CONDITIONING OF RANDOM CONIC SYSTEMS UNDER A GENERAL FAMILY OF INPUT DISTRIBUTIONS

RAPHAEL HAUSER* AND TOBIAS MÜLLER†

Communicated by Felipe Cucker and Mike Todd.

Abstract. We consider the conic feasibility problem associated with the linear homogeneous system $Ax \leq 0$, $x \neq 0$. The complexity of iterative algorithms for solving this problem depends on a condition number $\mathscr{C}(A)$. When studying the typical behaviour of algorithms under stochastic input one is therefore naturally led to investigate the fatness of the tails of the distribution of $\mathscr{C}(A)$. Introducing the very general class of *uniformly absolutely continuous* probability models for the random matrix A, we show that the distribution tails of $\mathscr{C}(A)$ decrease at algebraic rates, both for the Goffin-Cheung-Cucker number $\mathscr{C}_G$ and the Renegar number $\mathscr{C}_R$. The exponent that drives the decay arises naturally in the theory of *uniform absolute continuity*, which we also develop in this paper. In the case of $\mathscr{C}_G$ we also discuss lower bounds on the tail probabilities and show that there exist absolutely continuous input models for which the tail decay is subalgebraic.

AMS subject classifications. Primary 90C31, 15A52; secondary 90C05, 90C60, 62H10.

Key words. Condition numbers, random matrices, linear programming, conic feasibility problem, probabilistic analysis, smoothed analysis.

1. Introduction.

Any matrix $A \in \mathbb{R}^{n\times m}$ defines a pair of linear systems

$$\begin{aligned} &(\mathrm{D}(A)) \qquad Ax \leq 0,\ x \neq 0 \\ &(\mathrm{P}(A)) \qquad A^{\mathrm{T}}y = 0,\ y \geq 0,\ y \neq 0. \end{aligned}$$

If A has full column rank then one of these systems has a strict solution if and only if the other has no solution at all. In other words, the existence of x such that $Ax < 0$ yields a certificate of infeasibility of $(\mathrm{P}(A))$, and conversely, the existence of y such that $A^{\mathrm{T}}y = 0$, $y > 0$ proves the infeasibility of $(\mathrm{D}(A))$. In the first case we say that A is *strictly feasible*, and in the second case that A is *strictly infeasible*. If neither case occurs we say that A is *ill-posed*. In this case both $(\mathrm{D}(A))$ and $(\mathrm{P}(A))$ have solutions but none that will satisfy all inequalities strictly.

The conic feasibility problem (CFP) associated with A is to decide which of $(\mathrm{D}(A))$ or $(\mathrm{P}(A))$ is strictly feasible and to compute a solution for it. When A is ill-posed, then conic feasibility algorithms will usually fail, unless preprocessing is used to restrict the problem to a subspace where it is well-posed.

It is well known that from a complexity theoretic point of view linear programming is equivalent to the conic feasibility problem defined by a matrix A with special structure. As a consequence, the conic feasibility problem was studied extensively in the LP literature, where ellipsoid and interior-point methods have been established as polynomial-time algorithms under the rational number *Turing machine model*. In

*Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford, OX1 3QD, United Kingdom, (hauser@comlab.ox.ac.uk). This author was supported by a grant of the Nuffield Foundation under the "Newly Appointed Lecturers" grant scheme, (project number NAL/00720/G) and through grant GR/S34472 from the Engineering and Physical Sciences Research Council of the UK.

†EURANDOM, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, (t.muller@tue.nl). The research in this paper was conducted while this author was a research student at the University of Oxford. He was partially supported by EPSRC, the Oxford University Department of Statistics, Bekker-la-Bastide fonds, Dr. Hendrik Muller's Vaderlandsch fonds, and Prins Bernhard Cultuurfonds

the *real number Turing machine model* of Blum et al. [3] the complexity of the same algorithms is typically of order

$$\mathscr{O}\big(\log \mathscr{C}(A)\big) \cdot q(m,n), \tag{1.1}$$

where q is a polynomial and $\mathscr{C}(A)$ is either Renegar's condition number $\mathscr{C}_R(A)$ [17], the Goffin-Cheung-Cucker condition number $\mathscr{C}_G(A)$ [12, 13, 5] or another related concept of geometric measure. See Renegar [17] for an extensive discussion and examples of condition-number based complexity analyses. The complexity of the CFP is also known to be polynomial as a function of $\log \mathscr{C}(A)$ under the *finite-precision complexity model*: Cucker-Peña [7] developed a complexity bound of $\mathscr{O}\big((\log \mathscr{C}(A))^3\big)$ for an interior-point method under this complexity model.

To avoid confusing the reader, we should point out that in the classical Turing machine model the complexity of CFP is also of order $\mathscr{O}\big(\log \mathscr{C}(A)\big)$, but $\log \mathscr{C}(A)$ is polynomially bounded in the size of the input data of A when the data are rational numbers. Thus, the dependence on $\mathscr{C}(A)$ under the real-number or finite-precision complexity models replaces the dependence on the input size of the problem data under the rational-number complexity model. Indeed, no a priori upper bound on the complexity of interior point and ellipsoid methods is known that does not inherently depend on a condition number. This is in stark contrast to the simplex method, where the real Turing machine complexity is bounded by the exponential of a function of the problem dimension.

Conic feasibility also appears in the machine learning literature but with the important difference that in this context the focus is entirely on (D(A)): The problem is now to produce a solution that lies deeply inside the feasible region when A is strictly feasible, or a solution that is "as feasible as possible" in an appropriate sense when A is not strictly feasible. Certificates of infeasibility play a lesser role in this context, and certain popular algorithms cannot actually produce these. Despite this difference, the complexity of these algorithms typically also depends on $\mathscr{C}(A)$ when A is feasible. For example, the improved perceptron algorithm of Dunagan and Vempala [10] has a $\mathscr{O}(\mathscr{C}_G(A))$ probabilistic complexity, where $\mathscr{C}_G(A)$ is the Goffin-Cheung-Cucker condition number, and the classical perceptron algorithm of Rosenblatt [18] has a $\mathscr{O}(\mathscr{C}_G(A)^2)$ deterministic complexity.

1.1. Scope of this paper. In this paper we introduce and study the notion of *uniformly absolutely continuous* distributions μ with respect to a reference measure ν. Associated with such measures is a notion of smoothness parameter α which guarantees that $\mu(B)$ is quantifiably small when $\nu(B)$ is small. A measure with smoothness parameter $\alpha \in (0, 1]$ can be seen as having a Radon-Nikodym density that is essentially bounded by a function that has a delocalised singularity of degree $1 - k\alpha$, where k is the dimension of the space in which it lives. Some of the relevant theory - which seems to be new - is developed in the technical report [14] of which this paper is a shortened version. Several examples of such distributions are discussed in Section 2.3. Applying this theory to random matrices A of fixed dimension, we find that the distribution tail of the random variables $\mathscr{C}_G(A)$ and $\mathscr{C}_R(A)$ respectively decay as follows,

$$\mathrm{P}[\mathscr{C}(A) > t] = \tilde{\mathscr{O}}(t^{-\alpha}), \tag{1.2}$$

where $f(t) = \tilde{\mathscr{O}}(t^{-\alpha})$ means that $f(t) = \mathscr{O}(t^{-\alpha+\varepsilon})$ for all $\varepsilon > 0$. Together with complexity estimates of the form (1.1) the bound (1.2) implies that when random CFP input data A are used in an interior-point or ellipsoid algorithm, then the (random) number of operations Op required on a real Turing machine (see Blum et al. [3]) has exponentially bounded tail event probabilities

$$\mathrm{P}[\mathrm{Op} > t] < \exp\left(-(\alpha - \varepsilon)t\right) \tag{1.3}$$

for large enough t. This bound implies that all moments of the random variable Op are finite, when $\alpha > 0$. More importantly, (1.3) provides an explanation for why long running times are not observed in practice, see Section 1.3 below. Since $\mathscr{C}_G(A)$ is invariant under row scalings of A whereas $\mathscr{C}_R(A)$ is not, we split the discussion of tail bounds for the distributions of these random variables into two separate sections, Section 3 and 4. Basic bounds of type (1.2) are obtained by "boosting" similar bounds obtained by Cheung-Cucker-Hauser [6] for the case of matrices with i.i.d. uniform unit row vectors to general continuous distributions of A, see Theorem 3.1. In the case of $\mathscr{C}_G$ we take the analysis a step deeper on several fronts: In Section 3.1 we tighten the tail decay bounds using a more technically involved approach, in Section 3.2 we derive a *lower* bound on the tail decay, and we use this to to show that there exist absolutely continuous distributions with $\alpha = 0$ for which $\mathrm{P}[\mathscr{C}_G(A) > t]$ does not decay at any algebraic rate.

1.2. Existing Literature. The problem studied in this paper has its roots in the study of *average case complexity analysis* of linear programming, which was first studied in the context of the simplex algorithm (see e.g. Borgwardt [4], and Smale [20]). This line of research was later extended to interior-point methods (see e.g. Huhn-Borgwardt [15, 16]), Anstreicher et al. [1], and Todd et al. [22]). The last paper made the transition to studying the average case complexity of LP via a probabilistic analysis of condition numbers. All remaining cited authors followed this approach and derived decay results of the form (1.2): Cucker-Wschebor [8] considered the case where A is a matrix with i.i.d. Gaussian entries. Cheung-Cucker-Hauser analysed the case where A has i.i.d. rows that are uniformly distributed on the unit sphere. Dunagan-Spielman-Teng [9] studied the case where A is a random perturbation of a given matrix with i.i.d. Gaussian perturbation terms $\sim \mathscr{N}(0, \sigma^2)$. By letting $\sigma^2 \to 0$, this framework allows to study the transition between average case and worst case analyses. Hence, the new terminology of *smoothed analysis* has been introduced for studies of this kind, a framework that has also been applied to the complexity analysis of the simplex and perceptron algorithms (see Spielman-Teng [21] and Blum-Dunagan [2] respectively), as well as in other contexts. We also remark that since $\mathscr{C}_G(A)$ is independent of row scaling, the framework of study of [9] becomes the same as [8], [6] when $\sigma^2 \to \infty$.

In the mentioned papers [8, 9] and [6] the distributions of the random matrix A have smoothness parameter 1, so that the tails of $\mathscr{C}(A)$ decay like $\tilde{\mathscr{O}}(t^{-1})$. In fact, all of these distributions fall under the framework of Example 1 of Section 2.3. The focus in these papers is therefore on developing good bounds on the constants appearing in the $\tilde{\mathscr{O}}$ terms as functions of the matrix dimensions (n, m) and - in the case of smoothed analysis - on the variance σ^2 of the perturbation terms. In all of these cases the constants in question can be seen as functions of the distribution of A itself, but since these distributions are fully characterised by the parameters (n, m) and - in the last case - σ^2, the dependence is expressed through the latter. The scope of the present paper is more general and fundamental: Considering random matrices with distribution density functions that are not essentially bounded, our goal is to characterise the correct exponent in the $\tilde{\mathscr{O}}$ term of the tail decay formula (1.2). The hidden multiplicative constants remain functions of the distribution of A, but because of the generality of the family of distributions studied we limit ourselves to establishing the existence of such constants.

