choosing which model will best fit the data.

The Generalization Paradox



Model selection: choosing which model will best fit the data.

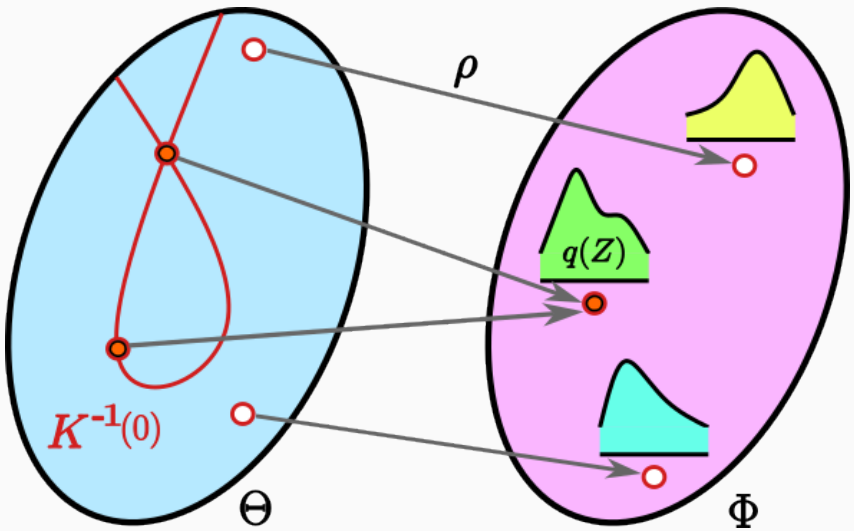
The Generalization Paradox



Model selection: choosing which model will best fit the data.

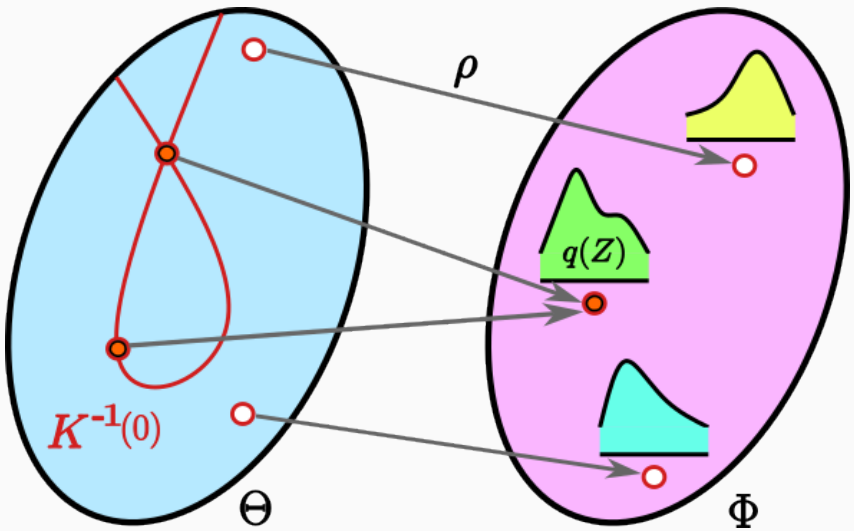
Deep learning (artificial neural networks, transformers) challenges the assumptions of **classical learning theory**. Why don't neural networks overfit when the number of layers is increased? Why do they generalize so well? Parameter count is not a good approach.

Parameterized Statistical Models



Param. Stat. Models:

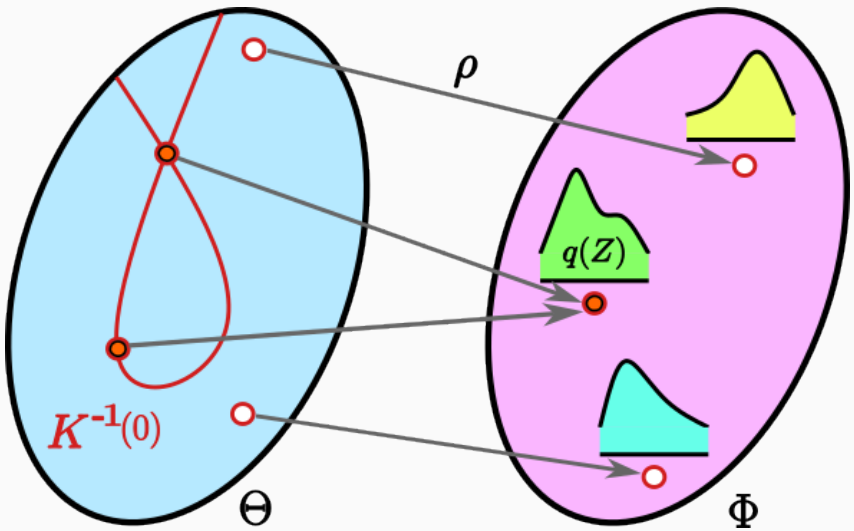
Parameterized Statistical Models



Param. Stat. Models:

- A **true distribution** $q(Z) \in \mathcal{P}(\mathbb{R}^m)$;

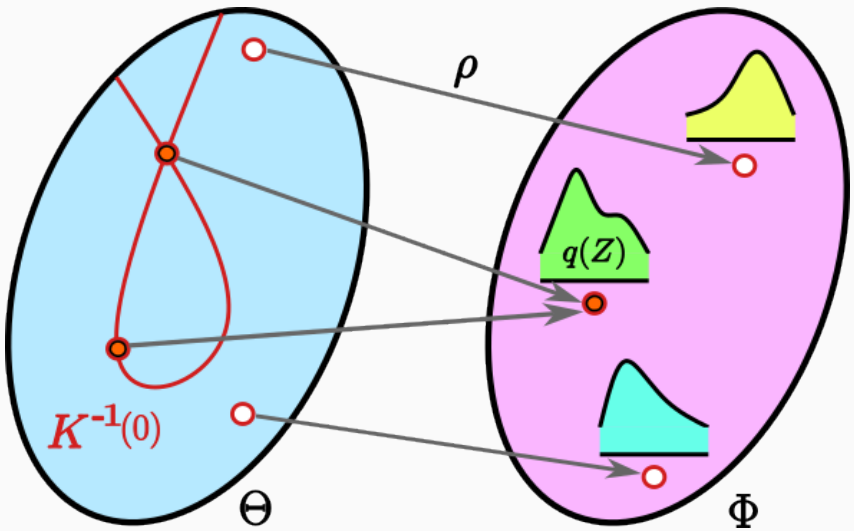
Parameterized Statistical Models



Param. Stat. Models:

- A **true distribution** $q(Z) \in \mathcal{P}(\mathbb{R}^m)$;
- A (compact) **parameter space** $\Theta \subset \mathbb{R}^d$;

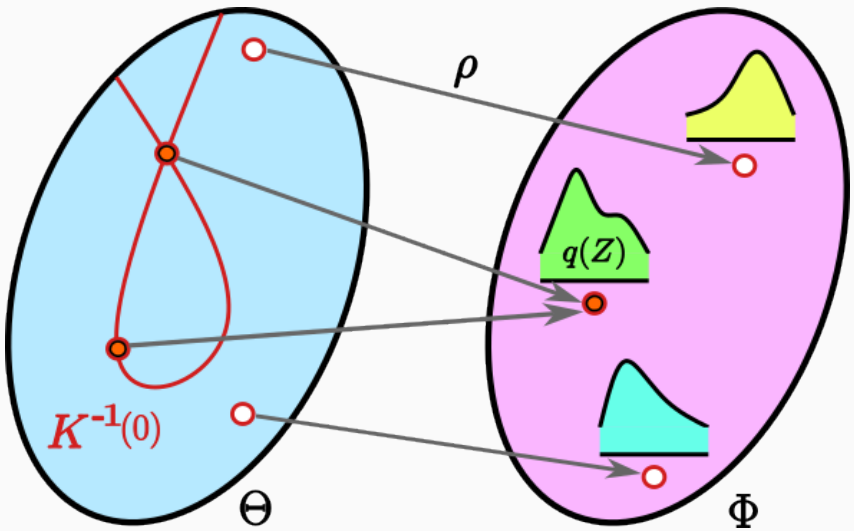
Parameterized Statistical Models



Param. Stat. Models:

- A **true distribution** $q(Z) \in \mathcal{P}(\mathbb{R}^m)$;
- A (compact) **parameter space** $\Theta \subset \mathbb{R}^d$;
- A **parameter-to-model map** $\rho : \Theta \to \Phi \subseteq \mathcal{P}(\mathbb{R}^m)$;

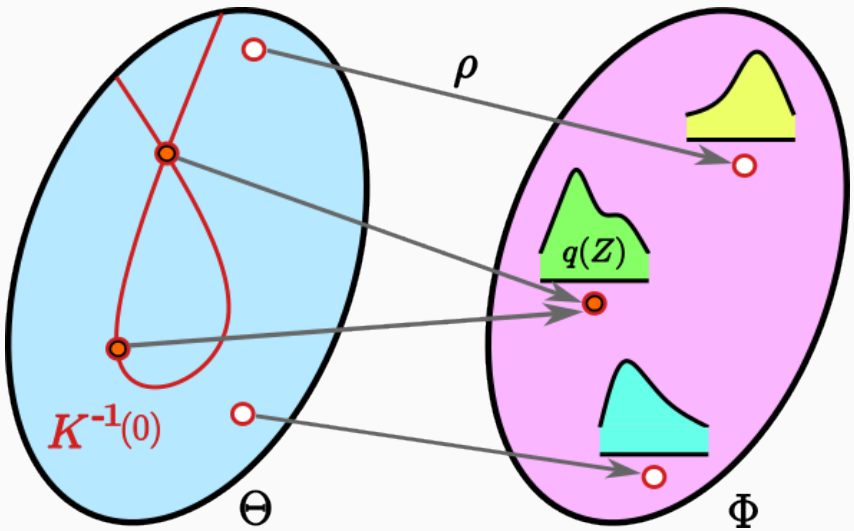
Parameterized Statistical Models



Param. Stat. Models:

- A **true distribution** $q(Z) \in \mathcal{P}(\mathbb{R}^m)$;
- A (compact) **parameter space** $\Theta \subset \mathbb{R}^d$;
- A **parameter-to-model map** $\rho : \Theta \to \Phi \subseteq \mathcal{P}(\mathbb{R}^m)$;
- The **Kullback-Leibler divergence** $K(\theta) = D_{KL}(q(Z) \parallel \rho(\theta)(Z))$. $K(\theta) = 0 \iff \rho(\theta)(Z) = q(Z)$.

