Proceedings of the International Congress of Mathematicians
Hyderabad, India, 2010

Smoothed Analysis of Condition Numbers

Peter Bürgisser *

Abstract. We present some recent results on the probabilistic behaviour of interior point methods for the convex conic feasibility problem and for homotopy methods solving complex polynomial equations. As suggested by Spielman and Teng, the goal is to prove that for all inputs (even ill-posed ones), and all slight random perturbations of that input, it is unlikely that the running time will be large. These results are obtained through a probabilistic analysis of the condition of the corresponding computational problems.

Mathematics Subject Classification (2000). 65H20, 65Y20, 68Q25, 90C31

Keywords. condition number, distance to ill-posedness, analysis of algorithms, smoothed analysis, volume of tubes, convex conic feasibility problem, Renegar's condition number, interior point methods, polynomial equation solving, homotopy methods, polynomial time, Smale's 17th problem

1. Introduction

In computer science, the most common theoretical approach to understanding the behaviour of algorithms is *worst-case analysis*. This means proving a bound on the worst possible performance an algorithm can have. In many situations this gives satisfactory answers. However, 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. In an attempt to rectify this discrepancy, researchers have introduced the concept of *average-case analysis*, which means bounding the expected performance of an algorithm on random inputs. For the simplex algorithm, average-case analyses have been first given by Borgwardt [13] and Smale [63]. However, while a proof of good average performance yields an indication of a good performance in practice, it can rarely explain it convincingly. The problem is that the results of an average-case analysis strongly depend on the distribution of the inputs, which is unknown, and usually assumed to be Gaussian for rendering the mathematical analysis feasible.

Spielman and Teng [67] suggested in 2001 the concept of *smoothed analysis* as a new form of analysis of algorithms that arguably blends the best of both worst-case and average-case. They used this new framework to give a more compelling explanation of the simplex method (for the shadow vertex pivot rule), see [69].

*Supported by DFG grants BU 1371/2-1, BU 1371/3-1, and Paderborn Institute for Scientific Computation (PaSCo).

The general idea of smoothed analysis is easy to explain. Let $T\colon \mathbb{R}^p \supseteq \mathcal{D} \to \mathbb{R}_+ \cup \{\infty\}$ be any function (measuring running time etc). Instead of showing "it is unlikely that $T(a)$ will be large," one shows that "for all $\overline{a}$ and all slight random perturbations a of $\overline{a}$, it is unlikely that $T(a)$ will be large." We model the perturbation a by a normal distribution $N(\overline{a}, \sigma^2\mathrm{I})$ with center $\overline{a}$ and covariance matrix $\sigma^2\mathrm{I}$, given by the density

$$\rho(a) = \Big(\frac{1}{\sigma\sqrt{2\pi}}\Big)^p \cdot \exp\Big(-\frac{\|a-\overline{a}\|^2}{2\sigma^2}\Big).$$

The goal of a smoothed analysis of T is to give good estimates of

$$\sup_{\overline{a}\in\mathcal{D}} \operatorname*{Prob}_{a\sim N(\overline{a},\sigma^2\mathrm{I})}\{T(a) \geq \varepsilon^{-1}\}.$$

In a first approach, one may focus on expectations, that is, on bounding

$$\sup_{\overline{a}\in\mathcal{D}} \operatorname*{\mathbb{E}}_{a\sim\in N(\overline{a},\sigma^2\mathrm{I})} T(a).$$

Figure 1 succinctly summarizes the three types of analysis of algorithms.

Worst-case analysis	Average-case analysis	Smoothed analysis
$\sup_{a\in\mathcal{D}} T(a)$	$\operatorname*{\mathbb{E}}_{a\sim N(0,\mathrm{I})} T(a)$	$\sup_{\overline{a}\in\mathcal{D}} \operatorname*{\mathbb{E}}_{a\sim N(\overline{a},\sigma^2)} T(a)$

Figure 1. Three types of analysis of algorithms.

Smoothed analysis is not only useful for analyzing the simplex algorithm, but can be applied to a wide variety of numerical algorithms. For doing so, understanding the concept of condition numbers is an important intermediate step.

A distinctive feature of the computations considered in numerical analysis is that they are affected by errors. A main character in the understanding of the effects of these errors is the *condition number* of the input. This is a positive number which, roughly speaking, quantifies the errors when computations are performed with infinite precision but the input has been modified by a small perturbation. The condition number depends only on the data and the problem at hand (but not on the algorithm). The best known condition number is that for matrix inversion and linear equation solving. For a square matrix A it takes the form $\kappa(A) = \|A\|\cdot\|A^{-1}\|$ and was independently introduced by von Neumann and Goldstine [46] and Turing [71].

Condition numbers are omnipresent in round-off analysis. They also appear as a parameter in complexity bounds for a variety of efficient iterative algorithms in linear algebra, linear and convex optimization, as well as homotopy methods for solving systems of polynomial equations. The running time $T(a,\varepsilon)$ of these algorithms, measured as the number of arithmetic operations, can often be bounded in the form

$$T(a,\varepsilon) \;\leq\; \big(\mathrm{size}(a) + \mu(a) + \log\varepsilon^{-1}\big)^c, \tag{1}$$

with some universal constant $c > 0$. Here the input is a vector $a \in \mathbb{R}^n$ of real numbers, $\mathrm{size}(a) = n$ is the dimension of a, the positive parameter ε measures the required accuracy, and $\mu(a)$ is some measure of conditioning of a. (Depending on the situation, $\mu(a)$ may be either a condition number or its logarithm. Moreover, $\log \varepsilon^{-1}$ might be replaced by $\log \log \varepsilon^{-1}$.)

Smale [65] proposed a two-part scheme for dealing with *complexity upper bounds* in numerical analysis. The first part consists of establishing bounds of the form (1). The second part of the scheme is to analyze the distribution of $\mu(a)$ under the assumption that the inputs a are random with respect to some probability distribution. More specifically, we aim at tail estimates of the form

$$\mathrm{Prob}\left\{\mu(a) \geq \varepsilon^{-1}\right\} \;\leq\; \mathrm{size}(a)^c\, \varepsilon^{\alpha} \quad (\varepsilon > 0)$$

with universal constants $c, \alpha > 0$. In a first attempt, one may try to show upper bounds on the expectation of $\mu(a)$ (or $\log \mu(a)$, depending on the situation). Combining the two parts of the scheme, we arrive at upper bounds for the average running time of our specific numerical algorithms considered. So if we content ourselves with statements about the probabilistic average-case, we can eliminate the dependence on $\mu(a)$ in (1). This approach was elaborated upon for average-case complexity by Blum and Shub [11], Renegar [47], Demmel [29], Kostlan [40], Edelman [33, 34], Shub [54], Shub and Smale [59, 60], Cheung and Cucker [23], Cucker and Wschebor [26], Cheung et al. [24], Beltrán and Pardo [6], Bürgisser et al. [20], and others.

Spielman and Teng in their ICM 2002 paper [68] proposed to refine part two of Smale's scheme by performing a smoothed analysis of the condition number $\mu(a)$ involved for obtaining more meaningful probabilistic upper complexity bounds. The implementation of this idea has been a success story. The goal of this survey is to present some of the recent results in this direction. Beside the original papers the interested reader may also consult the survey [14] and the forthcoming book [17].

Acknowledgments. I thank Dennis Amelunxen, Felipe Cucker, and Javier Peña for constructive comments on the manuscript.

2. Conic Condition Numbers

Often, a probabilistic analysis of condition numbers can be done in a systematic way by geometric tools. Let us explain this approach for Turing's condition number $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ of a matrix $A \in \mathbb{R}^{n\times n}$. This quantity measures the sensitivity or errors for the tasks of inverting A or of solving the linear system $Ax = b$. We interpret $\Sigma = \{B \in \mathbb{R}^{n\times n} \mid \det B = 0\}$ as the "set of ill-posed inputs" for these tasks. It is mathematically convenient to measure distances between matrices with the Euclidean or Frobenius norm $\|A\|_F := (\mathrm{trace}(AA^T))^{1/2}$. We replace the spectral norm $\|A\|$ by the larger Frobenius norm $\|A\|_F$ and, instead of $\kappa(A)$, study the larger quantity $\kappa_F(A) := \|A\|_F \cdot \|A^{-1}\|$. The Eckart-Young Theorem [32] states that $\kappa_F(A)$ is inversely proportional to the distance of A to Σ.