1.3. Complexity Theoretic Contribution. Bounds of the type (1.3) contribute to the complexity theory of the CFP by providing an explanation for why long running times are empirically nonexistent: They are exponentially rare!

The existing literature on the complexity of randomised problem instances (of CFP or other computational problems) often focuses on proving that the complexity is low *on average*. For instance, in the above mentioned papers [6, 8, 9] it is shown that under the probability models considered, the expected running time of interior point and ellipsoid methods is bounded by a polynomial in the problem dimensions (n, m). Such results are referred to as *strong polynomiality on average*.

However, tail probability estimates are a more useful measure of probabilistic complexity than averages, as they contain more information: A bound of the form $\mathrm{P}[\mathrm{Op} > t] \leq \exp(-\alpha t)$ for $t \geq t_0$ immediately implies

$$\mathrm{E}[\mathrm{Op}] = \int_0^\infty \mathrm{P}[\mathrm{Op} > t]dt \leq t_0 + \frac{1}{\alpha}\exp(-\alpha t_0). \tag{1.4}$$

In particular, if t_0 and α are polynomially bounded in the problem dimension, (1.4) implies strong polynomiality on average. On the other hand, strong polynomiality on average or $\mathrm{E}[\mathrm{Op}] < \infty$ does not imply an exponentially decreasing bound on the tail probabilities. As mentioned above, the generality of input distributions considered in this paper does not allow for explicit expressions of t_0 and α as functions of the problem dimension, but the fact that $\mathrm{P}[\mathrm{Op} > t]$ decreases exponentially is very significant, as it implies that values of Op much larger than t_0 are not observed empirically, i.e., Op is de facto bounded, while the same could not be said if the probability tails decreased algebraically.

2. Preliminaries. We denote the probability measure defined by the distribution of any random variable or vector X on its image space by $\mathscr{L}(X)$. Following the usual practice, we use upper case letters for random variables and vectors, and lower case letters for deterministic variables wherever possible. When two random vectors X and Y have identical distribution, we write $X \stackrel{\mathscr{D}}{=} Y$. Inner products are denoted by $\langle \cdot, \cdot \rangle$, whereas $\cdot$ denotes a scalar multiplication and is used in places where it improves the readability of formulae. $\mathscr{B}$ always denotes a completed Borel σ-algebra. The topological space it resides in is usually clear from the context. In particular, the uniform measure $\mathscr{U}_{m-1} := \mathscr{U}(\mathrm{S}^{m-1})$ on the unit sphere $\mathrm{S}^{m-1} \subset \mathbb{R}^m$ is a Borel measure under the usual topology. Writing

$$\mathscr{I}_k(\rho) := \int_0^\rho \sin^k \tau d\tau,$$

it can easily be established that $\mathscr{U}_{m-1}\big(\mathrm{cap}(p, \rho)\big) = \mathscr{I}_{m-2}(\rho)/\mathscr{I}_{m-2}(\pi)$, where $\mathrm{cap}(p, \rho)$ is the circular cap with half opening angle ρ centered at p. Throughout the paper, we denote the open ball of radius r in $\mathbb{R}^m$ centered at the origin by B_r, and we consider $\arccos(t)$ to be a function from $[-1, 1]$ into $[0, \pi]$, so that both cos and arccos are decreasing.

2.1. Condition Numbers for the CFP. In this section we will briefly summarise the definitions and essential properties of two of the standard condition numbers that appear in the LP and CFP literature.

Let $A_{i,:}$ be the ith row of A. Then the Cheung-Cucker [5] condition number is defined as

$$\mathscr{C}_G(A) := \left| \min_{x \neq 0} \max_i \frac{\langle A_{i,:}, x \rangle}{\|A_{i,:}\| \cdot \|x\|} \right|^{-1}. \tag{2.1}$$

(Here we interpret $1/0 = \infty$.) $\mathscr{C}_G(A)$ is a generalisation of Goffin's condition number [12, 13] which was defined for strictly feasible A only. The following is straightforward to see:

$$\min_{x \neq 0} \max_i \frac{\langle A_{i,:}, x \rangle}{\|A_{i,:}\| \cdot \|x\|} \begin{cases} < 0 & \Leftrightarrow A \text{ strictly feasible,} \\ = 0 & \Leftrightarrow A \text{ ill posed,} \\ > 0 & \Leftrightarrow A \text{ infeasible.} \end{cases}$$

Note that $\mathscr{C}_G(A)$ is invariant under positive scaling of the rows of A. Hence, when studying $\mathscr{C}_G(A)$ we may assume without loss of generality that all rows of A have been scaled to unit length.

The second condition number we consider relates ill-conditioning to a notion of distance to ill-posedness. Recall from the introduction that if the matrix A is well-posed then (P(A)) has a strict solution if and only if (D(A)) has no nontrivial solution and vice versa. Writing $\|A\|_{1,\infty} := \sup_{x \in \mathbb{R}^n \setminus \{0\}} \|Ax\|_\infty / \|x\|_1$, we define

$$\varrho_P(A) := \inf \left\{ \|\Delta A\|_{1,\infty} : \ (\mathrm{P}(A + \Delta A)) \text{ is infeasible} \right\},$$
$$\varrho_D(A) := \inf \left\{ \|\Delta A\|_{1,\infty} : \ (\mathrm{D}(A + \Delta A)) \text{ is infeasible} \right\}.$$

Then $\varrho(A) := \max \{\varrho_P(A), \varrho_D(A)\}$ yields a notion of how far A is located from the set of ill-posed matrices, or by how much A can be perturbed before it switches from feasible to infeasible or vice versa. Renegar's condition number [17] is defined as the inverse relative distance to ill-posedness

$$\mathscr{C}_R(A) := \frac{\|A^{\mathrm{T}}\|_{1,\infty}}{\varrho(A)}.$$

An important difference between the two condition numbers introduced above is that unlike $\mathscr{C}_G(A)$, $\mathscr{C}_R(A)$ *does* depend on the scaling of the rows of A. Cheung-Cucker [5] established the following inequalities linking the two condition numbers,

$$\frac{1}{\sqrt{m}} \cdot \mathscr{C}_G(A) \leq \mathscr{C}_R(A) \leq \frac{\|A\|_2}{\min_i \|A_{i,:}\|_2} \cdot \mathscr{C}_G(A). \tag{2.2}$$

2.2. Uniformly Absolutely Continuous Distributions. In this section we will build up the measure-theoretic tools that are necessary to conduct the tail decay analysis of Sections 3 and 4. Let $\mathscr{B}$ be the Borel σ-algebra of a sigma-compact Hausdorff space $(E, \mathcal{O})$ whose topology has a locally countable basis, and let $\nu \neq 0$ be a sigma-finite atom-free measure on $\mathscr{B}$. For simplicity, the reader may keep in mind the example that will play a role in later sections in which E is chosen as the $(m-1)$-dimensional unit sphere S^{m-1} endowed with the subspace topology inherited from $\mathbb{R}^m$, and ν is chosen as the uniform measure or the Hausdorff measure. The following lemma follows easily from standard results in measure theory, see e.g. Federer [11], Theorem 2.2.2. For a self-contained proof, see the technical report [14].

LEMMA 2.1. *For any $\mathscr{B}$-measurable set A with $\nu(A) > 0$ and for any $\delta > 0$ there exists $B \subseteq A$ such that* $0 < \nu(B) \leq \delta$.

Next, let μ be a ν-absolutely continuous probability measure on $\mathscr{B}$. In other words, the assumption is that $\nu(B) = 0$ implies $\mu(B) = 0$ for all $\mathscr{B}$-measurable B. By the Radon-Nikodym Theorem this is equivalent to the existence of a $\mathscr{B}$-measurable density function $f : E \to \mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$ such that $\mu(B) = \int_B f d\nu$ for all $\mathscr{B}$-measurable sets B. In what follows we will use the convention $\ln(0) := -\infty$.

THEOREM 2.2. *For measures μ and ν as above, let*

$$\inf(\delta) := \inf \left\{ \frac{\ln \mu(B)}{\ln \nu(B)} : B \text{ is } \mathscr{B}\text{-measurable and } 0 < \nu(B) \leq \delta \right\}$$

for all $\delta \in (0, 1)$. Then $\alpha_\nu(\mu) := \lim_{\delta \to 0} \inf(\delta) \in [0, 1]$ is well defined.

Proof. Since μ is a probability measure, $\ln \mu(B) \leq 0$ for any $\mathscr{B}$-measurable B, giving $\inf(\delta) \geq 0$ for all $\delta \in (0, 1)$. Furthermore, since $\inf(\delta)$ is increasing as δ decreases, we have $\lim_{\delta \to 0} \inf(\delta) = \sup_{\delta \in (0,1)} \inf(\delta)$. It only remains to show that the right-hand side of this equation is bounded by 1. Since $\int_E f d\nu = 1$, there exists some constant $c \in (0, 1)$ such that $\nu(\{f > c\}) > 0$. By Lemma 2.1, for all $\delta > 0$ there exists a $\mathscr{B}$-measurable set $B_\delta \subseteq \{f > c\}$ such that $0 < \nu(B_\delta) \leq \delta$. Finally, since $\mu(B_\delta) = \int_{B_\delta} f d\nu > c\nu(B_\delta)$, we have

$$\inf(\delta) \leq \frac{\ln \mu(B_\delta)}{\ln \nu(B_\delta)} < 1 + \frac{\ln(c)}{\ln(\delta)},$$

from where the claim follows by letting $\delta \to 0$. □

The following proposition is easy to prove and illustrates the meaning of $\alpha_\nu(\mu)$.

PROPOSITION 2.3. *The following statements are equivalent:*

i) $\alpha = \alpha_\nu(\mu)$,

ii) *α is the smallest nonnegative real number for which it is true that for all $\varepsilon > 0$ and $\delta > 0$ there exists a $\mathscr{B}$-measurable set $B_{\varepsilon,\delta}$ such that $0 < \nu(B_{\varepsilon,\delta}) \leq \delta$, yet $\nu(B_{\varepsilon,\delta})^{\alpha+\varepsilon} \leq \mu(B_{\varepsilon,\delta})$,*

iii) *α is the largest nonnegative real number for which it is true that for all $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that $\nu(B) \leq \delta_\varepsilon$ implies $\mu(B) \leq \nu(B)^{\alpha-\varepsilon}$,*

iv) *α is the largest nonnegative real number for which it is true that for all $\varepsilon \in (0, \alpha)$ there exists $c_\varepsilon > 0$ such that $\mu(B) \leq c_\varepsilon \cdot \nu(B)^{\alpha-\varepsilon}$ for all $\mathscr{B}$-measurable B.*

A further characterisation of α is given by the following result, in which we use the convention $-\infty/-\infty := 0/0 := 1$. The proof is relatively straightforward and is left to the reader, but can also be found in the technical report [14].

