# Smoothed Analysis of Condition Numbers

Peter Bürgisser

Universität Paderborn

Jo60: A Modern Computer Algebraist
Celebrating the Research and Influence of Joachim von zur Gathen

Bonn, June 2010

# Motivation

- Most common theoretical approach to understanding behaviour of algorithms: worst-case analysis.
- Sometimes algorithms perform well in practice and still have bad worst-case behaviour. Famous example: Dantzig's simplex algorithm.
- Average-case analysis tries to rectify this discrepancy: bound expected performance of an algorithm on random inputs. Average-case analyses for simplex algorithm: Borgwardt (1982) and Smale (1983).
- Disadvantage: strong dependence on unknown distribution of inputs.

# Smoothed analysis

New form of analysis of algorithms, proposed by Spielman and Teng. Smoothed analysis of simplex algorithm (Gödel Prize 2008, Fulkerson Prize 2009).

Let $T\colon \mathbb{R}^p \to \mathbb{R}_+ \cup \{\infty\}$ be a function (running time, condition number). Instead of showing

> *"It is unlikely that $T(a)$ will be large."*

one shows that

> *"For all $\overline{a}$ and all slight random perturbations $\overline{a} + \Delta a$, it is unlikely that $T(\overline{a} + \Delta a)$ will be large."*

| Worst case analysis | Average case analysis | Smoothed analysis |
|---|---|---|
| $\sup_{a \in \mathbb{R}^p} T(a)$ | $\mathbb{E}_{a \in \mathcal{D}} T(a)$ | $\sup_{\overline{a} \in \mathbb{R}^p} \mathbb{E}_{a \in N(\overline{a}, \sigma^2)} T(a)$ |

$\mathcal{D}$ distribution on $\mathbb{R}^p$, $N(\overline{a}, \sigma^2)$ Gaussian distribution centered at $\overline{a}$.

# Condition based analysis

- Smoothed analysis can be applied to a wide variety of numerical algorithms.
- Understanding condition numbers is important intermediate step.
- Condition numbers quantify errors when input is modified by small perturbation. Relevant for finite precision.
- Running time $T(x)$ of iterative numerical algorithms on input $x \in \mathbb{N}^n$ (measured by number of arithmetic operations), can often be effectively bounded by a polynomial in the size $n$ of $x$ and some measure $\mu(x)$ of conditioning of $x$.

# Stochastic analysis of condition numbers

- Two-part scheme for dealing with complexity upper bounds in numerical analysis (Smale):
  - I Condition based analysis: $T(x) \;\leq\; \big(\mathrm{size}(x) + \mu(x) + \big)^c$
  - II Stochastic analysis of condition number $\mu(x)$ for random inputs $x$.
- Approach elaborated for average-case complexity since eighties by many researchers, the pioneers being: Demmel, Edelman, Renegar, Shub, Smale, Todd, Vavasis, Ye, and others.
- Part two of Smale's scheme can be naturally refined by performing a smoothed analysis of the condition number $\mu(x)$ involved.
- Smoothed analysis for condition numbers since 2004: Amelunxen, Bürgisser, Cucker, Dunagan, Hauser, Lotz, Sankar, Spielman, Tao, Teng, Vu, Wschebor and others.

# Part I: Linear Equalities

## Turing's condition number of a matrix

A. Turing, 1948

J. von Neumann and H. Goldstine, 1947

# General definition of condition number

- Numerical computation problem

$$f : \mathbb{R}^p \to \mathbb{R}^q,\ x \mapsto y = f(x).$$

  Fix norms $\| \ \|$ on $\mathbb{R}^p, \mathbb{R}^q$.
- Relative error $\|\Delta y\|/\|y\|$ of output, relative error $\|\Delta x\|/\|x\|$ of input.
- Condition number $\kappa(f, x)$ of $x$:

$$\|\Delta y\|/\|y\| \ \lesssim \ \kappa(f, x)\ \|\Delta x\|/\|x\|.$$

- If $f$ is differentiable:

$$\kappa(f, x) \ := \ \|Df(x)\| \ \frac{\|x\|}{\|f(x)\|}$$

  where $\|Df(x)\|$ denotes the operator norm of the Jacobian of $f$ at $x$.

