

# 1

## Smoothed Analysis of Condition Numbers

Peter Bürgisser

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

### Abstract

The running time of many iterative numerical algorithms is dominated by the condition number of the input, a quantity measuring the sensitivity of the solution with regard to small perturbations of the input. Examples are iterative methods of linear algebra, interior-point methods of linear and convex optimization, as well as homotopy methods for solving systems of polynomial equations. Thus a probabilistic analysis of these algorithms can be reduced to the analysis of the distribution of the condition number for a random input. This approach was elaborated for average-case complexity by many researchers.

The goal of this survey is to explain how average-case analysis can be naturally refined in the sense of smoothed analysis. The latter concept, introduced by Spielman and Teng in 2001, aims at showing that for all real inputs (even ill-posed ones), and all slight random perturbations of that input, it is unlikely that the running time will be large. A recent general result of Bürgisser, Cucker and Lotz (2008) gives smoothed analysis estimates for a variety of applications. Its proof boils down to local bounds on the volume of tubes around a real algebraic hypersurface in a sphere. This is achieved by bounding the integrals of absolute curvature of smooth hypersurfaces in terms of their degree via the principal kinematic formula of integral geometry and Bézout's theorem.

### 1.1 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 (1982) and Smale (1983). 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 suggested in 2001 the concept of *smoothed analysis of algorithms*, which is 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). For this work they were recently awarded the 2008 Gödel prize. See Spielman and Teng (2004) for the full paper.

The general idea of smoothed analysis is easy to explain. Let  $T: \mathbb{R}^p \rightarrow \mathbb{R}_+ \cup \{\infty\}$  be any function (measuring running time etc). Instead of showing “it is unlikely that  $T(a)$  will be large,” one shows that “for all  $\bar{a}$  and all slight random perturbations  $\bar{a} + \delta a$ , it is unlikely that  $T(\bar{a} + \delta a)$  will be large.” If we assume that the perturbation  $\delta a$  is centered (multivariate) standard normal with variance  $\sigma^2$ , in short  $\delta a \in N(0, \sigma^2)$ , then the goal of a smoothed analysis of  $T$  is to give good estimates of

$$\sup_{\bar{a} \in \mathbb{R}^p} \text{Prob}_{\delta a \in N(0, \sigma^2)} \{T(\bar{a} + \delta a) \geq \epsilon^{-1}\}.$$

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

$$\sup_{\bar{a} \in \mathbb{R}^p} \mathbb{E}_{\delta a \in N(0, \sigma^2)} T(\bar{a} + \delta a).$$

Figure 1.1 succinctly summarizes the 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 anal-

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

Fig. 1.1. Three types of analysis of algorithms.  $\mathcal{D}$  denotes a probability distribution on  $\mathbb{R}^p$ .

ysis 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. The best known condition number is that for matrix inversion and linear equation solving. For a square matrix  $A$  it takes the form  $\kappa(A) = \|A\| \|A^{-1}\|$  and was independently introduced by Goldstine and von Neumann (1947) and Turing (1948).

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  $T(x, \epsilon)$  of these algorithms, measured as the number of arithmetic operations, can often be bounded in the form

$$T(x, \epsilon) \leq (\text{size}(x) + \mu(x) + \log \epsilon^{-1})^c, \quad (1.1)$$

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

We discuss the issue of condition-based analysis of algorithms in the Sections 1.2–1.4, by elaborating a bit on the case of convex optimization and putting special focus on generalizations of Renegar’s (1995a, 1995b) condition number for linear programming. We also discuss Shub and Smale’s (1993a) condition number for polynomial equation solving.

Let us mention that L. Blum (1990) suggested to extend the complexity theory of real computation due to Blum, Shub, Smale 1989 by

measuring the performance of algorithms in terms of the size and the condition of inputs. However, up to now, no complexity theory over the reals has been developed that incorporates the concepts of approximation and conditioning and allows to speak about lower bounds or completeness results in that context.

Smale (1997) proposed a two-part scheme for dealing with complexity *upper bounds* in numerical analysis. The first part consists of establishing bounds of the form (1.1). The second part of the scheme is to analyze the distribution of  $\mu(x)$  under the assumption that the inputs  $x$  are random with respect to some probability distribution. More specifically, we aim at tail estimates of the form

$$\text{Prob} \{ \mu(x) \geq \epsilon^{-1} \} \leq \text{size}(x)^c \epsilon^\alpha \quad (\epsilon > 0)$$

with universal constants  $c, \alpha > 0$ . In a first attempt, one may try to show upper bounds on the expectation of  $\mu(x)$  (or  $\log \mu(x)$ , 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  $\mu(x)$  in (1.1). This approach was elaborated for average-case complexity by Blum and Shub (1986), Renegar (1987), Demmel (1988), Kostlan (1988), Edelman (1988, 1992), Shub and Smale (1993b, 1994, 1996), Cheung and Cucker (2002), Cucker and Wschebor (2003), Cheung et al. (2005), Beltrán and Pardo (2007), and others. We only briefly discuss a few of these results in Section 1.5. Instead, we put emphasis on the analysis of the GCC-condition number  $\mathcal{C}(A)$  of linear programming introduced by Goffin (1980) and Cheung and Cucker (2001), see (1.11). This is a variation of the condition number introduced by Renegar (1995a, 1995b). We discuss a recently found connection between the average-case analysis of the GCC-condition number and covering processes on spheres, and we present a sharp result on the probability distribution of  $\mathcal{C}(A)$  for feasible inputs due to Bürgisser et al. (2007).

The main goal of this survey is to show that part two of Smale's scheme can be naturally refined by performing a smoothed analysis of the condition number  $\mu(x)$  involved. This was already suggested by Spielman and Teng in their ICM 2002 paper. For the matrix condition number, results in this direction were obtained by Wschebor (2004) and Sankar et al. (2006). A recent paper by Tao and Vu (2007) deals with the matrix condition number under random discrete perturbations. Dunagan et al. (2003) gave a smoothed analysis of Renegar's condition

number of linear programming, thereby obtaining for the first time a smoothed analysis for the running time of interior-point methods, see also Spielman and Teng (2003).

A paper by Demmel (1988) has the remarkable feature that the probabilistic average-case analysis performed there for a variety of problems is not done with ad-hoc arguments adapted to the problem considered. Instead, these applications are all derived from a single result bounding the tail of the distribution of a conic condition number in terms of geometric invariants of the corresponding set of ill-posed inputs. Bürgisser et al. (2006, 2008) 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*. This result will be presented in Section 1.6. The critical parameter entering these estimates turned out to be the degree of the defining equations of the set of ill-posed inputs. This result has a wide range of applications to linear and polynomial equation solving, as explained in Section 1.6.1. In particular, it easily gives a smoothed analysis of the condition number of a matrix. Moreover, Amelunxen and Bürgisser (2008) showed that this result, after suitable modification to a spherical convex setting, also allows a smoothed analysis of the GCC-condition number of linear programming.

The mathematical setting of uniform smoothed analysis has a clean and simple description. The set of ill-posed inputs to a computational problem is modelled as a subset  $\Sigma_S$  of a sphere  $S^p$ , which is considered the data space. In most of our applications,  $\Sigma_S$  is an algebraic hypersurface, but for optimization problems  $\Sigma_S$  will be semialgebraic. The corresponding conic condition number  $\mathcal{C}(a)$  of an input  $a \in S^p$  is defined as

$$\mathcal{C}(a) = \frac{1}{\sin d_S(a, \Sigma_S)},$$

where  $d_S$  refers to the angular distance on  $S^p$ . For  $0 \leq \sigma \leq 1$  let  $B(\bar{a}, \sigma)$  denote the spherical cap in the sphere  $S^p$  centered at  $\bar{a} \in S^p$  and having angular radius  $\arcsin \sigma$ . Moreover, we define for  $0 < \epsilon \leq 1$  the  $\epsilon$ -neighborhood of  $\Sigma_S$  as

$$T(\Sigma_S, \epsilon) := \{a \in S^p \mid d_S(a, \Sigma_S) < \arcsin \epsilon\}.$$

The task of a uniform smoothed analysis of  $\mathcal{C}$  consists of providing good upper bounds on

$$\sup_{\bar{a} \in S^p} \text{Prob}_{a \in B(\bar{a}, \sigma)} \{\mathcal{C}(a) \geq \epsilon^{-1}\},$$

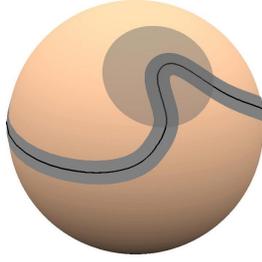


Fig. 1.2. Neighborhood of the curve  $\Sigma_S$  intersected with a spherical disk.

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

$$\text{Prob}_{a \in B(\bar{a}, \sigma)} \{\mathcal{C}(a) \geq \epsilon^{-1}\} = \frac{\text{vol}(T(\Sigma_S, \epsilon) \cap B(\bar{a}, \sigma))}{\text{vol}(B(\bar{a}, \sigma))}. \quad (1.2)$$

Thus uniform smoothed analysis means to provide bounds on the relative volume of the intersection of  $\epsilon$ -neighborhoods of  $\Sigma_S$  with small spherical disks, see Figure 1.2. 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. (Note that  $S^p = B(\bar{a}, 1) \cup B(-\bar{a}, 1)$  for any  $\bar{a}$ .)

In Section 1.7 we explain the rich mathematical background behind our uniform smoothed analysis estimates. We first review classical results on the volume of tubes and then state the principal kinematic formula of integral geometry for spheres. Finally, in Section 1.7.3, we outline the proof of the main Theorem 1.2, which proceeds by estimating the integrals of absolute curvature arising in Weyl's tube formula (1939) with the help of Chern's (1966) principal kinematic formula and Bézout's theorem.

## 1.2 Condition numbers for linear algebra

A numerical computation problem can often be formalized by a mapping  $f: U \rightarrow Y$  between finite-dimensional real or complex vector spaces  $X$  and  $Y$ , where  $U$  is an open subset of  $X$ . The space  $X$  is interpreted as the set of inputs to the problem,  $Y$  is the set of solutions, and  $f$  is the solution map. Small perturbations  $\delta x$  of an input  $x$  result in a

perturbation  $\delta y$  of the output  $y = f(x)$ . In order to quantify this effect with regard to small relative errors, we choose norms on the spaces  $X$  and  $Y$  and define the *relative condition number* of  $f$  at  $x$  by

$$\kappa(f, x) := \lim_{\epsilon \rightarrow 0} \sup_{\|\delta x\| \leq \epsilon \|x\|} \frac{\|f(x + \delta x) - f(x)\| / \|f(x)\|}{\|\delta x\| / \|x\|}.$$

If  $f$  is differentiable at  $x$ , this can be expressed in terms of the operator norm of the Jacobian  $Df(x)$  with respect to the chosen norms:

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

In the case  $X = Y = \mathbb{R}$ , the logarithm of the condition number measures the *loss of precision* when evaluating  $f$ : if we know  $x$  up to  $\ell$  decimal digits, then we know  $f(x)$  roughly up to  $\ell - \log_{10} \kappa(f, x)$  decimal digits.

Consider matrix inversion  $f: \text{GL}(m, \mathbb{R}) \rightarrow \mathbb{R}^{m \times m}$ ,  $A \mapsto A^{-1}$ , measuring errors with respect to the  $L_2$ -operator norm. A perturbation argument shows that the condition number of  $f$  at  $A$  equals the *classical matrix condition number*

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

of the matrix  $A$ . It is easy to see that  $\kappa(A)$  also equals the condition number of the map  $\text{GL}(m, \mathbb{R}) \rightarrow \mathbb{R}^m$ ,  $A \mapsto A^{-1}b$  for fixed nonzero  $b \in \mathbb{R}^m$ . In fact,  $\kappa(A)$  determines the condition number for solving a quadratic linear system of equations. It is also known that  $\kappa(A)$  dominates the condition number of the several other problems of numerical linear algebra, like the Cholesky and QR decomposition of matrices, see Amodè and Dedieu (2008). Moreover, the condition number  $\kappa(A)$  appears in Wilkinson's round-off analysis of Gaussian elimination with partial pivoting (together with the so-called growth factor), see Wilkinson (1963) and Higham (1996).

Let us return to the problem of matrix inversion. We can interpret the set of singular matrices  $\Sigma := \{A \in \mathbb{R}^{m \times m} \mid \det A = 0\}$  as its *set of ill-posed instances*. Let  $\text{dist}(A, \Sigma)$  denote the distance of the matrix  $A$  to  $\Sigma$ , measured with respect to the  $L_2$ -operator norm. The distance of  $A$  to  $\Sigma$  with respect to the Frobenius norm  $\|A\|_F := (\sum_{i,j} a_{ij}^2)^{1/2}$  shall be denoted by  $\text{dist}_F(A, \Sigma)$ . The theorem of Eckart and Young (1936) states that  $\text{dist}(A, \Sigma) = \text{dist}_F(A, \Sigma) = \|A^{-1}\|^{-1}$ . As the right-hand side equals the smallest singular value of  $A$ , this is just a special case of the well-known fact that the  $k$ th largest singular value of  $A$  equals the distance of  $A$  to the set of matrices of rank less than  $k$  (with respect to

the  $L_2$ -operator norm). We rephrase Eckart and Young's result as the following *condition number theorem*

$$\kappa(A) = \frac{\|A\|}{\text{dist}(A, \Sigma)} = \frac{\|A\|}{\text{dist}_F(A, \Sigma)}. \quad (1.3)$$

It is remarkable that the condition number  $\kappa(A)$ , which was defined using local properties, can be characterized in this global geometric way.

Demmel (1987) realized that this observation for the classical matrix condition number actually holds in much larger generality. For numerous computation problems, the condition number of an input  $x$  of norm one, say, can be bounded up to a constant factor by the inverse distance of  $x$  to a corresponding set of ill-posed inputs  $\Sigma$ . It is this key insight that allows to perform probabilistic analyses of condition numbers by geometric tools.

To further illustrate this connection, consider *eigenvalue computations*. Let  $\lambda \in \mathbb{C}$  be a simple eigenvalue of  $A \in \mathbb{C}^{m \times m}$ . The sensitivity to compute  $\lambda$  from  $A$  is captured by a condition number  $\kappa(A, \lambda)$ , see (1.22). Wilkinson (1972) proved that

$$\kappa(A, \lambda) \leq \frac{\sqrt{2} \|A\|_F}{\text{dist}_F(A, \Sigma_{\text{eigen}})},$$

where  $\Sigma_{\text{eigen}}$  is the set of matrices in  $\mathbb{C}^{m \times m}$  having a multiple eigenvalue.