PROPOSITION 2.4. $\alpha_\nu(\mu) = \liminf_{t \to \infty} \ln(\mu(\{f > t\}))/\ln(\nu(\{f > t\}))$.

Absolute continuity alone says that all ν-null-sets must be μ-null-sets, but this does not imply that $\mu(B)$ is small when $\nu(B)$ is small and strictly positive. In contrast, when $\alpha_\nu(\mu) > 0$ then Proposition 2.3 gives uniform upper bounds on $\mu(B)$ in terms of $\nu(B)$. Further, for smaller α the variation of μ in terms of ν is larger.

DEFINITION 2.5. *If μ is ν-absolutely continuous and $\alpha_\nu(\mu) > 0$, then we say that μ is* uniformly *ν-absolutely continuous and call $\alpha_\nu(\mu)$ the* smoothness parameter *of μ*

with respect to ν.

Next, let $(E_1, \mathcal{O}_1)$ and $(E_2, \mathcal{O}_2)$ be two sigma-compact Hausdorff spaces whose topologies have locally countable bases, and let us call the associated Borel σ-algebras $\mathscr{B}_1$ and $\mathscr{B}_2$. We endow the space $E_1 \times E_2$ with the usual product topology $\mathcal{O}_1 \otimes \mathcal{O}_2$ generated by $\mathcal{O}_1 \times \mathcal{O}_2$. The product space is then sigma-compact and Hausdorff with locally countable basis, and the corresponding Borel σ-algebra $\mathscr{B}_1 \otimes \mathscr{B}_2$ is generated by $\mathscr{B}_1 \times \mathscr{B}_2$. For $(i = 1, 2)$, let ν_i be a sigma-finite, atom-free measure and μ_i a ν_i-absolutely continuous probability measure on $\mathscr{B}_i$ with smoothness parameter $\alpha_i := \alpha_{\nu_i}(\mu_i)$. And finally, let $\nu_1 \otimes \nu_2$ and $\mu_1 \otimes \mu_2$ be the corresponding product measures. It is well known and trivial to prove that $\nu_1 \otimes \nu_2$ is sigma-finite and atom-free, and that $\mu_1 \otimes \mu_2$ is a $(\nu_1 \otimes \nu_2)$-absolutely continuous probability measure.

THEOREM 2.6. *Under the above made assumptions, $\alpha_{\nu_1 \otimes \nu_2}(\mu_1 \otimes \mu_2) = \min(\alpha_1, \alpha_2)$.*

Proof. Without loss of generality we may assume that $\alpha_1 = \min(\alpha_1, \alpha_2)$. Let an arbitrary $\varepsilon > 0$ be fixed. By Proposition 2.3 ii), for all $\delta > 0$ and $(i = 1, 2)$ there exist $\mathscr{B}_i$-measurable sets $B^i_{\varepsilon,\delta}$ such that $0 < \nu_i(B^i_{\varepsilon,\delta}) \leq \delta$ and $\nu_i(B^i_{\varepsilon,\delta})^{\alpha_i+\varepsilon} \leq \mu_i(B^i_{\varepsilon,\delta})$. Let $\delta_0 \in (0, 1)$ be chosen such that

$$\nu_2(B^2_{\varepsilon,1})^{\alpha_1+\varepsilon} \cdot \delta_0^{\frac{\varepsilon}{2}} \leq \mu_2(B^2_{\varepsilon,1}).$$

For $\delta \leq \delta_0$, let us set $B := B^1_{\frac{\varepsilon}{2},\delta} \times B^2_{\varepsilon,1}$, so that

$$\nu_1 \otimes \nu_2(B) = \nu_1(B^1_{\frac{\varepsilon}{2},\delta}) \cdot \nu_2(B^2_{\varepsilon,1}) \leq \delta \cdot 1,$$

and

$$\begin{aligned}\mu_1 \otimes \mu_2(B) = \mu_1(B^1_{\frac{\varepsilon}{2},\delta}) \cdot \mu_2(B^2_{\varepsilon,1}) &\geq \nu_1(B^1_{\frac{\varepsilon}{2},\delta})^{\alpha_1+\frac{\varepsilon}{2}} \cdot \nu_2(B^2_{\varepsilon,1})^{\alpha_1+\varepsilon} \cdot \delta_0^{\frac{\varepsilon}{2}} \\ &\geq \left(\nu_1(B^1_{\frac{\varepsilon}{2},\delta}) \cdot \nu_2(B^2_{\varepsilon,1})\right)^{\alpha_1+\varepsilon} = \left(\nu_1 \otimes \nu_2(B)\right)^{\alpha_1+\varepsilon}.\end{aligned}$$

Proposition 2.3 ii) now implies $\alpha_{\nu_1 \otimes \nu_2}(\mu_1 \otimes \mu_2) \leq \alpha_1$. It remains to prove that

$$\alpha_1 \leq \alpha_{\nu_1 \otimes \nu_2}(\mu_1 \otimes \mu_2). \tag{2.3}$$

For this purpose, let $0 < \varepsilon < \alpha_1$, and let $F := (C_1 \times D_1) \cup \cdots \cup (C_N \times D_N)$ be a finite union of elements from $\mathscr{B}_1 \times \mathscr{B}_2$. It is easy to see that without loss of generality we may assume that the D_i are disjoint. Let $\eta := \nu_1 \otimes \nu_2(F)$ and

$$\begin{aligned} I_0 &:= \left\{i : \nu_1(C_i) \leq \eta\right\}, \\ I_k &:= \left\{i : \nu_1(C_i) \in [\eta^{1-(k-1)\frac{\varepsilon}{2}}, \eta^{1-k\frac{\varepsilon}{2}})\right\} \quad (k = 1, \ldots, \lfloor 2/\varepsilon \rfloor), \\ I_{\lfloor \frac{2}{\varepsilon} \rfloor + 1} &:= \left\{i : \nu_1(C_i) \geq \eta^{1-\lfloor \frac{2}{\varepsilon} \rfloor \frac{\varepsilon}{2}}\right\}. \end{aligned}$$

For all k let $G_k := \bigcup_{i \in I_k} D_i$ and $F_k := F \cap (E_1 \times G_k)$. For $k \geq 1$ it must then be true that

$$\eta \geq \nu_1 \otimes \nu_2(F_k) = \sum_{i \in I_k} \nu_1(C_i)\nu_2(D_i) \geq \eta^{1-(k-1)\frac{\varepsilon}{2}} \sum_{i \in I_k} \nu_2(D_i) = \eta^{1-(k-1)\frac{\varepsilon}{2}} \nu_2(G_k), \tag{2.4}$$

which establishes that $\nu_2(G_k) \le \eta^{(k-1)\frac{\varepsilon}{2}}$. The assumption that $\alpha_1 \le \alpha_2$ together with Proposition 2.3 iv) implies that for $(i = 1, 2)$ there exist $c_i \in (0, 1)$ such that $\mu_i \le c_i \nu_i^{\alpha_1 - \frac{\varepsilon}{2}}$. In particular, we have $\mu_2(G_k) \le c_2 \eta^{(\alpha_1 - \frac{\varepsilon}{2})(k-1)\frac{\varepsilon}{2}}$. It follows that

$$\begin{aligned}\mu_1 \otimes \mu_2(F_k) &= \sum_{i \in I_k} \mu_1(C_i)\mu_2(D_i) \le c_1(\eta^{1-k\frac{\varepsilon}{2}})^{\alpha_1 - \frac{\varepsilon}{2}} \sum_{i \in I_k} \mu_2(D_i) \\ &= c_1(\eta^{1-k\frac{\varepsilon}{2}})^{\alpha_1 - \frac{\varepsilon}{2}} \mu_2(G_k) \le c_1 c_2 \eta^{(\alpha_1 - \frac{\varepsilon}{2})(1 - \frac{\varepsilon}{2})} \le c_1 c_2 \eta^{\alpha_1 - \varepsilon}.\end{aligned}$$

For $k = 0$ we find similarly,

$$\mu_1 \otimes \mu_2(F_0) = \sum_{i \in I_0} \mu_1(C_i)\mu_2(D_i) \le c_1 \eta^{\alpha_1 - \frac{\varepsilon}{2}} \mu_2(G_0) \le c_1 c_2 \eta^{\alpha_1 - \varepsilon},$$

so that

$$\begin{aligned}\mu_1 \otimes \mu_2(F) &= \sum_k \mu_1 \otimes \mu_2(F_k) \le (2 + \lfloor 2/\varepsilon \rfloor)\, c_1 c_2 \eta^{\alpha_1 - \varepsilon} \\ &= (2 + \lfloor 2/\varepsilon \rfloor)\, c_1 c_2 \nu_1 \otimes \nu_2(F)^{\alpha_1 - \varepsilon}. \end{aligned} \tag{2.5}$$

Next, let F be an arbitrary $\mathscr{B}_1 \otimes \mathscr{B}_2$-measurable set. Since $\mathscr{B}_1 \times \mathscr{B}_2$ generates $\mathscr{B}_1 \otimes \mathscr{B}_2$, the outer measure construction tells us that $\nu_1 \otimes \nu_2(F) = \inf_{\mathcal{F}} \sum_{F' \in \mathcal{F}} \nu_1 \otimes \nu_2(F')$, where the infimum is over all countable collections $\mathcal{F} \subset \mathscr{B}_1 \times \mathscr{B}_2$ that satisfy $F \subseteq \bigcup_{F' \in \mathcal{F}} F'$. Hence, there exists a countable family $(F_i)_{\mathbb{N}} \subset \mathscr{B}_1 \times \mathscr{B}_2$ such that $F \subseteq \bigcup_i F_i$ and

$$\nu_1 \otimes \nu_2(F) \ge (1 - \varepsilon)\nu_1 \otimes \nu_2\Big(\bigcup_{i=1}^{\infty} F_i\Big). \tag{2.6}$$

It must also hold that for some $i_0 \in \mathbb{N}$,

$$\mu_1 \otimes \mu_2\Big(\bigcup_{i=1}^{i_0} F_i\Big) \ge (1 - \varepsilon)\mu_1 \otimes \mu_2\Big(\bigcup_{i=1}^{\infty} F_i\Big). \tag{2.7}$$

Therefore, we have

$$\begin{aligned}\mu_1 \otimes \mu_2(F) \quad &\le \mu_1 \otimes \mu_2\Big(\textstyle\bigcup_{i=1}^{\infty} F_i\Big) \overset{(2.7)}{\le} \frac{1}{1-\varepsilon} \cdot \mu_1 \otimes \mu_2\Big(\textstyle\bigcup_{i=1}^{i_0} F_i\Big) \\ &\overset{(2.5)}{\le} \frac{\Big(2 + \lfloor \frac{2}{\varepsilon} \rfloor\Big) c_1 c_2}{1-\varepsilon} \cdot \Big(\nu\Big(\textstyle\bigcup_{i=1}^{i_0} F_i\Big)\Big)^{\alpha_1 - \varepsilon} \le \frac{\Big(2 + \lfloor \frac{2}{\varepsilon} \rfloor\Big) c_1 c_2}{1-\varepsilon} \cdot \Big(\nu\Big(\textstyle\bigcup_{i=1}^{\infty} F_i\Big)\Big)^{\alpha_1 - \varepsilon} \\ &\overset{(2.6)}{\le} \frac{\Big(2 + \lfloor \frac{2}{\varepsilon} \rfloor\Big) c_1 c_2}{(1-\varepsilon)^{1 + \alpha_1 - \varepsilon}} \cdot \Big(\nu(F)\Big)^{\alpha_1 - \varepsilon}.\end{aligned}$$

Since this is true for all $\varepsilon \in (0, \alpha)$ and all $\mathscr{B}$-measurable sets F, Proposition 2.3 iv) implies that (2.3) holds. □

2.3. Examples. Let us now give a few nontrivial examples of uniformly absolutely continuous distributions. For proofs of the claims made in this section, we refer the reader to the technical report [14]. Having the application of the above theory to random input matrices for $\mathscr{C}_G(A)$ in mind, the probability measures chosen in the examples live on the unit sphere $\mathrm{S}^{m-1} \subset \mathbb{R}^m$, and the reference measure ν is chosen as the uniform measure $\mathscr{U}_{m-1}$ on S^{m-1}. Similar examples could of course be derived for arbitrary varieties.

