Solving Polynomial Equations in Smoothed Polynomial Time

Peter Bürgisser

Technische Universität Berlin

joint work with Felipe Cucker (CityU Hong Kong)

Workshop on Polynomial Equation Solving

Simons Institute for the Theory of Computing, October 16, 2014

Context and Motivation

- Solving polynomial equations is a fundamental mathematical problem, studied for several hundred years.
- The problem is NP-complete over the field $\mathbb{F}_2$ (equivalent to SAT).
- Traditionally, the problem is studied over $\mathbb{C}$. There, it is NP-complete in the model of Blum-Shub-Smale.
- Methods of symbolic computation (Gröbner bases etc) solve polynomial equations, but the running time is exponential. And these algorithms are also slow in practice.
- Numerical methods provide less information on the solutions, but perform much better in practice.
  - Theoretical explanation?

Smale's 17th Problem

- The 17th of Steve Smale's problems for the 21st century asks:

  Can a zero of n complex polynomial equations in n unknowns be found approximately, on the average, in polynomial time with a uniform algorithm?

- The problem has its origins in the series of papers "Complexity of Bezout's Theorem I-V" by Shub and Smale (1993-1996).
- Beltrán and Pardo (2008) answered Smale's 17th problem affirmatively, when allowing randomized algorithms.

Our Contributions

Near solution to Smale's 17th problem

We design a deterministic numerical algorithm for Smale's 17th problem with expected running time $N^{\mathcal{O}(\log \log N)}$, where N denotes input size.

For systems of bounded degree the expected running time is polynomial. E.g., $\mathcal{O}(N^2)$ for quadratic polynomials.

Smoothed analysis is a blend of average-case and worst-case analysis. It was proposed by Spielman and Teng (2001) and successfully applied to the simplex algorithm.

Smoothed polynomial time

We perform a smoothed analysis of the randomized algorithm of Beltrán and Pardo, proving that its smoothed expected running time is polynomial.

Setting

- For degree vector $d = (d_1, \ldots, d_n)$ define input space

  $\mathcal{H}_d := \{f = (f_1, \ldots, f_n) \mid f_i \in \mathbb{C}[X_0, \ldots, X_n] \text{ homogeneous of degree } d_i\}.$

  Input size $N := \dim_{\mathbb{C}} \mathcal{H}_d$.
- Output space is complex projective space $\mathbb{P}^n$: Look for zero $\zeta \in \mathbb{P}^n$ with $f(\zeta) = 0$.
- Metric d on $\mathbb{P}^n$ (angle).
- Fix unitary invariant hermitian inner product $\langle \ , \ \rangle$ on $\mathcal{H}_d$ (Weyl). This defines a norm $\|f\| := \langle f, f \rangle^{1/2}$ and an (angular) distance d on the projective space $\mathbb{P}(\mathcal{H}_d)$, respectively on the sphere $S(\mathcal{H}_d)$.
- Solution variety (smooth manifold)

$$V := \{(f, \zeta) \mid f(\zeta) = 0\} \subseteq \mathcal{H}_d \times \mathbb{P}^n.$$

Condition number

- Let $f(\zeta) = 0$. How much does ζ change when we perturb f a little?
- This can be quantified by the condition number of (f, ζ):

$$\mu(f, \zeta) := \|f\| \cdot \|M^\dagger\|,$$

where ($\|\zeta\| = 1$, $M^\dagger$ stands for pseudo-inverse)

$$M := \mathrm{diag}(\sqrt{d_1}, \ldots, \sqrt{d_n})^{-1} Df(\zeta) \in \mathbb{C}^{n \times (n+1)}.$$

- μ is well defined on $\mathbb{P}(\mathcal{H}_d) \times \mathbb{P}^n$: $\mu(tf, \zeta) = \mu(f, \zeta)$ for $t \in \mathbb{C}^*$.

Newton iteration and approximate zeros

- Projective Newton iteration

$$x_{k+1} = N_f(x_k)$$

  with Newton operator $N_f : \mathbb{P}^n \to \mathbb{P}^n$ and starting point x_0.
- Gamma Theorem (Smale): Put $D := \max_i d_i$. If

$$d(x_0, \zeta) \le \frac{0.3}{D^{3/2}\,\mu(f,\zeta)},$$

  then immediate convergence of $x_{k+1} = N_f(x_k)$ with quadratic speed:

$$d(x_k, \zeta) \le \frac{1}{2^{2^k - 1}}\, d(x_0, \zeta).$$

  Call x_0 approximate zero of f.

From local to global search: homotopy continuation

- Given a start system

$$(g,\zeta) \in V := \Big\{(f,\zeta) \mid f(\zeta) = 0\Big\} \subseteq \mathcal{H}_d \times \mathbb{P}^n.$$

  in the solution manifold V.
- Connect input $f \in \mathcal{H}_d$ to g by line segment $[g,f] = \{q_t \mid t \in [0,1]\}$.
- If none of the q_t has multiple zero, there exists unique lifting of $t \mapsto q_t$ to a solution path in V

$$\gamma\colon [0,1] \to V, t \mapsto (q_t, \zeta_t)$$

  such that $(q_0, \zeta_0) = (g, \zeta)$.

