CONDITION: THE GEOMETRY OF NUMERICAL ALGORITHMS

MATERIAL ON A COURSE GIVEN AT JOURNÉES NATIONALES DE CALCUL FORMEL 2017
LUMINY, JANUARY 16–20, 2017

PETER BÜRGISSER

Abstract. The performance of numerical algorithms, both regarding stability and complexity, can be understood in a unified way in terms of condition numbers. This requires to identify the appropriate geometric settings and to characterize condition in geometric ways. A probabilistic analysis of numerical algorithms can be reduced to a corresponding analysis of condition numbers, which leads to fascinating problems of geometric probability and integral geometry.

This is the theme of my recent monograph *Condition*, written with Felipe Cucker, that appeared in 2013 in Springer's Grundlehren series. The monograph is divided into three parts. Its first part deals with the solution of linear systems of equations, where many of the concepts can be explained in an elementary way. The second part is devoted to linear programming, i.e., the solution of systems of linear inequalities (there exist natural extensions to convex programming). The third part is devoted to the solution of systems of polynomial equations, focusing on Smale's 17th problem, which asks to find a solution of a given system of n complex homogeneous polynomial equations in $n+1$ unknowns. This problem can be solved in average (and even smoothed) polynomial time. Recently, Pierre Lairez succeeded in providing a complete solution of Smale's 17th problem ("A deterministic algorithm to compute approximate roots of polynomial systems in polynomial average time," to appear in *J. FoCM*).

The enclosed course material in the form of slides follows the three part structure of the monograph and attempts to illustrate the main unifying concepts and key ideas. The framework seems quite generally applicable. For instance, a numerical algorithm for computing eigenpairs of matrices, that is numerically stable and provably runs in average polynomial time, was recently developed along these lines (Armentano, Beltrán, Bürgisser, Cucker, and Shub, "A stable, polynomial-time algorithm for the eigenpair problem," accepted for *J. EMS*).

Date: December 30, 2016.

Condition: The Geometry of Numerical Algorithms

Peter Bürgisser

Technical University of Berlin

Journées Nationales de Calcul Formel 2017

Luminy, January 16–20, 2017

Outline

Overview

Motivation

- In computer science, the most common theoretical approach to understanding the behaviour of algorithms is worst-case analysis.
- There are cases of algorithms that perform exceedingly well in practice and still have a provably bad worst-case behaviour. A famous example is Dantzig's simplex algorithm.
- To rectify this discrepancy, the concept of average-case analysis was introduced. This means bounding the expected performance of an algorithm on random inputs. For the simplex algorithm: average-case analyses by Borgwardt (1982) and Smale (1983).
- However, average analysis can rarely explain a good performance in practice. Its results strongly depend on the distribution of the inputs, which is unknown, and usually assumed to be Gaussian for rendering the mathematical analysis feasible.

Smoothed analysis

Smoothed analysis is a new form of analysis of algorithms, that arguably blends the best of both worst-case and average-case. It was proposed by Spielman and Teng who performed a smoothed analysis of the running time of the 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
$\sup\limits_{a\in\mathbb{R}^p} T(a)$	$\mathbb{E}_{a\in\mathcal{D}}\, T(a)$	$\sup\limits_{\overline{a}\in\mathbb{R}^p} \mathbb{E}_{a\in N(\overline{a},\sigma^2)}\, T(a)$

$\mathcal{D}$ 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.
- For doing so, understanding the concept of condition numbers is an important intermediate step.
- Condition numbers quantify the errors when the input has been modified by a small perturbation.
- The best known condition number is $\kappa(A) = \|A\|\|A^{-1}\|$ for matrix inversion and linear equation solving.
- The running time $T(x, \varepsilon)$ of iterative numerical algorithms, measured as the number of arithmetic operations, can often be bounded in the form

$$T(x, \varepsilon) \ \leq \ \left(\mathrm{size}(x) + \mu(x) + \log \varepsilon^{-1}\right)^c,$$

  - input $x \in \mathbb{R}^n$ of $\mathrm{size}(x) := n$
  - $\mu(x)$ measure of conditioning of x
  - ε required accuracy.

Stochastic analysis of condition numbers

- Two-part scheme for dealing with complexity upper bounds in numerical analysis (Smale):
  - I Condition based analysis: $T(x,\varepsilon) \ \le \ \left(\mathrm{size}(x) + \mu(x) + \log \varepsilon^{-1}\right)^c$
  - II Stochastic analysis of condition number $\mu(x)$ for random inputs x.
- This approach was elaborated for average-case complexity since the 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

- Suppose we have a numerical computation problem

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

  We fix norms $\| \ \|$ on $\mathbb{R}^p, \mathbb{R}^q$.

- Suppose the input x has a small relative error $\|\Delta x\|/\|x\|$. We want to bound the relative error $\|\Delta y\|/\|y\|$ of the output.
- This is done by the condition number $\kappa(f, x)$ of x:

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

- Formal definition for differentiable f:

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

  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.

- We show by a perturbation argument that the 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.

Connection to eigenvalues

- Let $\lambda_1 \geq \dots \lambda_n$ be the eigenvalues of $A^T A$.
- Then
$$\|A\|^2 = \sup_{\|x\|=1} \|Ax\|^2 = \sup_{\|x\|=1} x^T A^T A x.$$
Hence $\|A\|^2 = \lambda_1$ is the largest eigenvalue of $A^T A$.
- Since Let $\lambda_n^{-1} \geq \dots \lambda_1^{-1}$ are the eigenvalues of $A^{-1}(A^{-1})^T$, we get
$$\|A^{-1}\|^2 = \|(A^{-1})^T\|^2 = \lambda_n^{-1}.$$
- We obtain
$$\kappa(A) = \|A\| \, \|A^{-1}\| = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_n}} \geq 1.$$
- $\sqrt{\lambda_1}$ and $\sqrt{\lambda_n}$ are called largest and smallest singular value of A.

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

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

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

$$\|A\|_F := \left(\mathrm{tr}(AA^T)\right)^{1/2} = \Big(\sum_{i,j} a_{ij}^2\Big)^{1/2}.$$

- Hence

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

Finite precision

- Digital computers operate with floating-point numbers, and every arithmetic operations produces a round-off error.
- Let ϵ_{mach} denote the round-off unit (e.g., 10^{-12}).
- Suppose we compute the approximation $\widetilde{x}$ of $x \in \mathbb{R}$ with relative error δ, i.e. $\widetilde{x} = x(1+\delta)$.
- The best we can hope for is $\delta \leq \frac{1}{2}\epsilon_{\mathsf{mach}}$.
- One calls $\log_{10}\left(\frac{\delta}{\epsilon_{\mathsf{mach}}}\right)$ the loss of precision in decimal digits.

Condition number bounds loss of precision

- Turing's condition number is relevant for finite precision analysis of linear algebra.
- For instance, QR factorization is one of the main engines in numerical linear algebra.
- Let $A \in \mathbb{R}^{n\times n}$ be invertible and $b \in \mathbb{R}^n$. If the system $Ax = b$ is solved using the Householder QR factorization, the computed solution $\widetilde{x}$ has a loss of precision bounded by

$$\log_{10}\left(\frac{\|\widetilde{x} - x\|}{\epsilon_{\mathrm{mach}}\|x\|}\right) \leq 2\log_{10} n + \log_{10}\kappa(A) + c,$$

where c denotes a universal constant c.

Method of conjugate gradients

- Consider a full-rank rectangular matrix $A \in \mathbb{R}^{m\times n}$ with $m > n$, a vector $c \in \mathbb{R}^m$, and the least squares problem

$$\min_{v\in\mathbb{R}^n} \|Av - c\|.$$

- The solution $x^* \in \mathbb{R}^n$ is given by the solution of the system $Sx = b$ with

$$S := A^T A \in \mathbb{R}^{n\times n}, \quad b := A^T c.$$

- By construction, S is symmetric and positive definite.
- The method of conjugate gradients is a powerful iterative method of numerical linear algebra. Upon input S, b and a start vector x_0 it produces a sequence of iterates $x_1, \times_2, \dots, x_n = x^*$.
- In order to achieve a relative error ε, it suffices to execute

$$\frac{1}{2}\sqrt{\kappa(S)}\,\ln\left(\frac{1}{\varepsilon}\right)$$

iterations (Stiefel and Hestenes).

Probabilistic Analysis of Turing's Condition Number

Average-case Analysis

J. von Neumann and H. Goldstine

Numerical inverting matrices of high order, II, 1951

Wishart distribution

- Suppose $A \in \mathbb{R}^{n \times n}$ is a random matrix with independent standard Gaussian entries.
- What can we say about the random variable $\kappa(A)$?
- The distribution of $A^T A$ is known as Wishart distribution which is of relevance in multivariate statistics.
- The joint probability density of the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$ of $A^T A$ is known (Fisher, Hsu, Roy 1939) and equals

$$\rho(\lambda) = c_n \; e^{-\frac{1}{2}\sum_i \lambda_i} \prod_i \lambda_i^{-\frac{1}{2}} \prod_{i<j} (\lambda_i - \lambda_j),$$

  with some normalizing constant c_n.
- It plays an important role in physics (Wigner 1962).

(Limit) distribution of $\kappa(A)$

- From the joint distribution of the eigenvalues, it is possible to derive the distribution of $\kappa(A) = \sqrt{\lambda_1/\lambda_n}$.

Edelman, 1988

$$\lim_{n\to\infty} \mathrm{Prob}\{\kappa(A) \geq nx\} = 1 - e^{-2/x-2/x^2} = \frac{2}{x} + \mathcal{O}(\frac{1}{x^2}).$$

- This implies for the expectation

$$\mathbb{E}_A(\log \kappa(A)) = \log n + c + o(1).$$

- Application: the QR factorization has an average loss of precision $3\log_{10} n + \mathcal{O}(1)$. Satisfactory result!
- There is an intuitive geometric way of deriving such results, that also has the virtue of generalizing to a wide variety of situations.