Example 1. *If the density f of μ is $\mathscr{U}_{m-1}$-essentially bounded (i.e., $\exists\, M > 0$ s.t. $\nu(\{f > M\}) = 0$), then $\alpha_\nu(\mu) = 1$.*

Note that in this case Theorem 3.1 below shows that $\mathrm{P}[\mathscr{C}_G(A) > t] = \tilde{\mathscr{O}}(t^{-1})$. As mentioned earlier, special cases of this result were already established in [6, 8, 9].

Next, let $g \in C^0(\mathrm{S}^{m-1} \setminus \{x_0\}, \mathbb{R}_+)$. We say that g has a singularity of degree ς at x_0 if there exists a C^1-coordinate map $\varphi : D \to \mathrm{B}_1(\mathbb{R}^{m-1})$, where $D \subset \mathrm{S}^{m-1}$ is an open domain containing x_0 and $\mathrm{B}_1(\mathbb{R}^{m-1})$ is the open unit ball in $\mathbb{R}^m$, such that $\varphi(x_0) = 0$ and the limit $a_0 := \lim_{x\to x_0} \|\varphi(x)\|^\varsigma g(x)$ is well defined with $a_0 \in (0, +\infty)$. Functions g with this property can easily be constructed using a partition of unity. The following example shows that absolutely continuous probability distributions exist for all values of the smoothnesss parameter in $(0, 1]$.

Example 2. *Let g be as above, let ν denote $\mathscr{U}_{m-1}$, and let $\varsigma \in (0, m-1)$.*

- *i) If μ is a ν–absolutely continuous probability measure on S^{m-1} such that $d\mu/d\nu$ is essentially bounded by g then μ is has smoothness parameter $\alpha_\nu(\mu) \geq 1 - \varsigma/(m-1)$.*
- *ii) If μ is the ν–absolutely continuous probability measure on S^{m-1} defined by the density function $d\mu/d\nu \equiv g/\int_{\mathrm{S}^{m-1}} g(x)\nu(dx)$ then μ has smoothness parameter $\alpha_\nu(\mu) = 1 - \varsigma/(m-1)$.*

We remark that if $d\mu/d\mathscr{U}_{m-1}$ has a pole of degree $\varsigma \geq m-1$ then μ cannot be a finite measure and hence not a probability measure either. Therefore, this case need not be considered. Example 2 ii) provides an intuitive way of thinking about nontrivial values $\alpha < 1$ of the smoothness parameter as arising due to a singularity of the density function. It can even be established that all uniformly $\mathscr{U}_{m-1}$–absolutely continuous probability measures arise as the composition of a measure of the type exhibited in Example 2 i) with measure preserving maps from S^{m-1} to itself. Thus, in the general case the density is essentially bounded by a "delocalised" singularity. We cannot enter the details of this discussion here, as it would deviate too far from the central theme of this paper.

The following construction yields an example of an absolutely continuous measure that is not uniformly so.

Example 3. *By virtue of Example 2 we know that there exists a sequence $(\mu_i)_\mathbb{N}$ of $\mathscr{U}_{m-1}$-absolutely continuous probability measures with smoothness parameters $\alpha_{\mathscr{U}_{m-1}}(\mu_i) \leq i^{-1}$. For all $i \in \mathbb{N}$ let X_i be a random vector on S^{m-1} with law $\mathscr{L}(X_i) = \mu_i$, and let N be a random variable independent of the X_i taking values in $\mathbb{N}$ such that $\mathrm{P}[N = k] > 0$ for all k. Then the distribution $\mu = \mathscr{L}(X_N)$ of the random vector X_N is $\mathscr{U}_{m-1}$-absolutely continuous with smoothness parameter $\alpha(\mu, \mathscr{U}_{m-1}) = 0$.*

3. Tail Events of the Goffin-Cheung-Cucker Number. In this section we will analyse the tail behaviour of $\mathscr{C}_G(A)$. Recall that A is a random matrix of size $n \times m$, and that we denote the uniform measure on the unit sphere S^{m-1} by $\mathscr{U}_{m-1}$. Cheung-Cucker-Hauser [6] considered the case where A has i.i.d. rows with $\mathscr{L}(A_{i,:}) = \mathscr{U}_{m-1}$, or in other words, $\mathscr{L}(A) = \otimes^n \mathscr{U}_{m-1}$, and showed that in this case

$$\mathrm{P}[\mathscr{C}_G(A) \geq t] \leq c(n,m)t^{-1}, \quad t \geq 1, \tag{3.1}$$

where $c(n,m) := \binom{n}{m} 2m^{\frac{5}{2}}(1-\mathscr{I}_{m-2}(\arccos 1/t)/\mathscr{I}_{m-2}(\pi))^{n-m}$. The bound (3.1) also holds when A is a random matrix with i.i.d. Gaussian components, since

$$\mathrm{Diag}(\|A_{i,:}\|^{-1})A \stackrel{\mathscr{D}}{=} \otimes^n \mathscr{U}_{m-1}$$

and $\mathscr{C}_G(A)$ is invariant under row scalings of A. We now set out to boosting this result to general random matrices with $\otimes^n \mathscr{U}_{m-1}$-absolutely continuous distributions on $\otimes^n \mathrm{S}^{m-1}$ or $\otimes^{n\times m}\mathscr{N}(0,1)$-absolutely continuous distributions on $\mathbb{R}^{n\times m}$.

Theorem 3.1. *Let A be a random $n \times m$ matrix with distribution $\mathscr{L}(A) = \mu$. If μ is $\otimes^n \mathscr{U}_{m-1}$-absolutely continuous with smoothness parameter α, then*

$$\mathrm{P}[\mathscr{C}_G(A) > t] = \tilde{\mathscr{O}}\left(t^{-\alpha}\right).$$

Proof. Consider the $\mathscr{B}$-measurable set

$$W_t := \left\{X \in \mathbb{R}^{n\times m} : \, X_{i,:} \in \mathrm{S}^{m-1} \; \forall i, \, \mathscr{C}_G(X) > t\right\}.$$

Equation (3.1) shows that $\otimes^n \mathscr{U}_{m-1}(W_t) \to 0$ for $t \to \infty$. By Proposition 2.3 iii), for all $\varepsilon > 0$ there exists therefore $t_\varepsilon \geq 1$ such that for all $t \geq t_\varepsilon$,

$$\mathrm{P}[\mathscr{C}_G(A) > t] = \mu(W_t) \leq \left(\otimes^n \mathscr{U}_{m-1}(W_t)\right)^{\alpha-\varepsilon} \leq c(n,m)^{\alpha-\varepsilon} t^{-\alpha+\varepsilon},$$

where the last inequality follows from (3.1). □

Theorem 3.2. *Let A be a random $n \times m$ matrix that decomposes into disjoint blocks A_i $(i = 1, \ldots, k)$ of size $n_i \times m_i$ which are mutually independently distributed with laws $\mathscr{L}(A_i) = \mu_i$. If for all i μ_i is $\otimes^{n_i\times m_i}\mathscr{N}(0,1)$-absolutely continuous with smoothmess parameter α_i, then*

$$\mathrm{P}[\mathscr{C}_G(A) > t] = \tilde{\mathscr{O}}\left(t^{-\min_i \alpha_i}\right).$$

Proof. Applying Theorem 2.6 k times, we find that $\mathscr{L}(A)$ is $\otimes^{n\times m}\mathscr{N}(0,1)$-absolutely continuous with smoothness parameter $\min_i \alpha_i$. The rest of the proof is identical with that of Theorem 3.1 when W_t is replaced by $\{X \in \mathbb{R}^{n\times m} : \mathscr{C}_G(X) > t\}$. □

We remark that the block decomposition assumed in Theorem 3.2 need not be of grid-like type; these blocks can lie in arbitrary positions.

Example 4. *Let $A = \bar{A} + B$, where $\bar{A}$ is an arbitrary fixed $n \times m$ matrix and $\mathscr{B} = \otimes^{n\times m}\mathscr{N}(0,\sigma^2)$ is a Gaussian perturbation term. This is the probabilistic model studied in smoothed analysis. Since $\mathscr{N}(\bar{A}_{ij},\sigma^2)$ is $\mathscr{N}(0,1)$-absolutely continuous with smoothness parameter 1 for all i,j, Theorem 3.2 shows that $\mathrm{P}[\mathscr{C}_G(A) > t] = \tilde{\mathscr{O}}\left(t^{-1}\right)$.*

Example 5. *Let A be a random $n \times m$ matrix with independently distributed unit row vectors $A_{i,:}$ and such that $\mu_i := \mathscr{L}(A_{i,:})$ is $\mathscr{U}_{m-1}$-absolutely continuous with smoothness parameter α_i. In particular, this model arises when each row of a fixed matrix $\bar{A}$ is smoothed with a different uniformly absolutely continuous perturbation term. Applying Theorem 2.6 n times, we find that $\mathscr{L}(A)$ is $\otimes^n \mathscr{U}_{m-1}$-absolutely continuous with smoothness parameter $\min_i \alpha_i$, and by Theorem 3.1 we find $\mathrm{P}[\mathscr{C}_G(A) > t] = \tilde{\mathscr{O}}\left(t^{-\min_i \alpha_i}\right)$.*

Example 6. *Let A decompose into blocks as in Theorem 3.2, and let $A_i \sim \mathscr{N}(\bar{A}_i, Q_i)$ for all i, where $\bar{A}_i$ is an arbitrary fixed matrix of size $n_i \times m_i$ and Q_i is a positive definite covariance matrix of size $n_i m_i \times n_i m_i$. This generalises the probabilistic model of Example 4, as the components of A_i are no longer independent random variables. Since $\mathscr{N}(\bar{A}_i, Q_i)$ is $\otimes^{n_i \times m_i}\mathscr{N}(0,1)$-absolutely continuous with smoothness parameter 1, Theorem 3.2 once again shows that $\mathrm{P}[\mathscr{C}_G(A) > t] = \tilde{\mathscr{O}}\left(t^{-1}\right)$.*

Example 7. *Continuing the spirit of Example 6, the matrix in Theorem 3.2 can be thought of as arising as a block-wise smoothing of an arbitrary fixed matrix $\bar{A}$, where each block perturbation term $B_i = A_i - \bar{A}_i$ has a $\otimes^{n_i \times m_i}\mathscr{N}(0,1)$-absolutely continuous distribution with smoothmess parameter α_i.*

3.1. Tightness of Bounds. The upper bound in Theorem 3.1 already establishes that if random input data $A = [\, x_1 \dots x_n \,]^{\mathrm{T}} \in \mathbb{R}^{n \times m}$ of the conic feasibility problem has smoothness parameter $\alpha > 0$, then the random running time Op of several interior-point methods for these random problem instances has exponential tail-decay $\mathrm{P}[\mathrm{Op} > t] < \exp(-\gamma t)$ for some $\gamma > 0$. This result was obtained through a simple mechanism provided by Theorem 2.6 which allowed us to boost the corresponding result for the case where the problem input matrix A has rows that are i.i.d. uniformly distributed on the sphere. The following construction shows that Theorem 3.1 is essentially the best possible.

