Condition Numbers for the Cube.
I: Univariate Polynomials and Hypersurfaces

Josué TONELLI-CUETO (Inria Paris & IMJ-PRG)
together with
Elias TSIGARIDAS (Inria Paris & IMJ-PRG)

June 30, 2020

OURAGAN
Seminar

Slides at *https://tonellicueto.xyz/pdf/OURAGAN30062020_slides.pdf*

This presentation is about the accepted paper

Condition Numbers for the Cube.
I: Univariate Polynomials and Hypersurfaces

authored by

- Elias Tsigaridas (Inria Paris & IMJ-PRG), and
- Josué Tonelli-Cueto (Inria Paris & IMJ-PRG)

The authors were partially supported by

- ANR JCJC GALOP (ANR-17-CE40-0009),
- the PGMO grant ALMA, and
- the PHC GRAPE.

Complexity of numerical algorithms

Numerical algorithms

What do characterize numerical algorithms?

- Inexact input data
- Approximate operations with numbers

Which problems arise when working with numerical algorithms?

- Behaviour is not uniform
- Some inputs (*ill-posed*) are intractable

Why do we want numerical algorithms?

- More stable, i.e., robust with respect errors
- They can be faster in practice

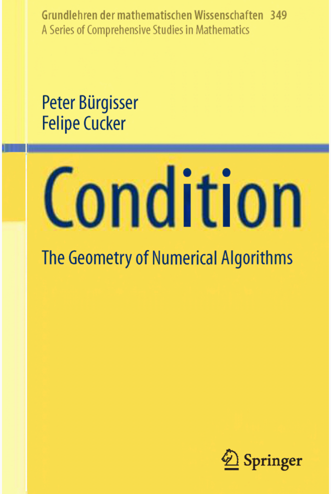
Complexity: Condition numbers I

ALL INPUTS ARE EQUAL
BUT SOME INPUTS ARE MORE EQUAL
THAN OTHERS

Condition number

- Measure of the numerical sensitivity
  - The bigger the worse!
  - It depends on the metric!
- Controls the complexity. This is what happens in:
  - Linear algebra
  - Linear programming and optimization
  - Algebraic geometry

Details in the Book!



...and some other papers!

Uniform complexity of numerical algorithms I

Worst-case complexity analysis:

What is the worst running time?

Average complexity analysis:

What is the expectation of the running time on a random input?

Smoothed complexity analysis: (Spielman, Teng; 2002)

What is the worst running time after perturbing the input with a random perturbation (with weight σ)?

Smoothed lies between worst-case and average complexity

- $\sigma \to 0$: We recover worst-case complexity
- $\sigma \to \infty$: We recover average analysis

Uniform complexity of numerical algorithms II

Worst-case complexity analysis:

Infinite for numerical algorithms!

Average complexity analysis: (Goldstein & von Neumann, Demmel, Smale)

It allows to derive complexity estimates that do not depend on the condition number

Smoothed complexity analysis:

Explains the success of numerical algorithms in practice

The long-term goal

Algorithms are faster and simpler on the cube,
but geometry is easier on the sphere!

Example:
Covering the cube efficiently is easy,
but covering the sphere is not so easy.

Cubes are better for subdivisions!

$$\begin{array}{rcll} \text{Geometry on the sphere} & = & \text{Euclidean norm} & \|x\| := \sqrt{\sum_i |x_i|^2} \\ \text{Geometry on the cube} & = & \infty\text{-norm} & \|x\|_\infty := \max_i |x_i| \end{array}$$

Goal:

$$\begin{array}{rcl} \text{Geometry on the sphere} & \to & \text{Geometry on the cube} \\ \text{Euclidean norm} & \to & \infty\text{-norm} \end{array}$$

Warning: The ∞-norm does not come from an inner product!

Hopes:

- Better complexity estimates
- Faster algorithms
- Better understanding of subdivision methods

Antecedent exploring other norms: (Cucker, Ergür, T.C.; SIAM AG'19)

Our local achievement

- Condition theory for hypersurfaces in the cube
- Gaussian polynomials
- Polynomials with restricted support (up to assumptions)

We showcase our results with:

- Separation bounds for roots of univariate polynomials in $(0, 1)$
- Plantinga-Vegter algorithm

Let’s see some details!

