Adversarial Smoothed Analysis

Cucker, Felipe and Hauser, Raphael and Lotz, Martin

2010

MIMS EPrint: **2012.105**

ISSN 1749-9097

Journal of Complexity 26 (2010) 255–262

Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

Adversarial smoothed analysis

Felipe Cucker [b,*], Raphael Hauser [a], Martin Lotz [a]

[a] *Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford, OX1 3QD, United Kingdom*

[b] *City University of Hong Kong, Department of Mathematics, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong*

ARTICLE INFO

Article history:
Received 9 December 2008
Accepted 15 December 2009
Available online 11 May 2010

Keywords:
Condition numbers
Random matrices
Average case analysis
Smoothed analysis

ABSTRACT

The purpose of this note is to extend the results on uniform smoothed analysis of condition numbers from Bürgisser et al. (2008) [1] to the case where the perturbation follows a radially symmetric probability distribution. In particular, we will show that the bounds derived in [1] still hold in the case of distributions whose density has a singularity at the center of the perturbation, which we call *adversarial*.

1. Introduction

Condition numbers play a central role in numerical analysis. They occur in error analysis for finite-precision algorithms (this being historically the reason for their introduction in the late 1940s by von Neumann and Goldstine [10] and Turing [9]) as well as a parameter in expressions bounding the number of iterations in a variety of algorithms (a paradigmatic example being the conjugate gradient method [8, Theorem 38.5]). In practice, however, a difficulty appears: it would seem that to know the condition number of a given data one needs to solve the problem at hand on this data. An inconvenient circularity. A way out of it, proposed by Steve Smale (see [5] for a review), is to assume a probability measure on the space of data and to study the condition number $\mathscr{C}(a)$ at data a as a random variable. In other words, to study the condition number of random data.

In doing so Demmel [2] noticed that most condition numbers could be written as (or at least reasonably sharply bounded by) the relativized inverse of the distance from the data $a \in \mathbb{R}^{n+1}$ to a set of ill-posed instances $\Sigma \subset \mathbb{R}^{n+1}$. That is, one could write

$$\mathscr{C}(a) = \frac{\|a\|}{\mathrm{dist}(a, \Sigma)}. \tag{1.1}$$

* Corresponding author.
E-mail addresses: macucker@cityu.edu.hk, macucker@math.cityu.edu.hk (F. Cucker), hauser@comlab.ox.ac.uk (R. Hauser).

0885-064X/$ – see front matter
doi:10.1016/j.jco.2009.12.005

The simplest example of this phenomenon is given by the condition number for matrix inversion and linear equation solving. For a non-singular $k \times k$ matrix A it takes the form $\kappa(A) := \|A\|\|A^{-1}\|$, where $\|\ \|$ denotes the operator norm. The Condition Number Theorem by Eckart and Young states that $\|A^{-1}\| = d(A, \Sigma)^{-1}$, where Σ is the set of singular matrices. Then, for $n = k^2 - 1$ and $a = A$ we obtain (1.1).

In most applications, Σ is a pointed cone. Therefore, one could normalize so that a belongs to the n-dimensional unit sphere S^n. Note that the usual assumption that a has a Gaussian distribution in $\mathbb{R}^{n+1}$ yields a uniform distribution in S^n after this normalization. It is for condition numbers as in (1.1) – which we shall call *conic* – with inputs drawn from the uniform distribution on S^n that Demmel proved in [3] (shortly after [2]) a general result bounding their tail as a function of n and the degree of an algebraic hypersurface containing Σ.