Theorem 3.3. *For all $0 < \alpha \leq 1$ there exists a probability measure μ on $(\mathrm{S}^{m-1})^n$ which has smoothness parameter α wrt. $\otimes^n \mathscr{U}_{m-1}$ and satisfies*

$$\mathrm{P}[\mathscr{C}_G(A) > t] = \Omega(t^{-\alpha}).$$

Proof. For convenience let us write $\nu := \otimes^m \mathscr{U}_{m-1}$, and let B be a random matrix with $\mathscr{L}(B) = \nu$. In [6] it was already shown that

$$\mathrm{P}[\mathscr{C}_G(B) > t] = \Omega(t^{-1}). \tag{3.2}$$

So if $\alpha = 1$ then the choice $\mu := \nu$ works. Next, suppose that $0 < \alpha < 1$. The probability density f of μ shall be given by

$$f(x_1, \dots, x_n) := c \cdot \mathscr{C}_G([x_1 \dots x_n]^T)^{1-\alpha} \cdot \mathbf{1}_{\{\mathscr{C}_G([x_1 \dots x_n]^T) < \infty\}}.$$

Here $c > 0$ is chosen such that $\int_{(S^{m-1})^n} f = 1$ (we shall see shortly that f and c are well-defined). Also notice that $\mathrm{P}[\mathscr{C}_G(B) = \infty] = 0$, as it can be seen from (2.1) that $\mathscr{C}_G([x_1 \dots x_n]^T) = \infty$ only if there are m collinear rows among $x_1, \dots, x_n$. With this choice of μ we find for $t > 0$:

$$
\begin{aligned}
\mathrm{P}[\mathscr{C}_G(A) > t] &= \int_{(S^{m-1})^n} \{\mathscr{C}_G([x_1 \dots x_n]^T) > t\} f(x_1, \dots, x_n) d\nu \\
&= c\,\mathrm{E}\left[\mathscr{C}_G(B)^{1-\alpha} \mathbf{1}_{\{\mathscr{C}_G(B)^{1-\alpha} > t^{1-\alpha}\}}\right] \\
&= c \int_0^\infty \mathrm{P}[\mathscr{C}_G(B)^{1-\alpha} \mathbf{1}_{\{\mathscr{C}_G(B)^{1-\alpha} > t^{1-\alpha}\}} > s] ds \\
&= c\left(t^{1-\alpha}\,\mathrm{P}[\mathscr{C}_G(B)^{1-\alpha} > t^{1-\alpha}] + \int_{t^{1-\alpha}}^\infty \mathrm{P}[\mathscr{C}_G(B)^{1-\alpha} > s] ds\right) \\
&= c\left(t^{1-\alpha}\,\mathrm{P}[\mathscr{C}_G(B) > t] + \int_{t^{1-\alpha}}^\infty \mathrm{P}[\mathscr{C}_G(B) > s^{1/(1-\alpha)}] ds\right) \\
&\overset{(3.1),(3.2)}{=} \Theta\left(t^{-\alpha} + \int_{t^{1-\alpha}}^\infty s^{-1/(1-\alpha)} ds\right) = \Theta(t^{-\alpha}).
\end{aligned}
$$

Notice that the above computation also shows that

$$\mathscr{C}_G([x_1 \dots x_n]^T)^{1-\alpha} \mathbf{1}_{\{\mathscr{C}_G([x_1 \dots x_n]^T) < \infty\}}$$

is integrable wrt. ν, so that c and f are indeed well-defined. The result for $0 < \alpha < 1$ now follows from Proposition 2.4 together with (3.1) and (3.2), and by noting that the definition of f implies

$$\{f > t\} = \{(x_1, \dots, x_n) \in (S^{m-1})^n : s_t < \mathscr{C}_G([x_1 \dots, x_n]^T) < \infty\}$$

for a suitably chosen s_t, and – as mentioned above –

$$\nu(\{(x_1, \dots, x_n) \in (S^{m-1})^n : \mathscr{C}_G([x_1 \dots, x_n]^T) = \infty\}) = 0.$$

□

A straighforward adaptation of the construction from Example 3 also yields the following result.

Corollary 3.4. *There exists a $\otimes^n \mathscr{U}_{m-1}$-absolutely continuous probability measure μ on $(S^{m-1})^n$ such that*

$$\mathrm{P}\left[\mathscr{C}_G(A) > t\right] = \Omega(t^{-\gamma}),$$

for all $\gamma > 0$.

In other words, this shows that for absolutely continuous input distributions with $\alpha = 0$ the tail probabilities $\mathrm{P}[\mathscr{C}_G(A) > t]$ do not decay at an algebraic rate in general. Similar constructions show that the exponents in Theorem 3.2 are also essentially best possible, and that an analogue of Corollary 3.4 holds.

The measure μ constructed in the proof of Theorem 3.3 might seem a bit artifical. A natural restriction is to consider distributions corresponding to matrices with independent rows. In this case the decay rate of Theorem 3.1 is not tight and can be further improved at the expense of working a bit harder. In the remainder of this section we will show that when $n \geq m$ and the rows of A are i.i.d. with smoothness parameter α wrt. $\mathscr{U}_{m-1}$,

$$\mathrm{P}\left[\mathscr{C}_G(A) > t\right] = \tilde{\mathscr{O}}\left(t^{-\min(1,2\alpha)}\right),$$

and that there exist measures for which

$$\mathrm{P}\left[\mathscr{C}_G(A) > t\right] = \Omega\left(t^{-m\alpha}\right).$$

We conjecture that the upper bound can be further improved to

$$\mathrm{P}\left[\mathscr{C}_G(A) > t\right] = \tilde{\mathscr{O}}\left(t^{-\min(1,m\alpha)}\right).$$

Let $n \geq m$ and denote $\mathscr{P}_m := \{S \subseteq \{1,\dots,n\} : |S| = m\}$. For $S \in \mathscr{P}_m$ let A_S be the $m \times m$ matrix obtained by removing all rows from A with index not in S. Since the probability models of A we study are all with i.i.d. $\mathscr{U}_{m-1}$-absolutely continuous row vectors, A_S is nonsingular with probability one, $U_S := A_S^{-1}\mathbf{1}$ is well defined, where $\mathbf{1} := [1 \dots 1]^{\mathrm{T}} \in \mathbb{R}^m$. Proposition 4.2 and Lemmas 4.3 and 4.4 of [6] show the inclusion of events

$$\{\mathscr{C}_G > t\} \subseteq \{\exists\, S \in \mathscr{P}_m \text{ s.t. } \|U_S\|_2 > t\}. \tag{3.3}$$

Since $A_S U_S = \mathbf{1}$, it is the case that $\langle X_i, U_S\rangle = 1$ for all $i \in S$, and this implies

$$U_S = \sum_{i\in S} \frac{Y_i}{\langle X_i, Y_i\rangle},$$

where Y_i is the unique unit vector in $\mathrm{Span}(\{X_j : j \in S \setminus \{i\}\})^{\perp}$ that turns $\{X_j : j \in S \setminus \{i\}\} \cup \{Y_i\}$ into a positively oriented basis of $\mathbb{R}^m$ when ordered according to increasing indices (this latter convention is only necessary to render Y_i well defined, i.e., to make a definite choice between Y_i and $-Y_i$). Hence, if $\mathscr{C}_G(A) > t$ then there must exist $S \in \mathscr{P}_m$ such that

$$\sqrt{m} \cdot \sum_{i\in S} \left|\frac{1}{\langle X_i, Y_i\rangle}\right| \geq \|U_S\|_2 > t,$$

and then at least one of the terms on the left must exceed $t/m^{3/2}$. Using the fact that the X_i are i.i.d., the previous discussion implies that

$$\mathrm{P}[\mathscr{C}_G(A) > t] \leq m\binom{n}{m} \cdot \mathrm{P}\left[|\langle Y(X_1,\dots,X_{m-1}), X_m\rangle| < \frac{m^{3/2}}{t}\right], \tag{3.4}$$

where $Y(X_1,\dots,X_{m-1})$ equals the vector Y_m defined for $S = \{1,\dots,m\}$.

Lemma 3.5. *If X_i $(i = 1,\dots,m-1)$ are i.i.d. random unit vectors in $\mathbb{R}^m$ with $\mathscr{U}_{m-1}$-absolutely continuous distribution $\mathscr{L}(X_i) = \mu$ of smoothness parameter $\alpha = \alpha_{\mathscr{U}_{m-1}}(\mu)$, then the distribution $\mathscr{L}(Y(X_1,\dots,X_{m-1}))$ is $\mathscr{U}_{m-1}$-absolutely continuous with smoothness parameter value at most α.*

Proof. This follows quite straightforwardly from Theorem 2.6: Let $V_1,\dots,V_{m-1}$ be i.i.d. uniformly distributed on S^{m-1}. Then $Y(V_1,\dots,V_{m-1})$ is also uniformly distributed on S^{m-1}. For any $\mathscr{B}$-measurable $W \subseteq \mathrm{S}^{m-1}$ and $\varepsilon > 0$ we have

$$\begin{aligned}
\mathrm{P}[Y(X_1,\dots,X_{m-1}) \in W] &= \mathrm{P}\left[(X_1,\dots,X_{m-1}) \in Y^{-1}(W)\right] = \otimes^{m-1}\mu\left(Y^{-1}(W)\right) \\
&\leq c_\varepsilon \cdot \otimes^{m-1}\mathscr{U}_{m-1}\left(Y^{-1}(W)\right)^{\alpha-\varepsilon} \\
&= c_\varepsilon \cdot \mathrm{P}\left[(V_1,\dots,V_{m-1}) \in Y^{-1}(W)\right]^{\alpha-\varepsilon} \\
&= c_\varepsilon \cdot \mathrm{P}\left[Y(V_1,\dots,V_{m-1}) \in W\right]^{\alpha-\varepsilon} = c_\varepsilon \cdot \mathscr{U}_{m-1}(W)^{\alpha-\varepsilon},
\end{aligned}$$

where c_ε is chosen as in Proposition 2.3 iii) applied to the measure $\otimes^{m-1}\mu$. □