Polynomial inequalities and condition

Some notation

$$
\begin{aligned}
\mathcal{P}_{n,d} &: \text{Polynomials of degree} \leq d \text{ in the variables } X_1, \ldots, X_n \\
B_n &: \text{Euclidean ball in } \mathbb{R}^n \\
I^n &: \text{Unit } \infty\text{-ball } ([-1,1]^n) \text{ in } \mathbb{R}^n
\end{aligned}
$$

$f = \sum_\alpha f_\alpha X^\alpha \in \mathcal{P}_{n,d}$, $x \in \mathbb{R}^n$

$$
\begin{aligned}
\|f\|_W &: \text{Weyl norm, given by } \sqrt{\textstyle\sum_\alpha \binom{d}{\alpha, d-|\alpha|}^{-1/2} f_\alpha} \\
\|f\|_1 &: \text{1-norm, given by } \textstyle\sum_\alpha |f_\alpha| \\
f(x) &: \text{Evaluation of } f \text{ at } x \\
\nabla f &: \text{Formal gradient of } f\text{, element of } \mathcal{P}_{n,d-1}^n \\
\nabla_x f &: \text{Gradient vector of } f \text{ at } x
\end{aligned}
$$

Idea: Controlling size of evaluation

Proposition

Let $f \in \mathcal{P}_{n,d}$ *and* $x \in B_n$. *Then* $|f(x)| \leq \|f\|_W \|(1,x)\|^d$.

Proof.

$$
\begin{aligned}
|f(x)| &= \left| \left\langle \left(\binom{d}{\alpha, d-|\alpha|}^{-1/2} f_\alpha \right), \left(\binom{d}{\alpha, d-|\alpha|}^{1/2} x_\alpha \right) \right\rangle \right| \\
&\leq \left\| \left(\binom{d}{\alpha, d-|\alpha|}^{-1/2} f_\alpha \right) \right\| \left\| \left(\binom{d}{\alpha, d-|\alpha|}^{1/2} x_\alpha \right) \right\| \\
&= \|f\|_W \sqrt{\sum_\alpha \binom{d}{\alpha, d-|\alpha|} x^{2\alpha}} \\
&= \|f\|_W \sqrt{(1 + \sum_i x_i^2)^d} \\
&= \|f\|_W \|(1,x)\|^d
\end{aligned}
$$

$\square$

Idea: Controlling size of evaluation

Proposition
Let $f \in \mathcal{P}_{n,d}$, $x \in B_{q,n}$ and $p, q \geq 1$ such that $1/p + 1/q = 1$. Then $|f(x)| \leq \|f\|_{W,p} \|(1,x)\|_q^d$.

Proof.

$$\begin{aligned}
|f(x)| &= \left| \left\langle \left(\binom{d}{\alpha, d-|\alpha|}^{1/p-1} f_\alpha \right), \left(\binom{d}{\alpha, d-|\alpha|}^{1/q} x_\alpha \right) \right\rangle \right| \\
&\leq \left\| \left(\binom{d}{\alpha, d-|\alpha|}^{1/p-1} f_\alpha \right) \right\|_p \left\| \left(\binom{d}{\alpha, d-|\alpha|}^{1/q} x_\alpha \right) \right\|_q \\
&= \|f\|_{W,p} \sqrt[q]{\sum_\alpha \binom{d}{\alpha, d-|\alpha|} x^{q\alpha}} \\
&= \|f\|_{W,p} \sqrt[q]{(1 + \sum_i x_i^q)^d} \\
&= \|f\|_{W,p} \|(1,x)\|_q^d
\end{aligned}$$

$\square$

Idea: Controlling size of evaluation

Taking $p = 1$ and $q = \infty$...

Proposition
Let $f \in \mathcal{P}_{n,d}$, $x \in I^n$. Then $|f(x)| \leq \|f\|_1$.