Reduction to sphere

- Define Frobenius condition number $\kappa_F(A) := \|A\|_F \, \|A^{-1}\| \geq \kappa(A)$.
- The standard Gaussian distribution on $\mathbb{R}^{n\times n}$ induces the uniform distribution on the sphere $\mathbb{S} := S^{n^2-1} := \{A \in \mathbb{R}^{n\times n} \mid \|A\|_F = 1\}$ via

$$\mathbb{R}^{n\times n} \setminus \{0\} \to \mathbb{S}, \; A \mapsto B = \tfrac{1}{\|A\|_F} A$$

- By the characterization by inverse distance to ill-posedness

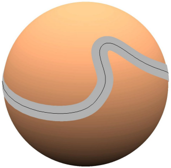
$$\kappa_F(A) = \kappa_F(B) = \|B\|_F \, \|B^{-1}\| = \frac{\|B\|_F}{\mathrm{dist}(B,\Sigma)} = \frac{1}{\mathrm{dist}(B,\Sigma)},$$

  where $\Sigma := \{A' \in \mathbb{R}^{n\times n} \mid \det A' = 0\}$ and dist is measured by Frobenius norm.
- Hence

$$\operatorname*{Prob}_A\{\kappa_F(A) \geq \varepsilon^{-1}\} = \operatorname*{Prob}_B\{\kappa_F(B) \geq \varepsilon^{-1}\} = \operatorname*{Prob}_B\{\mathrm{dist}(B,\Sigma) \leq \varepsilon\}.$$

Volume of tubes

Let $T(\Sigma_{\mathbb{S}}, \varepsilon)$ denote the neighborhood (or tube) of $\Sigma_{\mathbb{S}} := \Sigma \cap \mathbb{S}$ of radius $\arcsin \varepsilon$ in the sphere $\mathbb{S}$.



$$\begin{aligned} \operatorname*{Prob}_B\{\kappa_F(B) \geq \varepsilon^{-1}\} &= \operatorname*{Prob}_B\{\mathrm{dist}(B, \Sigma) \leq \varepsilon\} \\ = \frac{\mathrm{vol}(T(\Sigma_{\mathbb{S}}, \varepsilon))}{\mathrm{vol}(\mathbb{S})} &= \frac{\mathrm{vol}(\Sigma_{\mathbb{S}}) \cdot 2\varepsilon}{\mathrm{vol}(\mathbb{S})} + o(\varepsilon). \end{aligned}$$

Volume of determinant hypersurface $\Sigma_{\mathbb{S}}$

- Σ is the zero set of the determinant, a homogeneous polynomial of degree $d = n$.
- Let P be a plane (two-dimensional subspace) in $\mathbb{R}^{n\times n}$. How about the intersection $P \cap \Sigma$?
- Either $P \cap \Sigma = P$ (degenerate case) or $P \cap \Sigma$ is a union of k lines through the origin, for $k \leq d$.
- Hence, almost surely, $P \cap \Sigma_{\mathbb{S}}$ is either empty or consists of $2d$ points, two of which are diametral.
- Let $\mathbb{S}'$ be the intersection of $\mathbb{S}$ with a hyperplane (hyperequator). Poincaré's formula of integral geometry states

$$\frac{\mathrm{vol}(\Sigma_{\mathbb{S}})}{\mathrm{vol}(\mathbb{S}')} = \mathbb{E}_P\left(\frac{\#(P \cap \Sigma_{\mathbb{S}})}{2}\right),$$

  where the expectation is over random planes P.
- Therefore,

$$\frac{\mathrm{vol}(\Sigma_{\mathbb{S}})}{\mathrm{vol}(\mathbb{S}')} = d \cdot \operatorname*{Prob}_P\{P \cap \Sigma_{\mathbb{S}} \neq \emptyset\} \leq d.$$

Volume of determinant hypersurface $\Sigma_{\mathbb{S}}$

- From

$$\frac{\mathrm{vol}(\Sigma_{\mathbb{S}})}{\mathrm{vol}(\mathbb{S})} = \frac{\mathrm{vol}(\Sigma_{\mathbb{S}})}{\mathrm{vol}(\mathbb{S}')} \cdot \frac{\mathrm{vol}(\mathbb{S}')}{\mathrm{vol}(\mathbb{S})} \leq d\sqrt{\frac{\dim \mathbb{S}}{2\pi}} \leq \frac{n^2}{\sqrt{2\pi}}.$$

  we obtain the asymptotic tail bound

$$\operatorname*{Prob}_B\{\kappa_F(B) \geq \varepsilon^{-1}\} = \frac{\mathrm{vol}(\Sigma_{\mathbb{S}}) \cdot 2\varepsilon}{\mathrm{vol}(\mathbb{S})} + o(\varepsilon) \leq \sqrt{\frac{2}{\pi}}\, n^2\, \varepsilon + o(\varepsilon).$$

- This bound is larger by a factor$\approx n$ than Edelman's bound

$$\operatorname*{Prob}_B\{\kappa(B) \geq \varepsilon^{-1}\} = 2n\, \varepsilon + o(\varepsilon).$$

- By a more careful estimation of tube volumes one can derive nonasymptotic bounds.
- These ideas have been developed in detail by J. Demmel (1988).

Application to method of conjugate gradients

- The method of conjugate gradients, on input $S = A^T A$, takes

$$\frac{1}{2}\sqrt{\kappa(S)}\,\ln\Big(\frac{1}{\varepsilon}\Big) = \frac{1}{2}\,\kappa(A)\,\ln\Big(\frac{1}{\varepsilon}\Big).$$

  iterations to achieve relative error ε.

- However,

$$\mathrm{Prob}\{\kappa(A) \geq t\} = \mathcal{O}\Big(\frac{n}{t}\Big)$$

  implies

$$\mathbb{E}(\kappa(A)) = \infty.$$

  This is inconsistent with the success of CGM in practice!

Explanation?

Condition of rectangular matrices

- CGM is usually applied to $S = R^T R$, where $R \in \mathbb{R}^{m \times n}$ is rectangular with $m \geq n$. (E.g., overdetermined least square problem with m linear constraints in n variables.)
- Let $q \in (0, 1)$. It is known (Geman, Silverstein) that for standard Gaussian R of size $m_n \times n$ and $m_n/n \to q$ for $n \to \infty$

$$\kappa(R) \xrightarrow{a.s.} \frac{1 + \sqrt{q}}{1 - \sqrt{q}}.$$

- Hence: The expected number of iterations of CGM is independent of n and only depends on the ratio q.
- E.g., for $4n \times n$ matrices R and large n, $\kappa(A) \simeq 3$.

Probabilistic Analysis of Turing's Condition Number

Smoothed Analysis

Smoothed analysis of $\kappa(A)$

- Take now any $\overline{A} \in \mathbb{R}^{n\times n}$, $0 < \sigma \leq 1$ and consider the isotropic Gaussian density

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

  with mean $\overline{A}$ and covariance matrix $\sigma^2 I$. Notation: $A \sim N(\overline{A}, \sigma^2 I)$.
- This models a slight perturbation of A due to noise, round-off, etc.
- The goal of a smoothed analysis of $\kappa(A)$ is to derive tail bounds on it that are independent of the center $\overline{A}$.
- Due to scale invariance of $\kappa(A)$ we assume $\|\overline{A}\| = 1$.
- Improving results by Sankar, Spielman, and Teng, Wschebor showed:

Theorem (Wschebor, 2004)

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

Smoothed analysis of $\kappa(A)$: rectangular case

- Wschebor's tail bound implies

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

- This gives a more compelling probabilistic interpretation of the success of several procedures in numerical linear algebra.
- For the rectangular case $R\in\mathbb{R}^{m\times n}$, tail bounds have been derived by B & Cucker. They imply

$$\sup_{\|\overline{R}\|=1} \mathbb{E}_{R\sim N(\overline{R},\sigma^2 I)}\big(\kappa(R)\big) \;\le\; \frac{20.1}{1-q}$$

for $q\in(0,1)$, $m/n\le q$, and sufficiently large n.

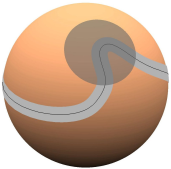
- As in the average case, the bound is independent of n. Interestingly, it is also independent of σ (for large n)!
- Has obvious consequence for the probabilistic analysis of CGM.

Geometric ideas for smoothed analysis

- The mentioned smoothed analysis bounds were derived by direct, problem adapted methods from probability.
- As for the average-case analysis, it is possible to give smoothed analysis bounds in a geometrically intuitive way that apply to a wide variety of situations.
- Think of κ as a function defined on the sphere $\mathbb{S} = S^{n^2-1}$.
- Let $B(\overline{A}, \sigma)$ denote the spherical cap in $\mathbb{S}$ of angular radius $\arcsin \sigma$ with center $\overline{A} \in \mathbb{S}$, where $0 \leq \sigma \leq 1$.
- We model now perturbations by A chosen from a uniform distribution on $B(\overline{A}, \sigma)$: uniform smoothed analysis.
  - $\sigma = 0$ yields worst-case analysis
  - $\sigma = 1$ yields average-case analysis

Geometric ideas for smoothed analysis

Let $T(\Sigma_{\mathbb{S}}, \varepsilon)$ denote the neighborhood (or tube) of Σ_S of radius $\arcsin \varepsilon$.



$$\operatorname*{Prob}_{A\in B(\overline{A},\sigma)} \{\kappa_F(A) \geq \varepsilon^{-1}\} = \operatorname*{Prob}_{A\in B(\overline{A},\sigma)} \{\mathsf{dist}(A,\Sigma) \leq \varepsilon\}$$

$$= \operatorname*{Prob}_{A\in B(\overline{A},\sigma)} \{A \in T(\Sigma_{\mathbb{S}}, \varepsilon)\} = \frac{\mathrm{vol}(T(\Sigma_{\mathbb{S}}, \varepsilon) \cap B(\overline{A}, \sigma))}{\mathrm{vol}(B(\overline{A}, \sigma))}$$

Uniform smoothed analysis means to provide relative bounds on the volume of tubes intersected with small spherical caps!

Heuristic estimation (1)

- Write $B := B(\overline{A}, \sigma)$. Then

$$\frac{\mathrm{vol}(T(\Sigma_{\mathbb{S}}, \varepsilon) \cap B)}{\mathrm{vol}(B)} \approx \frac{\mathrm{vol}(\Sigma_{\mathbb{S}} \cap B) \cdot 2\varepsilon}{\mathrm{vol}(B)}.$$

- Poincaré's formula yields as before, with $d = n$,

$$\frac{\mathrm{vol}(\Sigma_{\mathbb{S}} \cap B)}{\mathrm{vol}(\mathbb{S}')} = \mathbb{E}_P\left(\frac{\#(P \cap \Sigma_{\mathbb{S}} \cap B)}{2}\right) \leq d \cdot \mathop{\mathrm{Prob}}_P\{P \cap B \neq \emptyset\}.$$

  where the expectation is over random planes P.
- Therefore, writing $p := \dim \mathbb{S}$,

$$\begin{aligned} \frac{\mathrm{vol}(\Sigma_{\mathbb{S}} \cap B)}{\mathrm{vol}(B)} &= \frac{\mathrm{vol}(\Sigma_{\mathbb{S}} \cap B)}{\mathrm{vol}(\mathbb{S}')} \cdot \frac{\mathrm{vol}(\mathbb{S}')}{\mathrm{vol}(\mathbb{S})} \cdot \frac{\mathrm{vol}(\mathbb{S})}{\mathrm{vol} B} \\ &\lesssim d \cdot \mathop{\mathrm{Prob}}_P\{P \cap B \neq \emptyset\} \cdot \sqrt{p} \cdot \frac{1}{\sigma^p}. \end{aligned}$$

-

Heuristic estimation (2)

- $P \cap \mathbb{S} := S^1$ is a circle. We may as well fix this circle and take a random cap B of radius $\arcsin \sigma$.
- The cap B meets S^1 iff the center of B is σ-close to S^1. Therefore,

$$\operatorname*{Prob}_B\{S^1 \cap B \neq \emptyset\} = \frac{\mathrm{vol}(T(S^1, \sigma))}{\mathrm{vol}(\mathbb{S})}.$$

- This is roughly 2π times the volume of a $(p-1)$-dimensional ball of radius σ in the cross section to S^1, divided by $\mathrm{vol}(\mathbb{S})$. It is roughly σ^{p-1}.
- Hence

$$\begin{aligned} \frac{\mathrm{vol}(\Sigma_{\mathbb{S}} \cap B)}{\mathrm{vol}(B)} &\lesssim d \cdot \operatorname*{Prob}_P\{P \cap B \neq \emptyset\} \cdot \sqrt{p} \cdot \frac{1}{\sigma^p} \\ &\approx d \cdot \sigma^{p-1} \cdot \sqrt{p} \cdot \frac{1}{\sigma^p} = \frac{d\sqrt{p}}{\sigma}. \end{aligned}$$

Heuristic estimation (3)

- Altogether

$$\operatorname*{Prob}_{A\in B(\overline{A},\sigma)}\{\kappa_F(A)\geq\varepsilon^{-1}\} = \frac{\mathrm{vol}(T(\Sigma_{\mathbb{S}},\varepsilon)\cap B)}{\mathrm{vol}(B)}$$
$$\approx \quad \frac{\mathrm{vol}(\Sigma_{\mathbb{S}}\cap B)\cdot 2\varepsilon}{\mathrm{vol}(B)} \lesssim \frac{d\sqrt{p}\,\varepsilon}{\sigma} = \mathcal{O}\Big(\frac{n^2\varepsilon}{\sigma}\Big).$$

- Using some differential and integral geometry, this can be turned into a proof, yielding a bound of essentially this order of magnitude.
- The bound is worse by a factor n compared to Wschebor's result. But it has the advantage to be true in a much more general situation.

