MATHEMATICS OF COMPUTATION
Volume 77, Number 263, July 2008, Pages 1559–1583
S 0025-5718(08)02060-7
Article electronically published on January 30, 2008

THE PROBABILITY THAT A SLIGHTLY PERTURBED NUMERICAL ANALYSIS PROBLEM IS DIFFICULT

PETER BÜRGISSER, FELIPE CUCKER, AND MARTIN LOTZ

Abstract. We prove a general theorem providing smoothed analysis estimates for conic condition numbers of problems of numerical analysis. Our probability estimates depend only on geometric invariants of the corresponding sets of ill-posed inputs. Several applications to linear and polynomial equation solving show that the estimates obtained in this way are easy to derive and quite accurate. The main theorem is based on a volume estimate of ε-tubular neighborhoods around a real algebraic subvariety of a sphere, intersected with a spherical disk of radius σ. Besides ε and σ, this bound depends only on the dimension of the sphere and on the degree of the defining equations.

1. Introduction

In a seminal article [13] J. Demmel suggested that "to investigate the probability that a numerical analysis problem is difficult, we need to do three things:

(1) Choose a measure of difficulty,
(2) Choose a probability distribution on the set of problems,
(3) Compute the distribution of the measure of difficulty induced by the distribution on the set of problems."

Then, for the measure of difficulty, Demmel proposed the *condition number*. This is a positive number which, roughly speaking, measures the sensitivity of the output to small perturbations of the input. It depends only on the input data and the function being computed. Condition numbers occur in endless instances of round-off analysis. They also appear as a parameter in complexity bounds for a variety of iterative algorithms.

The main results in [13] carry out an analysis as sketched in (1)–(3) above for the condition number $\mathscr{C}$ of several problems. This analysis exhibits bounds on the tail of the distribution of $\mathscr{C}(a)$, showing that it is unlikely that $\mathscr{C}(a)$ will be large. From these bounds one can obtain, using standard methods in probability theory, bounds on the expected value of $\ln(\mathscr{C}(a))$, estimating the average loss of precision and average running time for algorithms solving the considered problem. Demmel's results thus yield prime instances of *average-case* analysis of algorithms in numerical analysis.

Received by the editor September 28, 2006 and, in revised form, April 23, 2007.
2000 *Mathematics Subject Classification.* Primary 65Y20; Secondary 65F35, 65H20.
Key words and phrases. Condition numbers, smooth analysis, tubular neighborhoods of algebraic surfaces.
Part of these results were announced in C.R. Acad. Sci. Paris, Ser. I 343 (2006) 145–150.
The first author was partially supported by DFG grant BU 1371.
The second and third authors were partially supported by CityU SRG grant 7001860.

While average-case analysis undoubtedly has advantages over worst-case analysis, it is not itself without shortcomings, the most noticeable being the arbitrariness of the selected probability distribution on the set of inputs. To find a way out of these shortcomings, D. Spielman and S.-H. Teng [31, §3] proposed a new form of analysis that arguably blends the best of both worst-case and average-case. The idea is to replace showing that

> "it is unlikely that $\mathscr{C}(a)$ will be large"

by showing that

> "for all a and all slight random perturbations Δa, it is unlikely that $\mathscr{C}(a + \Delta a)$ will be large."

A survey of this approach, called *smoothed analysis*, can be found in [31, 34]. If $\mathscr{D}(c, \sigma)$ denotes a probability distribution centered at $c \in \mathbb{R}^{p+1}$ with covariance matrix $\sigma^2 \mathrm{id}_{p+1}$, and $\mathbf{E}$ denotes mathematical expectation, we may summarize the objects of study of worst-case, average-case, and smoothed analyses, for a function $\psi : \mathbb{R}^{p+1} \to \mathbb{R}$, in the following table.

worst-case analysis	average-case analysis	smoothed analysis
$\sup_{a \in \mathbb{R}^{p+1}} \psi(a)$	$\mathbf{E}_{a \in \mathscr{D}(0,\sigma)} \psi(a)$	$\sup_{a \in \mathbb{R}^{p+1}} \mathbf{E}_{z \in \mathscr{D}(a,\sigma)} \psi(z)$

A remarkable feature of [13] is that the 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 $\mathscr{C}(a)$ in terms of geometric invariants (degree and dimension) of the set of ill-posed inputs of the problem for which $\mathscr{C}$ is a condition number.

A first goal of this paper is to extend the results of [13] from average-case to smoothed analysis. We will, however, also prove average-case bounds. Demmel's 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 A. Ocneanu on the volumes of tubes around real algebraic varieties. A second goal of this paper is to prove a result akin to Ocneanu's (Theorem 1.2). We are not the first to do so. In [45], R. Wongkew gave a bound for the volume of tubes around real algebraic varieties. A number of constants in his bounds, however, are not explicit and the only thing we know about them is that they are independent of the variety.

1.1. **Statement of the main result.** We assume our data space is $\mathbb{R}^{p+1}$, endowed with a scalar product $\langle\ ,\ \rangle$. By a semi-algebraic cone $\Sigma \subseteq \mathbb{R}^{p+1}$ we understand a semi-algebraic set $\Sigma \neq \{0\}$ that is closed under multiplication with positive scalars. We say that $\mathscr{C}$ is a *conic condition number* if there exists a semi-algebraic cone $\Sigma \subseteq \mathbb{R}^{p+1}$, the set of *ill-posed inputs*, such that, for all data $a \in \mathbb{R}^{p+1} \setminus \{0\}$,

$$\mathscr{C}(a) = \frac{\|a\|}{\mathsf{dist}(a, \Sigma)},$$

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

The best known condition number is that used for matrix inversion and linear equation solving. For a square matrix A it takes the form $\kappa(A) = \|A\|\|A^{-1}\|$

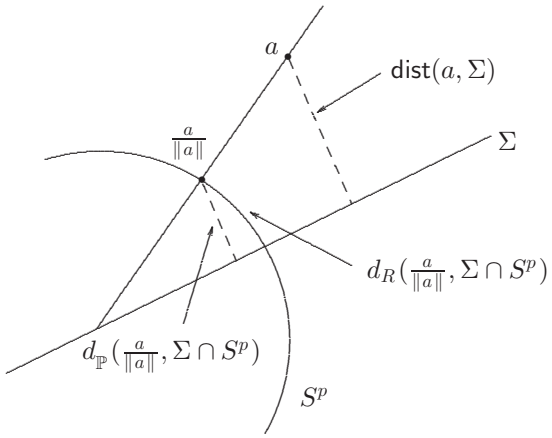

FIGURE 1. Three distances

and was independently introduced by H. Goldstine and J. von Neumann [41] and A. Turing [40]. Strictly speaking, $\kappa(A)$ is not conic since the operator norm $\|\ \|$ is not induced by a scalar product. Replacing this norm by the Frobenius norm $\|\ \|_F$ yields the (commonly considered) version $\kappa_F(A) := \|A\|_F\|A^{-1}\|$ of $\kappa(A)$. The Condition Number Theorem of C. Eckart and G. Young [15] then states that $\kappa_F(A)$ is conic, with Σ the set of singular matrices. Other examples can be found in [10], where a certain property (related with the so-called level-2 condition numbers) is proved for conic condition numbers. Furthermore, it is argued by Demmel in [12] that the condition numbers for many problems can be bounded by conic ones.

Note that, since Σ is a cone, for all $\lambda > 0$, $\mathscr{C}(a) = \mathscr{C}(\lambda a)$. Hence, we may restrict to data a lying in the sphere $S^p := \{x \in \mathbb{R}^{p+1} \mid \|x\| = 1\}$. If we set $\Sigma_s := \Sigma \cup (-\Sigma)$, then the conic condition number $\mathscr{C}$ can be estimated as

$$\mathscr{C}(a) \leq \frac{\|a\|}{\mathsf{dist}(a, \Sigma_s)} = \frac{1}{d_{\mathbb{P}}(a, \Sigma_s \cap S^p)}, \tag{1}$$

where $d_{\mathbb{P}}$ denotes the projective distance in S^p, which is defined as $d_{\mathbb{P}}(x, y) = \sin d_R(x, y)$ with d_R being the Riemannian (or angular) distance in S^p (cf. Figure 1).

Let $B_{\mathbb{P}}(a, \sigma)$ denote the open ball of radius σ, with respect to $d_{\mathbb{P}}$, around a in S^p. Note that

$$B_{\mathbb{P}}(a, \sigma) = B_R(a, \arcsin \sigma) \cup B_R(-a, \arcsin \sigma),$$

where B_R denotes a ball with respect to the angular distance. We will endow $B_{\mathbb{P}}(a, \sigma)$ with the uniform probability measure. Moreover, let

$$\mathcal{O}_p := \mathsf{vol}_p(S^p) = \frac{2\pi^{\frac{p+1}{2}}}{\Gamma\left(\frac{p+1}{2}\right)}$$

denote the p-dimensional volume of the sphere S^p. Our main result is the following.

Theorem 1.1. *Let $\mathscr{C}$ be a conic condition number with a set of ill-posed inputs Σ. Assume that $\Sigma \cap S^p \subseteq W$ where $W \subseteq S^p$ is the zero set in S^p of homogeneous*

polynomials of degree at most $d \geq 1$ and $W \neq S^p$. Then, for all $a \in S^p$, all $\sigma \in (0,1]$, and all $t \geq 1$,

$$\operatorname*{Prob}_{z \in B_{\mathbb{P}}(a,\sigma)}\{\mathscr{C}(z) \geq t\} \leq 2\sum_{k=1}^{p-1} \frac{p}{k}(2d)^k \left(1+\frac{1}{t\sigma}\right)^{p-k} \left(\frac{1}{t\sigma}\right)^k + \frac{p\mathcal{O}_p}{\mathcal{O}_{p-1}}(2d)^p \left(\frac{1}{t\sigma}\right)^p$$

and, for all $\sigma \in (0,1]$,

$$\sup_{a \in S^p} \mathop{\mathbf{E}}_{z \in B_{\mathbb{P}}(a,\sigma)} (\ln \mathscr{C}(z)) \leq 2\ln p + 2\ln d + 2\ln\frac{1}{\sigma} + 4.7.$$

In particular, for all $t \geq 1$ (take $\sigma = 1$),

$$\operatorname*{Prob}_{z \in S^p}\{\mathscr{C}(z) \geq t\} \leq 2\sum_{k=1}^{p-1} \frac{p}{k}(2d)^k \left(1+\frac{1}{t}\right)^{p-k} \frac{1}{t^k} + \frac{p\mathcal{O}_p}{\mathcal{O}_{p-1}}(2d)^p \frac{1}{t^p}$$

and

$$\mathop{\mathbf{E}}_{z \in S^p} (\ln \mathscr{C}(z)) \leq 2\ln p + 2\ln d + 4.7.$$

The main idea towards the proof of Theorem 1.1 is to reformulate the probability distribution of a conic condition number as a geometric problem in a sphere. We next see how. For a measurable subset V of S^p we denote by $\mathsf{vol}_p(V)$ the p-dimensional volume of V. If $-V = V$ we define the *ε-neighborhood* around V in S^p by

$$T_{\mathbb{P}}(V,\varepsilon) := \{x \in S^p \mid d_{\mathbb{P}}(x,V) < \varepsilon\}.$$

With this notation, using $\Sigma_s \cap S^p \subseteq W$, we obtain from (1) for $a \in S^p$ and $\sigma \in (0,1]$

$$\begin{aligned}\operatorname*{Prob}_{z \in B_{\mathbb{P}}(a,\sigma)} \left\{\mathscr{C}(z) \geq \frac{1}{\varepsilon}\right\} &\leq \operatorname*{Prob}_{z \in B_{\mathbb{P}}(a,\sigma)} \{d_{\mathbb{P}}(z, \Sigma_s \cap S^p) \leq \varepsilon\} \leq \operatorname*{Prob}_{z \in B_{\mathbb{P}}(a,\sigma)} \{d_{\mathbb{P}}(z,W) \leq \varepsilon\} \\ &= \frac{\mathsf{vol}_p(T_{\mathbb{P}}(W,\varepsilon) \cap B_{\mathbb{P}}(a,\sigma))}{\mathsf{vol}_p(B_{\mathbb{P}}(a,\sigma))}.\end{aligned}$$

The tail bounds in Theorem 1.1 will thus follow from the following purely geometric statement.