This, by duality, justifies our use of the 1-norm for polynomials when we use the ∞-norm for points.

In a similar way...

$f \in \mathcal{P}_{n,d}$, $x \in I^n$, $v \in \mathbb{R}^n$

- Control of the derivative I:

$$\|\langle \nabla f, v\rangle\|_1 \leq d\|f\|_1\|v\|_\infty$$

- Control of the derivative II:

$$\|\nabla_x f\|_1 \leq d\|f\|_1$$

- Lipschitz properties for f and its derivatives

Local condition number

Definition (T.C., Tsigaridas; ISSAC'20)

Let $f \in \mathcal{P}_{n,d}$ and $x \in I^n$, the *local condition number of f at x* is the quantity

$$C(f,x) := \frac{\|f\|_1}{\max\left\{|f(x)|, \frac{1}{d}\|\nabla_x f\|_1\right\}}.$$

Important observation: $C(f,x) = \infty$ iff x is a singular zero of f

Properties of the local condition number

- Regularity inequality

  $$\text{either } |f(x)|/\|f\|_1 \geq 1/\,C(f,x) \text{ or } \|\nabla_x f\|_1/(d\|f\|_1) \geq 1/\,C(f,x).$$

- 1st Lipschitz property

  $$f \mapsto \|f\|_1/\,C(f,x) \text{ is 1-Lipschitz}$$

- 2nd Lipschitz property

  $$I^n \ni x \mapsto 1/\,C(f,x) \text{ is } d\text{-Lipschitz}$$

- Condition Number Theorem

  $$\|f\|_1/\mathrm{dist}_1(f,\Sigma_x) \leq C(f,x) \leq 2d\,\|f\|_1/\mathrm{dist}_1(f,\Sigma_x)$$

  where $\Sigma_x := \{g \in \mathcal{P}_{n,d} \mid x \text{ is a singular zero of } f\}$
- Higher Derivative Estimate. If $C(f,x)f(x)/\|f\|_1 < 1$, then

  $$\gamma(f,x) \leq \frac{1}{2}(d-1)\sqrt{n}\,C(f,x).$$

  where $\gamma(f,x)$ is Smale's γ

All we need for complexity analyses!

Application 1:
Separation of roots

Separation of roots

Recall...

$$\Delta_\alpha(f) := \mathrm{dist}(\alpha, f^{-1}(0) \setminus \{\alpha\})$$

Theorem (T.C., Tsigaridas; ISSAC'20)
Let $f \in \mathcal{P}_{1,d}$. Then, for every complex $\alpha \in f^{-1}(0)$ such that $\mathrm{dist}(\alpha, I) \leq 1/(3(d-1)\,C(f))$,

$$\Delta_\alpha(f) \geq \frac{1}{16(d-1)\,C(f)}$$

where

$$C(f) := \sup_{x \in I} C(f, x).$$

I.e., the condition number controls the separation of the roots

Probabilistic results

Randomness model I: Two properties

(SG) We call a random variable $\mathfrak{x}$ *subgaussian*, if there exist a $K > 0$ such that for all $t \geq K$,

$$\mathbb{P}(|\mathfrak{x}| > t) \leq 2\exp(-t^2/K^2).$$

The smallest such K is the *subgaussian constant* of $\mathfrak{x}$.

(AC) A random variable $\mathfrak{x}$ has the *anti-concentration property*, if there exists a $\rho > 0$, such that for all $\varepsilon > 0$,

$$\max\{\mathbb{P}\left(|\mathfrak{x} - u| \leq \varepsilon\right) \mid u \in \mathbb{R}\} \leq 2\rho\varepsilon.$$

The smallest such ρ is the anti-concentration constant of $\mathfrak{x}$.

Randomness model II: Zintzo random polynomials I

Definition

Let $M \subseteq \mathbb{N}^n$ be a finite set such that $0, e_1, \ldots, e_n \in M$. A *zintzo random polynomial supported on M* is a random polynomial

$$\mathfrak{f} = \sum_{\alpha \in M} \mathfrak{f}_\alpha X^\alpha \in \mathcal{P}_{n,d}$$

such that the coefficients $\mathfrak{f}_\alpha$ are independent subgaussian random variables with the anti-concentration property.