A general result for smoothed analysis

- Assume that $\Sigma \subset \mathbb{R}^{p+1}$ is given as the zero set of a homogeneous polynomial of degree d.
- For $a \in \mathbb{R}^{p+1}$ define the conic condition number of a abstractly by

$$\mathscr{C}(a) = \frac{\|a\|}{\mathrm{dist}(a, \mathbb{S})}.$$

Theorem (B, Cucker, Lotz, 2008)

For all $\sigma \in (0,1]$ and all $t \geq (2d+1)\frac{p}{\sigma}$,

$$\sup_{\overline{a} \in S^p} \operatorname*{Prob}_{a \in B(\overline{a},\sigma)} \{\mathscr{C}(a) \geq t\} \;\leq\; 26\, dp\, \frac{1}{\sigma t}.$$

$$\sup_{\overline{a} \in S^p} \mathbb{E}_{a \in B(\overline{a},\sigma)}(\ln \mathscr{C}(a)) \;\leq\; 2 \ln\left(\frac{dp}{\sigma}\right) + 4.7.$$

-

Application: Eigenvalue computation

- A similar result can be shown over the complex numbers, where the set Σ of ill-posed inputs is a complex algebraic hypersurface. (Considerably simpler proof.)
- Problem: Compute the (complex) eigenvalues of a matrix $A \in \mathbb{C}^{n\times n}$
- Set of ill-posed inputs: Set Σ of matrices A having multiple eigenvalues. This is the zero set of the discriminant polynomial of the characteristic polynomial, which has degree $d = n^2 - n$.
- Condition number (Wilkinson): Satisfies $\kappa_{\mathrm{eigen}}(A) \leq \frac{\sqrt{2}\,\|A\|_F}{\mathrm{dist}(A,\Sigma)}$
- Corollary: For all $\overline{A} \in \mathbb{R}^{n\times n}$ of Frobenius norm one and $0 < \sigma \leq 1$

$$\mathbb{E}_{A\in B(\overline{A},\sigma)}(\ln \kappa_{\mathrm{eigen}}(A)) \leq 2\ln \frac{n^4}{\sigma} + 5.$$

Random Triangular Matrices:

The classical condition number is not always appropriate!

Random triangular matrices are ill-conditioned

- Practitioners observed since long that triangular systems of equations are generally solved to high accuracy in spite of being, in general, ill-conditioned.
- Let $L = (\ell_{ij}) \in \mathbb{R}^{n \times n}$ be a random lower-triangular matrix with independent standard Gaussian random entries ℓ_{ij} for $i \geq j$.

Theorem (Viswanathan and Trefethen, 1998)

$$\mathbb{E}(\ln \kappa(L)) \geq \Omega(n).$$

- We give a simple proof of a related result later on.
- Would the loss of precision in the solution of triangular systems conform to this bound, we would not be able to accurately find these solutions!

Explanation?

Componentwise relative errors

The classical condition number is the condition number of matrix inversion $A \mapsto A^{-1}$:

$$\kappa(A) = \lim_{\delta \to 0} \sup_{\mathsf{RelError}(A) \leq \delta} \frac{\mathsf{RelError}(A^{-1})}{\mathsf{RelError}(A)}.$$

Here, we use the normwise relative error

$$\mathsf{RelError}(A) := \frac{\|\widetilde{A} - A\|}{\|A\|},$$

with the spectral norm $\| \; \|$.

Componentwise condition number

- Instead of RelError we may use the possibly much larger componentwise relative error

$$\mathrm{CwRelError}(A) := \max_{i,j} \frac{\|\widetilde{a_{ij}} - a_{ij}\|}{\|a_{ij}\|}.$$

- We define the componentwise condition number of matrix inversion correspondingly as

$$\mathrm{Cw}^{\dagger}(A) := \lim_{\delta \to 0} \sup_{\mathrm{CwRelError}(A) \le \delta} \frac{\mathrm{CwRelError}(A^{-1})}{\mathrm{CwRelError}(A)}.$$

Backward substitution is componentwise stable

- Backward substitution is the obvious algorithm for solving a triangular linear system $Lx = b$.
- The loss of precision of backward substitution can be shown to be bounded by $\mathcal{O}(\log \mathrm{Cw}^{\dagger}(L) + \log n)$,
- Recent result:

Theorem (Cheung & Cucker)

$$\mathbb{E}(\log \mathrm{Cw}^{\dagger}(L)) = \mathcal{O}(\log n)$$

for a random lower-triangular matrix $L \in \mathbb{R}^{n\times n}$ with independent standard Gaussian random entries ℓ_{ij}

- This explains why linear triangular systems can be solved by backward substitution with high accuracy.

Why random triang. matrices are ill-conditioned (1)

Let $L = (\ell_{ij})$ denote a random unit lower-triangular matrix with $\ell_{ii} = 1$ and with independent standard Gaussian random entries ℓ_{ij} for $i > j$. Then we have

$$\mathbb{E}(\|L^{-1}\|_F^2) = 2^n - 1.$$

In particular, $\mathbb{E}(\|L\|_F^2 \|L^{-1}\|_F^2) \geq n(2^n - 1)$, hence $\mathbb{E}(\kappa(L)^2)$ grows exponentially in n.

Proof.

- The first column $(s_1, \ldots, s_n)$ of L^{-1} is characterized by $s_1 = 1$ and the recursive relation

$$s_i = -\sum_{j=1}^{i-1} \ell_{ij} s_j \quad \text{for } i = 2, \ldots, n.$$

- Hence s_i is a function of the first i rows of L and thus independent of the entries of L in the rows with index larger than i.

Why random triang. matrices are ill-conditioned (2)

- By squaring we obtain for $i \geq 2$

$$s_i^2 = \sum_{\substack{j \neq k \\ j,k<i}} \ell_{ij}\ell_{ik} s_j s_k + \sum_{j<i} \ell_{ij}^2 s_j^2.$$

- By the preceding observation, $s_j s_k$ is independent of $\ell_{ij}\ell_{ik}$ for $j, k < i$. If additionally $j \neq k$, we get

$$\mathbb{E}(\ell_{ij}\ell_{ik} s_j s_k) = \mathbb{E}(\ell_{ij}\ell_{ik})\,\mathbb{E}(s_j s_k) = \mathbb{E}(\ell_{ij})\,\mathbb{E}(\ell_{ik})\,\mathbb{E}(s_j s_k) = 0$$

as ℓ_{ij} and ℓ_{ik} are independent and centered.

- So the expectations of the mixed terms vanish and we obtain, using $\mathbb{E}(\ell_{ij}^2) = 1$, that

$$\mathbb{E}(s_i^2) = \sum_{j=1}^{i-1} \mathbb{E}(s_j^2) \quad \text{for } i \geq 2.$$

- Solving this recursion with $\mathbb{E}(s_1^2) = 1$ yields

$$\mathbb{E}(s_i^2) = 2^{i-2} \quad \text{for } i \geq 2.$$

Why random triang. matrices are ill-conditioned (3)

- Therefore, the first column v_1 of L^{-1} satisfies

$$\mathbb{E}(\|v_1\|^2) = \mathbb{E}\Big(\sum_{i=1}^{n} s_i^2\Big) = 2^{n-1}.$$

- By an analogous argument one shows that

$$\mathbb{E}(\|v_k\|^2) = 2^{n-k}$$

for the kth column v_k of L^{-1}. Altogether, we obtain

$$\mathbb{E}(\|L^{-1}\|_F^2) = \mathbb{E}\Big(\sum_{k=1}^{n} \|v_k\|^2\Big) = \sum_{k=1}^{n} \mathbb{E}(\|v_k\|^2) = 2^n - 1.$$

$\square$

Part II: Linear Inequalities

Interior-point methods for linear programming

Linear programming (1)

- Standard primal form of linear programs: given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$; look for optimal $x \in \mathbb{R}^n$

$$\min c^{\mathrm{T}} x \quad \text{subject to } Ax = b\,,\; x \geq 0 \tag{P}$$

- Standard dual form of linear programs: Given A, b, c, look for optimal $y \in \mathbb{R}^m$.

$$\max b^{\mathrm{T}} y \quad \text{subject to } A^{\mathrm{T}} y \leq c \tag{D}$$

- It is known that $\max b^{\mathrm{T}} y = \min c^{\mathrm{T}} x$ if (P) and (D) are both feasible (duality).
- We always assume $n \geq m$.

Linear programming (2)

- Suppose that (P) and (D) are both feasible, The vector $s := c - A^{\mathrm{T}}y$ of slack variables satisfies

$$A^{\mathrm{T}}y + s = c, \quad s \geq 0,$$

hence, using $Ax = b$,

$$c^{\mathrm{T}}x - b^{\mathrm{T}}y = (s^{\mathrm{T}} + y^{\mathrm{T}}A)x - b^{\mathrm{T}}y = s^{\mathrm{T}}x + y^{\mathrm{T}}(Ax - b) = s^{\mathrm{T}}x \geq 0.$$

- Optimality is equivalent to $s^{\mathrm{T}}x = 0$, which is equivalent to the complementary slackness condition

$$x_i s_i = 0 \quad i = 1, 2, \ldots, n. \tag{1}$$

Idea of primal-dual interior point methods (1)

- Dantzig's simplex method follows a path of vertices on the boundary of the polyhedron of solutions.
- By contrast, interior point methods follow a path in the interior of the polyhedron, hence the name. This path is a nonlinear curve that is approximately followed by a variant of Newton's method.
- More specifically, primal-dual interior point methods follow the central path in the strictly feasible set $\mathcal{F}^\circ \subseteq \mathbb{R}^{n+m+n}$ defined by

$$Ax = b\,,\ A^{\mathrm{T}}y + s = c\,,\ x > 0\,,\ s > 0.$$

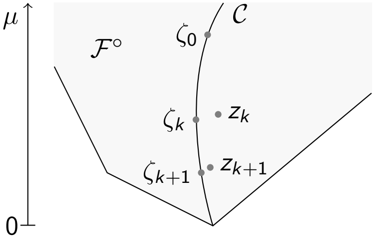
  with the additional quadratic constraints for $\mu > 0$

$$x_1 s_1 = \mu\,,\ \ldots\,,\ x_n s_n = \mu.$$

- It can be shown that, if $\operatorname{rank} A = m$, there is exactly one solution ζ_μ of this system, for all $\mu > 0$.