Theorem 1.2. *Let $W \subseteq S^p$ be a real algebraic variety defined by homogeneous polynomials of degree at most $d \geq 1$ such that $W \neq S^p$. Then we have for $a \in S^p$ and $0 < \varepsilon, \sigma \leq 1$*

$$\frac{\mathsf{vol}_p\left(T_{\mathbb{P}}(W,\varepsilon) \cap B_{\mathbb{P}}(a,\sigma)\right)}{\mathsf{vol}_p B_{\mathbb{P}}(a,\sigma)} \leq 2\sum_{k=1}^{p-1} \frac{p}{k}(2d)^k \left(1+\frac{\varepsilon}{\sigma}\right)^{p-k} \left(\frac{\varepsilon}{\sigma}\right)^k + \frac{p\mathcal{O}_p}{\mathcal{O}_{p-1}}(2d)^p \left(\frac{\varepsilon}{\sigma}\right)^p.$$

In particular (take $\sigma = 1$),

$$\frac{\mathsf{vol}_p T_{\mathbb{P}}(W,\varepsilon)}{\mathcal{O}_p} \leq 2\sum_{k=1}^{p-1} \frac{p}{k}(2d)^k (1+\varepsilon)^{p-k} \varepsilon^k + \frac{p\mathcal{O}_p}{\mathcal{O}_{p-1}}(2d)^p \varepsilon^p.$$

Here is a brief outline of the proof of Theorem 1.2: The first step is an upper bound on the volume of an ε-neighborhood of a smooth hypersurface in terms of integrals of absolute curvature (Proposition 3.1). This is a variation of H. Weyl's [42] exact formula for the volume of tubes, a formula which, however, only holds for sufficiently small ε. Then (Proposition 3.2) we derive a degree bound on these integrals of absolute curvature based on the kinematic formula of integral geometry

and Bézout's theorem. Finally we get rid of the smoothness assumption by some perturbation argument.

We will devote Section 4 to derive applications of Theorem 1.1 to several condition numbers occuring in the literature, namely, those for linear equation solving, eigenvalue computation, polynomial system zero finding, and zero counting.

Remark 1.3. Theorem 1.2 could be stated for real projective space $\mathbb{P}^p$ with the same bounds. While such a statement is the most natural over the complex numbers (cf. [1] and [7, Theorem 1.3]) it does not follow the tradition over the reals (cf. [13, 42, 45]) and it is not a natural ambient space for real conic condition numbers. Note that Σ is not necessarily symmetrical around the coordinate origin and that the use of a symmetric W (an algebraic cone containing the semi-algebraic cone Σ) is just an artifice of our proofs.

Theorem 1.1 was announced in [6] (in the projective setting and with slightly worse constants).

1.2. **Relation to previous work.** Consider a symmetric function $\psi : \mathbb{R}^{p+1} \to \mathbb{R}$ (i.e., satisfying $\psi(x) = \psi(-x)$). In most instances of smoothed analysis (e.g., [11, 14, 31, 32, 33, 46]) one studies the behaviour of

$$\sup_{a\in\mathbb{R}^{p+1}} \mathop{\mathbf{E}}_{z\in N^{p+1}(a,\sigma^2)} \psi(z) \tag{2}$$

(possibly for sufficiently small σ) where $N^{p+1}(a,\sigma^2)$ denotes the $p+1$-dimensional Gaussian distribution over $\mathbb{R}$ with mean a and variance σ^2. It has been argued that smoothed analysis interpolates between worst and average cases since it amounts to the first for $\sigma = 0$ and it approaches the second for large σ.

When $\psi(\lambda x) = \psi(x)$ for all $\lambda > 0$ —e.g., a conic condition number— it makes sense to restrict ψ to the sphere S^p. In this case, it also makes sense to replace the distribution $N^{p+1}(a,\sigma^2)$ by the uniform distribution supported on the disk $B_{\mathbb{P}}(a,\sigma) \subseteq S^p$ and to consider, instead of (2), the following quantity:

$$\sup_{a\in S^p} \mathop{\mathbf{E}}_{z\in B_{\mathbb{P}}(a,\sigma)} \psi(z). \tag{3}$$

Note that in this case, the interpolation mentioned above is transparent. When $\sigma = 0$ the expected value amounts to $\psi(a)$ and we obtain worst-case analysis, while if $\sigma = 1$ the expected value is independent of a and we obtain average-case analysis.

It is this version of smoothed analysis, introduced in [7], which we deal with in this paper. Note that while, technically, this "uniform smoothed analysis" differs from the Gaussian one considered so far, both share the viewpoint described above.

We have already mentioned the references [11, 14, 31, 32, 33, 46] as instances on previous work in smoothed analysis. In all these cases, an *ad hoc* argument is used to obtain the desired bounds. This is in contrast with the goal of this paper which is to provide general estimates which can be applied to a large class of condition numbers. We believe the applications in Section 4 give substance to this goal.

We finish this section with a brief overview of previous work on the relations between complexity, conditioning and probabilistic analysis. In [3], L. Blum suggested a complexity theory for numerical algorithms parameterized by a condition number $\mathscr{C}(a)$ for the input data (in addition to input size). S. Smale [30, §1] extended this suggestion by proposing to obtain estimates on the probability distribution of $\mathscr{C}(a)$. Combining both ideas, he argued, one can give probabilistic bounds on the complexity of numerical algorithms.

The idea of reformulating probability distributions as quotients of volumes in projective spaces (or spheres) to estimate condition measures goes back at least to Smale [29] and Renegar [21]. In particular, J. Renegar [21] uses this idea to show bounds on the probability distribution of a certain random variable in the average-case analysis of the complexity of Newton's method. Central to his argument is the fact that this random variable can be bounded by a conic condition number. The set of ill-posed inputs in [21] is a hypersurface. An extension of these results to the case of codimension greater than one was done by Demmel [13] where, in addition, an average-case analysis of several conic condition numbers is performed. Most of these results are for problems over the complex numbers. An extension in another direction, namely, to possibly singular ambient spaces, was done by C. Beltrán and L.M. Pardo [1]. Another extension of Demmel's result, now to smoothed analysis for complex problems, was achieved in [7]. At this point we want to emphasize that the mentioned work in [1, 7, 21], in contrast with the contents of this paper, is over the complex numbers (and the work over the reals in [13] rely on unproved results). That is, problem data are assumed to be vectors in $\mathbb{C}^p$ (an assumption that excludes a large number of problems in numerical analysis). We will return to this issue in Remark 4.2(i). Furthermore, these works make use of properties of complex varieties which do not hold in the real case.

The remainder of the paper is organized as follows. In Section 2 we provide the preliminary notations and results needed to prove Theorem 1.2. These are mostly taken from differential and integral geometry. In Section 3 we prove Theorem 1.2 and Theorem 1.1. Section 4 is devoted to several applications of our main result. Finally, the precise value of some constants —whose existence is well-documented in the literature but whose magnitude is not—is derived in the Appendix.

2. Preliminaries

2.1. **Distances, volumes, and tubes on the sphere.** The p-dimensional sphere S^p carries the structure of a compact Riemannian manifold. Correspondingly, there is a Riemannian distance $d_R(x, y) \in [0, \pi]$ between two points $x, y \in S^p$, which is just the angle between these points. We denote by $B_R(a, \alpha) = \{x \in S^p \mid d_R(a, x) < \alpha\}$ the open ball of radius α centered at $a \in S^p$.

It will be more natural for us to work with the related notion of *projective distance*, which is defined as $d_{\mathbb{P}}(x, y) := \sin d_R(x, y) \in [0, 1]$ (cf. Figure 1). We note that $d_{\mathbb{P}}$ satisfies all the axioms of a metric, except that $d_{\mathbb{P}}(x, y) = 0$ iff $x \in \{-y, y\}$. In fact, $d_{\mathbb{P}}$ induces a metric on the real projective space $\mathbb{P}^p$ (obtained from S^p by identifying antipodal points). However, we prefer to work on the sphere, which seems more intuitive to us.

Let V be a subset of S^p. For $0 < \varepsilon \leq 1$ we define the *ε-neighborhood* around V by

$$T_{\mathbb{P}}(V, \varepsilon) := \{x \in S^p \mid d_{\mathbb{P}}(x, V) < \varepsilon\},$$

where $d_{\mathbb{P}}(x, V) := \inf_{y \in V} d_{\mathbb{P}}(x, y)$. This equals the α-neighborhood of $V \cup -V$ defined with respect to d_R, where $\alpha = \arcsin \varepsilon$.

For a measurable subset $A \subseteq S^p$ we write $\mathsf{vol}_p A = \int_A dS^p$ for the *p-dimensional volume*, where dS^p denotes the volume form induced by the Riemannian metric. In order to compute volumes of balls and tubes around subspheres, the following

functions $J_{p,k}(\alpha)$ are relevant:

$$J_{p,k}(\alpha) := \int_0^\alpha (\sin\rho)^{k-1}\,(\cos\rho)^{p-k}\,d\rho \quad (1 \le k \le p).$$

Lemma 2.1. *For $1 \le k \le p$, $0 < \alpha \le \pi/2$, and $\varepsilon = \sin\alpha$, we have*

$$\mathsf{vol}_p T_{\mathbb{P}}(S^{p-k}, \varepsilon) = \mathcal{O}_{p-k}\mathcal{O}_{k-1}J_{p,k}(\alpha), \quad \mathsf{vol}_p B_R(a,\alpha) = \mathcal{O}_{p-1}J_{p,p}(\alpha).$$

Proof. This follows from [42] or by straightforward calculation. □

The quantity $J_{p,k}(\alpha)$ can be easily bounded. Recall that $\mathcal{O}_p$ denotes the p-dimensional volume of S^p.

Lemma 2.2. *The following estimates hold ($1 \le k \le p$, $0 < \alpha \le \pi/2$, $\varepsilon = \sin\alpha$):*

$$J_{p,k}(\alpha) \le \frac{\varepsilon^k}{k} \quad \text{if } k < p, \quad \frac{\varepsilon^p}{p} \le J_{p,p}(\alpha) \le \frac{\mathcal{O}_p}{2\mathcal{O}_{p-1}}\,\varepsilon^p$$

with equality when $\alpha = \pi/2$ in the last upper bound.

Proof. To settle the first inequality note that for $k < p$

$$J_{p,k}(\alpha) \le \int_0^\alpha (\sin\rho)^{k-1}\,(\cos\rho)\,d\rho = \int_0^\varepsilon u^{k-1}\,du = \frac{\varepsilon^k}{k}.$$

Similarly,

$$J_{p,p}(\alpha) = \int_0^\alpha (\sin\rho)^{p-1}\,d\rho \ge \int_0^\alpha (\sin\rho)^{p-1}\cos\rho\,d\rho = \frac{\varepsilon^p}{p}.$$

It is easy to check that $\alpha \to J_{p,p}(\alpha)(\sin\alpha)^{-p}$ is monotonically increasing on $[0, \pi/2]$ by computing the derivative of this function. Hence, $J_p(\alpha)(\sin\alpha)^{-p} \le J_{p,p}(\pi/2)$. From Lemma 2.1 we get $\frac{1}{2}\mathcal{O}_p = \mathsf{vol}_p B_R(a, \pi/2) = \mathcal{O}_{p-1}J_{p,p}(\pi/2)$ from which it follows that $J_{p,p}(\pi/2) = \frac{\mathcal{O}_p}{2\mathcal{O}_{p-1}}$. □

In this paper, the notions of manifold and differentiability always refer to C^∞-differentiability. For a submanifold M of S^p and $0 < \alpha \le \pi/2$ we define the *α-tube* $T_R^\perp(M,\alpha)$ around M by (compare with [17, p. 34])

$$T_R^\perp(M,\alpha) := \{x \in S^p \mid \text{there is a great circle segment in } S^p \text{ of length} < \alpha \text{ from } x \text{ to } M \text{ that intersects } M \text{ orthogonally}\}.$$

Here we used the Riemannian distance. Sometimes, when thinking in terms of the projective distance and $M = -M$, it will be convient to use the notation of *ε-tube* $T_{\mathbb{P}}^\perp(M,\varepsilon) := T_R^\perp(M, \arcsin\varepsilon)$ defined for $0 < \varepsilon \le 1$. Clearly, $T_{\mathbb{P}}^\perp(M,\varepsilon)$ can be characterized in a way similar to $T_R^\perp(M,\alpha)$. We note that $T_{\mathbb{P}}^\perp(M,\varepsilon) \subseteq T_{\mathbb{P}}(M,\varepsilon)$ and the inclusion is in general strict. It can be shown, however, that if M is a compact submanifold, then equality holds.

