ALGEBRAIC TAIL DECAY OF CONDITION NUMBERS FOR RANDOM CONIC SYSTEMS UNDER A GENERAL FAMILY OF INPUT DISTRIBUTIONS

RAPHAEL HAUSER* AND TOBIAS MÜLLER†

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 distribution tails of $\mathscr{C}(A)$. We study an unprecedently general class of probability models for the random input matrix A and show that the tails of the Goffin-Cheung-Cucker condition number decay at algebraic rates. Furthermore, the exponent naturally emerges when applying a theory of *uniform absolute continuity* – which we also develop in this paper – to the distribution of A. We then develop similar results for Renegar's condition number. Finally we discuss lower bounds on the tail decay of $\mathscr{C}(A)$ 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, probabilistic 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 solutions at all. In other words, the existence of x such that $Ax < 0$ yields a certficate 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 $(\mathrm{D}(A))$ and $(\mathrm{P}(A))$ 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 general matrix A. 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 the *real number*

*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.

†Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, United Kingdom, (muller@stats.ox.ac.uk). This author was partially supported by EPSRC, the Department of Statistics, Bekker-la-Bastide fonds, Dr. Hendrik Muller's Vaderlandsch fonds, and Prins Bernhard Cultuurfonds

Turing machine model (see [1]) the complexity of the same algorithms is typically bounded by a linear function of $\log \mathscr{C}(A)$, i.e., the worst case running time is of order

$$\mathscr{O}\big(\log \mathscr{C}(A)\big), \tag{1.1}$$

where $\mathscr{C}(A)$ is either Renegar's condition number $\mathscr{C}_R(A)$ [10], the Goffin-Cheung-Cucker condition number $\mathscr{C}_{GCC}(A)$ [8, 9, 2] or another related concept of geometric measure, and where the multiplicative constant is polynomial in the matrix dimensions n and m. See [10] 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*, see [4] where a bound of $\mathscr{O}\big((\log \mathscr{C}(A))^3\big)$ was established for an interior-point method. 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.

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 no 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 [7] has a $\mathscr{O}(\mathscr{C}_{GCC}(A))$ probabilistic complexity, where $\mathscr{C}_{GCC}(A)$ is the Goffin-Cheung-Cucker condition number, and the classical perceptron algorithm [11] has a $\mathscr{O}(\mathscr{C}_{GCC}(A)^2)$ deterministic complexity.

One of the main theorems of this paper establishes that when A is a random matrix whose rows are i.i.d. and have an absolutely continuous distribution on the $(m-1)$-dimensional unit sphere, then

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

where α is a parameter that depends on the distribution of the rows of A and $f(t) = \tilde{\mathscr{O}}(t^{-\alpha})$ means that $f(t) = \mathscr{O}(t^{-\alpha+\varepsilon})$ for all $\varepsilon > 0$. Similar results are established for the condition number $\mathscr{C}_R(A)$, see Theorems 5.1 and 7.2. Such results are interesting because together with complexity estimates of the form (1.1) they imply that when random input data A are fed to interior-point and ellipsoid algorithms for the conic feasibility problem, then the distributions of the (random) running times RT have tails that satisfy

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

when t is large enough. In other words, if $\alpha > 0$ then extremely long running times are exponentially rare and thus not observable in practice. The bound (1.3) also implies that for $\alpha > 0$ all moments of RT are finite. This can be used to establish that the conic feasibility problem is strongly polynomial *on average* for all input distributions with $\alpha > 0$. For the sake of brevity we will not discuss this further, but the arguments are similar to the special case where A has uniformly distributed unit row vectors, discussed in [3].

Special cases of tail decay results of the form (1.2) were already obtained in [5, 6] and [3], however for much more restricted families of distributions for which α is always equal to 1: Cucker-Wschebor [5] considered the case where A is a matrix with i.i.d. Normal entries, Dunagan-Spielman-Teng [6] the case where A has a multivariate normal distribution centred at a given matrix and Cheung-Cucker-Hauser analysed the case where A has i.i.d. rows that are uniformly distributed on the sphere. All of these cases fall under the framework of Example 1 of Section 6. In the present paper we extend the family of input models considerably: In the case of $\mathscr{C}_{GCC}$, our random input matrices have i.i.d. rows that are absolutely continuous with respect to the uniform measure on the $(m-1)$-dimensional sphere S^{m-1}. In such measures an event has probability zero if its probability under the uniform measure is also zero. To derive bounds of the form (1.2) we need a slightly stronger condition that allows us to give bounds on the probability of an event if its probability under the uniform measure is small. This leads to a notion of *uniformly absolutely continuous measure* and an associated *smoothness parameter* that seems to be new. 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 $1-(m-1)\alpha$. We develop some of the relevant theory in the appendix of this paper. A related family of distributions is used in the context of $\mathscr{C}_R$, but since this condition number is not invariant under row scaling, the rows of A are scaled with i.i.d. positive random variables R_i whose tail decay rates now play a similar role as α. In the last two sections we discuss a further tightening of the tail decay bounds of $\mathscr{C}_{GCC}$, and we prove a lower bound which is then used to show that there exist absolutely continuous distributions with $\alpha = 0$ for which $\mathrm{P}[\mathscr{C}_{GCC}(A) > t]$ does not decay at any algebraic rate.

2. Notation. 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. Let

$$I_k(\rho) := \int_0^\rho \sin^k \tau d\tau.$$

Then it can easily be shown that

$$I_m(\pi) I_{m-1}(\pi) = \frac{2\pi}{m}, \qquad \forall m \geq 1, \tag{2.1}$$

$$V_m(\mathrm{B}_r) = 2r^m \prod_{j=2}^{m} I_j(\pi), \quad m \geq 2, \qquad \text{and} \tag{2.2}$$

$$A_{m-1}\big(\mathrm{cap}(p,\rho)\big) = 2\pi I_{m-2}(\rho) \prod_{j=1}^{m-3} I_j(\pi), \quad m \geq 2, \tag{2.3}$$

where V_m and A_{m-1} denote the standard volume and area in $\mathbb{R}^m$ and on S^{m-1} respectively, i.e., the Lebesgue and Hausdorff measures, B_r denotes the open ball of radius r in $\mathbb{R}^m$ centered at the origin, and $\mathrm{cap}(p, \rho)$ denotes the circular cap with half opening angle ρ which is centered at p.
The uniform measure of a circular cap can therefore be expressed as

$$\nu_{m-1}\big(\mathrm{cap}(p, \rho)\big) = I_{m-2}(\rho)/I_{m-2}(\pi).$$

Using integration by parts and induction on k it is easy to show that

$$\frac{I_k(\pi)\sqrt{k}}{2} \geq 1 \qquad \forall k \in \mathbb{N}. \tag{2.4}$$

3. Condition Numbers for 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.

Considering $\arccos(t)$ as a function from $[-1, 1]$ into $[0, \pi]$, both cos and arccos are decreasing functions. Let a_i be the ith row of A. Denote by $\theta_i(A, x)$ the angle between a_i and x, that is, $\arccos(a_i \cdot x/\|a_i\|\|x\|)$. Let $\theta(A, x) = \min_{1 \leq i \leq n} \theta_i(A, x)$ and $\bar{x}$ be any vector in $\mathbb{R}^m \setminus \{0\}$, s.t. $\theta(A) = \theta(A, \bar{x}) = \sup_{x \in \mathbb{R}^m} \theta(A, x)$. A compactness argument shows that such a vector $\bar{x}$ exists. The Cheung-Cucker [2] condition number $\mathscr{C}_{GCC}(A)$ is defined as

$$\mathscr{C}_{GCC}(A) = |\cos(\theta(A))|^{-1}.$$

$\mathscr{C}_{GCC}(A)$ is a generalisation of Goffin's condition number [8, 9] which was defined for strictly feasible A only. It is not difficult to see that A is strictly feasible iff $\theta(A) > \pi/2$, ill-posed iff $\theta(A) = \pi/2$ and infeasible iff $\theta(A) < \pi/2$. Note that since $\mathscr{C}_{GCC}(A)$ is defined purely in terms of angles between vectors, this condition number is invariant under positive scaling of the rows of A. Hence, 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 $(\mathrm{P}(A))$ has a strict solution if and only if $(\mathrm{D}(A))$ has no nontrivial solution and vice versa. Let us also recall that $\|A\|_{1,\infty} := \sup_{x \in \mathbb{R}^n \setminus \{0\}} \|Ax\|_\infty / \|x\|_1$. Let

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

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 [10] 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 $\mathscr{C}_{GCC}(A)$, $\mathscr{C}_R(A)$ *does* depend on the scaling of the rows of A. Cheung-Cucker [2] established the following inequalities linking the two condition numbers,

$$\frac{1}{\sqrt{m}} \cdot \mathscr{C}_{GCC}(A) \leq \mathscr{C}_R(A) \leq \frac{\|A\|_2}{\min_i \|a_i\|_2} \cdot \mathscr{C}_{GCC}(A). \tag{3.1}$$

4. 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 5, 7 and 8. 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.