Very recently, a new paradigm for probabilistic analysis was proposed by Spielman and Teng [6,7]. Called *smoothed analysis*, it consists of replacing the idea of "random data" by that of "random perturbation of a given data" and study the worst case (w.r.t. data a) of the latter. In its original formulation, and in the case of a condition number $\mathscr{C}(a)$, this amounts to study the tail

$$\sup_{a\in\mathbb{R}^{n+1}} \operatorname*{Prob}_{z\in N(a,\sigma^2)} \{\mathscr{C}(z) \geq t\}$$

or the expected value

$$\sup_{a\in\mathbb{R}^{n+1}} \mathop{\mathbf{E}}_{z\in N(a,\sigma^2)} [\ln \mathscr{C}(z)]$$

where $N(a, \sigma^2)$ is a Gaussian distribution centered at a with covariance matrix $\sigma^2\mathrm{Id}$ and σ^2 small (with respect to $\|a\|$). In [1], to obtain general results as in [3], data was again restricted to S^n and the expressions above replaced by

$$\sup_{a\in S^n} \operatorname*{Prob}_{z\in B(a,\sigma)} \{\mathscr{C}(z) \geq t\}$$

and

$$\sup_{a\in S^n} \mathop{\mathbf{E}}_{z\in B(a,\sigma)} [\ln \mathscr{C}(z)]$$

where $B(a, \sigma)$ is the open ball (that is, the spherical cap) in S^n centered at a and of radius σ, and z is drawn from a uniform distribution on this ball.

One of the claimed advantages of smoothed analysis is a smaller dependence on the underlying distribution. It follows from this claim that the replacement of Gaussian perturbations by uniform ones should not significantly affect the smoothed analysis of $\mathscr{C}(a)$. The goal of this note is to further pursue this claim by extending the main result in [1], combining it with ideas from [4], to a class of distributions we call *adversarial*. The support of such a distribution is, as in the uniform case, the ball $B(a, \sigma)$ and they are radially symmetric as well. But their density increases when approaching a and has a pole at a.

2. Preliminaries

We assume our data space is $\mathbb{R}^{n+1}$, endowed with a scalar product $\langle\ ,\ \rangle$. In all that follows we consider problems whose set of ill-posed inputs Σ is a point-symmetric cone in $\mathbb{R}^{n+1}$. That is, if $x \in \Sigma$ then $\lambda x \in \Sigma$ for all $\lambda \in \mathbb{R}$. By a *conic condition number* we understand a function $\mathscr{C}\colon \mathbb{R}^{n+1} \to [1, \infty]$ such that for all $a \in \mathbb{R}^{n+1}$ we have

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

where $\|\ \|$ and dist are the norm and distance induced by $\langle\ ,\ \rangle$. Note that for $\lambda \neq 0$ we have $\mathscr{C}(\lambda a) = \mathscr{C}(a)$. We can therefore work with the n-dimensional real projective space $\mathbb{P}^n$ as ambient space. If we also denote by $\Sigma \subset \mathbb{P}^n$ the image of the ill-posed cone in projective space, then for $a \in \mathbb{P}^n$ it follows that

$$\mathscr{C}(a) = \frac{1}{d_{\mathbb{P}}(a, \Sigma)},$$

where $d_{\mathbb{P}}(x, y) = \sin\alpha$, denotes the projective distance between $x, y \in \mathbb{P}^n$ (α being the angle between x and y).

The two-fold covering $p: S^n \to \mathbb{P}^n$ induces a measure ν on $\mathbb{P}^n$ by means of $\nu(B) := \frac{1}{2}\operatorname{Vol}_n(p^{-1}(B))$ for $B \subseteq \mathbb{P}^n$, where Vol_n is the n-dimensional volume on the sphere. Thus $\nu(\mathbb{P}^n) = \mathscr{O}_n/2$, where $\mathscr{O}_n := \operatorname{Vol}_n(S^n) = \frac{2\pi^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})}$.