Clearly, condition numbers are a crucial issue when dealing with finite precision computations and round-off errors. When considering iterative methods (instead of direct methods), it turns out that, even when assuming infinite precision arithmetic, the condition of an input often bounds the number of iterations required to achieve a certain precision. A famous example for this phenomenon is the *conjugate gradient method* of Hestenes and Stiefel (1952). For a given linear system  $Ax = b$ ,  $A$  a symmetric positive definite matrix, the conjugate gradient method starts with an initial value  $x_0 \in \mathbb{R}^n$  and produces a sequence of iterates  $x_0, x_1, \dots, x_n = x^*$  satisfying

$$\|x_k - x^*\|_A \leq 2 \left( \frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k \|x_0 - x^*\|_A,$$

where the  $A$ -norm of a vector  $v$  is defined as  $\|v\|_A := (v^T A v)^{1/2}$ . Therefore, roughly  $\frac{1}{2} \sqrt{\kappa(A)} \ln \frac{1}{\epsilon}$  iterations are sufficient in order to achieve  $\|x_k - x^*\|_A \leq \epsilon \|x_0 - x^*\|_A$ .

### 1.3 Condition numbers for convex optimization

We restrict our discussion to feasibility problems in convex conic form. 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  $L(Y, X)$  the space of linear maps from  $Y$  to  $X$  endowed with the operator norm. Given  $A \in L(Y, X)$ , consider the feasibility problem of deciding

$$\exists y \in Y \setminus \{0\} \quad Ay \in K. \quad (1.4)$$

Two special cases of this general framework should be kept in mind. For  $K = \mathbb{R}_+^n$ , the nonnegative orthant in  $\mathbb{R}^n$ , 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  $\mathbb{R}^{n \times n}$ .

The feasibility problem dual to (1.4) is

$$\exists x^* \in X^* \setminus \{0\} \quad A^*x^* = 0, \quad x^* \in K^*. \quad (1.5)$$

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

We denote by  $\mathcal{D}$  the set of instances  $A \in L(Y, X)$  for which the problem (1.4) is strictly feasible, i.e., there exists  $y \in Y$  such that  $Ay \in \text{int}(K)$ . Likewise, we denote by  $\mathcal{P}$  the set of  $A \in L(Y, X)$  such that (1.5) is strictly feasible, i.e., there exists  $x^* \in \text{int}(K^*)$  with  $A^*x^* = 0$ .

$\mathcal{D}$  and  $\mathcal{P}$  are disjoint open subsets of  $L(Y, X)$  and duality in convex optimization implies that  $\mathcal{P}$  is the complement of the closure of  $\overline{\mathcal{D}}$  in  $L(Y, X)$ , cf. Boyd and Vandenberghe (2004). The *conic feasibility problem* is to decide for given  $A \in L(Y, X)$  whether (1.4) or (1.5) holds. The common boundary  $\Sigma := \partial\mathcal{D} = \partial\mathcal{P}$  of the sets  $\mathcal{D}$  and  $\mathcal{P}$  can be considered as the set of ill-posed instances. Indeed, for given  $A \in \Sigma$ , arbitrarily small perturbations of  $A$  may yield instances in both  $\mathcal{D}$  and  $\mathcal{P}$ .

Renegar (1995a) defined the *condition number* of the conic feasibility problem by

$$C(A) := \frac{\|A\|}{\text{dist}(A, \Sigma)}. \quad (1.6)$$

He observed that the number of steps of interior-point algorithms solving the conic feasibility problem can be effectively bounded in terms

of  $C(A)$ . Before elaborating on this important issue, let us characterize the condition number  $C(A)$  in a different way. Suppose there exists  $e \in \text{int}(K)$  such that the unit ball  $B(e, 1)$  centered at  $e$  is contained in  $K$ . We define  $\lambda_{\min}: X \rightarrow \mathbb{R}$  by  $\lambda_{\min}(x) := \max\{t \in \mathbb{R} \mid x - te \in K\}$  and note that  $x \in K \Leftrightarrow \lambda_{\min}(x) \geq 0$ . For  $K = \mathbb{R}_+^n$  and  $e = (1, \dots, 1)$  we have  $\lambda_{\min}(x) = \min_i x_i$ , while in the case  $K = \mathcal{S}_+^n$  and  $e$  being the unit matrix,  $\lambda_{\min}(x)$  equals the minimum eigenvalue of  $x$ .

The problem (1.4) is feasible iff there exists  $y \in Y$  of norm one such that  $\lambda_{\min}(Ay) \geq 0$ . A vector  $y$  maximizing  $\lambda_{\min}(Ay)/\|y\|$  may be interpreted as a best-conditioned solution, due to the following max-min characterization in Cheung et al. (2008):

$$\text{dist}(A, \Sigma) = \left| \max_{\|y\|=1} \lambda_{\min}(Ay) \right|. \quad (1.7)$$

Actually, in that paper a more general result is shown. Suppose we have a multifold conic structure:  $X = X_1 \times \dots \times X_r$ , where  $X_i$  is a normed vector space,  $K = K_1 \times \dots \times K_r$  with regular closed convex cones  $K_i$  in  $X_i$ , and  $e_i \in \text{int}(K_i)$  such that the unit ball centered at  $e_i$  is contained in  $K_i$ . We have a corresponding function  $\lambda_{\min}^i: X_i \rightarrow \mathbb{R}$ . Then (1.4) can be written as

$$\exists y \in Y \setminus \{0\} \quad A_1 y_1 \in K_1, \dots, A_r y_r \in K_r,$$

where  $A_i \in L(Y, X_i)$  is the composition of  $A$  with the projection onto  $X_i$ . Generalizing (1.6), we define the corresponding *multifold condition number*  $\mathcal{C}(A)$  by

$$\mathcal{C}(A) := \left( \min_{B \in \Sigma} \max_i \frac{\|A_i - B_i\|}{\|A_i\|} \right)^{-1}.$$

It is easy to see that  $\mathcal{C}(A) \leq C(A)$  when taking  $\|A\| = \max_i \|A_i\|$ . Note that in the case  $r = 1$  of just one factor, we retrieve  $C(A) = \mathcal{C}(A)$ . The condition number  $\mathcal{C}(A)$  seems a more natural measure of conditioning in the multifold setting, when allowing component normalization as preconditioning. Cheung et al. (2008) proved the following *condition number theorem*, extending (1.7),

$$\frac{1}{\mathcal{C}(A)} = \left| \max_{\|y\|=1} \min_i \frac{\lambda_{\min}^i(A_i y)}{\|A_i\|} \right|. \quad (1.8)$$

Let us now have a closer look at the important special case of  $X_i = \mathbb{R}$ ,  $K_i = \mathbb{R}_+$ ,  $e_i = 1$  for  $i = 1, \dots, n$ . We endow  $X = \mathbb{R}^n$  with the  $L_\infty$ -norm and  $Y := \mathbb{R}^{m+1}$  with the  $L_2$ -norm. The problem (1.4) now reads as the

linear programming feasibility problem

$$\exists y \in \mathbb{R}^{m+1} \setminus \{0\} \quad a_1 y \geq 0, \dots, a_n y \geq 0, \quad (1.9)$$

where  $a_i \in \mathbb{R}^{m+1}$  denote the rows of the given matrix  $A \in \mathbb{R}^{n \times (m+1)}$ , which we may assume to be scaled to euclidean length one. We can therefore interpret the input  $A$  as a sequence of  $n$  points  $a_1, \dots, a_n$  on the unit sphere  $S^m := \{y \in \mathbb{R}^{m+1} \mid \|y\| = 1\}$ . By self-duality of the nonnegative orthant, the feasibility problem (1.5) translates to

$$\exists x \in \mathbb{R}^n \setminus \{0\} \quad A^T x = 0, \quad x \in \mathbb{R}_+^n. \quad (1.10)$$

The multifold condition number  $\mathcal{C}(A)$  corresponding to this setting has been introduced and investigated by Goffin (1980), and Cheung and Cucker (2001). We will refer to it as the *GCC-condition number*.

There is a nice geometric characterization of  $\mathcal{C}(A)$ : Fix an input  $A$ , interpreted as a sequence of points  $a_1, \dots, a_n \in S^m$ . For  $y \in S^m$  we have  $a_i y = \cos \theta_i(y)$ , where  $\theta_i(y) \in [0, \pi]$  denotes the angle between  $y$  and  $a_i$ . Put  $\theta(y) := \max_i \theta_i(y)$ . Then  $\rho(A) := \min_{y \in S^m} \theta(y)$  is the angular radius of a smallest spherical cap enclosing all the points  $a_i$ . This quantity captures the GCC-condition number. Indeed, using  $\lambda_{\min}^i(x_i) = x_i$ , the condition number theorem (1.8) translates to

$$\mathcal{C}(A)^{-1} = |\cos \rho(A)|. \quad (1.11)$$

Moreover, we note that (1.9) is feasible iff  $\rho(A) \leq \pi/2$  and hence  $A \in \Sigma$  iff  $\rho(A) = \pi/2$ .

### 1.3.1 Condition based analysis

We turn now to the relation of conditioning to complexity. Freund and Vera (1999) gave a condition based analysis of Khachyian's (1979) ellipsoid method. The essence of their argument is rather simple so that we are going to sketch it briefly.

Suppose we are in the general conic setting and  $A$  is a feasible instance of (1.4). We define the *width*  $\tau(A)$  of the cone of solutions  $A^{-1}(K)$  as the maximum ratio  $r/\|y\|$  over all balls  $B(y, r)$  contained in  $A^{-1}(K)$ . Let  $y_0 \in Y$  be a best conditioned solution of norm 1, that is, maximizing the right hand side of (1.7). Then it is not hard to see that  $B(y_0, C(A)^{-1}) \subseteq A^{-1}(K)$ , hence  $C(A)^{-1} \leq \tau(A)$ . (In the case  $K = \mathbb{R}_+^n$  we even have  $\mathcal{C}(A)^{-1} \leq \tau(A)$ .)

Suppose now that  $Y = \mathbb{R}^m$  is endowed with the  $L_2$ -norm and consider

the compact convex set  $\tilde{K}_A$  obtained by homogenizing with an additional variable  $t$  and intersecting with the unit ball  $B_{m+1}$ :

$$\tilde{K}_A := \{(y, t) \in Y \times \mathbb{R}_+ \mid Ay \in K, \|y\|^2 + t^2 \leq 1\}.$$

Freund and Vera (1999) showed that

$$\ln \left( \frac{\text{vol } B_{m+1}}{\text{vol } \tilde{K}_A} \right) \leq (m+1) \ln \left( 2 + \frac{6}{\tau(A)} \right),$$

where  $\text{vol}$  denotes the  $(m+1)$ -dimensional volume.

We assume now  $X = \mathbb{R}^n$  and that the convex cone  $K \subseteq \mathbb{R}^n$  is given by a separation oracle, i.e., for a given  $x_0 \in \mathbb{R}^n$  the oracle either answers  $x_0 \in K$  or provides a hyperplane separating  $x_0$  from  $K$ . (Note that for  $K = \mathbb{R}^n$  the separation oracle is trivial.) Running the ellipsoid method (see Grötschel, Lovász and Schrijver (1988)) on the convex set  $\tilde{K}_A$ , starting with the enclosing ball  $B_{m+1}$ , we arrive at the following:

**Theorem 1.1** *The ellipsoid method, applied to the homogenized convex set  $\tilde{K}_A$ , either finds a feasible point  $y \in A^{-1}(K)$  or decides  $A^{-1}(K) = \emptyset$  with a number of iterations bounded by  $2(m+1)^2 \ln(2 + 6C(A))$ . Each iteration step involves one call of the separation oracle plus  $\mathcal{O}(m^2)$  arithmetic operations and one square root for the computation of the next ellipsoid.*

This general result is impractical, but it has the beauty of showing by a simple argument that the complexity of rather general conic feasibility problems is polynomially bounded in the dimensions  $m, n$  and  $\ln C(A)$ . (Of course we assume that the cost of one call to the separation oracle is polynomially bounded in  $n, m$ .)

A great deal of motivation for the work described so far in this section comes from the major open problem whether the linear programming feasibility problem LPF (1.9) can be algorithmically solved with a number of arithmetic operations polynomial in  $m$  and  $n$ . In fact this problem is listed as one of Smale's problems (2000) for the next century. Motivated by this question, Renegar (1995a, 1995b) introduced the condition number  $C(A)$  and proved by interior-point methods that the complexity of LPF is polynomially bounded in  $m, n$  and  $\ln C(A)$ . This considerably added to our understanding of the complexity of LPF. The well-known fact that LPF for rational inputs is solvable in polynomial time in the Turing model is a simple consequence of this. Indeed, it is sufficient to note that for rational matrices  $A \notin \Sigma$ ,  $\ln C(A)$  is polynomially bounded

in the bitsize of  $A$ . (One also has to check that there is no explosion of bit size in the computations, which is straightforward.)

The most efficient known algorithms for solving convex optimization problems in theory and practice are interior-point methods, cf. Nesterov and Nemirovskii (1994). We do not want to enter this vast field and just mention that Cucker and Peña (2002) gave a condition based analysis of a primal-dual interior-point method for solving the linear programming feasibility problem LPF (1.9) with a number of iterations bounded by

$$\mathcal{O}(\sqrt{m+n} (\log(m+n) + \log \mathcal{C}(A))).$$

Hereby, each iteration costs at most  $\mathcal{O}((m+n)^3)$  arithmetic operations. In that paper, for the first time, a round-off analysis of an interior-point algorithm was performed, and it was shown that the amount of precision required can be bounded in terms of  $\mathcal{C}(A)$ . For an early condition based analysis of LPF (in terms of another condition number) we refer to Vavasis and Ye (1995).

A solution to LPF (1.9) can be found by the perceptron method with a number of iterations bounded by  $\mathcal{O}(1/\tau(A)^2)$ , see Rosenblatt (1962). A more efficient, re-scaled version of the perceptron algorithm has been developed by Dunagan and Vempala (2004), which uses only  $\mathcal{O}(n \ln(1/\tau(A)))$  iterations. Recently, this result was extended to conic systems by Belloni, Freund and Vempala (2007).

#### 1.4 Condition numbers for polynomial equation solving

Condition numbers for solving systems of complex polynomial equations were introduced and studied by Shub and Smale (1993a). The geometric viewpoint of looking for roots of homogeneous equations in complex projective space adds a lot to the elegance and mathematical feasibility of the theory.