Lemma 4.1. *For any $\mathscr{B}$-measurable set A with $\nu(A) > 0$ and for any $\delta > 0$ there exists $B \subseteq A$ such that $0 < \nu(B) \leq \delta$.*

Proof. Since $(E, \mathcal{O})$ is sigma-compact and ν is sigma-finite we may assume w.l.o.g. that $\nu(A) < \infty$ and $A \subseteq C$ for some compact set C. We claim that for all $x \in E$ there exist an open neighbourhood O_x of x such that $\nu(O_x \cap A) \leq \delta$. In fact, since $(E, \mathcal{O})$ has a locally countable basis, there exists a nested countable collection $\{O_{x,n} : n \in \mathbb{N}\}$ of open neighbourhoods of x such that $\{x\} = \bigcap_{n \in \mathbb{N}} O_{x,n}$, and since ν is atom-free we have

$$0 = \nu\left(\bigcap_{n \in \mathbb{N}} O_{x,n} \cap A\right) = \lim_{n \to \infty} \nu(O_{x,n} \cap A).$$

The sets $\{O_x : x \in C\}$ form an open cover of C. By compactness there exists a finite subcover $O_{x_1} \cup \cdots \cup O_{x_m} \supseteq C$. Writing $B_i := O_{x_i} \cap A$, we find $0 < \nu(A) \leq \sum_{i=1}^m \nu(B_i)$, so that at least one of the sets B_i satisfies $0 < \nu(B_i) \leq \delta$. □

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 4.2. *The limit*

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

is well-defined and takes a value in the interval $[0, 1]$.

Proof. For all $\delta \in (0, 1)$ let

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

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 decreasing in δ, 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 4.1 there exists a $\mathscr{B}$-measurable set $B_\delta \subseteq \{f > c\}$ such that $0 < \nu(B_\delta) \le \delta$ for all $\delta > 0$. Finally, since $\mu(B_\delta) = \int_{B_\delta} f d\nu > c\nu(B_\delta)$, we have

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

Letting $\delta \to 0$ now establishes the result. □

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

Proposition 4.3. *The following statements are equivalent:*

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

ii) α *is the smallest nonnegative real 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* $0 < \nu(B_{\varepsilon,\delta}) \le \delta$, *yet* $\nu(B_{\varepsilon,\delta})^{\alpha+\varepsilon} \le \mu(B_{\varepsilon,\delta})$,

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

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

A further characterisation of α is given by the following result whose proof is given in the appendix. Here we use the convention that $-\infty/-\infty := 0/0 := 1$.

Proposition 4.4. $\alpha_\nu(\mu) = \liminf_{n\to\infty} \frac{\ln(\mu(\{f>n\}))}{\ln(\nu(\{f>n\}))}$.

Absolute continuity tells us that all ν-null-sets must be μ-null-sets, but it does not tell us that $\mu(B)$ is small when $\nu(B)$ is small but positive. However, if $\alpha > 0$ then Proposition 4.3 gives uniform upper bounds on $\mu(B)$ in terms of $\nu(B)$. In this case we say that μ is *uniformly* ν-absolutely continuous. Furthermore, for smaller α the variation of μ in terms of ν is larger. We call $\alpha_\nu(\mu)$ the *smoothness parameter* of μ with respect to ν.

5. Tail Events of the Goffin-Cheung-Cucker Number. In this section we will analyse the tail behaviour of the Goffin-Cheung-Cucker number $\mathscr{C}_{GCC}(A)$. To derive any meaningful information about $\mathscr{C}_{GCC}(A)$, we require a family of distributions for the random $n \times m$ matrix A. Each row vector of A determines a constraint in the system of linear inequalities $Ax \le 0$. It is therefore natural to consider matrices $A = [\,X_1 \, \dots \, X_n\,]^{\mathrm{T}}$ where the X_i are i.i.d. m-dimensional random vectors. Furthermore, since $\mathscr{C}_{GCC}(A) = \mathscr{C}_{GCC}(DA)$ for any strictly positive diagonal matrix D, we may restrict the model to random vectors X_i on the $(m-1)$-dimensional unit sphere S^{m-1} without losing any generality.

We endow S^{m-1} with the subspace topology inherited from $\mathbb{R}^m$ and the associated Borel σ-algebra $\mathscr{B}$. The uniform measure ν_{m-1} on S^{m-1} is then a Borel measure. Let μ be any probability measure on S^{m-1} that is ν_{m-1}-absolutely continuous and has smoothness parameter $\alpha := \alpha_\nu(\mu) > 0$. The random matrix A is then well-defined by choosing the law $\mathscr{L}(X_i) = \mu$ for its rows. We also refer to α as the *smoothness* *parameter of A.*

Next we will derive upper bounds on $\mathrm{P}[\mathscr{C}_{GCC}(A) \geq t]$. In [3] it was shown that in the special case where $\mu = \nu_{m-1}$ the bound

$$\mathrm{P}[\mathscr{C}_{GCC}(A) \geq t] \leq c(n,m)t^{-1} \tag{5.1}$$

holds, where $c(n,m) := \binom{n}{m} 2m^{\frac{5}{2}}$. Theorem 5.2 below will yield a tool that allows one to boost this results to the general case discussed here. Recall from Section 1 that we write $f(t) = \tilde{\mathscr{O}}(t^{-\alpha})$ if $f(t) = \mathscr{O}(t^{-\alpha+\varepsilon})$ for all $\varepsilon > 0$.

THEOREM 5.1. *The tails of the Goffin-Cheung-Cucker condition number of an arbitrary random $n \times m$ matrix A with smoothness parameter α are bounded by*

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

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

$$W := \left\{(x_1, \dots, x_n) \in (\mathrm{S}^{m-1})^n : \mathscr{C}_{GCC}([\, x_1 \cdots x_n \,]^T) > t\right\}.$$

Let $X_1, \dots, X_n$ be i.i.d. random unit vectors with law $\mathscr{L}(X_i) = \mu$ and let $U_1, \dots, U_n$ be i.i.d. uniformly distributed unit vectors, $\mathscr{L}(U_i) = \mathscr{U}(\mathrm{S}^{m-1})$. From (5.1) we know that $\mathrm{P}[(U_1, \dots, U_n) \in W] \leq c(n,m)t^{-1}$. Furthermore, Theorem 5.2 below establishes that for all $\varepsilon > 0$ there exists $t_\varepsilon > 0$ such that for $t \geq t_\varepsilon$,

$$\otimes^n \mu(W) \leq \left(\otimes^n \nu_{m-1}(W)\right)^{\alpha-\varepsilon}. \tag{5.2}$$

Therefore, for $t \geq t_\varepsilon$,

$$\begin{aligned}\mathrm{P}\left[\mathscr{C}_{GCC}(A) > t\right] &= \mathrm{P}\left[(X_1, \dots, X_n) \in W\right] = \otimes^n \mu(W) \\ &\leq \left(\otimes^n \nu_{m-1}(W)\right)^{\alpha-\varepsilon} = \left(\mathrm{P}\left[(U_1, \dots, U_n) \in W\right]\right)^{\alpha-\varepsilon} \leq c(n,m)^{\alpha-\varepsilon} t^{-\alpha+\varepsilon}.\end{aligned}$$