For $0 < \sigma \leq 1$, we denote by $B_{\mathbb{P}}(a, \sigma)$ the open ball of projective radius σ around $a \in \mathbb{P}^n$. It is known that

$$\nu(\mathrm{B}_{\mathbb{P}}(a, \sigma)) = \mathscr{O}_{n-1} \cdot I_n(\sigma),$$

where

$$I_n(\sigma) := \int_0^\sigma \frac{r^{n-1}}{\sqrt{1 - r^2}} \mathrm{d}r. \tag{2.1}$$

The following bounds will prove useful on several occasions:

$$\frac{\sigma^n}{n} \leq I_n(\sigma) \leq \min\left\{\frac{1}{\sqrt{1 - \sigma^2}}, \sqrt{\frac{\pi n}{2}}\right\} \cdot \frac{\sigma^n}{n}. \tag{2.2}$$

For $a \in \mathbb{P}^n$ and $\sigma \in (0, 1]$, the uniform measure on $\mathrm{B}_{\mathbb{P}}(a, \sigma)$ is defined by

$$\nu_{a,\sigma}(B) = \frac{\nu(B \cap \mathrm{B}_{\mathbb{P}}(a, \sigma))}{\nu(\mathrm{B}_{\mathbb{P}}(a, \sigma))} \tag{2.3}$$

for all Borel-measurable $B \subseteq \mathbb{P}^n$.

2.1. *Uniform smoothed analysis*

A reformulation of the main result in [1] in the projective space setting can be written as follows.

Theorem 2.1. *Let $\mathscr{C}$ be a conic condition number with set of ill-posed inputs $\Sigma \subset \mathbb{P}^n$. Assume that Σ is contained in the zero set in $\mathbb{P}^n$ of homogeneous polynomials of degree at most d. Then, for all $\sigma \in (0, 1]$ and all $t \geq t_0 = (2d + 1)\frac{n}{\sigma}$,*

$$\sup_{a \in \mathbb{P}^n} \operatorname*{Prob}_{z \in B_{\mathbb{P}}(a,\sigma)} \{\mathscr{C}(z) \geq t\} \leq 13\, dn\, \frac{1}{\sigma t}.$$

and

$$\sup_{a \in \mathbb{P}^n} \operatorname*{\mathbf{E}}_{z \in B_{\mathbb{P}}(a,\sigma)} [\ln \mathscr{C}(z)] \leq 2\ln n + 2\ln d + 2\ln\frac{1}{\sigma} + 5,$$

where Prob *and* **E** *are taken with respect to $\nu_{a,\sigma}$.*

As a consequence of this result, uniform smoothed analysis results for the condition numbers of a variety of problems are obtained, including linear equation solving, Moore–Penrose inversion, eigenvalue computation and polynomial system solving. The bounds obtained are consistently of the same order of magnitude as the best bounds obtained previously by ad hoc methods.

2.2. Uniformly absolutely continuous distributions

In [4] a general boosting mechanism was developed that allows extending any probabilistic analysis of a condition number with respect to some chosen probability distribution over the input data to a more general class of distributions.

Let μ be a $\nu_{a,\sigma}$-absolutely continuous probability measure. Using the convention $\ln(0) := -\infty$ we define, for $\delta \in (0, 1)$,

$$\inf(\delta) := \inf\left\{ \frac{\ln \mu(B)}{\ln \nu_{a,\sigma}(B)} : B \text{ is Borel-measurable and } 0 < \nu_{a,\sigma}(B) \le \delta \right\}.$$

With these conventions, Theorem 2.2 of [4] shows that

$$\alpha_{\nu_{a,\sigma}}(\mu) := \lim_{\delta\to 0} \inf(\delta) \in [0, 1]. \tag{2.4}$$

Absolute continuity alone ensures that all $\nu_{a,\sigma}$-null-sets must be μ-null-sets, but this does not imply that $\mu(B)$ is small when $\nu_{a,\sigma}(B)$ is small and strictly positive. In contrast, when $\alpha_{\nu_{a,\sigma}}(\mu) > 0$ then (2.4) gives uniform upper bounds on $\mu(B)$ in terms of $\nu_{a,\sigma}(B)$. Furthermore, the smaller α gets, the larger the variation of μ in terms of $\nu_{a,\sigma}$. If μ is $\nu_{a,\sigma}$-absolutely continuous and $\alpha_{\nu_{a,\sigma}}(\mu) > 0$, we therefore say that μ is *uniformly* $\nu_{a,\sigma}$-absolutely continuous and call $\alpha_{\nu_{a,\sigma}}(\mu)$ the *smoothness parameter* of μ with respect to $\nu_{a,\sigma}$.