We briefly review the setting, for more details and a simplified treatment see BCSS (Blum, Cucker, Shub, and Smale, 1998). Fix  $d_1, \dots, d_n \in \mathbb{N} \setminus \{0\}$  and denote by  $\mathcal{H}_d$  the vector space of polynomial systems  $f = (f_1, \dots, f_n)$  with  $f_i \in \mathbb{C}[X_0, \dots, X_n]$  homogeneous of degree  $d_i$ . For  $f, g \in \mathcal{H}_d$  we write

$$f_i(x) = \sum_{\alpha} a_{\alpha}^i X^{\alpha}, \quad g_i(x) = \sum_{\alpha} b_{\alpha}^i X^{\alpha},$$

where  $\alpha = (\alpha_0, \dots, \alpha_n)$  is assumed to range over all multi-indices such that  $|\alpha| = \sum_{k=0}^n \alpha_k = d_i$  and  $X^{\alpha} := X_0^{\alpha_0} X_1^{\alpha_1} \dots X_n^{\alpha_n}$ . The space  $\mathcal{H}_d$  is

endowed with a Hermitian inner product  $\langle f, g \rangle = \sum_{i=1}^n \langle f_i, g_i \rangle$ , where

$$\langle f_i, g_i \rangle = \sum_{|\alpha|=d_i} a_\alpha^i \overline{b_\alpha^i} \binom{d_i}{\alpha}^{-1}.$$

Here, the bar denotes complex conjugate and  $\binom{d_i}{\alpha}$  denotes the multinomial coefficients. The reason for choosing this inner product is that it is invariant under the natural action of the unitary group  $U(n+1)$  on  $\mathcal{H}_d$ . This property is crucial for the whole development. We denote by  $\|f\|$  the corresponding norm of  $f \in \mathcal{H}_d$ .

Let  $\mathbb{P}^n := \mathbb{P}(\mathbb{C}^{n+1})$  and  $\mathbb{P}(\mathcal{H}_d)$  denote the complex projective spaces associated to  $\mathbb{C}^{n+1}$  and  $\mathcal{H}_d$ , respectively. These are complex manifolds that naturally carry the structure of a Riemannian manifold. The solution variety defined as  $V := \{(f, \zeta) \in \mathbb{P}(\mathcal{H}_d) \times \mathbb{P}^n \mid f(\zeta) = 0\}$  is a smooth submanifold of  $\mathbb{P}(\mathcal{H}_d) \times \mathbb{P}^n$  and hence also carries a Riemannian structure. (We identify  $f \in \mathcal{H}_d$  and its corresponding element in  $\mathbb{P}(\mathcal{H}_d)$ .)

The computational problem under investigation is now the following: given  $f \in \mathbb{P}(\mathcal{H}_d)$ , find  $\zeta \in \mathbb{P}^n$  such that  $f(\zeta) = 0$ . Suppose that  $\zeta$  is a simple solution of  $f$ . By the implicit function theorem, the projection map  $V \rightarrow \mathbb{P}(\mathcal{H}_d)$ ,  $(f', \zeta') \mapsto f'$  can be locally inverted around  $(f, \zeta)$ . The solution map  $G$  is the local inverse of this projection. Following the scheme of Section 1.2, it is natural to define the condition number at  $(f, \zeta)$  as the operator norm of the derivative of  $G$  at  $\zeta$ :  $\mu(f, \zeta) := \|DG(\zeta)\|$ . A calculation shows

$$\mu(f, \zeta) = \|f\| \left\| (Df(\zeta)|_{T_\zeta})^{-1} \text{diag}(\|\zeta\|^{d_1-1}, \dots, \|\zeta\|^{d_n-1}) \right\|,$$

where  $Df(\zeta)|_{T_\zeta}$  denotes restriction of the derivative of  $f: \mathbb{C}^{n+1} \rightarrow \mathbb{C}^n$  at  $\zeta$  to the tangent space  $T_\zeta := \{v \in \mathbb{C}^{n+1} \mid \langle v, \zeta \rangle = 0\}$  of  $\mathbb{P}^n$  at  $\zeta$ . Note that  $\mu(f, \zeta)$  is homogeneous of degree 0 in both arguments and hence defined for  $(f, \zeta) \in V$  outside the *subvariety of ill-posed pairs*

$$\Sigma' := \{(f, \zeta) \in V \mid \text{rank } Df(\zeta)|_{T_\zeta} < n\}. \quad (1.12)$$

We remark that  $(f, \zeta) \in \Sigma'$  means that  $\zeta$  is a multiple root of  $f$ .

In order to simplify the statement of the condition number theorem below, one considers the (*normalized*) *condition number* defined as

$$\mu_{\text{norm}}(f, \zeta) := \|f\| \left\| (Df(\zeta)|_{T_\zeta})^{-1} \text{diag}(\sqrt{d_1} \|\zeta\|^{d_1-1}, \dots, \sqrt{d_n} \|\zeta\|^{d_n-1}) \right\|.$$

The *condition number theorem* in Shub and Smale (1993a) gives a characterization of  $\mu_{\text{norm}}$  in terms of the inverse distance to the nearest

ill-posed input. It states that for  $(f, \zeta) \in V \setminus \Sigma'$

$$\mu_{\text{norm}}(f, \zeta) = \frac{1}{\sin d_{\zeta}(f, \Sigma_{\zeta})}. \quad (1.13)$$

Here  $d_{\zeta}(f, \Sigma_{\zeta})$  denotes the distance of  $f$  to  $\Sigma_{\zeta} := \{f' \in \mathbb{P}(\mathcal{H}_d) \mid (f', \zeta) \in \Sigma'\}$  measured in the Riemannian metric of the “fiber”  $\{f' \in \mathbb{P}(\mathcal{H}_d) \mid f'(\zeta) = 0\}$ .

If  $f \in \mathcal{H}_d$  has only simple zeros  $\zeta_1, \dots, \zeta_q$  we define the *condition number of  $f$  for approximating roots* as

$$\mu_{\text{norm}}(f) := \max_{i \leq q} \mu_{\text{norm}}(f, \zeta_i),$$

otherwise we set  $\mu_{\text{norm}}(f) := \infty$ .

In closing this discussion we mention that Wilkinson (1963), Wozniakowski (1977), and Demmel (1987) studied condition numbers for finding the roots of polynomials in one variable.

We now briefly discuss how to compute an approximate zero of a system of polynomial equations by homotopy continuation and how condition numbers enter the complexity estimates.

By an *approximate zero* of  $f$  (in the strict sense) associated with a zero  $\zeta$  of  $f$  we understand a point  $z$  such that the sequence of Newton iterates (adapted to projective space)

$$z_{i+1} := N_f(z_i) := z_i - (Df(z_i)|_{T_{z_i}})^{-1}f(z_i)$$

with initial point  $z_0 := z$  converges immediately quadratically to  $\zeta$ , i.e.,

$$d_R(z_i, \zeta) \leq \left(\frac{1}{2}\right)^{2^i - 1} d_R(z_0, \zeta),$$

for all  $i \in \mathbb{N}$ , where  $d_R$  refers to the Riemannian distance of  $\mathbb{P}^n$ .

Suppose we are looking for a root of  $f \in \mathbb{P}(\mathcal{H}_d)$ . We use a “start system”  $(f_0, \zeta_0) \in V$  and define  $f_t := tf + (1-t)f_0$  for  $t \in [0, 1]$ . If the line segment  $\{f_t \mid t \in [0, 1]\}$  does not meet the *discriminant variety*

$$\Sigma := \{f' \in \mathbb{P}(\mathcal{H}_d) \mid \exists \zeta' (f', \zeta') \in \Sigma'\}, \quad (1.14)$$

then there exists a unique lifting to a “solution curve”  $[0, 1] \rightarrow V, t \mapsto (f_t, \zeta_t)$ . Since  $f_1 = f$ ,  $\zeta_1$  is the root of  $f$  we are looking for. The idea is now to partition  $[0, 1]$  into  $k$  parts by  $t_i = i/k$  for  $i = 0, \dots, k$  and to successively compute approximations  $z_i$  of  $\zeta_{t_i}$  by Newton’s method. More specifically, we set  $z_0 := \zeta_0$  and for  $1 \leq i \leq k$

$$z_i := N_{f_{t_i}}(z_{i-1}).$$

Put  $D := \max_i d_i$  and let  $L$  denote the length of the curve  $(f_t)_{0 \leq t \leq 1}$  in  $\mathbb{P}(\mathcal{H}_d)$ . The main result in Shub and Smale (1993a) states that

$$k = \mathcal{O}\left(L D^2 \max_{t \in [0,1]} \mu_{\text{norm}}(f_t, \zeta_t)^2\right) \quad (1.15)$$

iterations are sufficient to achieve that  $z_i$  is an approximate zero of  $f_{t_i}$ , for all  $1 \leq i \leq k$ . In particular,  $z_k$  is an approximate zero of  $f$ .

So by the condition number theorem (1.13), the number  $k$  of Newton iterations depends on how close the solution curve  $(f_t, \zeta_t)_{t \in [0,1]}$  approaches the variety  $\Sigma'$  of ill-posed pairs. As a suitable start system it has been proposed to take  $g = (X_0^{d_1-1} X_1, \dots, X_0^{d_n-1} X_n)$  together with its zero  $e = (1, 0, \dots, 0)$ .

Let us mention some recent improvements. Shub (to appear) introduced the *condition metric* on the solution variety  $V$  by scaling the Riemannian metric on  $V$  with  $\mu_{\text{norm}}^2$ . He showed that for a given smooth curve  $\gamma: [0, 1] \rightarrow V$  in  $V$ , the number of Newton steps sufficient to follow a homotopy along  $\gamma$  is bounded by  $\mathcal{O}(D^{3/2} \text{Length}(\gamma))$ , where  $\text{Length}(\gamma) := \int_0^1 \mu_{\text{norm}}(\gamma(t)) \|\dot{\gamma}\| dt$  is the length of  $\gamma$  in the condition metric. Beltrán and Shub (to appear) proved that any  $(f, \zeta) \in V$  can be connected to  $(g, e)$  by a curve  $\gamma$  with

$$\text{Length}(\gamma) \leq 9nD^{3/2} + 2\sqrt{n} \ln(\mu_{\text{norm}}(f, \zeta)/\sqrt{n}).$$

Note that this is a much better bound than (1.15), as the condition number  $\mu_{\text{norm}}(f, \zeta)$  has been replaced by its logarithm. Unfortunately the above result is not algorithmic, so that it only suggests, but does not imply a considerable complexity improvement.

### 1.5 Average-case probabilistic analysis

Recall Smale's two part scheme for analyzing iterative numerical algorithms from the introduction. In the previous three sections, we illustrated the first part of this scheme in several important examples. We continue now the discussion of these examples with regard to the second part of the scheme.

The first example is the classical condition number  $\kappa(A) = \|A\| \cdot \|A^{-1}\|$  of a random matrix  $A \in \mathbb{R}^{m \times m}$ . Suppose that the entries of  $A$  are independent standard normal distributed. Edelman (1988) derived sharp estimates on the distribution of  $\kappa(A)$  by analyzing the distribution of the smallest and largest singular value of random matrices. In particular,

he showed that

$$\mathbb{E}(\ln \kappa(A)) = \ln m + c + o(1), \quad m \rightarrow \infty$$

where  $c \approx 1.537$ . Edelman (1992) also gave closed formula expressions for the distribution of the related quantity  $\kappa_F(A) := \|A\|_F \cdot \|A^{-1}\|$ . In the case where the entries of  $A$  are complex numbers with independent standard normal distributed real and imaginary part, the resulting closed form is so amazingly simple, that we state it here:

$$\text{Prob}\{\kappa_F(A) \geq \epsilon^{-1}\} = 1 - (1 - \min\{1, m\epsilon^2\})^{m-1}.$$

We move now to the probabilistic analysis of the GCC-condition number  $\mathcal{C}(A)$  for the linear programming feasibility problem (1.9). The input  $A = (a_1, \dots, a_n)$  is assumed to be uniformly distributed in the product  $(S^m)^n$  of spheres and our goal is to determine the induced probability distribution of  $\mathcal{C}(A)$ . Let  $\rho(A)$  denote the angular radius of a smallest spherical cap enclosing all the points  $a_i$ . The geometric characterization (1.11) states that  $\mathcal{C}(A)^{-1} = |\cos \rho(A)|$ . Moreover, (1.9) is feasible iff  $\rho(A) \leq \pi/2$ . We denote by  $\mathcal{F}_{n,m}$  the set of all feasible instances.

Our problem can be restated as one concerning coverage processes on spheres. For  $\alpha \in [0, \pi]$  let  $p(n, m, \alpha)$  denote the probability that randomly chosen spherical caps with centers  $a_1, \dots, a_n$  and angular radius  $\alpha$  do *not* cover the sphere  $S^m$ . Of course, we assume  $a_i$  to be uniformly and independently chosen from the uniform distribution of  $S^m$ .

We claim that for  $0 < \epsilon \leq 1$

$$\begin{aligned} \text{Prob}\{A \in \mathcal{F}_{n,m} \text{ and } \mathcal{C}(A) \geq \epsilon^{-1}\} &= p(n, m, \pi/2) - p(n, m, \alpha_f(\epsilon)) \\ \text{Prob}\{A \notin \mathcal{F}_{n,m} \text{ and } \mathcal{C}(A) \geq \epsilon^{-1}\} &= p(n, m, \alpha_i(\epsilon)) - p(n, m, \pi/2), \end{aligned}$$

where  $\alpha_i(\epsilon) := \arccos \epsilon \leq \pi/2$  and  $\alpha_f(\epsilon) := \arccos(-\epsilon) \geq \pi/2$ . Indeed, the caps of radius  $\alpha$  with center  $a_1, \dots, a_n$  do not cover  $S^m$  iff there exists  $y \in S^m$  having distance greater than  $\alpha$  from all  $a_i$ . The latter means that the cap of radius  $\pi - \alpha$  centered at  $-y$  contains all the  $a_i$ . Hence  $p(n, m, \alpha) = \text{Prob}\{\rho(A) \leq \pi - \alpha\}$ . This implies

$$p(n, m, \pi/2) - p(n, m, \alpha_f(\epsilon)) = \text{Prob}\{\pi - \alpha_f(\epsilon) \leq \rho(A) \leq \pi/2\}.$$

This equals the probability that  $A \in \mathcal{F}_{n,m}$  and  $\cos \rho(A) \leq \epsilon$ , which can be rewritten as  $A \in \mathcal{F}_{n,m}$  and  $\mathcal{C}(A) \geq \epsilon^{-1}$ . Hence the first assertion follows. The second one is shown similarly.

