

Isolation of real roots of polynomial systems, complexity and condition number

B. Mourrain

GALAAD, INRIA Méditerranée, Sophia Antipolis

`mourrain@sophia.inria.fr`

October 24, 2009

A general scheme

Algorithm (A generic subdivision algorithm)

INPUT: *An algebraic description of a semi-algebraic set.*

OUTPUT: *A topological description of the semi-algebraic set.*

Create a subdivision tree \mathcal{T} and set its root to B_0 .

Create a list of cells \mathcal{C} and initialize it with $[B_0]$.

While $\mathcal{C} \neq \emptyset$

- $c = \text{pop } \mathcal{C}$
- **If** **regular**(c) $\mathcal{T} \leftarrow \text{process}(c)$ **else** $\mathcal{C} \leftarrow \text{subdivide}(c)$

return **assemble** (\mathcal{T})

👉 **The problem:** *Given a system of polynomial equations with real (rational, integer) coefficients, isolate (approximate within a given precision ε) the real roots of the system in a domain $D \subset \mathbb{R}^n$.*

👉 **Regularity:** we will use

- an **exclusion** test to remove cells with no root;
- an **inclusion** test to check if the cell contains a unique root.

👉 **Analysis** will be performed in terms of

- d maximal degree of the equations;
- τ maximal size of the coefficients.
- intrinsic quantities of the system not necessarily computed by the algorithm.

How hard is the isolation problem?

Theorem (Separation bound)

$$\Delta = \text{sep}(A) = \min_{i \neq j} |\gamma_i - \gamma_j| \sim 2^{-\mathcal{O}(d^2 + d\tau)}$$

Example: Consider the Wilkinson polynomial

$$A = (x - 1)(x - 2) \cdots (x - 20)$$

- Lower bound:

$$\Delta \geq 10^{-344}$$

- but actually

$$\text{sep}(A) = 1$$

Not all can be bad!

Theorem (Separation bound)

$$\Delta = \text{sep}(A) = \min_{i \neq j} |\gamma_i - \gamma_j| \sim 2^{-\mathcal{O}(d^2 + d\tau)}$$

$$\Delta_j := \min_{k \neq j} \text{dist}(\zeta_j, \zeta_k)$$

Theorem (DMM₁)

$$\prod \Delta_j = \prod_j |\gamma_j - \gamma_{c_j}| \sim 2^{-\mathcal{O}(d^2 + d\tau)}$$

where γ_{c_j} is the closest root to γ_j [Davenport; 1985].

Not all can be bad, in dimension n

Theorem (Separation bound)

$$\Delta = \text{sep}(A) = \min_{i \neq j} |\gamma_i - \gamma_j| \sim 2^{-\mathcal{O}(nd^{2n-1}\tau)}$$

Theorem (DMM_n [EMT'09])

$$\prod \Delta_j = \prod_j |\gamma_j - \gamma_{c_j}| \sim 2^{-\mathcal{O}(nd^{2n-1}\tau)}$$

where γ_{c_j} is the closest root to γ_j .

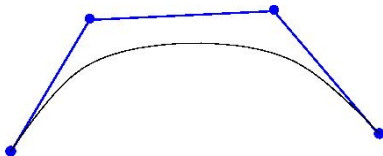
Univariate polynomials

Univariate Bernstein representation

For any $f(x) \in \mathbb{Q}[x]$ of degree d , with

$$f(x) = \sum_{i=0}^d c_i \binom{d}{i} (x-a)^i (b-x)^{d-i} (b-a)^{-d} = \sum_{i=0}^d c_i B_d^i(x; a, b),$$

The $\mathbf{c} = [c_i]_{i=0,\dots,d}$ are the *control coefficients* of f on $[a, b]$.



Properties:

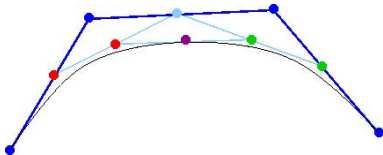
- $\sum_{i=0}^d B_d^i(x; a, b) = 1$; $\sum_{i=0}^d (a \frac{d-i}{d} + b \frac{i}{d}) B_d^i(x; a, b) = x$;
- $f(a) = c_0, f(b) = c_d$;
- $d f'(x) = \sum_{i=0}^{d-1} \Delta(\mathbf{c})_i B_{d-1}^i(x; a, b)$ where $\Delta(\mathbf{c})_i = c_{i+1} - c_i$;
- $(x, f(x))_{x \in [a, b]} \in \text{convex hull of the points } (a \frac{d-i}{d} + b \frac{i}{d}, c_i)_{i=0..d}$
- $\#\{f(x) = 0; x \in [a, b]\} = V(\mathbf{c}) - 2p, p \in \mathbb{N}$.