Parameterized Statistical Models



Param. Stat. Models:

- A **true distribution** $q(Z) \in \mathcal{P}(\mathbb{R}^m)$;
- A (compact) **parameter space** $\Theta \subset \mathbb{R}^d$;
- A **parameter-to-model map** $\rho : \Theta \to \Phi \subseteq \mathcal{P}(\mathbb{R}^m)$;
- The **Kullback-Leibler divergence** $K(\theta) = D_{KL}(q(Z) \parallel \rho(\theta)(Z))$. $K(\theta) = 0 \iff \rho(\theta)(Z) = q(Z)$.

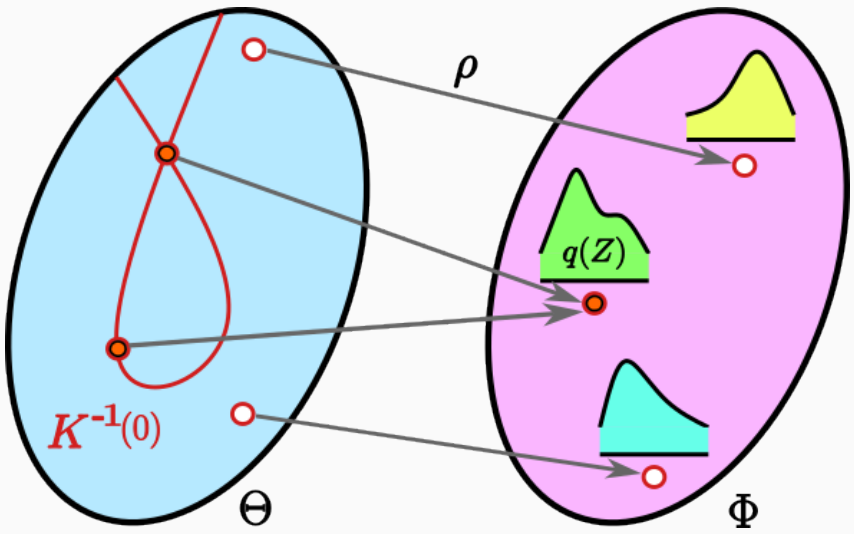
Regular and singular models

A pair $(\Phi, q(Z))$ is **regular** if it satisfies:

1. (**Identifiability**) $\rho : \Theta \to \Phi$ is injective AND;
2. (**Positive-definiteness**) $\nabla^2 K(\theta) \succ 0$ for all $\theta \in \Theta$.

It is singular otherwise.

Parameterized Statistical Models



Param. Stat. Models:

- A **true distribution** $q(Z) \in \mathcal{P}(\mathbb{R}^m)$;
- A (compact) **parameter space** $\Theta \subset \mathbb{R}^d$;
- A **parameter-to-model map** $\rho : \Theta \to \Phi \subseteq \mathcal{P}(\mathbb{R}^m)$;
- The **Kullback-Leibler divergence** $K(\theta) = D_{KL}(q(Z) \parallel \rho(\theta)(Z))$. $K(\theta) = 0 \iff \rho(\theta)(Z) = q(Z)$.

Regular and singular models

A pair $(\Phi, q(Z))$ is **regular** if it satisfies:

1. (**Identifiability**) $\rho : \Theta \to \Phi$ is injective AND;
2. (**Positive-definiteness**) $\nabla^2 K(\theta) \succ 0$ for all $\theta \in \Theta$.

It is singular otherwise.

Most models in the wild are singular.

Generalization, Model Selection & Information Criteria

Statistical Learning Theory:

- A **dataset** $D_n = \{z_1, \ldots, z_n\} \in (\mathbb{R}^m)^n$, sampled i.i.d. from $q(Z)$,

Generalization, Model Selection & Information Criteria

Statistical Learning Theory:

- A **dataset** $D_n = \{z_1, \ldots, z_n\} \in (\mathbb{R}^m)^n$, sampled i.i.d. from $q(Z)$,
- A collection of models $\Phi_1, \ldots, \Phi_T$.

Generalization, Model Selection & Information Criteria

Statistical Learning Theory:

- A **dataset** $D_n = \{z_1, \ldots, z_n\} \in (\mathbb{R}^m)^n$, sampled i.i.d. from $q(Z)$,
- A collection of models $\Phi_1, \ldots, \Phi_T$.

Question: How do we select the best model?

Generalization, Model Selection & Information Criteria

Statistical Learning Theory:

- A **dataset** $D_n = \{z_1, \ldots, z_n\} \in (\mathbb{R}^m)^n$, sampled i.i.d. from $q(Z)$,
- A collection of models $\Phi_1, \ldots, \Phi_T$.

Question: How do we select the best model?
Answer: By minimizing $-\log p(D_n \mid \Phi_i)$ where $p(D_n \mid \Phi_i)$ is the **marginal likelihood** (or **evidence**) of Φ_i.

Generalization, Model Selection & Information Criteria

Statistical Learning Theory:

- A **dataset** $D_n = \{z_1, \ldots, z_n\} \in (\mathbb{R}^m)^n$, sampled i.i.d. from $q(Z)$,
- A collection of models $\Phi_1, \ldots, \Phi_T$.

Question: How do we select the best model?
Answer: By minimizing $-\log p(D_n \mid \Phi_i)$ where $p(D_n \mid \Phi_i)$ is the **marginal likelihood** (or **evidence**) of Φ_i.
Problem: The marginal likelihood is an integral over the parameter space $\Theta_i \subset \mathbb{R}^d$ which is **hard** to compute.

Generalization, Model Selection & Information Criteria

Statistical Learning Theory:

- A **dataset** $D_n = \{z_1, \ldots, z_n\} \in (\mathbb{R}^m)^n$, sampled i.i.d. from $q(Z)$,
- A collection of models $\Phi_1, \ldots, \Phi_T$.

Question: How do we select the best model?

Answer: By minimizing $-\log p(D_n \mid \Phi_i)$ where $p(D_n \mid \Phi_i)$ is the **marginal likelihood** (or **evidence**) of Φ_i.

Problem: The marginal likelihood is an integral over the parameter space $\Theta_i \subset \mathbb{R}^d$ which is **hard** to compute.

Approach: Asymptotic approximation of the ML leads to the:

Bayesian Information Criterion [Schwarz'78]

$$\mathrm{BIC}(\Phi) = \underbrace{-nL_n(\theta_{\mathrm{MLE}})}_{\text{bias}} + \underbrace{\tfrac{d}{2}\log n}_{\text{variance}},$$

only when the model is regular!

Generalization, Model Selection & Information Criteria

Statistical Learning Theory:

- A **dataset** $D_n = \{z_1, \ldots, z_n\} \in (\mathbb{R}^m)^n$, sampled i.i.d. from $q(Z)$,
- A collection of models $\Phi_1, \ldots, \Phi_T$.

Question: How do we select the best model?

Answer: By minimizing $-\log p(D_n \mid \Phi_i)$ where $p(D_n \mid \Phi_i)$ is the **marginal likelihood** (or **evidence**) of Φ_i.

Problem: The marginal likelihood is an integral over the parameter space $\Theta_i \subset \mathbb{R}^d$ which is **hard** to compute.

Approach: Asymptotic approximation of the ML leads to the:

Singular Bayesian Information Criterion [Watanabe'13]

$$\mathrm{WBIC}(\Phi) = \underbrace{-nL_n(\theta_0)}_{\text{bias}} + \underbrace{\lambda \log n}_{\text{variance}},$$

where the **learning coefficient** $\lambda \leq \frac{d}{2}$, with equality for regular models.

Learning Coefficients as Real Log Canonical Thresholds

Singular Bayesian Information Criterion [Watanabe'13]

$$\mathrm{WBIC}(\Phi) = -nL_n(\theta_0) + \lambda \log n$$

Learning Coefficients as Real Log Canonical Thresholds

Singular Bayesian Information Criterion [Watanabe'13]

$$\mathrm{WBIC}(\Phi) = -nL_n(\theta_0) + \lambda \log n$$

- The **learning coefficient** λ depends on Φ and $q(Z)$.

Learning Coefficients as Real Log Canonical Thresholds

Singular Bayesian Information Criterion [Watanabe'13]

$$\mathrm{WBIC}(\Phi) = -nL_n(\theta_0) + \lambda \log n$$

- The **learning coefficient** λ depends on Φ and $q(Z)$.
- It coincides with the **real log canonical threshold** of $K(\theta)$ [W'09]:

Learning Coefficients as Real Log Canonical Thresholds

Singular Bayesian Information Criterion [Watanabe'13]

$$\mathrm{WBIC}(\Phi) = -nL_n(\theta_0) + \lambda \log n$$

- The **learning coefficient** λ depends on Φ and $q(Z)$.
- It coincides with the **real log canonical threshold** of $K(\theta)$ [W'09]:

Real Log Canonical Threshold

$$\mathrm{RLCT}(K) = \sup\left\{ s \in \mathbb{R}_{\geq 0} \mid \int_\Theta |K(\theta)|^{-s} d\theta < \infty \right\}.$$

Learning Coefficients as Real Log Canonical Thresholds

Singular Bayesian Information Criterion [Watanabe'13]

$$\mathrm{WBIC}(\Phi) = -nL_n(\theta_0) + \lambda \log n$$

- The local **learning coefficient** λ depends on Φ and $q(Z)$.
- It coincides with the local **real log canonical threshold** of $K(\theta)$ [W'09]:

(Local) Real Log Canonical Threshold

$$\mathrm{RLCT}_{\theta_0}(K) = \sup\left\{ s \in \mathbb{R}_{\geq 0} \mid \exists \epsilon > 0 \text{ s.t. } \int_{B_\epsilon(\theta_0)} |K(\theta)|^{-s} d\theta < \infty \right\}.$$

- RLCTs can be computed using Hironaka's **resolution of singularities** (existence of RoS [Hironaka'64]).