The problem to determine  $p(n, m, \alpha)$  is classic, see Solomon (1978). It has been completely solved for  $m = 1, 2$ , but little is known for  $m \geq 3$ . If  $n \leq m + 1$  and  $\alpha \leq \pi/2$  it is not hard to see that  $p(n, m, \alpha) = 1$ .

We therefore focus on the more interesting case  $n > m$ . Wendel (1962) showed that

$$\text{Prob}\{A \in \mathcal{F}_{n,m}\} = p(n, m, \pi/2) = 2^{1-n} \sum_{k=0}^m \binom{n-1}{k}. \quad (1.16)$$

Further, a general result by Janson (1986) implies an asymptotic estimate of  $p(n, m, \alpha)$  for  $\alpha \rightarrow 0$ .

Motivated by the probabilistic analysis of the linear programming feasibility problem, Bürgisser et al. (2007) 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  $\alpha \leq \pi/2$  that is asymptotically sharp for  $\alpha \rightarrow \pi/2$ . To state this result, let  $\lambda_m(t)$  denote the relative volume of a spherical cap of radius  $\arccos t \in [0, \pi/2]$  in  $S^m$ . It is well known that for  $t \in [0, 1]$

$$\lambda_m(t) = \frac{\mathcal{O}_{m-1}}{\mathcal{O}_m} \int_0^{\arccos t} (\sin \theta)^{m-1} d\theta,$$

where  $\mathcal{O}_m := \text{vol}(S^m) = 2\pi^{\frac{m+1}{2}}/\Gamma(\frac{m+1}{2})$  denotes the  $m$ -dimensional volume of the sphere  $S^m$ . Put  $\epsilon := |\cos(\alpha)|$ . Bürgisser et al. (2007) proved that for  $\alpha \geq \pi/2$ ,

$$p(n, m, \alpha) = \sum_{k=1}^m \binom{n}{k+1} C(m, k) \int_0^\epsilon t^{m-k} (1-t^2)^{\frac{1}{2}km-1} \lambda_m(t)^{n-k-1} dt.$$

Moreover for  $\alpha \leq \pi/2$ ,  $p(n, m, \alpha) - p(n, m, \pi/2)$  is upper bounded by

$$\binom{n}{m+1} C(m, m) \int_0^\epsilon (1-t^2)^{\frac{m^2-2}{2}} (1-\lambda_m(t))^{n-m-1} dt.$$

The constants  $C(m, k)$  occurring in this formula describe higher moments of the volume of certain random simplices. Their definition is somewhat complicated, but we shall give it for the sake of completeness:

$$C(m, k) := \frac{(k!)^{m-k+1}}{\mathcal{O}_m^k} \text{vol } G_k(\mathbb{R}^m) \int_{M_k} (\text{vol } \Delta)^{m-k+1} d(S^{k-1})^{k+1},$$

where the integral is over the set  $M_k$  of all  $(b_1, \dots, b_{k+1}) \in (S^{k-1})^{k+1}$  containing the origin in their convex hull  $\Delta$ . Further,  $\text{vol } G_k(\mathbb{R}^m)$  denotes the volume of the Grassmannian consisting of the  $k$ -dimensional linear subspaces of  $\mathbb{R}^m$ .

By analyzing the above formulas, Bürgisser et al. (2007) proved that for a random  $A \in \mathbb{R}^{n \times (m+1)}$  with independent standard normal entries ( $n > m$ )

$$\mathbb{E}(\ln \mathcal{E}(A)) \leq 2 \ln(m+1) + 3.31, \quad (1.17)$$

which is the sharpest bound for this expectancy as of today. (Note that the number of inequalities  $n$  does not occur in the upper bound.) Previous results on this were obtained by Cheung and Cucker (2002), Cucker and Wschebor (2003), and Cheung et al. (2005).

Condition number theorems allow a probabilistic analysis of condition numbers in a systematic way by geometric tools. Let us explain this approach for the matrix condition number  $\kappa(A)$  for  $A \in \mathbb{R}^{m \times m}$ . The first step is to replace  $\kappa(A)$  by the slightly larger quantity  $\kappa_F(A) := \|A\|_F \cdot \|A^{-1}\|$ . Note that  $\kappa(A) \leq \kappa_F(A) \leq \sqrt{m} \kappa(A)$ . The point is that by the Eckart-Young Theorem (1.3)

$$\kappa_F(A) = \frac{\|A\|_F}{\text{dist}_F(A, \Sigma)}, \quad (1.18)$$

where  $\Sigma = \{B \in \mathbb{R}^{m \times m} \mid \det B = 0\}$ . If the entries of  $A$  are independent standard normal, then  $A/\|A\|$  is uniformly distributed on the sphere  $S^{m^2-1}$ . Since  $\kappa_F$  is scale-invariant, we may assume that the inputs  $A$  are chosen uniformly at random in  $S^{m^2-1}$ . We also write  $\Sigma_S := \Sigma \cap S^{m^2-1}$ .

The  $\epsilon$ -neighborhood of  $\Sigma_S$ , for  $0 < \epsilon \leq 1$ , is defined as

$$T(\Sigma_S, \epsilon) := \{A \in S^{m^2-1} \mid d_S(A, \Sigma_S) < \arcsin \epsilon\},$$

where  $d_S(A, \Sigma_S) := \inf\{d_S(A, B) \mid B \in \Sigma_S\}$  and  $d_S(A, B)$  denotes the angular (or Riemannian) distance of  $A$  and  $B$  in  $S^{m^2-1}$ . Using  $d_F(A, \Sigma) = \sin d_S(A, \Sigma_S)$  we obtain from (1.18) for  $0 < \epsilon \leq 1$

$$\text{Prob}\{\kappa_F(A) \geq \epsilon^{-1}\} = \frac{\text{vol } T(\Sigma_S, \epsilon)}{\text{vol } S^{m^2-1}}.$$

The task is therefore to compute or to estimate the volume of neighborhoods of  $\Sigma_S$ . This can be achieved by combining Weyl's (1939) tube formula with techniques from integral geometry, as explained in more detail in the next two sections.

It is important to realize that this approach applies to a much more general context than just the matrix condition number. In the context of one variable polynomial equation solving, one can already find the core of these ideas in Smale's early AMS bulletin article dating from 1981. This approach has been elaborated for the average-case probabilistic analysis of various problems by many researchers, as mentioned already in the introduction. The remainder of this survey will be devoted to show how these results on average-case analysis can be naturally refined in the sense of smoothed analysis.

Before doing so, we would like to say a word about what is known on

the average-case analysis of the condition number  $\mu_{\text{norm}}$  for polynomial equation solving introduced in Section 1.4. In Shub and Smale (1993b) it was shown that if  $f \in \mathbb{P}(\mathcal{H}_d)$  is chosen uniformly at random, then we have for  $0 < \epsilon \leq 1/\sqrt{n}$

$$\text{Prob}\{\mu_{\text{norm}}(f) \geq \epsilon^{-1}\} \leq 0.25 n^2(n+1)N(N-1) d_1 \cdots d_n \epsilon^4, \quad (1.19)$$

where  $N = \dim \mathcal{H}_d - 1$ . By combining this with an improvement of (1.15), Shub and Smale (1994) derived the existence of a “nonuniform” algorithm for finding an approximate zero of  $f \in \mathcal{H}_d$  in average polynomial time. The nonuniformity was due to the fact that good starting points of the homotopy were only proven to exist, but were not constructed. Beltrán and Pardo (2008) succeeded to replace the nonuniformity by randomness and described a randomized average polynomial time algorithm for this problem.

### 1.6 Smoothed probabilistic analysis

The condition numbers we have encountered so far fit within the following abstract framework. We assume our data space is a finite-dimensional real Hilbert space, say  $\mathbb{R}^{p+1}$  with the standard scalar product  $\langle \cdot, \cdot \rangle$ . By a semi-algebraic cone  $\Sigma \subseteq \mathbb{R}^{p+1}$  we understand a semi-algebraic set  $\Sigma \neq \{0\}$  that is closed by multiplications with positive scalars. We interpret  $\Sigma$  as a set of *ill-posed inputs* and abstractly define the associated *conic condition number*  $\mathcal{C}(a)$  of  $a \in \mathbb{R}^{p+1} \setminus \{0\}$  as

$$\mathcal{C}(a) := \frac{\|a\|}{\text{dist}(a, \Sigma)}, \quad (1.20)$$

where  $\|\cdot\|$  and  $\text{dist}$  are the norm and distance induced by  $\langle \cdot, \cdot \rangle$ .

The classical matrix condition number  $\kappa(A)$  is not conic since the operator norm  $\|\cdot\|$  is not induced by an inner product. However,  $\kappa(A)$  is upper bounded by  $\kappa_F(A) = \|A\|_F \|A^{-1}\|$ , which, due to the Eckart-Young Theorem (1.18), is conic with respect to the set  $\Sigma \subseteq \mathbb{R}^{m \times m}$  of singular matrices. Likewise, by replacing the operator norm by the Frobenius norm, Renegar’s condition number  $C(A)$  of  $A \in \mathbb{R}^{n \times (m+1)}$ , cf. (1.6), can be replaced by a conic condition number, which differs from  $C(A)$  by at most a factor of  $\sqrt{m+1}$ . Also the condition number  $\mu_{\text{norm}}(f)$  for polynomial equation solving can be analyzed in this general framework, as we will see soon.

Let us continue with the general discussion. Since  $\mathcal{C}(\lambda a) = \mathcal{C}(a)$  for  $\lambda > 0$  we restrict the input data  $a$  to the sphere  $S^p$  and set  $\Sigma_S := \Sigma \cap S^p$ .

Let  $d_S(a, b)$  denote the angular distance of two points  $a$  and  $b$  in  $S^p$  and set  $d_S(a, \Sigma_S) := \inf\{d(a, b) \mid b \in \Sigma_S\}$ . We further assume that  $\Sigma$  is symmetric, i.e.,  $-\Sigma = \Sigma$ , which is actually the case in the examples considered so far, except for the linear programming setting. Then it is easy to see that for  $a \in S^p$  we have

$$\text{dist}(a, \Sigma) = \sin d_S(a, \Sigma_S). \quad (1.21)$$

Recall that for  $0 < \epsilon \leq 1$ , the  $\epsilon$ -neighborhood of  $\Sigma_S$  is defined as

$$T(\Sigma_S, \epsilon) := \{a \in S^p \mid d_S(a, \Sigma_S) < \arcsin \epsilon\}.$$

By the Definition (1.20) we have  $\mathcal{C}(a) > \epsilon^{-1}$  iff  $a \in T(\Sigma_S, \epsilon)$ , for  $a \in S^p$ . Thus an average-case analysis of  $\mathcal{C}(a)$  for  $a$  chosen uniformly at random in  $S^p$  boils down to estimating the volume of  $T(\Sigma_S, \epsilon)$ .

The following model for a smoothed analysis of  $\mathcal{C}(a)$ , proposed in Bürgisser et al. (2006), naturally fits into this geometric framework. Recall that  $B(\bar{a}, \sigma)$  denotes the spherical cap centered at  $\bar{a}$  with angular radius  $\arcsin \sigma$ , for  $0 \leq \sigma \leq 1$ . *Uniform smoothed analysis* of  $\mathcal{C}$  consists of providing good upper bounds on

$$\sup_{\bar{a} \in S^p} \text{Prob}_{a \in B(\bar{a}, \sigma)} \{\mathcal{C}(a) \geq \epsilon^{-1}\},$$

where  $a$  is assumed to be chosen uniformly at random in  $B(\bar{a}, \sigma)$ . The geometric meaning is to provide bounds on the relative volume of the intersection of  $\epsilon$ -neighborhoods of  $\Sigma_S$  with small spherical disks, see Equation (1.2) and Figure 1.2.

The following result from Bürgisser et al. (2008) extends the previously mentioned result by Demmel (1988) from average-case to smoothed analysis. Actually, a sharper bound is proven, for more precise statements see Section 1.7.

**Theorem 1.2** *Let  $\mathcal{C}$  be a conic condition number with set  $\Sigma$  of ill-posed inputs. Assume that  $\Sigma$  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  $0 < \epsilon \leq \sigma/(p(2d+1))$  we have*

$$\sup_{\bar{a} \in S^p} \text{Prob}_{a \in B(\bar{a}, \sigma)} \{\mathcal{C}(a) \geq \epsilon^{-1}\} \leq 26 dp \frac{\epsilon}{\sigma},$$

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

Demmel's 1988 paper 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 Oceanu on the volumes of tubes around real algebraic varieties. A second goal of Bürgisser et al. (2008) was to prove a result akin to Oceanu's. We will outline this proof in Section 1.7.

The setting of conic condition numbers has a natural counterpart over the complex numbers that we want to sketch briefly. Assume that the data space is a finite-dimensional complex Hilbert space, say  $\mathbb{C}^{p+1}$  with the standard hermitean inner product  $\langle \cdot, \cdot \rangle$ . Fix an algebraic cone  $\Sigma \subseteq \mathbb{C}^{p+1}$ , i.e., a zero set of homogeneous complex polynomials, that is interpreted as a set of ill-posed inputs to some computational problem. We define the associated conic condition number  $\mathcal{C}(a)$  of a nonzero  $a \in \mathbb{C}^{p+1}$  as in (1.20). It should be clear that the examples of linear and polynomial equation solving have a natural formulation over  $\mathbb{C}$ .