More specifically, we have

$$\kappa_F(A) = \frac{\|A\|_F}{\operatorname{dist}(A,\Sigma)}. \tag{2}$$

If the entries of A are independent standard normal distributed, then $A/\|A\|_F$ is uniformly distributed on the unit sphere S^{n^2-1}. Since κ_F is scale-invariant, we may assume that the inputs A are chosen uniformly at random in S^{n^2-1}. We also write $\Sigma_S := \Sigma \cap S^{n^2-1}$. The *$\varepsilon$-neighborhood* of Σ_S, for $0 < \varepsilon \leq 1$, is defined as

$$T(\Sigma_S, \varepsilon) := \{A \in S^{n^2-1} \mid d_S(A, \Sigma_S) < \arcsin \varepsilon\}, \tag{3}$$

where $d_S(A, \Sigma_S) := \inf\{d_S(A, B) \mid B \in \Sigma_S\}$ and $d_S(A, B)$ denotes the angular distance of A and B in S^{n^2-1}. Using $\operatorname{dist}(A, \Sigma) = \sin d_S(A, \Sigma_S)$ we obtain from (2) for $0 < \varepsilon \leq 1$

$$\operatorname{Prob}\{\kappa_F(A) \geq \varepsilon^{-1}\} = \frac{\mathsf{vol}\, T(\Sigma_S, \varepsilon)}{\mathsf{vol}\, S^{n^2-1}}.$$

The task is therefore to compute or to estimate the volume of neighborhoods of Σ_S.

This approach applies to a much more general context than just the matrix condition number. Assume that $\mathbb{R}^{p+1}$ is the data space of a computational problem under consideration and the set of "ill-posed inputs" $\Sigma \subseteq \mathbb{R}^{p+1}$ is an algebraic cone, i.e., a real algebraic set that is closed by multiplications with scalars. We associate with Σ the *conic condition number* function defined as

$$\mathscr{C} \colon \mathbb{R}^{p+1} \setminus \{0\} \to \mathbb{R},\ \mathscr{C}(a) := \frac{\|a\|}{\operatorname{dist}(a, \Sigma)},$$

where $\|\ \|$ and dist refer to the Euclidean norm. For instance the matrix condition number κ_F is conic due to the Eckart-Young Theorem (2). The homogeneity of $\mathscr{C}$ allows us to restrict to inputs a lying in the unit sphere S^p, so that the conic condition number $\mathscr{C}(a)$ takes the form

$$\mathscr{C}(a) = \frac{1}{\operatorname{dist}(a, \Sigma)} = \frac{1}{\sin d_S(a, \Sigma_S)},$$

where $\Sigma_S := \Sigma \cap S^p$ and d_S refers to the angular distance on S^p.

Demmel [29] derived a general result giving an average-case analysis for conic condition numbers in terms of geometric invariants of the corresponding set of ill-posed inputs Σ. This is based on general estimates on the volume of neighborhoods of Σ_S obtained with integral-geometric tools. The core of these ideas, in the context of one variable polynomial equation solving, can already be found in Smale's early AMS bulletin article [62] dating from 1981.

Bürgisser et al. [18, 19] recently extended Demmel's result from average-case analysis to a natural geometric framework of smoothed analysis of conic condition numbers, called *uniform smoothed analysis*. Suppose that $\mathscr{C}$ is a conic condition number as above associated with the set Σ of ill-posed inputs. For $0 \leq \sigma \leq 1$ let $B(\overline{a}, \sigma)$ denote the spherical cap in the sphere S^p centered at $\overline{a} \in S^p$ and having angular radius $\arcsin \sigma$. Moreover, we define for $0 < \varepsilon \leq 1$ the *ε-neighborhood* of Σ_S

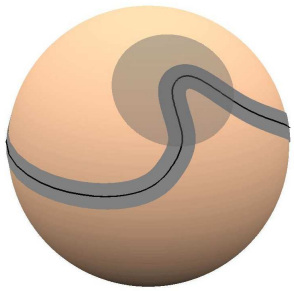

Figure 2. Neighborhood of the curve Σ_S intersected with a spherical disk.

as in (3). The task of a uniform smoothed analysis of $\mathscr{C}$ consists of providing good upper bounds on

$$\sup_{\overline{a}\in S^p} \operatorname*{Prob}_{a\in B(\overline{a},\sigma)} \{\mathscr{C}(a) \geq \varepsilon^{-1}\},$$

where a is assumed to be chosen uniformly at random in $B(\overline{a},\sigma)$. The probability occurring here has an immediate geometric meaning:

$$\operatorname*{Prob}_{a\in B(\overline{a},\sigma)} \{\mathscr{C}(a) \geq \varepsilon^{-1}\} = \frac{\mathsf{vol}\,(T(\Sigma_S,\varepsilon)\cap B(\overline{a},\sigma))}{\mathsf{vol}\,(B(\overline{a},\sigma))}. \tag{4}$$

Thus uniform smoothed analysis means to provide bounds on the relative volume of the intersection of ε-neighborhoods of Σ_S with small spherical disks, see Figure 2. We note that uniform smoothed analysis interpolates transparently between worst-case and average-case analysis. Indeed, when $\sigma = 0$ we get worst-case analysis, while for $\sigma = 1$ we obtain average-case analysis.

The following result from Bürgisser et al. [19] extends the previously mentioned result by Demmel [29] from average-case to smoothed analysis.

Theorem 2.1. *Let $\mathscr{C}$ be a conic condition number with set Σ_S of ill-posed inputs. Assume that Σ_S is contained in a real algebraic hypersurface, given as the zero set of a homogeneous polynomial of degree d. Then, for all $0 < \sigma \leq 1$ and all $0 < \varepsilon \leq \sigma/(p(2d+1))$ we have*

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

$$\sup_{\overline{a}\in S^p} \mathop{\mathbb{E}}_{a\in B(\overline{a},\sigma)} (\ln \mathscr{C}(a)) \leq 2\ln(dp) + 2\ln\frac{1}{\sigma} + 4.7.$$

The proof relies on a classical paper by Weyl [75] in which a formula for the volume of the ε-neighborhood of a submanifold M of the sphere S^p was derived. In this formula, integrals of (absolute) curvature of M enter. In [19], the integrals of absolute curvature of a smooth algebraic hypersurface M of S^p were bounded in terms of the degree d of M by means of Chern's principal kinematic formula

of integral geometry [21] and Bézout's Theorem. The smoothness assumption can then be removed by a perturbation argument.

In Cucker at al. [25] it was shown that Theorem 2.1 is quite robust with respect to the assumption on the distribution modeling the perturbations. The bound on the expectation extends (in order of magnitude) to any radially symmetric probability distributions supported on a spherical disk of radius σ whose density may even have a mild singularity at the center of the perturbation.

The setting of conic condition numbers has a natural counterpart over the complex numbers. In this setting, a result similar to Theorem 2.1 was obtained in Bürgisser et al. [18]. The critical parameter entering the estimates is again the degree but now algebraic varieties of higher codimension are taken into account as well.

Demmel's paper [29] also dealt with both complex and real problems. For complex problems he provided complete proofs. For real problems, Demmel's bounds rely on an unpublished (and apparently unavailable) result by Ocneanu on the volumes of tubes around real algebraic varieties. A second goal of Bürgisser et al. [18] was to prove a result akin to Ocneanu's.

Theorem 2.1 has a wide range of applications to linear equation solving, eigenvalue computation, and polynomial equation solving. It easily gives the following uniform smoothed analysis of the condition number of a matrix $A \in \mathbb{R}^{n\times n}$:

$$\sup_{\|\overline{A}\|_F=1} \mathop{\mathbb{E}}_{A\in B(\overline{A},\sigma)} (\ln \kappa(A)) = \mathcal{O}\big(\frac{n}{\sigma}\big).$$

Sharper results (for Gaussian perturbations) were obtained by Wschebor [79] and Sankar et al. [53]. A paper by Tao and Vu [70] deals with the condition number of integer matrices under random discrete perturbations.

3. Convex Conic Feasibility Problem

For simplicity we focus on the complexity of feasibility problems and leave out the discussion of the related convex optimization problems.

3.1. Renegar's condition number. Let X and Y be real finite-dimensional vector spaces endowed with norms. Further, let $K \subseteq X$ be a closed convex cone that is assumed to be regular, that is $K \cap (-K) = \{0\}$ and K has nonempty interior. We denote by $L(X,Y)$ the space of linear maps from X to Y endowed with the operator norm. Given $A \in L(X,Y)$, consider the feasibility problem in primal form of deciding

$$\exists x \in X \setminus \{0\} \quad Ax = 0,\ x \in K. \tag{5}$$

Two special cases of this general framework should be kept in mind. For $K = \mathbb{R}^n_+$, the nonnegative orthant in $X = \mathbb{R}^n$, one obtains the homogeneous *linear programming feasibility problem*. The feasibility version of homogeneous *semidefinite*

programming corresponds to the cone $K = \mathcal{S}^n_+$ consisting of the positive semidefinite matrices in $X = \{x \in \mathbb{R}^{n\times n} \mid x = x^T\}$.

The feasibility problem dual to (5) is

$$\exists y^* \in Y^* \setminus \{0\} \quad A^* y^* \in K^*. \tag{6}$$

Here X^*, Y^* are the dual spaces of X, Y, respectively, $A^* \in L(Y^*, X^*)$ denotes the map adjoint to A, and $K^* := \{x^* \in X^* \mid \forall x \in K \ \langle x^*, x\rangle \geq 0\}$ denotes the cone dual to K.

We denote by $\mathcal{P}$ the set of instances $A \in L(X, Y)$ for which the primal problem (5) is feasible. Likewise, we denote by $\mathcal{D}$ the set of $A \in L(X, Y)$ for which the dual problem (6) is feasible.