Idea of primal-dual interior point methods (2)

- Suppose we know ζ_{μ_0} for some $\mu_0 > 0$.
- We choose a centering parameter $\sigma \in (0,1)$ and consider $\mu_k = \sigma^k \mu_0$ converging to 0.
- We successively compute approximations z_k of $\zeta_k := \zeta_{\mu_k}$ for $k = 0, 1, 2, \ldots$ until a certain accuracy is reached.



- The duality measure of $z = (x, y, s) \in \mathcal{F}^\circ$ is defined as

$$\mu(z) := \frac{1}{n} \sum_{i=1}^{n} x_i s_i.$$

Derivation of the algorithm (1)

- We get the approximations z_k by Newton's method, one of the most fundamental methods in computational mathematics.
- Consider the map $F\colon \mathbb{R}^{n+m+n} \to \mathbb{R}^{n+m+n}$,

$$z = (x, y, s) \mapsto F(z) = (A^{\mathrm{T}}y + s - c, Ax - b, x_1 s_1, \ldots, x_n s_n)$$

  satisfying $\{\zeta_\mu\} = F^{-1}(0, 0, \mu e_n)$, where $e_n := (1, \ldots, 1) \in \mathbb{R}^n$. The Jacobian matrix of F at z equals

$$DF(z) = \begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix},$$

  where here and in the following we set

$$S = \mathrm{diag}(s_1, \ldots, s_n)\,,\ X = \mathrm{diag}(x_1, \ldots, x_n).$$

- Fact: $DF(z)$ is invertible if $\operatorname{rank} A = m$ and $s_i x_i \neq 0$ for all i.

Derivation of the algorithm (2)

- Set $\zeta_k = \zeta_{\mu_k}$. Then $F(\zeta_k) = (0, 0, \mu_k\, e_n)$ for all $k \in \mathbb{N}$. A first order approximation gives

$$F(\zeta_{k+1}) \approx F(\zeta_k) + DF(\zeta_k)(\zeta_{k+1} - \zeta_k). \tag{2}$$

- Suppose now that $z_k = (x, y, s) \in \mathcal{F}^\circ$ is an approximation of ζ_k. Then $F(z_k) = (0, 0, x_1 s_1, \ldots, x_n s_n) = (0, 0, X\,S\,e_n)$. We obtain from (2), replacing the unknowns ζ_k by z_k,

$$(0, 0, \mu_{k+1}\, e_n) = F(\zeta_{k+1}) \approx F(z_k) + DF(z_k)(\zeta_{k+1} - z_k).$$

- This leads to the following choice of the approximation of ζ_{k+1}.

$$z_{k+1} := z_k + DF(z_k)^{-1}(0, 0, \mu_{k+1}\, e_n - X\,S\,e_n)$$

- One easilys checks for $z_{k+1} = z_k + (\Delta x, \Delta y, \Delta s)$

$$A^{\mathrm{T}}(y + \Delta y) + (s + \Delta s) = c, \quad A(x + \Delta x) = b.$$

Primal-dual IPM

We choose $\sigma := 1 - \frac{1}{4\sqrt{n}}$.

Algorithm: Primal-Dual IPM
Input: $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ s.t. $\operatorname{rank} A = m \leq n$.
Choose starting point $z_0 = (x^0, y^0, s^0) \in \mathcal{F}^\circ$ with duality measure μ_0.
for $k = 0, 1, 2, \ldots$
Solve

$$\begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \cdot \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma^{k+1}\mu_0\, e_n - X^k\, S^k\, e_n \end{bmatrix},$$

where $X^k = \operatorname{diag}(x_1^k, \ldots, x_n^k)$, $S^k = \operatorname{diag}(s_1^k, \ldots, s_n^k)$.
Set

$$(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + (\Delta x^k, \Delta y^k, \Delta s^k).$$

until some stopping criterion is matched

Analysis of IPM

Theorem.

Primal-Dual IPM produces, on a strictly feasible starting point z_0 on the central path (or close to it), a sequence of iterates $z_k \in \mathcal{F}^\circ$ such that $\mu(z_k) = \sigma^k \mu(z_0)$. After

$$k \geq 4\sqrt{n} \ln \frac{\mu_0}{\varepsilon}.$$

iterations we have $\mu(z^k) \leq \varepsilon$.

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

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

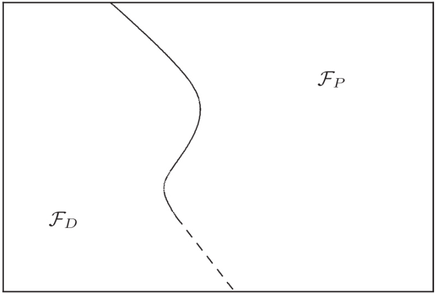
- Let $\mathcal{F}_P^\circ$ and $\mathcal{F}_D^\circ$ denote the set of instances where P and D are solvable, respectively.
- We have a disjoint union

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

where the set of ill-posed instances Σ 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 $A \in \Sigma$.
- HLPF can be solved by solving a related linear programming optimization problem up to a certain accuracy. More specifically,

$$\mu(z_k) = \mathcal{O}\Big(\frac{1}{n^2 \mathcal{C}_R(A)}\Big)$$

suffices for the decision $A \in \mathcal{F}_P^\circ$ or $A \in \mathcal{F}_D^\circ$ (Renegar 1995).
- By the previous analysis

Primal-Dual IPM can be solved 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.

Characterization of ill-posedness

- Let $A \in \mathbb{R}^{m \times n}$ be of full rank, $n > m$. Denote by $a_1, \ldots, a_n$ the columns of A and Δ its convex hull.
- Primal feasibility
$$\exists x \in \mathbb{R}^n \; Ax = 0, x > 0 \tag{P}$$
means that $x_1 a_1 + \cdots + x_n a_n = 0$ for some $x_i > 0$, that is, $0 \in \mathrm{int}\Delta$.
- Dual feasibility
$$\exists y \in \mathbb{R}^m \; A^T y < 0 \tag{D}$$
means that $\langle a_i, y \rangle < 0$ for some y, that is, Δ lies in some open halfspace.
- Recall $\Sigma = \overline{\mathcal{F}_P^\circ} \cap \overline{\mathcal{F}_D^\circ}$.
- Hence A is ill-posed, $A \in \Sigma$, iff Δ is contained in a closed halfspace and $0 \in \Delta$.

GCC condition number (1)

- We are going to define a variant of Renegar's condition number, that is better suited for probabilistic analysis.
- Suppose $A \in \mathcal{F}_S^\circ$ for $S \in \{P, D\}$. We define

$$\Delta(A) := \sup\left\{\delta > 0 \;\middle|\; \forall A' \in \mathbb{R}^{m\times n} \left(\max_{i\le n} \frac{\|a_i' - a_i\|}{\|a_i\|} < \delta \Rightarrow A' \in \mathcal{F}_S^\circ\right)\right\},$$

  where a_i' stands for the ith column of A'.
- The GCC-condition number of A (Goffin, Cheung, Cucker) is defined as

$$\mathscr{C}(A) := 1/\Delta(A).$$

- Note that we measure the relative size of the perturbation for each column a_i with respect to the norm of a_i.
- Also, $\Delta(A)$ is scale invariant. We may therefore assume, without loss of generality, that $\|a_i\| = 1$ for all i.
- Hence we can interpret A with columns $a_1, \ldots, a_n$ as an element in the product $\mathbb{S}^n = \mathbb{S} \times \cdots \times \mathbb{S}$ of spheres $\mathbb{S} := \mathbb{S}^{m-1}$.

GCC condition number (2)

- Let d denote angular distance on $\mathbb{S}$. Define a metric on $\mathbb{S}^n$ by

$$d_{\mathbb{S}}(A, B) \;:=\; \max_{1\leq i\leq n} d(a_i, b_i)\;.$$

- It is straightforward to show

$$\mathscr{C}(A) = \frac{1}{\sin d_{\mathbb{S}}(A, \Sigma)}.$$

- We note that HLPF can be solved by a primal-dual interior-point method with a number of iterations

$$\mathcal{O}\Big(\sqrt{n}\;\log(n\mathscr{C}(A))\Big).$$

Minimal spherical caps

- Let $\rho(A)$ be the angular radius of a spherical cap of minimal radius containing $a_1, \dots, a_n \in \mathbb{S}$.
- Easy to see: $\rho(A) < \frac{\pi}{2}$ iff $A \in \mathcal{F}_D^\circ$. Hence, $\rho(A) = \frac{\pi}{2}$ iff $A \in \Sigma$.

Theorem (Cheung & Cucker)

$$d_{\mathbb{S}}(A, \Sigma) \; = \; \begin{cases} \frac{\pi}{2} - \rho(A) & \text{if } A \in \mathcal{F}_D^\circ \\ \rho(A) - \frac{\pi}{2} & \text{if } A \in \mathbb{S}^n \setminus \mathcal{F}_D^\circ \end{cases} \; .$$

- In particular, $d_{\mathbb{S}}(A, \Sigma) \leq \frac{\pi}{2}$ and

$$\mathscr{C}(A)^{-1} = \sin d_{\mathbb{S}}(A, \Sigma) \; = \; |\cos \rho(A)| \; .$$

Average Analysis of GCC condition number

GCC condition number and coverage processes (1)

- Suppose $A \in \mathbb{R}^{n \times m}$ is standard Gaussian.
- After normalization, this means that each column a_i is independently chosen from the uniform distribution on the sphere $\mathbb{S}$.
- The probability distribution of $\mathscr{C}(A)$ is related to a classical question on covering a sphere by random spherical caps.
- Let $p(n, m, \alpha)$ denote the probability that randomly chosen spherical caps with centers $a_1, \ldots, a_n$ and angular radius α do not cover the sphere $\mathbb{S} = S^{m-1}$.
- We claim that

$$p(n, m, \alpha) = \mathrm{Prob}\left\{\rho(A) \leq \pi - \alpha\right\}.$$

GCC condition number and coverage processes (2)

- Claim: $p(n, m, \alpha) = \mathrm{Prob}\left\{\rho(A) \leq \pi - \alpha\right\}$.
- **Proof.** The caps of radius α with center $a_1, \ldots, a_n$ do not cover $\mathbb{S}$ iff there exists $y \in \mathbb{S}$ having distance greater than α from all a_i.
- This means that the cap of radius $\pi - \alpha$ centered at $-y$ contains all the a_i. Hence

$$p(n, m, \alpha) = \mathrm{Prob}\left\{\rho(A) \leq \pi - \alpha\right\}.$$

$\square$

Average analysis of $\mathscr{C}$

- The problem to determine the coverage probabilities $p(n,m,\alpha)$ is classical and completely solved only for $m-1 = \dim \mathbb{S} \leq 2$ (Gilbert '65, Miles '69).
- For $m > 3$ little was known except (Wendel '62)

$$p(n,m,\pi/2) = \frac{1}{2^{n-1}} \sum_{k=0}^{m-1} \binom{n-1}{k}$$

  and asymptotic formulas for $p(n,m,\alpha)$ for $\alpha \to 0$ (Janson '86).
- B, Cucker, Lotz (2007, Ann. Prob. to appear) recently discovered a closed formula for $p(n,m,\alpha)$ in the case $\alpha \geq \pi/2$
  and an upper bound for $p(n,m,\alpha)$ in the case $\alpha \leq \pi/2$.
- This implies

$$\mathbb{E}(\ln \mathscr{C}(A)) \leq 2 \ln m + 3.31.$$

- Consequence: the expected number of iterations of interior point methods for HLPF is $\mathcal{O}(\sqrt{n} \log n)$.