Learning Coefficients as Real Log Canonical Thresholds

Singular Bayesian Information Criterion [Watanabe'13]

$$\mathrm{WBIC}(\Phi) = -nL_n(\theta_0) + \lambda \log n$$

- The local **learning coefficient** λ depends on Φ and $q(Z)$.
- It coincides with the local **real log canonical threshold** of $K(\theta)$ [W'09]:

(Local) Real Log Canonical Threshold

$$\mathrm{RLCT}_{\theta_0}(K) = \sup\left\{ s \in \mathbb{R}_{\geq 0} \mid \exists \epsilon > 0 \text{ s.t. } \int_{B_\epsilon(\theta_0)} |K(\theta)|^{-s} d\theta < \infty \right\}.$$

- RLCTs can be computed using Hironaka's **resolution of singularities** (existence of RoS [Hironaka'64]).
- But resolutions of singularities are notoriously hard to compute!

Prior and Present Work

- resolution of singularities [H'64],

Prior and Present Work

- resolution of singularities [H'64],
- an iterative procedure to compute RLCTs for $d = 2$ [Varchenko'76], [Phong et al.'99], [Collins'18]

Prior and Present Work

- resolution of singularities [H'64],
- an iterative procedure to compute RLCTs for $d = 2$ [Varchenko'76], [Phong et al.'99], [Collins'18]
  - Problem: termination is not guaranteed. No complexity analysis.

Prior and Present Work

- resolution of singularities [H'64],
- an iterative procedure to compute RLCTs for $d = 2$ [Varchenko'76], [Phong et al.'99], [Collins'18]
  - Problem: termination is not guaranteed. No complexity analysis.

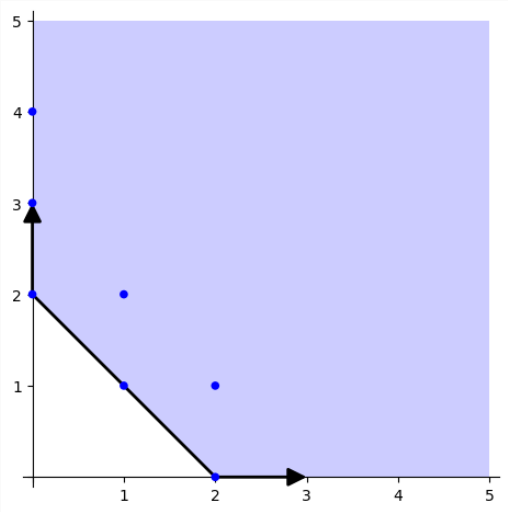
This work

We modify the works of [V'76],[P et al'99],[C'18] to guarantee termination. This gives an exact algorithm to compute RLCTs in $d = 2$. We discuss its complexity.

Table of Contents

Newton Polygon

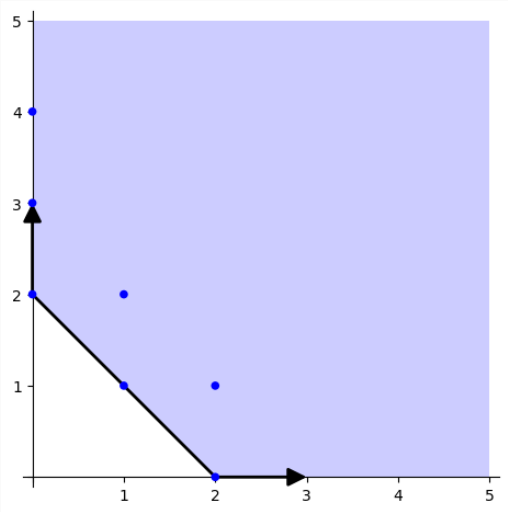
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X, Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$

Newton Polygon

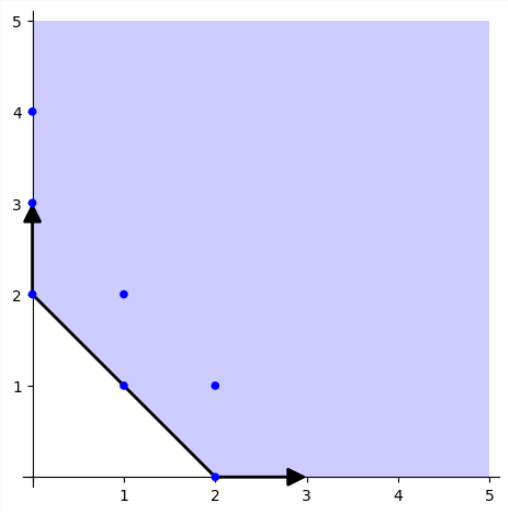
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$

Newton Polygon

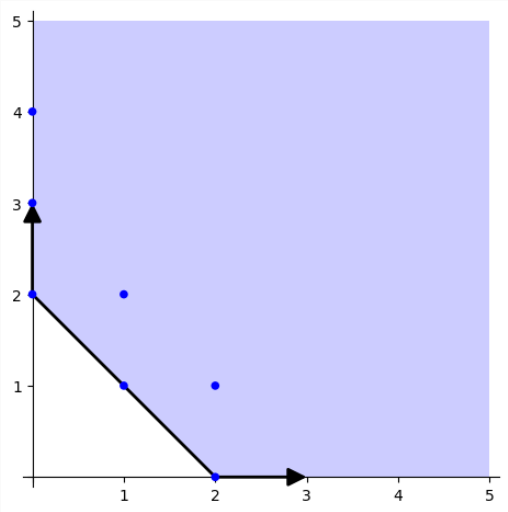
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$

Newton Polygon

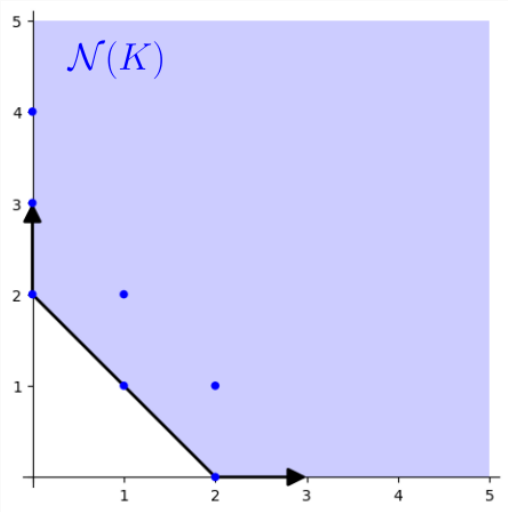
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$
  - $S_K = \{(0, 4), (0, 3), (0, 2), (1, 2), (1, 1), (2, 1), (2, 0)\}$

Newton Polygon

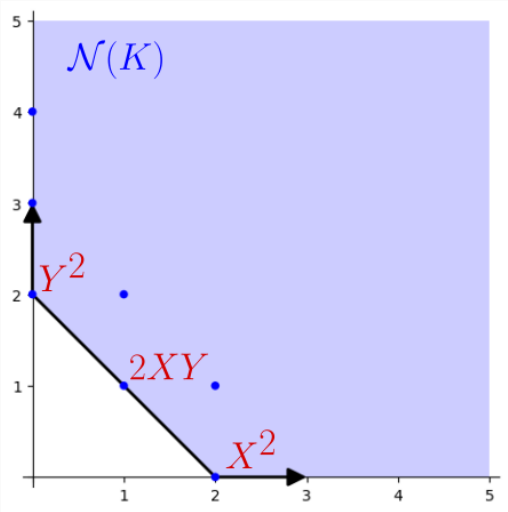
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$
  - $S_K = \{(0, 4), (0, 3), (0, 2), (1, 2), (1, 1), (2, 1), (2, 0)\}$
- **Newton polygon:** $\mathcal{N}(K) = \bigcup_{\boldsymbol{p} \in S_K} \boldsymbol{p} \oplus \mathbb{N}^2$

Newton Polygon

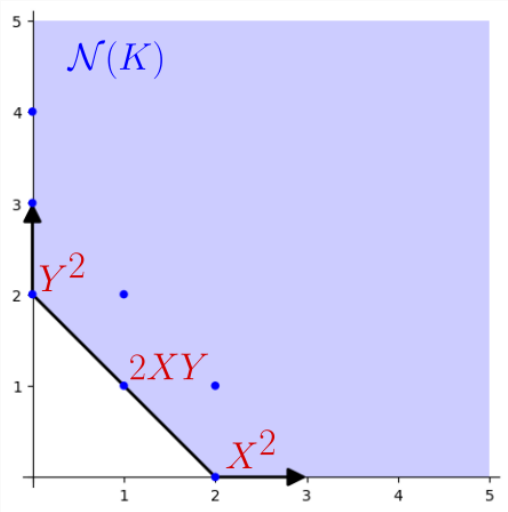
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$
  - $S_K = \{(0, 4), (0, 3), (0, 2), (1, 2), (1, 1), (2, 1), (2, 0)\}$
- **Newton polygon:** $\mathcal{N}(K) = \bigcup_{\boldsymbol{p} \in S_K} \boldsymbol{p} \oplus \mathbb{N}^2$
- **Facet polynomial:** $K_\Delta = \sum_{(p,q) \in S_K \cap \Delta} a_{pq} X^p Y^q$