$\mathcal{P}$ and $\mathcal{D}$ are closed subsets of $L(X, Y)$ and the separation theorem implies that $L(X, Y) = \mathcal{P} \cup \mathcal{D}$, cf. Rockafellar [52]. One can show that $\Sigma := \mathcal{P} \cap \mathcal{D}$ is the common boundary of both of the sets $\mathcal{P}$ and $\mathcal{D}$. The *conic feasibility problem for* K is to decide for given $A \in L(X, Y)$ whether $A \in \mathcal{P}$ or $A \in \mathcal{D}$. The set Σ can be considered as the *set of ill-posed instances* for the conic feasibility problem. Indeed, for given $A \in \Sigma$, arbitrarily small perturbations of A may yield instances in both $\mathcal{P}$ and $\mathcal{D}$. We note that Σ is a cone that is neither convex nor an algebraic set.

Jim Renegar [49, 50, 51] defined the *condition number* $\mathscr{C}_R(A)$ of an instance A of the conic feasibility problem for K by

$$\mathscr{C}_R(A) := \frac{\|A\|}{\operatorname{dist}(A, \Sigma)}, \tag{7}$$

where $\|\ \|$ denotes the spectral norm and dist refers to the corresponding metric. This definition can be rephrased as follows. Suppose $A \in \mathcal{P}$. Then $1/\mathscr{C}_R(A)$ is the supremum over all $\delta > 0$ such that

$$\forall A' \in \mathbb{R}^{m\times n} : \quad \frac{\|A' - A\|}{\|A\|} < \delta \implies A' \in \mathcal{P}. \tag{8}$$

Roughly, $1/\mathscr{C}_R(A)$ is the largest normwise relative error of A that makes A stay in $\mathcal{P}$. An analogous characterization applies for $A \in \mathcal{D}$.

The most efficient known algorithms for solving convex optimization problems in theory and practice are interior-point methods, cf. Nesterov and Nemirovskii [44]. Renegar [50, 51] was the first to realize that the number of steps of interior-point algorithms solving the conic feasibility problem can be effectively bounded in terms of $\mathscr{C}_R(A)$. Early work related to this is Vavasis and Ye [72] and Nesterov et al. [45]. Condition-based analyses also exist for other algorithms in convex optimization. For Khachiyan's ellipsoid method [41] such analysis was performed by Freund and Vera [35]. For the perceptron algorithm, condition-based analyses were given for linear programming by Dunagan and Vempala [31] and for general convex conic systems by Belloni, Freund and Vempala [5].

Vera at al. [73] recently showed the following general result. One can relax the above pair of conic feasibility problems (5), (6), to a primal-dual pair of conic

optimization problems. When K is a self-scaled cone with a known self-scaled barrier function, the conic programming relaxation can be solved via a primal-dual interior-point algorithm. Moreover, for a well-posed instance A, a strict solution to one of the two original conic systems can be obtained in $\mathcal{O}\big(\sqrt{\nu}\log(\nu\mathscr{C}_R(A))\big)$ interior-point iterations. Here ν is a complexity parameter of the self-scaled barrier function of K that equals n in the interesting cases $K=\mathbb{R}^n_+$ and $K=\mathcal{S}^n_+$. An important feature of this algorithm is that the condition of the systems of equations that arise at each interior-point iteration grows in a controlled manner and remains bounded by a constant factor times $\mathscr{C}_R(A)^2$ throughout the entire algorithm.

We specialize now the discussion to the case of linear programming. That is, we consider the cone $K=\mathbb{R}^n_+$ in $X=\mathbb{R}^n$. Note that K is self-dual, i.e., $K^*=K$ when identifying X with its dual space. We set $Y=\mathbb{R}^m$ with $n\geq m$ and view $A\in L(X,Y)$ as an $m\times n$-matrix with the columns $a_1,\ldots,a_n\in\mathbb{R}^m$.

The primal feasibility problem (5) now reads as

$$\exists x\in\mathbb{R}^n\setminus\{0\}\ Ax=0, x\geq 0.$$

Geometrically, this means that 0 lies in the interior of the convex hull Δ of $a_1,\ldots,a_n$. The dual feasibility problem (6) translates to

$$\exists y\in\mathbb{R}^m\setminus\{0\}\ A^Ty\geq 0$$

meaning that Δ lies in some closed halfspace H (with $0\in\partial H$). Since $\Sigma=\mathcal{P}\cap\mathcal{D}$, an instance A is ill-posed iff 0 lies in the convex hull Δ of $a_1,\ldots,a_n$ and Δ is contained in some closed halfspace.

We note that individual scaling of the columns a_i does not change membership of A in $\mathcal{P}$ or $\mathcal{D}$, respectively. It therefore makes sense to measure relative errors of A componentwise. The resulting *GCC-condition number* $\mathscr{C}(A)$ has been introduced and investigated by Goffin [38] and Cheung and Cucker [22]. Formally, $1/\mathscr{C}(A)$ is defined as the supremum over all $\delta>0$ such that (8) holds with $\|A'-A\|/\|A\|$ replaced by $\max_i\|a_i'-a_i\|/\|a_i\|$. We remark that $\mathscr{C}(A)$ differs from $\mathscr{C}_R(A)$ by at most a factor of $\sqrt{n}$ if the a_i have equal norms.

In the following we will assume the normalization $\|a_i\|=1$. Hence we can interpret the matrix A with columns $a_1,\ldots,a_n$ as an element in the product $\mathbb{S}:=S^{m-1}\times\cdots\times S^{m-1}$ of the spheres S^{m-1}. An advantage of the GCC-condition number is that it has various nice geometric characterizations that greatly facilitate its probabilistic analysis. When introducing the metric $d_{\mathbb{S}}$ on $\mathbb{S}$ by $d(A,B):=\max_i d_S(a_i,b_i)$ with d_S denoting angular distance on S^{m-1}, and writing $\Sigma_{\mathbb{S}}:=\Sigma\cap\mathbb{S}$, the definition of $\mathscr{C}(A)$ can be rephrased as

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

This characterization can be turned into a more specific form. Let $\rho(A)$ be the angular radius of a *spherical cap of minimal radius* containing $a_1,\ldots,a_n\in S^{m-1}$. It is easy to see that $\rho(A)\leq\frac{\pi}{2}$ iff $A\in\mathcal{D}$. Hence, $\rho(A)=\frac{\pi}{2}$ iff $A\in\Sigma$. The

following characterization is due to Cheung and Cucker [22]

$$d_{\mathbb{S}}(A, \Sigma_{\mathbb{S}}) \;=\; \begin{cases} \frac{\pi}{2} - \rho(A) & \text{if } A \in \mathcal{D} \\ \rho(A) - \frac{\pi}{2} & \text{if } A \in \mathcal{P}. \end{cases}$$

It follows that $\mathscr{C}(A)^{-1} = \sin d_{\mathbb{S}}(A, \Sigma_{\mathbb{S}}) = |\cos \rho(A)|$.

3.2. Average and smoothed analysis. The average-case analysis of the GCC condition number is intimately related to a classical question on covering a sphere by random spherical caps.

Suppose that the entries of the matrix $A \in \mathbb{R}^{m \times n}$ are independent standard Gaussian random variables. After normalization this means that each column a_i is independently chosen from the uniform distribution on the sphere S^{m-1}. 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 S^{m-1}. We claim that

$$p(n, m, \alpha) = \text{Prob}\big\{\rho(A) \leq \pi - \alpha\big\}.$$

Indeed, the caps of radius α with center $a_1, \ldots, a_n$ do not cover S^{m-1} iff there exists $y \in S^{m-1}$ having distance greater than α from all a_i. The latter means that the cap of radius $\pi - \alpha$ centered at $-y$ contains all the a_i, which implies $\rho(A) \leq \pi - \alpha$ and vice versa.

The problem of determining the coverage probabilities $p(n, m, \alpha)$ is classical and completely solved only for $m \leq 2$ (Gilbert [37], Miles [43]). For $m > 2$ little was known except

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

due to Wendel [74] and asymptotic formulas for $p(n, m, \alpha)$ for $\alpha \to 0$ due to Janson [39]. Bürgisser et al. [20] 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 $\alpha \leq \pi/2$. In particular, this implies

$$\mathbb{E}(\ln \mathscr{C}(A)) \leq 2 \ln m + 3.31. \tag{9}$$

A smoothed analysis of a condition number of linear programming was first obtained by Dunagan et al. [30]. They obtained the following excellent result for Renegar's condition number $\mathscr{C}_R(A)$ of a matrix $A \in \mathbb{R}^{m \times n}$ with $n \geq m$:

$$\sup_{\|\overline{A}\|=1} \mathop{\mathbb{E}}_{A \sim N(\overline{A}, \sigma^2 I)} \big(\ln \mathscr{C}_R(A)\big) = \mathcal{O}\big(\frac{n}{\sigma}\big). \tag{10}$$

This implies the bound $\mathcal{O}(\sqrt{n} \ln \frac{n}{\sigma})$ on the smoothed expected number of iterations of the above mentioned interior-point algorithms for the conic feasibility problem in the LP-case $K = \mathbb{R}^n_+$.