Note: ‘zintzo’, from Basuqe, means honest, upright, righteous.

Observation: No scaling in the coefficients, as it happens with dobro random polynomials (Cucker, Ergür, TC; ISSAC’19)

Randomness model II: Zintzo random polynomials II

For $\mathfrak{f}$ a zintzo random polynomial, we define:

1. the *subgaussian constant* of $\mathfrak{f}$ which is given by
$$K_{\mathfrak{f}} := \sum\nolimits_{\alpha \in M} K_\alpha, \tag{5.1}$$
where K_α is the subgaussian constant of $\mathfrak{f}_\alpha$, and
2. the *anti-concentration constants* of $\mathfrak{f}$ which is given by
$$\rho_{\mathfrak{f}} := \sqrt[n+1]{\rho_0 \rho_{e_1} \cdots \rho_{e_n}}, \tag{5.2}$$
where ρ_0 is the anti-concentration constant of $\mathfrak{f}_0$ and for each i, ρ_{e_i} is the anti-concentration constant of $\mathfrak{f}_{e_i}$.

$K_{\mathfrak{f}}$ and $\rho_{\mathfrak{f}}$ will control the complexity estimates

Randomness model II: Zintzo random polynomials III

Let $M \subseteq \mathbb{N}^n$ be such that it contains $0, e_1, \ldots, e_n$. These are two important cases of zintzo random polynomials:

G A *Gaussian polynomial supported on M* is a zintzo random polynomial $\mathfrak{f}$ supported on M, the coefficients of which are i.i.d. Gaussian random variables.
In this case, $\rho_{\mathfrak{f}} = 1/\sqrt{2\pi}$ and $K_{\mathfrak{f}} \leq |M|$.

U A *uniform random polynomial supported on M* is a zintzo random polynomial $\mathfrak{f}$ supported on M, the coefficients of which are i.i.d. uniform random variables on $[-1, 1]$.
In this case, $\rho_{\mathfrak{f}} = 1/2$ and $K_{\mathfrak{f}} \leq |M|$.

Randomness model III: Smoothed case

Proposition

Let $\mathfrak{f}$ be a zintzo random polynomial supported on M, $f \in \mathcal{P}_{n,d}$ a polynomial supported on M, and $\sigma > 0$. Then,

$$\mathfrak{f}_\sigma := f + \sigma \|f\|_1 \mathfrak{f}$$

is a zintzo random polynomial supported on M such that

$$K_{\mathfrak{f}_\sigma} \leq \|f\|_1 (1 + \sigma K_{\mathfrak{f}}) \text{ and } \rho_{\mathfrak{f}_\sigma} \leq \rho_{\mathfrak{f}} / (\sigma \|f\|_1).$$

In particular,

$$K_{\mathfrak{f}_\sigma} \rho_{\mathfrak{f}_\sigma} = (K_{\mathfrak{f}} + 1/\sigma) \rho_{\mathfrak{f}}.$$

I.e., the smoothed case is included in our average case!

Probabilistic bound

Theorem (T.C., Tsigaridas; ISSAC'20)

Let $\mathfrak{f} \in \mathcal{P}_{n,d}$ a zintzo random polynomial supported on M. Then for all $t \geq e$,

$$\mathbb{P}(C(\mathfrak{f},x) \geq t) \leq \sqrt{n} d^n |M| \left(8K_{\mathfrak{f}}\rho_{\mathfrak{f}}\right)^{n+1} \frac{\ln^{\frac{n+1}{2}} t}{t^{n+1}}.$$

Corollary (T.C., Tsigaridas; ISSAC'20)

Let $\mathfrak{f} \in \mathcal{P}_{n,d}$ be a zintzo random polynomial supported on M. Then, for all $t > 2e$,

$$\mathbb{P}(C(\mathfrak{f}) \geq t) \leq \frac{1}{4}\sqrt{n} d^{2n} |M| \left(64K_{\mathfrak{f}}\rho_{\mathfrak{f}}\right)^{n+1} \frac{\ln^{\frac{n+1}{2}} t}{t}.$$

Application 2:
Plantinga-Vegter algorithm

Setting

What do we have?