The following result, which easily follows from (2.4), can be used to boost bounds on tail probabilities with respect to $\nu_{a,\sigma}$ (as those in Theorem 2.1) to obtain similar bounds on any uniformly $\nu_{a,\sigma}$-absolutely continuous probability measure μ.

Proposition 2.2. *$\alpha_{\nu_{a,\sigma}}(\mu)$ is the largest nonnegative real number α for which it is true that for all $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that $\nu_{a,\sigma}(B) \le \delta_\varepsilon$ implies $\mu(B) \le \nu_{a,\sigma}(B)^{\alpha-\varepsilon}$.*

3. Smoothed analysis for adversarial distributions

In this section we present our main result, namely an extension of Theorem 2.1 to the case where we have a radially symmetric distribution whose density has a pole at the point being perturbed. We begin by introducing some notation.

Let $a \in \mathbb{P}^n$ and $\sigma \in (0, 1]$, and let $\nu_{a,\sigma}$ be the uniform measure on $\mathrm{B}_{\mathbb{P}}(a, \sigma)$, as defined in (2.3). Let μ be a $\nu_{a,\sigma}$-absolutely continuous probability measure on $\mathbb{P}^n$ with density f. In other words,

$$\mu(B) = \int_B f(x)\ \nu_{a,\sigma}(\mathrm{d}x)$$

for all events B. Assume further that $f: \mathbb{P}^n \to [0, \infty]$ is of the form $f(x) = g(d_{\mathbb{P}}(x, a))$, with a monotonically decreasing function $g: [0, \sigma] \to [0, \infty]$ of the form

$$g(r) = C_{\beta,\sigma} \cdot r^{-\beta} \cdot h(r),$$

with $\beta < n$, where $C_{\beta,\sigma} = I_n(\sigma)/I_{n-\beta}(\sigma)$ and $h: [0, \sigma] \to \mathbb{R}_+$ is a continuous function satisfying $h(0) \neq 0$ and

$$\int_0^\sigma h(r) \frac{r^{n-\beta-1}}{\sqrt{1-r^2}} \mathrm{d}r = I_{n-\beta}(\sigma),$$

so that μ is a probability measure on $B_{\mathbb{P}}(a, \sigma)$. In other words, f is radially symmetric around a with respect to $d_{\mathbb{P}}$ and has a pole of order $-\beta$ at 0 in case $\beta > 0$. The normalizing factor $C_{\beta,\sigma}$ is chosen to make $h(r) = 1$ a valid choice. Set $H := \sup_{0\le r\le\sigma} h(r)$. Note that $H \ge 1$, and that $H = 1$ implies $h \equiv 1$. Note as well that the measure μ depends on σ, β and H.

The main result of this note is the following.

Theorem 3.1. *Let $\mathscr{C}$ be a conic condition number with set of ill-posed inputs $\Sigma \subseteq \mathbb{P}^n$, and assume Σ is contained in a projective hypersurface of degree at most d. Then*

$$\mathop{\mathbf{E}}_{\mu}[\ln \mathscr{C}] \le 2\ln(n) + \ln(d) + \ln\left(\frac{1}{\sigma}\right) + \ln\left(\frac{13\pi}{2}\right) + \frac{1}{1-\frac{\beta}{n}}\left(\ln \frac{2eH^2 n}{\ln(\pi n/2)}\right).$$

This result applies to the variety of problems mentioned after Theorem 2.1. The statement of the theorem follows from calculating the smoothness parameter $\alpha_\nu(\mu)$ and the constants in Proposition 2.2. These are given by the following two lemmas, to be proven later.

