On Tail Decay and Moment Estimates of a Condition Number for Random Linear Conic Systems

Dennis Cheung, Felipe Cucker*
Department of Mathematics
City University of Hong Kong
83 Tat Chee Avenue, Kowloon
HONG KONG
e-mail: {50003110@student,macucker@math}.cityu.edu.hk

Raphael Hauser†
Oxford University Computing Laboratory
Wolfson Building, Parks Road, Oxford, OX1 3QD
UNITED KINGDOM
e-mail: hauser@comlab.ox.ac.uk

Abstract. In this paper we study the distribution tails and the moments of $\mathscr{C}(A)$ and $\log \mathscr{C}(A)$, where $\mathscr{C}(A)$ is a condition number for the linear conic system $Ax \leq 0$, $x \neq 0$, with $A \in \mathbb{R}^{n \times m}$. We consider the case where A is a Gaussian random matrix. For this input model we characterise the exact decay rates of the distribution tails, we improve the existing moment estimates, and we prove various limit theorems for the cases where either n or m and n tend to infinity. Our results are of complexity theoretic interest, because interior-point methods and relaxation methods for the solution of $Ax \leq 0$, $x \neq 0$ have running times that are bounded in terms of $\log \mathscr{C}(A)$ and $\mathscr{C}(A)^2$ respectively.

AMS Classification: primary 90C31,15A52; secondary 90C05,90C60,62H10.

Key Words: condition number, random matrices, linear programming, probabilistic analysis, complexity theory.

1 Introduction

Let $A \in \mathbb{R}^{n \times m}$ be given and consider the two systems[1]

$$Ax \leq 0,\ x \neq 0 \tag{1}$$

*This author has been substantially funded by a grant from the Research Grants Council of the Hong Kong SAR (project number CityU 1085/02P).

†This author has been supported by Felipe Cucker's grant from the Research Grants Council of the Hong Kong SAR (project number CityU 1085/02P) and through a grant of the Nuffield Foundation under the "Newly Appointed Lecturers" grant scheme, (project number NAL/00720/G).

[1]Usually, in the literature, one considers a $m \times n$ matrix A appearing in (2), the "primal system", and its transpose A^{T} appears in (1), the "dual system." We revert this notation here since in most of this paper we will deal with system (1) and we do not want to burden the notation with the transpose superscript.

and

$$A^{\mathrm{T}} y = 0,\ y \geq 0,\ y \neq 0. \tag{2}$$

It is well-known that, if A is full row rank, one of these systems has a strict solution (one for which the satisfied inequality is strict in all coordinates) if and only if the other has no solutions at all.

The following feasibility problem is a standard subproblem in linear programming:

> **(FP)** Given a $n \times m$ full row rank real matrix A, decide which of (1) or (2) is strictly feasible and return a strict solution for it.

Iterative algorithms that solve this problem, such as variants of interior-point or ellipsoid methods, have a cost which depends on some measure of conditioning of the matrix A besides the natural dependence on n and m. For instance, a finite-precision algorithm solving the problem above is analysed in [14] to show a complexity of

$$\mathcal{O}\left((m+n)^{3.5}\left(\log(m+n) + \log C(A)\right)^3\right). \tag{3}$$

Here $C(A)$ may be either the condition number introduced by Renegar in [32, 33, 34], which we denote by $C_R(A)$, or a generalisation of Goffin's "inner measure" [20, 21] introduced in [10], which we denote by $\mathscr{C}(A)$ in the sequel.

Another family of algorithms whose complexity can be analysed in terms of $\mathscr{C}(A)$ are the many variants of the Agmon-Motzkin-Schönberg (AMS) relaxation method [2, 29] for solving systems of linear inequalities. This includes the cyclic projection method, the perceptron algorithm and certain types of subgradient algorithms. The complexity of these methods is typically proportional to $\mathscr{C}(A)^2$. For example, for feasible systems the perceptron algorithm is guaranteed to find a solution to $Ax < 0$ in

$$\mathcal{O}\left(\mathscr{C}(A)^2\right) \tag{4}$$

iterations (see Appendix B for further remarks). Although less interesting from a complexity perspective, such algorithms have appealing aspects in applications where m and n are both very large, for example in tumour radiation therapy planning. AMS relaxation is also the historic context in which the condition number $\mathscr{C}(A)$ has first been studied, albeit only in the case of feasible systems, see [20, 21].

It is also worthwhile mentioning that there exist close links between the condition number $\mathscr{C}(A)$ and the notion of *margin* that plays a key role in the learning theory literature. Furthermore, the condition number $\mathscr{C}(A)$ has applications in the backward error analysis and in estimating the stability of linear feasibility problems. For all these reasons, studying the moments of both $\log \mathscr{C}(A)$ and $\mathscr{C}(A)$ will be interesting in the probabilistic setting outlined below.