- An implicit curve C inside $[-1,1]^2$ given by a C^1 function $f : [-1,1]^2 \to \mathbb{R}$
- Interval approximations $\square f$ of f and $\square \nabla f$ of ∇f

What do we want?

- Piecewise-linear approximation L of C in $[-1,1]^2$such that $([-1,1]^2, C)$ and $([-1,1]^2, L)$ are isotopic

Any assumptions?

- C smooth
- C Intersects the boundary of $[-1,1]^2$ transversely

Plantinga-Vegter algorithm for curves I

Algorithm: PV Algorithm for curves (Plantinga, Vegter; 2004)

Input: $f : \mathbb{R}^2 \to \mathbb{R}$
with interval approximations $\square[f]$ and $\langle \square[\nabla f], \square[\nabla f] \rangle$

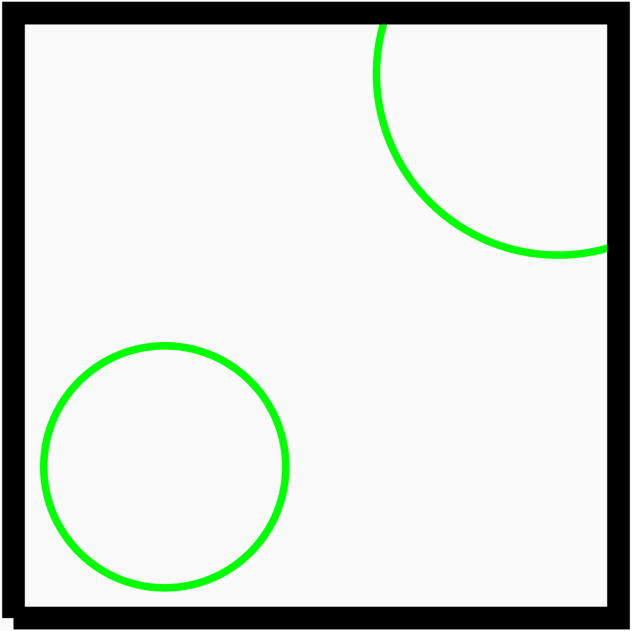
SUBDIVISION:
Starting with the trivial subdivision $\mathcal{S} := \{[-1, 1]^n\}$, repeatedly subdivide each $J \in \mathcal{S}$ into 4 squares until for all $J \in \mathcal{S}$,

$$0 \notin \square f(J) \text{ or } 0 \notin \langle \square \nabla f(J), \square \nabla f(J) \rangle$$

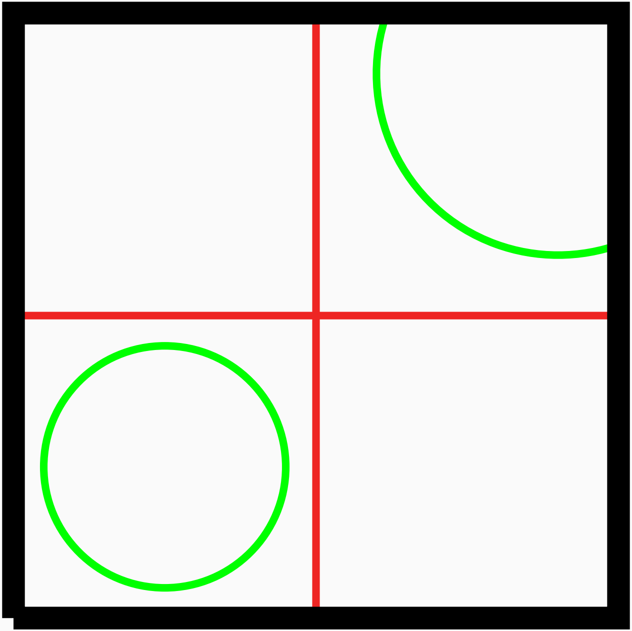
CONSTRUCTION:
Construct piecewise-linear curve L
joining the midpoints of “small” edges of each $J \in \mathcal{S}$ with oposite f-signs at their vertices