□

It remains to prove the claim (5.2), which follows readily by applying Theorem 5.2 below $n-1$ times. 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 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 $\alpha_i := \alpha_{\nu_i}(\mu_i)$. Finally, let $\nu_1 \otimes \nu_2$ and $\mu_1 \otimes \mu_2$ be the corresponding product measures. It is well known that then $\nu_1 \otimes \nu_2$ 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 5.2. *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 4.3 ii), for all $\delta > 0$ and $(i = 1, 2)$ there exist $\mathscr{B}_i$-measurable sets $B^i_{\varepsilon,\delta}$ such that $0 < \nu_i(B^i_{\varepsilon,\delta}) \le \delta$ and $\nu_i(B^i_{\varepsilon,\delta})^{\alpha_i+\varepsilon} \le \mu_i(B^i_{\varepsilon,\delta})$. Let $\delta_0 \in (0, 1)$ be chosen such that

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

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

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

and

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

Proposition 4.3 ii) now implies $\alpha_{\nu_1\otimes\nu_2}(\mu_1 \otimes \mu_2) \le \alpha_1$. It remains to prove that

$$\alpha_1 \le \alpha_{\nu_1\otimes\nu_2}(\mu_1 \otimes \mu_2). \tag{5.3}$$

For this purpose, let $0 < \varepsilon < \alpha_1$, and let $B := (C_1 \times D_1) \cup \cdots \cup (C_N \times D_N)$ 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 $\eta := \nu_1 \otimes \nu_2(B)$ and

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

For all k let $A_k := \bigcup_{i\in I_k} D_i$ and $B_k := B \cap (E_1 \times A_k)$. For $k \ge 1$ it must then be true that

$$\eta \ge \nu_1 \otimes \nu_2(B_k) = \sum_{i\in I_k} \nu_1(C_i)\nu_2(D_i) \ge \eta^{1-(k-1)\frac{\varepsilon}{2}} \sum_{i\in I_k} \nu_2(D_i) = \eta^{1-(k-1)\frac{\varepsilon}{2}} \nu_2(A_k), \tag{5.4}$$

which establishes that $\nu_2(A_k) \le \eta^{(k-1)\frac{\varepsilon}{2}}$. The assumption that $\alpha_1 \le \alpha_2$ together with Proposition 4.3 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 $\mu_2(A_k) \le c_2 \eta^{(\alpha_1-\frac{\varepsilon}{2})(k-1)\frac{\varepsilon}{2}}$. It follows that

$$\begin{aligned}\mu_1 \otimes \mu_2(B_k) = \sum_{i\in I_k} \mu_1(C_i)\mu_2(D_i) &\le c_1(\eta^{1-k\frac{\varepsilon}{2}})^{\alpha_1-\frac{\varepsilon}{2}} \sum_{i\in I_k} \mu_2(D_i) \\ &= c_1(\eta^{1-k\frac{\varepsilon}{2}})^{\alpha_1-\frac{\varepsilon}{2}} \mu_2(A_k) \le c_1c_2\eta^{(\alpha_1-\frac{\varepsilon}{2})(1-\frac{\varepsilon}{2})} \le c_1c_2\eta^{\alpha_1-\varepsilon}.\end{aligned}$$

For $k = 0$ we find similarly,

$$\mu_1 \otimes \mu_2(B_0) = \sum_{i\in I_0} \mu_1(C_i)\mu_2(D_i) \le c_1\eta^{\alpha_1-\frac{\varepsilon}{2}} \mu_2(A_0) \le c_1c_2\eta^{\alpha_1-\varepsilon},$$

so that

$$\mu_1\otimes\mu_2(B) = \sum_k \mu_1\otimes\mu_2(B_k) \le (2 + \lfloor 2/\varepsilon \rfloor)\, c_1c_2\eta^{\alpha_1-\varepsilon} = (2 + \lfloor 2/\varepsilon \rfloor)\, c_1c_2\nu_1\otimes\nu_2(B)^{\alpha_1-\varepsilon}. \tag{5.5}$$

Next, let B 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 $\nu_1 \otimes \nu_2(B) = \inf_{\mathcal{B}} \sum_{B' \in \mathcal{B}} \nu_1 \otimes \nu_2(B')$, where the infimum is over all countable collections $\mathcal{B} \subset \mathscr{B}_1 \times \mathscr{B}_2$ that satisfy $B \subseteq \bigcup_{B' \in \mathcal{B}} B'$. Hence, there exists a countable family $(B_i)_{\mathbb{N}} \subset \mathscr{B}_1 \times \mathscr{B}_2$ such that $B \subseteq \bigcup_i B_i$ and

$$\nu_1 \otimes \nu_2(B) \geq (1-\varepsilon)\nu_1 \otimes \nu_2 \left(\bigcup_{i=1}^{\infty} B_i \right). \tag{5.6}$$

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

$$\mu_1 \otimes \mu_2 \left(\bigcup_{i=1}^{i_0} B_i \right) \geq (1-\varepsilon)\mu_1 \otimes \mu_2 \left(\bigcup_{i=1}^{\infty} B_i \right). \tag{5.7}$$

Therefore, we have

$$\begin{aligned}
\mu_1 \otimes \mu_2(B) &\leq \mu_1 \otimes \mu_2 \left(\bigcup_{i=1}^{\infty} B_i \right) \\
&\overset{(5.7)}{\leq} \frac{1}{1-\varepsilon} \cdot \mu_1 \otimes \mu_2 \left(\bigcup_{i=1}^{i_0} B_i \right) \\
&\overset{(5.5)}{\leq} \frac{\left(2 + \lfloor \frac{2}{\varepsilon} \rfloor\right) c_1 c_2}{1-\varepsilon} \cdot \left(\nu \left(\bigcup_{i=1}^{i_0} B_i \right) \right)^{\alpha_1 - \varepsilon} \\
&\leq \frac{\left(2 + \lfloor \frac{2}{\varepsilon} \rfloor\right) c_1 c_2}{1-\varepsilon} \cdot \left(\nu \left(\bigcup_{i=1}^{\infty} B_i \right) \right)^{\alpha_1 - \varepsilon} \\
&\overset{(5.6)}{\leq} \frac{\left(2 + \lfloor \frac{2}{\varepsilon} \rfloor\right) c_1 c_2}{(1-\varepsilon)^{1+\alpha_1-\varepsilon}} \cdot (\nu(B))^{\alpha_1 - \varepsilon}.
\end{aligned}$$

Since this holds for all $\varepsilon \in (0, \alpha)$ and $\mathscr{B}$-measurable sets B, Proposition 4.3 iv) implies that (5.3) holds. □

6. Examples. The class of input models for the CFP analysed in Section 5 is considerably more general than the family of distributions with continuous density functions. To illustrate this, we will now give a few nontrivial examples. The proofs of most statements made in this section are sketched in the appendix (see Section 11). The reference measure ν is chosen as the uniform measure ν_{m-1} on S^{m-1}.

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

We remark that when μ as in Example 1, Theorem 5.1 shows that $\mathrm{P}[\mathscr{C}_{GCC}(A) > t] = \tilde{\mathscr{O}}(t^{-1})$. As mentioned in the introduction, special cases of this result were already established in [5, 6] and [3].

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 next 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 and $\varsigma \in (0, m-1)$.*

i) If μ is a ν_{m-1}–absolutely continuous probability measure on S^{m-1} such that $d\mu/d\nu_{m-1}$ is essentially bounded by g then μ is has smoothness parameter $\alpha \geq 1 - \varsigma/(m-1)$.

ii) If μ is the ν_{m-1}–absolutely continuous probability measure on S^{m-1} defined by the density function $d\mu/d\nu_{m-1} \equiv g/\int_{\mathrm{S}^{m-1}} g(x)\nu_{m-1}(dx)$ then μ has smoothness parameter $\alpha = 1 - \varsigma/(m-1)$.