Newton Polygon

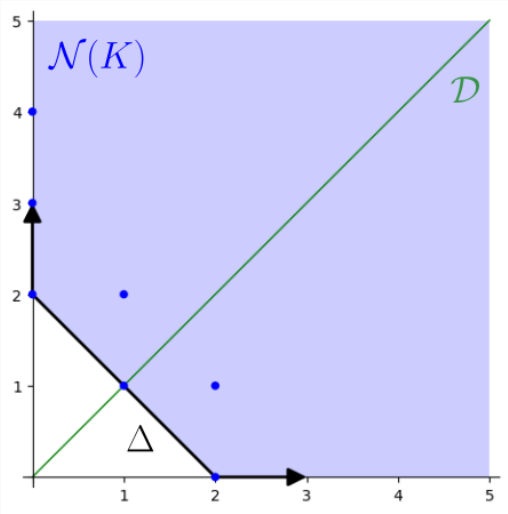
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$
  - $S_K = \{(0, 4), (0, 3), (0, 2), (1, 2), (1, 1), (2, 1), (2, 0)\}$
- **Newton polygon:** $\mathcal{N}(K) = \bigcup_{\boldsymbol{p} \in S_K} \boldsymbol{p} \oplus \mathbb{N}^2$
- **Facet polynomial:** $K_\Delta = \sum_{(p,q) \in S_K \cap \Delta} a_{pq} X^p Y^q$
  - $K_\Delta = Y^2 + 2XY + X^2$

Newton Polygon

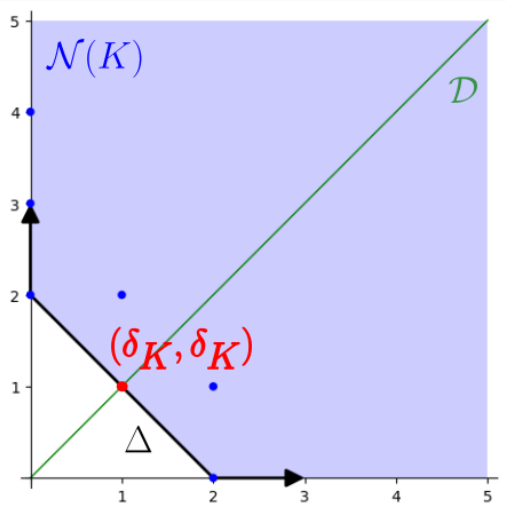
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$
  - $S_K = \{(0, 4), (0, 3), (0, 2), (1, 2), (1, 1), (2, 1), (2, 0)\}$
- **Newton polygon:** $\mathcal{N}(K) = \bigcup_{\boldsymbol{p} \in S_K} \boldsymbol{p} \oplus \mathbb{N}^2$
- **Facet polynomial:** $K_\Delta = \sum_{(p,q) \in S_K \cap \Delta} a_{pq} X^p Y^q$
  - $K_\Delta = Y^2 + 2XY + X^2$
- **Main facet:** $\Delta =$ facet intersected by $\mathcal{D}$.

Newton Polygon

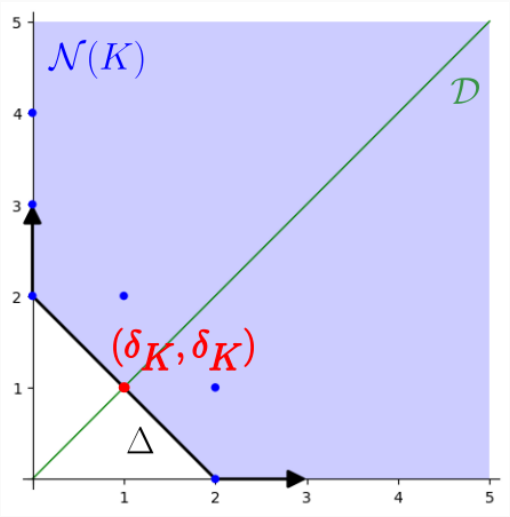
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$
  - $S_K = \{(0, 4), (0, 3), (0, 2), (1, 2), (1, 1), (2, 1), (2, 0)\}$
- **Newton polygon:** $\mathcal{N}(K) = \bigcup_{\boldsymbol{p} \in S_K} \boldsymbol{p} \oplus \mathbb{N}^2$
- **Facet polynomial:** $K_\Delta = \sum_{(p,q) \in S_K \cap \Delta} a_{pq} X^p Y^q$
  - $K_\Delta = Y^2 + 2XY + X^2$
- **Main facet:** $\Delta =$ facet intersected by $\mathcal{D}$.
- **Newton distance:** $\delta_K =$ coordinate of $\mathcal{D} \cap \partial\mathcal{N}(K)$.

Newton Polygon

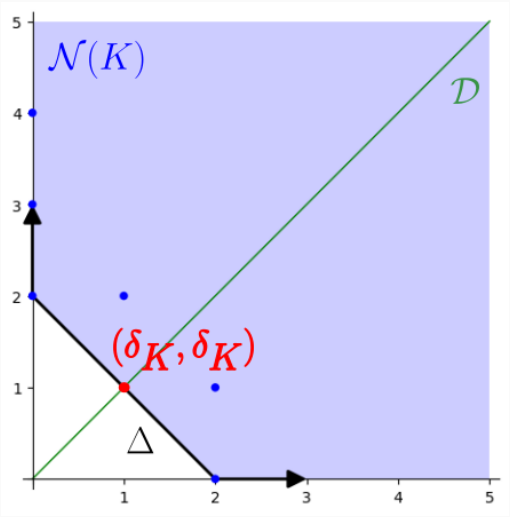
Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$
  - $S_K = \{(0, 4), (0, 3), (0, 2), (1, 2), (1, 1), (2, 1), (2, 0)\}$
- **Newton polygon:** $\mathcal{N}(K) = \bigcup_{\boldsymbol{p} \in S_K} \boldsymbol{p} \oplus \mathbb{N}^2$
- **Facet polynomial:** $K_\Delta = \sum_{(p,q) \in S_K \cap \Delta} a_{pq} X^p Y^q$
  - $K_\Delta = Y^2 + 2XY + X^2$
- **Main facet:** $\Delta =$ facet intersected by $\mathcal{D}$.
- **Newton distance:** $\delta_K =$ coordinate of $\mathcal{D} \cap \partial\mathcal{N}(K)$.
  - $\delta_K = 1$

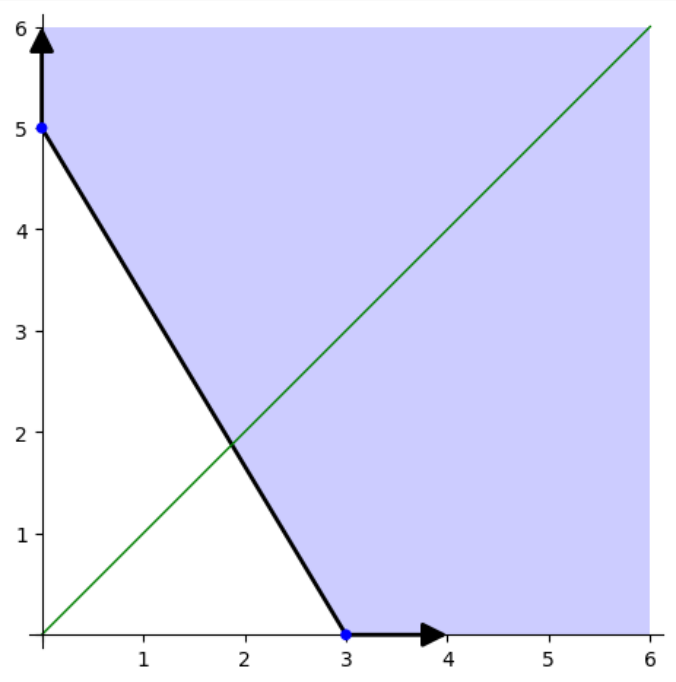
Newton Polygon

Let $d = 2$, and $\theta = (X, Y)$. Let $K(\theta) = K(X, Y) = \sum a_{pq} X^p Y^q \in \mathbb{Q}[X][Y]$.



- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- **Support:** $S_K = \{(p, q) \mid a_{pq} \neq 0\} \subseteq \mathbb{N}^2$
  - $S_K = \{(0, 4), (0, 3), (0, 2), (1, 2), (1, 1), (2, 1), (2, 0)\}$
- **Newton polygon:** $\mathcal{N}(K) = \bigcup_{\boldsymbol{p} \in S_K} \boldsymbol{p} \oplus \mathbb{N}^2$
- **Facet polynomial:** $K_\Delta = \sum_{(p,q) \in S_K \cap \Delta} a_{pq} X^p Y^q$
  - $K_\Delta = Y^2 + 2XY + X^2$
- **Main facet:** $\Delta =$ facet intersected by $\mathcal{D}$.
- **Newton distance:** $\delta_K =$ coordinate of $\mathcal{D} \cap \partial\mathcal{N}(K)$.
  - $\delta_K = 1$

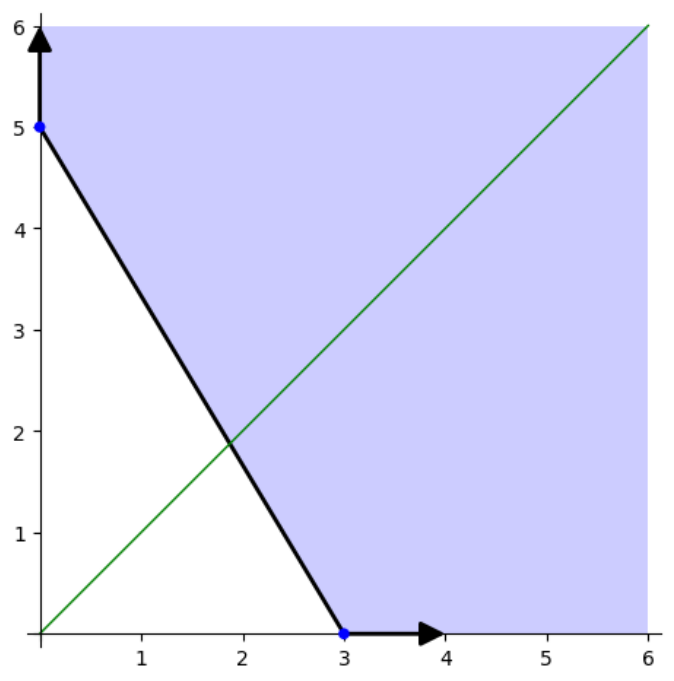
Varchenko's Theorem



Normalized

We say that K is normalized if K_Δ has no factor in $\bar{\mathbb{Q}}[X, Y]$ of multiplicity $m > \delta_K$.