Closed formula for $p(n, m, \alpha)$

- For $\alpha \geq \pi/2$, setting $\varepsilon := |\cos(\alpha)|$,

$$p(n, m+1, \alpha) = \sum_{k=1}^{m} \binom{n}{k+1} C(m,k) \int_{\varepsilon}^{1} t^{m-k}(1-t^2)^{\frac{1}{2}km-1} \lambda_m(t)^{n-k-1}\, dt.$$

Here, $\lambda_m(t)$ denotes the relative volume of a spherical cap of radius $\arccos t \in [0, \pi/2]$ in S^m and the constants $C(m,k)$ describe higher moments of the volume of certain random simplices.

- Let $\mathcal{O}_m$ denote the m-dimensional volume of the sphere S^m.

$$\begin{aligned}
&\frac{\mathrm{vol}(\Sigma) \cdot \varepsilon}{\mathrm{vol}(\mathbb{S})^n} + o(\varepsilon^2) = \\
&= \quad \mathrm{Prob}\left\{A \in \mathcal{F}_D^{\circ},\ \mathscr{C}(A)^{-1} \leq \varepsilon\right\} = p(n, m, \pi/2) - p(n, m, \alpha) \\
&= \quad \binom{n}{m+1}(m+1)\frac{\mathcal{O}_{m-1}}{\mathcal{O}_m}\,\frac{1}{2^{n-2}}\,\varepsilon + o(\varepsilon^2).
\end{aligned}$$

Smoothed Analysis of GCC condition number

Gaussian smoothed analysis

- Model for local perturbations: $\overline{A} \in \mathbb{R}^{m\times n}$, Gaussians $A \in \mathbb{R}^{m\times n}$.

Theorem (Dunagan, Spielman & Teng)

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

-
- 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. Excellent result!

Uniform smoothed analysis of $\mathscr{C}$

- Model for smoothed analysis on product of spheres: $\overline{a}_1, \dots, \overline{a}_n \in \mathbb{S}$, independently choose a_i uniformly at random in spherical cap $B(\overline{a}_i, \sigma)$ of $\mathbb{S}$ centered at $\overline{a}_i$ with angular radius $\arcsin \sigma$. That is, choose $A \in B(\overline{A}, \sigma) := \prod_i B(\overline{a}_i, \sigma)$ uniformly.
- Amelunxen and B (2008): For $0 < \varepsilon \le \sigma/(2m(m+1))$

$$\sup_{\overline{A} \in \mathbb{S}^n} \operatorname*{Prob}_{A \in B(\overline{A}, \sigma)} \{A \in \mathcal{F}_D^\circ,\ \mathscr{C}(A) \ge \varepsilon^{-1}\} \ \le\ 6.5\, nm^2\, \frac{\varepsilon}{\sigma}.$$

- For the infeasible case ($A \notin \mathcal{F}_P^\circ$) a slightly worse tail estimate is obtained. Moreover,

$$\sup_{\overline{A} \in \mathbb{S}^n} \mathbb{E}_{A \in B(\overline{A}, \sigma)} \big(\ln \mathscr{C}(A)\big) = \mathcal{O}\Big(\ln \frac{n}{\sigma}\Big).$$

- We even obtain robustness results.

Sketch of proof (1)

- By a convex body K in the sphere $\mathbb{S}$ we understand the intersection with $\mathbb{S}$ of a closed regular convex cone C in $\mathbb{R}^m$.
- We call $T_o(\partial K, \varepsilon) := T(\partial K, \varepsilon) \setminus K$ the outer ε-neighborhood of the boundary ∂K. Then

$$\frac{\mathrm{vol}(T_o(\partial K, \varepsilon) \cap B(\overline{a}, \sigma))}{\mathrm{vol} B(\overline{a}, \sigma)} \le 6.5\, m \frac{\varepsilon}{\sigma} \quad \text{if } \varepsilon \le \frac{\sigma}{2m}, \qquad (*)$$

  and the same upper bound holds for the relative volume of the inner ε-neighborhood of ∂K.
- The proof idea is similar to the previously mentioned (volume of tubes, integral geometry, counting argument).
- In particular, Poincaré's formula implies

$$\frac{\mathrm{vol}(\partial K)}{\mathrm{vol}(\mathbb{S}')} \le 1.$$

  Indeed, by convexity, the intersection of ∂K with a hyperequator $\mathbb{S}'$ of $\mathbb{S}$ in general position consists of at most two points.

Sketch of proof (2)

- **Crucial Lemma.** Let $A = (a_1, \ldots, a_n) \in \mathcal{F}_D^\circ$ and $\mathscr{C}(A) \geq m\varepsilon^{-1}$. Then there exists $i \in \{1, \ldots, n\}$ such that $a_i \in T_o(\partial K_i, \varepsilon)$, where $-K_i$ is the spherical convex hull of $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$.
- The Lemma yields with $t = m/\varepsilon$

$$\mathrm{Prob}\{A \in \mathcal{F}_D^\circ,\ \mathscr{C}(A) \geq t\} \leq \sum_{i=1}^{n} \mathrm{Prob}\{A \in \mathcal{F}_D^\circ,\ a_i \in T_o(\partial K_i, \varepsilon)\}.$$

  Note that $B(\overline{A}, \sigma) = B(\overline{A}', \sigma) \times B(\overline{a}_n, \sigma)$ where $\overline{A'} := (\overline{a}_1, \ldots, \overline{a}_{n-1})$.
- We bound the probability on the right-hand side for $i = n$ by an integral of probabilities conditioned on $A' := (a_1, \ldots, a_{n-1})$:

$$\mathrm{Prob}\{A' \in \mathcal{F}_D^\circ \text{ and } a_n \in T_o(\partial K_n, \varepsilon)\}$$
$$= \frac{1}{\mathrm{vol} B(\overline{A'}, \sigma)} \int_{A' \in \mathcal{F}_D^\circ \cap B(\overline{A'}, \sigma)} \mathrm{Prob}\{a_n \in T_o(\partial K_n, \varepsilon) \mid A'\}\, dA'.$$

Sketch of proof (3)

- Fix now $A' \in \mathcal{F}_{n-1,m}$ and consider the convex set K_n in $\mathbb{S}$. The volume bound (*) yields

$$\mathrm{Prob}\{a_n \in T_o(\partial K_n, \varepsilon) \mid A'\} = \frac{\mathrm{vol}(T_o(\partial K_n, \varepsilon) \cap B(\overline{a}_n, \sigma))}{\mathrm{vol} B(\overline{a}_n, \sigma)} \leq 6.5\, m\, \frac{\varepsilon}{\sigma}.$$

We conclude that

$$\mathrm{Prob}\{A \in \mathcal{F}_D^\circ,\ a_n \in T_o(\partial K_n, \phi)\} \ \leq \ 6.5\, m\, \frac{\varepsilon}{\sigma}.$$

- The same upper bound holds for any K_i. Altogether, we obtain

$$\mathrm{Prob}\{A \in \mathcal{F}_D^\circ \text{ and } \mathscr{C}(A) \geq t\} \leq 6.5\, nm^2\, \frac{1}{\sigma t},$$

$\square$

Condition Numbers of Convex Optimization

Convex homogeneous feasibility problem

- Much of what has been said for linear optimization can be generalized to convex optimization.
- Fix a closed regular convex cone $C \subseteq \mathbb{R}^n$ with dual cone

$$\breve{C} := \{y \in \mathbb{R}^n \mid \forall x \in C : \langle y, x \rangle \geq 0\}$$

- Homogeneous convex feasibility problem (HCFP)

  Input $A \in \mathbb{R}^{m \times n}$ $(n > m)$

  Decide the alternative

$$\exists x \in \mathbb{R}^n \setminus \{0\} : \quad Ax = 0,\ x \in \breve{C} \qquad \text{(P)}$$

$$\exists y \in \mathbb{R}^m \setminus \{0\} : \quad A^T y \in C \qquad \text{(D)}$$

Convex homogeneous feasibility problem

- Most important cases:

$$\begin{array}{lcl} \text{Linear Programming} & : & C = \mathbb{R}^n_+ = \mathbb{R}_+ \times \ldots \times \mathbb{R}_+ \\ \text{Semidefinite Programming} & : & C = \{M \in \mathbb{R}^{\ell \times \ell},\ M \text{ is pos. semidef.}\} \end{array}$$

- Define

$$\begin{aligned} \mathcal{F}_P &:= \{A \mid \text{(P) is feasible}\}, \\ \mathcal{F}_D &:= \{A \mid \text{(D) is feasible}\}, \\ \Sigma &:= \mathcal{F}_P \cap \mathcal{F}_D. \end{aligned}$$

- Renegar's condition number is defined as:

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

Convex homogeneous feasibility problem

- The probabilistic analyses for LP-condition numbers relie on the product structure of the cone $C = \mathbb{R}^n_+ = \mathbb{R}_+ \times \ldots \times \mathbb{R}_+$.
- For general cones (like SDP), we look for a different, more coordinate-free approach.
- Suppose $A \in \mathbb{R}^{m\times n}$ has rank m. Consider the m-dimensional linear subspace $W := \mathrm{im} A^T$ of $\mathbb{R}^n$.

$$\exists x \in \mathbb{R}^n \setminus \{0\} : \quad \begin{array}{l} Ax = 0 \\ x \in \breve{C} \end{array} \qquad \text{(P)}$$

$$\Leftrightarrow$$

$$\underbrace{\ker A}_{=W^\perp} \cap \breve{C} \neq \{0\}$$

$$\exists y \in \mathbb{R}^m \setminus \{0\} : \quad A^T y \in C \qquad \text{(D)}$$

$$\Leftrightarrow$$

$$\underbrace{\mathrm{im} A^T}_{=:W} \cap C \neq \{0\}$$

Grassmann condition number (1)

- Consider the inputs as an element of the Grassmann manifold

$$W \in \mathbb{G}_{n,m} := \{W \subseteq \mathbb{R}^n \mid W \text{ lin. subspace}\,,\ \dim W = m\}.$$

We have to decide the alternative

$$W^{\perp} \cap \breve{C} \neq \{0\} \quad \text{(P)} \qquad \text{or} \qquad W \cap C \neq \{0\} \quad \text{(D)}$$

- Define

$$\begin{array}{lcll} \mathcal{F}_P & := & \{W \in \mathbb{G}_{n,m} \mid W^{\perp} \cap \breve{C} \neq \{0\}\} & \text{(primal feasible)} \\ \mathcal{F}_D & := & \{W \in \mathbb{G}_{n,m} \mid W \cap C \neq \{0\}\} & \text{(dual feasible)} \\ \Sigma_{\mathbb{G}} & := & \mathcal{F}_P \cap \mathcal{F}_D & \text{(ill-posed)} \end{array}$$

- $\mathbb{G}_{n,m}$ is a compact Riemannian manifold. We have thus well-defined notions of (geodesic) distance ("angle") and volume.

Grassmann condition number (2)

- We define the Grassmann condition number for $W \in \mathbb{G}_{n,m}$ as

$$\mathscr{C}_{\mathbb{G}}(W) \;:=\; \frac{1}{\sin d(W, \Sigma_{\mathbb{G}})},$$

where d denotes the geodesic distance in $\mathbb{G}_{n,m}$.

- The following result (Amelunxen, Belloni & Freund) separates Renegar's condition number into an "intrinsic" and "extrinsic" part.