In passing, let us note that if $d\mu/d\nu_{m-1}$ has a pole of degree $\varsigma \geq m-1$ then μ cannot be a finite measure and hence not a probability measure either. Hence, 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 ν_{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.

Not all ν_{m-1}-absolutely continuous measures are *uniformly* ν_{m-1}-absolutely continuous. The following construction yields a counterexample:

EXAMPLE 3. *By virtue of Example 2 we know that there exists a sequence $(\mu_i)_{\mathbb{N}}$ of ν_{m-1}-absolutely continuous probability measures with smoothness parameters $\alpha(\mu_i, \nu_{m-1}) \leq i^{-1}$. For all $i \in \mathbb{N}$ let X_i be a random vector on S^{m-1} with law $\mathscr{L}(X_i) \sim \mu_i$, 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 (e.g., a Poisson variable). Then the distribution $\mu = \mathscr{L}(X_N)$ of the random vector X_N is ν_{m-1}-absolutely continuous with smoothness parameter $\alpha(\mu, \nu_{m-1}) = 0$.*

7. Tail Events of the Renegar Number. In the present section we take the analysis of Section 5 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 $\mathscr{C}_{GCC}(A)$, 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 $\mathscr{C}_{GCC}(A)$, 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 4. *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 $R_i^2 \sim \chi_m^2$ 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 $y \leq z \in \mathbb{R}^k$ (component-wise) implies $f(y) \geq f(z)$ (respectively $f(y) \leq f(z)$). The following lemma is a standard result. For a proof of part i) see e.g. [12]. The proof of part ii) is completely analogous.

LEMMA 7.1. *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

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

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

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

THEOREM 7.2. *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 (3.1) and $\|A\|_2 \leq \max_i \|A_i\|_2 \cdot \sqrt{n}$, it suffices to show that

$$\mathrm{P}\big[Z \cdot \mathscr{C}_{GCC}(X) > t\big] = \mathscr{O}(t^{-\xi_\varepsilon}). \tag{7.1}$$

By virtue of Theorem 5.1 there exists a constant c_ε such that for all $t > 0$,

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

Writing f_Z for the density function of Z, we have

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

In order to prove (7.1) it remains to establish 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 7.1 ii) yields

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

But note that

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

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

$$\begin{aligned}\mathrm{E}\left[(\max_i R_i)^{\xi_\varepsilon}\right] &= \int_0^\infty \mathrm{P}\left[(\max_i R_i)^{\xi_\varepsilon}>t\right]dt\\ &\le c_\varepsilon\int_0^\infty t^{-\frac{\beta-\frac{\varepsilon}{2}}{\xi_\varepsilon}}dt\\ &< +\infty.\end{aligned}$$

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. ◻

8. Tightness of Bounds. The upper bound in Theorem 5.1 already establishes that if the random input $A=[\,x_1\;\dots\;x_n\,]^{\mathrm{T}}\in\mathbb{R}^{n\times m}$ of a random family of conic feasibility problems has smoothness parameter $\alpha>0$, then the random running time RT of several interior-point methods for this family of problems has exponential tail-decay $\mathrm{P}[RT>t]<\exp(-\gamma t)$ for some $\gamma>0$. This result was obtained through a simple mechanism provided by Theorem 5.2 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. However, the decay rates of Theorem 5.1 are not tight and can be further improved at the expense of working a bit harder. In this section we will show that when $n\ge m$,

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

and that there exist measures for which

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

We conjecture that the upper bound can be further improved to

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

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. ν_{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 [3] show the inclusion of events

$$\{\mathscr{C}_{GCC} > t\} \subseteq \{\exists\, S \in \mathscr{P}_m \text{ s.t. } \|U_S\|_2 > t\}. \tag{8.1}$$

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 $\mathscr{C}_{GCC}(A) > t$ then there must exist $S \in \mathscr{P}_m$ such that

$$\sum_{i=1}^{m} \left|\frac{1}{\langle X_i, Y_i\rangle}\right| \geq \|U_S\|_2 > t,$$

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

$$\mathrm{P}[\mathscr{C}_{GCC}(A) > t] \leq m\binom{n}{m} \cdot \mathrm{P}\left[|\langle Y(X-1,\dots,X_{m-1}), X_m\rangle| < \frac{m}{t}\right], \tag{8.2}$$

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

Lemma 8.1. *If X_i $(i = 1,\dots,m-1)$ are i.i.d. random unit vectors in $\mathbb{R}^m$ with ν_{m-1}-absolutely continuous distribution $\mathscr{L}(X_i) = \mu$ of smoothness parameter $\alpha = \alpha_{\nu_{m-1}}(\mu)$, then the distribution $\mathscr{L}(Y(X_1,\dots,X_{m-1}))$ is ν_{m-1}-absolutely continuous with the same smoothness parameter value α.*

Proof. This follows quite straightforwardly from Theorem 5.2: Let $V_1,\dots,V_{m-1}$ be uniformly distributed on S^{m-1}, and note that 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

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

where c_ε is chosen as in Proposition 4.3 iii) applied to the measure $\otimes^{m-1}\mu$. □

PROPOSITION 8.2. *Let X, Y be independent random vectors with ν_{m-1}-absolutely continuous distributions on S^{m-1}, and such that the smoothness parameter takes the same value $\alpha > 0$ for both. 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,

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

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 $\alpha \le 1/2$. For $r > 0$ let $\rho(r) := 2\arcsin(r/2)$. For a fixed $r > 0$ let Let $x_1, \dots, x_N \in \mathrm{S}^{m-1}$ be chosen [1] so that

$$\mathrm{cap}\,(x_i, \rho(r/2)) \cap \mathrm{cap}\,(x_j, \rho(r/2)) = \emptyset \quad (i \ne j), \tag{8.3}$$

$$\mathrm{S}^{m-1} \subseteq \bigcup_i \mathrm{cap}\,(x_i; \rho(r)). \tag{8.4}$$

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

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

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

$$c_1 \cdot r^{m-1} \le \nu_{m-1}\left(\mathrm{cap}(x, \rho(r))\right) \le c_2 \cdot r^{m-1}.$$

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

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

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,

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

We can conclude that

$$\mathrm{P}\left[|\langle X, Y\rangle| < r\right] \le \sum_{ij \in E(G)} p_i p_j, \tag{8.5}$$

[1]This can be achieved by iteratively adding points x_i as long as (8.3) can be satisfied. Since the area of S^{m-1} is finite, this process must end after $N < \infty$ choices have been made. Criterion (8.4) 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 $\|x - x_i\|_2 \le \|x - y\|_2 + \|y - x_i\|_2 < r$, 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 $y \in \mathrm{S}^{m-1}$, $\nu_{m-1}(\{x : |\langle x, y\rangle| < r\}) = 1 - 2 \cdot I_{m-2}(\arccos(r))/I_{m-2}(\pi)$, so that there exist constants $0 < d_1 < d_2$ such that for all $r \in (0,1)$,

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

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

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

Without loss of generality we may assume that the x_i were ordered so that $p_1 \geq p_2 \geq \cdots \geq p_N$. 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 $\geq D+1$. 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, lk, and note that $2p_ip_k \geq p_ip_j + p_lp_k$.

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

$$\sum_{ij \in E(G)} p_ip_j \leq 2 \sum_{ij \in E(G')} p_ip_j \leq 2 \sum_l \sum_{i,j \in I_l} p_ip_j < \sum_l \mu(A_l)^2, \tag{8.6}$$

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

$$\nu_{m-1}(A_k) \leq \gamma \cdot r.$$

Let $\varepsilon > 0$, and set

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

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

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

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

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

as $N = \mathscr{O}(r^{-(m-1)})$ and $D = \Theta(r^{-(m-2)})$. Combining (8.5) and (8.6) with the above estimates we find

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

This shows that $\mathrm{P}[|\langle X, Y\rangle| < r] = \tilde{O}(r^{2\alpha})$, provided that $\alpha \leq 1/2$. 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 4.3, 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. □

Corollary 8.3. *If A is a random $n \times m$ matrix with smoothness parameter $\alpha > 0$ and $m \geq n$, then*

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

Proof. This follows immediately from inequality (8.2), Lemma 8.1 and Proposition 8.2. □

It is now natural to ask whether the upper bound on $\mathrm{P}[\mathscr{C}_{GCC}(A) > t]$ can be further improved. While we suspect that there is indeed room for further improvements when $m > 2$, Theorem 8.6 below establishes that in the case $m = 2$ the exponent of Theorem 8.3 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 $\alpha > 0$) $\mathrm{P}[\mathscr{C}_{GCC}(A) > t]$ does not decay faster than at an algebraic rate. Finally, a variant of Theorem 8.6 will be presented in Section 9 to show that for absolutely continuous input distributions with $\alpha = 0$ the tail probabilities $\mathrm{P}[\mathscr{C}_{GCC}(A) > t]$ do not decay at an algebraic rate in general. Before we can present these results, we need two lemmas.