PROPOSITION 3.6. *Let X, Y be independent random vectors with $\mathscr{U}_{m-1}$-absolutely continuous distributions on S^{m-1}, and such that the smoothness parameters are at most $\alpha > 0$ for both variables. Then $\mathrm{P}[|\langle X, Y\rangle| < r] = \tilde{\mathscr{O}}(r^{\min(1,2\alpha)})$.*

Proof. To reduce the amount of notation required in the proof, we first show that it suffices to establish the result for X, Y identically distributed. Let $W_{0,i}$ $(i = 1, 2)$ be independent copies of X, $W_{1,i}$ $(i = 1, 2)$ independent copies of Y and N_i $(i = 1, 2)$ independent Bernoulli variables with parameter 1/2. Then $W_i := W_{N_i,i}$ $(i = 1, 2)$ are i.i.d. random vectors with smoothness parameter α, and furthermore,

$$\mathrm{P}\left[|\langle X, Y\rangle| < r\right] \leq 2\,\mathrm{P}\left[|\langle W_1, W_2\rangle| < r\right],$$

so that it suffices to show the right-hand side is $\tilde{\mathscr{O}}(r^{\min(1,2\alpha)})$. In what follows we will thus assume that X and Y are identically distributed and denote their common distribution by μ. Let us first assume that $\alpha \leq 1/2$. For $r > 0$ let $\rho(r) := \arcsin(r)$, so that $x \in \mathrm{cap}(y, \rho(r))$ iff $\|x - y\| < r$ for $x, y \in S^{m-1}$. For a fixed $r > 0$, let $x_1, \dots, x_N \in \mathrm{S}^{m-1}$ be chosen[1] so that

$$\mathrm{cap}\left(x_i, \rho(r/2)\right) \cap \mathrm{cap}\left(x_j, \rho(r/2)\right) = \emptyset \quad (i \neq j), \tag{3.5}$$

$$\mathrm{S}^{m-1} \subseteq \bigcup_i \mathrm{cap}\left(x_i; \rho(r)\right). \tag{3.6}$$

Thus we can partition the sphere S^{m-1} into disjoint sets $C_1, \dots, C_N$ such that

$$\mathrm{cap}\left(x_i, \rho(r/2)\right) \subseteq C_i \subseteq \mathrm{cap}\left(x_i, \rho(r)\right).$$

Since $\mathscr{U}_{m-1}(\mathrm{cap}(x, \rho)) = \mathscr{I}_{m-2}(\rho)/\mathscr{I}_{m-2}(\pi)$, there exist constants $c_1 < c_2$ such that for all $r \in (0, 1)$,

$$c_1 \cdot r^{m-1} \leq \mathscr{U}_{m-1}\left(\mathrm{cap}(x, \rho(r))\right) \leq c_2 \cdot r^{m-1}.$$

Note that this gives that $N = \mathscr{O}(r^{-(m-1)})$. Next, we define an undirected graph G with vertex set $\{1, \dots, N\}$ and an edge $ij \in E(G)$ if and only if $|\langle x_i, x_j\rangle| < 4r$. We remark that

$$\mathrm{P}\left[|\langle X, Y\rangle| < r\right] \leq \mathrm{P}\left[\text{There exists an edge } ij \in E(G) \text{ such that } X \in C_i, Y \in C_j\right].$$

To see this, note that if $|\langle X, Y\rangle| < r$ holds and $X \in C_i$, $Y \in C_j$, then, using the Cauchy-Schwartz inequality,

$$\begin{aligned}|\langle x_i, x_j\rangle| &= |\langle (x_i - X) + X, (x_j - Y) + Y\rangle| \\ &\leq |\langle (x_i - X), (x_j - Y)\rangle| + |\langle (x_i - X), Y\rangle| + |\langle X, (x_j - Y)\rangle| + |\langle X, Y\rangle| \\ &< r^2 + 3r.\end{aligned}$$

We can conclude that

$$\mathrm{P}\left[|\langle X, Y\rangle| < r\right] \leq \sum_{ij \in E(G)} p_i p_j, \tag{3.7}$$

[1]This can be achieved by iteratively adding points x_i as long as (3.5) can be satisfied. Since the area of S^{m-1} is finite, this process must end after $N < \infty$ choices have been made. Criterion (3.6) is now automatically satisfied, because for each $x \in \mathrm{S}^{m-1}$ there exists $y \in \mathrm{cap}(x, \rho(r/2)) \cap \mathrm{cap}(x_i, \rho(r/2))$ for some i, and then $\|x - x_i\|_2 \leq \|x - y\|_2 + \|y - x_i\|_2 < r$, showing that $x \in \mathrm{cap}(x_i, \rho(r))$.

where $p_i := \mathrm{P}[X \in C_i]$. The formulas of Section 2 imply that for all $y \in \mathrm{S}^{m-1}$, $\mathscr{U}_{m-1}(\{x : |\langle x, y\rangle| < r\}) = 1 - 2 \cdot \mathscr{I}_{m-2}(\arccos(r))/\mathscr{I}_{m-2}(\pi)$, so that there exist constants $0 < d_1 < d_2$ such that for all $r \in (0,1)$,

$$d_1 \cdot r \le \mathscr{U}_{m-1}\left(\{x : |\langle x, y\rangle| < r\}\right) = \mathscr{U}_{m-1}\left(\{x : |x_m| \le r\}\right) \le d_2 \cdot r.$$

Also observe that if $ij \in E(G)$ then, by an inner product computation similar to the one above, $C_j \subseteq \{x : |\langle x_i, x\rangle| < 5r\}$. It follows that the degree of any vertex in G is bounded above by

$$D := \left\lfloor \frac{5 d_2 r}{c_1 \left(\frac{r}{2}\right)^{m-1}} \right\rfloor = \Theta\left(r^{-(m-2)}\right).$$

Without loss of generality we may assume that the x_i were ordered so that $p_1 \ge p_2 \ge \cdots \ge p_N$. Let us write

$$\begin{aligned} J_k &:= \{(k-1)(D+1)+1, \dots, k(D+1)\}, \quad (k = 1, \dots, \lfloor N/D+1 \rfloor), \\ J_{\lfloor N/(D+1)\rfloor+1} &:= \{\lfloor N/(D+1)\rfloor \cdot (D+1)+1, \dots, N\}, \end{aligned}$$

where the last index set is obviously empty if $D+1$ divides N. Let us now apply the following rule until exhaustion of candidate edges:

> If $ij \in E(G)$ for some $i \in J_1, j \notin J_1$, then there exists $k \in J_1 \setminus \{i\}$ such that $ik \notin E(G)$, for otherwise i would have degree $\ge D+1$. Node k either has degree $< D$ or it has a neighbour ℓ with $\ell \notin J_1$. In the first case, add the edge ik and remove ij from $E(G)$. In the second case, add ik and remove $ij, \ell k$, and note that $2p_ip_k \ge p_ip_j + p_\ell p_k$.

After this process has finished it is still the case that the degree of none of the nodes of the new graph exceeds D, and furthermore, nodes with indices in J_1 are only joined to nodes with indices in J_1. Therefore, if we next apply the same procedure to the nodes $i \in J_2$, none of the edges incident to nodes in J_1 will change again. After applying the procedure to $J_2, \dots, J_{\lfloor N/(D+1)\rfloor+1}$, we end up with a graph G' that satisfies

$$\sum_{ij \in E(G)} p_ip_j \le 2 \sum_{ij \in E(G')} p_ip_j \le 2 \sum_l \sum_{i,j \in J_l} p_ip_j < \sum_l \mu(A_l)^2, \tag{3.8}$$

where $A_k := \bigcup_{i \in J_k} C_i$. The A_k form a partition of the sphere S^{m-1}, and $\mathscr{U}_{m-1}(A_k) \le (D+1)c_2 r^{m-1}$. Hence, setting $\gamma := 5c_2d_2c_1^{-1}2^{m-1}+1$ we find that for all $0 < r < 1$,

$$\mathscr{U}_{m-1}(A_k) \le \gamma \cdot r.$$

Let $\varepsilon > 0$, and set

$$\begin{aligned} L_\ell &:= \left\{k : \mu(A_k) \in [r^{\alpha+\ell\varepsilon}, r^{\alpha+(\ell-1)\varepsilon})\right\}, \quad (\ell = 0, \dots, \lceil (1-\alpha)/\varepsilon \rceil), \\ L_{\lceil \frac{1-\alpha}{\varepsilon} \rceil + 1} &:= \{k : \mu(A_k) \le r\}. \end{aligned}$$

Note that when r is small enough then every A_k is contained in some L_ℓ. Now, for $\ell \le \lceil (1-\alpha)/\varepsilon \rceil$ we have,

$$|L_\ell| \cdot r^{\alpha+\ell\varepsilon} \le \mu\left(\bigcup_{k \in L_\ell} A_k\right) \le c_\varepsilon \cdot (|L_\ell| \cdot \gamma \cdot r)^{\alpha-\varepsilon},$$

giving $|L_\ell| = \mathscr{O}(r^{-\frac{(\ell+1)\varepsilon}{(1-\alpha+\varepsilon)}})$. Since $\alpha \le 1/2$ by assumption, we find $|L_\ell| = \mathscr{O}(r^{-2(\ell+1)\varepsilon})$. On the other hand,

$$\left|L_{\lceil \frac{1-\alpha}{\varepsilon} \rceil + 1}\right| \le \left\lceil \frac{N}{D} \right\rceil = \mathscr{O}\left(r^{-1}\right),$$

as $N = \mathscr{O}(r^{-(m-1)})$ and $D = \Theta(r^{-(m-2)})$. Combining (3.7) and (3.8) with the above estimates we find

$$\mathrm{P}\left[|\langle X, Y\rangle| < r\right] \le \sum_{\ell=0}^{\lceil \frac{1-\alpha}{\varepsilon} \rceil} |L_\ell| \cdot r^{2(\alpha+(\ell-1)\varepsilon)} + \left|L_{\lceil \frac{1-\alpha}{\varepsilon} \rceil + 1}\right| \cdot r^2 = \mathscr{O}\left(r^{2\alpha - 4\varepsilon}\right).$$

This shows that $\mathrm{P}[|\langle X, Y\rangle| < r] = \tilde{O}(r^{2\alpha})$, provided that $\alpha \le 1/2$. Finally, notice that in the computations so far the only facts used about α are the upper bounds provided by parts iii) and iv) of Proposition 2.3, which also hold if we replace α by $\alpha' < \alpha$. Hence, if $\alpha > 1/2$ the computations still carry through using $\alpha' = 1/2$ instead, and we get $\mathrm{P}[|\langle X, Y\rangle| < r] = \tilde{\mathscr{O}}(r)$ in this case. ◻

Theorem 3.7. *If A is a random $n \times m$ matrix with i.i.d. rows, having smoothness parameter $\alpha > 0$ wrt. $\mathscr{U}_{m-1}$ and $m \ge n$, then*

$$\mathrm{P}\left[\mathscr{C}_G(A) > t\right] = \tilde{\mathscr{O}}\left(t^{-\min(1, 2\alpha)}\right).$$

Proof. This follows immediately from inequality (3.4), Lemma 3.5 and Proposition 3.6. ◻

3.2. Lower Bounds on Tail Probabilities for independent rows. It is now natural to ask whether the upper bound on $\mathrm{P}[\mathscr{C}_G(A) > t]$ under the assumption of independent rows provided by Theorem 3.7 can be further improved. While we suspect that there is indeed room for further improvements when $m > 2$, Theorem 3.10 below establishes that in the case $m = 2$ the exponent of Theorem 3.7 cannot be improved for general input distributions with smoothness parameter α. Furthermore, the same result establishes that in general (for arbitrary m and arbitrary input distributions with smoothness parameter $\alpha > 0$) $\mathrm{P}[\mathscr{C}_G(A) > t]$ does not decay faster than at an algebraic rate. Before we can present these results, we need two lemmas.