We also need the notion of the m-dimensional volume (Hausdorff measure) of subsets T of S^p. For simplicity, we restrict ourselves to semialgebraic sets; cf. [5]. Let T be an m-dimensional semialgebraic subset of S^p. The Zariski closure W of T in S^p is a real algebraic variety of dimension m, and its regular locus $\mathrm{Reg}(W)$ is an m-dimensional submanifold of S^p. We define the m-dimensional volume of T by $\mathsf{vol}_m T := \mathsf{vol}_m(T \cap \mathrm{Reg}(W))$. This makes sense since $T \setminus \mathrm{Reg}(W)$ has dimension strictly less than m. For $k < m$ we set $\mathsf{vol}_k T := 0$.

2.2. **A useful transformation formula.** We will repeatedly use the following special case of the coarea formula. A proof can be found in [38, III§2, Folgerung 1] or [18, Appendix].

Here and in what follows, $\#A$ denotes the cardinality of a set A (being ∞ if A is infinite).

Proposition 2.3. *Let M, N be Riemannian manifolds of the same dimension and $\varphi\colon M \to N$ be differentiable. Suppose that $\int_M |\det D\varphi|\, dM$ is finite. Then the fiber $\varphi^{-1}(y)$ is finite for almost all $y \in N$ and we have*

$$\int_M |\det D\varphi|\, dM = \int_{y\in N} \#\varphi^{-1}(y)\, dN(y).$$

2.3. **Some differential geometry of hypersurfaces on spheres.** For the following material from differential geometry we refer to [36, 39].

In the following let M be a compact oriented smooth hypersurface of S^p interpreted as a Riemannian submanifold. The orientation corresponds to the choice of a unit normal vector field $\nu\colon M \to \mathbb{R}^{p+1}$ on M. The *Weingarten map* $L_M(x)\colon T_xM \to T_xM$ of M at x is the linear map defined by $L_M(x) := -D\nu(x)$ (it is easy to verify that this is a well-defined map). The *second fundamental form* of M at $x \in M$ is the corresponding bilinear map $\mathrm{II}_M(x)\colon T_xM \times T_xM \to \mathbb{R}$, defined by $\mathrm{II}_M(x)(Y, Z) = \langle L_M(x)Y, Z\rangle$ for all $Y, Z \in T_xM$. We are going to describe these notions in terms of local coordinates of M. Thus let $v = (v_1, \ldots, v_{p-1}) \mapsto x(v_1, \ldots, v_{p-1}) \in M \subset \mathbb{R}^{p+1}$ be a local parametrization of M. Then

$$\mathrm{II}_M(x)\left(\frac{\partial}{\partial v_i}, \frac{\partial}{\partial v_j}\right) = -\left\langle \frac{\partial \nu}{\partial v_i}, \frac{\partial x}{\partial v_j}\right\rangle = \left\langle \nu, \frac{\partial^2 x}{\partial v_i \partial v_j}\right\rangle,$$

where the last equality follows from deriving $\langle \partial x/\partial v_j, \nu\rangle = 0$. In particular, $\mathrm{II}_M(x)$ and $L_M(x)$ are symmetric. The eigenvalues $\kappa_1(x), \ldots, \kappa_{p-1}(x)$ of $L_M(x)$ are called *principal curvatures* at x of the hypersurface M.

Example 2.4. Consider the case of $M = S^{p-1}$, the subsphere of S^p given by the equation $x_p = 0$. Then it easy to see that $\mathrm{II}_M(x) = 0$ for all $x \in M$. Hence all the principal curvatures of M are zero. This example makes clear that the principal curvatures are relative to the ambient space S^p. (Of course, S^{p-1} is curved; however, its "curvature relative to the ambient sphere" is zero.)

For $1 \le i < p$ we define the *ith curvature* $K_{M,i}(x)$ of M at x as the ith elementary symmetric polynomial in $\kappa_1(x), \ldots, \kappa_{p-1}(x)$, and put $K_{M,0}(x) := 1$. In particular, $K_{M,p-1}(x) = \det L_M(x)$. Note that the ith curvatures are essentially the coefficients of the characteristic polynomial of the Weingarten map:

$$\det(\mathrm{id}_{p-1} + tL_M(x)) = \sum_{i=0}^{p-1} t^i\, K_{M,i}(x). \tag{4}$$

Definition 2.5. Let M be a compact oriented smooth hypersurface M of S^p and U be an open subset of M. The *integral $\mu_i(U)$ of the ith curvature* and the *integral $|\mu_i|(U)$ of the ith absolute curvature* over U, with respect to the ambient space M, are defined as $(0 \le i \le p-1)$

$$\mu_i(U) := \int_U K_{M,i}\, dM \qquad \text{and} \qquad |\mu_i|(U) := \int_U |K_{M,i}|\, dM.$$

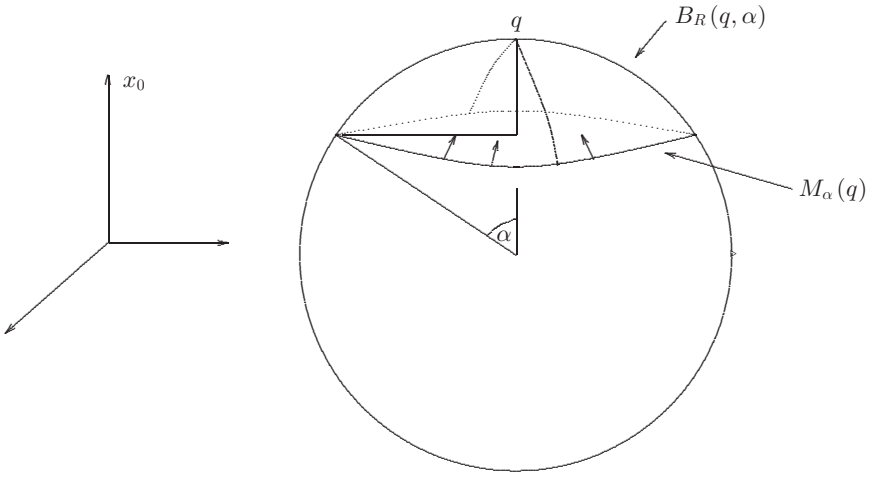

FIGURE 2. The manifold M_α

A few remarks: $|\mu_i|$ is monotone in the sense that $|\mu_i|(U_1) \le |\mu_i|(U_2)$ for $U_1 \subseteq U_2$. Also, note that $|\mu_i(U)| \le |\mu_i|(U)$ and $\mu_i(\emptyset) = |\mu_i|(\emptyset) = 0$. Moreover, $\mu_0(U) = |\mu_0(U)| = \mathsf{vol}_{p-1}(U)$. By Example 2.4, $|\mu_i|(S^{p-1}) = 0$ for $i > 0$.

Example 2.6. Consider the boundary M_α of the ball $B_R(q,\alpha)$ in S^p of radius $0 < \alpha \le \pi/2$ centered at q. Clearly, M_α is a $(p-1)$-dimensional sphere of radius $\sin\alpha$ that is described by the equations

$$x_0 = \cos\alpha,\ x_1^2 + \cdots + x_p^2 = \sin^2\alpha$$

if $q = (1, 0, \ldots, 0)$. We orient M_α by the unit normal vector field on S^p pointing towards q. It is straightforward to see that the second fundamental form of M_α satisfies $\mathrm{II}_{M_\alpha}(x) = (\cot\alpha)\,\mathrm{id}_{p-1}$ for all $x \in M_\alpha$. Hence all the principal curvatures of M_α at x are equal to $\cot\alpha$. Therefore the ith curvature of M_α satisfies $K_{M_\alpha,i}(x) = \frac{p-1}{i}(\cot\alpha)^i$, a quantity independent of $x \in M_\alpha$. For the integral of the ith curvature we obtain

$$\mu_i(M_\alpha) = K_{M_\alpha,i}\,\mathsf{vol}_{p-1}M_\alpha = \frac{p-1}{i}\mathcal{O}_{p-1}(\sin\alpha)^{p-i-1}(\cos\alpha)^i, \tag{5}$$

using that $\mathsf{vol}_{p-1}M_\alpha = \mathcal{O}_{p-1}(\sin\alpha)^{p-1}$. In particular, note that $\mu_{p-1}(M_\alpha) = \mathcal{O}_{p-1}(\cos\alpha)^{p-1}$. Finally note that $\mu_i(U) = |\mu_i|(U)$ for all open subsets U of M_α, since all the principal curvatures are nonnegative.

2.4. **A kinematic formula from integral geometry for spheres.** Here we recall a basic formula of integral geometry. The orthogonal group $G = O(p+1)$ operates on S^p in the natural way. We will denote by dG the volume element on the compact Lie group G normalized such that the volume of G equals one. We will interpret S^i as a submanifold of S^p for $i \le p$, e.g., given by the equations $x_{i+1} = \cdots = x_p = 0$.

Let M be a compact oriented smooth hypersurface of S^p. It follows by standard methods from Sard's lemma [35] that gM intersects S^{i+1} transversally for almost all $g \in G$. Hence, for almost all $g \in G$, the intersection $gM \cap S^{i+1}$ is either empty or is a smooth hypersurface of S^{i+1}. Moreover, this intersection inherits

an orientation from M in a natural way as follows: let ν be the distinguished unit normal vector field of M. Then we require that the distinguished unit normal vector of the hypersurface $gM \cap S^{i+1}$ in S^{i+1} at x lies in the positive halfspace of T_xM determined by ν.

Therefore, for almost all $g \in G$, the integral of the ith curvature $\mu_i(gM \cap S^{i+1})$ of $gM \cap S^{i+1}$, considered as a submanifold of S^{i+1}, is well defined, and this is also the case for $\mu_i(gU \cap S^{i+1})$ when U denotes an open subset of M.

We will need the following special case of the principal kinematic formula of integral geometry for spheres. A proof can be found in [18]. For Euclidean space a corresponding result was stated by Chern [9].

Theorem 2.7. *Let U be an open subset of a compact oriented smooth hypersurface M of S^p and $0 \le i < p-1$. Then we have*

$$\mu_i(U) = \mathcal{C}(p,i) \int_{g\in G} \mu_i(gU \cap S^{i+1})\, dG(g),$$

where $\mathcal{C}(p,i) = (p-i-1)\frac{p-1}{i} \frac{\mathcal{O}_{p-1}\mathcal{O}_p}{\mathcal{O}_i\mathcal{O}_{i+1}\mathcal{O}_{p-i-2}}$. □

While the existence of the constants $\mathcal{C}(p,i)$ follows from [9, 18], it is quite cumbersome to extract explicit formulas for $\mathcal{C}(p,i)$ from these sources. This is partly due to the fact that the quantities μ_i (and even $\mathcal{O}_i$) have slightly different meanings in the literature. For the convenience of the reader, we have therefore included a short derivation of these constants in the Appendix.

For future reference, we state the case $i = 0$ of Theorem 2.7 in a slightly more general form.

Corollary 2.8. *For any semialgebraic subset T of S^p such that* $\dim T \le p-1$ *we have*

$$\mathsf{vol}_{p-1} T = \frac{\mathcal{O}_{p-1}}{2} \int_{g\in G} \#(T \cap gS^1)\, dG(g).$$

Proof. Using the comments given at the end of Section 2.1, it is easy to reduce to the case where T is an open subset of a hypersurface of S^p. Now apply Theorem 2.7 for $i = 0$, taking into account that neither the compactness nor the orientability assumption are necessary in that case; cf. [22, §15.2]. □

3. On the volume of tubes around real algebraic sets

The goal of this section is to provide the proof of Theorem 1.2.

3.1. Bounding the volume of tubes of smooth hypersurfaces.

In an important paper, Weyl [42] derived a formula for the volume of tubes around a compact submanifold of a Euclidean space or a sphere. However, this formula only holds for a sufficiently small radius. The following proposition gives an upper bound on the volume of tubes around a hypersurface that holds for any radius. Compare also Gray [17, Theorem 8.4, (8.6), p. 162].