Varchenko's Theorem

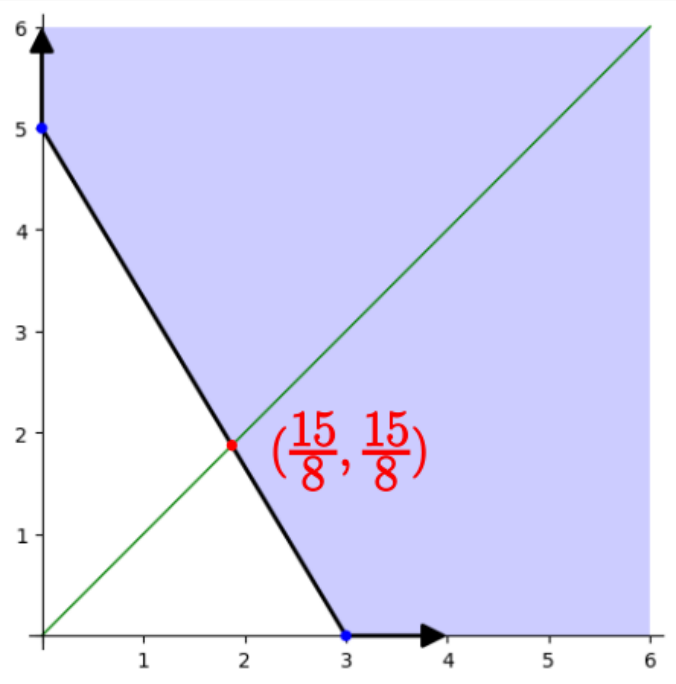


Normalized

We say that K is normalized if K_Δ has no factor in $\bar{\mathbb{Q}}[X, Y]$ of multiplicity $m > \delta_K$.

- $H(X, Y) = H_\Delta(X, Y) = Y^5 + X^3$

Varchenko's Theorem

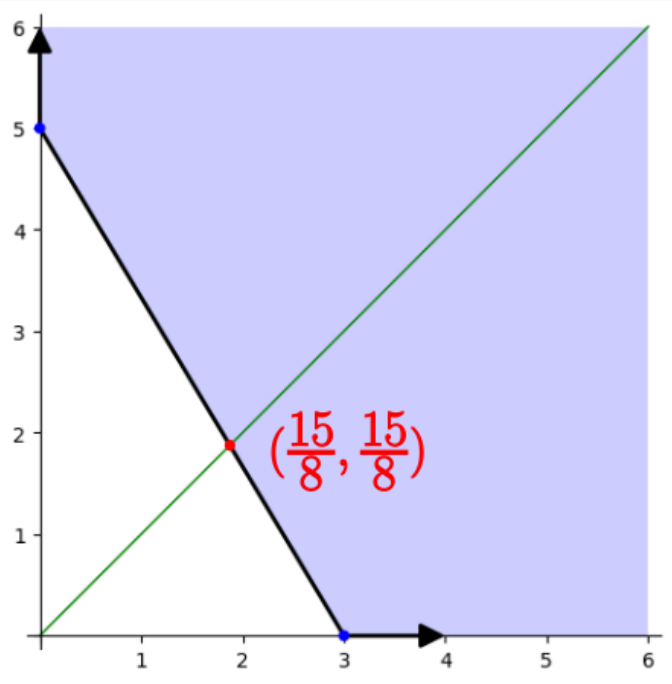


Normalized

We say that K is normalized if K_Δ has no factor in $\bar{\mathbb{Q}}[X, Y]$ of multiplicity $m > \delta_K$.

- $H(X, Y) = H_\Delta(X, Y) = Y^5 + X^3$
- $m = 1, \delta_H = \frac{15}{8} > m$

Varchenko's Theorem

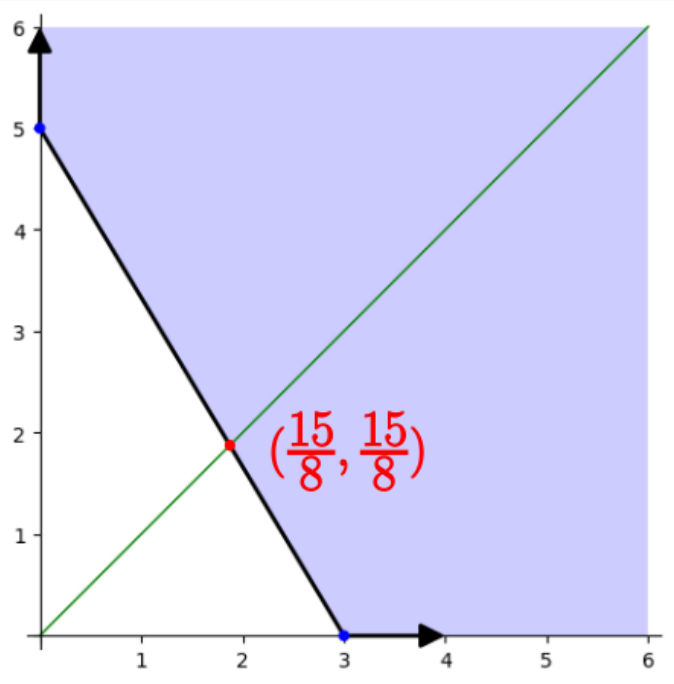


Normalized

We say that K is normalized if K_Δ has no factor in $\bar{\mathbb{Q}}[X, Y]$ of multiplicity $m > \delta_K$.

- $H(X, Y) = H_\Delta(X, Y) = Y^5 + X^3$
- $m = 1, \delta_H = \frac{15}{8} > m$
- $H(X, Y)$ is normalized.

Varchenko's Theorem



Normalized

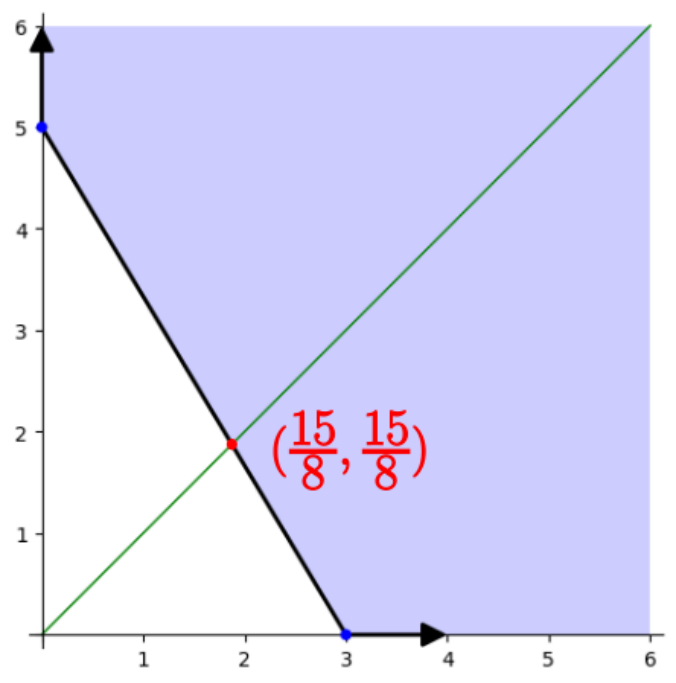
We say that K is normalized if K_Δ has no factor in $\bar{\mathbb{Q}}[X, Y]$ of multiplicity $m > \delta_K$.

- $H(X, Y) = H_\Delta(X, Y) = Y^5 + X^3$
- $m = 1, \delta_H = \frac{15}{8} > m$
- $H(X, Y)$ is normalized.

Theorem [Varchenko, PSS, Collins]

If $K(X, Y)$ is normalized, then $\mathrm{RLCT}_0(K) = \frac{1}{\delta_K}$. If $K(X, Y)$ is not normalized, there exists an change of variables $\Psi : (X, Y) \mapsto (X, Y - Q(X))$, with $Q(X) \in \bar{\mathbb{Q}}[[X]]$ such that $\tilde{K}(X, Y) = K \circ \Psi(X, Y)$ is normalized and $\mathrm{RLCT}_0(\tilde{K}) = \mathrm{RLCT}_0(K) = \frac{1}{\delta_{\tilde{K}}}$.

Varchenko's Theorem



Normalized

We say that K is normalized if K_Δ has no factor in $\bar{\mathbb{Q}}[X, Y]$ of multiplicity $m > \delta_K$.

- $H(X, Y) = H_\Delta(X, Y) = Y^5 + X^3$
- $m = 1, \delta_H = \frac{15}{8} > m$
- $H(X, Y)$ is normalized.

Theorem [Varchenko, PSS, Collins]

If $K(X, Y)$ is normalized, then $\mathrm{RLCT}_0(K) = \frac{1}{\delta_K}$. If $K(X, Y)$ is not normalized, there exists an change of variables $\Psi : (X, Y) \mapsto (X, Y - Q(X))$, with $Q(X) \in \bar{\mathbb{Q}}[[X]]$ such that $\tilde{K}(X, Y) = K \circ \Psi(X, Y)$ is normalized and $\mathrm{RLCT}_0(\tilde{K}) = \mathrm{RLCT}_0(K) = \frac{1}{\delta_{\tilde{K}}}$.

Ψ can be computed with the Newton-Puiseux algorithm + Normalized check!

Puiseux Theorem

Puiseux Theorem

Let $K(X, Y)$ be a polynomial in $\mathbb{Q}[X][Y]$. The algebraic closure of $\mathbb{Q}[X]$ is contained in the field of Puiseux series $\mathbb{C}\langle\langle X\rangle\rangle = \bigcup_n \mathbb{C}((X^{1/n}))$.

Puiseux Theorem

Puiseux Theorem

Let $K(X, Y)$ be a polynomial in $\mathbb{Q}[X][Y]$. The algebraic closure of $\mathbb{Q}[X]$ is contained in the field of Puiseux series $\mathbb{C}\langle\langle X\rangle\rangle = \bigcup_n \mathbb{C}((X^{1/n}))$.