Adaptive linear homotopy

- Adaptive Linear Homotopy ALH: follow solution path γ numerically. Put $t_0 = 0, q_0 := g, z_0 := \zeta$.
  Compute $t_{i+1}, q_{i+1}, z_{i+1}$ adaptively from t_i, $q_i := q_{t_i}, z_i$ by Newton's method:

$$\begin{aligned} d(q_{i+1}, q_i) &= \frac{7.5 \cdot 10^{-3}}{D^{3/2}\mu(q_i, z_i)^2}, \\ z_{i+1} &= N_{q_{i+1}}(z_i). \end{aligned}$$

- Let $K(f, g, \zeta)$ denote the number k of Newton continuation steps needed to follow the homotopy.
- Shub-Smale & Shub (2007): z_i is approximate zero of ζ_{t_i} and

$$K(f, g, \zeta) \ \leq \ 217\, D^{3/2} \int_0^1 \mu(\gamma(t))^2 \, \|\dot{q}_t\| \, dt.$$

Randomization

- How to choose the start system?
- Almost all $(g, \zeta) \in V$ are "good": $\mu(g, \zeta) = N^{\mathcal{O}(1)}$ (Shub-Smale).
- Unknown how to efficiently construct such (g, ζ):
  "problem to find hay in a haystack."

- We may choose $g \in S(\mathcal{H}_d)$ uniformly at random.
- Alternatively, we may choose g according to the standard Gaussian distribution on $\mathcal{H}_d$: it has the density

$$\rho(g) = (2\pi)^{-N} \exp(-\frac{1}{2}\|g\|^2).$$

A Las Vegas algorithm

- Standard distribution on solution variety V:
  - choose $g \in \mathcal{H}_d$ from standard Gaussian,
  - choose one of the $d_1 \cdots d_n$ many zeros ζ of g uniformly at random.
- Efficient sampling of $(g, \zeta) \in V$ is possible (Beltrán & Pardo 2008).

- Las Vegas algorithm LV: on input f, draw $(g, \zeta) \in V$ at random, run ALH on (f, g, ζ)
- LV has expected "running time" $K(f) := \mathbb{E}_{g,\zeta} K(f, g, \zeta)$.

Average of LV (Beltrán and Pardo)

$$\mathbb{E}_f K(f) \; = \; \mathcal{O}\big(D^{3/2} N n\big)$$

for standard Gaussian $f \in \mathcal{H}_d$.

Smoothed expected polynomial time

Smoothed analysis: Fix $\overline{f} \in \mathcal{H}_d$ and $\sigma > 0$. The isotropic Gaussian on $\mathcal{H}_d$ with mean $\overline{f}$ and variance σ^2 has the density

$$\rho(f) = \frac{1}{(2\pi\sigma^2)^N} \exp\Big(-\frac{1}{2\sigma^2}\|f-\overline{f}\|^2\Big).$$

We write $f \sim N(\overline{f}, \sigma^2 I)$.
Technical issue: we truncate this Gaussian by requiring $\|f-\overline{f}\| \le \sqrt{2N}$, obtaining the distribution $N_T(\overline{f}, \sigma^2 I)$.

Smoothed analysis of LV

$$\sup_{\|\overline{f}\|=1} \mathbb{E}_{f\sim N_T(\overline{f},\sigma^2 I)} K(f) \;=\; \mathcal{O}\Big(\frac{D^{3/2}Nn}{\sigma}\Big).$$

Near solution to Smale's 17th problem

The deterministic algorithm below computes an approximate zero of $f \in \mathcal{H}_d$ with an expected number of arithmetic operations $N^{\mathcal{O}(\log\log N)}$, for standard Gaussian input $f \in \mathcal{H}_d$.

- **(I)** $D \leq n$**:** Run ALH with the start system (g, ζ), where

$$g_i = X_i^{d_i} - X_0^{d_i}, \quad \zeta = (1, \ldots, 1)$$

$$\mu(g, \zeta)^2 \leq 2(n+1)^D.$$

- **(II)** $D \geq n$**:** Use known method from computer algebra (Renegar), taking roughly D^n steps.

If $D \leq n^{1-\varepsilon}$, for fixed $\varepsilon > 0$, then n^D and hence the running time is polynomially bounded in N. Similarly for $D \geq n^{1+\varepsilon}$.

On the proof

- Reduce to smoothed analysis of mean square condition number defined as

$$\mu_2(q) := \Big(\frac{1}{d_1 \cdots d_n} \sum_{q(\zeta)=0} \mu(q,\zeta)^2\Big)^{1/2} \quad \text{for } q \in \mathcal{H}_d.$$

- Main auxiliary result:

$$\sup_{\|\overline{q}\|=1} \mathbb{E}_{q\sim N(\overline{q},\sigma^2 I)}\Big(\frac{\mu_2(q)^2}{\|q\|^2}\Big) = \mathcal{O}\Big(\frac{n}{\sigma^2}\Big).$$

- Proof is involved and proceeds by the analysis of certain probability distributions on fiber bundles (coarea formula etc).
- This way, the proof essentially reduces to a smoothed analysis of a matrix condition number.