Lemma 3.2. *The smoothness parameter of μ with respect to $\nu_{a,\sigma}$ is given by $\alpha_{\nu_{a,\sigma}}(\mu) = 1 - \beta/n$.*

For the statement of the next lemma, let $\varepsilon \in (0, 1-\beta/n)$, and let

$$\rho_\varepsilon := \sigma \cdot \left(\frac{1}{H}\cdot\sqrt{1-\left(\frac{2}{\pi n}\right)^{(1-\frac{\beta}{n}-\varepsilon)/(n\varepsilon)}}\right)^{\frac{1}{\varepsilon n}} \left(\sqrt{\frac{2}{\pi n}}\right)^{(1-\frac{\beta}{n}-\varepsilon)\frac{1}{\varepsilon n}}.$$

Set $\delta_\varepsilon := I_n(\rho_\varepsilon)/I_n(\sigma)$.

Lemma 3.3. *Let $B \subseteq \mathbb{P}^n$ be such that $\nu_{a,\sigma}(B) \le \delta_\varepsilon$. Then $\mu(B) \le (\nu_{a,\sigma}(B))^{1-\frac{\beta}{n}-\varepsilon}$.*

We are now ready to prove the main result.

Proof of Theorem 3.1. Setting $\varepsilon = \frac{1}{2}(1-\frac{\beta}{n})$ and using the bounds (2.2) we obtain

$$\frac{2}{\pi n}\left(\frac{1}{H}\cdot\sqrt{1-\left(\frac{2}{\pi n}\right)^{\frac{1}{n}}}\right)^{\frac{2}{1-\frac{\beta}{n}}} \le \delta_\varepsilon \le \left(\frac{1}{H}\cdot\sqrt{1-\left(\frac{2}{\pi n}\right)^{\frac{1}{n}}}\right)^{\frac{2}{1-\frac{\beta}{n}}}. \tag{3.1}$$

From Theorem 2.1 it follows that for all $t \ge t_0 := \ln[(1+2d)n/\sigma]$,

$$\mathop{\mathrm{Prob}}_{\nu_{a,\sigma}}\{\ln \mathscr{C} > t\} \le \frac{13dn}{\sigma}e^{-t}. \tag{3.2}$$

Set

$$t_\varepsilon := \ln\left(\frac{13dn}{\sigma\cdot\delta_\varepsilon}\right) = \ln\left(\frac{13dn}{\sigma}\right) + \ln(\delta_\varepsilon^{-1}).$$

Using (3.1) we obtain

$$\ln\left(13\frac{dn}{\sigma}\right) \le t_\varepsilon - \frac{2}{1-\frac{\beta}{n}}\ln\left(\frac{H}{\sqrt{1-\left(\frac{2}{\pi n}\right)^{\frac{1}{n}}}}\right) \le \ln\left(13\frac{\pi}{2}\frac{dn^2}{\sigma}\right).$$

The lower bound shows that $t_\varepsilon > t_0$, so that for all $t \ge t_\varepsilon$,

$$\nu_{a,\sigma}(\{x : \ln \mathscr{C}(x) > t\}) = \mathop{\mathrm{Prob}}_{\nu_{a,\sigma}}\{\ln \mathscr{C} > t\} \le \frac{13dn}{\sigma}e^{-t} \le \delta_\varepsilon.$$

Applying Lemma 3.3, it follows that for $t \ge t_\varepsilon$,

$$\mathop{\mathrm{Prob}}_{\mu}\{\ln \mathscr{C} > t\} = \mu(\{x : \ln \mathscr{C}(x) > t\}) \le \left(\frac{13dn}{\sigma}e^{-t}\right)^{\frac{1}{2}(1-\frac{\beta}{n})},$$

and hence,

$$\begin{aligned}\mathop{\mathbf{E}}_{\mu}[\ln\mathscr{C}] &= \int_0^\infty \mathop{\mathrm{Prob}}_{\mu}\{\ln\mathscr{C} > t\}\mathrm{d}t\\ &\le \int_0^{t_\varepsilon} 1\mathrm{d}t + \int_{t_\varepsilon}^\infty \left(\frac{13dn}{\sigma}e^{-t}\right)^{\frac{1}{2}(1-\frac{\beta}{n})}\mathrm{d}t\\ &= t_\varepsilon + \frac{2\delta_\varepsilon^{\frac{1}{2}(1-\frac{\beta}{n})}}{1-\frac{\beta}{n}}.\end{aligned}$$