For $A \in \mathbb{R}^{m\times n}$ of rank m and $W := \mathrm{im} A^T$ we have

$$\mathscr{C}_{\mathbb{G}}(A) \;\le\; \mathscr{C}_R(A) \;\le\; \kappa(A) \cdot \mathscr{C}_{\mathbb{G}}(A)\,.$$

Average analysis of Grassmann condition number

- Fix any closed regular convex cone $C \subset \mathbb{R}^n$.
- If $A \in \mathbb{R}^{m\times n}$ is standard Gaussian, then $W := \mathrm{im} A^T$ is uniformly distributed in $\mathbb{G}_{n,m}$ (w.r.t. orthogonal invariant volume form).
- With the volume of tube interpretation and some differential geometry, B and Amelunxen showed

$$\mathrm{Prob}\left(\mathscr{C}_{\mathbb{G}}(A) \geq \frac{1}{\varepsilon}\right) \;\leq\; 6 \cdot n\,\varepsilon \quad \text{if } \varepsilon < n^{-\frac{3}{2}}.$$

$$\mathbb{E}(\ln \mathscr{C}_{\mathbb{G}}(A)) \;\leq\; 2.5 \cdot \ln(n) + 2.8.$$

- We are currently extending this result to a uniform smoothed analysis.

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

- Let us explain this question in detail.
- For a degree vector $d = (d_1, \ldots, d_n)$ we define

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

- The input size is $N := \dim_{\mathbb{C}} \mathcal{H}_d$.
- We look for zeros ζ 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 ζ 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 ζ 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
- By the implicit function theorem, the projection map $V \to \mathbb{P}(\mathcal{H}_d), (f', \zeta') \mapsto f'$ can be locally inverted around (f, ζ) if ζ is a simple solution of f. The solution map G is the local inverse of this projection.
- The condition number of f at (f, ζ),

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

is essentially the operator norm of the derivative of G at ζ, where

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

(choose representative of ζ 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 $d(x,\zeta) \leq \frac{0.3}{D^{3/2}\,\mu(f,\zeta)}$, then x is an approximate zero of f associated with ζ.

Adaptive linear homotopy continuation

- Given a start system $(g, \zeta) \in V$ and an input $f \in \mathcal{H}_d$.
- Consider the line segment $[g, f]$ connecting g and f that consists of the systems
$$q_t := (1-t)g + tf \qquad \text{for } t \in [0,1].$$
- 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
$$\gamma\colon [0,1] \to V, t \mapsto (f_t, \zeta_t)$$
such that $f_0 = g$.
- The idea is to follow the path γ numerically: partition $[0,1]$ into $t_0 = 0, \ldots, t_k = 1$. Writing $q_i := q_{t_i}$, successively compute approximations z_i of ζ_{t_i} by Newton's method starting with $z_0 := \zeta$. More specifically, compute
$$z_{i+1} := N_{q_{i+1}}(z_i).$$

Complexity of adaptive linear homotopy continuation

- We 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 ζ_i for all i. Moreover,

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

-

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, ζ).
- Since we don't know how to construct a good start system (g, ζ_0), we choose it at random:
  - choose $g \in \mathcal{H}_d$ from standard Gaussian,
  - choose one of the $\mathcal{D}$ many zeros ζ of g uniformly at random.

  Here $\mathcal{D} := d_1 \cdots d_n$ is the Bezout number.
- Efficient sampling of (g, ζ) is possible (Beltrán & Pardo 2008).
- Las Vegas Algorithm LV
  draw $(g, \zeta) \in V$ at random
  run ALH on input (f, g, ζ)
- 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 σ^2.
- Recently, I obtained with F. Cucker the following

Smoothed analysis of ALH

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

- Byproduct of this work

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, ζ), where

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

(the zeros of g consist of roots of unity). We have

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

- 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.

Proof idea for smoothed analysis of ALH (1)

- A main innovation is the systematic use of Gaussians, which were previously not used in this context.
- Consider the mean square condition number

$$\mu_2(q)^2 := \frac{1}{\mathcal{D}} \sum_{\zeta \in V(q)} \mu(q,\zeta)^2 \quad \text{for } g \in \mathcal{H}_d.$$

- The analysis of ALH gives

$$\begin{aligned} \mathbb{E}_{\zeta \in V(g)} K(f,g,\zeta) &\leq c\, D^{3/2} \int_0^1 \mu(q_t)^2 \, \|\dot{\gamma}(t)\| \, dt \\ &\leq c\, D^{3/2} \int_0^1 \mu(q_t)^2 \, \frac{\|f\| \cdot \|g\|}{\|q_t\|^2} \, dt. \end{aligned}$$

- $\mathbb{E}(\|f\|^2) = 2N$ (chi-square). Replace $\|f\|$ by $\sqrt{N}$ (cheating a bit).

Proof idea for smoothed analysis of ALH (2)

- $$\mathbb{E}_{\zeta\in V(g)}K(f,g,\zeta) \leq c\, D^{3/2}N\int_0^1 \frac{\mu_2(q_t)^2}{\|q_t\|^2}\,dt.$$
- By Fubini,
$$\mathbb{E}_{f\sim N(\overline{f},\sigma^2 I)}E_{g\sim N(0,I)}\mathbb{E}_{\zeta\in V(g)}K(f,g,\zeta) \leq c\, D^{3/2}N\int_0^1 \mathbb{E}\Big(\frac{\mu_2(q_t)^2}{\|q_t\|^2}\Big)\,dt.$$
- For fixed t, $q_t=(1-t)g+tf$ is again Gaussian, $q_t\sim N(\overline{q}_t,\sigma_t^2 I)$, with
$$\overline{q}_t = t\overline{f},\quad \sigma_t^2=(1-t)^2+\sigma^2t^2.$$

Proof idea for smoothed analysis of ALH (3)

Main technical contribution of proof

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

Using this,

$$\mathbb{E}_{f\sim N(\overline{f},\sigma^2 I)} K(f) \;\le\; c\,D^{3/2} N \int_0^1 \frac{n}{(1-t)^2+\sigma^2 t^2}\,dt = c\,D^{3/2} N\,\frac{n}{\sigma}. \quad \square$$

On proving the main technical contribution (1)

- Put $\mathcal{M} := \mathbb{C}^{n\times(n+1)}$ and consider the map (slightly cheating ...)

$$\Psi\colon V \to \mathcal{M}, (q,\zeta) \mapsto M := \mathrm{diag}(\sqrt{d_1},\dots,\sqrt{d_n})^{-1} Df(\zeta).$$

  Recall $\mu(q,\zeta)/\|q\| = \|M^\dagger\|$.

- The noncentered Gaussian on $\mathcal{H}_d$ defines a distribution on V (choose q and then one of its $\mathcal{D}$ zeros uniformly at random). Then

$$\mathbb{E}_{\mathcal{H}_d}\Big(\frac{\mu_2(q)^2}{\|q\|^2}\Big) = \mathbb{E}_V\Big(\frac{\mu(q,\zeta)^2}{\|q\|^2}\Big) = \mathbb{E}_{\mathcal{M}}(\|M^\dagger\|^2)$$

  where the last expectation is w.r.t. the distribution on $\mathcal{M}$ induced by Ψ.

On proving the main technical contribution (2)

- For $\zeta \in \mathbb{P}^n$ let R_ζ be the set of those $q \in \mathcal{H}_d$ that vanish at ζ of order> 1.
- Further, let L_ζ be the orthogonal complement in R_ζ in the space of $q \in \mathcal{H}_d$ vanishing at ζ.
- We obtain an orthogonal decomposition

$$\mathcal{H}_d = C_\zeta \oplus L_\zeta \oplus R_\zeta, \quad \overline{q} = \overline{k}_\zeta + \overline{g}_\zeta + \overline{h}_\zeta.$$

- The density of $N(\overline{q}, \sigma^2 I)$ factors into Gaussians:

$$\rho_{\mathcal{H}_d}(k + g + h) = \rho_{C_\zeta}(k) \cdot \rho_{L_\zeta}(g) \cdot \rho_{R_\zeta}(h).$$

- L_ζ is isometrically isomorphic to $\mathcal{M}_\zeta := \{M \in \mathcal{M} \colon M\zeta = 0\}$ inducing a Gaussian $N(\overline{M}_\zeta, \sigma^2 I)$ on the fiber $\mathcal{M}_\zeta$.

On proving the main technical contribution (3)

- For $M \in \mathcal{M}$ of full rank with zero ζ one can show that

$$\rho_{\mathcal{M}}(M) = \rho_{C_\zeta}(0) \cdot \rho_{\mathcal{M}_\zeta}(M)$$

- With the coarea formula (transformation of integrals) one shows

$$\begin{aligned} \mathbb{E}_{\mathcal{M}}(\|M^\dagger\|^2) &= \int_{\mathcal{M}} \|M^\dagger\|^2 \rho_{\mathcal{M}}(M)\, dM \\ &= \mathbb{E}_{\zeta \in \mathbb{P}^n}\Big(\mathbb{E}_{\widetilde{\rho}_{\mathcal{M}_\zeta}}(\|M^\dagger\|^2)\Big) \end{aligned}$$

  First expection is over induced distribution of the zeros ζ of M, second expectation is w.r.t. the following conditional density on $\mathcal{M}_\zeta$:

$$\widetilde{\rho}_{\mathcal{M}_\zeta}(M) = c_\zeta\, \rho_{\mathcal{M}_\zeta}(M) \det(MM^*).$$

- As for the smoothed analysis of matrix condition numbers one can show

$$\mathbb{E}_{\widetilde{\rho}_{\mathcal{M}_\zeta}}\big(\|M^\dagger\|^2\big)\Big) = \mathcal{O}\Big(\frac{n}{\sigma^2}\Big).$$

- Hence $\mathbb{E}_{\mathcal{M}}(\|M^\dagger\|^2) = \mathcal{O}\Big(\frac{n}{\sigma^2}\Big)$. □

Thank you for your attention!

References

[1] E.L. Allgower and K. Georg. *Numerical Continuation Methods*. Springer-Verlag, 1990.

[2] D. Amelunxen. *Geometric analysis of the condition of the convex feasibility problem*. PhD thesis, University of Paderborn, 2011.

[3] D. Amelunxen and P. Bürgisser. Probabilistic analysis of the Grassmann condition number. *Found. Comput. Math.*, 15(1): 3–51, 2015.

[4] D. Amelunxen and P. Bürgisser. A coordinate-free condition number for convex programming. *SIAM J. Optim.*, 22(3):1029–1041, 2012.

[5] D. Amelunxen and P. Bürgisser. Robust smoothed analysis of a condition number for linear programming. *Math. Programming*, 131(1, Ser. A):221–251, 2012.

[6] D. Amelunxen and M. Lotz. Average-case analysis without the black swans. Accepted for *J. Compl.*, arXiv:1512.09290.

[7] D. Armentano. Stochastic perturbations and smooth condition numbers. *J. Complexity*, 26(2):161–171, 2010.

[8] D. Armentano. Complexity of path-following methods for the eigenvalue problem. *Found. Comput. Math.*, 14(2): 185–236, 2014.

[9] D. Armentano, C. Beltrán, P. Bürgisser, F. Cucker, and M. Shub. Condition length and complexity for the solution of polynomial systems. *Found. Comput. Math.*, 16(6): 1401–1422, 2016.

[10] E. Barbier. Note sur le problème de l'aguille et le jeu du joint couvert. *J. Math. Pures et Appl.*, 5(2):273–286, 1860.