Proposition 3.1. *Let M be a compact oriented smooth hypersurface of S^p and U be an open subset of M. Then we have for all $0 < \alpha \le \pi/2$*

$$\mathsf{vol}_p T_R^{\perp}(U,\alpha) \le 2 \sum_{i=0}^{p-1} J_{p,i+1}(\alpha)\, |\mu_i|(U).$$

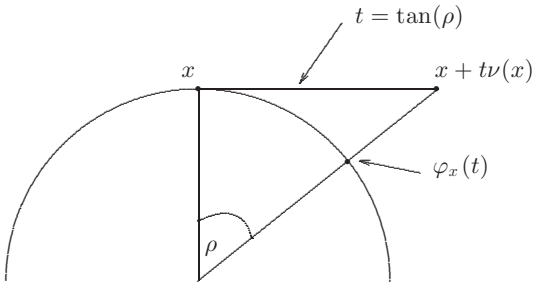

FIGURE 3. The point $\varphi_x(t)$ and the magnitudes t and ρ

Proof. Let $\nu\colon M \to S^p$ be the unit normal vector field on M corresponding to its orientation. For $x \in M$ consider the following parametrization:

$$\varphi_x\colon \mathbb{R} \to S^p,\ \varphi_x(t) = \frac{x + t\nu(x)}{\|x + t\nu(x)\|} = \frac{x + t\nu(x)}{(1+t^2)^{\frac{1}{2}}}$$

of the half great circle intersecting M at x orthogonally. Note that if we denote $\rho = d_R(x, \varphi_x(t))$, then $\rho = \arctan t$.

Consider the following differentiable map of Riemannian manifolds:

$$\varphi\colon M \times \mathbb{R} \to S^p, (x,t) \mapsto \varphi_x(t).$$

Let U be an open subset of M, $0 < \alpha \le \pi/2$, and put $\tau = \tan\alpha$. We denote by $T_R^+(U,\alpha)$ and $T_R^-(U,\alpha)$ the images of $U \times (0,\tau)$ and $U \times (-\tau, 0)$ under the map φ, respectively. Clearly, $T_R^\perp(U,\alpha) = U \cup T_R^+(U,\alpha) \cup T_R^-(U,\alpha)$.

We apply the transformation formula of Proposition 2.3 to the surjective differentiable map $\varphi\colon U \times (0,\tau) \to T_R^+(U,\alpha)$ of Riemannian manifolds. This yields

$$\int_{U\times(0,\tau)} |\det D\varphi|\, d(M \times \mathbb{R}) = \int_{y \in T_R^+(U,\alpha)} \#\varphi^{-1}(y)\, dS^p \ \ge\ \mathsf{vol}_p T_R^+(U,\alpha).$$

By Fubini, $\int_{U\times(0,\tau)} |\det D\varphi| d(M\times\mathbb{R}) = \int_0^\tau g(t)\,dt$, where

$$g(t) := \int_{x\in U} |\det D\varphi|(x,t)\, dM(x).$$

Claim A. The determinant of the derivative $D\varphi(x,t)$ of φ at $(x,t) \in M\times\mathbb{R}$ satisfies

$$|\det D\varphi(x,t)| = \frac{1}{(1+t^2)^{(p+1)/2}} \left|\det(\mathrm{id}_{T_xM} - tL_M(x))\right|. \tag{6}$$

Using this claim, whose proof is postponed to the end, we obtain

$$\begin{aligned} g(t) &= \int_{x\in U} \frac{1}{(1+t^2)^{\frac{p+1}{2}}} |\det(\mathrm{id}_{T_xM} - tL_M(x))|\, dM(x) \qquad \text{(by Claim A)} \\ &\le \sum_{i=0}^{p-1} \frac{|t|^i}{(1+t^2)^{\frac{p+1}{2}}} \int_U |K_{M,i}|\, dM \qquad \text{(by (4))} \\ &= \sum_{i=0}^{p-1} \frac{|t|^i}{(1+t^2)^{\frac{p+1}{2}}}\ |\mu_i|(U). \end{aligned}$$

By making the substitution $t = \tan\rho$ (recall $\tau = \tan\alpha$) we get

$$\int_0^\tau \frac{t^i}{(1+t^2)^{(p+1)/2}}\,dt = \int_0^\alpha (\cos\rho)^{p-i-1}(\sin\rho)^i\,d\rho = J_{p,i+1}(\alpha).$$

Altogether we conclude that

$$\mathsf{vol}_p T_R^+(U,\alpha) \le \int_0^\tau g(t)\,dt \le \sum_{i=0}^{p-1} J_{p,i+1}(\alpha)|\mu_i|(U).$$

The same estimate can be shown for $\mathsf{vol}_p T_R^-(U,\alpha)$, which implies the desired estimate of $\mathsf{vol}_p T_R^\perp(U,\alpha)$.

It remains to prove Claim A. Choose a local parametrization $x = x(v) \in \mathbb{R}^{p+1}$ of M in terms of coordinates $v_1,\dots,v_{p-1}$ defined in a neighborhood of 0. We assume that $\partial_{v_1}x,\dots,\partial_{v_{p-1}}x$ are orthonormal at 0. Abusing notation we will interpret $\nu = \nu(v)$ as a function of v. The matrix (λ_{ij}) of L_M with respect to the basis $\partial_{v_j}x$ of T_xM is given by $-\partial_{v_i}\nu = \sum_j \lambda_{ij}\partial_{v_j}x$.

Consider the map $(v,t) \mapsto R(v,t) := x(v) + t\nu(v) \in \mathbb{R}^{p+1}$. Then

$$(v,t) \mapsto \psi(v,t) := \varphi(x(v),t) = \frac{R(v,t)}{(1+t^2)^{\frac12}}$$

is a local parametrization of S^p. In the following let $[R,\partial_t R,\partial_{v_1}R,\dots,\partial_{v_{p-1}}R]$ denote the square matrix of size $p+1$ whose rows are R and the partial derivatives of R. Using the multilinearity of the determinant and the fact that

$$\partial_t\psi = (1+t^2)^{-1/2}\partial_t R - t(1+t^2)^{-3/2}R$$

we obtain by a short calculation that

$$\begin{aligned} |\det D\psi(v,t)| &= |\det[\psi,\partial_t\psi,\partial_{v_1}\psi,\dots,\partial_{v_{p-1}}\psi]| \\ &= \frac{1}{(1+t^2)^{(p+1)/2}}\,\big|\det[R,\partial_t R,\partial_{v_1}R,\dots,\partial_{v_{p-1}}R]\big|. \end{aligned} \tag{7}$$

Computing partial derivatives we get $\partial_{v_i}R = \partial_{v_i}x + t\partial_{v_i}\nu$ and $\partial_t R = \nu$. Using $\partial_{v_i}\nu = -\sum_j \lambda_{ij}\partial_{v_j}x$, we get

$$\partial_{v_i}R = \textstyle\sum_j(\delta_{ij} - t\lambda_{ij})\partial_{v_j}x.$$

Hence we obtain

$$\begin{aligned} \det[R,\partial_t R,\partial_{v_1}R,\dots,\partial_{v_{p-1}}R] &= \det[x+t\nu,\nu,\partial_{v_1}R,\dots,\partial_{v_{p-1}}R] \\ &= \det[x,\nu,\partial_{v_1}R,\dots,\partial_{v_{p-1}}R] \\ &= \det(\delta_{ij}-t\lambda_{ij})\det[x,\nu,\partial_{v_1}x,\dots,\partial_{v_{p-1}}x]. \end{aligned} \tag{8}$$

Since we assume that $\partial_{v_1}x,\dots,\partial_{v_{p-1}}x$ are orthonormal at $v=0$ (and also orthogonal to x and ν since $T_M(x)$ is so) we conclude (using the chain rule $D\psi = D\varphi Dx$, together with (7) and (8)) that

$$|\det D\varphi(x,t)| = |\det D\psi(0,t)| = \frac{1}{(1+t^2)^{(p+1)/2}}|\det(\delta_{ij}-t\lambda_{ij})|,$$

which shows Claim A. □

3.2. **Bounding integrals of absolute curvature in terms of degree.** In this section let $f \in \mathbb{R}[X_0, \ldots, X_p]$ be homogeneous of degree $d > 0$ with a nonempty zero set $V \subseteq S^p$ such that the derivative of the restriction of f to S^p does not vanish on V. Then V is a compact smooth hypersurface of S^p. We orient V by the following unit normal vector field (Gauss map):

$$\nu\colon V \to S^p,\ \nu(x) = \|\operatorname{grad} f(x)\|^{-1} \operatorname{grad} f(x).$$

The goal of this section is to bound the integrals of absolute curvature on patches of V.

Proposition 3.2. *For $a \in S^p$, $0 < \sigma \leq 1$, and $0 \leq i < p$ we have*

$$|\mu_i|(V \cap B_{\mathbb{P}}(a, \sigma)) \ \leq\ 2\frac{p-1}{i}\mathcal{O}_{p-1}\, d^{i+1}\, \sigma^{p-i-1}.$$

The proof is based on the following lemma.

Lemma 3.3. *We have $|\mu_{p-1}|(V) \leq \mathcal{O}_{p-1} d^p$.*

Proof. Recall that the determinant of the linear map $L_V(x) = -D\nu(x)\colon T_xV \to T_xV$ equals $K_{V,p-1}$ (cf. Section 2.3). We may assume without loss of generality that the open subset $U := \{x \in V \mid \mathsf{rank}(D\nu(x)) = p-1\}$ of V is nonempty (otherwise $\mu_{p-1}(V) = 0$). We would like to apply Proposition 2.3 to the restriction of ν to U, but face the problem that ν is only an immersion. Hence $\nu(U)$ might not be a submanifold of S^p. In order to circumvent this, we use some standard facts of real algebraic geometry [5].

Consider the Zariski closure W of $\nu(U)$ in S^p, which is a real algebraic variety of dimension $p-1$. Its regular locus W_1 is a submanifold of S^p of dimension $p-1$. Consider the open subset $V_1 := U \cap \nu^{-1}(W_1)$ of V and the restriction $\nu_1\colon V_1 \to W_1$ of ν. The singular locus $\operatorname{Sing}(W) = W \setminus W_1$ is an algebraic subset of dimension strictly less than $\dim W$. Since ν_1 is an immersion, we conclude that $U \setminus V_1 = \nu_1^{-1}(\operatorname{Sing}(W))$ has dimension strictly less than $p-1$. We therefore obtain

$$|\mu_{p-1}(V)| = |\mu_{p-1}(U)| = |\mu_{p-1}(V_1)|.$$

Applying the transformation formula of Proposition 2.3 to ν_1 we get

$$|\mu_{p-1}(V)| = \int_{V_1} |\det D\nu_1|\, dV = \int_{y \in W_1} \#\nu_1^{-1}(y)\, dW_1(y).$$

Consider for $\ell \in \mathbb{N} \cup \{\infty\}$ the semialgebraic sets $F_\ell := \{y \in W_1 \mid |\nu_1^{-1}(y)| = \ell\}$. Since the above integral is finite, $\mathsf{vol}_{p-1} F_\infty = 0$, and therefore $\dim F_\infty < p-1$. We obtain

$$\int_{y \in W_1} \#\nu_1^{-1}(y)\, dW_1(y) = \sum_{\ell \geq 0} \ell\, \mathsf{vol}_{p-1}\, F_\ell.$$

Corollary 2.8 applied to F_ℓ yields the following:

$$\mathsf{vol}_{p-1} F_\ell = \frac{\mathcal{O}_{p-1}}{2} \int_{g \in G} \#(F_\ell \cap gS^1)\, dG(g).$$

Combining these findings we get

$$|\mu_{p-1}|(V) = \frac{\mathcal{O}_{p-1}}{2} \int_{g \in G} \sum_{\ell \geq 0} \ell\, \#(F_\ell \cap gS^1)\, dG(g) = \frac{\mathcal{O}_{p-1}}{2} \int_{g \in G} \#\nu_1^{-1}(gS^1)\, dG(g).$$

(In order to see the last equality use the fact that gS^1 does not intersect F_∞ almost surely.)