Lemma 8.4. *Let X be a random variable on $\mathbb{R}$ with cumulative distribution function F_X and let Y be a random variable on $\mathbb{R}$ with cumulative distribution function $F_Y(x) = F_X(x)^\alpha$ with $0 < \alpha \leq 1$. Then*

i) $\mathrm{P}[Y \in B] \leq \mathrm{P}[X \in B]^\alpha$ *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 $\mathrm{P}[Y \in B] \leq 2^\alpha \cdot \mathrm{P}[Z \in B]^\alpha$ *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 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

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

where U is a random variable with uniform distribution on $[0,1]$, and $\mathrm{P}[Y \in B] = \mathrm{P}[U \in F_Y(B)]$. Setting $C := F_X(B)$ 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 of 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 $\{y : \alpha y^{\alpha-1} \geq c\}$, this can be shown via an argument similar to the proof of Lemma 11.1 in the appendix). 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 $B = F_X^{-1}([0,p]) = (-\infty, c]$. Part ii) is an immediate extension of the same argument. □

LEMMA 8.5. *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* ν_{m-1}*-absolutely continuous measure on* S^{m-1} *with smoothness parameter* $\alpha_{\nu_{m-1}}(\mu) = \alpha$.

Proof. Firstly, note that $\mathscr{L}([\, \sqrt{1-Z^2} \cdot V \;\; Z \,]) = \mathscr{L}(U) = \nu_{m-1}$. Lemma 8.4 shows that $\alpha_{\mathscr{L}(|Z|)}(\mathscr{L}(W)) = \alpha$, and by Theorem 5.2 this implies

$$\alpha_{\mathscr{L}([\, V \;\; Z \,])}\left(\mathscr{L}\left(\begin{bmatrix} V & W \end{bmatrix}\right)\right) = \min\left(1, \alpha_{\mathscr{L}(Z)}\left(\mathscr{L}(W)\right)\right) = \alpha.$$

It follows that for all $\mathscr{B}$-measurable $B \subseteq \mathrm{S}^{m-1}$ and $\varepsilon > 0$,

$$\begin{aligned} \mu(B) &= \mathrm{P}\left[\begin{bmatrix} \sqrt{1-W^2} \cdot V & W \end{bmatrix} \in B\right] \\ &= \mathrm{P}\left[\begin{bmatrix} V & W \end{bmatrix} \in B'\right] \\ &\leq c_\varepsilon \cdot \mathrm{P}\left[\begin{bmatrix} V & Z \end{bmatrix} \in B'\right]^{\alpha-\varepsilon} \\ &= c_\varepsilon \cdot \mathrm{P}\left[\begin{bmatrix} \sqrt{1-Z^2} \cdot V & Z \end{bmatrix} \in B\right]^{\alpha-\varepsilon} \\ &= c_\varepsilon \cdot \nu_{m-1}(B)^{\alpha-\varepsilon}, \end{aligned}$$

where c_ε is chosen as in Proposition 4.3 iii) and

$$B' := \left\{(x,z) \in \mathrm{S}^{m-2} \times [-1,1] : \begin{bmatrix} \sqrt{1-z^2} \cdot x & z \end{bmatrix} \in B\right\}.$$

Thus we see that $\alpha_{\nu_{m-1}}(\mu) = \alpha$ as claimed. □

THEOREM 8.6. *For any* $\alpha \in (0,1)$ *there exists a* ν_{m-1}*-absolutely continuous measure* μ *on* S^{m-1} *with smoothness parameter* α *and such that*

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

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

[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 $\langle x, p_m \rangle \geq (1/\sqrt{m-1} - (m-2)c)/\sqrt{m-1} = c$.

the random vector V and the random variable W respectively that were defined in Lemma 8.5, and let

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

for all i. 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 $\tilde{c} := (I_{m-3}(2\arcsin(c/4))/I_{m-3}(\pi))^m$, where I_k are the functions defined in Section 2. We remark that $\mathrm{P}[|Z| \leq t^{-1}] = \Omega(t^{-1})$. [3] By Lemma 8.4 ii) we therefore have

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

To prove our claim, it thus suffices to show that $B_t \subseteq \{\mathscr{C}(A) \geq t\}$. By the definition of $\mathscr{C}(A)$,

$$\mathscr{C}(A)^{-1} = \left|\cos\left(\sup_{x\in \mathrm{S}^{m-1}} \min_i \arccos\langle X_i, x\rangle\right)\right| = \left|\inf_{x\in\mathrm{S}^{m-1}} \max_i \langle X_i, x\rangle\right|,$$

where we used that arccos and cos are decreasing on $[0,\pi]$. Writing $x = [\,\sqrt{1-z^2}\cdot u\ \ z\,]$, 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

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

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

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

and consequently,

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

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

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 $-\min(1, m\alpha)$ is the exponent that corresponds to tight bounds for general random matrices A with smoothness parameter α.

Proposition 8.7. *For $\alpha > m^{-1}$ there exists a distribution μ with $\alpha_{\nu_{m-1}}(\mu) = \alpha$ and such that*

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

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

Proof. In [3] it was shown that in the case where the rows of A are i.i.d. uniformly distributed on S^{m-1}, there exists a lower bound $\mathrm{P}[\mathscr{C}(A) > t] = \Omega(t^{-1})$. Let X_i^0 $(i = 1, \dots, n)$ be i.i.d. random vectors with distribution ν_{m-1}, let X_i^1 $(i = 1, \dots, n)$ be i.i.d. random vectors with distribution μ on S^{m-1} with $\alpha_{\nu_{m-1}}(\mu) = \alpha$, and let N_i $(i = 1, \dots, n)$ be independent Bernoulli variables with parameter $1/2$. Then the random matrix

$$\tilde{A} = \begin{bmatrix} X_1^{N_1} & \dots & X_n^{N_n} \end{bmatrix}^{\mathrm{T}}$$

also has smoothness parameter α (as can be shown using an argument similar to Example 9), and

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

□

9. Tail Decay for the Case $\alpha = 0$. In the case where A has rows with ν_{m-1}-absolutely continuous distribution but not *uniformly* so, i.e., when $\alpha = 0$, the tail decay of $\mathscr{C}_{GCC}(A)$ is subalgebraic in general. In other words, although

$$\lim_{t \to \infty} \mathrm{P}[\mathscr{C}(A) > t] = 0,$$

there does not exist an exponent $\gamma > 0$ such that $\mathrm{P}[\mathscr{C}(A) > t] = \mathscr{O}(t^{-\gamma})$. This is established by the following result.

THEOREM 9.1. *There exists a random $n \times m$ matrix with i.i.d. ν_{m-1}-absolutely continuous rows $X_i \sim \mu$ and smoothness parameter $\alpha_{\nu_{m-1}}(\mu) = 0$ such that*

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

for all $\gamma > 0$.