Example:

- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
  - $\alpha_i(X) = \xi_i + (2\xi_i + 2)X - X^2 + (-2\xi_i - 2)X^3 + \ldots$, where ξ_i are roots of $T^2 + T + 1$, for $i = 1, 2$
  - $\beta_j(X) = -X + X^2 + \nu_j X^{\frac{5}{2}} + X^3 + \ldots$, where ν_j are roots of $T^2 - 3$, for $j = 1, 2$

Puiseux Theorem

Puiseux Theorem

Let $K(X, Y)$ be a polynomial in $\mathbb{Q}[X][Y]$. The algebraic closure of $\mathbb{Q}[X]$ is contained in the field of Puiseux series $\mathbb{C}\langle\langle X\rangle\rangle = \bigcup_n \mathbb{C}((X^{1/n}))$.

Example:

- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
  - $\alpha_i(X) = \xi_i + (2\xi_i + 2)X - X^2 + (-2\xi_i - 2)X^3 + \ldots$, where ξ_i are roots of $T^2 + T + 1$, for $i = 1, 2$
  - $\beta_j(X) = -X + X^2 + \nu_j X^{\frac{5}{2}} + X^3 + \ldots$, where ν_j are roots of $T^2 - 3$, for $j = 1, 2$

Recall the change of variables $\Psi : (X, Y) \mapsto (X, Y - Q(X))$.

Puiseux Theorem

Puiseux Theorem

Let $K(X, Y)$ be a polynomial in $\mathbb{Q}[X][Y]$. The algebraic closure of $\mathbb{Q}[X]$ is contained in the field of Puiseux series $\mathbb{C}\langle\langle X\rangle\rangle = \bigcup_n \mathbb{C}((X^{1/n}))$.

Example:

- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
  - $\alpha_i(X) = \xi_i + (2\xi_i + 2)X - X^2 + (-2\xi_i - 2)X^3 + \ldots$, where ξ_i are roots of $T^2 + T + 1$, for $i = 1, 2$
  - $\beta_j(X) = -X + X^2 + \nu_j X^{\frac{5}{2}} + X^3 + \ldots$, where ν_j are roots of $T^2 - 3$, for $j = 1, 2$

Recall the change of variables $\Psi : (X, Y) \mapsto (X, Y - Q(X))$.

$Q(X) \in \bar{\mathbb{Q}}[[X]]$ is given by (a subpart of) the polynomial part of a certain Puiseux root of $K(X, Y)$.

Puiseux Theorem

Puiseux Theorem

Let $K(X, Y)$ be a polynomial in $\mathbb{Q}[X][Y]$. The algebraic closure of $\mathbb{Q}[X]$ is contained in the field of Puiseux series $\mathbb{C}\langle\langle X\rangle\rangle = \bigcup_n \mathbb{C}((X^{1/n}))$.

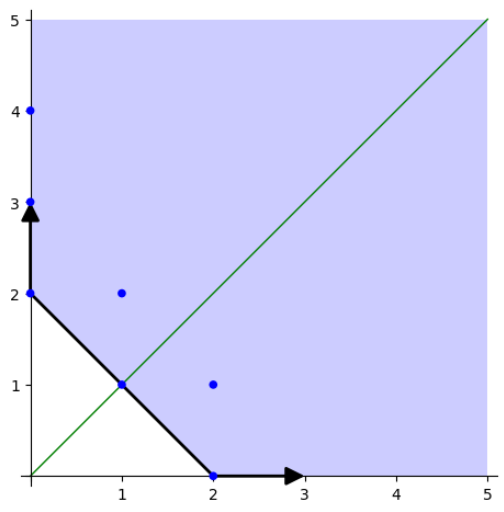
Example:

- $K(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
  - $\alpha_i(X) = \xi_i + (2\xi_i + 2)X - X^2 + (-2\xi_i - 2)X^3 + \ldots$, where ξ_i are roots of $T^2 + T + 1$, for $i = 1, 2$
  - $\beta_j(X) = -X + X^2 + \nu_j X^{\frac{5}{2}} + X^3 + \ldots$, where ν_j are roots of $T^2 - 3$, for $j = 1, 2$

Recall the change of variables $\Psi : (X, Y) \mapsto (X, Y - Q(X))$.

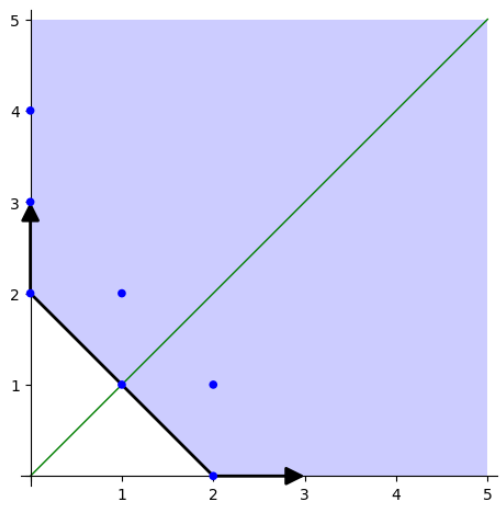
$Q(X) \in \bar{\mathbb{Q}}[[X]]$ is given by (a subpart of) the polynomial part of a certain Puiseux root of $K(X, Y)$.

Varchenko's Procedure



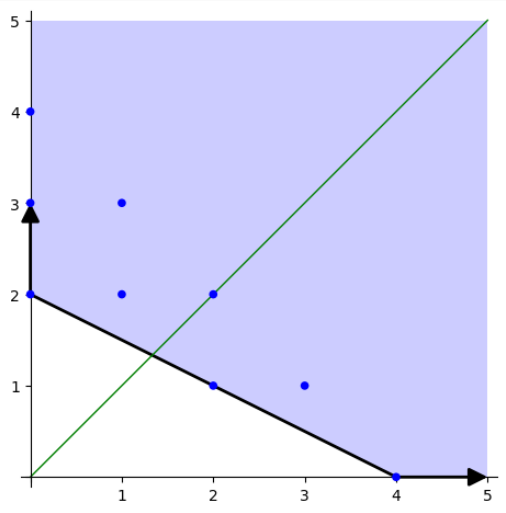
- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 + (\text{higher-order terms})$
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Varchenko's Procedure



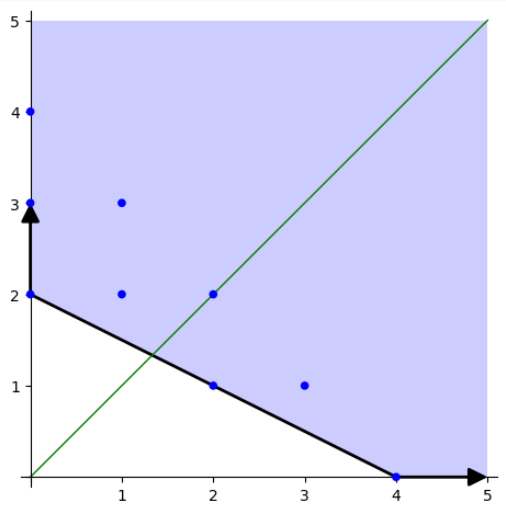
- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 + (\text{higher-order terms})$
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Varchenko's Procedure



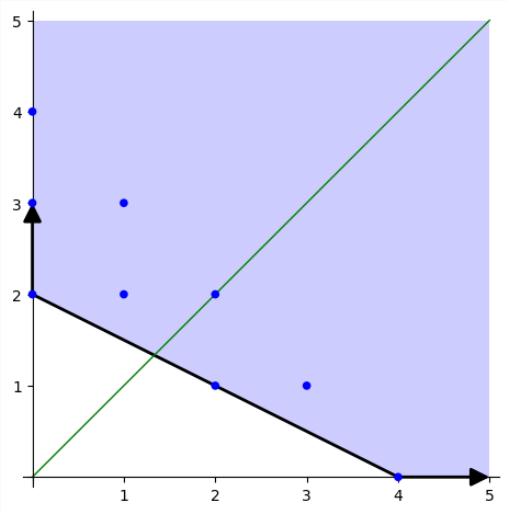
- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 + (\text{higher-order terms})$
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Varchenko's Procedure



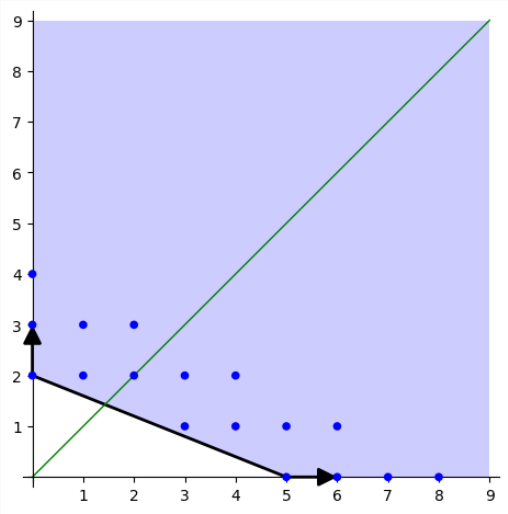
- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 +$ (higher-order terms)
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Varchenko's Procedure



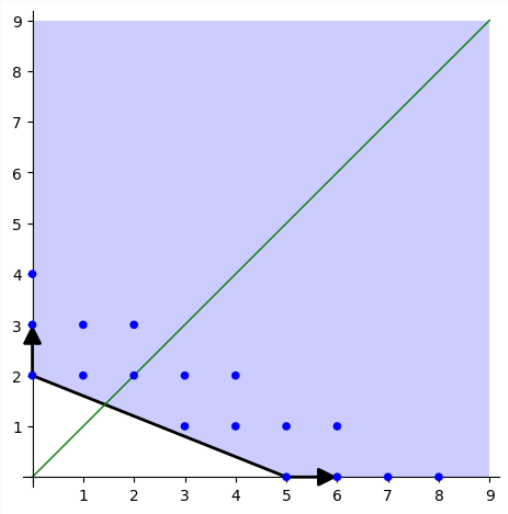
- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 +$ (higher-order terms)
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Varchenko's Procedure