Using the bounds on t_ε and δ_ε we get

$$\begin{aligned}\mathop{\mathbf{E}}_{\mu}[\ln\mathscr{C}] &\le 2\ln(n) + \ln(d) + \ln\left(\frac{1}{\sigma}\right) + \ln\left(\frac{13\pi}{2}\right)\\ &\quad + \frac{2}{1-\frac{\beta}{n}}\left(\ln\left(\frac{H}{\sqrt{1-\left(\frac{2}{\pi n}\right)^{\frac{1}{n}}}}\right) + \frac{\sqrt{1-\left(\frac{2}{\pi n}\right)^{\frac{1}{n}}}}{H}\right).\end{aligned}$$

A small calculation shows that $\left(1-\left(\frac{2}{\pi n}\right)^{\frac{1}{n}}\right)^{-1/2} \le \sqrt{\frac{2n}{\ln(\pi n/2)}}$. This completes the proof. $\square$

3.1. Proofs of Lemmas 3.2 and 3.3

The content of the following lemma, needed for calculating the smoothness parameter, should be intuitively clear.

Lemma 3.4. *Let $0 < \delta < 1$. Then among all measurable sets $B \subseteq \mathrm{B}_{\mathbb{P}}(a,\sigma)$ with $0 < \nu_{a,\sigma}(B) \le \delta$, the quantity $\mu(B)$ is maximized by $\mathrm{B}_{\mathbb{P}}(a,\rho)$ where $\rho \in (0,\sigma)$ is chosen so that $\nu_{a,\sigma}(\mathrm{B}_{\mathbb{P}}(a,\rho)) = \delta$.*

Proof. It clearly suffices to show that

$$\int_B f(x)\ \nu_{a,\sigma}(\mathrm{d}x) \le \int_{\mathrm{B}_{\mathbb{P}}(a,\rho)} f(x)\ \nu_{a,\sigma}(\mathrm{d}x)$$

for all Borel sets $B \subset \mathrm{B}_{\mathbb{P}}(a,\sigma)$ such that $\nu_{a,\sigma}(B) = \delta$. Indeed, we have

$$\begin{aligned}\int_B f(x)\ \nu_{a,\sigma}(\mathrm{d}x) &= \int_{B\cap \mathrm{B}_{\mathbb{P}}(a,\rho)} f(x)\ \nu_{a,\sigma}(\mathrm{d}x) + \int_{B\setminus \mathrm{B}_{\mathbb{P}}(a,\rho)} f(x)\ \nu_{a,\sigma}(\mathrm{d}x)\\ &\le \int_{B\cap \mathrm{B}_{\mathbb{P}}(a,\rho)} f(x)\ \nu_{a,\sigma}(\mathrm{d}x) + g(\rho)\ \nu_{a,\sigma}(B\setminus \mathrm{B}_{\mathbb{P}}(a,\rho))\\ &= \int_{B\cap \mathrm{B}_{\mathbb{P}}(a,\rho)} f(x)\ \nu_{a,\sigma}(\mathrm{d}x) + g(\rho)\ \nu_{a,\sigma}(\mathrm{B}_{\mathbb{P}}(a,\rho)\setminus B)\\ &\le \int_{B\cap \mathrm{B}_{\mathbb{P}}(a,\rho)} f(x)\ \nu_{a,\sigma}(\mathrm{d}x) + \int_{\mathrm{B}_{\mathbb{P}}(a,\rho)\setminus B} f(x)\ \nu_{a,\sigma}(\mathrm{d}x)\\ &= \int_{\mathrm{B}_{\mathbb{P}}(a,\rho)} f(x)\ \nu_{a,\sigma}(\mathrm{d}x), \end{aligned} \tag{3.3}$$

where we have used $\nu_{a,\sigma}(\mathrm{B}_{\mathbb{P}}(a,\rho)) = \delta = \nu_{a,\sigma}(B)$ in (3.3). This proves our claim. $\square$

Even though ρ is a function of δ, we will not reflect this notationally in what follows.