Proof. We repeat the construction of Theorem 8.6 with a small modification. Let Z be as in Lemma 8.5, and for all $j \in \mathbb{N}$ let W_j be defined as in Lemma 8.5 when $\alpha = j^{-1}$. For $(i = 1, \dots, n)$ and $j \in \mathbb{N}$ let $W_{i,j}$ be i.i.d. copies of W_j, let $V_{i,j}$ be i.i.d. copies of the random vector V defined in Lemma 8.5, and let

$$X_{i,j} = \left[\sqrt{1 - W_{i,j}^2} \cdot V_{i,j} \quad W_{i,j}\right].$$

Finally, let $p_1, \dots, p_m \in \mathrm{S}^{m-1}$ and $c > 0$ be chosen as in the proof of Theorem 8.6, let N_i $(i = 1, \dots, n)$ be i.i.d. Poisson variables with parameter 1 and let

$$A = \begin{bmatrix} X_{1,N_1} & \dots & X_{n,N_n} \end{bmatrix}^{\mathrm{T}}.$$

For $t > 0$ consider the events

$$\begin{aligned} B_t &= \left\{ \|V_{i,N_i} - p_i\|_2 < \frac{c}{2}, \, W_{i,N_i} \leq \frac{1}{t} \; (i = 1, \dots, m) \right\}, \\ B_{t,j} &= B_t \cap \{N_i = j \; (i = 1, \dots, n)\}, \end{aligned}$$

and let $\tilde{c}$ be as in the proof of Theorem 8.6. Then $B_{t,j} \subset B_t \subseteq \{\mathscr{C}(A) \geq t\}$, where the second inclusion can be shown exactly like in the proof of Theorem 8.6, and

$$\begin{aligned} \mathrm{P}\left[B_{t,j}\right] &= \tilde{c}\left(\frac{\mathrm{e}^{-1}}{j!}\right)^{m} \cdot \mathrm{P}\left[W_j \leq t^{-1}\right]^{m} \\ &= \tilde{c}\left(\frac{\mathrm{e}^{-1}}{j!}\right)^{m} \cdot \mathrm{P}\left[|Z| \leq t^{-1}\right]^{\frac{m}{j}} \\ &= \Omega\left(t^{-\frac{m}{j}}\right). \end{aligned}$$

Since this holds for all j, the claim of the theorem is established. □

10. Conclusions. This paper shows that the distribution tails of the Goffin-Cheung-Cucker condition number and the Renegar condition number decays algebraically for a very large class of random matrices A. Furthermore, the rate of the tail decay is governed by the smoothness parameter of A, a simple and natural quantity defined by the distribution of A. Our findings explain why extremely long running times in ellipsoid and interior-point methods for the conic feasibility problem are exponentially rare on random input of the discussed kind, that is, CFP is “empirically” strongly polynomial for a large class of random input.

REFERENCES

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

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

[3] 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.

[4] 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.

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

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

[7] 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.

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

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

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

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

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

11. Appendix. In this appendix we fill in the missing proofs of Proposition 4.4 and of Examples 1–3.

Proof of Proposition 4.4

Proof. Let $B_n := \{f > n\}$, where $f = d\mu/d\nu$, and set

$$\beta := \liminf_{n\to\infty} \frac{\ln(\mu(B_n))}{\ln(\nu(B_n))}.$$

By definition of α we have $\beta \geq \alpha$. Now let $\delta > 0$ be an arbitrary small number and $A \in \mathscr{B}$ an arbitrary Borel-measurable set such that $0 < \nu(A) < \delta$. Then there exists $n \in \mathbb{N}$ such that $\nu(B_{n+1}) \leq \nu(A) < \nu(B_n)$. In the case where $\mu(B_{n+1}) \geq \mu(A \backslash B_{n+1})$, we have

$$\frac{\ln(\mu(A))}{\ln(\nu(A))} \geq \frac{\ln(2\mu(B_{n+1}))}{\ln(\nu(A))} \geq \frac{\ln(2\mu(B_{n+1}))}{\ln(\nu(B_{n+1}))} = \frac{\ln(\mu(B_{n+1})) + \ln(2)}{\ln(\nu(B_{n+1}))}. \tag{11.1}$$

On the other hand, in the case where $\mu(B_{n+1}) < \mu(A \setminus B_{n+1})$ we have

$$nu(B_{n+1}) < \nu(A \setminus B_{n+1}),$$

since $f|_{B_{n+1}} > f|_{B_n \setminus B_{n+1}}$, and hence,

$$\begin{aligned}\frac{\ln(\mu(A))}{\ln(\nu(A))} &\geq \frac{\ln(2\mu(A \setminus B_{n+1}))}{\ln(\nu(A))} \geq \frac{\ln(2\mu(A \setminus B_{n+1}))}{\ln(\nu(A \setminus B_{n+1}))} \\ &\geq \frac{\ln(2(n+1)\nu(A \setminus B_{n+1}))}{\ln(\nu(A \setminus B_{n+1}))} \geq \frac{\ln(2(n+1)\nu(B_n))}{\ln(\nu(B_n))}. \end{aligned} \tag{11.2}$$

But note that

$$(n+1)\nu(B_n) \leq \frac{n+1}{n}\mu(B_n \setminus B_{n+1}) + \mu(B_{n+1}) \leq 2\mu(B_n),$$

so that (11.2) implies

$$\frac{\ln(\mu(A))}{\ln(\nu(A))} \geq \frac{\ln(4\mu(B_n))}{\ln(\nu(B_n))} = \frac{\ln(\mu(B_n)) + \ln(4)}{\ln(\nu(B_n))}. \tag{11.3}$$

Inequalities (11.1) and (11.3) show

$$\frac{\ln(\mu(A))}{\ln(\nu(A))} \geq \min\left(\frac{\ln(\mu(B_{n+1})) + \ln(2)}{\ln(\nu(B_{n+1}))}, \frac{\ln(\mu(B_n)) + \ln(4)}{\ln(\nu(B_n))}\right).$$

Therefore, if $inf(\delta)$ is as in the proof of Theorem 4.2, we have

$$\alpha = \lim_{\delta\to 0} \inf(\delta) \geq \liminf_{n\to\infty} \frac{\ln(\mu(B_n))}{\ln(\nu(B_n))} = \beta.$$

□

In the remainder of this section, let η denote the standard Lebesgue measure on $\mathbb{R}^k$ and $\mathscr{B}$ the standard Borel σ-algebra of $\mathbb{R}^k$ completed with the subsets of Borel sets of ν-measure zero. The dimension of the space is usually clear from the context, so we drop the dependence on k in the notation. We write $\mathscr{B}_{m-1}$ for the standard

Borel σ-algebra and ν_{m-1} for the uniform probability measure on S^{m-1}. Also recall the functions $I_k(\rho)$ introduced in Section 2 and the associated volume and area formulae.

Proof of the Claims of Example 1

Proof. For all $B \in \mathscr{B}_{m-1}$ we have

$$\mu(B) = \int_B f(x)\nu_{m-1}(dx) \leq M \int_B \nu_{m-1}(dx) = M\nu_{m-1}(B).$$

Therefore, $\ln \mu(B) \leq \ln M + \ln \nu_{m-1}(B)$. If $\nu_{m-1}(B) < 1$ this implies

$$\frac{\ln \mu(B)}{\ln \nu_{m-1}(B)} \geq \frac{\ln M}{\ln \nu_{m-1}(B)} + 1,$$

and hence,

$$\alpha = \lim_{\delta \to 0} \inf \left\{ \frac{\ln \mu(B)}{\ln \nu_{m-1}(B)} : B \in \mathscr{B}_{m-1}, 0 < \nu_{m-1}(B) \leq \delta \right\} \geq 1.$$

It was also established in Theorem 4.2 that $\alpha \leq 1$, so that $\alpha = 1$. □

Lemma 11.1. *Let $\varsigma, \delta > 0$, $k \in \mathbb{N}$ and $r = \left(2 \prod_{j=2}^{k} I_j(\pi)/\delta\right)^{-1/k}$. Then the open ball $\mathrm{B}_r = \{y \in \mathbb{R}^k : \|y\| < r\}$ is a global maximiser of the optimization problem*

$$\max \left\{ \int_B \|x\|^{-\varsigma} \eta(dx) : B \in \mathscr{B}, \eta(B) \leq \delta \right\}. \tag{11.4}$$