It is now sufficient to prove that $\#\nu^{-1}(gS^1) \leq 2d^p$ for almost all $g \in G$. To simplify notation suppose first that $g = \mathrm{id}$. Let $y_0, \ldots, y_p$ denote coordinate functions on $\mathbb{R}^{p+1}$ and $S^1 = \{y \in S^p \mid y_2 = \cdots = y_p = 0\}$. A point $x \in \mathbb{R}^{p+1}$ lies in $\nu^{-1}(S^1)$ iff it satisfies the following system of equations:

$$\sum_i x_i^2 - 1 = 0,\ f(x) = 0,\ \partial_2 f(x) = \cdots = \partial_p f(x) = 0.$$

If all real solutions of this system of equations are nondegenerated, then they are isolated in $\mathbb{C}^{p+1}$. By Bézout's theorem [23] the number of these solutions is bounded by $2\,d\,(d-1)^{p-1} \leq 2d^p$. Furthermore, one can show along the lines in [20] that the nondegeneracy condition is satisfied for almost all $g \in G$. This finishes the proof. □

Proof of Proposition 3.2. Put $U := V \cap B_{\mathbb{P}}(a, \sigma)$. The case $i = p-1$ is already settled by Lemma 3.3, as $|\mu_i|(U) \leq |\mu_i|(V)$. So we may assume $i < p-1$.

Let U_+ be the set of points of U where $K_{V,i}$ is positive, and similarly define U_- where $K_{V,i}$ is negative. Then $|\mu_i|(U) = |\mu_i(U_+)| + |\mu_i(U_-)|$.

Fix a subsphere $S^{i+1} \subset S^p$. Let $g \in G = O(p+1)$ such that V intersects gS^{i+1} transversally and such that the intersection is nonempty. Then $V \cap gS^{i+1}$ is the zero set in gS^{i+1} of the homogeneous polynomial f of degree d. By transversality, the derivative of the restriction of f to gS^{i+1} does not vanish on $V \cap gS^{i+1}$. Hence we may apply Lemma 3.3 (now refering to the hypersurface $V \cap gS^{i+1}$ of gS^{i+1}) which gives the estimate

$$|\mu_i|(V \cap gS^{i+1}) \leq \mathcal{O}_i d^{i+1}.$$

By the monotonicity of $|\mu_i|$, we have

$$|\mu_i(U_+ \cap gS^{i+1})| \leq |\mu_i|(U_+ \cap gS^{i+1}) \leq |\mu_i|(V \cap gS^{i+1}) \leq \mathcal{O}_i d^{i+1}.$$

The Kinematic Formula of Theorem 2.7 implies that

$$|\mu_i(U_+)| \leq \mathcal{C}(p,i) \int_{g \in G} |\mu_i(gU_+ \cap S^{i+1})|\, dG(g).$$

Therefore, using the fact that $\mu_i(gU_+ \cap S^{i+1}) = \mu_i(U_+ \cap g^{-1}S^{i+1})$, we obtain

$$|\mu_i(U_+)| \ \leq\ \mathcal{C}(p,i)\ \mathcal{O}_i d^{i+1}\ \operatorname*{Prob}_{g \in G}\{gU_+ \cap S^{i+1} \neq \emptyset\},$$

where the probability is taken with respect to the uniform distribution in G. We may estimate this probability as follows (put $\alpha = \arcsin \sigma$ and note that $gU_+ \subseteq B_{\mathbb{P}}(ga, \sigma)$):

$$\begin{aligned}\operatorname*{Prob}_{g \in G}\{B_{\mathbb{P}}(ga,\sigma) \cap S^{i+1} \neq \emptyset\} &= \operatorname*{Prob}_{a' \in S^p}\{B_{\mathbb{P}}(a',\sigma) \cap S^{i+1} \neq \emptyset\} \\ &= \frac{1}{\mathcal{O}_p}\mathsf{vol}_p T_{\mathbb{P}}(S^{i+1},\sigma) = \frac{\mathcal{O}_{i+1}\mathcal{O}_{p-i-2}}{\mathcal{O}_p}\, J_{p,p-i-1}(\alpha) \leq \frac{\mathcal{O}_{i+1}\mathcal{O}_{p-i-2}}{\mathcal{O}_p}\ \frac{\sigma^{p-i-1}}{p-i-1},\end{aligned}$$

where we used Lemmas 2.1 and 2.2 for the last two steps. Multiplying this with the formula for $\mathcal{C}(p,i)$, the expression miraculously simplifies and we get

$$|\mu_i|(U_+) \ \leq\ \frac{p-1}{i}\mathcal{O}_{p-1}\ d^{i+1}\ \sigma^{p-i-1}.$$

The same estimate can be shown for $|\mu_i|(U_-)$, which proves the proposition. □

3.3. **Extension to the nonsmooth case: proof of Theorem 1.2.** The following proposition estimates the volume of the tube around a patch of a smooth hypersurface in the sphere.

Proposition 3.4. *Let $f \in \mathbb{R}[X_0, \dots, X_p]$ be homogeneous of degree $d > 0$ with zero set $V = \mathcal{Z}(f)$ in S^p. Assume that the derivative of the restriction of f to S^p does not vanish on V. (Thus V is a smooth hypersurface in S^p.) Let $a \in S^p$ and $0 < \varepsilon, \sigma \le 1$. Then*

$$\mathsf{vol}_p T^\perp_{\mathbb{P}}(V \cap B_{\mathbb{P}}(a,\sigma),\varepsilon) \;\le\; \frac{4\mathcal{O}_{p-1}}{p} \sum_{k=1}^{p-1} \frac{p}{k}\, d^k\, \varepsilon^k\, \sigma^{p-k} + 2\mathcal{O}_p\, d^p\, \varepsilon^p.$$

Proof. Put $U := V \cap B_{\mathbb{P}}(a,\sigma)$. Take $0 < \alpha \le \pi/2$ such that $\varepsilon = \sin\alpha$. Proposition 3.1 and the symmetry of U imply

$$\mathsf{vol}_p T^\perp_{\mathbb{P}}(U,\varepsilon) = \mathsf{vol}_p T^\perp_R(U,\alpha) \le 2 \sum_{i=0}^{p-1} J_{p,i+1}(\alpha)\, |\mu_i|(U).$$

Combining this with Proposition 3.2 we obtain

$$\mathsf{vol}_p T^\perp_{\mathbb{P}}(U,\varepsilon) \le\; 4 \sum_{i=0}^{p-1} \frac{p-1}{i} \mathcal{O}_{p-1} d^{i+1} \sigma^{p-i-1} J_{p,i+1}(\alpha).$$

Using the estimates of Lemma 2.2 we obtain (put $k = i+1$ and consider separately the term for $k = p$)

$$\mathsf{vol}_p T^\perp_{\mathbb{P}}(U,\varepsilon) \le\; 4 \sum_{k=1}^{p-1} \frac{p-1}{k-1} \mathcal{O}_{p-1}\, d^k\, \sigma^{p-k}\, \frac{\varepsilon^k}{k} + 4\mathcal{O}_{p-1}\, d^p \frac{\mathcal{O}_p}{2\mathcal{O}_{p-1}}\, \varepsilon^p.$$

Now use $\frac{p-1}{k-1} = \frac{k}{p}\frac{p}{k}$ to get the desired upper bound on $\mathsf{vol}_p T^\perp_{\mathbb{P}}(U,\varepsilon)$. □

Proof of Theorem 1.2. We have to remove the smoothness assumption in Proposition 3.4 and to estimate the volume of the ε-neighborhood instead of the ε-tube.

Assume $W = \mathcal{Z}(f_1, \dots, f_r)$ with homogeneous polynomials f_i of degree d_i. Then W is the zero set in S^p of the polynomial

$$f(X) := \sum_{i=1}^{r} f_i(X)^2 \|X\|^{2d-2d_i},$$

which is homogeneous of degree $2d$. Our assumption $W \ne S^p$ implies $\dim W < p$.

Let $\delta > 0$ be smaller than any positive critical value of the restriction $\widetilde{f}\colon S^p \to \mathbb{R}$ of f to S^p. Then $D_\delta := \{\xi \in S^p \mid \widetilde{f}(\xi) \le \delta\}$ is a compact domain in S^p with smooth boundary

$$\partial D_\delta = \{\xi \in S^p \mid \widetilde{f}(\xi) = \delta\}.$$

Indeed, the derivative of $\widetilde{f} - \delta$ does not vanish on ∂D_δ (use $\sum_i x_i \partial_i f(x) = 2df(x)$). Moreover, note that $W = \bigcap_{\delta>0} D_\delta$ and hence $\lim_{\delta\to 0} \mathsf{vol}_p D_\delta = \mathsf{vol}_p(W) = 0$, since $\dim W < p$.

Claim B. We have $T_{\mathbb{P}}(W,\varepsilon) \subseteq D_\delta \cup T_{\mathbb{P}}(\partial D_\delta, \varepsilon)$ for $0 < \varepsilon \le 1$.

In order to see this, let $x \in T_{\mathbb{P}}(W,\varepsilon) \setminus D_\delta$ and $\gamma\colon [0,1] \to S^p$ be a segment of Riemannian length less than $\arcsin\varepsilon$ such that $\gamma(1) = x$ and $\gamma(0) \in W$. Consider $F\colon [0,1] \to \mathbb{R}, F(t) := \widetilde{f}(\gamma(t))$. By assumption $F(1) = \widetilde{f}(x) > \delta$ and $F(0) = 0$. Hence

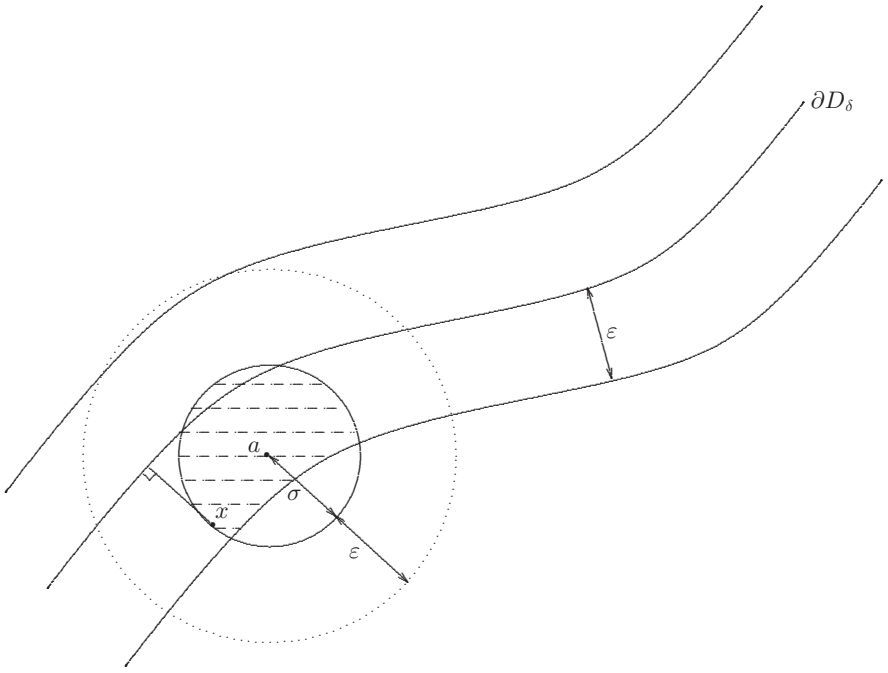

FIGURE 4. The shaded area is $T_{\mathbb{P}}(\partial D_\delta, \varepsilon) \cap B_{\mathbb{P}}(a, \sigma)$

there exists $\tau \in (0,1)$ such that $F(\tau) = \delta$. Thus $\gamma(\tau) \in \partial D_\delta$ and $d_{\mathbb{P}}(x, \partial D_\delta) \leq d_{\mathbb{P}}(x, \gamma(\tau)) < \varepsilon$, which shows the claim.