Lemma 3.8. *Let X be a non-atomic random variable on $\mathbb{R}$ with cumulative distribution function $\mathscr{F}_X$ and let Y be a random variable on $\mathbb{R}$ with cumulative distribution function $\mathscr{F}_Y(x) = \mathscr{F}_X(x)^\alpha$ with $0 < \alpha \le 1$. Then*

- *i)* $\mathrm{P}[Y \in B] \le \mathrm{P}[X \in B]^\alpha$ *holds true for all Borel-measurable $B \subseteq \mathbb{R}$ and* $\alpha_{\mathscr{L}(X)}(\mathscr{L}(Y)) = \alpha$.
- *ii) If $X = |Z|$, where Z is a symmetric random variable on $\mathbb{R}$, then* $\mathrm{P}[Y \in B] \le 2^\alpha \cdot \mathrm{P}[Z \in B]^\alpha$ *holds and* $\alpha_{\mathscr{L}(Z)}(\mathscr{L}(Y)) = \alpha$. *Furthermore, for B of the form $B = [0, c]$ we have* $\mathrm{P}[Y \in B] = 2^\alpha \cdot \mathrm{P}[Z \in B]^\alpha$.

Proof. First note that $\mathscr{F}_Y$ in fact determines a unique probability distribution on $\mathbb{R}$. The case $\alpha = 1$ is trivial, so we may assume that $\alpha \in (0, 1)$. We notice that

$$\mathrm{P}[X \in B] = \mathrm{P}[\mathscr{F}_X(X) \in \mathscr{F}_X(B)] = \mathrm{P}[U \in \mathscr{F}_X(B)],$$

where U is a random variable with uniform distribution on $[0, 1]$, and $\mathrm{P}[Y \in B] = \mathrm{P}[U \in \mathscr{F}_Y(B)]$. Setting $C := \mathscr{F}_X(B)$ and $\phi(x) := x^\alpha$, we find $\mathrm{P}[X \in B] = \int_C 1 dx$

and

$$\mathrm{P}[Y \in B] = \int_{\phi[C]} 1 dx = \int_C \phi'(y) dy = \int_C \alpha y^{\alpha-1} dy,$$

where we used the substitution $y = \phi^{-1}(x)$. Now note that among all the sets $C \subseteq [0,1]$ of Lebesgue measure $p := \mathrm{P}[X \in B]$ the set $[0,p]$ maximises $\int_C \alpha y^{\alpha-1} dy$ (using that $[0,p]$ is of the form $\{y : \alpha y^{\alpha-1} \geq c\}$). Therefore, $\mathrm{P}[Y \in B] \leq p^\alpha$, as required in part i). Furthermore, the above argument shows that equality is achieved for sets of the form $B = \mathscr{F}_X^{-1}([0,p]) = (-\infty, c]$. Part ii) is an immediate extension of the same argument. □

LEMMA 3.9. *Let* $V := [\, V_1 \cdots V_{m-1} \,]$ *be a random vector with uniform distribution on* S^{m-2} *and* Z *a random variable independent of* V *and identically distributed as the* m*-th component* U_m *of a random vector* U *with uniform distribution on* S^{m-1}. *Finally, let* W *be a random variable independent of* V *and such that* $\mathrm{P}[W \leq z] = \mathrm{P}[|Z| \leq z]^\alpha$ *for some* $\alpha \in (0,1)$. *Then* $\mu := \mathscr{L}([\, \sqrt{1-W^2} V \;\; W \,])$ *is a* $\mathscr{U}_{m-1}$*-absolutely continuous measure on* S^{m-1} *with smoothness parameter* $\alpha_{\mathscr{U}_{m-1}}(\mu) = \alpha$.

The proof is a straightforward combination of Lemma 3.8 with an argument analogous to the proof of Lemma 3.5 and is left to the reader (but it can again be found in the technical report [14]).

THEOREM 3.10. *For any* $\alpha \in (0,1)$ *there exists a* $\mathscr{U}_{m-1}$*-absolutely continuous measure* μ *on* S^{m-1} *with smoothness parameter* $\alpha_{\mathscr{U}_{m-1}}(\mu) = \alpha$ *such that if* $\mathscr{L}(A) = \otimes^n \mu$ *then*

$$\mathrm{P}\left[\mathscr{C}_G(A) > t\right] = \Omega\left(t^{-m\alpha}\right).$$

Proof. Let $p_1, \ldots, p_m \in \mathrm{S}^{m-2}$ and $c > 0$ be chosen such that for all $x \in \mathrm{S}^{m-2}$ there is an i such that $\langle x, p_i \rangle \geq c$[2]. Let V_i and W_i $(i = 1, \ldots, m)$ be i.i.d. copies of the random vector V and the random variable W respectively that were defined in Lemma 3.9, and let

$$X_i = \left[\sqrt{1 - W_i^2} \cdot V_i \quad W_i\right]$$

for $i = 1, \ldots, m$ and $A := [X_1 \ldots X_m]^T$. For $t > 0$ let us consider the event

$$B_t := \left\{ \|V_i - p_i\|_2 < \frac{c}{2},\, W_i \leq \frac{1}{t} \; (i = 1, \ldots, m) \right\},$$

and let $\tilde{c} := (\mathscr{I}_{m-3}(2 \arcsin(c/4))/\mathscr{I}_{m-3}(\pi))^m$, where $\mathscr{I}_k$ are the functions defined in Section 2. We remark that $\mathrm{P}[|Z| \leq t^{-1}] = \Omega(t^{-1})$[3]. By Lemma 3.8 ii) we therefore have

$$\mathrm{P}\left[B_t\right] = \tilde{c} \cdot \mathrm{P}\left[W \leq t^{-1}\right]^m = \tilde{c} \cdot \mathrm{P}\left[|Z| \leq t^{-1}\right]^{m\alpha} = \Omega\left(t^{-m\alpha}\right).$$

[2]It is easily checked that $p_i = e_i$ $(i = 1, \ldots, m-1)$, $p_m = -(e_1 + \cdots + e_{m-1})/\sqrt{m-1}$, and $c = ((m-2)\sqrt{m-1} + m - 1)^{-1}$ is an example of a valid choice: Suppose that $x = [\, x_1 \cdots x_{m-1} \,] \in \mathrm{S}^{m-2}$ satisfies $\langle x, p_i \rangle < c$ for all i. Then, since $\|x\|_2 = 1$, there exists at least one i with $|x_i| \geq 1/\sqrt{m-1}$ and thus $x_i < -1/\sqrt{m-1} < -c$. But then $\langle x, p_m \rangle \geq (1/\sqrt{m-1} - (m-2)c)/\sqrt{m-1} = c$.

[3]In fact, $\mathrm{P}[|Z| \leq t^{-1}] = 1 - 2 \cdot \mathscr{I}_{m-2}(\arccos(1/t))/\mathscr{I}_{m-2}(\pi)$.

To prove our claim, it thus suffices to show that $B_t \subseteq \{\mathscr{C}(A) \geq t\}$. Recall that by definition $\mathscr{C}_G(A)^{-1} = |\min_{x\in \mathrm{S}^{m-1}} \max_i \langle X_i, x\rangle|$. Writing $x = \left[\begin{smallmatrix} u\cdot\sqrt{1-z^2} & z \end{smallmatrix}\right]$, we have

$$\langle X_i, x\rangle = \sqrt{(1-W_i^2)(1-z^2)} \cdot \langle V_i, u\rangle + W_i \cdot z$$

and

$$\min_{x\in \mathrm{S}^{m-1}} \max_i \langle X_i, x\rangle \leq \max_i \langle X_i, -e_m\rangle = \max_i -W_i \leq 0.$$

By construction of the p_i there exists an index i such that $\langle p_i, u\rangle \geq c$, and hence, by the Cauchy-Schwartz inequality, $\langle V_i, u\rangle > c/2$ when B_t occurs, and in that case we also have

$$\langle X_i, x\rangle > \sqrt{(1-W_i^2)(1-z^2)} \cdot \frac{c}{2} + W_i \cdot z \geq -W_i \geq -t^{-1}.$$

Consequently,

$$0 \geq \min_{x\in \mathrm{S}^{m-1}} \max_i \langle X_i, x\rangle \geq -t^{-1}.$$

This shows that the occurrence of B_t implies $\mathscr{C}(A) \geq t$, as required. □

Again, a straightforward adaptation of the construction from Example 3establishes the following result.

Corollary 3.11. *There exists a $\mathscr{U}_{m-1}$-absolutely continuous measure μ on S^{m-1} with smoothness parameter $\alpha_{\mathscr{U}_{m-1}}(\mu) = 0$ such that if $\mathscr{L}(A) = \otimes^n \mu$ then*

$$\mathrm{P}[\mathscr{C}_G(A) > t] = \Omega(t^{-\gamma})$$

for all $\gamma > 0$.

We conclude this section with a result that implies that in general there does not exists an upper bound on $\mathrm{P}[\mathscr{C}(A) > t]$ better than $\tilde{\mathscr{O}}(t^{-\min(1,m\alpha)})$. We conjecture that the exponent $-\min(1, m\alpha)$ is tight, i.e. the upper bound in Theorem 3.7 can be sharpened to $\mathrm{P}[\mathscr{C}_G(A) > t] = \tilde{\mathscr{O}}(t^{-\min(1,m\alpha)})$.