For the GCC condition number a similar result can be obtained in the model of uniform smoothed analysis by different methods. More specifically, fix $\overline{a}_i \in S^{m-1}$

for $i = 1, \ldots, n$ and, independently for each i, choose a_i uniformly at random in the spherical cap $B(\overline{a}_i, \sigma)$ of S^{m-1} centered at $\overline{a}_i$ with angular radius $\arcsin \sigma$. That is, we choose $A \in B(\overline{A}, \sigma) := \prod_i B(\overline{a}_i, \sigma)$ uniformly at random. Amelunxen and Bürgisser [2] showed the following uniform tail bound: for $0 < \varepsilon \leq \sigma/(2m(m+1))$ we have

$$\sup_{\overline{A} \in \mathbb{S}} \operatorname*{Prob}_{A \in B(\overline{A}, \sigma)} \{A \in \mathcal{D},\ \mathscr{C}(A) \geq \varepsilon^{-1}\} \ \leq\ 6.5\, nm^2\, \frac{\varepsilon}{\sigma}.$$

For the primal feasible case ($A \in \mathcal{P}$) a slightly worse tail estimate was obtained. This implies a bound on the expectation similar to (10)

$$\sup_{\overline{A} \in \mathbb{S}} \mathop{\mathbb{E}}_{A \in B(\overline{A}, \sigma)} \big(\ln \mathscr{C}(A)\big) = \mathcal{O}\big(\ln \frac{n}{\sigma}\big). \tag{11}$$

The proof of this result is based on similar ideas as for Theorem 2.1. One of the points of [2] was to show that the bound (11) is *robust* in the sense of [25]: it extends to radially symmetric probability distributions supported on $B(\overline{a}_i, \sigma)$ whose density may even have a mild singularity at the center of the perturbation.

3.3. Grassmann condition number. In view of the great relevance of semidefinite programming [77] it would be desirable to have a smoothed analysis of Renegar's condition number for the cone of semidefinite matrices. However, the proofs of (10) as well as of (11) crucially rely on the product structure of the cone $\mathbb{R}_+ \times \cdots \times \mathbb{R}_+$. We therefore try to address the problem for a general regular closed convex cone $K \subseteq \mathbb{R}^n$ in a different, coordinate-free way, following Amelunxen's PhD thesis [1].

We assign to an instance $A \in \mathbb{R}^{m \times n}$ of full rank $m < n$ its kernel $W := \ker A$. This is an element of the Grassmann manifold $\mathbb{G} := \mathrm{Gr}(n-m, n)$, which is defined as the set of $(n-m)$-dimensional linear subspaces of $\mathbb{R}^n$. We note that $\mathrm{im} A^T$ equals the orthogonal complement $W^\perp$ of W. The conic feasibility problem for K on instance A can thus be rephrased as deciding the alternative

$$(P) \quad W \cap K \neq 0 \quad \text{or} \quad (D) \quad W^\perp \cap K^* \neq 0,$$

for given W, compare (5) and (6). Since A enters this decision problem only through W, we view A as a particular way of representing the object W in the Grassmann manifold of inputs. In this setting we define the set $P_\mathbb{G}$ of primal feasible instances and the set $D_\mathbb{G}$ of dual feasible instances by

$$P_\mathbb{G} := \big\{W \in \mathbb{G} \mid W \cap K \neq 0\big\}, \quad D_\mathbb{G} := \big\{W \in \mathbb{G} \mid W^\perp \cap K^* \neq 0\big\}.$$

Let us point out that, unlike in the conic feasibility problem, we have here perfect symmetry with regard to switching from the primal to the dual given by the isometry $\mathrm{Gr}(n-m, n) \to \mathrm{Gr}(m, n), W \mapsto W^\perp$. The set of ill-posed instances, defined as $\Sigma_\mathbb{G} := P_\mathbb{G} \cap D_\mathbb{G}$, can be shown to be a hypersurface in $\mathbb{G}$. It is easily seen that W is ill-posed iff W touches the cone K.

The Grassmann manifold $\mathbb{G}$ is a compact manifold with a well-defined Riemannian metric that is orthogonally invariant. Therefore the (geodesic) distance

between two elements of $\mathbb{G}$ is well-defined. In analogy with the previous developments, Amelunxen defined the *Grassmann condition number* of $W \in \mathbb{G}$ as

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

where d denotes the geodesic distance in $\mathbb{G}$. In the case $W \cap K = 0$ the distance $d(W, \Sigma_{\mathbb{G}})$ has a more intuitive interpretation: it equals the angular distance between the subsets $W \cap S^{n-1}$ and $K \cap S^{n-1}$ of the sphere S^{n-1}.

The following result cleanly separates Renegar's condition number into the intrinsic Grassmann condition $\mathscr{C}_{\mathbb{G}}(W)$ and the representation-dependent matrix condition number $\kappa(A) = \|A\| \cdot \|A^{\dagger}\|$ (where $A^{\dagger}$ denotes the Moore-Penrose inverse of A). For $A \in \mathbb{R}^{m \times n}$ of rank m and $W = \ker A$ we have

$$\mathscr{C}_{\mathbb{G}}(W) \;\leq\; \mathscr{C}_R(A) \;\leq\; \kappa(A) \cdot \mathscr{C}_{\mathbb{G}}(W). \tag{12}$$

This was shown by Belloni and Freund [4] for the dual feasible case and extended to the primal feasible case by Amelunxen [1].

The Grassmann manifold $\mathbb{G}$ has an orthogonally invariant volume form that defines a probability measure on $\mathbb{G}$. In particular it makes sense to talk about the uniform distribution on $\mathbb{G}$. This distribution arises naturally for $W = \ker A$ when we assume that the entries of $A \in \mathbb{R}^{m \times n}$ are independent standard Gaussian.

Amelunxen and Bürgisser [1, 3] obtained the following average-case analysis of the Grassmann condition number, which holds for *any* regular closed convex cone $K \subseteq \mathbb{R}^n$:

$$\begin{aligned} \operatorname*{Prob}_{W \in \mathbb{G}} \left\{ \mathscr{C}_{\mathbb{G}}(W) \geq \varepsilon^{-1} \right\} &\leq 6\, n\varepsilon \quad \text{if } \varepsilon < n^{-\frac{3}{2}}, \\ \operatorname*{\mathbb{E}}_{W \in \mathbb{G}} \left(\ln \mathscr{C}_{\mathbb{G}}(W) \right) &\leq 2.5\, \ln n + 2.8. \end{aligned}$$

Here is a very brief indication of the ideas of proof. Showing the first statement means bounding the volume of the ε-neighborhood of $\Sigma_{\mathbb{G}}$ in $\mathbb{G}$. By a perturbation argument, it suffices to consider cones with smooth boundary ∂K of positive Gaussian curvature so that K is strictly convex. Then $M := \partial K \cap S^{n-1}$ is a smooth hypersurface in the sphere S^{n-1}. By assumption, each $W \in \Sigma_{\mathbb{G}}$ touches the cone K along a unique ray $\mathbb{R}_+ p_W$ determined by a point $p_W \in M$. The fiber over $p \in M$ of the map $\Sigma_{\mathbb{G}} \to M, W \mapsto p_W$ consists of the $(n-m)$-dimensional subspaces W of the tangent space of $T_p \partial K$ containing the line $\mathbb{R}p$. The set of these W can be identified with the Grassmann manifold of $(n-m-1)$-dimensional subspaces of $T_p M$. This way, one sees that $\Sigma_{\mathbb{G}} \to M$ has the structure of a Grassmann bundle over M. With some work it is possible to extend Weyl's formula [75] for the volume of ε-neighorhoods of M in S^{n-1} to obtain a formula for the ε-neighborhood of $\Sigma_{\mathbb{G}}$ in $\mathbb{G}$.

This approach should also yield a uniform smoothed analysis of the Grassmann condition number and we are currently elaborating the details.

We close this section with a few further remarks. In the case $K = \mathbb{R}^n_+$ we get the better bounds $\mathbb{E}\left(\ln \mathscr{C}_{\mathbb{G}}(A)\right) \leq 1.5 \ln m + 6$ only depending on m as in (9).

In view of the inequality $\ln \mathscr{C}_R(A) \leq \ln \kappa(A) + \ln \mathscr{C}_{\mathbb{G}}(W)$ resulting from (12) one may ask about the contribution of the data-dependent $\ln \kappa(A)$. Surprisingly, this contribution turns out to be bounded if m/n is bounded away from 1. It is a known fact (Geman [36], Silverstein [61]) that for standard Gaussian matrices A_n of size $m_n \times n$ with m_n/n converging to a fixed number $q \in (0,1)$, the condition number $\kappa(A_n)$ converges to $\frac{1+\sqrt{q}}{1-\sqrt{q}}$ almost surely. Recently, this average-case analysis was complemented by a smoothed analysis by Bürgisser and Cucker [16] who showed

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

for $q \in (0,1)$, $m/n \leq q$, 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.