Using the triangle inequality for the projective distance, it is easy to see that (cf. Figure 4)

$$T_{\mathbb{P}}(\partial D_\delta, \varepsilon) \cap B_{\mathbb{P}}(a, \sigma) \subseteq T^{\perp}_{\mathbb{P}}(\partial D_\delta \cap B_{\mathbb{P}}(a, \sigma + \varepsilon), \varepsilon). \tag{9}$$

Combining (9) with Claim B we arrive at

$$T_{\mathbb{P}}(W, \varepsilon) \cap B_{\mathbb{P}}(a, \sigma) \subseteq D_\delta \cup T^{\perp}_{\mathbb{P}}(\partial D_\delta \cap B_{\mathbb{P}}(a, \sigma + \varepsilon), \varepsilon).$$

We apply Proposition 3.4 to $V = \partial D_\delta = \mathcal{Z}(f - \delta\|x\|^{2d})$ intersected with the ball $B_{\mathbb{P}}(a, \sigma + \varepsilon)$. This implies

$$\mathsf{vol}_p T^{\perp}_{\mathbb{P}}(\partial D_\delta \cap B_{\mathbb{P}}(a, \sigma + \varepsilon), \varepsilon) \leq \frac{4\mathcal{O}_{p-1}}{p} \sum_{k=1}^{p-1} \frac{p}{k} (2d)^k \varepsilon^k (\sigma + \varepsilon)^{p-k} + 2\mathcal{O}_p (2d)^p \varepsilon^p.$$

Taking into account that $\mathsf{vol}_p B_{\mathbb{P}}(a,\sigma) \geq 2\mathcal{O}_{p-1}\frac{\sigma^p}{p}$ (cf. Lemmas 2.2 and 2.1) we obtain

$$\begin{aligned}\frac{\mathsf{vol}_p\left(T_{\mathbb{P}}(W,\varepsilon)\cap B_{\mathbb{P}}(a,\sigma)\right)}{\mathsf{vol}_p B_{\mathbb{P}}(a,\sigma)} &\leq \frac{\mathsf{vol}_p D_\delta}{\mathsf{vol}_p B_{\mathbb{P}}(a,\sigma)} + \frac{\mathsf{vol}_p T_{\mathbb{P}}^{\perp}(\partial D_\delta \cap B_{\mathbb{P}}(a,\sigma+\varepsilon),\varepsilon)}{\mathsf{vol}_p B_{\mathbb{P}}(a,\sigma)}\\ &\leq \frac{\mathsf{vol}_p D_\delta}{\mathsf{vol}_p B_{\mathbb{P}}(a,\sigma)} + 2\sum_{k=1}^{p-1}\frac{p}{k}(2d)^k\left(1+\frac{\varepsilon}{\sigma}\right)^{p-k}\left(\frac{\varepsilon}{\sigma}\right)^k\\ &\quad + \frac{p\mathcal{O}_p}{\mathcal{O}_{p-1}}(2d)^p\left(\frac{\varepsilon}{\sigma}\right)^p.\end{aligned}$$

Taking the limit for $\delta\to 0$ the first term vanishes and the assertion follows. □

3.4. **Estimating expected values: proof of Theorem 1.1.** The tail bounds in Theorem 1.1 follow from Theorem 1.2 as indicated in the Introduction. It thus suffices to show how to derive the claim on expected values from the tail bounds. This is achieved by the following proposition.

We use the inequality $\frac{\mathcal{O}_p}{2\mathcal{O}_{p-1}}-\frac{1}{p}\leq\frac{1}{2}$, valid for $p\geq 2$, which implies $\frac{p\mathcal{O}_p}{\mathcal{O}_{p-1}}\leq p+2$.

Proposition 3.5. *For $0<\sigma\leq 1$ let $X_\sigma\geq 1$ be a random variable satisfying, for all $0<\varepsilon\leq 1$ and $p\geq 2$,*

$$\mathrm{Prob}\{X_\sigma\geq 1/\varepsilon\}\leq 2\sum_{k=1}^{p-1}\frac{p}{k}(2d)^k\left(1+\frac{\varepsilon}{\sigma}\right)^{p-k}\left(\frac{\varepsilon}{\sigma}\right)^k+(p+2)(2d)^p\left(\frac{\varepsilon}{\sigma}\right)^p.$$

Then, for $\varepsilon\leq\frac{\sigma}{(1+2d)(p-1)}$, we have $\mathrm{Prob}\{X_\sigma\geq 1/\varepsilon\}\leq(8e+4)dp\frac{\varepsilon}{\sigma}$ and

$$\mathbf{E}(\ln X_\sigma)\leq 2\ln p+2\ln d+2\ln\frac{1}{\sigma}+4.7.$$

Proof. $\mathrm{Prob}\{X_\sigma\geq 1/\varepsilon\}$ is bounded by

$$\begin{aligned}&2\left[\sum_{k=1}^{p}\frac{p}{k}(2d)^k\left(1+\frac{\varepsilon}{\sigma}\right)^{p-k}\left(\frac{\varepsilon}{\sigma}\right)^k+\frac{p}{2}(2d)^p\left(\frac{\varepsilon}{\sigma}\right)^p\right]\\ &= 4d\frac{\varepsilon}{\sigma}\left[\sum_{k=1}^{p}\frac{p}{k}(2d)^{k-1}\left(1+\frac{\varepsilon}{\sigma}\right)^{p-k}\left(\frac{\varepsilon}{\sigma}\right)^{k-1}+\frac{p}{2}(2d)^{p-1}\left(\frac{\varepsilon}{\sigma}\right)^{p-1}\right]\\ &\leq \frac{4dp\varepsilon}{\sigma}\left[\sum_{k=0}^{p-1}\frac{p-1}{k}(2d)^k\left(1+\frac{\varepsilon}{\sigma}\right)^{p-1-k}\left(\frac{\varepsilon}{\sigma}\right)^k+\frac{1}{2}(2d)^{p-1}\left(\frac{\varepsilon}{\sigma}\right)^{p-1}\right].\end{aligned}$$

Using that $\varepsilon\leq\frac{\sigma}{(1+2d)(p-1)}$, this can be further bounded to obtain

$$\mathrm{Prob}\{X_\sigma\geq 1/\varepsilon\}\leq\frac{4dp\varepsilon}{\sigma}\left[\left(1+\frac{1}{p-1}\right)^{p-1}+\frac{1}{2}\right]\leq\frac{4dp}{\sigma}(e+\frac{1}{2})\,\varepsilon.$$

For the bound on the expectation we use Proposition 2.4 in [7] which implies

$$\begin{aligned}\mathbf{E}(\ln X_\sigma) &\leq \ln\left(\frac{(1+2d)(p-1)}{\sigma}\right)+\ln\left(\frac{4dp}{\sigma}(e+\frac{1}{2})\right)+1\\ &\leq 2\ln p+2\ln d+2\ln\frac{1}{\sigma}+\ln(3\cdot 4(e+\frac{1}{2})e) \qquad \text{since } 1+2d\leq 3d\\ &\leq 2\ln p+2\ln d+2\ln\frac{1}{\sigma}+4.7,\end{aligned}$$

where we used $\ln(12e(e+\frac{1}{2}))\approx 4.7$. □

4. Applications

We give several applications of Theorem 1.1 to smooth analysis estimates for the condition numbers of the following problems: linear equation solving, Moore-Penrose inversion, eigenvalue computations, real polynomial equation solving, and zero counting.

4.1. **Linear equation solving.** The first natural application of our result is for the classical condition number $\kappa(A)$. In [46], M. Wschebor showed (solving a conjecture posed in [31]) that, for all $n \times n$ real matrices A with $\|A\| \leq 1$, all $0 < \sigma \leq 1$ and all $t > 0$,

$$\operatorname*{Prob}_{Z \in N^{n^2}(A,\sigma^2)} (\kappa(Z) \geq t) \leq \frac{Kn}{\sigma t},$$

with K a universal constant. Hereby, $\|A\|$ stands for the operator norm with respect to euclidean norm. Note that, by Proposition 2.4 in [7], this implies

$$\sup_{\|A\| \leq 1} \mathop{\mathbf{E}}_{Z \in N^{n^2}(A,\sigma^2)} (\ln \kappa(Z)) \leq \ln n + \ln \frac{1}{\sigma} + \ln K + 1.$$

We next compare Wschebor's result with what can be obtained from Theorem 1.1. Let $\|A\|_F$ denote the Frobenius norm of a matrix $A \in \mathbb{R}^{n \times n}$, which is induced by the inner product $(A, B) \mapsto \mathsf{trace}(AB^T)$. We have

$$\kappa(A) = \|A\|\|A^{-1}\| \leq \|A\|_F \|A^{-1}\| =: \kappa_F(A).$$

The Condition Number Theorem of Eckart and Young [15] states that $\|A^{-1}\| = d_F(A, \Sigma)^{-1}$ where $\Sigma \subseteq \mathbb{R}^{n \times n}$ denotes the set of singular matrices and d_F is the distance induced by the Frobenius norm (see also [4, Thm. 1, Ch. 11]). It follows that κ_F is a conic condition number. We can thus give upper bounds for $\kappa_F(A)$ and they will hold as well for $\kappa(A)$.

Corollary 4.1. *For all $n \geq 1$, $0 < \sigma \leq 1$, we have*

$$\sup_{Z \in S^{n^2-1}} \mathop{\mathbf{E}}_{Z \in B_{\mathbb{P}}(A,\sigma)} (\ln \kappa_F(Z)) \leq 6 \ln n + 2 \ln \frac{1}{\sigma} + 4.7,$$

where the expectation is over all Z uniformly distributed in the disk of radius σ centered at A in the sphere S^{n^2-1} endowed with the projective distance.

Proof. The variety Σ of singular matrices is the zero set of the determinant, which is a homogeneous polynomial of degree n. We now apply Theorem 1.1 with $p = n^2 - 1$. □

Remark 4.2. (i) In [7] a bound is proved on

$$\sup_{Z \in \mathbb{P}^{n^2-1}(\mathbb{C})} \mathop{\mathbf{E}}_{Z \in B_{\mathbb{P}}(A,\sigma)} (\ln \kappa_F(Z))$$

where $\mathbb{P}^{n^2-1}(\mathbb{C})$ denotes the (n^2-1)-dimensional complex projective space and $\|A\|_F = \mathsf{trace}(AA^*)$. While Corollary 4.1 is close to this result it does not follow from it. Note that the balls $B_{\mathbb{P}}(A, \sigma)$ over which one takes the expected value $\mathbf{E}(\ln \kappa_F(Z))$ are different, and there is no obvious way to bound the expected value for the real case in terms of that for the complex one. For the converse, however, one may identify $\mathbb{C}$ with $\mathbb{R}^2$ and do a smoothed analysis for a complex problem using Theorem 1.1. We will see instances of this in Sections 4.3 and 4.4 below.

(ii) Note, the bound in Corollary 4.1 is of the same order of magnitude as Wschebor's, worse by just a constant factor. On the other hand, its derivation from Theorem 1.1 is rather immediate.

We next extend the bound in Corollary 4.1 to rectangular matrices.

4.2. **Moore-Penrose inversion.** Let $\ell \geq m$ and consider the space $\mathbb{R}^{\ell \times m}$ of $\ell \times m$ rectangular matrices. Denote by $\Sigma \subset \mathbb{R}^{\ell \times m}$ the subset of rank-deficient matrices. For $A \notin \Sigma$ let $A^\dagger$ denote its Moore-Penrose inverse (see, e.g., [2, 8]). The condition number of A (for the computation of $A^\dagger$) is defined as

$$\mathsf{cond}^\dagger(A) = \lim_{\varepsilon \to 0} \sup_{\|\Delta A\| \leq \varepsilon} \frac{\|(A + \Delta A)^\dagger - A^\dagger\| \cdot \|A\|}{\|A^\dagger\| \cdot \|\Delta A\|}.$$

This is not a conic condition number but it happens to be close to one. One defines $\kappa^\dagger(A) = \|A\|\|A^\dagger\|$ and, since $\|A^\dagger\| = d_F(A, \Sigma)^{-1}$ [16], we obtain

$$\kappa^\dagger(A) = \frac{\|A\|}{d_F(A, \Sigma)}.$$

In addition (see [37, §III.3]),

$$\kappa^\dagger(A) \leq \mathsf{cond}^\dagger(A) \leq \frac{1 + \sqrt{5}}{2} \kappa^\dagger(A).$$

Thus, $\ln(\mathsf{cond}^\dagger(A))$ differs from $\ln(\kappa^\dagger(A))$ by just a small additive constant. As for square matrices, $\kappa^\dagger(A)$ is not conic since the operator norm is not induced by an inner product in $\mathbb{R}^{\ell \times m}$. But, again, we can bound $\kappa^\dagger(A)$ by the conic condition number $\kappa_F^\dagger(A) := \|A\|_F \|A^\dagger\| \geq \kappa^\dagger(A)$.