It will be important to have expressions for $\nu_{a,\sigma}(B)$ and $\mu(B)$ when $B = B_{\mathbb{P}}(a,\rho)$ is a projective ball. In this situation we have

$$\begin{aligned}
\mu(B_{\mathbb{P}}(a,\rho)) &= \frac{1}{\nu(B_{\mathbb{P}}(a,\sigma))}\int_{B_{\mathbb{P}}(a,\rho)} f(x)\ \nu(\mathrm{d}x) \\
&= \frac{1}{\mathscr{O}_{n-1}I_n(\sigma)}\cdot C_{\beta,\sigma}\cdot\mathscr{O}_{n-1}\int_0^\rho r^{-\beta}h(r)\frac{r^{n-1}}{\sqrt{1-r^2}}\mathrm{d}r \\
&= \frac{1}{I_{n-\beta}(\sigma)}\int_0^\rho h(r)\frac{r^{n-\beta-1}}{\sqrt{1-r^2}}\mathrm{d}r \\
&\le \left(\sup_{0\le r\le\rho} h(r)\right)\cdot\frac{I_{n-\beta}(\rho)}{I_{n-\beta}(\sigma)}.
\end{aligned} \tag{3.4}$$

Similarly,

$$\mu(B_{\mathbb{P}}(a,\rho)) \ge \left(\inf_{0\le r\le\rho} h(r)\right)\cdot\frac{I_{n-\beta}(\rho)}{I_{n-\beta}(\sigma)}.$$

In particular,

$$\nu_{a,\sigma}(B_{\mathbb{P}}(a,\rho)) = \frac{I_n(\rho)}{I_n(\sigma)}. \tag{3.5}$$

Proof of Lemma 3.2. From (3.4), (3.5) and (2.2) we get the bounds of the form

$$\frac{1}{C_1}\cdot\rho^n \le \nu_{a,\sigma}(\mathrm{B}_{\mathbb{P}}(a,\rho)) \le C_1\cdot\rho^n, \tag{3.6}$$

$$\inf_{0\le r\le\rho} h(r)\cdot\frac{1}{C_2}\cdot\rho^{n-\beta} \le \mu(\mathrm{B}_{\mathbb{P}}(a,\rho)) \le \sup_{0\le r\le\rho} h(r)\cdot C_2\cdot\rho^{n-\beta}, \tag{3.7}$$

where the constants C_i do not depend on ρ.

We thus have (using Lemma 3.4)

$$\begin{aligned}
\alpha_{\nu_{a,\sigma}}(\mu) &= \lim_{\delta\to 0}\inf\left\{\frac{\ln\mu(B)}{\ln\nu_{a,\sigma}(B)} : B \text{ measurable},\ 0 < \nu_{a,\sigma}(B)\le\delta\right\} \\
&= \lim_{\rho\to 0}\frac{\ln\mu(\mathrm{B}_{\mathbb{P}}(a,\rho))}{\ln\nu_{a,\sigma}(\mathrm{B}_{\mathbb{P}}(a,\rho))}\begin{cases}\le \lim_{\rho\to 0}\dfrac{\ln(\inf h(r)/C_2)+(n-\beta)\ln\rho}{\ln(C_1)+n\ln\rho} = 1-\dfrac{\beta}{n} \\[2ex] \ge \lim_{\rho\to 0}\dfrac{\ln(C_2\cdot\sup h(r))+(n-\beta)\ln\rho}{-\ln C_1+n\ln\rho} = 1-\dfrac{\beta}{n}.\end{cases}
\end{aligned}$$