Since  $\mathcal{C}(a) = \mathcal{C}(\lambda a)$  for  $\lambda \in \mathbb{C}^*$  it is natural to think of the inputs as elements of the complex projective space  $\mathbb{P}^p := \mathbb{P}^p(\mathbb{C})$  and to define their condition number correspondingly. On the space  $\mathbb{P}^p$ , the Fubini-Study metric is a natural way to measure distances, angles and volumes. We do not formally define it, but just note that the induced Riemannian distance  $d_R(a, b)$  of two points  $a, b \in \mathbb{P}^p$  satisfies  $\cos d_R(a, b) = |\langle \hat{a}, \hat{b} \rangle| / (\|\hat{a}\| \|\hat{b}\|)$ , where  $\hat{a}$  and  $\hat{b}$  are affine representatives in  $\mathbb{C}^{p+1}$  of  $a$  and  $b$  in  $\mathbb{P}^p$ . (Hence  $d_R$  has the meaning of an angle as  $d_S$  in the situation over  $\mathbb{R}$ .) Besides the Riemannian metric  $d_R$  on  $\mathbb{P}^p$ , one considers the so-called *projective distance* of points  $a, b \in \mathbb{P}^p$  defined by

$$d_{\mathbb{P}}(a, b) = \sin d_R(a, b).$$

This is motivated by the definition of conic condition numbers. In fact, as for (1.21), one shows that the condition number of  $a \in \mathbb{P}^p$  then takes the form

$$\mathcal{C}(a) = 1/d_{\mathbb{P}}(a, \Sigma)$$

where, abusing notation,  $\Sigma$  is interpreted now as a subset of  $\mathbb{P}^p$ . (In the following we will not distinguish anymore between affine representatives and their corresponding elements of  $\mathbb{P}^p$ .) We denote by  $B_{\mathbb{P}}(a, \sigma)$  the ball of radius  $\sigma$  around  $a$  in  $\mathbb{P}^p$  with respect to projective distance.

In what follows we assume that  $\Sigma$  is purely  $m$ -dimensional, that is, all of its irreducible components are of dimension  $m$ . We recall that the *degree*  $\deg \Sigma$  of  $\Sigma$  in the sense of algebraic geometry can be defined as

the number of intersection points of  $\Sigma$  with a linear subspace of  $\mathbb{P}^p$  of dimension  $p - m$  in general position.

The following general result by Bürgisser et al. (2006) gives a smoothed analysis of conic condition numbers over the complex numbers. We remark that this result, unlike Theorem 1.2, also appropriately covers the case where  $\Sigma$  has codimension greater than one.

**Theorem 1.3** *Let  $\mathcal{C}$  be a conic condition number with set of ill-posed inputs  $\Sigma \subset \mathbb{P}^p$  that is purely  $m$ -dimensional. Then, for all  $\bar{a} \in \mathbb{P}^p$ , all  $0 < \sigma \leq 1$ , and all  $0 < \epsilon \leq (p - m)/(p\sqrt{2})$ , we have*

$$\text{Prob}_{a \in B_{\mathbb{P}}(\bar{a}, \sigma)} \{\mathcal{C}(a) \geq \epsilon^{-1}\} \leq K(p, m) \deg \Sigma \left(\frac{\epsilon}{\sigma}\right)^{2(p-m)} \left(1 + \frac{p}{p-m} \frac{\epsilon}{\sigma}\right)^{2m}$$

and

$$\mathbb{E}_{a \in B_{\mathbb{P}}(\bar{a}, \sigma)} (\ln \mathcal{C}(a)) \leq \frac{\ln K(p, m) + 3 + \ln \deg \Sigma}{2(p-m)} + \ln \frac{pm}{p-m} + 2 \ln \frac{1}{\sigma},$$

with the constant  $K(p, m) := 2 \frac{p^{3p}}{m^{3m}(p-m)^{3(p-m)}}$ .

The proof of this result is based on ideas in Renegar (1987) and Beltrán and Pardo (2007).

### 1.6.1 Applications

Theorem 1.2 and Theorem 1.3 easily imply a smoothed analysis of several of the conic condition numbers we encountered earlier. The next three corollaries are from Bürgisser et al. (2006, 2008).

**Corollary 1.1** *The matrix condition number  $\kappa(A)$  for  $A \in \mathbb{R}^{m \times m}$  satisfies for all  $0 < \sigma \leq 1$*

$$\sup_{\|A\|_F=1} \mathbb{E}_{A \in B(\bar{A}, \sigma)} (\ln \kappa(A)) \leq 6 \ln m + 2 \ln \frac{1}{\sigma} + 4.7.$$

*Proof* We have  $\kappa(A) \leq \kappa_F(A)$ , where  $\kappa_F(A)$  is the conic condition number whose set  $\Sigma$  of ill-posed inputs is the zero set of the determinant, which is a homogeneous polynomial of degree  $m$ . Now apply Theorem 1.2.  $\square$

A smoothed analysis of  $\kappa(A)$  for Gaussian perturbations was previously given by Wschebor (2004) and Sankar et al. (2006) by direct methods.

We discuss now eigenvalue computations. Let  $A \in \mathbb{C}^{m \times m}$  and  $\lambda \in \mathbb{C}$  be a simple eigenvalue of  $A$ . Suppose that  $x \in \mathbb{C}^m$  and  $y \in \mathbb{C}^m$  are right and left eigenvectors associated to  $\lambda$ , respectively (i.e., nonzero and satisfying  $Ax = \lambda x$  and  $y^T A = \lambda y^T$ ). From the fact that  $\lambda$  is a simple eigenvalue, one can deduce that  $\langle x, y \rangle \neq 0$ , cf. Wilkinson (1972). For any sufficiently small perturbation  $\delta A \in \mathbb{C}^{m \times m}$  there exists a unique eigenvalue  $\lambda + \delta\lambda$  of  $A + \delta A$  close to  $\lambda$ . We thus have

$$(A + \delta A)(x + \delta x) = (\lambda + \delta\lambda)(x + \delta x),$$

which implies up to second order terms  $\delta A x + A \delta x \approx \delta\lambda x + \lambda \delta x$ . By multiplying with  $y^T$  from the left we get

$$\delta\lambda = \frac{1}{\langle x, y \rangle} y^T \delta A x + o(\|\delta A\|).$$

Moreover,  $\sup_{\|\delta A\|_F \leq 1} |y^T \delta A x| = \|x\| \|y\|$ .

It therefore makes sense to define the *condition number of  $A$  for the computation of  $\lambda$*  as follows

$$\kappa(A, \lambda) := \frac{\|x\| \|y\|}{|\langle x, y \rangle|} \quad (1.22)$$

and to set  $\kappa(A, \lambda) := \infty$  if  $\lambda$  is a multiple eigenvalue of  $A$ . We further define the *condition number of  $A$  for eigenvalue computations* by

$$\kappa_{\text{eigen}}(A) := \max_{\lambda} \kappa(A, \lambda),$$

where the maximum is over all the complex eigenvalues  $\lambda$  of  $A$ . The set of ill-posed inputs  $\Sigma_{\text{eigen}} := \{A \in \mathbb{C}^{m \times m} \mid \kappa_{\text{eigen}}(A) = \infty\}$  consists of the matrices having multiple eigenvalues. Wilkinson (1972) proved that

$$\kappa_{\text{eigen}}(A) \leq \frac{\sqrt{2} \|A\|_F}{\text{dist}(A, \Sigma)}. \quad (1.23)$$

**Corollary 1.2** *The condition number  $\kappa_{\text{eigen}}(A)$  for  $A \in \mathbb{C}^{m \times m}$  satisfies for all  $0 < \sigma \leq 1$*

$$\sup_{\|\bar{A}\|_F=1} \mathbb{E}_{A \in B(\bar{A}, \sigma)} (\ln \kappa_{\text{eigen}}(A)) \leq 8 \ln m + 2 \ln \frac{1}{\sigma} + 5.$$

*Proof* According to (1.23),  $2^{-1/2} \kappa_{\text{eigen}}$  is bounded by the conic condition number, whose associated set  $\Sigma_{\text{eigen}}$  of ill-posed inputs consists of the matrices  $A$  having multiple eigenvalues.  $\Sigma_{\text{eigen}}$  is the zero set of the discriminant polynomial of the characteristic polynomial, which can be

shown to be homogeneous of degree  $m^2 - m$ . Now apply Theorem 1.3.  $\square$

We remark that it is also possible to derive a corresponding statement for the computation of real eigenvalues of real matrices. However, some care has to be taken when defining the corresponding condition number.

Our next application is concerned with the condition number  $\mu_{\text{norm}}(f)$  for finding an approximate solution of the multivariate polynomial equation  $f(\zeta) = 0$ , where  $f \in \mathcal{H}_d$  (see Section 1.4).

**Corollary 1.3** *For all  $\bar{f} \in \mathcal{H}_d$  of norm one and  $0 < \sigma \leq 1$  we have*

$$\mathbb{E}_{f \in B(\bar{f}, \sigma)}(\ln \mu_{\text{norm}}(f)) \leq 3.5 \ln N + \ln \mathcal{D} + 0.5 \ln n + 2 \ln \frac{1}{\sigma} + 5.$$

where  $N = \dim \mathcal{H}_d - 1$  and  $\mathcal{D} = d_1 \cdots d_n$  is the Bézout number.

Shub and Smale (1993b) obtained similar estimates for the average of  $\ln \mu_{\text{norm}}$ , see (1.19).

*Proof* The discriminant variety  $\Sigma$  consists of the systems  $f \in \mathbb{P}(\mathcal{H}_d)$  having multiple zeros. It is a well-known fact that  $\Sigma$  is a hypersurface in  $\mathbb{P}(\mathcal{H}_d)$  defined by a homogeneous polynomial of total degree at most  $2n\mathcal{D}^2$ , see Bürgisser et al. (2006).

Recall from (1.12) the variety  $\Sigma'$  of ill-posed pairs. The discriminant variety  $\Sigma$  is the projection of  $\Sigma'$  onto the first factor. By the condition number theorem (1.13) we have for all  $(f, \zeta) \in V \setminus \Sigma'$

$$\mu_{\text{norm}}(f, \zeta) = \frac{1}{d_{\mathbb{P}}(f, \Sigma_{\zeta})},$$

where  $d_{\mathbb{P}}(f, \Sigma_{\zeta})$  denotes the projective distance of  $f$  to  $\Sigma_{\zeta} := \{f' \in \mathbb{P}(\mathcal{H}_d) \mid (f', \zeta) \in \Sigma'\}$  measured in the fiber  $\{f' \in \mathbb{P}(\mathcal{H}_d) \mid f'(\zeta) = 0\}$ . Since  $\Sigma_{\zeta} \subseteq \Sigma$  we have  $d_{\mathbb{P}}(f, \Sigma_{\zeta}) \geq d_{\mathbb{P}}(f, \Sigma)$ . Therefore

$$\mu_{\text{norm}}(f, \zeta) \leq \frac{1}{d_{\mathbb{P}}(f, \Sigma)},$$

which implies  $\mu_{\text{norm}}(f) \leq 1/d_{\mathbb{P}}(f, \Sigma)$ . Now apply Theorem 1.3.  $\square$

We remark that it is also possible to derive a corresponding statement for real polynomial systems.

Let us move now to applications to condition numbers of convex optimization. When trying to directly apply Theorem 1.2 we obtain bad bounds. The reason is that the corresponding sets  $\Sigma$  of ill-posed inputs are semialgebraic (of codimension one). Inequalities are essential here

and by enclosing  $\Sigma$  in algebraic hypersurfaces essential information gets lost.

Nevertheless, the proof ideas behind Theorem 1.2 turned out to be useful for obtaining a uniform smoothed analysis of the GCC-condition number  $\mathcal{C}(A)$  of the linear programming feasibility problem (1.9). (For the average-case analysis of  $\mathcal{C}(A)$  see Section 1.5.) The point is that the conclusion of Theorem 1.2 is also true when  $\Sigma$  is the boundary of a spherical convex set.

For stating this precisely let us introduce some notation. By a *convex body*  $K$  in the sphere  $S^m$  we understand the intersection with  $S^m$  of a closed regular convex cone  $C$  in  $\mathbb{R}^{m+1}$ . We call  $T_o(\partial K, \epsilon) := T(\partial K, \epsilon) \setminus K$  the *outer  $\epsilon$ -neighborhood* of the boundary  $\partial K$ . The assertion is

$$\frac{\text{vol}(T_o(\partial K, \epsilon) \cap B(a, \sigma))}{\text{vol} B(a, \sigma)} \leq 6.5 m \frac{\epsilon}{\sigma} \quad \text{if } \epsilon \leq \frac{\sigma}{2m}, \quad (1.24)$$

and the same upper bound holds for the relative volume of the inner  $\epsilon$ -neighborhood of  $\partial K$ .

The relation of this bound to Theorem 1.2 is the following. By convexity, the intersection of  $\partial K$  with a hyperequator of  $S^m$  in general position consists of at most two points. In that sense we may think of  $\partial K$  as a set of “degree” at most two. Of course this analogy has to be taken with a grain of salt. For instance, if  $K$  corresponds to a polyhedral cone  $C$ , then  $\partial K$  can be expressed as the zeroset of a polynomial equation and inequality constraints. However, the degree of this equation would be the number of facets of  $C$ , which is in general a huge number. We will outline the proof of (1.24) in Section 1.7.3.

The smoothed analysis of the GCC-condition number is performed in the following model. Fix  $0 < \sigma \leq 1$  and  $\bar{a}_1, \dots, \bar{a}_n \in S^m$ . Independently choose points  $a_i$  uniformly at random in the spherical caps  $B(\bar{a}_i, \epsilon)$  of  $S^m$  centered at  $\bar{a}_i$  with angular radius  $\arcsin \sigma$ . In other words,  $A = (a_1, \dots, a_n)$  is chosen uniformly at random in  $B(\bar{A}, \alpha) := \prod_i B(\bar{a}_i, \epsilon)$ . We recall that  $\mathcal{F}_{n,m}$  denotes the set of feasible instances in  $(S^m)^n$ .

The following recent result is from Amelunxen and Bürgisser (2008).

**Theorem 1.4** *For  $n > m$  and  $0 < \sigma \leq 1$  we have*

$$\sup_{\bar{A} \in (S^m)^n} \mathbb{E}_{A \in B(\bar{A}, \alpha)} (\ln \mathcal{C}(A)) = \mathcal{O}\left(\ln\left(\frac{nm}{\sigma}\right)\right).$$

For the average-case ( $\sigma = 1$ ) we even get  $\mathbb{E}(\ln \mathcal{C}(A)) = \mathcal{O}(\log m)$ , as already stated in (1.17). Moreover, we have for  $0 < \epsilon \leq \sigma/(2m(m+1))$

$$\sup_{\bar{A} \in (S^m)^n} \text{Prob}\{A \in \mathcal{F}_{n,m}, \mathcal{C}(A) \geq \epsilon^{-1}\} \leq 6.5 nm(m+1) \frac{\epsilon}{\sigma}.$$

For the infeasible case ( $A \notin \mathcal{F}_{n,m}$ ) a slightly worse tail estimate holds.