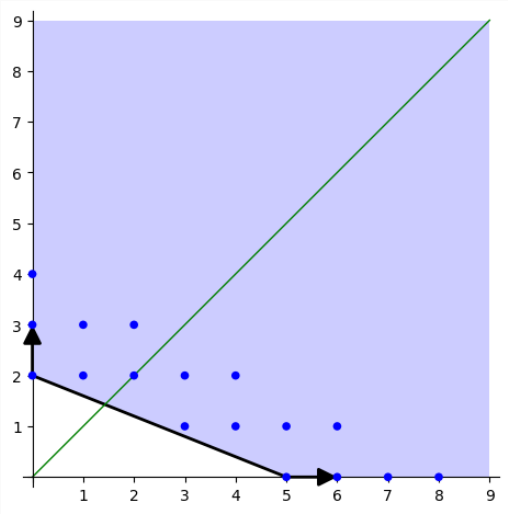
- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 + (\text{higher-order terms})$
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Varchenko's Procedure



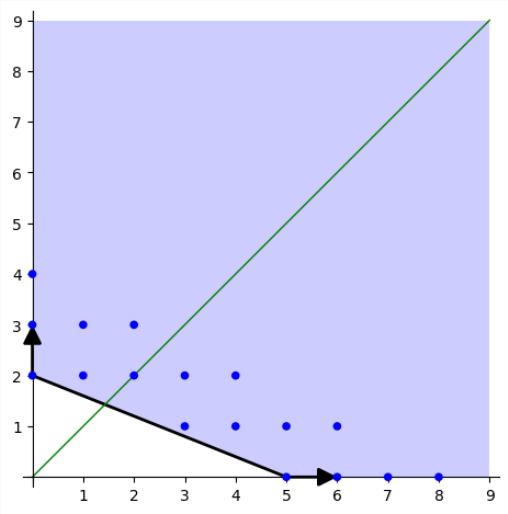
- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 +$ (higher-order terms)
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Varchenko's Procedure



- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 +$ (higher-order terms)
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Varchenko's Procedure



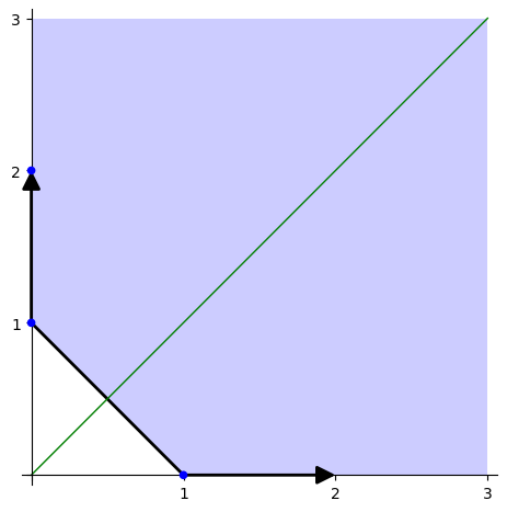
- Take $K(X, Y) = K_0(X, Y) = Y^2 + 2XY + X^2 + Y^3 + 4XY^2 + 3X^2Y + Y^4$
- $\delta_{K_0} = 1$, $(K_0)_\Delta = Y^2 + 2XY + X^2 = (Y + X)^2$, $m_0 = 2 > \delta_{K_0}$
- Apply the change of variables $K_1(X, Y) = K_0(X, Y - X)$
- Get $K_1(X, Y) = Y^2 + (-2)X^2Y + X^4 + (-4)X^3Y + 6X^2Y^2 + (-4)XY^3 + Y^4 + XY^2 + Y^3$
- $\delta_{K_1} = 4/3$, $(K_1)_\Delta = Y^2 - 2X^2Y + X^4 = (Y - X^2)^2$, $m_1 = 2 > \delta_{K_1}$
- Apply the change of variables $K_2(X, Y) = K_1(X, Y + X^2)$
- Get $K_2(X, Y) = Y^2 - 3X^5 +$ (higher-order terms)
- $\delta_{K_2} = 10/7$, $(K_2)_\Delta = Y^2 - 3X^5$, $m_2 = 1 < \delta_{K_2} = 10/7$
- $\mathrm{RLCT}_0(K) = 7/10$

Termination and Complexity

But termination is not guaranteed.

Termination and Complexity

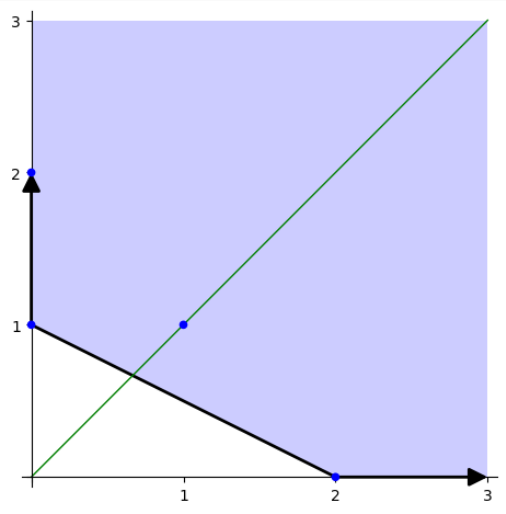
But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$

Termination and Complexity

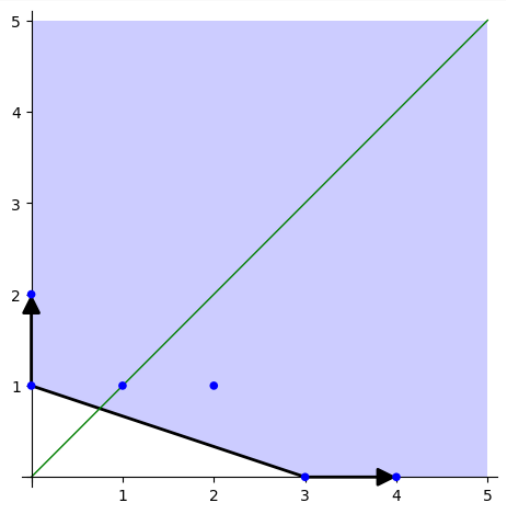
But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$

Termination and Complexity

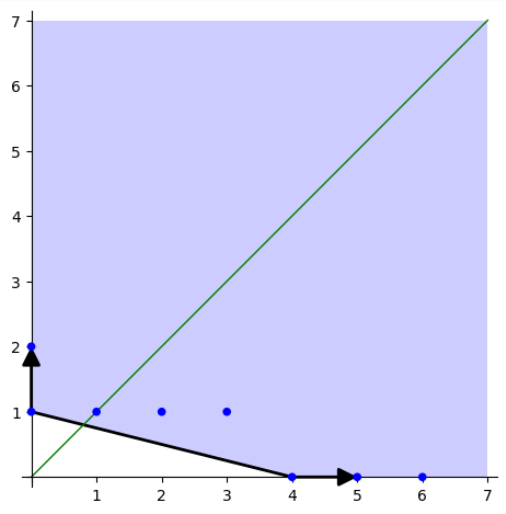
But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$

Termination and Complexity

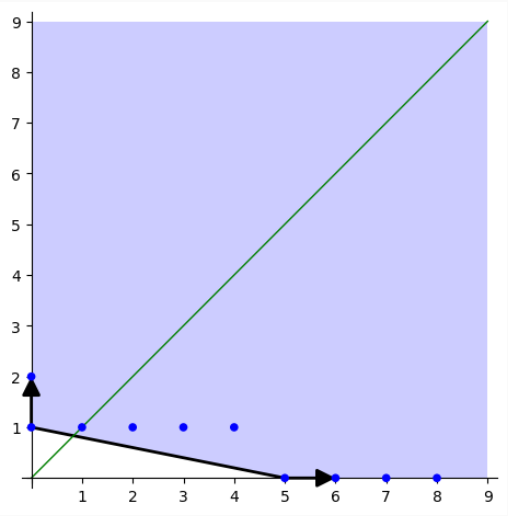
But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$

Termination and Complexity

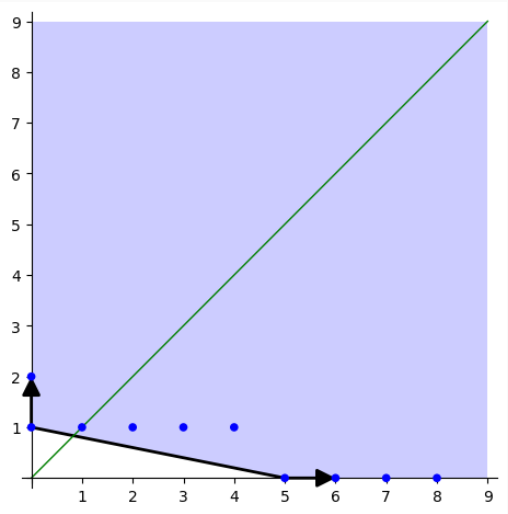
But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$

Termination and Complexity

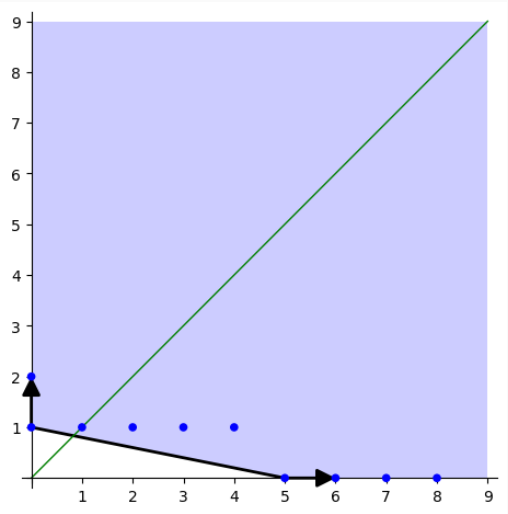
But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$
- Roots of H_0:
  - $\alpha(X) = -1 + X + X^2 + 2X^3 + 5X^4 + \ldots$
  - $\beta(X) = -X - X^2 - 2X^3 - 5X^4 - 14X^5 + \ldots$

Termination and Complexity

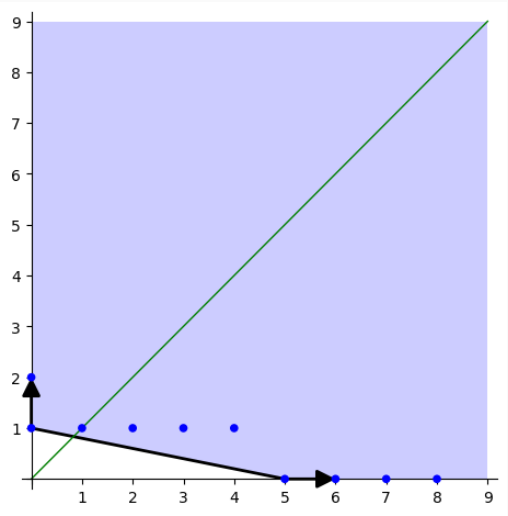
But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$
- Roots of H_0:
  - $\alpha(X) = -1 + X + X^2 + 2X^3 + 5X^4 + \ldots$
  - $\beta(X) = -X - X^2 - 2X^3 - 5X^4 - 14X^5 + \ldots$
- $1 = m_i > \delta_{H_i} = \frac{i+1}{i+2}$ for all $i \in \mathbb{N}$.