This concludes the proof. □

Proof of Lemma 3.3. Since sets of the form $\mathrm{B}_{\mathbb{P}}(a,\rho)$ maximize $\mu(B)$ among all measurable sets $B\subseteq \mathrm{B}_{\mathbb{P}}(a,\sigma)$ such that $\nu_{a,\sigma}(B)\le\delta$ for any δ, we may w.l.o.g. assume $B = B_{\mathbb{P}}(a,\rho)$. By (3.4) and (3.5) our task amounts to showing

$$H\cdot\frac{I_{n-\beta}(\rho)}{I_{n-\beta}(\sigma)} \le \left(\frac{I_n(\rho)}{I_n(\sigma)}\right)^{1-\frac{\beta}{n}-\varepsilon}$$

for $\rho\le\rho_\varepsilon$. And indeed, using the bounds (2.2), we get

$$\begin{aligned}
H\cdot\frac{I_{n-\beta}(\rho)}{I_{n-\beta}(\sigma)} &\le H\frac{1}{\sqrt{1-\rho^2}}\cdot\left(\frac{\rho}{\sigma}\right)^{n-\beta} \\
&\le H\frac{1}{\sqrt{1-\rho^2}}\cdot\left(\left(\frac{\rho}{\sigma}\right)^n\right)^{1-\frac{\beta}{n}-\varepsilon}\left(\frac{\rho_\varepsilon}{\sigma}\right)^{\varepsilon n}
\end{aligned}$$

$$\leq \frac{\sqrt{1-\left(\frac{2}{\pi n}\right)^{(1-\frac{\beta}{n}-\varepsilon)/(n\varepsilon)}}}{\sqrt{1-\rho^2}} \cdot \left(\sqrt{\frac{2}{\pi n}}\left(\frac{\rho}{\sigma}\right)^n\right)^{1-\frac{\beta}{n}-\varepsilon}$$

$$\leq \frac{\sqrt{1-\left(\frac{2}{\pi n}\right)^{(1-\frac{\beta}{n}-\varepsilon)/(n\varepsilon)}}}{\sqrt{1-\rho^2}} \cdot \left(\frac{I_n(\rho)}{I_n(\sigma)}\right)^{1-\frac{\beta}{n}-\varepsilon},$$

where for the last inequality we use the bounds (2.2) again. Moreover, we have

$$\rho \leq \rho_\varepsilon \leq \left(\sqrt{\frac{2}{\pi n}}\right)^{(1-\frac{\beta}{n}-\varepsilon)\frac{1}{\varepsilon n}}.$$

Therefore, $\sqrt{1-\left(\frac{2}{\pi n}\right)^{(1-\frac{\beta}{n}-\varepsilon)\frac{1}{\varepsilon n}}} \leq \sqrt{1-\rho^2}$, completing the proof. □

Acknowledgment

The first author was partially supported by the GRF grant CityU 100808.

References

[1] P. Bürgisser, F. Cucker, M. Lotz, The probability that a slightly perturbed numerical analysis problem is difficult, Math. Comp. 77 (2008) 1559–1583.
[2] J. Demmel, On condition numbers and the distance to the nearest ill-posed problem, Numer. Math. 51 (1987) 251–289.
[3] J. Demmel, The probability that a numerical analysis problem is difficult, Math. Comp. 50 (1988) 449–480.
[4] R. Hauser, T. Müller, Conditioning of random conic systems under a general family of input distributions, Found. Comput. Math. 9 (2009) 335–358.
[5] S. Smale, Complexity theory and numerical analysis, in: A. Iserles (Ed.), Acta Numerica, Cambridge University Press, 1997, pp. 523–551.
[6] D.A. Spielman, S.-H. Teng, Smoothed analysis of algorithms, in: Proceedings of the International Congress of Mathematicians, vol. I, 2002, pp. 597–606.
[7] D.A. Spielman, S.-H. Teng, Smoothed analysis of algorithms and heuristics, in: Foundations of Computational Mathematics, Santander 2005, in: Lecture Notes of the London Mathematical Society, vol. 331, 2006, pp. 274–342.
[8] L.N. Trefethen, D. Bau III, Numerical Linear Algebra, SIAM, 1997.
[9] A.M. Turing, Rounding-off errors in matrix processes, Quart. J. Mech. Appl. Math. 1 (1948) 287–308.
[10] J. von Neumann, H.H. Goldstine, Numerical inverting matrices of high order, Bull. Amer. Math. Soc. 53 (1947) 1021–1099.