Dunagan et al. (2003) previously gave a smoothed analysis of Renegar's condition number. The crucial ingredient of their proof is a result due to Ball (1993) about the measure of Gaussians on boundaries of convex sets in euclidean space. Our proof of Theorem 1.4 roughly uses the same overall strategy as Dunagan et al. (2003). However, we substitute Ball's result by the volume estimate (1.24) on neighborhoods of boundaries of spherically convex sets. A relevant observation that enables us to successfully apply this estimate is the following result. Let  $\mathcal{F}_{n,m}^\circ := \mathcal{F}_{n,m} \setminus \partial \mathcal{F}_{n,m}$  denote the set of strictly feasible instances.

**Lemma 1.1** *Let  $A = (a_1, \dots, a_n) \in \mathcal{F}_{n,m}^\circ$  and  $\mathcal{C}(A) \geq (m+1)\epsilon^{-1}$ . Then there exists  $i \in \{1, \dots, n\}$  such that  $a_i \in T_o(\partial K_i, \epsilon)$ , where  $-K_i$  is the spherical convex hull of  $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n$ .*

The proof of the probability tail estimate in Theorem 1.4 for the feasible case is now easy. Suppose that  $A$  is chosen uniformly at random in  $B(\bar{A}, \sigma)$ . Lemma 1.1 yields with  $t = (m+1)\epsilon^{-1}$

$$\text{Prob}\{A \in \mathcal{F}_{n,m}^\circ, \mathcal{C}(A) \geq t\} \leq \sum_{i=1}^n \text{Prob}\{A \in \mathcal{F}_{n,m}^\circ, a_i \in T_o(\partial K_i, \epsilon)\}.$$

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

$$\begin{aligned} & \text{Prob}\{A' \in \mathcal{F}_{n-1,m}^\circ \text{ and } a_n \in T_o(\partial K_n, \epsilon)\} \\ &= \frac{1}{\text{vol } B(\bar{A}', \sigma)} \int_{A' \in \mathcal{F}_{n-1,m}^\circ \cap B(\bar{A}', \sigma)} \text{Prob}\{a_n \in T_o(\partial K_n, \epsilon) \mid A'\} dA'. \end{aligned}$$

Fix now  $A' \in \mathcal{F}_{n-1,m}^\circ$  and consider the convex set  $K_n$  in  $S^m$ . The volume bound (1.24) on the outer neighborhood of  $\partial K_n$  yields

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

We conclude that

$$\text{Prob}\{A \in \mathcal{F}_{n,m}^\circ, a_n \in T_o(\partial K_n, \epsilon)\} \leq 6.5 m \frac{\epsilon}{\sigma}.$$

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

$$\text{Prob}\{A \in \mathcal{F}_{n,m}^\circ \text{ and } \mathcal{C}(A) \geq t\} \leq 6.5 nm(m+1) \frac{1}{\sigma t},$$

which is the bound claimed in Theorem 1.4.

One of the goals of our current research with Amelunxen is to find a general result providing smoothed analysis estimates of condition numbers for convex optimization, in particular for semidefinite programming. The proof just presented heavily relies on the product structure of the cone  $\mathbb{R}_+^{m+1}$  and does not generalize.

### 1.7 Tools from integral geometry

As already pointed out, uniform smoothed analysis boils down to the task of providing bounds on the relative volume of the intersection of  $\epsilon$ -neighborhoods of  $\Sigma_S$  with small spherical disks, see Figure 1.2.

Even though the volume of neighborhoods of subsets of euclidean spaces or spheres is a rich and thoroughly studied mathematical topic, as can be seen from the textbook by Gray (1990), further developments were needed to arrive at the results mentioned in Section 1.6. In the following we describe first some of the classic results on the volume of neighborhoods, then we discuss the principal kinematic formula and finally indicate how to combine these tools in order to prove Theorem 1.2.

#### 1.7.1 On the volume of tubes

To warm up, assume that  $K$  is a convex compact subset of  $\mathbb{R}^n$ . Consider the  $\epsilon$ -neighborhood  $K_\epsilon$  of  $K$  consisting of the points in  $\mathbb{R}^n$  having (euclidean) distance at most  $\epsilon$  from  $K$ . Steiner (1840) observed that the volume of  $K_\epsilon$  is a polynomial function in  $\epsilon$ :  $\text{vol } K_\epsilon = \sum_{i=0}^n c_i(K) \epsilon^i$ . Clearly,  $c_0(K)$  equals the volume of  $K$ , and it should be intuitively clear that  $c_1(K)$  equals the  $(n-1)$ -dimensional volume of the boundary  $\partial K$  of  $K$ . It is an easy and instructive exercise to prove Steiner's result for convex polytopes in  $\mathbb{R}^2$  and  $\mathbb{R}^3$ . This exercise also reveals the meaning of the coefficients  $c_i(K)$ . (For instance,  $c_n(K)$  always equals the volume  $\mathcal{O}_{n-1}/n$  of the  $n$ -dimensional unit ball.) In Minkowski's theory of convex bodies, the coefficients  $c_i(K)$  are called cross-sectional measures of  $K$  (Quermassintegrale), see Bonnesen and Fenchel (1974) for more information. So the volume of the outer  $\epsilon$ -neighborhood  $T_o(\partial K, \epsilon)$  of  $\partial K$  satisfies  $\text{vol } T_o(\partial K, \epsilon) = \sum_{i=1}^n c_i(K) \epsilon^i$ . Weyl (1939) considerably extended this

observation by showing that the volume of the  $\epsilon$ -neighborhood  $T(M, \epsilon)$  of a compact smooth submanifold  $M$  of  $\mathbb{R}^n$  is a polynomial function in  $\epsilon$ , for sufficiently small values of  $\epsilon$ .

For our purposes, we need to study the case where the ambient space is a sphere  $S^n$ . Weyl (1939) also analyzed this case. We will only state his result in the special case where  $M$  is a smooth oriented hypersurface of  $S^n$ . In order to do so, we first need to review a few elementary concepts from differential geometry, see for instance Thorpe (1993) or do Carmo (1992), p. 129.

We assume that a unit normal vector field  $\nu$  has been chosen on  $M$  (which corresponds to the choice of an orientation of  $M$ ). Let  $T_x M$  denote the tangent space of  $M$  at  $x \in M$ . The *second fundamental form*  $\Pi_M(x): T_x M \times T_x M \rightarrow \mathbb{R}$  of  $M$  at  $x$  is defined as  $\Pi_M(x)(u, w) := -\langle \nabla_u \nu(x), w \rangle$ . Here,  $\nabla_u \nu(x)$  denotes the covariant derivative of  $\nu$  at  $x$  in direction  $u$ . It can be computed by taking the derivative of  $\nu: M \rightarrow \mathbb{R}^{n+1}$  at  $x$  in direction  $u$  and projecting orthogonally onto  $T_x S^n$ . It is well known that  $\Pi_M(x)$  is a symmetric bilinear form. Its eigenvalues  $\kappa_1(x), \dots, \kappa_{n-1}(x)$  are called the *principal curvatures* at  $x$  of the hypersurface  $M$ . For  $1 \leq i < n$  we define the  *$i$ th curvature*  $K_{M,i}(x)$  of  $M$  at  $x$  as the  $i$ th elementary symmetric polynomial in  $\kappa_1(x), \dots, \kappa_{n-1}(x)$ , and put  $K_{M,0}(x) := 1$ . In particular,  $K_{M,n-1}(x) = \det L_M(x)$ . We define the *integral*  $\mu_i(M)$  of  *$i$ th curvature* and the *integral*  $|\mu_i|(M)$  of  *$i$ th absolute curvature* of  $M$  as follows ( $0 \leq i \leq n-1$ ):

$$\mu_i(M) := \int_M K_{M,i} dM, \quad |\mu_i|(M) := \int_M |K_{M,i}| dM.$$

For reasons that will become apparent soon, it is more convenient to think in terms of the *normalized integrals of (absolute) curvature* of  $M$  defined by

$$\mu_i^{\text{no}}(M) := \frac{2}{\mathcal{O}_{n-i-1} \mathcal{O}_i} \mu_i(M), \quad |\mu_i^{\text{no}}|(M) := \frac{2}{\mathcal{O}_{n-i-1} \mathcal{O}_i} |\mu_i|(M).$$

Note that  $\mu_0^{\text{no}}(M) = |\mu_0^{\text{no}}|(M) = \mathcal{O}_{n-1}^{-1} \text{vol } M$  is the volume of  $M$  relative to the volume of  $S^{n-1}$ .

For  $0 < \epsilon \leq 1$  we define the  $\epsilon$ -tube  $T^\perp(M, \epsilon)$  around  $M$  as the set of points in  $S^n$  such that there exists a great circle segment in  $S^n$  of angular length less than  $\arcsin \epsilon$  that connects  $x$  with a point in  $M$  and intersects  $M$  orthogonally in that point, see Figure 1.3. Note that  $T^\perp(M, \epsilon) \subseteq T(M, \epsilon)$ . If  $M$  has a smooth boundary, then  $T(M, \epsilon)$  is the

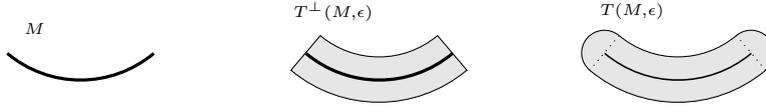


Fig. 1.3.  $\epsilon$ -tube  $T^\perp(M, \epsilon)$  and  $\epsilon$ -neighborhood  $T(M, \epsilon)$  around the curve  $M$ .

union of  $T^\perp(M, \epsilon)$  and a “half-tube” around the boundary of  $\partial M$ . If  $\partial M = \emptyset$ , then  $T^\perp(M, \epsilon) = T(M, \epsilon)$ .

Weyl’s formula now states that for sufficiently small  $\epsilon$  we have

$$\text{vol } T^\perp(M, \epsilon) = \sum_{\substack{0 \leq i \leq n-1 \\ i \text{ even}}} \mu_i^{\text{no}}(M) \text{vol } T(S^{n-i-1}, \epsilon). \quad (1.25)$$

Here  $S^{n-i-1}$  is interpreted as a subset of  $S^n$ . (There is a cancellation effect between the contributions of “outer” and “inner” neighborhoods which results in the sum being only over even indices  $i$ .) If  $M$  is an open subset of  $S^{n-1}$ , then  $\mu_i^{\text{no}}(M) = 0$  for  $i > 0$  and (1.25) specializes to the obvious formula  $\text{vol } T^\perp(M, \epsilon) = \mu_0^{\text{no}}(M) \text{vol } T(S^{n-1}, \epsilon)$ , which is asymptotically equal to  $2\epsilon \text{vol } M$  for  $\epsilon \rightarrow 0$ . For completeness, let us also mention that

$$\text{vol } T(S^{n-i-1}, \epsilon) = \mathcal{O}_{n-i-1} \mathcal{O}_i \int_0^{\arcsin \epsilon} (\cos \rho)^{n-i-1} (\sin \rho)^i d\rho.$$

By tracing Weyl’s proof, it is not hard to see that the following upper bound is valid for all  $0 < \epsilon \leq 1$ :

$$\text{vol } T^\perp(M, \epsilon) \leq \sum_{i=0}^{n-1} |\mu_i^{\text{no}}|(M) \text{vol } T(S^{n-i-1}, \epsilon). \quad (1.26)$$

The question is now how to bound the normalized integrals  $|\mu_i^{\text{no}}|(M)$  of absolute curvature in specific situations. It turns out that this can be effectively done with tools from integral geometry. In a first step, we focus on  $|\mu_0^{\text{no}}|(M)$ , that is, we need to bound the volume of  $M$ .

### 1.7.2 The principal kinematic formula

The orthogonal group  $G := O(n+1)$  is a compact Lie group. It has an invariant Riemannian metric (induced by the euclidean metric on the space of real  $n+1$ -matrices) and a corresponding invariant volume form (Haar measure). So we can talk about random elements of  $G$  chosen

with respect to the uniform distribution. Note that  $G$  acts on  $S^n$  in a straightforward way.

Now suppose that  $M, N \subseteq S^n$  are smooth submanifolds of dimension  $m$  and  $p$ , respectively, such that  $m + p \geq n$ . By transversality principles, the intersection of  $M$  with a random translate  $gN$  of  $N$  is almost surely empty or a submanifold of dimension  $m + p - n$ . So the volume of  $M \cap gN$  in this dimension is well defined (we put  $\text{vol } \emptyset := 0$ ). A key result in integral geometry states that

$$\mathbb{E}_{g \in G} \left( \frac{\text{vol}(M \cap gN)}{\mathcal{O}_{m+p-n}} \right) = \frac{\text{vol } M}{\mathcal{O}_m} \cdot \frac{\text{vol } N}{\mathcal{O}_p}. \quad (1.27)$$

In fact the smoothness assumption on  $M$  and  $N$  in this formula is not important since removing lower dimensional parts does not change the volume (for instance it is sufficient to require that  $M, N$  are semialgebraic).

Santaló (1976), which is the standard reference on integral geometry, refers to (1.27) as *Poincaré's formula* (cf. §7.1 in Santaló). Apparently, Poincaré stated this result for the case of  $S^2$  and in such form it was also known to Barbier (1860). Formula (1.27) is stated in §18.6 of Santaló's book, but a proof is only given in §15.2 for an analogous statement for euclidean space. The book by Howard (1993) states and proves formulas like (1.27) for homogeneous spaces in great generality.

The following corollary of (1.27) allows to reduce the estimation of volumes to counting arguments:

$$\frac{\text{vol } M}{\mathcal{O}_m} = \frac{1}{2} \mathbb{E}_{g \in G} (\#(M \cap gS^{n-m})). \quad (1.28)$$

To illustrate this with a simple example, assume that  $M$  is the real zero set in  $S^n$  of a homogeneous polynomial  $f$  of degree  $d$ . Suppose that  $\dim M = n - 1$ . We claim that if  $M \cap gS^1$  is finite, then it has at most  $2d$  points. In order to see this assume w.l.o.g.  $g = \text{id}$  and  $f(1, 0, \dots, 0) \neq 0$ . Suppose that  $S^1$  is given by  $x_2 = \dots = x_n = 0$ . For each  $x_0 \in \mathbb{R}$  such that  $f(x_0, 1, 0, \dots, 0) = 0$  there are two points  $\pm(1 + x_0^2)^{-1/2}(x_0, 1, 0, \dots, 0)$  in  $M \cap S^1$  and these are all the points in  $M \cap S^1$ . Equation (1.28) then implies that

$$\frac{\text{vol } M}{\mathcal{O}_{n-1}} = \frac{1}{2} \mathbb{E}_{g \in G} (\#(M \cap gS^1)) \leq d \cdot \text{Prob}_{g \in G} \{M \cap gS^1 \neq \emptyset\} \leq d.$$

More generally, we obtain for  $a \in S^n$  and  $0 < \sigma \leq 1$

$$\begin{aligned} \frac{\text{vol}(M \cap B(a, \sigma))}{\mathcal{O}_{n-1}} &\leq d \cdot \text{Prob}_{g \in G} \{M \cap B(a, \sigma) \cap gS^1 \neq \emptyset\} \\ &\leq d \cdot \text{Prob}_{g \in G} \{B(ga, \sigma) \cap S^1 \neq \emptyset\} = d \frac{\text{vol}T(S^1, \sigma)}{\mathcal{O}_n}. \end{aligned}$$

We remark that a statement analogous to (1.27) holds for complex projective spaces  $\mathbb{P}^n(\mathbb{C})$ . The same argument as before then shows that for a complex  $m$ -dimensional algebraic subvariety  $M$  of  $\mathbb{P}^n(\mathbb{C})$  we have  $\text{vol}M = \text{deg}M \cdot \text{vol}(\mathbb{P}^m(\mathbb{C}))$ .