4. Solving Complex Polynomial Equations

4.1. Smale's 17th problem. In 2000, Steve Smale published a list of mathematical problems for the 21st century [66]. The 17th problem in the list reads as follows:

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

This is the guiding problem underlying the series of papers [56, 57, 58, 60, 59] — commonly referred to as "the Bézout series"— written by Shub and Smale during the first half of the 1990s, a collection of ideas, methods, and results that pervade all the research done in Smale's 17th problem since it was proposed.

We make now precise the different notions intervening in Smale's 17th problem. Fix a degree pattern $\mathbf{d} = (d_1, \ldots, d_n)$. The input space is the vector space $\mathcal{H}_{\mathbf{d}}$ of polynomial systems $f = (f_1, \ldots, f_n)$ with $f_i = \sum_\alpha a^i_\alpha X^\alpha \in \mathbb{C}[X_0, \ldots, X_n]$ homogeneous of degree d_i. We endow $\mathcal{H}_{\mathbf{d}}$ with the *Bombieri-Weyl Hermitian inner product* that is associated with the norm

$$\|f\|^2 := \sum_{|\alpha|=d_i} |a^i_\alpha|^2 \binom{d_i}{\alpha}^{-1}.$$

The reason to do so is that this inner product is invariant under the natural action of the unitary group $U(n+1)$. The quantity $N := \dim_{\mathbb{C}} \mathcal{H}_{\mathbf{d}}$ measures the size of the input system f and we further put $D := \max_i d_i$ and let $\mathcal{D} = \prod_i d_i$ be the Bézout number.

We look for solutions ζ of the equation $f(\zeta) = 0$ in the complex projective space $\mathbb{P}^n := \mathbb{P}(\mathbb{C}^{n+1})$. The expression "on the average" in Smale's 17th problem refers to the expectation with respect to the uniform distribution on the unit sphere $S(\mathcal{H}_{\mathbf{d}})$ of $\mathcal{H}_{\mathbf{d}}$. For $f, g \in \mathcal{H}_{\mathbf{d}} \setminus \{0\}$, we denote by $d_{\mathbb{S}}(f,g)$ the angle between f and g. Similarly we define $d_{\mathbb{P}}(x,y)$ for $x, y \in \mathbb{P}^n$.

In [54], Mike Shub introduced the following *projective version of Newton's method*. Let $Df(\zeta)_{|T_\zeta}$ denote the restriction of the derivative of $f\colon \mathbb{C}^{n+1} \to \mathbb{C}^n$ at ζ to the tangent space $T_\zeta := \{v \in \mathbb{C}^{n+1} \mid \langle v, \zeta\rangle = 0\}$ of $\mathbb{P}^n$ at ζ. We associate to $f \in \mathcal{H}_{\mathbf{d}}$ a map $N_f : \mathbb{C}^{n+1} \setminus \{0\} \to \mathbb{C}^{n+1} \setminus \{0\}$ defined (almost everywhere) by

$$N_f(x) = x - Df(x)^{-1}_{|T_x} f(x).$$

Note that $N_f(x)$ is homogeneous of degree 0 in f so that N_f induces a rational map from $\mathbb{P}^n$ to $\mathbb{P}^n$.

The expression "approximate zero" in Smale's 17th problem has the following precise meaning. By an *approximate zero* of $f \in \mathcal{H}_{\mathbf{d}}$ associated with a zero $\zeta \in \mathbb{P}^n$ of f we understand a point $z \in \mathbb{P}^n$ such that the sequence of Newton iterates $z_{i+1} := N_f(z_i)$ with initial point $z_0 := z$ converges immediately quadratically to ζ, i.e., $d_{\mathbb{P}}(z_i, \zeta) \leq 2^{-(2^i-1)}\, d_{\mathbb{P}}(z_0, \zeta)$ for all $i \in \mathbb{N}$.

The *condition number of f at the zero ζ* measures how much does ζ change when we perturb f a little. More specifically, we consider the *solution variety* $V_{\mathbb{P}} := \big\{(f, \zeta) \mid f(\zeta) = 0\big\} \subseteq \mathcal{H}_{\mathbf{d}} \times \mathbb{P}^n$, which is a smooth Riemannian submanifold. By the implicit function theorem, the projection map $V_{\mathbb{P}} \to \mathcal{H}_{\mathbf{d}}, (g, x) \mapsto g$ can be locally inverted around (f, ζ) if ζ is a simple solution of f: let us denote by G its local inverse. The *condition number* $\mu(f, \zeta)$ of (f, ζ) is defined as the operator norm of the derivative of G at ζ. After some rescaling it takes the following form:

$$\mu(f, \zeta) = \|f\| \cdot \|M^\dagger\|, \tag{13}$$

where (choosing a representative of ζ with $\|\zeta\| = 1$)

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

Here $M^\dagger = M^*(MM^*)^{-1}$ denotes the Moore-Penrose inverse of M and $\|M^\dagger\|$ its spectral norm.

We remark that before Shub and Smale's work, condition numbers for finding the roots of polynomials in one variable were defined and studied by Wilkinson [76], Woźniakowski [78], and Demmel [28].

Smale's α-theory [64] shows that the size of the basin of attraction of a simple zero ζ for Newton's operator N_f is controlled by $\mu(f, \zeta)$. More specifically, for z being an approximate zero of f associated with ζ, it is sufficient to have (cf. [55, 15])

$$d_{\mathbb{P}}(z, \zeta) \leq \frac{0.3}{D^{3/2}\mu(f, \zeta)}. \tag{14}$$

Finally, the notion of "uniform polynomial time algorithm" in Smale's 17th problem refers to the so-called BSS-model [10], which is essentially a model of a random access machine operating with real numbers with infinite precision and at unit cost.

The overall idea in the Bézout series is to use a *linear homotopy*. Given a *start system* $(g, \zeta) \in V_{\mathbb{P}}$ and an input $f \in \mathcal{H}_{\mathbf{d}}$ we consider the line segment $[g, f]$ connecting g and f that consists of the systems

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

If $[g, f]$ does not meet the discriminant variety (i.e., none of the q_t has a multiple zero), then there exists a unique lifting to the solution variety $V_{\mathbb{P}}$

$$\gamma\colon [0,1] \to V, t \mapsto (q_t, \zeta_t) \tag{15}$$

such that $q_0 = g$. The root of f we are looking for is ζ_1 (note $q_1 = f$).

The idea is to follow the path γ numerically: we choose a partition $t_0 = 0, t_1, \ldots, t_k = 1$ and, writing $q_i := q_{t_i}$ and $\zeta_i := \zeta_{t_i}$, we successively compute approximations z_i of ζ_i by Newton's method starting with $z_0 := \zeta$. More specifically, we compute

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

Two questions arise: how do we choose the start system (g, ζ) and how do we find the subdivision points t_i?

The state of the art at the end of the Bézout series, i.e., in [59], showed an incomplete picture. For the choice of the subdivision, the rule consisted of taking a regular subdivision of $[g, f]$ for a given k, executing the path-following procedure, and repeating with k replaced by $2k$ if the final point could not be shown to be an approximate zero of f (a criterion for checking this follows from (14)).

As for the question of the choice of the start system (g, ζ), Shub and Smale proved in [59] that good start systems (g, ζ) existed for each degree pattern $\mathbf{d}$ (in the sense that the average number of iterations for the rule above was polynomial in the input size N), but they could not exhibit a procedure to generate one such start system. They conjectured in [59] that the system $g \in \mathcal{H}_{\mathbf{d}}$ given by $g_i = X_0^{d_i - 1} X_i$ is a good start system. While this conjecture is supported by numerical experiments, a proof remains elusive.

After the Bézout series, the next breakthrough took a decade to come. Beltrán and Pardo proposed in [7, 8, 9] that the start system (g, ζ) should be randomly chosen. We consider the following probability distribution $\rho_{\mathbf{st}}$ on $V_{\mathbb{P}}$ for the start system (g, ζ). It consists of drawing g in the sphere $S(\mathcal{H}_{\mathbf{d}}) := \{g \in \mathcal{H}_{\mathbf{d}} \mid \|g\| = 1\}$ from the uniform distribution and then choosing one of the (almost surely) $\mathcal{D}$ zeros of g from the uniform distribution on $\{1, \ldots, \mathcal{D}\}$. This procedure is clearly non-constructive, as computing a zero of a system is the problem we wanted to solve in the first place. One of the major contributions in [7] was to show that the distribution $\rho_{\mathbf{st}}$ can be efficiently sampled.