[11] V. Bargmann, D. Montgomery, and J. von Neumann. Solution of linear systems of high order (Princeton, 1946). In A.H. Taub, editor, *John von Neumann Collected Works*, volume 5. Pergamon, Elmsford, NY, 1963.

[12] D.J. Bates, J.D. Hauenstein, A.J. Sommese, and C.W. Wampler. Software for numerical algebraic geometry: a paradigm and progress towards its implementation. In *Software for algebraic geometry*, volume 148 of *IMA Vol. Math. Appl.*, pages 1–14. Springer, New York, 2008.

[13] A. Belloni, R.M. Freund, and S. Vempala. An efficient rescaled perceptron algorithm for conic systems. *Math. Oper. Res.*, 34:621–641, 2009.

[14] C. Beltrán. A continuation method to solve polynomial systems and its complexity. *Numer. Math.*, 117(1):89–113, 2011.

[15] C. Beltrán, J.-P. Dedieu, G. Malajovich, and M. Shub. Convexity properties of the condition number II. *SIAM J. Matrix Anal. Appl.*, 33(3): 905–939, 2012.

[16] C. Beltrán, J.-P. Dedieu, G. Malajovich, and M. Shub. Convexity properties of the condition number. *SIAM J. Matrix Anal. Appl.*, 31(3):1491–1506, 2009.

[17] C. Beltrán and A. Leykin. Certified numerical homotopy tracking. *Experimental Mathematics*, 21(1):69–83, 2012.

[18] C. Beltrán and L. M. Pardo. Estimates on the distribution of the condition number of singular matrices. *Found. Comput. Math.*, 7(1):87–134, 2007.

[19] C. Beltrán and L.M. Pardo. On the complexity of non universal polynomial equation solving: old and new results. In *Foundations of computational mathematics, Santander 2005*, volume 331 of *London Math. Soc. Lecture Note Ser.*, pages 1–35. Cambridge Univ. Press, Cambridge, 2006.

[20] C. Beltrán and L.M. Pardo. On Smale's 17 problem: a probabilistic positive solution. *Found. Comput. Math.*, 8:1–43, 2008.

[21] C. Beltrán and L.M. Pardo. Smale's 17th problem: average polynomial time to compute affine and projective solutions. *J. Amer. Math. Soc.*, 22(2):363–385, 2009.

[22] C. Beltrán and L.M. Pardo. Fast linear homotopy to find approximate zeros of polynomial systems. *Found. Comput. Math.*, 11(1):95–129, 2011.

[23] C. Beltrán and M. Shub. Complexity of Bézout's Theorem VII: distance estimates in the condition metric. *Found. Comput. Math.*, 9:179–195, 2009.

[24] C. Beltrán and M. Shub. On the geometry and topology of the solution variety for polynomial system solving. *Found. Comput. Math.*, 12(6): 719–763, 2012.

[25] Å. Björck. Component-wise perturbation analysis and error bounds for linear least squares solutions. *BIT*, 31(2):238–244, 1991.

[26] L. Blum. Lectures on a theory of computation and complexity over the reals (or an arbitrary ring). In E. Jen, editor, *Lectures in the Sciences of Complexity II*, pages 1–47. Addison-Wesley, 1990.

[27] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and real computation*. Springer-Verlag, New York, 1998. With a foreword by R.M. Karp.

[28] L. Blum and M. Shub. Evaluating rational functions: infinite precision is finite cost and tractable on average. *SIAM J. Comput.*, 15(2):384–398, 1986.

[29] Comte de Buffon, G.-L. Leclerc. Essai d'arithmétique morale. In *Supplément à l'Histoire naturelle*, volume 4, pages 46–148. Imprimerie Royale, Paris, 1777.

[30] P. Bürgisser. Condition of intersecting a projective variety with a varying linear subspace. *SIAM Journal on Applied Algebra and Geometry*, 2017. To appear, arXiv:1510.04142v2.

[31] P. Bürgisser and F. Cucker. Smoothed analysis of Moore-Penrose inversion. *SIAM J. Matrix Anal. Appl.*, 31(5):2769–2783, 2010.

[32] P. Bürgisser and F. Cucker. On a problem posed by Steve Smale. *Annals of Mathematics*, 174:1785–1836, 2011.

[33] P. Bürgisser and F. Cucker. *Condition: The geometry of numerical algorithms*, volume 349 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Heidelberg, 2013.

[34] P. Bürgisser, F. Cucker, and E.R. Cardozo. On the condition of the zeros of characteristic polynomials. arXiv:1510.04419.

[35] P. Bürgisser, F. Cucker, and M. Lotz. Smoothed analysis of complex conic condition numbers. *J. Math. Pures Appl. (9)*, 86(4):293–309, 2006.

[36] P. Bürgisser, F. Cucker, and M. Lotz. The probability that a slightly perturbed numerical analysis problem is difficult. *Mathematics of Computation*, 77:1559–1583, 2008.

[37] P. Bürgisser, F. Cucker, and M. Lotz. Coverage processes on spheres and condition numbers for linear programming. *Annals of Probability*, 38:570–604, 2010.

[38] Z.-Z. Cheng and J.J. Dongarra. Condition numbers of Gaussian random matrices. *SIAM J. Matrix Anal. Appl.*, 27:603–620, 2005.

[39] S. Chern. On the kinematic formula in integral geometry. *J. Math. Mech.*, 16:101–118, 1966.

[40] D. Cheung and F. Cucker. Smoothed analysis of componentwise condition numbers for sparse matrices. *IMA J. Numer. Anal.*, 35(1):74–88, 2015.

[41] D. Cheung and F. Cucker. A new condition number for linear programming. *Math. Program.*, 91(1, Ser. A):163–174, 2001.

[42] D. Cheung and F. Cucker. Probabilistic analysis of condition numbers for linear programming. *Journal of Optimization Theory and Applications*, 114:55–67, 2002.

[43] D. Cheung and F. Cucker. Solving linear programs with finite precision: I. Condition numbers and random programs. *Math. Program.*, 99:175–196, 2004.

[44] D. Cheung and F. Cucker. Solving linear programs with finite precision: II. Algorithms. *Journal of Complexity*, 22:305–335, 2006.

[45] D. Cheung and F. Cucker. Componentwise condition numbers of random sparse matrices. *SIAM J. Matrix Anal. Appl.*, 31:721–731, 2009.

[46] D. Cheung and F. Cucker. On the average condition of random linear programs. *SIAM J. Optim.*, 23(2): 799–810, 2013.

[47] D. Cheung, F. Cucker, and R. Hauser. Tail decay and moment estimates of a condition number for random linear conic systems. *SIAM J. Optim.*, 15(4):1237–1261 (electronic), 2005.

[48] D. Cheung, F. Cucker, and J. Peña. Unifying condition numbers for linear programming. *Math. Oper. Res.*, 28(4):609–624, 2003.

[49] D. Cheung, F. Cucker, and Ye. Y. Linear programming and condition numbers under the real number computation model. In Ph. Ciarlet and F. Cucker, editors, *Handbook of Numerical Analysis*, volume XI, pages 141–207. North-Holland, 2003.

[50] F. Cucker. Approximate zeros and condition numbers. *J. of Complexity*, 15:214–226, 1999.

[51] F. Cucker, H. Diao, and Y. Wei. On mixed and componentwise condition numbers for Moore-Penrose inverse and linear least squares problems. *Mathematics of Computation*, 76:947–963, 2007.

[52] F. Cucker and J. Peña. A primal-dual algorithm for solving polyhedral conic systems with a finite-precision machine. *SIAM J. Optim.*, 12(2):522–554 (electronic), 2001/02.
[53] F. Cucker and S. Smale. Complexity estimates depending on condition and round-off error. *Journal of the ACM*, 46:113–184, 1999.
[54] F. Cucker and M. Wschebor. On the expected condition number of linear programming problems. *Numer. Math.*, 94:419–478, 2003.
[55] J.-P. Dedieu. *Points fixes, zéros et la méthode de Newton*, volume 54 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Berlin, 2006. With a preface by Steve Smale.
[56] J.-P. Dedieu, G. Malajovich, and M. Shub. On the curvature of the central path of linear programming theory. *Found. Comput. Math.*, 5(2):145–171, 2005.
[57] J.-P. Dedieu, P. Priouret, and G. Malajovich. Newton's method on Riemannian manifolds: convariant alpha theory. *IMA J. Numer. Anal.*, 23(3):395–419, 2003.
[58] J.P. Dedieu. Condition number analysis for sparse polynomial systems. In *Foundations of computational mathematics (Rio de Janeiro, 1997)*, pages 75–101. Springer, Berlin, 1997.
[59] J.W. Demmel. On condition numbers and the distance to the nearest ill-posed problem. *Numer. Math.*, 51:251–289, 1987.
[60] J.W. Demmel. The probability that a numerical analysis problem is difficult. *Math. Comp.*, 50:449–480, 1988.
[61] J.W. Demmel. *Applied Numerical Linear Algebra*. SIAM, 1997.
[62] J. Dunagan, D.A. Spielman, and S.-H. Teng. Smoothed analysis of condition numbers and complexity implications for linear programming. *Math. Program.*, 126(2, Ser. A):315–350, 2011.
[63] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.
[64] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.
[65] A. Edelman. On the distribution of a scaled condition number. *Math. Comp.*, 58(197):185–190, 1992.
[66] M. Epelman and R.M. Freund. A new condition measure, preconditioners, and relations between different measures of conditioning for conic linear systems. *SIAM J. Optim.*, 12(3):627–655 (electronic), 2002.
[67] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.
[68] R. Fletcher. Expected conditioning. *IMA J. Numer. Anal.*, 5(3):247–273, 1985.
[69] R.M. Freund and J.R. Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM J. Optim.*, 10(1):155–176 (electronic), 1999.
[70] R.M. Freund and J.R. Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Math. Program.*, 86:225–260, 1999.
[71] F. Gao, D. Hug, and R. Schneider. Intrinsic volumes and polar sets in spherical space. *Math. Notae*, 41:159–176 (2003), 2001/02. Homage to Luis Santaló. Vol. 1 (Spanish).
[72] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.
[73] A.J. Geurts. A contribution to the theory of condition. *Numer. Math.*, 39:85–96, 1982.
[74] S. Glasauer. Integral geometry of spherically convex bodies. *Diss. Summ. Math.*, 1(1-2):219–226, 1996.
[75] J.-L. Goffin. The relaxation method for solving systems of linear inequalities. *Math. Oper. Res.*, 5(3):388–414, 1980.
[76] I. Gohberg and I. Koltracht. Mixed, componentwise, and structured condition numbers. *SIAM J. Matrix Anal. Appl.*, 14:688–704, 1993.
[77] H.H. Goldstine and J. von Neumann. Numerical inverting matrices of high order, II. *Proc. Amer. Math. Soc.*, 2:188–202, 1951.
[78] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.

[79] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics: Study and Research Texts*. Springer-Verlag, Berlin, 1988.

[80] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436 (1953), 1952.

[81] N.J. Higham. *Accuracy and stability of numerical algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2002.

[82] H. Hotelling. Some new methods in matrix calculation. *Ann. Math. Statistics*, 14:1–34, 1943.

[83] R. Howard. The kinematic formula in Riemannian homogeneous spaces. *Mem. Amer. Math. Soc.*, 106(509):vi+69, 1993.

[84] W. Kahan. Numerical linear algebra. *Canad. Math. Bull.*, 9:757–801, 1966.