Formula (1.27) has a natural extension involving the normalized integrals  $\mu_i^{\text{no}}(M)$  of curvature. This is called the principal kinematic formula of integral geometry, which is considered the most important result of integral geometry. We will only need the following special case extending (1.28). As before let  $M$  be an oriented smooth hypersurface  $M$  in  $S^n$ . Then we have for  $0 \leq i \leq n-1$

$$\mu_i^{\text{no}}(M) = \mathbb{E}_{g \in G} (\mu_i^{\text{no}}(M \cap gS^{i+1})). \quad (1.29)$$

Note that for almost all  $g$ ,  $M \cap gS^{i+1}$  is either empty or a smooth hypersurface of the sphere  $gS^{i+1}$  (with a canonical orientation inherited from  $M$ ). With this interpretation  $\mu_i^{\text{no}}(M \cap gS^{i+1})$  is well defined (setting  $\mu_i^{\text{no}}(\emptyset) := 0$ ).

The general principal kinematic formula for spheres is so beautiful that we cannot resist to state it here. Also, this could be useful for future applications, when the set of ill-posed inputs has higher codimension.

Let  $M \subseteq S^n$  be a smooth submanifold of dimension  $m$ . For  $x \in M$  let  $S_x := S(T_x M^\perp)$  denote the sphere of unit normal vectors  $v$  in  $T_x S^n$  that are perpendicular to  $T_x M$ . Let us denote by  $K_{M,i}(x, v)$  the  $i$ th elementary symmetric polynomial in the eigenvalues of the second fundamental form of the embedding  $M \hookrightarrow S^n$  at  $x$  in direction  $v$ , see do Carmo (1992), p. 128. We now define the *normalized integral*  $\mu_i^{\text{no}}(M)$  of  $i$ th curvature of  $M$  as ( $0 \leq i \leq m$ ):

$$\mu_i^{\text{no}}(M) := \frac{1}{\mathcal{O}_{m-i} \mathcal{O}_{n-m+i-1}} \int_{x \in M} \int_{v \in S_x} K_{M,i}(x, v) dS_x(v) dM(x).$$

This value is easily seen to vanish if  $i$  is odd (consider  $v \mapsto -v$ ). It follows from Weyl (1939) that  $\mu_i^{\text{no}}(M)$  is a relative isometric invariant of  $M$  in the sense that  $\mu_i^{\text{no}}(M)$  can be written as an integral over  $M$  of a function whose value at  $x \in M$  only depends on the difference of the

values at  $x$  of the curvature tensor of  $M$  and the curvature tensor of  $S^n$  restricted to  $M$ .

Weyl's formula, extending (1.25), states that for sufficiently small  $\epsilon$  we have

$$T^\perp(M, \epsilon) = \sum_{\substack{0 \leq i \leq m \\ i \text{ even}}} \mu_i^{\text{no}}(M) \text{vol} T(S^{m-i}, \epsilon). \quad (1.30)$$

The principal kinematic formula for spheres is best stated in terms of the *curvature polynomial*  $\mu^{\text{no}}(M; X)$  of  $M$  defined as

$$\mu^{\text{no}}(M; X) := \sum_{i=0}^m \mu_i^{\text{no}}(M) X^i,$$

where  $X$  denotes a formal variable. The degree of  $\mu^{\text{no}}(M; X)$  is at most the dimension  $m$  of  $M$ . For example we have  $\mu^{\text{no}}(S^m; X) = 1$ .

The *principal kinematic formula* says that for smooth submanifolds  $M$  and  $N$  of  $S^n$  having dimension  $m$  and  $p$ , respectively, such that  $m + p \geq n$ , we have

$$\mathbb{E}_{g \in G}(\mu^{\text{no}}(M \cap gN; X)) \equiv \mu^{\text{no}}(M; X) \cdot \mu^{\text{no}}(N; X) \text{ mod } X^{m+p-n+1}.$$

Here, the expectation on the left-hand side is defined coefficientwise, while on the right-hand side we have a *polynomial multiplication modulo*  $X^{m+p-n+1}$ . This makes perfect sense as  $m + p - n$  is the expected dimension of  $M \cap gN$ . We note that the principal kinematic formula contains (1.29) as a special case (for even  $i$ ).

It is not all easy to locate the principal kinematic formula for spheres in the above explicit form in the literature. Santaló in his book attributes the principal kinematic formula in the plane to Blaschke, and in euclidean spaces to Chern (1966) and Federer (1959). An elementary and unconventional introduction to geometric probability and the kinematic formula for euclidean spaces can be found in Klain and Rota (1997). The normalization of integrals of curvatures leading to the simple formula of reduced polynomial multiplication was discovered by Nijenhuis (1974), again for euclidean space. Santaló derives the principal kinematic formula for the special case of intersections of domains in spheres, but he does not care about the scaling coefficients. In fact, the principal kinematic formulas for submanifolds of spheres and euclidean spaces take exactly the same form. An indication of this at first glance astonishing fact can be found, somewhat hidden, in Santaló's book on page 320. The situation was clarified by Howard (1993), who gave a unified treatment of kinematic formulas in homogeneous spaces. But Howard does

not care about the scaling constants either. For the purpose of explicitly bounding the volumes of tubes, a good understanding of the scaling factors is relevant. To our best knowledge, the general principal kinematic formula for spheres in the above form was first explicitly stated in Bürgisser (2007).

### 1.7.3 Bounding integrals of absolute curvature

The basic idea is best explained with the example of a hypersurface in  $\mathbb{R}^n$ . The approach is inspired by Spivak (1979), p. 409ff. Suppose that  $f$  is a polynomial of degree  $d$  with compact zero set  $Z \subseteq \mathbb{R}^n$  such that the gradient of  $f$  does not vanish on  $Z$ . Consider the Gauss map  $\nu: Z \rightarrow S^{n-1}$ ,  $x \mapsto \text{grad}f(x)/\|\text{grad}f(x)\|$ . The Jacobian determinant of  $\nu$  at  $x \in Z$  yields the Gaussian curvature:  $K_{Z,n-1}(x) = \det D\nu(x)$ . Let  $\varphi(y) := \#\nu^{-1}(y)$  denote the size of the fiber of  $y \in S^{n-1}$ . and put  $\varphi(\pm y) := \varphi(y) + \varphi(-y)$ . The transformation formula implies

$$\int_Z |\det D\nu| dZ = \int_{S^{n-1}} \varphi(y) dS^{n-1} = \int_{S^{n-1}} \varphi(-y) dS^{n-1}.$$

Hence we obtain  $\int_Z |\det D\nu| dZ = \frac{1}{2} \int_{S^{n-1}} \varphi(\pm y) dS^{n-1}$ .

A point  $x \in \mathbb{R}^n$  satisfies  $\nu(x) = (\pm 1, 0, \dots, 0)$  iff

$$f(x) = 0, \partial_2 f(x) = 0, \dots, \partial_n f(x) = 0.$$

If  $y$  is a regular point of  $\nu$ , then all real solutions of this system of equations are nondegenerate, hence they are isolated in  $\mathbb{C}^n$ . We conclude  $\varphi(\pm y) \leq d(d-1)^{n-1}$  from Bézout's theorem, which is a standard result from algebraic geometry, see Mumford (1976). This estimate holds for any regular value  $y \in S^{n-1}$ . We therefore obtain that

$$\int_Z |K_{Z,n-1}| dZ \leq \frac{\mathcal{O}_{n-1}}{2} d(d-1)^{n-1}.$$

Note that this bound is sharp for  $d = 2$  and  $Z = S^{n-1}$ .

The previous reasoning can be extended to hypersurfaces in  $S^n$  as follows. Suppose now that  $f \in \mathbb{R}[X_0, \dots, X_n]$  is homogeneous of degree  $d$  with zero set  $M \subseteq S^n$  such that the derivative of the restriction of  $f$  to  $S^n$  does not vanish on  $M$ . Then  $M$  is a compact smooth hypersurface of  $S^n$  oriented by the gradient of  $f$ . We claim that

$$|\mu_{n-1}^{\text{no}}|(M) \leq d(d-1)^{n-1}. \quad (1.31)$$

Before showing this bound, let us illustrate it with a simple example.

The zero set of  $f = \sum_{i=1}^n X_i^2 - \epsilon^2 X_0^2$  consists of two small circles in  $S^n$  centered at  $(\pm 1, 0, \dots, 0)$  with a radius going to zero as  $\epsilon \rightarrow 0$ . For each of the circles  $C_\epsilon$ , the total Gaussian curvature  $\mu_{n-1}(C_\epsilon) = |\mu_{n-1}|(C_\epsilon)$  converges to  $\mathcal{O}_{n-1}$  as  $\epsilon \rightarrow 0$ . Hence  $|\mu_{n-1}^{\text{no}}|(C_\epsilon) \rightarrow 1$ . This shows that (1.31) is a sharp bound for  $d = 2$ .

In order to prove (1.31), consider the Gauss map  $\nu: M \rightarrow S^n$  defined as before. For simplicity, we assume that the image  $N$  of  $\nu$  is a smooth hypersurface of  $S^n$  (this can be achieved by removing lower dimensional parts). Again put  $\varphi(y) := \#\nu^{-1}(y)$  for  $y \in N$ . The transformation formula gives

$$|\mu_{n-1}|(M) = \int_M |\det D\nu| dM = \int_N \varphi dN = \sum_{\ell \in \mathbb{N}} \ell \text{vol } F_\ell,$$

where  $F_\ell := \{y \in N \mid \varphi(y) = \ell\}$ . Poincaré's formula (1.28) implies

$$\text{vol } F_\ell = \frac{\mathcal{O}_{n-1}}{2} \mathbb{E}_{g \in G} (\#(F_\ell \cap gS^1)).$$

Therefore,

$$\sum_{\ell \in \mathbb{N}} \ell \text{vol } F_\ell = \frac{\mathcal{O}_{n-1}}{2} \mathbb{E} \left( \sum_{\ell \in \mathbb{N}} \ell \#(F_\ell \cap gS^1) \right) = \frac{\mathcal{O}_{n-1}}{2} \mathbb{E} (\#\nu^{-1}(gS^1)).$$

Now  $gS^1$  intersects  $N$  transversally for almost all  $g \in G$ . To simplify notation suppose this is the case for  $g = \text{id}$ . A point  $x \in \mathbb{R}^{n+1}$  lies in  $\nu^{-1}(S^1)$  iff it satisfies the following system of equations

$$\sum_{i=0}^n x_i^2 - 1 = 0, \quad f(x) = 0, \quad \partial_2 f(x) = \dots = \partial_n f(x) = 0.$$

By Bézout's theorem, the number of solutions to this system of equations is bounded by  $2d(d-1)^{n-1}$ . Altogether,  $\#\nu^{-1}(gS^1) \leq 2d(d-1)^{n-1}$  for almost all  $g$  and the assertion (1.31) follows.

Similarly, one shows that if  $K$  is a convex body in  $S^n$  with smooth boundary  $\partial K$ , then  $|\mu_{n-1}^{\text{no}}|(\partial K) \leq 1$ , which is an optimal bound. The argument is as before, replacing Bézout's theorem by the fact that if  $\partial K \cap gS^1$  is finite, then it consists of at most two points by convexity. (Compared to (1.31) we save here a factor 2 since  $K$  does not contain diametral points.)

Now let the hypersurface  $M$  of  $S^n$  be given as before as the zero set of the homogeneous polynomial  $f$  of degree  $d$ . Let  $a \in S^n$  and  $0 < \sigma \leq 1$ . We can bound the  $i$ th integral of absolute curvature  $|\mu_i^{\text{no}}|(M \cap B(a, \sigma))$

in terms of the degree  $d$  and the dimension parameters  $n, i$  as follows:

$$|\mu_i^{\text{no}}|(M \cap B(a, \sigma)) \leq 2d(d-1)^i \frac{\text{vol} T(S^{i+1}, \sigma)}{\mathcal{O}_n}. \quad (1.32)$$

In order to show this, put  $U := M \cap B(a, \sigma)$  and let  $U_+$  be the set of points of  $U$  where  $K_{M,i}$  is positive and  $U_-$  be the set of points of  $U$  where  $K_{M,i}$  is negative. Then  $|\mu_i|(U) = |\mu_i(U_+)| + |\mu_i(U_-)|$ .

Let  $g \in G$  be such that  $M$  intersects  $gS^{i+1}$  transversally. We apply the bound (1.31) to the hypersurface  $M \cap gS^{i+1}$  of the sphere  $gS^{i+1}$ , which yields  $|\mu_i^{\text{no}}|(M \cap gS^{i+1}) \leq d(d-1)^i$ . By monotonicity we obtain

$$|\mu_i^{\text{no}}(U_+ \cap gS^{i+1})| \leq |\mu_i^{\text{no}}|(U_+ \cap gS^{i+1}) \leq |\mu_i^{\text{no}}|(M \cap gS^{i+1}) \leq d(d-1)^i.$$

The kinematic formula (1.29) applied to  $U_+$  implies that

$$\begin{aligned} |\mu_i^{\text{no}}(U_+)| &\leq \mathbb{E}_{g \in G} (|\mu_i^{\text{no}}(U_+ \cap gS^{i+1})|) \\ &\leq d(d-1)^i \text{Prob}_{g \in G} \{U_+ \cap gS^{i+1} \neq \emptyset\} \\ &\leq d(d-1)^i \text{Prob}_{g \in G} \{B(a, \sigma) \cap gS^{i+1} \neq \emptyset\} \\ &= d(d-1)^i \frac{\text{vol} T(S^{i+1}, \sigma)}{\mathcal{O}_n}. \end{aligned}$$

The same upper bound holds for  $|\mu_i(U_-)|$  and hence the assertion (1.32) follows.