4.2. Average and smoothed analysis. Following a result in Shub [55], the following specific adaptive choice of the subdivision was proposed in [15]. We reparametrize the curve γ from (15) using a parameter $0 \le \tau \le 1$ which measures a ratio of angles. More specifically, let $\alpha = d_{\mathbb{S}}(g, f)$ and $\alpha\tau(t)$ be the angle between $g/\|g\|$ and $q_t/\|q_t\|$. As the stepsize we choose, with the parameter $\lambda = 7.53 \cdot 10^{-3}$,

$$\alpha(\tau_{i+1} - \tau_i) = \frac{\lambda}{D^{3/2}\mu(q_i, z_i)^2}$$

and call the resulting algorithm ALH (*Adaptive Linear Homotopy*). An analysis [55, 15] shows that ALH finds an approximate zero of $f = q_1$ with a number $K(f, g, \zeta)$

of steps bounded by

$$K(f,g,\zeta) \;\le\; 217\, D^{3/2}\, d_{\mathbb{S}}(f,g) \int_0^1 \mu(q_\tau,\zeta_\tau)^2\, d\tau. \tag{16}$$

Consider the *Las Vegas algorithm* LV that on input $f \in \mathcal{H}_{\mathbf{d}}$ draws the start system $(g,\zeta) \in V_{\mathbb{P}}$ at random from the distribution $\rho_{\mathbf{st}}$ and then runs ALH on input (f,g,ζ). The algorithm LV either outputs an approximate zero z of f or loops forever. We write

$$K(f) := \mathop{\mathbb{E}}_{(g,\zeta)\sim\rho_{\mathbf{st}}} (K(f,g,\zeta))$$

for the *expected number of iterations* of LV on input f. The expected running time (i.e., number of arithmetic operations) of LV is given by $K(f)$ times the cost of one iteration, the latter being dominated by that of computing one Newton iterate (which is $\mathcal{O}(N+n^3)$).

Beltrán and Pardo [9] performed an average-case analysis of LV showing that

$$E_{f\in S(\mathcal{H}_{\mathbf{d}})} K(f) = \mathcal{O}(nND^{3/2}).$$

We note that in this result, randomness enters in two ways: as a computational technique (choice of the start system) and as a way of analyzing the algorithm (average over all inputs).

Bürgisser and Cucker [15] succeeded in giving a smoothed analysis of the algorithm LV. For making such analysis possible it was essential to model random perturbations by Gaussians. For $\overline{f} \in \mathcal{H}_{\mathbf{d}}$ and $\sigma > 0$ we denote by $\rho_{\overline{f},\sigma}$ the density of the Gaussian distribution $N(\overline{f},\sigma^2\mathrm{I})$ on $\mathcal{H}_{\mathbf{d}}$ with mean $\overline{f}$ and covariance matrix $\sigma^2\mathrm{I}$. For technical simplicity, the smoothed analysis of LV assumes that the local perturbations follow a *truncated Gaussian distribution* $N_A(\overline{f},\sigma^2\mathrm{I})$ with center $\overline{f} \in \mathcal{H}_{\mathbf{d}}$ that is defined by the following density

$$\rho(f) = \begin{cases} \frac{\rho_{\overline{f},\sigma}(f)}{P_{A,\sigma}} & \text{if } \|f-\overline{f}\| \le A \\ 0 & \text{otherwise.} \end{cases}$$

Here $A := \sqrt{2N}$ and $P_{A,\sigma} := \mathrm{Prob}\{\|g\| \le A \mid g \sim N(0,\sigma^2\mathrm{I})\}$: one can show that $P_{A,\sigma} \ge \frac{1}{2}$ for all $\sigma \le 1$.

Here is the smoothed analysis result for LV from Bürgisser and Cucker [15]:

Theorem 4.1. *For any $0 < \sigma \le 1$, the algorithm* LV *satisfies*

$$\sup_{\overline{f}\in S(\mathcal{H}_{\mathbf{d}})} \mathop{\mathbb{E}}_{f\sim N_A(\overline{f},\sigma^2\mathrm{I})} K(f) = \mathcal{O}\Big(\frac{nND^{3/2}}{\sigma}\Big).$$

Average (or smoothed) complexity results do not provide information on the running time of an algorithm for the instance at hand. In [15] a condition based analysis for LV was achieved. It bounds the number of iterations on input f in terms of the *maximum condition* of f defined in [56] as

$$\mu_{\max}(f) := \max_{\zeta\,|\,f(\zeta)=0} \mu(f,\zeta).$$

Theorem 4.2. *The expected number of iterations of Algorithm* LV *with input* $f \in S(\mathcal{H}_{\mathbf{d}})$ *is bounded as*

$$K(f) = \mathcal{O}\big(D^3 n N \mu_{\max}^2(f)\big).$$

All previously known complexity bounds depended also on the condition of the intermediate systems q_t encountered along the homotopy.

The polynomials occurring in practice often have a special structure. For instance one might be interested in polynomial systems lying in a certain fixed linear subspace of $\mathcal{H}_{\mathbf{d}}$. Important examples are provided by *sparse polynomial systems*, where the set of occurring monomials is prescribed (and usually small). It is a challenging research problem to analyze the behaviour of homotopy algorithms for sparse random input systems in a meaningful way. For work in this direction we refer to Dedieu [27] and Malajovich and Rojas [42].

4.3. A near solution to Smale's 17th problem. Even though randomized algorithms are efficient in theory and reliable in practice they do not offer an answer to the question of the existence of a *deterministic* algorithm computing approximate zeros of complex polynomial systems in average polynomial time. We shall exhibit a deterministic algorithm finding an approximate zero of a given polynomial system that works in nearly-polynomial average time, more precisely in average time $N^{\mathcal{O}(\log\log N)}$.

In the case $D \leq n$ we apply algorithm ALH with the starting system $(U, \boldsymbol{z})$, where $U_i = X_i^{d_i} - X_0^{d_i}$ and $\boldsymbol{z} = (1 : 1 : \ldots : 1)$. Let $K_U(f)$ denote the number of iterations of the resulting deterministic algorithm. One can show that $\mu_{\max}(U)^2 \leq 2\,(n+1)^D$. Using this and employing the same technique as for Theorem 4.2 one can show that (cf. [15])

$$\mathop{\mathbb{E}}_{f \in S(\mathcal{H}_{\mathbf{d}})} K_{\overline{U}}(f) = \mathcal{O}(D^3 N n^{D+1}).$$

For $D > n$ we use another approach, namely, a real number algorithm designed by Renegar [48] which in this case has a performance similar to the above algorithm when $D \leq n$. Putting both pieces together Bürgisser and Cucker [15] obtained a near solution to Smale's 17th problem.

Theorem 4.3. *There is a* deterministic *real number algorithm that on input* $f \in \mathcal{H}_{\mathbf{d}}$ *computes an approximate zero of* f *in average time* $N^{\mathcal{O}(\log\log N)}$*, where* $N = \dim \mathcal{H}_{\mathbf{d}}$ *measures the size of the input* f*. Moreover, if we restrict data to polynomials satisfying*

$$D \leq n^{\frac{1}{1+\varepsilon}} \quad \text{or} \quad D \geq n^{1+\varepsilon},$$

for some fixed $\varepsilon > 0$*, then the average time of the algorithm is polynomial in the input size* N*.*

4.4. Some ideas of the proofs. It is essential that, for fixed t, $q_t = (1-t)g + tf$ follows a Gaussian law if f and g do so. Note that the variance σ_t^2 of q_t is given by $\sigma_t^2 = (1-t)^2\sigma_g^2 + t^2\sigma_f^2$, where σ_f^2 and σ_g^2 denote the variances of f

and g, respectively. By a change of parameter, the integral in (16) bounding the number of steps of ALH can be estimated as

$$d_{\mathbb{S}}(f,g)\int_0^1 \mu_2(q_\tau)^2\,d\tau \le \int_0^1 \|f\|\,\|g\|\,\frac{\mu_2^2(q_t)}{\|q_t\|^2}\,dt, \tag{17}$$

where the *mean square condition number* $\mu_2(q)$ of $q \in \mathcal{H}_{\mathbf{d}}$ is defined as

$$\mu_2(q) := \Big(\frac{1}{\mathcal{D}}\sum_{\zeta\,|\,q(\zeta)=0}\mu(q,\zeta)^2\Big)^{1/2}.$$

The factor $\|f\|\,\|g\|$ in (17) can be easily bounded and factored out the expectation. So by exchanging the expectation (over f and/or g) with the integral over t we face the problem of estimating expectations of $\mu_2^2(q_t)/\|q_t\|^2$ for different choices of the mean $\overline{q}_t$ and the variance σ_t^2. This is achieved by the following smoothed analysis of the mean square condition number, which is the technical heart of the proofs in [15].

Theorem 4.4. *Let $\overline{q} \in \mathcal{H}_{\mathbf{d}}$ and $\sigma > 0$. For $q \in \mathcal{H}_{\mathbf{d}}$ drawn from $N(\overline{q}, \sigma^2\mathrm{I})$ we have*

$$\mathop{\mathbb{E}}_{q}\Big(\frac{\mu_2^2(q)}{\|q\|^2}\Big) \;\le\; \frac{e(n+1)}{2\sigma^2}.$$

Sketch of Proof. We distinguish points $[\zeta] \in \mathbb{P}^n$ from their representatives ζ in the sphere $\mathbb{S}^n := \{\zeta \in \mathbb{C}^{n+1} \mid \|\zeta\| = 1\}$. Note that $[\zeta] \cap \mathbb{S}^n$ is a circle with radius one. We work with the "lifting"

$$V := \{(q,\zeta) \in \mathcal{H}_{\mathbf{d}} \times \mathbb{S}^n \mid q(\zeta) = 0\}$$

of the solution variety $V_{\mathbb{P}}$, which is a vector bundle over $\mathbb{S}^n$ with respect to the projection $\pi_2 \colon V \to \mathbb{S}^n, (q,\zeta) \mapsto \zeta$.