Output: Piecewise-linear approximation L of $C = f^{-1}(0) \cap [-a, a]^2$ isotopic to it

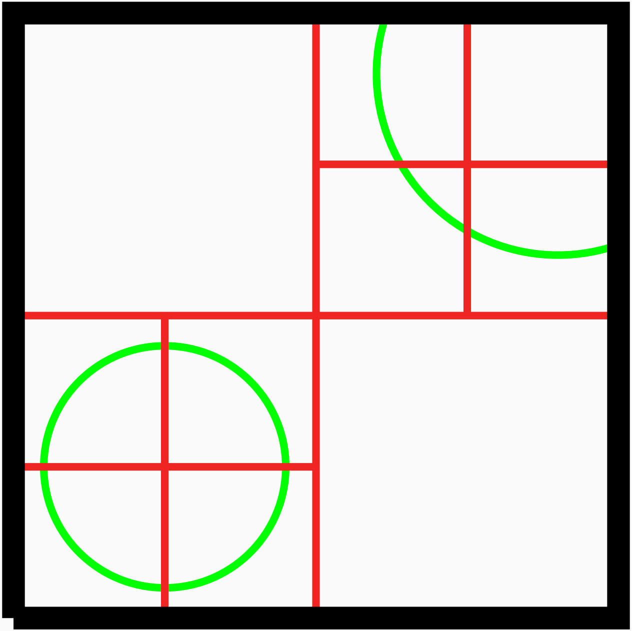
Plantinga-Vegter algorithm for curves II: Example



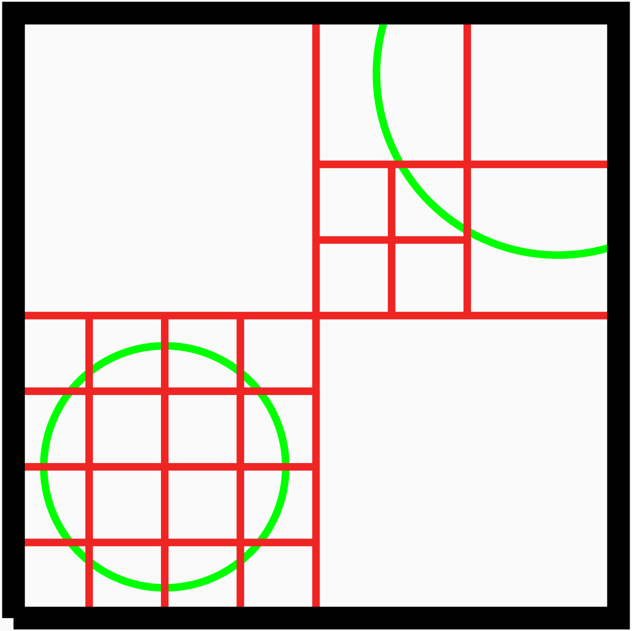
Plantinga-Vegter algorithm for curves II: Example



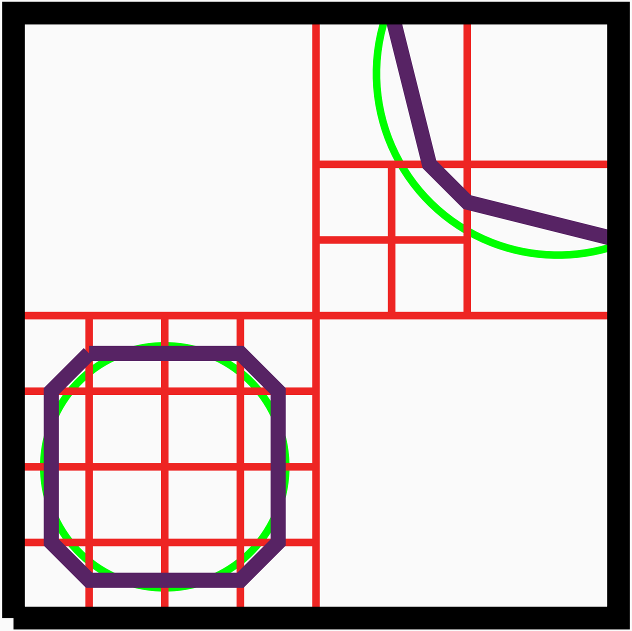
Plantinga-Vegter algorithm for curves II: Example