Corollary 4.3. *For all $\ell \geq m \geq 1$ and $0 < \sigma \leq 1$ we have*

$$\sup_{A \in S^{\ell m - 1}} \mathop{\mathbf{E}}_{Z \in B_{\mathbb{P}}(A, \sigma)} (\ln \kappa_F^\dagger(Z)) \leq 2 \ln \ell + 4 \ln m + 2 \ln \frac{1}{\sigma} + 4.7.$$

Proof. If a matrix A is rank deficient, then $\det A_0 = 0$ where A_0 is the $m \times m$ matrix obtained by removing all rows of A with index greater than m. Therefore $\Sigma \subseteq \Sigma_0 = \{A \in \mathbb{R}^{\ell \times m} \mid \det A_0 = 0\}$. This implies that $\kappa_F^\dagger(A) \leq \frac{1}{d_F(A, \Sigma_0)}$ for $\|A\|_F = 1$. Since Σ_0 is the zero set of a homogeneous polynomial of degree m, an immediate application of Theorem 1.1 with $W = \Sigma_0$ yields the claimed bound. □

4.3. **Eigenvalue computations.** Let $A \in \mathbb{K}^{n \times n}$ ($\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$) and consider the problem of computing the eigenvalues of A. Even though these eigenvalues (and A when $\mathbb{K} = \mathbb{C}$) may be complex, we can still consider the problem as being over the reals by identifying $\mathbb{C}$ with $\mathbb{R}^2$ (and hence, consider $A \in \mathbb{R}^{2n^2}$ when $\mathbb{K} = \mathbb{C}$). Note that this is actually the way computers deal with complex numbers.

Let $\lambda \in \mathbb{C}$ be a simple eigenvalue of A. Suppose that $x \in \mathbb{C}^n$ and $y \in \mathbb{C}^n$ are right and left eigenvectors associated to λ, respectively (i.e., nonzero and satisfying $Ax = \lambda x$ and $y^* A = \lambda y^*$ (here y^* is the transpose conjugate of y)). From the fact that λ is a simple eigenvalue one can deduce that $y^* x = \langle x, y \rangle \neq 0$; cf. Wilkinson [44].

For any sufficiently small perturbation $\Delta A \in \mathbb{K}^{n \times n}$ there exists a unique eigenvalue $\lambda + \Delta\lambda$ of $A + \Delta A$ close to λ. 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^* from the left we get

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

Moreover, $\sup_{\|\Delta A\|_F \le 1} |y^* \Delta A x| = \|x\|\, \|y\|$.

It therefore makes sense to define the condition number of A for the computation of λ as follows:

$$\kappa(A, \lambda) := \frac{\|x\|\, \|y\|}{|\langle x, y\rangle|} \tag{10}$$

and to set $\kappa(A, \lambda) := \infty$ if λ is a multiple eigenvalue of A. Then, one takes

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

where the maximum is over all eigenvalues λ of A.

Let $\Sigma \subseteq \mathbb{K}^{n\times n}$ be the set of matrices having multiple eigenvalues. A result by Wilkinson [44] shows that

$$\kappa_{\mathsf{eigen}}(A) \le \frac{\sqrt{2}\|A\|_F}{\mathsf{dist}(A, \Sigma)}. \tag{11}$$

In [13], Demmel used the fact that the right-hand side of (11) is conic in order to obtain bounds on the tail of $\kappa_{\mathsf{eigen}}(A)$ for random A. We next use this same fact to obtain smoothed analysis estimates.

Proposition 4.4. *For all $n \ge 1$ and $0 < \sigma \le 1$ we have*

(i) *For real matrices A,*

$$\sup_{A\in S^{n^2-1}} \mathop{\mathbf{E}}_{Z\in B(A,\sigma)} (\ln \kappa_{\mathsf{eigen}}(Z)) \le 8 \ln n + 2 \ln \frac{1}{\sigma} + 5.1.$$

(ii) *For complex matrices A,*

$$\sup_{A\in S^{2n^2-1}} \mathop{\mathbf{E}}_{Z\in B(A,\sigma)} (\ln \kappa_{\mathsf{eigen}}(Z)) \le 8 \ln n + 2 \ln \frac{1}{\sigma} + 6.5.$$

Proof. (i) It is well known that Σ is the zero set of the discriminant of the characteristic polynomial of A, which is a homogeneous polynomial of degree $n^2 - n$ in the entries of A (compare [7, Prop. 3.4]). Theorem 1.1 applied to the conic condition number $\frac{\|A\|_F}{\mathsf{dist}(A,\Sigma)}$ implies the stated bound.

(ii) Note that Σ, as a subset of $\mathbb{R}^{2n^2}$, is the zero set of the real and imaginary parts of the discriminant polynomial, which both have degree $n^2 - n$. Then apply Theorem 1.1 with $p = 2n^2$. □

4.4. **Solving polynomial systems.** Let $d_1, \dots, d_n \in \mathbb{N} \setminus \{0\}$. We denote by $\mathcal{H}_{\mathbf{d}}$ the vector space of polynomial systems $f = (f_1, \dots, f_n)$ with $f_i \in \mathbb{R}[X_0, \dots, X_n]$ homogeneous of degree d_i. For $f, g \in \mathcal{H}_{\mathbf{d}}$ we write

$$f_i(x) = \sum_{\alpha} a^i_\alpha X^\alpha, \quad g_i(x) = \sum_{\alpha} b^i_\alpha X^\alpha,$$

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 $X^\alpha := X_0^{\alpha_0} X_1^{\alpha_1} \cdots X_n^{\alpha_n}$. We endow the space $\mathcal{H}_{\mathbf{d}}$ with the inner

product $\langle f, g\rangle := \sum_{i=1}^{n} \langle f_i, g_i\rangle$, where

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

Hereby, the multinomial coefficients are defined as

$$\frac{d_i}{\alpha} = \frac{d_i!}{\alpha_0!\alpha_1!\cdots\alpha_n!}.$$

This inner product has the beautiful property of being invariant under the natural action of the orthogonal group $O(n+1)$ on $\mathcal{H}_{\mathbf{d}}$. In the case of one variable, it was introduced by Weyl [43]. Its use in computational mathematics goes back at least to Kostlan [19].

Throughout this section, let $\|f\|$ denote the corresponding norm of $f \in \mathcal{H}_{\mathbf{d}}$. The Weyl inner product defines a Riemannian structure on the sphere $S(\mathcal{H}_{\mathbf{d}}) := \{f \in \mathcal{H}_{\mathbf{d}} \mid \|f\| = 1\}$. As in the previous sections, we endow this sphere with the corresponding projective distance $d_{\mathbb{P}}$.

In a seminal series of papers, M. Shub and S. Smale [24, 25, 26, 28, 27] studied the problem of, given $f \in \mathcal{H}_{\mathbf{d}} \otimes_{\mathbb{R}} \mathbb{C}$, computing an approximation of a complex zero of f. They proposed an algorithm and studied its complexity in terms of, among other parameters, a condition number $\mu_{\mathrm{norm}}(f)$ for f. In the following we will recall the definition of $\mu_{\mathrm{norm}}(f)$ adapted to the case of real systems and real zeros (see [4, Chapter 12] for details).

For a simple zero $\zeta \in S^n$ of $f \in \mathcal{H}_{\mathbf{d}}$ one defines

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

where $Df(\zeta)_{|T_\zeta}$ denotes the restriction of the derivative of $f\colon \mathbb{R}^{n+1} \to \mathbb{R}^n$ at ζ to the tangent space $T_\zeta S^n = \{v \in \mathbb{R}^{n+1} \mid \langle v, \zeta\rangle = 0\}$ of S^n at ζ. (The norm on the right is the operator norm with respect to the Euclidean norm.) If ζ is not a simple root of f we set $\mu_{\mathrm{norm}}(f) = \infty$. Note that $\mu_{\mathrm{norm}}(f,\zeta)$ is homogeneous of degree 0 in f and ζ.

Shub and Smale [25] proved a condition number theorem for the condition number $\mu_{\mathrm{norm}}(f,\zeta)$ for complex polynomial systems f and complex roots ζ. To describe a corresponding result in the real situation, consider for $\zeta \in S^n$

$$\Sigma_\zeta := \{g \in \mathcal{H}_{\mathbf{d}} \mid \zeta \text{ is a multiple zero of } g\}.$$

Theorem 4.5. *For a zero $\zeta \in S^n$ of $f \in S(\mathcal{H}_{\mathbf{d}})$ we have*

$$\mu_{\mathrm{norm}}(f,\zeta) = \frac{1}{d_{\mathbb{P}}(f, \Sigma_\zeta \cap S(\mathcal{H}_{\mathbf{d}}))}.$$

Proof. The proof of [25] (see also [4, §12.4]) carries over immediately to the real situation. □

Let $\Sigma \subseteq \mathcal{H}_{\mathbf{d}}$ be the set of systems $f \in \mathcal{H}_{\mathbf{d}}$ having a real multiple zero. Note that $\Sigma_\zeta \subseteq \Sigma$. Therefore, by Theorem 4.5,

$$\mu_{\mathrm{norm}}(f,\zeta) \le \frac{1}{d_{\mathbb{P}}(f, \Sigma \cap S(\mathcal{H}_{\mathbf{d}}))}.$$

We define the condition number for real polynomial solving as the right-hand side:

$$\mu_{\mathrm{norm},\mathbb{R}}(f) := \frac{1}{d_{\mathbb{P}}(f, \Sigma \cap S(\mathcal{H}_{\mathbf{d}}))}.$$

This is by definition a conic condition number: it varies continuously with f and takes the value ∞ when $f \in \Sigma$.

Proposition 4.6. *For all $d_1, \ldots, d_n \in \mathbb{N} \setminus \{0\}$ and all $\sigma \in (0,1]$ we have*

$$\sup_{f \in S(\mathcal{H}_{\mathbf{d}})} \mathop{\mathbf{E}}_{g \in B(f,\sigma)} (\ln \mu_{\mathrm{norm},\mathbb{R}}(g)) \leq 2 \ln N + 4 \ln \mathcal{D} + 2 \ln n + 2 \ln \frac{1}{\sigma} + 6.1,$$

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

Proof. Let W be the *discriminant variety* consisting of the systems $f \in \mathcal{H}_{\mathbf{d}}$ having a complex multiple zero. Then $\Sigma \subseteq W$. In addition, it is well known that W is the zero set of a multihomogeneous polynomial of total degree bounded by $2n\mathcal{D}^2$, where $\mathcal{D} = d_1 \cdots d_n$ is the Bézout number (see, e.g., [7, Lemma 3.6]). The statement now follows immediately from Theorem 1.1. □

Remark 4.7.
(i) By the results in [24], the condition number $\mu_{\mathrm{norm}}(f)$ not only measures the maximum sensitivity of complex solutions to the input f, but it is also a crucial complexity parameter for algorithms approximating such solutions. While $\mu_{\mathrm{norm},\mathbb{R}}(f)$ shares with $\mu_{\mathrm{norm}}(f)$ the first property, it is not clear whether this is also the case for the second one.
(ii) By considering complex numbers as pairs of real numbers one can see the problem of, given $f \in \mathcal{H}_{\mathbf{d}}$ (or in $\mathcal{H}_{\mathbf{d}} \otimes \mathbb{C}$), computing an approximation of a complex zero of f (studied by Shub and Smale [24, 25, 26, 28, 27]) as a problem over the reals. Proceeding as in Section 4.3, one can find bounds for the smoothed analysis of this problem similar to those in Proposition 4.6.