Proposition 3.12. *For $\alpha > m^{-1}$ there exists a distribution μ with $\alpha_{\mathscr{U}_{m-1}}(\mu) = \alpha$ and such that if $\mathscr{L}(A) = \otimes^n \mu$, then*

$$\mathrm{P}\left[\mathscr{C}_G(A) > t\right] = \Omega(t^{-1}).$$

Proof. In [6] it was shown that if $\mathscr{L}(A) = \otimes^n \mathscr{U}_{m-1}$ then $\mathrm{P}[\mathscr{C}(A) > t] = \Omega(t^{-1})$. Let X_i^0 $(i = 1, \ldots, n)$ be i.i.d. with $\mathscr{L}(X_i^0) = \mathscr{U}_{m-1}$, let X_i^1 $(i = 1, \ldots, n)$ be i.i.d. with $\mathscr{L}(X_i^1) = \mu$ where μ is a distribution on S^{m-1} with $\alpha_{\mathscr{U}_{m-1}}(\mu) = \alpha$, and let N_i $(i = 1, \ldots, n)$ be independent Bernoulli variables with parameter $1/2$. Then the random matrix $\tilde{A} := \begin{bmatrix} X_1^{N_1} & \ldots & X_n^{N_n} \end{bmatrix}^{\mathrm{T}}$ also has smoothness parameter α wrt. $\otimes^n \mathscr{U}_{m-1}$ (as can be shown using an argument similar to Example 3), and

$$\mathrm{P}\left[\mathscr{C}_G(\tilde{A}) > t\right] \geq 2^{-n} \cdot \mathrm{P}\left[\mathscr{C}_G(A) > t\right] = \Omega\left(t^{-1}\right).$$

□

4. Tail Events of the Renegar Number. In the present section we take the analysis of Section 3 a step further and establish similar results for the Renegar number $\mathscr{C}_R(A)$. Recall that, in contrast to the Goffin-Cheung-Cucker number $\mathscr{C}_G(A)$, the Renegar number $\mathscr{C}_R(A)$ is not invariant under row scaling. Thus, the assumption that A has unit row vectors is no longer justified. A natural extension of the framework studied above is to consider random matrices $A = DX$, where $X : \Omega \to \mathbb{R}^{n\times m}$ is an absolutely continuous matrix with i.i.d. unit row vectors and smoothness parameter α, and where $D = \mathrm{Diag}(R_1, \ldots, R_n)$ is a diagonal matrix whose diagonal elements are i.i.d. absolutely continuous positive random variables independent of X.

Like in the case of $\mathscr{C}_G(A)$, the tail decay of $\mathscr{C}_R(A)$ depends on the smoothness parameter α of X, but in addition the tails of R_i also play a similar role. We say that the diagonal matrix D is absolutely continuous with tail exponents $(\beta, \gamma) \in \mathbb{R}^2_+$ if the law $\mathscr{L}(R_i)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$ and furthermore,

$$\mathrm{P}[R_i > t] = \tilde{\mathscr{O}}\left(t^{-\beta}\right),$$
$$\mathrm{P}[R_i^{-1} > t] = \tilde{\mathscr{O}}\left(t^{-\gamma}\right).$$

Example 8. *If A has i.i.d. standard normal entries $A_{ij} \sim \mathcal{N}(0,1)$, then the rows of A are of the form $A_{i,:} = R_i X_i$, where $X_i \sim \mathscr{U}(\mathrm{S}^{m-1})$ are i.i.d. uniform random vectors on the unit sphere and $R_i^2 \sim \chi^2_m$ are i.i.d. chi-square distributed random variables with m degrees of freedom. Since the density of this latter distribution is*

$$f(t) = \frac{t^{\frac{m}{2}-1} e^{-\frac{t}{2}}}{2^{\frac{m}{2}}\Gamma(m/2)},$$

one finds that we can take $\gamma = m$ and β may be taken arbitrarily large.

Next, we present a result that will put us in a position to bound the tail decay of Renegar's condition number. We call $f : \mathbb{R}^k \to \mathbb{R}$ *non-increasing* (respectively non-decreasing) if whenever $y \le z \in \mathbb{R}^k$ (component-wise) implies $f(y) \ge f(z)$ (respectively $f(y) \le f(z)$). The following lemma is a standard result. For a proof of part i) see e.g. [19]. The proof of part ii) is completely analogous.

Lemma 4.1. *Let $Y = (Y_1, \ldots, Y_k)$ be a random vector consisting of independent random variables, and let $f, g : \mathbb{R}^k \to \mathbb{R}$ be measurable functions. Then the following hold true.*

i) If f, g are either both non-increasing or both non-decreasing then

$$\mathrm{E}[f(Y)g(Y)] \ge \mathrm{E}[f(Y)] \cdot \mathrm{E}[g(Y)].$$

ii) If f is non-increasing and g is nondecreasing or vice versa then

$$\mathrm{E}[f(Y)g(Y)] \le \mathrm{E}[f(Y)] \cdot \mathrm{E}[g(Y)].$$

Theorem 4.2. *Let $A = DX$ be a random matrix as defined above. Then*

$$\mathrm{P}[\mathscr{C}_R(A) > t] = \tilde{\mathscr{O}}\big(t^{-\min(\alpha,\beta,\gamma)}\big).$$

Proof. Let $\varepsilon > 0$ and set $\xi_\varepsilon := \min(\alpha, \beta, \gamma) - \varepsilon$. Let $Z := \frac{\max_i R_i}{\min_i R_i}$ and $X := [X_1 \ldots X_n]^{\mathrm{T}}$. Because of the inequalities (2.2) and $\|A\|_2 \le \max_i \|A_{i,:}\|_2 \cdot \sqrt{n}$, it suffices to show that

$$\mathrm{P}\big[Z \cdot \mathscr{C}_G(X) > t\big] = \mathscr{O}(t^{-\xi_\varepsilon}). \tag{4.1}$$

By virtue of Theorem 3.1 there exists a constant c_ε such that for all $t > 0$,

$$\mathrm{P}\big[\mathscr{C}_G(X) > t\big] \leq c_\varepsilon \cdot t^{-\xi_\varepsilon}.$$

Writing f_Z for the density function of Z, we have

$$\begin{aligned}\mathrm{P}\big[Z \cdot \mathscr{C}_G(X) > t\big] &= \int_0^\infty \mathrm{P}\left[\mathscr{C}_G(X) > \frac{t}{s}\right] \cdot f_Z(s)ds \leq \int_0^\infty c_\varepsilon \cdot \left(\frac{t}{s}\right)^{-\xi_\varepsilon} \cdot f_Z(s)ds \\ &\leq c_\varepsilon \cdot t^{-\xi_\varepsilon} \int_0^\infty s^{\xi_\varepsilon} \cdot f_Z(s)ds = c_\varepsilon \cdot t^{-\xi_\varepsilon} \cdot \mathrm{E}\big[Z^{\xi_\varepsilon}\big],\end{aligned}$$

In order to prove (4.1) it therefore suffices to show that $\mathrm{E}\big[Z^{\xi_\varepsilon}\big]$ is finite. Noting that $r \mapsto (\min_i r_i)^{-\xi_\varepsilon}$ and $r \mapsto (\max_i r_i)^{\xi_\varepsilon}$ are non-increasing respectively non-decreasing functions on $\mathbb{R}^n$, Lemma 4.1 ii) yields

$$\mathrm{E}\big[Z^{\xi_\varepsilon}\big] \leq \mathrm{E}\left[(\min_i R_i)^{-\xi_\varepsilon}\right] \cdot \mathrm{E}\left[(\max_i R_i)^{\xi_\varepsilon}\right].$$

Note that

$$\mathrm{P}\left[\max_i R_i > t\right] = \sum_{k=1}^n \binom{n}{k} \cdot \mathrm{P}\left[R_1 > t\right]^k \cdot \big(1 - \mathrm{P}[R_1 > t]\big)^{n-k} \leq c \cdot t^{-\beta + \frac{\varepsilon}{2}}$$

for some $c > 0$. Therefore, we have

$$\mathrm{E}\left[(\max_i R_i)^{\xi_\varepsilon}\right] = \int_0^\infty \mathrm{P}\left[(\max_i R_i)^{\xi_\varepsilon} > t\right] dt \leq 1 + c \int_1^\infty t^{-\frac{\beta - \frac{\varepsilon}{2}}{\xi_\varepsilon}} dt < +\infty.$$

Analogously, $\mathrm{P}\big[(\min_i R_i)^{-1} > t\big] = \mathscr{O}\big(t^{-\gamma + \frac{\varepsilon}{2}}\big)$, and $\mathrm{E}\big[(\min_i R_i)^{-\xi_\varepsilon}\big] < +\infty$. Therefore, $\mathrm{E}\big[Z^{\xi_\varepsilon}\big] < \infty$ as required. ∎

5. Acknowledgments. The authors are grateful to the referees and editors for comments that led to substantial improvements of the manuscript.

REFERENCES

[1] K.M. Anstreicher, J. Ji, F.A. Potra, and Y. Ye. Probabilistic analysis of an infeasible-interior-point algorithm for linear programming. *Math. Oper. Res.*, 24(1):176–192, 1999.

[2] A. Blum and J. Dunagan. Smoothed analysis of the perceptron algorithm for linear programming. *SODA'02*, pages 905–914, 2002.

[3] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and Real Computation.* Springer-Verlag, 1998.

[4] K.H. Borgwardt. Some distribution–independent results about the asymptotic order of the average number of pivot steps of the simplex method. *Mathematics of Operations Research*, 7:441–462, 1982.

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

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

[7] F. Cucker and J. Peña. A primal-dual algorithm for solving polyhedral conic systems with a finite-precision machine. *SIAM Journal on Optimization*, 12:522–554, 2002.

[8] F. Cucker and M. Wschebor. On the expected condition number of linear programming problems. *Numer. Math.*, 94:419–478, 2002.

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

[10] J. Dunagan and S. Vempala. A simple polynomial-time rescaling algorithm for solving linear programs. *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 315–320, 2004.

[11] H. Federer. *Geometric Measure Theory*. Springer-Verlag, 1969.

[12] J.-L. Goffin. *On the finite convergence of the relaxation method for solving systems of inequalities*. PhD thesis, University of California, Berkeley, 1971.

[13] J.-L. Goffin. The relaxation method for solving systems of linear inequalities. *Mathematics of Operations Research*, 5:388–414, 1980.

[14] R. Hauser and T. Müller. Algebraic tail decay of condition numbers for random conic systems under a general family of input distributions. Numerical Analysis Group Research Report NA-06/01, Oxford University Computing Laboratory, 2006.

[15] P. Huhn and K.H. Borgwardt. An upper bound for the average number of iterations required in phase II of an interior-point-method. In *Operations Research Proceedings 1997 (Jena)*, pages 19–24, Berlin, 1998. Springer.

[16] P. Huhn and K.H. Borgwardt. Interior-point methods: worst case and average case analysis of a phase-I algorithm and a termination procedure. *J. Complexity*, 18(3):833–910, 2002.

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

[18] R. Rosenblatt. *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms.* Spartan Books, 1962.

[19] Sheldon M. Ross. *Probability models in computer science*. Harcourt/Academic Press, 2002.

[20] S. Smale. On the average number of steps of the simplex method of linear programming. *Math. Program.*, 27:241–262, 1983.

[21] D. Spielman and S. Teng. Smoothed analysis: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51:385–463, 2004.

[22] M.J. Todd, L. Tunçel, and Y. Ye. Characterizations, bounds and probabilistic analysis of two complexity measures for linear programming problems. *Math. Program.*, 90:59–69, 2001.