De Casteljau subdivision algorithm:

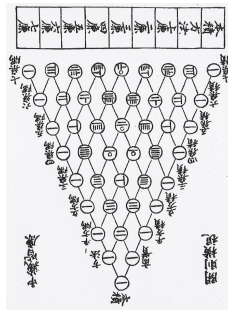
$$\begin{cases} c_i^0 = c_i, & i = 0, \dots, d, \\ c_i^r(t) = (1-t) c_i^{r-1}(t) + t c_{i+1}^{r-1}(t), & i = 0, \dots, d-r. \end{cases}$$

- $\mathbf{c}^-(t) = (c_0^0(t), c_0^1(t), \dots, c_0^d(t))$ represents f on $[a, (1-t)a + tb]$.
- $\mathbf{c}^+(t) = (c_0^d(t), c_1^{d-1}(t), \dots, c_d^0(t))$ represents f on $[(1-t)a + tb, b]$.

The geometric point of view.



The algebraic point of view.



Real root isolation for squarefree polynomials

□ **Regularity:**

- Count the number $V(\mathbf{c}; a, b)$ of coefficient sign changes.
- $V(\mathbf{c}; a, b) = 0 \Rightarrow$ no root.
- $V(\mathbf{c}; a, b) = 1 \Rightarrow$ a single root.

□ **Subdivision:**

If $V(\mathbf{c}) > 1$, split the interval in the middle using de Casteljau algorithm;

Continued Fraction solver [AC'76, ..., TE'08]

👉 Instead of changing the interval:

- Fix it: $]0, +\infty[$
- Change the fonction, by homography transformation:

$$\begin{aligned} H :]0, +\infty[&\rightarrow]\frac{a}{c}, \frac{b}{d}[\\ x &\mapsto \frac{a + bx}{c + dx} \end{aligned}$$

- Work with $(f \circ H, H)$

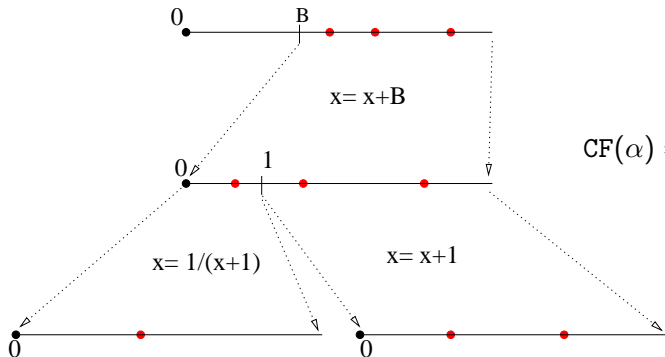
☐ **Regularity:**

- $V(f \circ H) = 0 \Rightarrow$ no root;
- $V(f \circ H) = 1 \Rightarrow$ a single root;

where $V(\cdot)$ is the number of sign changes of the coefficients in the monomial basis.

□ Subdivision:

- Compute a lower bound $b = L(f) \in \mathbb{N}$ of the roots of f in \mathbb{R}_+ ;
- Compute $f(x) := T_b(f) = f(x + n)$ and repeat until $L(f) = 0$;
- Split: $T_1(p) = p(x + 1)$, $R(p) = (x + 1)^d p(\frac{1}{x+1})$.



$$\text{CF}(\alpha) = \lfloor \alpha \rfloor + \frac{1}{\text{CF}\left(\frac{1}{\alpha - \lfloor \alpha \rfloor}\right)}$$

Continued Fraction expansion of the roots:

$$\alpha = b_0 + \frac{1}{b_1 + \frac{1}{b_2 + \dots}}$$

where b_i is the total shift between the i^{th} and $(i+1)^{\text{th}}$ inversions.