Unlike m and n, the condition number of A is not immediate to determine from the data A and seems to require a computation which is not easier than solving the feasibility problem instance described by A (see [31] for a discussion). In addition, there are no bounds on its magnitude as a function of m and n. It may actually be infinite. Thus, bounds such as (3) and (4) tell us little about the running time we can expect for a given input A.

A reasonable way to cope with this situation is to assume a probability measure on the space of $n \times m$ matrices A and to estimate the expected value of $\log \mathscr{C}(A)$ (or that of $\mathscr{C}(A)$). A standard choice of distribution is the Gaussian model. We say that a random matrix is *Gaussian* when its entries are i.i.d. random variables with standard normal distribution. The main result in [11] shows that if A is a Gaussian $n \times m$ matrix, then

$$\mathbf{E}\left[\log \mathscr{C}(A)\right] = \begin{cases} \mathcal{O}(\min\{m \log n, n\}) & \text{if } n > m \\ \mathcal{O}(\log n) & \text{otherwise} \end{cases} \tag{5}$$

and

$$\mathbf{E}[\log C_R(A)] \leq \mathbf{E}[\log \mathscr{C}(A)] + \frac{5 \log n}{2} + \frac{\log m}{2} + 2 \log 2.$$

In most practical occurrences of the feasibility problem **(FP)** one deals with matrices for which n is some orders of magnitude larger than m. The case $n \gg m$ (n much larger than m) is actually the case of interest among researchers in linear programming. In [16] the estimate (5) was refined to prove that, when n is moderately larger than m one has $\mathbf{E}[\log \mathscr{C}(A)] \leq \max\{\log m, \log\log n\} + \mathcal{O}(1)$.

The main contributions of this paper are the following:

(i) We further strengthen the existing bounds on $\mathbf{E}[\log \mathscr{C}(A)]$: In Corollary 8 we show that if $n \geq m$

$$\mathbf{E}[\log \mathscr{C}(A)] \leq m \log 2$$

asymptotically as $m \to \infty$, and in Corollary 9 we show that if $n \geq 5m$ and $\gamma > 0$ is an arbitrary constant, then

$$\mathbf{E}[\log \mathscr{C}(A)] \leq m^{\gamma}$$

asymptotically as $m \to \infty$.

(ii) We generalise the bounds to arbitrary moments of $\log \mathscr{C}(A)$ and derive similar bounds for the moments of $\mathscr{C}(A)$: see Corollaries 2,3 and 4.

(iii) We derive various limit theorems that are of interest in large-scale problems: in the case where m is fixed and $n \to \infty$, we show in particular that

$$\begin{aligned} \mathscr{C}(A) &\overset{n\to\infty}{\longrightarrow}_P 1, \\ \lim_{n\to\infty} \mathbf{E}[(\log \mathscr{C}(A))^{\gamma}] &= 0 \quad \forall \gamma > 0, \\ \lim_{n\to\infty} \mathbf{E}[(\mathscr{C}(A))^{\gamma}] &= 1 \quad \forall \gamma \in [0, 1), \end{aligned}$$

see Corollaries 5,6 and 7. In the case where $n \geq m$ and $m \to \infty$, it is again Corollaries 8 and 9 that bound the asymptotic growth rates of $\mathbf{E}[\log \mathscr{C}(A)]$.

(iv) Previous probabilistic analyses of $\log \mathscr{C}(A)$ relied on the fact that linear algebraic operations applied to Gaussian matrices (including vectors as a special case) lead again to Gaussian matrices. The analysis presented here is very different because it is based on geometry on high dimensional spheres. This approach makes it possible to derive not just upper bounds on the tail decay of $\log \mathscr{C}(A)$, but also lower bounds: Theorems 1 and 2 show that there exist functions $c(m,n) \geq d(m,n)$ of m and n such that

$$\frac{d(m,n)}{t} \leq \mathbf{P}\left[\mathscr{C}(A) \geq t\right] \leq \frac{c(m,n)}{t}$$

for all t large enough. This implies that the distribution tails of $\log \mathscr{C}(A)$ asymptotically decay exactly at the exponential rate e^{-t}, see Corollary 1. More importantly, the geometric analysis we developed here generalises to almost arbitrary probability measures that are absolutely continuous with respect to the uniform measure on the sphere, as we will show in a follow-up paper. In the general case, the tail decay rates of $\log \mathscr{C}(A)$ are again exponential, but the exponent depends on a parameter defined as a function of the distribution. This exponential decay is perhaps the most important conclusion of our analysis, as it explains why the polyhedral feasibility problem – and linear programming by extension – is "empirically strongly polynomial". We discuss the relevance of this notion in Section 2.