[85] L.V. Kantorovich. *On Newton's method*, volume 28 of *Trudy Mat. Inst. Steklov.*, pages 104–144. Acad. Sci. USSR, Moscow–Leningrad, 1949. In Russian.

[86] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[87] M. Karow, D. Kressner, and F. Tisseur. Structured eigenvalue condition numbers. *SIAM J. Matrix Anal. Appl.*, 28(4):1052–1068 (electronic), 2006.

[88] L.G. Khachiyan. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR*, 244:1093–1096, 1979. (In Russian, English translation in *Soviet Math. Dokl.*, 20:191–194, 1979.).

[89] D.A. Klain and G.-C. Rota. *Introduction to geometric probability*. Lezioni Lincee. [Lincei Lectures]. Cambridge University Press, Cambridge, 1997.

[90] E. Kostlan. On the distribution of the roots of random polynomials. In M. Hirsch, J.E. Marsden, and M. Shub, editors, *From Topology to Computation: Proceedings of the Smalefest*, pages 419–431. Springer-Verlag, 1993.

[91] P. Lairez. A Deterministic Algorithm to Compute Approximate Roots of Polynomial Systems in Polynomial Average Time. To appear in *Found. Comput. Math.*

[92] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[93] T.Y. Li. Numerical solution of polynomial systems by homotopy continuation methods. In Ph. Ciarlet and F. Cucker, editors, *Handbook of Numerical Analysis*, volume XI, pages 209–304. North-Holland, Amsterdam, 2003.

[94] J. De Loera, B. Sturmfels, and C. Vinzant. The central curve of linear programming. *Found. Comput. Math.*, 12(4): 509–540, 2012.

[95] M. Lotz. On the volume of tubular neighborhoods of real algebraic varieties. *Proc. Amer. Math. Soc.*, 143(5): 1875–1889, 2015.

[96] G. Malajovich and J.M. Rojas. High probability analysis of the condition number of sparse polynomial systems. *Theoret. Comput. Sci.*, 315(2-3):524–555, 2004.

[97] T. Motzkin and I.Y. Schönberg. The relaxation method for linear inequalitites. *Canadian Journal of Mathematics*, 6:393–404, 1954.

[98] R.J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.

[99] Y. Nesterov and A. Nemirovsky. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.

[100] J. von Neumann and H.H. Goldstine. Numerical inverting matrices of high order. *Bull. Amer. Math. Soc.*, 53:1021–1099, 1947.

[101] M. Nunez and R.M. Freund. Condition measures and properties of the central trajectory of a linear program. *Math. Program.*, 83:1–28, 1998.

[102] V.Y. Pan. Optimal and nearly optimal algorithms for approximating polynomial zeros. *Comput. Math. Appl.*, 31(12):97–138, 1996.

[103] V.Y. Pan. Solving a polynomial equation: some history and recent progress. *SIAM Rev.*, 39(2):187–220, 1997.

[104] J. Peña and J. Renegar. Computing approximate solutions for conic systems of constraints. *Math. Program.*, 87:351–383, 2000.

[105] J. Peña. Understanding the geometry on infeasible perturbations of a conic linear system. *SIAM J. Optim.*, 10:534–550, 2000.

[106] J. Peña. A characterization of the distance to infeasibility under block-structured perturbations. *Linear Algebra Appl.*, 370:193–216, 2003.

[107] J. Renegar. On the efficiency of Newton's method in approximating all zeros of a system of complex polynomials. *Math. Oper. Res.*, 12(1):121–148, 1987.

[108] J. Renegar. On the worst-case arithmetic complexity of approximating zeros of systems of polynomials. *SIAM J. Comput.*, 18:350–370, 1989.

[109] J. Renegar. Is it possible to know a problem instance is ill-posed? *Journal of Complexity*, 10:1–56, 1994.

[110] J. Renegar. Some perturbation theory for linear programming. *Math. Program.*, 65:73–91, 1994.

[111] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5:506–524, 1995.

[112] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Program.*, 70:279–351, 1995.

[113] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization.* SIAM, 2000.

[114] J.R. Rice. A theory of condition. *SIAM J. Numer. Anal.*, 3:217–232, 1966.

[115] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739, 2009.

[116] S.M. Rump. Structured perturbations part I: normwise distances. *SIAM J. Matrix Anal. Appl.*, 25:1–30, 2003.

[117] S.M. Rump. Structured perturbations part II: componentwise distances. *SIAM J. Matrix Anal. Appl.*, 25:31–56, 2003.

[118] A. Sankar, D.A. Spielman, and S.-H. Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476 (electronic), 2006.

[119] L.A. Santaló. *Integral geometry and geometric probability.* Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1976. With a foreword by Mark Kac, Encyclopedia of Mathematics and its Applications, Vol. 1.

[120] R. Schneider and W. Weil. *Stochastic and integral geometry.* Probability and its Applications (New York). Springer-Verlag, Berlin, 2008.

[121] M. Shub. Some remarks on Bezout's theorem and complexity theory. In *From Topology to Computation: Proceedings of the Smalefest (Berkeley, CA, 1990)*, pages 443–455. Springer, New York, 1993.

[122] M. Shub. Complexity of Bézout's Theorem VI: geodesics in the condition (number) metric. *Found. Comput. Math.*, 9:171–178, 2009.

[123] M. Shub and S. Smale. Computational complexity. On the geometry of polynomials and a theory of cost. I. *Ann. Sci. École Norm. Sup. (4)*, 18(1):107–142, 1985.

[124] M. Shub and S. Smale. Computational complexity: on the geometry of polynomials and a theory of cost. II. *SIAM J. Comput.*, 15(1):145–161, 1986.

[125] M. Shub and S. Smale. Complexity of Bézout's Theorem I: geometric aspects. *J. Amer. Math. Soc.*, 6:459–501, 1993.

[126] M. Shub and S. Smale. Complexity of Bézout's Theorem II: volumes and probabilities. In F. Eyssette and A. Galligo, editors, *Computational Algebraic Geometry*, volume 109 of *Progress in Mathematics*, pages 267–285. Birkhäuser, 1993.

[127] M. Shub and S. Smale. Complexity of Bézout's Theorem III: condition number and packing. *Journal of Complexity*, 9:4–14, 1993.

[128] M. Shub and S. Smale. Complexity of Bézout's Theorem V: polynomial time. *Theoretical Computer Science*, 133:141–164, 1994.

[129] M. Shub and S. Smale. Complexity of Bézout's Theorem IV: probability of success; extensions. *SIAM J. of Numer. Anal.*, 33:128–148, 1996.

[130] J.W. Silverstein. The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.

[131] S. Smale. The fundamental theorem of algebra and complexity theory. *Bull. Amer. Math. Soc.*, 4:1–36, 1981.

[132] S. Smale. Newton's method estimates from data at one point. In R. Ewing, K. Gross, and C. Martin, editors, *The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics*. Springer-Verlag, 1986.

[133] S. Smale. Complexity theory and numerical analysis. In A. Iserles, editor, *Acta Numerica*, pages 523–551. Cambridge University Press, 1997.

[134] S. Smale. Mathematical problems for the next century. *Math. Intelligencer*, 20(2):7–15, 1998.

[135] A.J. Sommese and C.W. Wampler, II. *The numerical solution of systems of polynomials*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005. Arising in engineering and science.

[136] D.A. Spielman and S.-H. Teng. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, pages 296–305 (electronic), New York, 2001. ACM.

[137] D.A. Spielman and S.-H. Teng. Smoothed analysis of algorithms. In *Proceedings of the International Congress of Mathematicians*, volume I, pages 597–606, 2002.

[138] D.A. Spielman and S.-H. Teng. Smoothed analysis: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004.

[139] J. Steiner. Über parallele Flächen. *Monatsber. preuss. Akad. Wiss.*, pages 114–118, 1840.

[140] G.W. Stewart. On the perturbation of pseudo-inverses, projections and linear least squares problems. *SIAM Rev.*, 19(4):634–662, 1977.

[141] G.W. Stewart. Stochastic perturbation theory. *SIAM Rev.*, 32(4):579–610, 1990.

[142] G.W. Stewart and J.-G. Sun. *Matrix perturbation theory*. Computer Science and Scientific Computing. Academic Press Inc., Boston, MA, 1990.

[143] T. Tao and V. Vu. Smooth analysis of the condition number and the least singular value. *Math. Comp.*, 79(272):2333–2352, 2010.

[144] L.N. Trefethen and R.S. Schreiber. Average-case stability of Gaussian elimination. *SIAM J. Matrix Anal. Appl.*, 11:335–360, 1990.

[145] A.M. Turing. Rounding-off errors in matrix processes. *Quart. J. Mech. Appl. Math.*, 1:287–308, 1948.

[146] S.A. Vavasis and Y. Ye. Condition numbers for polyhedra with real number data. *Oper. Res. Lett.*, 17:209–214, 1995.

[147] S.A. Vavasis and Y. Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Math. Program.*, 74:79–120, 1996.

[148] D. Viswanath and L.N. Trefethen. Condition numbers of random triangular matrices. *SIAM J. Matrix Anal. Appl.*, 19:564–581, 1998.

[149] P.-Å. Wedin. Perturbation theory for pseudo-inverses. *Nordisk Tidskr. Informationsbehandling (BIT)*, 13:217–232, 1973.

[150] N. Weiss, G.W. Wasilkowski, H. Woźniakowski, and M. Shub. Average condition number for solving linear equations. *Linear Algebra Appl.*, 83:79–102, 1986.

[151] J.G. Wendel. A problem in geometric probability. *Math. Scand.*, 11:109–111, 1962.

[152] H. Weyl. On the volume of tubes. *Amer. J. Math.*, 61(2):461–472, 1939.

[153] E. Wigner. Random matrices in physics. *SIAM Rev.*, 9:1–23, 1967.

[154] J.H. Wilkinson. Error analysis of direct methods of matrix inversion. *J. Assoc. Comput. Mach.*, 8:281–330, 1961.

[155] J.H. Wilkinson. *Rounding Errors in Algebraic Processes*. Prentice Hall, 1963.

[156] J.H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, 1965.

[157] J.H. Wilkinson. Modern error analyis. *SIAM Review*, 13:548–568, 1971.

[158] J.H. Wilkinson. Note on matrices with a very ill-conditioned eigenproblem. *Numer. Math.*, 19:176–178, 1972.

[159] J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(272):32–43, 1928.

[160] R. Wongkew. Volumes of tubular neighbourhoods of real algebraic varieties. *Pacific J. Math.*, 159(1):177–184, 1993.

[161] H. Woźniakowski. Numerical stability for solving nonlinear equations. *Numer. Math.*, 27(4):373–390, 1976/77.

[162] M.H. Wright. The interior-point revolution in optimization: history, recent developments, and lasting consequences. *Bull. Amer. Math. Soc. (N.S.)*, 42(1):39–56, 2005.

[163] S. Wright. *Primal-Dual Interior-Point Methods*. SIAM, 1997.

[164] M. Wschebor. Smoothed analysis of $\kappa(A)$. *Journal of Complexity*, 20(1):97–107, 2004.

[165] Y. Ye. Toward probabilistic analysis of interior-point algorithms for linear programming. *Math. of Oper. Res.*, 19:38–52, 1994.

[166] T.J. Ypma. Historical development of the Newton-Raphson method. *SIAM Rev.*, 37(4):531–551, 1995.

Institute of Mathematics, Technische Universität Berlin
E-mail address: pbuerg@math.tu-berlin.de