Theorem ([Vincent;1836], [Uspensky;1948], [Alesina,Galuzzi;1998])

Let $f \in \mathbb{Z}[x]$, and $b_0, b_1, \dots, b_n \in \mathbb{Z}_+$, $n > \mathcal{O}(d\tau)$. The map

$$x \mapsto b_0 + \frac{1}{b_1 + \frac{1}{\ddots b_n + \frac{1}{x}}}$$

transforms $f(x)$ to $\tilde{f}(x)$ such that

- ① $V(\tilde{f}) = 0 \Leftrightarrow f$ has no positive real roots.
- ② $V(\tilde{f}) = 1 \Leftrightarrow f$ has one positive real root.

$\Rightarrow 2^{\mathcal{O}(d\tau)}$ [Vincent; 1836], [Uspensky;1948] \dots ,
 $\mathcal{O}_B(d^5\tau^3)$ [Akritas;1980] \dots

Termination & Complexity

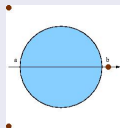
Proposition (Descartes' rule)

For $f := (\mathbf{c}, [a, b])$, $\#\{f(x) = 0; x \in [a, b]\} = V(\mathbf{c}) - 2p, p \in \mathbb{N}$.

Theorem

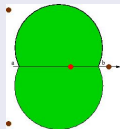
$$V(\mathbf{c}^-) + V(\mathbf{c}^+) \leq V(\mathbf{c}).$$

Theorem (Vincent)



If there is no complex root in the disc $D(m_{a,b}, \frac{|b-a|}{2}) \subset \mathbb{C}$, then $V(\mathbf{c}) = 0$.

Theorem (Two circles)



If there is no complex root in the union of the discs $D(T_{a,b}^+) \cup D(T_{a,b}^-) \subset \mathbb{C}$ except a simple real root, then $V(\mathbf{c}) = 1$.

Theorem (Mahler-Davenport-Mignotte)

Let $f \in \mathbb{Z}[x]$ (not necessarily square free),

$$\prod_{i=1}^k \Delta_k \geq \mathcal{M}(f)^{-d+1} d^{-\frac{d}{2}} \left(\frac{\sqrt{3}}{d}\right)^k.$$

Proposition

Let $f \in \mathbb{Z}[x]$ of degree d and coefficients of bit size $\leq \tau$, with simple roots. Then, the number of subdivisions to isolate its real roots is $\mathcal{O}(d\tau + d \log d)$.

Theorem ([ESY'06], [EMT'06])

Let $f \in \mathbb{Z}[x]$ of degree d and coefficients of bit size $\leq \tau$. The binary cost of the subdivision solver is $\tilde{\mathcal{O}}_B(d^4 \tau^2)$.

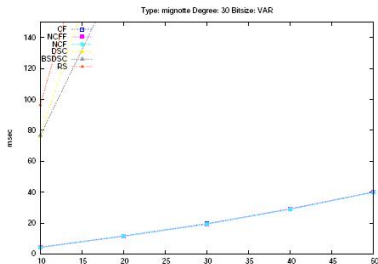
Average complexity [Tsigaridas, Emiris; 2008]

The **expected** complexity of **CF** is $\tilde{\mathcal{O}}_B(d^3 \tau)$.

Mignotte polynomials

- ▶ Separation is not known a priori
- ▶ Difficult for subdivision solvers
- ▶ Approximate methods failed
- ▶ Only CF is efficient

Figure: Mignotte polynomials



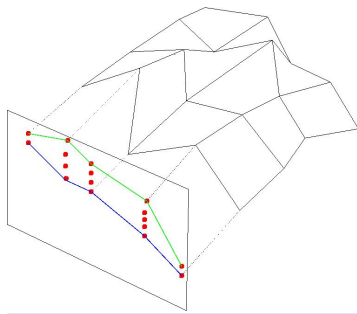
Multivariate polynomials

Multivariate Tensor product Bernstein representation

$$f(x_1, \dots, x_n) = \sum_{i_1=0}^{d_1} \cdots \sum_{i_n=0}^{d_n} c_{i_1, \dots, i_n} B_{d_1}^{i_1}(x_1; a_1, b_1) \cdots B_{d_n}^{i_n}(x_n; a_n, b_n)$$

associated with the box $\prod [a_i, b_i]$.