For $\zeta \in \mathbb{S}^n$ we consider the following subspace R_ζ of $\mathcal{H}_{\mathbf{d}}$ consisting of systems h that vanish at ζ of higher order in the following sense:

$$R_\zeta := \{h \in \mathcal{H}_{\mathbf{d}} \mid h(\zeta) = 0, Dh(\zeta) = 0\}.$$

We further decompose the orthogonal complement $R_\zeta^\perp$ of R_ζ in $\mathcal{H}_{\mathbf{d}}$ (defined with respect to the Bombieri-Weyl Hermitian inner product). Let L_ζ denote the subspace of $R_\zeta^\perp$ consisting of the systems vanishing at ζ and let C_ζ denote its orthogonal complement in $R_\zeta^\perp$. Then we have an orthogonal decomposition

$$\mathcal{H}_{\mathbf{d}} = C_\zeta \oplus L_\zeta \oplus R_\zeta \tag{18}$$

parameterized by $\zeta \in \mathbb{S}^n$. In fact, this can be interpreted as an orthogonal decomposition of the trivial Hermitian vector bundle $\mathcal{H}_{\mathbf{d}} \times \mathbb{S}^n \to \mathbb{S}^n$ into subbundles C, L, and R over $\mathbb{S}^n$. Moreover, the vector bundle V is the orthogonal sum of L and R: we have $V_\zeta = L_\zeta \oplus R_\zeta$ for all ζ.

Let $\mathcal{M}$ denote the space $\mathbb{C}^{n\times(n+1)}$ of matrices. In the special case, where all the degrees d_i are one, the solution manifold V specializes to the manifold

$$W := \big\{\big(M,\zeta\big) \in \mathcal{M}\times\mathbb{S}^n \mid M\zeta = 0\}$$

and π_2 specializes to the vector bundle $p_2\colon W \to \mathbb{S}^n, (M,\zeta)\mapsto\zeta$ with the fibers

$$W_\zeta := \{M\in\mathcal{M} \mid M\zeta = 0\}.$$

One can show that we have *isometrical* linear maps

$$W_\zeta \to L_\zeta,\ M=(m_{ij}) \mapsto g_{M,\zeta} := \big(\sqrt{d_i}\,\langle X,\zeta\rangle^{d_i-1}\textstyle\sum_j m_{ij}X_j\big). \tag{19}$$

In other words, the Hermitian vector bundles W and L over $\mathbb{S}^n$ are isometric.

We compose the orthogonal bundle projection $V_\zeta = L_\zeta\oplus R_\zeta \to L_\zeta$ with the bundle isometry $L_\zeta\simeq W_\zeta$ obtaining the map of vector bundles

$$\Psi\colon V\to W,\ (g_{M,\zeta}+h,\zeta)\mapsto(M,\zeta)$$

whose fibers $\Psi^{-1}(M,\zeta)$ are isometric to R_ζ. The map Ψ provides the link to the condition number: by the definition (13) we have

$$\frac{\mu(q,\zeta)}{\|q\|} \;=\; \|M^\dagger\|,\quad \text{where } (M,\zeta)=\Psi(q,\zeta). \tag{20}$$

(For showing this use $Dg_{M,\zeta}(\zeta) = \mathsf{diag}(\sqrt{d_1},\dots,\sqrt{d_n})M$.)

Let $\rho_{\mathcal{H}_\mathbf{d}}$ denote the density of the Gaussian $N(\overline{q},\sigma^2\mathrm{I})$ on $\mathcal{H}_\mathbf{d}$, where $\overline{q}\in\mathcal{H}_\mathbf{d}$ and $\sigma>0$. For fixed $\zeta\in\mathbb{S}^n$ we decompose the mean $\overline{q}$ as

$$\overline{q} = \overline{k}_\zeta + \overline{g}_\zeta + \overline{h}_\zeta \in C_\zeta\oplus L_\zeta\oplus R_\zeta$$

according to (18). If we denote by ρ_{C_ζ}, ρ_{L_ζ}, and ρ_{R_ζ} the densities of the Gaussian distributions in the spaces C_ζ, L_ζ, and R_ζ with covariance matrices $\sigma^2\mathrm{I}$ and means $\overline{k}_\zeta, \overline{M}_\zeta$, and $\overline{h}_\zeta$, respectively, then the density $\rho_{\mathcal{H}_\mathbf{d}}$ factors as

$$\rho_{\mathcal{H}_\mathbf{d}}(k+g+h) = \rho_{C_\zeta}(k)\cdot\rho_{L_\zeta}(g)\cdot\rho_{R_\zeta}(h). \tag{21}$$

The Gaussian density ρ_{L_ζ} on L_ζ induces a Gaussian density ρ_{W_ζ} on the fiber W_ζ with the covariance matrix $\sigma^2\mathrm{I}$ via the isometrical linear map (19), so that we have $\rho_{W_\zeta}(M) = \rho_{L_\zeta}(g_{M,\zeta})$.

Think now of choosing (q,ζ) at random from V by first choosing $q\in\mathcal{H}_\mathbf{d}$ from $N(\overline{q},\sigma^2\mathrm{I})$, then choosing one of its $\mathcal{D}$ zeros $[\zeta]\in\mathbb{P}^n$ at random from the uniform distribution on $\{1,\dots,\mathcal{D}\}$, and finally choosing a representative ζ in the unit circle $[\zeta]\cap\mathbb{S}^n$ uniformly at random. We denote the resulting probability density on V by ρ_V (this is a natural extension of $\rho_\mathbf{st}$). Then we have

$$\mathop{\mathbb{E}}_{\mathcal{H}_\mathbf{d}}\Big(\frac{\mu_2(q)^2}{\|q\|^2}\Big) \;=\; \mathop{\mathbb{E}}_{V}\Big(\frac{\mu(q,\zeta)^2}{\|q\|^2}\Big), \tag{22}$$

where $\mathbb{E}_{\mathcal{H}_\mathbf{d}}$ and $\mathbb{E}_V$ refer to the expectations with respect to the distribution $N(\overline{q}, \sigma^2 \mathrm{I})$ on $\mathcal{H}_\mathbf{d}$ and the probability density ρ_V on V, respectively.

To estimate the right-hand side in (22) we reduce the problem to one in the space $\mathcal{M}$ of matrices via the map Ψ. Equation (20) implies that

$$\mathop{\mathbb{E}}_{V}\Big(\frac{\mu(q,\zeta)^2}{\|q\|^2}\Big) \;=\; \mathop{\mathbb{E}}_{W}\Big(\|M^\dagger\|^2\Big), \tag{23}$$

where $\mathbb{E}_W$ denotes the expectation with respect to the *pushforward density* ρ_W of the density ρ_V via the map Ψ.

We have thus reduced our problem to a probability analysis of $\|M^\dagger\|$, the latter being a quantity closely related to the matrix condition number $\kappa(M) = \|M\| \cdot \|M^\dagger\|$. In order to proceed, we need to get some understanding of the probability density ρ_W.

The probability density ρ_W defines a pushforward density $\rho_{\mathbb{S}^n}$ on $\mathbb{S}^n$ via the projection $p_2 \colon W \to \mathbb{S}^n$, as well as *conditional probability densities* $\widetilde{\rho}_{W_\zeta}$ on the fibers W_ζ, and we have

$$\mathop{\mathbb{E}}_{W}\big(\|M^\dagger\|^2\big) \;=\; \mathop{\mathbb{E}}_{\zeta\sim\rho_{\mathbb{S}^n}}\Big(\mathop{\mathbb{E}}_{M\sim\widetilde{\rho}_{W_\zeta}}\big(\|M^\dagger\|^2\big)\Big). \tag{24}$$

(This is made formal by means of the coarea formula or Fubini's theorem for Riemannian manifolds.) For proving Theorem 4.4 it is therefore enough to show that for all $\zeta \in \mathbb{S}^n$

$$\mathop{\mathbb{E}}_{M\sim\widetilde{\rho}_{W_\zeta}}\big(\|M^\dagger\|^2\big) \;\leq\; \frac{e(n+1)}{2\sigma^2}. \tag{25}$$

The analysis of the situation reveals that the density $\widetilde{\rho}_{W_\zeta}$ is closely related to a Gaussian, namely it has the form (c_ζ denoting a normalization factor)

$$\widetilde{\rho}_{W_\zeta}(M) = c_\zeta^{-1} \cdot \det(MM^*)\, \rho_{W_\zeta}(M).$$

This finding allows one to prove tail bounds similarly as it was done in Sankar et al. [53, §3]. □

References

[1] D. Amelunxen (2010), 'Geometric analysis of the condition of the convex feasibility problem', PhD thesis, University of Paderborn, in preparation.