4.5. **Real zero counting.** Consider the problem of, given $f \in \mathcal{H}_{\mathbf{d}}$, counting the number of real zeros of f in S^n. Unlike the problems considered so far, this is a problem with a discrete output. This means that sensitivity considerations as described in the previous problems do not apply here. Yet, finite precision algorithms will require more precision to give a reliable output when the input f is close to the set Σ of systems with multiple real zeros, and will not give any such reliable output when $f \in \Sigma$ (since in this case, arbitrarily small perturbations of f will change the output). It therefore makes sense to define the condition number for the counting problem

$$\mathsf{cond}(f) = \frac{\|f\|}{\mathsf{dist}(f, \Sigma)}.$$

Since $\mathsf{cond}(f) = \mu_{\mathrm{norm},\mathbb{R}}(f)$ for $f \in S(\mathcal{H}_{\mathbf{d}})$, the bounds of Proposition 4.6 hold for $\mathsf{cond}(f)$ as well. We note, however, that in Section 4.4 we assumed a specific inner product on $\mathcal{H}_{\mathbf{d}}$ (whose properties are crucial in the proof of Theorem 4.5). In contrast, for the smoothed analysis of $\mathsf{cond}(f)$, any inner product on $\mathcal{H}_{\mathbf{d}}$ would do.

Appendix

The constants $\mathcal{C}(p,i)$ appearing in Theorem 2.7 depend only on p and i and are independent of the manifold M. We derive the expression for $\mathcal{C}(p,i)$ stated in Theorem 2.7 by selecting a simple enough M. Consider the boundary $M_\alpha = M_\alpha(q)$ of the ball $B_R(q,\alpha)$ in S^p of radius $0 < \alpha \leq \pi/2$ centered at q (recall Example 2.6). According to Theorem 2.7 we have for $0 \leq i < p-1$

$$\mu_i(M_\alpha) = \mathcal{C}(p,i) \frac{1}{\mathcal{O}_p} \int_{S^p} \mu_i(M_\alpha(z) \cap S^{i+1}) dS^p(z).$$

Let $\rho(z)$ denote the Riemannian distance from $z \in S^p$ to S^{i+1}. If $\rho(z) < \alpha$, then the intersection $M_\alpha(z) \cap S^{i+1}$ is the boundary of a ball in S^{i+1}. The radius $\delta(z)$ of the sphere $M_\alpha(z) \cap S^{i+1}$ satisfies $\cos\alpha = \cos\rho(z) \cdot \cos\delta(z)$ by a well known formula of spherical trigonometry. From Example 2.6 we know that $\mu_i(M_\alpha(z) \cap S^{i+1}) = \mathcal{O}_i\,(\cos\delta(z))^i$. On the other hand, $M_\alpha(z)$ does not intersect S^{i+1} if $\rho(z) > \alpha$.

From these reasonings we obtain

$$\frac{1}{\mathcal{O}_p}\int_{S^p} \mu_i(M_\alpha(z)\cap S^{i+1})dS^p(z) = \frac{\mathcal{O}_i}{\mathcal{O}_p}\int_0^\alpha \left(\frac{\cos\alpha}{\cos\rho}\right)^i \frac{d}{d\rho}\,\mathsf{vol}_p T_{\mathbb{P}}(S^{i+1},\rho)\,d\rho.$$

From Lemma 2.1 and the definition of $J_{p,p-i-1}(\rho)$ it follows that

$$\frac{d}{d\rho}\,\mathsf{vol}_p T_R(S^{i+1},\rho) = \mathcal{O}_{i+1}\mathcal{O}_{p-i-2}(\sin\rho)^{p-i-2}\,(\cos\rho)^{i+1}.$$

We obtain

$$\begin{aligned}\frac{1}{\mathcal{O}_p}\int_{S^p} \mu_i(M_\alpha(z)\cap S^{i+1})dS^p(z) &= \frac{\mathcal{O}_i\mathcal{O}_{i+1}\mathcal{O}_{p-i-2}}{\mathcal{O}_p}\int_0^\alpha (\cos\alpha)^i(\sin\rho)^{p-i-2}\,\cos\rho\,d\rho.\\ &= \frac{\mathcal{O}_i\mathcal{O}_{i+1}\mathcal{O}_{p-i-2}}{\mathcal{O}_p}(\cos\alpha)^i\;\frac{(\sin\alpha)^{p-i-1}}{p-i-1}.\end{aligned}$$

On the other hand, by equation (5),

$$\mu_i(M_\alpha) = \frac{p-1}{i}\mathcal{O}_{p-1}(\sin\alpha)^{p-i-1}(\cos\alpha)^i.$$

By comparing the last two equations, the asserted form of $\mathcal{C}(p,i)$ follows. □

References

[1] C. Beltrán and L.M. Pardo. Estimates on the distribution of the condition number of singular matrices. *Found. Comput. Math.*, 7:87–134, 2007. MR2283343 (2008b:65059)

[2] A. Ben-Israel and T.N.E. Greville. *Generalized Inverses: Theory and Applications*. Springer-Verlag, 2nd edition, 2003. MR1987382 (2004b:15008)

[3] L. Blum. Lectures on a theory of computation and complexity over the reals (or an arbitrary ring). In E. Jen, editor, *Lectures in the Sciences of Complexity II*, pages 1–47. Addison-Wesley, 1990. MR1104440 (92h:68027)

[4] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and Real Computation*. Springer, 1998. MR1479636 (99a:68070)

[5] J. Bochnak, M. Coste, and M.-F. Roy. *Real algebraic geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete* (3). Springer-Verlag, Berlin, 1998. MR1659509 (2000a:14067)

[6] P. Bürgisser, F. Cucker, and M. Lotz. General formulas for the smoothed analysis of condition numbers. *C. R. Acad. Sc. Paris*, 343:145–150, 2006. MR2243310 (2007b:65147)

[7] P. Bürgisser, F. Cucker, and M. Lotz. Smoothed analysis of complex conic condition numbers. *J. Math. Pures et Appl.*, 86:293–309, 2006. MR2257845 (2007h:65040)

[8] S.L. Campbell and C.D. Meyer. *Generalized Inverse of Linear Transformations*. Pitman, 1979.

[9] S.-S. Chern. On the kinematic formula in integral geometry. *J. Math. Mech.*, 16:101–118, 1966. MR0198406 (33:6564)

[10] D. Cheung and F. Cucker. A note on level-2 condition numbers. *J. of Complexity*, 21:314–319, 2005. MR2138441 (2006a:65201)

[11] F. Cucker, H. Diao, and Y. Wei. Smoothed analysis of some condition numbers. *Numer. Lin. Alg. Appl.*, 13:71–84, 2005. MR2194973 (2006k:65114)

[12] J. Demmel. On condition numbers and the distance to the nearest ill-posed problem. *Numer. Math.*, 51:251–289, 1987. MR895087 (88i:15014)

[13] J. Demmel. The probability that a numerical analysis problem is difficult. *Math. Comp.*, 50:449–480, 1988. MR929546 (89g:65062)

[14] J. Dunagan, D.A. Spielman, and S.-H. Teng. Smoothed analysis of Renegar's condition number for linear programming. Preprint available at `http://theory.lcs.mit.edu/~spielman`, 2003.

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

[16] G. Golub and C. Van Loan. *Matrix Computations*. John Hopkins Univ. Press, 1989. MR1002570 (90d:65055)

[17] A. Gray. *Tubes*. Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA, 1990. MR1044996 (92d:53002)

[18] R. Howard. The kinematic formula in Riemannian homogeneous spaces. *Mem. Amer. Math. Soc.*, 106(509):vi+69, 1993. MR1169230 (94d:53114)

[19] E. Kostlan. On the distribution of the roots of random polynomials. In M. Hirsch, J.E. Marsden, and M. Shub, editors, *From Topology to Computation: Proceedings of the Smalefest*, pages 419–431. Springer-Verlag, 1993. MR1246137

[20] J. Milnor. On the Betti numbers of real varieties. In *Proc. AMS*, volume 15, pages 275–280, 1964. MR0161339 (28:4547)

[21] J. Renegar. On the efficiency of Newton's method in approximating all zeros of systems of complex polynomials. *Math. of Oper. Research*, 12:121–148, 1987. MR882846 (88j:65112)

[22] L. A. Santaló. *Integral geometry and geometric probability*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1976. MR0433364 (55:6340)

[23] I. R. Shafarevich. *Basic algebraic geometry*. Springer-Verlag, New York, 1974. Translated from the Russian by K. A. Hirsch, Die Grundlehren der mathematischen Wissenschaften, Band 213. MR0366917 (51:3163)

[24] M. Shub and S. Smale. Complexity of Bézout's theorem I: geometric aspects. *J. Amer. Math.*, 6:459–501, 1993. MR1175980 (93k:65045)

[25] M. Shub and S. Smale. Complexity of Bézout's theorem II: volumes and probabilities. In F. Eyssette and A. Galligo, editors, *Computational Algebraic Geometry*, volume 109 of *Progress in Mathematics*, pages 267–285. Birkhäuser, 1993. MR1230872 (94m:68086)

[26] M. Shub and S. Smale. Complexity of Bézout's theorem III: condition number and packing. *Journal of Complexity*, 9:4–14, 1993. MR1213484 (94g:65152)

[27] M. Shub and S. Smale. Complexity of Bézout's theorem V: polynomial time. *Theoretical Computer Science*, 133:141–164, 1994. MR1294430 (96d:65091)

[28] M. Shub and S. Smale. Complexity of Bézout's theorem IV: probability of success; extensions. *SIAM J. of Numer. Anal.*, 33:128–148, 1996. MR1377247 (97k:65310)

[29] S. Smale. The fundamental theorem of algebra and complexity theory. *Bull. Amer. Math. Soc.*, 4:1–36, 1981. MR590817 (83i:65044)

[30] S. Smale. Complexity theory and numerical analysis. In A. Iserles, editor, *Acta Numerica*, pages 523–551. Cambridge University Press, 1997. MR1489262 (99d:65385)

[31] D.A. Spielman and S.-H. Teng. Smoothed analysis of algorithms. In *Proceedings of the International Congress of Mathematicians*, volume I, pages 597–606, 2002. MR1989210 (2004d:90138)

[32] D.A. Spielman and S.-H. Teng. Smoothed analysis of termination of linear programming algorithms. *Math. Programm. Series B*, 97:375–404, 2003. MR2004403 (2005b:90069)

[33] D.A. Spielman and S.-H. Teng. Smoothed analysis: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004. MR2145860 (2006f:90029)

[34] D.A.. Spielman and S.-H. Teng. Smoothed analysis of algorithms and heuristics. In *Foundations of Computational Mathematics, Santander 2005*, volume 331 of *London Mathematical Society, Lecture Note Series*, pages 274–342. Cambridge University Press, 2006. MR2277110

[35] M. Spivak. *A comprehensive introduction to differential geometry. Vol. I*. Publish or Perish Inc., Wilmington, Del., second edition, 1979.

[36] M. Spivak. *A comprehensive introduction to differential geometry. Vol. III*. Publish or Perish Inc., Wilmington, Del., second edition, 1979.

[37] G.W. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, 1990. MR1061154 (92a:65017)

[38] R. Sulanke and P. Wintgen. *Differentialgeometrie und Faserbündel*. Birkhäuser Verlag, Basel, 1972. Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften, Mathematische Reihe, Band 48. MR0413153 (54:1274)

[39] J. A. Thorpe. *Elementary topics in differential geometry*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1994. MR1329997 (95m:53002)
[40] A.M. Turing. Rounding-off errors in matrix processes. *Quart. J. Mech. Appl. Math.*, 1:287–308, 1948. MR0028100 (10:405c)
[41] J. von Neumann and H.H. Goldstine. Numerical inverting matrices of high order. *Bull. Amer. Math. Soc.*, 53:1021–1099, 1947. MR0024235 (9:471b)
[42] H. Weyl. On the Volume of Tubes. *Amer. J. Math.*, 61(2):461–472, 1939. MR1507388
[43] H. Weyl. *The Theory of Groups and Quantum Mechanics*. Dover, New York, 1950.
[44] J. Wilkinson. Note on matrices with a very ill-conditioned eigenproblem. *Numer. Math.*, 19:176–178, 1972. MR0311092 (46:10188)
[45] R. Wongkew. Volumes of tubular neighbourhoods of real algebraic varieties. *Pacific J. of Mathematics*, 159:177–184, 2003. MR1211391 (94e:14073)
[46] M. Wschebor. Smoothed analysis of $\kappa(a)$. *J. of Complexity*, 20:97–107, 2004. MR2031560 (2005g:15041)

Department of Mathematics, University of Paderborn, Germany
E-mail address: pbuerg@upb.de

Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong
E-mail address: macucker@cityu.edu.hk

Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong
E-mail address: lotzm@upb.de