- **Subdivision** for each direction, similar to the univariate case.
- Arithmetic **complexity** of a subdivision bounded by $\mathcal{O}(d^{n+1})$ ($d = \max(d_1, \dots, d_n)$), memory space $\mathcal{O}(d^n)$.



$$m_j(f; x_j) = \sum_{i_j=0}^{d_j} \min_{\{0 \leq i_k \leq d_k, k \neq j\}} b_{i_1, \dots, i_n} B_{d_j}^{i_j}(x_j; a_j, b_j)$$

$$M_j(f; x_j) = \sum_{i_j=0}^{d_j} \max_{\{0 \leq i_k \leq d_k, k \neq j\}} b_{i_1, \dots, i_n} B_{d_j}^{i_j}(x_j; a_j, b_j).$$

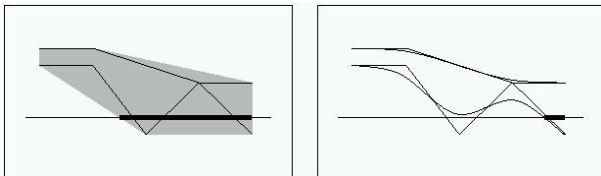
Proposition (PS93)

The intersection of the convex hull of the control polygon with the axis contains the projection of the zeroes of $\mathbf{f}(\mathbf{u}) = 0$.

Proposition

For any $\mathbf{u} = (u_1, \dots, u_n) \in \mathcal{D}$, and any $j = 1, \dots, n$, we have

$$m_j(f; u_j) \leq f(\mathbf{u}) \leq M_j(f; u_j).$$



Use the roots of $m_j(f, u_j) = 0$, $M_j(f, u_j) = 0$ to reduce the domain of search.

Multivariate Monomial Tensor Representation

Homography (or Möbius transformation)

Bijjective projective transformation $\mathcal{H} = (\mathcal{H}_1, \dots, \mathcal{H}_n)$ over $\mathbb{P}^1 \times \dots \times \mathbb{P}^1$,

$$x_k \mapsto \mathcal{H}_k(x_k) = \frac{\alpha_k x_k + \beta_k}{\gamma_k x_k + \delta_k}, \quad \alpha_k, \beta_k, \gamma_k, \delta_k \in \mathbb{Z}, \quad \alpha_k \delta_k - \beta_k \gamma_k \neq 0$$

$$H(f) := \prod_{k=1}^n (\gamma_k x_k + \delta_k)^{d_k} \cdot (f \circ \mathcal{H})(x)$$

Base homographies:

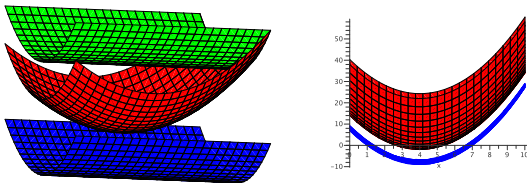
- translation by $c \in \mathbb{Z}$: $T_k^c(f) = f|_{x_k = x_k + c}$
- contraction by $c \in \mathbb{Z}$: $C_k^c(f) = f|_{x_k = cx_k}$
- reciprocal polynomial: $R_k(f) = x_k^{d_k} f|_{x_k = 1/x_k}$

Lemma

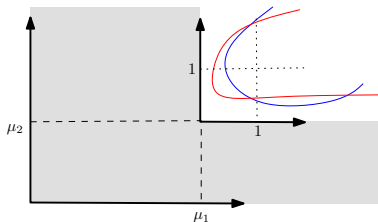
The group of homographies is generated by R_k, C_k^c, T_k^c , $k = 1, \dots, n$.

Reduction step

- Bounding the graph of f_i by cylinders in \mathbb{R}^{n+1} :



- Reducing the domain using univariate lower bounds:



$$m_k(f; x_k) = \sum_{i_k=0}^{d_k} \min_{i_1, \dots, \hat{i}_k, \dots, i_n} c_{i_1 \dots i_n} x_k^{i_k} \quad , \quad M_k(f; x_k) = \sum_{i_k=0}^{d_k} \max_{i_1, \dots, \hat{i}_k, \dots, i_n} c_{i_1 \dots i_n} x_k^{i_k}$$

Lemma

$$m_k(f; x_k) \leq \frac{f(x)}{\prod_{s \neq k} \sum_{i_s=0}^{d_s} x_s^{i_s}} \leq M_k(f; x_k) \quad , \quad k = 1, \dots, n$$