Proof. First, let us observe that $\mathrm{B}_r \in \mathscr{B}$ and that (2.2) shows $\eta(\mathrm{B}_r) = \delta$. Hence, the set B_r is feasible for (11.4). Moreover, for all feasible $B \in \mathscr{B}$,

$$\begin{aligned}
\int_B \|y\|^{-\varsigma}\eta(dy) &= \int_{B\cap \mathrm{B}_r} \|y\|^{-\varsigma}\eta(dy) + \int_{B\setminus \mathrm{B}_r} \|y\|^{-\varsigma}\eta(dy) \\
&\leq \int_{B\cap \mathrm{B}_r} \|y\|^{-\varsigma}\eta(dy) + r^{-\varsigma}\eta(B \setminus \mathrm{B}_r) \\
&\leq \int_{B\cap \mathrm{B}_r} \|y\|^{-\varsigma}\eta(dy) + r^{-\varsigma}\eta(\mathrm{B}_r \setminus B) \qquad (11.5) \\
&\leq \int_{B\cap \mathrm{B}_r} \|y\|^{-\varsigma}\eta(dy) + \int_{\mathrm{B}_r \setminus B} \|y\|^{-\varsigma}\eta(dy) \\
&= \int_{\mathrm{B}_r} \|y\|^{-\varsigma}\eta(dy),
\end{aligned}$$

where the inequality

$$\eta(B \cap \mathrm{B}_r) + \eta(B \setminus \mathrm{B}_r) = \eta(B) \leq \delta = \eta(\mathrm{B}_r) = \eta(B \cap \mathrm{B}_r) + \eta(\mathrm{B}_r \setminus B)$$

was used in (11.5). □

Lemma 11.2. *Let $C_1, C_2 > 0$ and $0 < \iota < 1$. Then for $\delta < (\iota C_2/C_1)^{1/(1-\iota)}$, $x = 0$ maximises the function $h : x \mapsto C_1 x + C_2(\delta - x)^{\iota}$ over the interval $[0, \delta]$.*

Proof. We have

$$h'(x) = C_1 - C_2\iota(\delta - x)^{\iota-1}, \qquad \text{and}$$
$$h''(x) = C_2\iota(\iota-1)(\delta - x)^{\iota-2} < 0 \qquad \forall\, x < \delta.$$

Hence, h is strictly concave on $[0,\delta]$ and has a unique maximiser at

$$x^* = \delta - (C_1/(\iota C_2))^{1/(\iota-1)}.$$

Note that x^* becomes negative when $\delta < (\iota C_2/C_1)^{1/(1-\iota)}$, and then we have $h'(x) < 0$ for all $x \in [0,\delta]$. Therefore, $x = 0$ maximises h on $[0,\delta]$. □

PROOF OF THE CLAIMS OF EXAMPLE 2

Proof. Let $\varphi : D \to \mathrm{B}_1(\mathbb{R}^{m-1})$ and a_0 be chosen as in the remarks preceeding Example 2. Let $\phi = \varphi^{-1}$ and $D_{1/2} = \phi(\mathrm{B}_{1/2}(\mathbb{R}^{m-1}))$. We regard S^{m-1} as embedded in $\mathbb{R}^m$ and ϕ as a function from $\mathbb{R}^{m-1}$ to $\mathbb{R}^m$. For $y \in \mathrm{B}_1(\mathbb{R}^{m-1})$ we use the abuse of notation

$$|\det \phi'(y)| := \left|\det \begin{bmatrix} \phi'(y) & v \end{bmatrix}\right|,$$

where v is a unit vector in $\operatorname{Span}\phi'(y)^{\perp}$ Likewise, we use the abuse of notation $|\det \varphi'(x)| := |\det \phi'(\varphi(x))|^{-1}$. Since g is continuous on the compact set $\mathrm{S}^{m-1} \setminus D_{1/2}$, it takes a maximum G_1 there. Let $B \in \mathscr{B}_{m-1}$, and set $\delta := \nu_{m-1}(B)$. We have

$$\begin{aligned}
\mu(B) &= \int_B \frac{d\mu}{d\nu_{m-1}}(x)\nu_{m-1}(dx) \\
&\le \int_B g(x)\nu_{m-1}(dx) \\
&= \int_{B\setminus D_{1/2}} g(x)\nu_{m-1}(dx) + \int_{B\cap D_{1/2}} g(x)\nu_{m-1}(dx) \\
&\le G_1 \cdot \nu_{m-1}(B \setminus D_{1/2}) + \int_{\varphi(B\cap D_{1/2})} g\big(\phi(y)\big) \cdot \frac{|\det \phi'(y)|}{A_{m-1}(\mathrm{S}^{m-1})} \cdot \eta(dy). \qquad (11.6)
\end{aligned}$$

Since $y \to \|y\|^{\varsigma} g\big(\phi(y)\big)$ is continuous on the compact closure of $\mathrm{B}_{1/2}(\mathbb{R}^{m-1})$ when prolonged by continuity at the origin, and since $\phi \in C^1$, the quantities

$$\begin{aligned}
supDet &:= \sup\big\{|\det \phi'(y)| :\, y \in \mathrm{B}_{1/2}(\mathbb{R}^{m-1})\big\}, \\
supdet &:= \sup\big\{|\det \phi'(y)| :\, y \in \mathrm{B}_{1/2}(\mathbb{R}^{m-1})\big\}, \\
M &:= \sup\big\{\|y\|^{\varsigma} g\big(\phi(y)\big) :\, y \in \mathrm{B}_{1/2}(\mathbb{R}^{m-1})\big\}
\end{aligned}$$

are all well-defined and finite, and we have $a_0 \le M$. Let $\omega := \nu_{m-1}(B \setminus D_{1/2})$. Then

$$\begin{aligned}
\eta\big(\varphi(B \cap D_{1/2})\big) &= \int_{B\cap D_{1/2}} |\det \varphi'(x)| \cdot A_{m-1}(\mathrm{S}^{m-1}) \cdot \nu_{m-1}(dx) \\
&\le A_{m-1}(\mathrm{S}^{m-1}) \cdot \nu_{m-1}(B \cap D_{1/2}) \cdot \sup_{x \in B\cap D_{1/2}} |\det \varphi'(x)| \\
&\le \frac{\delta - \omega}{supdet} \cdot A_{m-1}(\mathrm{S}^{m-1}).
\end{aligned}$$

This implies that the radius of the ball $\mathrm{B}_\rho(\mathbb{R}^{m-1})$ with volume $\eta\big(\varphi(B\cap D_{1/2})\big)$ satisfies

$$\rho = \left(\frac{\eta(\varphi(B\cap D_{1/2}))}{2\prod_{j=2}^{m-1} I_j(\pi)}\right)^{\frac{1}{m-1}} \leq \left(\frac{\pi\cdot(\delta-\omega)\cdot I_1(\pi)}{supdet\cdot I_{m-1}(\pi)}\right)^{\frac{1}{m-1}}, \tag{11.7}$$

and then Lemma 11.1 shows

$$\begin{aligned}
&\int_{\varphi(B\cap D_{1/2})} g\big(\phi(y)\big)\cdot\frac{|\det\phi'(y)|}{A_{m-1}(\mathrm{S}^{m-1})}\cdot\eta(dy)\\
&\qquad\leq \frac{supDet\cdot M}{A_{m-1}(\mathrm{S}^{m-1})}\cdot\int_{\varphi(B\cap D_{1/2})}\|y\|^{-\varsigma}\eta(dy)\\
&\qquad\leq \frac{supDet\cdot M}{A_{m-1}(\mathrm{S}^{m-1})}\cdot\int_{B_\rho}\|y\|^{-\varsigma}\eta(dy)\\
&\qquad= \frac{supDet\cdot M}{A_{m-1}(\mathrm{S}^{m-1})}\cdot A_{m-2}(\mathrm{S}^{m-2})\cdot\int_0^\rho r^{m-2-\varsigma}dr\\
&\qquad= \frac{supDet\cdot M}{(m-1-\varsigma)\cdot I_{m-2}(\pi)}\cdot\rho^{m-1-\varsigma}\\
&\qquad\overset{(11.7)}{\leq} G_2\cdot(\delta-\omega)^{1-\frac{\varsigma}{m-1}},
\end{aligned} \tag{11.8}$$

where

$$G_2 := \frac{supDet\cdot M}{(m-1-\varsigma)\cdot I_{m-2}(\pi)}\cdot\left(\frac{\pi\cdot I_1(\pi)}{supdet\cdot I_{m-1}(\pi)}\right)^{1-\frac{\varsigma}{m-1}}.$$