# Turing's condition number

- Consider matrix inversion

$$f\colon \mathrm{GL}(m,\mathbb{R}) \to \mathbb{R}^{m\times m}, A \mapsto A^{-1}.$$

  We measure errors with the spectral norm.

- Condition number of $A$ with respect to $f$ equals the classical condition number of $A$:

$$\kappa(A) := \kappa(f, A) = \|A\|\,\|A^{-1}\|.$$

- Note that $\kappa(\lambda A) = \kappa(A)$ for $\lambda \in \mathbb{R}$.
- $\kappa(A)$ was introduced by A. Turing in 1948.

# Distance to ill-posedness

- We call the set of singular matrices $\Sigma \subseteq \mathbb{R}^{m \times m}$ the set of ill-posed instances for matrix inversion. Clearly, $A \in \Sigma \Leftrightarrow \det A = 0$.
- The Eckart-Young Theorem from 1936 states that

$$\kappa(A) = \|A\| \, \|A^{-1}\| = \frac{\|A\|}{\mathrm{dist}(A, \Sigma)}.$$

where dist either refers to operator norm or to Frobenius norm (Euclidean norm on $\mathbb{R}^{n \times n}$).

# Smoothed analysis of $\kappa(A)$

- We model a slight perturbation of $A$ due to noise, round-off, etc. with isotropic Gaussian distributions $A \sim N(\overline{A}, \sigma^2 I)$.
- Consider the density

$$\rho(A) = \frac{1}{(\sigma\sqrt{2\pi})^{n^2}} \exp\Big(-\frac{\|A - \overline{A}\|_F^2}{2\sigma^2}\Big).$$

  with mean $\overline{A} \in \mathbb{R}^{n\times n}$ and covariance matrix $\sigma^2 I$.
- Improving results by Sankar, Spielman, and Teng, Wschebor showed:

**Theorem (Wschebor, 2004)**

$$\sup_{\|\overline{A}\|=1} \ \underset{A\sim N(\overline{A},\sigma^2 I)}{\mathrm{Prob}} \{\kappa(A) \geq t\} \ = \ \mathcal{O}\Big(\frac{n}{\sigma t}\Big).$$

# Part II: Linear Inequalities

# Condition numbers of linear programming

Jim Renegar, 1995

# Linear Programming Feasibility Problem (1)

- We focus on the homogeneous feasibility problem.
- For $A \in \mathbb{R}^{m \times n}$, $n > m$, consider the system of linear inequalities

$$\exists x \in \mathbb{R}^n \; Ax = 0, x > 0. \tag{P}$$

and its dual problem

$$\exists y \in \mathbb{R}^m \; A^T y < 0 \tag{D}$$

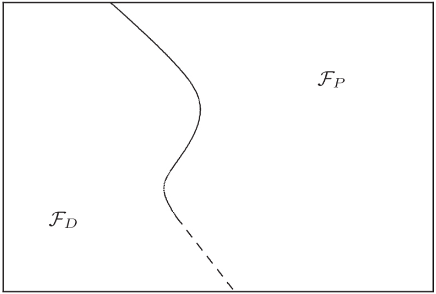
- Let $\mathcal{F}_P^\circ$ and $\mathcal{F}_D^\circ$ denote the set of instances where $P$ and $D$ are solvable, respectively.
- We have a disjoint union

$$\mathbb{R}^{n \times m} = \mathcal{F}_P^\circ \cup \mathcal{F}_D^\circ \cup \Sigma,$$

where the set of ill-posed instances $\Sigma$ is the common boundary of $\mathcal{F}_P^\circ$ and $\mathcal{F}_D^\circ$.

# Linear Programming Feasibility Problem (2)

$$\mathbb{R}^{n\times m} = \mathcal{F}_P^\circ \cup \mathcal{F}_D^\circ \cup \Sigma,$$




The Homogeneous Linear Programming Feasibility problem (HLPF) is to decide for given $A$, whether $A \in \mathcal{F}_P^\circ$ or $A \in \mathcal{F}_D^\circ$.