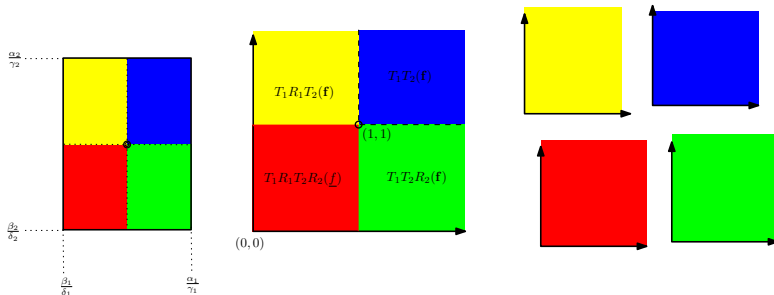
Corollary (lower bounds on the coordinates of the zeros)

$$\mu_k := \begin{cases} \text{min. pos. root of } M_k(f, x_k) & \text{if } M_k(f; 0) < 0 \\ \text{min. pos. root of } m_k(f, x_k) & \text{if } m_k(f; 0) > 0 \\ 0 & \text{otherwise} \end{cases}$$

All positive roots of f lie in $\mathbb{R}_{>\mu_1} \times \dots \times \mathbb{R}_{>\mu_n}$.

👉 Use the lowest root of $m_k(f_j, x_k)$ or $M_k(f_j, x_k)$ to reduce the domain.

Subdivision



Keep in memory:

- Transformed polynomials: $H(f_1), \dots, H(f_s)$ as coefficient *tensors*.
- $4n$ integers: $\alpha_k, \beta_k, \gamma_k, \delta_k$, $k = 1, \dots, n$ to keep track of the domain.

- No sign variation of the coefficients in the Bernstein/monomial basis \Rightarrow no real root in the domain \mathcal{D} .

or

- $|\mathbf{f}(\mathbf{m})| > |K_1(\mathbf{f})| |\mathcal{D}| \Rightarrow$ no root in \mathcal{D} ,
where \mathbf{m} is the center of \mathcal{D} and $K_1(\mathbf{f})$ is a bound on the Lipschitz constant of \mathbf{f} on \mathcal{D} .

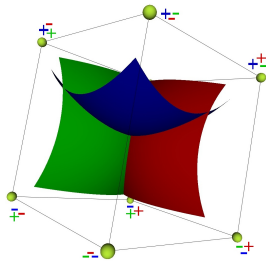
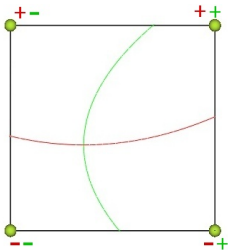
Inclusion criterion

Miranda Theorem

If for every pair of parallel faces there exists f_i that attains opposite signs on the faces, then f_1, \dots, f_n have at least one root inside the box.

Lemma

If the Jacobian has a constant sign in the box, then there is at most one root of f_1, \dots, f_n inside the box.



👉 or use α -theory [BCSS98]:

- $\beta := \beta(\mathbf{f}; \mathbf{x}) = \|D\mathbf{f}(\mathbf{x})^{-1}\mathbf{f}(\mathbf{x})\|$
- $\gamma := \gamma(\mathbf{f}; \mathbf{x}) = \sup_{k \geq 2} \left(\frac{1}{k!} \|D\mathbf{f}(\mathbf{x})^{-1} D^k \mathbf{f}(\mathbf{x}, y)\| \right)^{1/(k-1)}$
- $\alpha := \alpha(\mathbf{f}; \mathbf{x}) = \beta\gamma.$

Theorem

If $\alpha(\mathbf{f}; \mathbf{x}) < \alpha_0$ then

- \mathbf{x} is an approximate zero of \mathbf{f} ;
- Its associated zero ζ is in $B(\mathbf{x}; \frac{u_0}{\gamma(\mathbf{f}; \mathbf{x})})$;
- For any point $\mathbf{z} \in B(\mathbf{x}; \frac{u_0}{\gamma(\mathbf{f}; \mathbf{x})})$, Newton iteration converges quadratically from \mathbf{z} to ζ .

\Rightarrow Same root for all the points in a connected components of $\bigcup_{\alpha(\mathbf{f}; \mathbf{m}) < \alpha_0} B(\mathbf{m}; \frac{u_0}{\gamma(\mathbf{f}; \mathbf{m})})$.