[2] D. Amelunxen and P. Bürgisser (2008), 'Robust Smoothed Analysis of a Condition Number for Linear Programming', accepted for *Math. Program.* Ser. A, available at arXiv:0803.0925.

[3] D. Amelunxen and P. Bürgisser, 'Probabilistic analysis of the Grassmann condition number (working title)', in preparation.

[4] A. Belloni and R. M. Freund (2009), 'A geometric analysis of Renegar's condition number, and its interplay with conic curvature', *Math. Program.* Ser. A **119**, 95–107.

[5] A. Belloni, R. M. Freund and S. Vempala (2009), 'An efficient re-scaled perceptron algorithm for conic systems', *Math. Oper. Res.* **34** (3), 621–641.

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

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

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

[9] C. Beltrán and L. M. Pardo , 'Fast linear homotopy to find approximate zeros of polynomial systems', submitted.

[10] L. Blum, F. Cucker, M. Shub and S. Smale (1998), *Complexity and Real Computation*, Springer.

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

[12] L. Blum, M. Shub and S. Smale (1989), 'On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines', *Bull. Amer. Math. Soc. (N.S.)* **21**(1), 1–46.

[13] K.-H. Borgwardt (1982), 'The average number of pivot steps required by the simplex-method is polynomial', *Z. Oper. Res. Ser.* **26**(5), 157–177.

[14] P. Bürgisser (2009), 'Smoothed Analysis of Condition Numbers', in *Foundations of Computational Mathematics, Hong Kong 2008*, edited by F. Cucker, A. Pinkus, and M. J. Todd, London Mathematical Society Lecture Note Series 363, Cambridge University Press.

[15] P. Bürgisser, and F. Cucker (2009), 'Solving Polynomial Equations in Smoothed Polynomial Time and a Near Solution to Smale's 17th Problem', accepted for 42nd annual ACM Symposium on Theory of Computing (STOC 2010). Full version available at `http://arxiv.org/abs/0909.2114`.

[16] P. Bürgisser, and F. Cucker 'Smoothed Analysis of Moore-Penrose Inversion', submitted. Available at arXiv:1002.4690.

[17] P. Bürgisser, and F. Cucker 'Condition', book in preparation.

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

[19] P. Bürgisser, F. Cucker and M. Lotz (2008), 'The probability that a slightly perturbed numerical analysis problem is difficult', *Math. Comp.* **77**, 1559–1583.

[20] P. Bürgisser, F. Cucker, and M. Lotz (2010), 'Coverage processes on spheres and condition numbers for linear programming', *Ann. Probab.* **38**(2), 570–604.

[21] S.-S. Chern (1966), 'On the kinematic formula in integral geometry', *J. Math. Mech.* **16**, 101–118.

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

[23] D. Cheung and F. Cucker (2002), 'Probabilistic analysis of condition numbers for linear programming', *J. Optim. Theory Appl.* **114**(1), 55–67.

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

[25] F. Cucker, R. Hauser, and M. Lotz (2010), 'Adversarial smoothed analysis', to appear in *J. of Complexity.*

[26] F. Cucker and M. Wschebor (2003), 'On the expected condition number of linear programming problems', *Numer. Math.* **94**(3), 419–478.

[27] J.-P. Dedieu (1999), 'Condition number analysis for sparse polynomial systems', *Foundations of computational mathematics (Rio de Janeiro, 1997)*, 75–101, Springer, Berlin, 1997.

[28] J. Demmel (1987), 'On condition numbers and the distance to the nearest ill-posed problem', *Numer. Math.* **51**, 251–289.

[29] J. Demmel (1988), 'The probability that a numerical analysis problem is difficult', *Math. Comp.* **50**, 449–480.

[30] J. Dunagan, D. A. Spielman, and S.-H. Teng (2009), 'Smoothed analysis of condition numbers and complexity implications for linear programming', *Math. Program.* Ser. A, DOI 10.1007/s10107-009-0278-5.

[31] J. Dunagan and S. Vempala (2008), 'A simple polynomial-time rescaling algorithm for solving linear programs', *Math. Program.* **114** (1, Ser. A), 101–114.

[32] C. Eckart and G. Young (1936), 'The approximation of one matrix by another of lower rank', *Psychometrika* **1**, 211–218.

[33] A. Edelman (1988), 'Eigenvalues and condition numbers of random matrices', *SIAM J. of Matrix Anal. and Applic.* **9**, 543–556.

[34] A. Edelman (1992), 'On the distribution of a scaled condition number', *Math. Comp.* **58**, 185–190.

[35] R.M. Freund and J.R. Vera (1999), 'Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm', *SIAM J. Optim.* **10**(1), 155–176.

[36] S. Geman (1980), 'A limit theorem for the norm of random matrices', *Ann. Probab.* **8**(2), 252–261.

[37] E.N. Gilbert (1965), 'The probability of covering a sphere with N circular caps', *Biometrika* **52**, 323–330.

[38] J.-L. Goffin (1980), 'The relaxation method for solving systems of linear inequalities', *Math. Oper. Res.* **5**(3), 388–414.

[39] S. Janson (1986), 'Random coverings in several dimensions', *Acta Math.* **156**, 83–118.

[40] E. Kostlan (1988), 'Complexity theory of numerical linear algebra', *J. Comput. Appl. Math.* **22**, 219–230.

[41] L.G. Khachiyan (1979), 'A polynomial time algorithm in linear programming', *Dokl. Akad. Nauk. SSSR* **244**, 1093–1096.

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

[43] R.E. Miles (1969), 'The asymptotic values of certain coverage probabilities', *Biometrika*, **56**, 661–680.

[44] Y. Nesterov and A. Nemirovskii (1994), *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[45] Y. Nesterov and M. Todd and Y. Ye (1999), 'Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems', *Math. Program.* **84** (2, Ser. A), 227–267.

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

[47] J. Renegar (1987), 'On the efficiency of Newton's method in approximating all zeros of systems of complex polynomials', *Math. of Oper. Research* **12**, 121–148.

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

[49] J. Renegar (1994), 'Some perturbation theory for linear programming', *Math. Program.* **65** (1, Ser. A), 73–91.

[50] J. Renegar (1995a), 'Incorporating condition measures into the complexity theory of linear programming', *SIAM J. Optim.* **5**(3), 506–524.

[51] J. Renegar (1995b), 'Linear programming, complexity theory and elementary functional analysis', *Math. Program.* **70**(3, Ser. A), 279–351.

[52] R.T. Rockafellar (1970), *Convex Analysis*, Princeton University Press.

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

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

[55] M. Shub (2009), 'Complexity of Bézout's theorem VI: Geodesics in the condition (number) metric', *Found. Comput. Math.* **9**(2), 171–178.

[56] M. Shub and S. Smale (1993a), 'Complexity of Bézout's theorem. I. Geometric aspects', *J. Amer. Math. Soc.* **6**(2), 459–501.

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

[58] M. Shub and S. Smale (1993), 'Complexity of Bézout's theorem III: condition number and packing', *J. Complexity* **9**, 4–14.

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

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

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

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

[63] S. Smale (1983), 'On the average number of steps of the simplex method of linear programming', *Math. Programming* **27**(3), 241–262.

[64] S. Smale (1986), 'Newton's method estimates from data at one point', in *The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985)*, pages 185–196. Springer, New York, 1986.

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

[66] S. Smale (2000), 'Mathematical problems for the next century', In *Mathematics: frontiers and perspectives*, pages 271–294, Amer. Math. Soc., Providence, RI.

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

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

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

[70] T. Tao and V. Vu (2007), 'The condition number of a randomly perturbed matrix', *Proceedings 39th annual ACM Symposium on Theory of computing*, pages 248–255.

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

[72] S. A. Vavasis and Y. Ye (1995), 'Condition numbers for polyhedra with real number data', *Oper. Res. Lett.* **17**(5), 209–214.

[73] J. C. Vera, J. C. Rivera, J. Peña, and Y. Hui (2007), 'A primal-dual symmetric relaxation for homogeneous conic systems', *J. Complexity* **23**, 245–261.

[74] J. G. Wendel (1962), 'A problem in geometric probability', *Math. Scand.* **11**, 109–111.

[75] H. Weyl (1939), 'On the Volume of Tubes', *Amer. J. Math.* **61**(2), 461–472.

[76] J. H. Wilkinson (1963), *Rounding errors in algebraic processes*, Prentice-Hall Inc., Englewood Cliffs, N.J.

[77] H. Wolkowicz, R. Saigal, and L. Vandenberghe (2000), *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, Kluwer Academic Publishers.

[78] H. Woźniakowski (1977), 'Numerical stability for solving nonlinear equations', *Numer. Math.* **27**(4), 373–390.

[79] M. Wschebor (2004), 'Smoothed analysis of $\kappa(a)$', *J. of Complexity* **20**, 97–107.

Institute of Mathematics, University of Paderborn, D-33098 Paderborn, Germany
E-mail: pbuerg@upb.de