# Renegar's condition number

- For the HLPF problem, J. Renegar defined the condition number of the instance $A \in \mathbb{R}^{m \times n}$ as

$$\mathcal{C}_R(A) := \frac{\|A\|}{\mathrm{dist}(A, \Sigma)}.$$

- Note that $\mathcal{C}_R(A) = \infty$ iff $A \in \Sigma$.

### Renegar 1995

HLPF can be solved with an interior-point method with a number of iterations bounded by

$$\mathcal{O}\Big(\sqrt{n}\,\log(n\,\mathcal{C}_R(A))\Big).$$

# Condition-based complexity analysis

- L. Khachian: for an integer matrix $A$, HLPF can be solved in polynomial time (in the bit size of $A$).
- Notorious open problem: can HLPF be solved for real matrix $A$ with a number of arithmetic operations polynomial in $m, n$?
- Renegar's analysis bounds the number of arithmetic operations by a polynomial in both the
  - dimension $n$ of the problem
  - logarithm of its condition number.
- $\log \mathcal{C}_R(A)$ is polynomially bounded in bitsize of $A$ for integer matrices $A \notin \Sigma$.
- Consequence: HLPF can be solved in polynomial time for an integer matrix $A$, counting bit operations.

# Smoothed analysis of Renegar's condition number

Model for local perturbations: $\overline{A} \in \mathbb{R}^{m\times n}$, $A \sim N(\overline{A}, \sigma^2 I)$.

**Theorem (Dunagan, Spielman & Teng)**

$$\sup_{\|\overline{A}\|=1} \mathbb{E}_{A\sim N(\overline{A},\sigma^2 I)}\big(\log \mathcal{C}_R(A)\big) = \mathcal{O}\Big(\log \frac{n}{\sigma}\Big).$$

This implies the bound

$$\mathcal{O}(\sqrt{n}\log\frac{n}{\sigma})$$

on the smoothed expected number of iterations of the IPM considered for HLPF.

# Part III: Polynomial Equations

Complexity of Bezout's Theorem

(Shub and Smale 1993–1996)

# Smale's 17th problem

The 17th of S. 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?*

# Notations

- For a degree vector $d = (d_1, \dots, d_n)$ we define

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

- The input size is $N := \dim_{\mathbb{C}} \mathcal{H}_d$.
- We look for zeros $\zeta$ of $f$ in complex projective space $\mathbb{P}^n$: $f(\zeta) = 0$.
- The Bombieri-Weyl hermitian inner product $\langle\ \rangle$ on $\mathcal{H}_d$ is invariant under the natural action of the unitary group $U(n+1)$ on $\mathcal{H}_d$ and allows to define $\|f\| := \langle f, f\rangle^{1/2}$.
- We have a standard Gaussian distribution on $\mathcal{H}_d$ with density

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

# Approximate zeros

- Have a 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 $x_0$.

- Definition (Smale). $x \in \mathbb{P}^n$ is called approximate zero of $f$ with zero $\zeta$ iff

$$\forall i \in \mathbb{N} : \quad d(x_i, \zeta) \leq \frac{1}{2^{2^i - 1}}\, d(x_0, \zeta).$$

- Here the distance $d$ refers to the geodesic distance on the Riemannian manifold $\mathbb{P}^n$ (Fubini-Study metric). One may think of $d$ as an angle.

# Condition number

- Let $f(\zeta) = 0$. How much does $\zeta$ change when we perturb $f$ a little?
- Consider the solution variety $V := \Big\{(f, \zeta) \mid f(\zeta) = 0\Big\} \subseteq \mathcal{H}_d \times \mathbb{P}^n$, which is a smooth Riemannian submanifold
- The solution map $G$ is the local inverse of the projection map $V \to \mathbb{P}(\mathcal{H}_d), (f, \zeta) \mapsto f$.
- The condition number of $f$ at $(f, \zeta)$,

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

is essentially the operator norm of the derivative of $G$ at $\zeta$, where

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

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

# Radius of quadratic convergence

Put $D := \max_i d_i$.