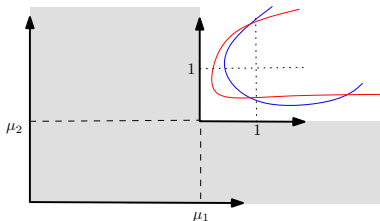
Subdivision speed

$\Delta_i(\zeta)$: local separation bound of ζ_i ,

$k_i(\zeta)$: # of steps that isolate ζ_i

- Continued fraction expansion:

$$\zeta_1 = b_0^{(1)} + \frac{1}{b_1^{(1)} + \frac{1}{b_2^{(1)} + \dots}} = \frac{P_{k_i(\zeta)}^{(1)}}{Q_{k_i(\zeta)}^{(1)}}$$



$$\left| \frac{P_{k_i(\zeta)}^{(1)}}{Q_{k_i(\zeta)}^{(1)}} - \zeta_j \right| < \phi^{-2k_i(\zeta)+1} \leq \Delta_i(\zeta),$$

- Bernstein binary subdivision:

$$|m_{k_i(\zeta)} - \zeta_i| < \sqrt{n} 2^{-k_i(\zeta)} |\mathcal{D}_0| \leq \Delta_i(\zeta),$$

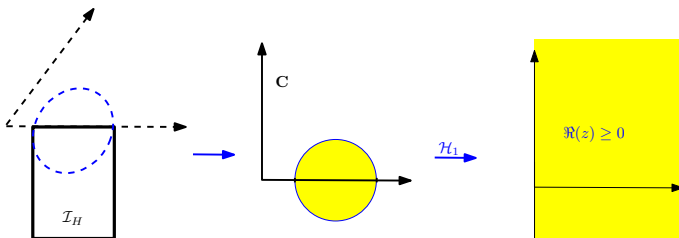
Complexity analysis

Vincent Theorem in several variables

Let $f(\mathbf{x}) = \sum_{i=0}^d c_i \mathbf{x}^i$ with $c_i \in \mathbb{R}$, without (complex) solutions s.t. $\Re(z_k) \geq 0$ for some k . Then all its coefficients c_i are of the same sign.

Corollary

If the complex multidisk associated to a domain \mathcal{I}_H does not intersect $\{z \in (\mathbb{P}^1)^n : f_i(z) = 0\}$ then the coeffs. of $H(f_i)$ have no sign changes.



Definition (ε -tubular neighborhood & “entropy”)

- $\tau_\varepsilon(f) = \{x \in \mathbb{R}^n : \exists z \in \mathbb{C}^n, f(z) = 0, \text{ s.t. } \|z - x\|_\infty < \varepsilon\}.$
- $\tau_\varepsilon(\mathbf{f}) := \cap_{i=1}^s \tau_\varepsilon(f_i)$ for $\mathbf{f} = (f_1, \dots, f_s).$
- $N_\varepsilon(\mathbf{f}) :=$ minimal number of boxes of size $< \varepsilon$ covering $\tau_\varepsilon(\mathbf{f})$ in a complete binary subdivision of $D_0.$

Proposition

The number of boxes of size ε not excluded is less than $N_\varepsilon(\mathbf{f}).$

Remark:

- $N_\varepsilon(\mathbf{f}) \leq \varepsilon^{-n} \text{Vol}(\tau_{2\varepsilon}(\mathbf{f})).$
- $N_\varepsilon(\mathbf{f})$ bounded for $\varepsilon > 0$: $N_*(\mathbf{f}) := \max_{\varepsilon > 0} N_\varepsilon(\mathbf{f}).$
- For a square system ($s = n$) with simple roots

$$\lim_{\varepsilon \rightarrow 0} N_\varepsilon(\mathbf{f}) \leq \lim_{\varepsilon \rightarrow 0} \varepsilon^{-n} \text{Vol}(\tau_{2\varepsilon}(\mathbf{f})) \leq c(n) \sum_{\zeta \in \mathcal{D}_0} \frac{\prod_i \|\nabla f_i(\zeta)\|}{|J_{\mathbf{f}}(\zeta)|}.$$

- By preconditionning $\mathbf{f}' := J_{\mathbf{f}}(\mathbf{m})^{-1} \mathbf{f}$, limit = $c(n) \sum_{\zeta \in \mathcal{D}} 1.$

For some $\rho > 0$, $\tau_\rho(\mathbf{f}) \subset \cup_{\zeta \in \mathcal{D}} B(\zeta, \frac{u_0}{\gamma(\mathbf{f}, \zeta)})$.