Plantinga-Vegter algorithm for curves II: Example



Plantinga-Vegter algorithm for curves II: Example



Plantinga-Vegter algorithm in higher dimensions

1. Plantinga-Vegter algorithm can be generalized to produce isotopic approximations of surfaces (Plantinga, Vegter; 2004)
   This is really why is called Plantinga-Vegter!
   Very efficient in practice
2. The subdivision method can be generalized to higher dimensions (Burr, Gao, Tsigaridas; ISSAC2017)

We will focus on the later, since...

complexity of the algorithm is mainly that of the subdivision part

We will mainly count the number of subdivisions, because...

$$\text{cost(subdivision algorithm)} \sim \#\,\text{(subdivisions)} \cdot \text{cost(evaluations)}$$

Subdivision in Plantinga-Vegter algorithm

Algorithm: Subdivision of PV Algorithm (Burr, Gao, Tsigaridas; ISSAC'17)

Input: $f : \mathbb{R}^n \to \mathbb{R}$
with interval approximations $\square[hf]$ and $\square[h'\nabla f]$
for some functions $h, h' : \mathbb{R}^n \to (0, \infty)$

Starting with the trivial subdivision $\mathcal{S} := \{[-a, a]^n\}$, repeatedly subdivide each $J \in \mathcal{S}$ into 2^n cubes until the condition

$$C_f(J) : 0 \notin \square[hf](J) \text{ or } 0 \notin \langle \square[h'\nabla f], \square[h'\nabla f] \rangle$$

holds for all $J \in \mathcal{S}$

Output: Subdivision $\mathcal{S} \subseteq \mathcal{I}_n$ of $[-a, a]^n$
such that for all $J \in \mathcal{S}$, $C_f(J)$ is true

h, h' depend on the setting and the interval arithmetic one uses

The complexity estimate

We had...

Theorem (Cucker, Ergür, T.C.; ISSAC'19)
Let $\mathfrak{f} \in \mathcal{P}_{n,d}$ be a dobro random polynomial with parameters K and ρ. The average number of boxes of the final subdivision of PV *algorithm on input $\mathfrak{f}$ is at most*

$$d^{\frac{n^2+3n}{2}} 2^{\frac{n^2+16n\log(n)}{2}} (c_1 c_2 K \rho)^{n+1}.$$

We get...

Theorem (T.C., Tsigaridas; ISSAC'20)
Let $\mathfrak{f} \in \mathcal{P}_{n,d}$ be a zintzo random polynomial supported on M. The average number of boxes of the final subdivision of PV *algorithm on input $\mathfrak{f}$ is at most*

$$n^{\frac{3}{2}} d^{2n} |M| \left(80\sqrt{n(n+1)} K_{\mathfrak{f}} \rho_{\mathfrak{f}}\right)^{n+1}.$$

An specific bound

Corollary (T.C., Tsigaridas; ISSAC'20)
Let $\mathfrak{f} \in \mathcal{P}_{n,d}$ be a random polynomial supported on M. The average number of boxes of the final subdivision of PV *algorithm on input $\mathfrak{f}$ is at most*

$$n^{\frac{3}{2}} \left(40\sqrt{n(n+1)}\right)^{n+1} d^{2n}|M|^{n+2}$$

if $\mathfrak{f}$ is Gaussian or uniform.

Bere arretagatik eskerrik asko!
Merci pour votre attention!

Galderak?
Des questions?