A similar reasoning shows  $|\mu_i^{\text{no}}|(\partial K \cap B(a, \sigma)) \leq \mathcal{O}_n^{-1} \text{vol} T(S^{i+1}, \sigma)$  for a convex body  $K$  in  $S^n$  with smooth boundary  $\partial K$ .

We outline now the proof of Theorem 1.2. By plugging in the estimate (1.32) into the upper bound on tube volumes (1.26), we obtain

$$\frac{\text{vol} T^\perp(M \cap B(a, \sigma), \epsilon)}{\text{vol} B(a, \sigma)} \leq 2d \sum_{i=0}^{n-1} d^i \frac{\text{vol} T(S^{i+1}, \sigma)}{\text{vol} B(a, \sigma)} \frac{\text{vol} T(S^{n-i-1}, \epsilon)}{\mathcal{O}_n}$$

and after some estimations one can arrive at the estimate

$$\frac{\text{vol} T^\perp(M \cap B(a, \sigma), \epsilon)}{\text{vol} B(a, \sigma)} \leq 4 \sum_{k=0}^{n-1} \binom{n}{k} \left(\frac{d\epsilon}{\sigma}\right)^k + \frac{2n\mathcal{O}_n}{\mathcal{O}_{n-1}} \left(\frac{d\epsilon}{\sigma}\right)^n.$$

From this, estimates of the volume of the set  $T(M, \epsilon) \cap B(a, \sigma)$  can be deduced by noting that the latter set is contained in the  $\epsilon$ -tube around  $M \cap B(\pm a, \sigma + \epsilon)$ . Another problem is that  $M$  is assumed to be smooth, but the real algebraic hypersurface  $M'$  in the statement of Theorem 1.2 may have singularities. Fortunately, this can be easily dealt with by

a perturbation argument. By some further estimations, one finally arrives at

$$\frac{\text{vol}(T(M', \epsilon) \cap B(a, \sigma))}{\text{vol} B(a, \sigma)} \leq 26 dn \frac{\epsilon}{\sigma}$$

for  $\epsilon \leq \sigma/(n(2d+1))$ , as claimed in Theorem 1.2. For details we refer to the original paper by Bürgisser et al. (2008). We note that the bound (1.24) on the volume of  $\epsilon$ -neighborhoods of  $\partial K$  follows in a similar way.

In order to deduce from the above the bound on the expectation stated in Theorem 1.2, we use the general observation that a tail bound of the form

$$\text{Prob}\{X \geq t\} \leq Kt^{-\alpha} \quad \text{for all } t \geq t_0 > 0$$

for a nonnegative absolutely continuous random variable  $X$  such that  $K, \alpha > 0$  implies

$$\mathbb{E}(\ln X) \leq \ln t_0 + \frac{1}{\alpha} (\ln K + 1).$$

We finally remark that the proof of Theorem (1.3), dealing with the situation over  $\mathbb{C}$ , is more direct and avoids curvatures. However, it is not possible to extend those arguments to the situation over  $\mathbb{R}$ .

**Acknowledgements** The surveys by Smale (1997) and Cucker (2002) were very helpful in writing this article. I thank Dennis Amelunxen for useful comments. This work was supported by DFG grant BU1371/2-1.

### Bibliography

- D. Amelunxen and P. Bürgisser (2008), ‘Uniform smoothed analysis of a condition number of linear programming’, arXiv:0803.0925.
- L. Amodi and J.-P. Dedieu, *Analyse numérique matricielle, Mathématiques Appliquées pour le Master/SMIAI*. Dunod. To appear.
- K. Ball (1993), ‘The reverse isoperimetric problem for Gaussian measure’, *Discrete Comput. Geom.* **10**(4), 411–420.
- E. Barbier (1860), ‘Note sur le problème de l’aiguille et le jeu du joint couvert’, *J. Math. Pures et Appl.* **5**(2), 273–286.
- A. Belloni, R. M. Freund and S. Vempala (2007), ‘An efficient re-scaled perceptron algorithm for conic systems’, *Proc. of 20th Conf. on Computational Learning Theory*, San Diego, 2007.
- 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.
- C. Beltrán and L. M. Pardo (2008), ‘On Smale’s 17th problem: a probabilistic positive solution’, *Found. Comput. Math.* **8**(1), 1–43.

- C. Beltrán and M. Shub, ‘Complexity of Bézout’s theorem VII: Distance estimates in the condition metric’, *Found. Comput. Math.* To appear.
- L. Blum (1990), ‘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.
- L. Blum, F. Cucker, M. Shub and S. Smale (1998), *Complexity and Real Computation*, Springer.
- 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.
- 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.
- T. Bonnesen and W. Fenchel (1974), *Theorie der konvexen Körper*, Springer-Verlag, Berlin, Berichtiger Reprint.
- 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.
- S. Boyd and L. Vandenberghe (2004), *Convex Optimization*, Cambridge University Press.
- P. Bürgisser (2007), ‘Average Euler characteristic of random real algebraic varieties’, *C. R. Math. Acad. Sci. Paris* **345**(9), 507–512.
- P. Bürgisser, F. Cucker, and M. Lotz (2007), ‘Coverage processes on spheres and condition numbers for linear programming’, arXiv:0712.2816.
- P. Bürgisser, F. Cucker and M. Lotz (2006), ‘General formulas for the smoothed analysis of condition numbers’, *C. R. Acad. Sci. Paris, Ser. I* **343**, 145–150.
- P. Bürgisser, F. Cucker and M. Lotz (2006), ‘Smoothed analysis of complex conic condition numbers’, *J. Math. Pures et Appl.* **86**, 293–309.
- P. Bürgisser, F. Cucker and M. Lotz (2008), ‘The probability that a slightly perturbed numerical analysis problem is difficult’, *Math. Comp.* **77**, 1559–1583.
- M. P. do Carmo (1992), *Riemannian Geometry*, Birkhäuser.
- S.-S. Chern (1966), ‘On the kinematic formula in integral geometry’, *J. Math. Mech.* **16**, 101–118.
- D. Cheung and F. Cucker (2001), ‘A new condition number for linear programming’, *Math. Program.* **91**(1, Ser. A), 163–174.
- D. Cheung and F. Cucker (2002), ‘Probabilistic analysis of condition numbers for linear programming’, *J. Optim. Theory Appl.* **114**(1), 55–67.
- 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.
- D. Cheung, F. Cucker and J. Peña (2008), ‘A condition number for multifold conic systems’, *SIAM J. Optim.* **19**(1), 261–280.
- F. Cucker (2002), ‘Real computations with fake numbers’, *J. Complexity* **18**(1), 104–134.
- F. Cucker and J. Peña (2002), ‘A primal-dual algorithm for solving polyhedral conic systems with a finite-precision machine’, *SIAM J. Optim.* **12**(2), 522–554.
- F. Cucker and M. Wschebor (2003), ‘On the expected condition number of linear programming problems’, *Numer. Math.* **94**(3), 419–478.
- J. Demmel (1987), ‘On condition numbers and the distance to the nearest ill-posed problem’, *Numer. Math.* **51**, 251–289.

- J. Demmel (1988), ‘The probability that a numerical analysis problem is difficult’, *Math. Comp.* **50**, 449–480.
- J. Dunagan, D. A. Spielman, and S.-H. Teng (2003), ‘Smoothed Analysis of Renegar’s Condition Number for Linear Programming’, arXiv:cs.DS/0302011v2.
- J. Dunagan and S. Vempala (2004), ‘A simple polynomial-time rescaling algorithm for solving linear programs’, *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 315–320, New York, ACM.
- C. Eckart and G. Young (1936), ‘The approximation of one matrix by another of lower rank’, *Psychometrika* **1**, 211–218.
- A. Edelman (1988), ‘Eigenvalues and condition numbers of random matrices’, *SIAM J. of Matrix Anal. and Applic.* **9**, 543–556.
- A. Edelman (1992), ‘On the distribution of a scaled condition number’, *Math. Comp.* **58**, 185–190.
- H. Federer (1959), ‘Curvature measures’, *Trans. Amer. Math. Soc.* **93**, 418–491.
- R. M. Freund and J. R. Vera (1999), ‘Some characterizations and properties of the “distance to ill-posedness” and the condition measure of a conic linear system’, *Math. Program.* **86**(2, Ser. A), 225–260.
- J.-L. Goffin (1980), ‘The relaxation method for solving systems of linear inequalities’, *Math. Oper. Res.* **5**(3), 388–414.
- A. Gray (1990), *Tubes*, Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA.
- M. Grötschel, L. Lovász and A. Schrijver (1988), *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics: Study and Research Texts*, Springer-Verlag, Berlin.
- M. R. Hestenes and E. Stiefel (1952), ‘Methods of conjugate gradients for solving linear systems’, *J. Research Nat. Bur. Standards* **49**, 409–436.
- N. J. Higham (1996), *Accuracy and stability of numerical algorithms*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- R. Howard (1993), ‘The kinematic formula in Riemannian homogeneous spaces’, *Mem. Amer. Math. Soc.* **106**.
- S. Janson (1986), ‘Random coverings in several dimensions’, *Acta Math.* **156**, 83–118.
- L. G. Khachyian (1979), ‘A polynomial algorithm in linear programming’, *Dokl. Akad. Nauk SSSR* **244**, 1093–1096.
- D. A. Klain and G.-C. Rota (1997), *Introduction to geometric probability*, Lezioni Lincee. [Lincei Lectures]. Cambridge University Press, Cambridge.
- E. Kostlan (1988), ‘Complexity theory of numerical linear algebra’, *J. Comput. Appl. Math.* **22**, 219–230.
- D. Mumford (1976), *Algebraic Geometry I: Complex Projective Varieties*, Springer.
- 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.
- A. Nijenhuis (1974), ‘On Chern’s kinematic formula in integral geometry’, *J. Differential Geometry* **9**, 475–482.
- 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.
- J. Renegar (1995a), ‘Incorporating condition measures into the complexity theory of linear programming’, *SIAM J. Optim.* **5**(3), 506–524.
- J. Renegar (1995b), ‘Linear programming, complexity theory and elementary functional analysis’, *Math. Programming* **70**(3, Ser. A), 279–351.
- F. Rosenblatt (1962), *Principles of neurodynamics. Perceptrons and the theory of brain mechanisms*, Spartan Books, Washington, D.C.
- 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.
- L. A. Santaló (1976), *Integral geometry and geometric probability*, Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam.
- M. Shub, ‘Complexity of Bézout’s theorem VI: Geodesics in the condition (number) metric’, *Found. Comput. Math.* To appear.
- M. Shub and S. Smale (1993a), ‘Complexity of Bézout’s theorem. I. Geometric aspects’, *J. Amer. Math. Soc.* **6**(2), 459–501.
- M. Shub and S. Smale (1993b), ‘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.
- M. Shub and S. Smale (1994), ‘Complexity of Bézout’s theorem V: polynomial time’, *Theoretical Computer Science* **133**, 141–164.
- M. Shub and S. Smale (1996), ‘Complexity of Bézout’s theorem IV: probability of success; extensions’, *SIAM J. of Numer. Anal.* **33**, 128–148.
- S. Smale (1981), ‘The fundamental theorem of algebra and complexity theory’, *Bull. Amer. Math. Soc.* **4**, 1–36.
- S. Smale (1983), ‘On the average number of steps of the simplex method of linear programming’, *Math. Programming* **27**(3), 241–262.
- S. Smale (1997), ‘Complexity theory and numerical analysis’, In A. Iserles, editor, *Acta Numerica*, pages 523–551. Cambridge University Press.
- S. Smale (2000), ‘Mathematical problems for the next century’, In *Mathematics: frontiers and perspectives*, pages 271–294, Amer. Math. Soc., Providence, RI.
- H. Solomon (1978), *Geometric probability*, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- 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.
- D. A. Spielman and S.-H. Teng (2002), ‘Smoothed analysis of algorithms’, *Proceedings of the International Congress of Mathematicians*, volume I, pages 597–606.
- D. A. Spielman and S.-H. Teng (2003), ‘Smoothed analysis of termination of linear programming algorithms’, *Math. Programm. Series B* **97**, 375–404.
- 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.
- D. A. Spielman and S.-H. Teng (2006), ‘Smoothed analysis of algorithms and heuristics’, *Foundations of Computational Mathematics, Santander 2005*, volume 331 of *London Mathematical Society, Lecture Note Series*, pages 274–342, Cambridge University Press.

- M. Spivak (1979), *A comprehensive introduction to differential geometry. Vol. III*, Publish or Perish Inc.
- J. Steiner (1840), ‘Über parallele Flächen’, *Monatsbericht der Akademie der Wissenschaften zu Berlin*, pages 114–118. Also Werke vol. 2 (1882), pp. 171–176.
- 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.
- J. A. Thorpe (1994), *Elementary topics in differential geometry*, Undergraduate Texts in Mathematics. Springer-Verlag, New York.
- A. M. Turing (1948), ‘Rounding-off errors in matrix processes’, *Quart. J. Mech. Appl. Math.* **1**, 287–308.
- S. A. Vavasis and Y. Ye (1995), ‘Condition numbers for polyhedra with real number data’, *Oper. Res. Lett.* **17**(5), 209–214.
- J. von Neumann and H. H. Goldstine (1947), ‘Numerical inverting of matrices of high order’, *Bull. Amer. Math. Soc.* **53**, 1021–1099.
- J. G. Wendel (1962), ‘A problem in geometric probability’, *Math. Scand.* **11**, 109–111.
- H. Weyl (1939), ‘On the Volume of Tubes’, *Amer. J. Math.* **61**(2), 461–472.
- J. H. Wilkinson (1963), *Rounding errors in algebraic processes*, Prentice-Hall Inc., Englewood Cliffs, N.J.
- J. H. Wilkinson (1972), ‘Note on matrices with a very ill-conditioned eigenproblem’, *Numer. Math.* **19**, 176–178.
- H. Woźniakowski (1977), ‘Numerical stability for solving nonlinear equations’, *Numer. Math.* **27**(4), 373–390.
- M. Wschebor (2004), ‘Smoothed analysis of  $\kappa(a)$ ’, *J. of Complexity* **20**, 97–107.