Definition (Lipshitz constant)

$$K_1(\mathbf{f}, \mathcal{D}) := \max(1, \frac{\text{Lipshitz constant}(\mathbf{f})}{\|\mathbf{f}\|}).$$

Definition (CKMW)

- $\kappa(\mathbf{f}, \mathbf{x}) := \frac{\|\mathbf{f}\|}{(\|\mathbf{f}\| \mu_{\mathbf{f}}(\mathbf{x})^{-2} + \|\mathbf{f}(\mathbf{x})\|_\infty)^{1/2}}$ where $\mu_{\mathbf{f}}(\mathbf{x}) = \|J_{\mathbf{f}}(\mathbf{x})\|$.
- $\kappa(\mathbf{f}) := \max_{\zeta \in \mathcal{D}; \mathbf{f}(\zeta)=0} \kappa(\mathbf{f}, \zeta)$.

Proposition

For $\varepsilon < \frac{cst(d)}{K_1(\mathbf{f}, \mathcal{D})^2 \kappa(\mathbf{f})^2}$, a retained box of size $\leq \varepsilon$ satisfies the inclusion test.

Proposition

The arithmetic complexity is $\tilde{O}(N_*(\mathbf{f}) d^{n+1} (\log \kappa(\mathbf{f}) + \log K_1(\mathbf{f})))$.

Complexity analysis for exact input over \mathbb{Z}

☞ **To simplify the complexity analysis, we assume that `exclude()` and `include()` test always give a correct answer.**

- Generalization of DMM bound [EMT'09]:

$$\prod_{\zeta \in V} \Delta_i(\zeta) \geq 2^{-2n\tau d^{2n-1} - d^{2n}/2} (nd^n)^{-nd^{2n}}$$

- Overall

$$\begin{aligned} \#STEPS &\leq n \sum_{\zeta \in V} k_i(\zeta) \leq n^2 \frac{1}{2} R - n^2 \frac{1}{2} \sum_{\zeta \in V} \lg \Delta_i(\zeta) \\ &\leq 2n^2 \tau d^{2n-1} + 2n^2 d^n \lg(nd^{2n}) \end{aligned}$$

Lemma

The number of reduction/subdivision steps is $\tilde{O}(n^2\tau d^{2n-1})$.

- Complexity of shifting ($\mathbf{x} = \mathbf{x} + \mathbf{u}$) [Gathen, Gerhard; 1997]: $\tilde{O}_B(n^2 d^n \tau + d^{n+1} n^3 \sigma)$, obtained as nd^{n-1} univariate shifts
- σ is bounding the bit size of partial quotients in the CF expansion of the roots: $E[\log b_i] = \mathcal{O}(\log \mathcal{K}) = \mathcal{O}(1)$.
- Bound computation with cost \mathcal{C}_1 ,
Tests evaluation with cost \mathcal{C}_2 .

Theorem

The total complexity is $\tilde{O}_B(2^n n^7 d^{5n-1} \tau^2 \sigma + (\mathcal{C}_1 + \mathcal{C}_2) n^2 \tau d^{n-1})$.

- Best rational approximation of the (coords. of the) real roots.
- Improvement by initial scaling: apply $C_k^{1/2^\ell}$ to the input.
 - The real roots are multiplied by 2^ℓ and their distance increases.
 - Total complexity improves by an order of $d^{2n\tau}$.
- $n = 1$: matches average complexity of [TE'08].
- **mCF** is implemented in MATHEMAGIX, in the C++ module `realroot`.
 - Uses GMP arithmetic to work with large integer coefficients.
 - Polynomials based on dense tensor (higher dimensional matrix) representation.
 - Univariate solving by classic CF algorithm, special case of **mCF**. DFS traversal of the subdivision tree returns only the (floor of the) first positive root.

MP Mourrain Bernard, Pavone Jean-Pascal:
Subdivision methods for solving polynomial equations Journal
of Symbolic Computation 44,3(2009) p. 292-306; (Preprint
version 2005).

CKMW Felipe Cucker, Teresa Krick, Gregorio Malajovich, Mario
Wschebor:
*A numerical algorithm for zero counting, I: Complexity and
accuracy.* J. Complexity 24(5-6): 582-605 (2008).

MMT Mantzaflaris Angelos, Mourrain Bernard, Tsigaridas Elias P.
*Continued Fraction Expansion of Real Roots of Polynomial
Systems*, In proc. of the conference on Symbolic-Numeric
Computation (2009) 85-94.