### Smale's Gamma Theorem

If

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

then $x$ is an approximate zero of $f$ associated with $\zeta$.

# Adaptive linear homotopy continuation

- Given a start system $(g, \zeta) \in V$ and an input $f \in \mathcal{H}_d$.
- Connecting $g$ and $f$ by line segment $[g, f]$ consisting of

$$q_t := (1-t)g + tf \qquad \text{for } t \in [0,1].$$

- If $[g, f]$ does not meet the discriminant variety (none of the $q_t$ has a multiple zero), then there exists a unique lifting to a path

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

such that $(f_0, \zeta_0) = (g, \zeta)$.

- Follow $\gamma$ numerically: Let $t_0 = 0, \ldots, t_k = 1$ and write $q_i := q_{t_i}$. Successively compute approximations $z_i$ of $\zeta_{t_i}$ by Newton's method

$$z_{i+1} := N_{q_{i+1}}(z_i)$$

starting with $z_0 := \zeta$.

# Complexity of adaptive linear homotopy continuation

- Compute $t_{i+1}$ adaptively from $t_i$ such that

$$d(q_{i+1}, q_i) = \frac{c}{D^{3/2}\mu^2(q_i, x_))}.$$

  This defines the Adaptive Linear Homotopy ALH algorithm.

- We denote by $K(f, g, \zeta)$ the number $k$ of Newton continuation steps that are needed to follow the homotopy.

**Shub & Smale, and Shub (2007)**

$x_i$ is an approximate zero of $\zeta_i$ for all $i$. Moreover,

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

# Randomized algorithm

- Shub and Smale had shown that almost all $(g, \zeta) \in V$ have a condition number polynomial bounded in $N, D$.
- However, it is unknown how to efficiently construct such $(g, \zeta)$.
- Since we don't know how to construct a good start system $(g, \zeta_0)$, we choose it at random:
  - choose $g \in \mathcal{H}_d$ from standard Gaussian,
  - choose one of the $\mathcal{D} := d_1 \cdots d_n$ many zeros $\zeta$ of $g$ uniformly at random.
- Efficient sampling of $(g, \zeta)$ is possible (Beltrán & Pardo 2008).
- Las Vegas Algorithm LV
  draw $(g, \zeta) \in V$ at random
  run ALH on input $(f, g, \zeta)$
- LV has the expected "running time"

$$K(f) := \mathbb{E}_{g,\zeta} K(f, g, \zeta).$$

# Average expected polynomial time

- LV runs in average expected polynomial time:

**Beltrán and Pardo**

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

where the expectation is over a standard Gaussian $f \in \mathcal{H}_d$.

- When allowing randomized algorithms, this is a solution to Smale's 17th problem.
- Note that randomness enters here in two ways: as an algorithmic tool and as a way to measure the performance of algorithms.

# Smoothed expected polynomial time

- Smoothed analysis: let $\overline{f} \in \mathcal{H}_d$ and suppose that $f$ is isotropic Gaussian with mean $\overline{f}$ and variance $\sigma^2$.
- Recently, I obtained with Felipe Cucker the following result

## Smoothed analysis of ALH

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

# A near solution to Smale's 17th problem

There is a deterministic algorithm for Smale's 17th problem taking on standard Gaussian input $f \in \mathcal{H}_d$ an expected number of arithmetic operations $T(f)$ bounded by

$$\mathbb{E}_f\, T(f) = N^{\mathcal{O}(\log\log N)}.$$

- If $D \leq n$, the algorithm runs ALH with the start system $(g, \zeta)$, where
$$g_i = X_i^{d_i} - X_0^{d_i}, \quad \zeta = (1, \ldots, 1)$$
$$\mu(g, \zeta)^2 \leq 2(n+1)^D.$$
- If $D \leq n^{1-\varepsilon}$, for fixed $\varepsilon > 0$, then $n^D$ is polynomially bounded in $N$. In this case we even get deterministic polynomial time.
- In the case $D \geq n$, the algorithm is a more or less known symbolic procedure that takes roughly $D^n$ steps.

# Thank you for your attention!