Substituting (11.8) into (11.6), we find

$$\mu(B)\leq G_1\cdot\omega + G_2\cdot(\delta-\omega)^{1-\frac{\varsigma}{m-1}}$$

and by Lemma 11.2 this implies

$$\mu(B)\leq G_2\cdot\delta^{1-\frac{\varsigma}{m-1}}$$

for $\delta\ll 1$. Taking logarithms on both sides of this inequality, we obtain

$$\frac{\ln\mu(B)}{\ln\delta}\geq\frac{\ln G_2}{\ln\delta}+1-\frac{\varsigma}{(m-1)},$$

and hence,

$$\liminf_{\delta\to 0}\left\{\frac{\ln\mu(B)}{\ln\nu_{m-1}(B)} : \nu_{m-1}(B)\leq\delta, B\in\mathscr{B}\right\}\geq 1-\frac{\varsigma}{m-1}.$$

This establishes that the claims of part i). To prove part ii), note that ς was chosen so that $\int_{\mathrm{S}^{m-1}} g(x)\nu_{m-1}(dx) < +\infty$. Therefore, μ is a well-defined probability measure on S^{m-1}. Moreover, it follows from part i) that its smoothness parameter satisfies $\alpha\geq 1-\varsigma/(m-1)$. It only remains to show that $\alpha\leq 1-\varsigma/(m-1)$. For all $\varepsilon>0$ there exists a radius $\rho_\varepsilon>0$ such that for all $y\in\mathrm{B}_{\rho_\varepsilon}$,

$$\Big||\det\phi'(y)|-|\det\phi'(0)|\Big|<\varepsilon,$$

$$\Big|\|y\|^\varsigma g(\phi(y))-a_0\Big|<\varepsilon.$$

Thus, on the one hand we have

$$\nu_{m-1}\big(\phi(\mathrm{B}_{\rho_\varepsilon})\big) = \int_{\mathrm{B}_{\rho_\varepsilon}} \frac{|\det \phi'(y)|}{A_{m-1}(\mathrm{S}^{m-1})} \cdot \eta(dy) \le \eta(\mathrm{B}_{\rho_\varepsilon}) \cdot \frac{|\det \phi'(0)| + \varepsilon}{A_{m-1}(\mathrm{S}^{m-1})}, \tag{11.9}$$

and on the other hand,

$$\begin{aligned}
\mu\big(\phi(\mathrm{B}_{\rho_\varepsilon})\big) &\ge \frac{\big(|\det \phi'(0)| - \varepsilon\big) \cdot \int_{\mathrm{B}_{\rho_\varepsilon}} g(\phi(y))\eta(dy)}{A_{m-1}(\mathrm{S}^{m-1}) \cdot \int_{\mathrm{S}^{m-1}} g(x)\nu_{m-1}(dx)} \\
&\ge \frac{\big(|\det \phi'(0)| - \varepsilon\big) \cdot (a_0 - \varepsilon) \cdot \int_{\mathrm{B}_{\rho_\varepsilon}} \|y\|^{-\varsigma}\eta(dy)}{A_{m-1}(\mathrm{S}^{m-1}) \cdot \int_{\mathrm{S}^{m-1}} g(x)\nu_{m-1}(dx)} \\
&= \frac{\big(|\det \phi'(0)| - \varepsilon\big) \cdot (a_0 - \varepsilon) \cdot \rho_\varepsilon^{m-1-\varsigma}}{I_{m-2}(\pi) \cdot \int_{\mathrm{S}^{m-1}} g(x)\nu_{m-1}(dx) \cdot (m-1-\varsigma)} \\
&\overset{(2.2)}{=} \frac{\big(|\det \phi'(0)| - \varepsilon\big)(a_0 - \varepsilon) \cdot \eta(\mathrm{B}_{\rho_\varepsilon})^{1-\frac{\varsigma}{m-1}}}{(m-1-\varsigma) I_{m-2}(\pi) \int_{\mathrm{S}^{m-1}} g(x)\nu_{m-1}(dx) \cdot \big(2\prod_{j=2}^{m-1} I_j(\pi)\big)^{1-\frac{\varsigma}{m-1}}} \\
&\overset{(11.9)}{\ge} G_3 \cdot \nu_{m-1}\big(\phi(\mathrm{B}_{\rho_\varepsilon})\big)^{\frac{m-1-\varsigma}{m-1}},
\end{aligned} \tag{11.10}$$

where

$$G_3 = \frac{(|\det \phi'(0)| - \varepsilon) \cdot (a_0 - \varepsilon)}{(m-1-\varsigma) I_{m-2}(\pi) \int_{\mathrm{S}^{m-1}} g(x)\nu_{m-1}(dx)} \cdot \left(\frac{\pi \cdot I_1(\pi)}{I_{m-1}(\pi) \cdot (|\det \phi'(0)| + \varepsilon)} \right)^{1-\frac{\varsigma}{m-1}}.$$

Taking logarithms on both sides of (11.10) and dividing by $\ln \nu_{m-1}(\phi(\mathrm{B}_{\rho_\varepsilon}))$, we obtain

$$\frac{\ln \mu\big(\phi(\mathrm{B}_{\rho_\varepsilon})\big)}{\ln \nu_{m-1}\big(\phi(\mathrm{B}_{\rho_\varepsilon})\big)} \le \frac{\ln G_3}{\ln \nu_{m-1}\big(\phi(\mathrm{B}_{\rho_\varepsilon})\big)} + 1 - \frac{\varsigma}{m-1}.$$

Letting $\varepsilon \to 0$, this implies that $\alpha \le 1 - \varsigma/(m-1)$, as claimed. □

Proof of the Claims of Example 3

Proof. For $B \in \mathscr{B}_{m-1}$ we have

$$\mu(B) = \sum_k \mathrm{P}[N = k] \cdot \mathrm{P}\big[X_N \in B \| N = k\big] = \sum_k \mathrm{P}[N = k] \cdot \mu_k(B).$$

If B is a ν_{m-1}-nullsets, it follows from this formula that $\mu(B) = 0$. Thus, μ is ν_{m-1}– absolutely continuous. Proposition 4.3 ii) implies that for any $\varepsilon, \delta > 0$ there exists $B_{\varepsilon,\delta} \in \mathscr{B}_{m-1}$ such that $\nu_{m-1}(B_{\varepsilon,\delta}) \in (0, \delta)$ and

$$\frac{\ln \mu_{\lceil \varepsilon^{-1} \rceil}(B_{\varepsilon,\delta})}{\ln \nu_{m-1}(B_{\varepsilon,\delta})} \le 2\varepsilon.$$

Since $\mu(B_{\varepsilon,\delta}) \ge \mathrm{P}[N = \lceil \varepsilon^{-1} \rceil] \cdot \mu_{\lceil \varepsilon^{-1} \rceil}(B_{\varepsilon,\delta})$, this implies that for

$$\delta \le \exp\left(\frac{\ln \mathrm{P}[N = \lceil \varepsilon^{-1} \rceil]}{\varepsilon} \right)$$

we have

$$\frac{\ln \mu(B_{\varepsilon,\delta})}{\ln \nu_{m-1}(B_\delta^{[\varepsilon]})} \le 2\varepsilon + \frac{\ln \mathrm{P}\left[N = \lceil \varepsilon^{-1} \rceil\right]}{\ln \nu_{m-1}(B_{\varepsilon,\delta})} \le 3\varepsilon.$$

Therefore, $alpha_{\nu_{m-1}}(\mu) \le 3\varepsilon$, and the claim follows by letting $\varepsilon \to 0$. □