Termination and Complexity

But termination is not guaranteed.

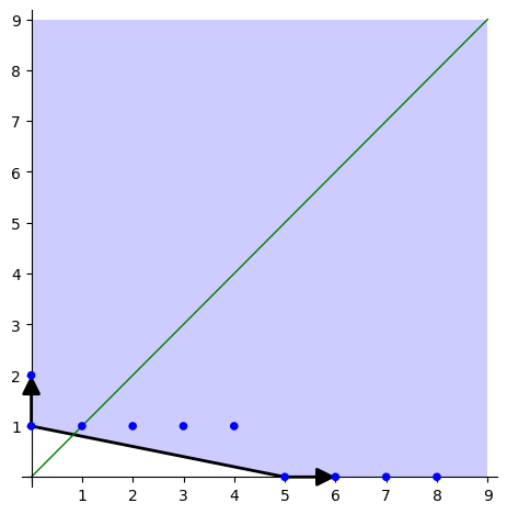


- $H_0(X, Y) = Y^2 + Y + X$
- Roots of H_0:
  - $\alpha(X) = -1 + X + X^2 + 2X^3 + 5X^4 + \ldots$
  - $\beta(X) = -X - X^2 - 2X^3 - 5X^4 - 14X^5 + \ldots$
- $1 = m_i > \delta_{H_i} = \frac{i+1}{i+2}$ for all $i \in \mathbb{N}$.

We can detect such counter-examples. Moreover,

Termination and Complexity

But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$
- Roots of H_0:
  - $\alpha(X) = -1 + X + X^2 + 2X^3 + 5X^4 + \ldots$
  - $\beta(X) = -X - X^2 - 2X^3 - 5X^4 - 14X^5 + \ldots$
- $1 = m_i > \delta_{H_i} = \frac{i+1}{i+2}$ for all $i \in \mathbb{N}$.

We can detect such counter-examples. Moreover,

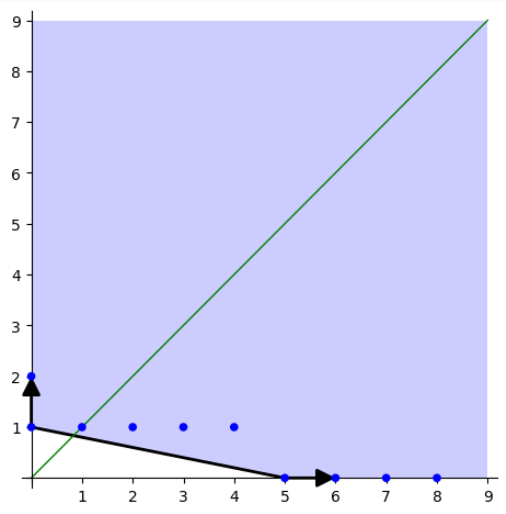
Theorem [Sergeant-Perthuis, Tsigaridas, T.]

Let $K(X, Y) \in \mathbb{Q}[X][Y]$, and $d_Y = \deg_Y K$, $d_X = \deg_X K$. Let $\alpha(X)$ and $\beta(X)$ be any two distinct roots of $K(X, Y)$. Then,

$$\mathrm{ord}(\alpha(X) - \beta(X)) \leq d_X(d_Y + \tfrac{1}{2})$$

Termination and Complexity

But termination is not guaranteed.



- $H_0(X, Y) = Y^2 + Y + X$
- Roots of H_0:
  - $\alpha(X) = -1 + X + X^2 + 2X^3 + 5X^4 + \ldots$
  - $\beta(X) = -X - X^2 - 2X^3 - 5X^4 - 14X^5 + \ldots$
- $1 = m_i > \delta_{H_i} = \frac{i+1}{i+2}$ for all $i \in \mathbb{N}$.

We can detect such counter-examples. Moreover,

Theorem [Sergeant-Perthuis, Tsigaridas, T.]

Let $K(X, Y) \in \mathbb{Q}[X][Y]$, and $d_Y = \deg_Y K$, $d_X = \deg_X K$. Let $\alpha(X)$ and $\beta(X)$ be any two distinct roots of $K(X, Y)$. Then,

$$\mathrm{ord}(\alpha(X) - \beta(X)) \leq d_X(d_Y + \tfrac{1}{2})$$

This can be used as an upper bound on the number of iterations of the algorithm!

Table of Contents

Application: Learning Coefficients of a PNN

We apply the algorithm to a family of **polynomial neural networks** (PNNs) with parameters $\theta = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^2$, activation degree $r = 2$, and a single weight $W(\theta)$ repeated L times (regression task). By [Aoyagi]'s theorem, the Kullback–Leibler divergence $K_L(\theta)$ is *contact equivalent* to an explicit sum-of-squares polynomial $H_L(\theta) \in \mathbb{Q}[\theta_1, \theta_2]$, so $\lambda_L = \mathrm{RLCT}_0(H_L)$.

Application: Learning Coefficients of a PNN

We apply the algorithm to a family of **polynomial neural networks** (PNNs) with parameters $\theta = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^2$, activation degree $r = 2$, and a single weight $W(\theta)$ repeated L times (regression task). By [Aoyagi]'s theorem, the Kullback–Leibler divergence $K_L(\theta)$ is *contact equivalent* to an explicit sum-of-squares polynomial $H_L(\theta) \in \mathbb{Q}[\theta_1, \theta_2]$, so $\lambda_L = \mathrm{RLCT}_0(H_L)$.

L	1	2	3	4
λ_L	$3/4$	$1/4$	$3/28$	$1/20$
Time	45.4ms	76.4ms	1.12s	43.8s

Table 1: Exact local RLCTs λ_L of $H_L(\theta)$ for $L \in [4]$, with wall-clock time.

Application: Learning Coefficients of a PNN

We apply the algorithm to a family of **polynomial neural networks** (PNNs) with parameters $\theta = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^2$, activation degree $r = 2$, and a single weight $W(\theta)$ repeated L times (regression task). By [Aoyagi]'s theorem, the Kullback–Leibler divergence $K_L(\theta)$ is *contact equivalent* to an explicit sum-of-squares polynomial $H_L(\theta) \in \mathbb{Q}[\theta_1, \theta_2]$, so $\lambda_L = \mathrm{RLCT}_0(H_L)$.

L	1	2	3	4
λ_L	$3/4$	$1/4$	$3/28$	$1/20$
Time	45.4ms	76.4ms	1.12s	43.8s

Table 1: Exact local RLCTs λ_L of $H_L(\theta)$ for $L \in [4]$, with wall-clock time.

The learning coefficient shrinks as depth grows.

Table of Contents

Conclusion

Recap:

- We showed how RLCTs appear naturally in the context of model selection and singular learning theory;

Conclusion

Recap:

- We showed how RLCTs appear naturally in the context of model selection and singular learning theory;
- We gave an algorithm which computes exactly the RLCT of bivariate polynomials with rational coefficients, and terminates in at most $d_X(d_Y + \frac{1}{2})$ steps.

Conclusion

Recap:

- We showed how RLCTs appear naturally in the context of model selection and singular learning theory;
- We gave an algorithm which computes exactly the RLCT of bivariate polynomials with rational coefficients, and terminates in at most $d_X(d_Y + \frac{1}{2})$ steps.

Perspectives:

Conclusion

Recap:

- We showed how RLCTs appear naturally in the context of model selection and singular learning theory;
- We gave an algorithm which computes exactly the RLCT of bivariate polynomials with rational coefficients, and terminates in at most $d_X(d_Y + \frac{1}{2})$ steps.

Perspectives:

- Thorough comparison of exact computation methods and estimation methods;

Conclusion

Recap:

- We showed how RLCTs appear naturally in the context of model selection and singular learning theory;
- We gave an algorithm which computes exactly the RLCT of bivariate polynomials with rational coefficients, and terminates in at most $d_X(d_Y + \frac{1}{2})$ steps.

Perspectives:

- Thorough comparison of exact computation methods and estimation methods;
- Generalization of the Newton-Puiseux to higher dimensions:

Conclusion

Recap:

- We showed how RLCTs appear naturally in the context of model selection and singular learning theory;
- We gave an algorithm which computes exactly the RLCT of bivariate polynomials with rational coefficients, and terminates in at most $d_X(d_Y + \frac{1}{2})$ steps.

Perspectives:

- Thorough comparison of exact computation methods and estimation methods;
- Generalization of the Newton-Puiseux to higher dimensions:
  - for quasi-ordinary singularities;

Conclusion

Recap:

- We showed how RLCTs appear naturally in the context of model selection and singular learning theory;
- We gave an algorithm which computes exactly the RLCT of bivariate polynomials with rational coefficients, and terminates in at most $d_X(d_Y + \frac{1}{2})$ steps.

Perspectives:

- Thorough comparison of exact computation methods and estimation methods;
- Generalization of the Newton-Puiseux to higher dimensions:
  - for quasi-ordinary singularities;
  - for general polynomials following [Collins, Green, Pramanik'11]

Thank You!