2 Complexity Theoretic Context

2.1 Background

The interest in the distribution of $\log \mathscr{C}(A)$ stems to a large extent from the conjecture that there exist so-called *strongly polynomial* algorithms for linear programming. Let us give a brief explanation for

the uninitiated reader: since the mid-1940s, variants of Dantzig's simplex method proved to be efficient algorithms for solving linear programming problems in practice, despite the fact that in the worst case these methods terminate only after a number of iterations that is exponential in the "size", or the total input data length of the problem. As interest in complexity theory grew, many researchers believed that a good algorithm should terminate within a number of iterations that is bounded by a polynomial in the input size. Thus, the simplex method is not a polynomial algorithm.

Surprising new approaches to linear programming subsequently proved to be polynomial time algorithms for linear programming under the Turing model: Khachiyan's ellipsoid method [25], Karmarkar's method [24] and the many interior-point methods developed since then are algorithms of this kind. These algorithms are guaranteed to terminate in polynomial time when the input data are rational and the input size is measured by the total bit-length of the data.

On the other hand, under the so-called *real* complexity model one considers linear programming problems whose input data are real numbers and imagines a hypothetical computer that can perform operations on real numbers. In this model the complexity of an algorithm is the number of operations that are needed in the worst case to solve the problem. Such an algorithm is called *strongly polynomial* if its complexity is a polynomial function in the number of constraints and variables (the "dimension") of the underlying problem. Neither the ellipsoid method nor any of the known interior-point algorithms for linear programming is known to be strongly polynomial. In fact, their running time is theoretically unbounded! This is in stark contrast to the simplex method which is guaranteed to terminate in exponential time.

The situation is not hopeless, however, for the real complexity of ellipsoid and interior-point methods can be bounded by a polynomial in the problem dimension and the logarithm of a condition number, see the results cited in the introduction. Earlier relevant papers on this subject and on other applications of LP condition numbers include (among others) [32, 33, 34, 19, 18, 45, 14, 43, 44].

This condition-based complexity analysis is not new. In numerical linear algebra it occurs, for instance, in the analysis of the conjugate gradient method (cf. [30, 42]); in polynomial equation solving it occurs in the analysis of homotopy methods [36] or of grid-based methods [15, 13]. A recent survey for linear programming is [12]. A conceptually related idea in discrete mathematics is that of *parameterized complexity* [17].

The question of whether linear programming is strongly polynomial time solvable is considered an important open problem and has many ramifications within complexity theory; in his list of 18 mathematical problems for the XXI century [38] Steve Smale includes this question as Problem 9.

An interesting approach to get around the difficult issue of strong polynomiality of linear programming is the average case analysis of algorithms. The average case analysis reveals that linear programming is strongly polynomial time solvable *on average*, that is, there exist algorithms whose average running times are polynomial under the real model when the input data are normally distributed. The simplex method was known to possess this property since the early 1980s [6, 7, 8, 37]. Similar work was continued in [1, 27, 39] and [9]. More recently, the attention has shifted to the average case analysis of interior-point methods [3, 40, 28, 22] and [23]. While all of these papers followed an analysis pertaining to particular algorithms, it is also possible to derive similar results by analysing the expected value of condition numbers under Gaussian (or other) input data. Combined with a condition-based complexity analysis this yields average case running time bounds for particular algorithms. This was the approach pursued in [41, 11] and [16], and it is also the approach we pursue in the present paper.

We should point out that the relevance of average case analyses is subject to some justified scrutiny which we will further address and respond to in the next paragraph.

2.2 Discussion: Relevance of Our Results

As mentioned in the synopsis of the introduction, the results we will derive in this paper include in particular polynomial bounds of $\mathbf{E}[\log \mathscr{C}(A)]$ in m and n. In the literature on the probabilistic analysis of linear programming such results are considered important because they show that LP is "strongly polynomial on average". Precisely how significant are such statements from a complexity theoretic view point?

On the one hand, the average behaviour of an algorithm on random input data yields in itself an interesting complexity measure which, as many would argue, can be more relevant than the study of the worst case scenario. The weakness of this argument is that the relevance of average case results depends on how well the assumed probabilistic model describes the distribution of the input data that one might observe in particular applications. Without doubt, uniform or Gaussian matrices are an inadequate model in most but a few applications. In a follow-up paper we will therefore show how the techniques developed here extend to matrices with much more general probability distributions. We chose to treat the Gaussian case separately because it can be presented in a non-measure-theoretic setting which is accessible to a wider audience, and because it allows to directly compare our results, which were obtained by arguments based on spherical geometry, with the results obtained in earlier papers via the very different techniques of linear algebra on Gaussian matrices.

On the other hand, it is sometimes argued that understanding the average behaviour of interior-point algorithms constitutes a step towards proving strong polynomiality of linear programming. This argument is somewhat weaker than the first, because it may of course be that linear programming is "strongly polynomial on average" for a wide range of input distributions whilst not allowing a strongly polynomial time algorithm. Thus, the two phenomena might be unrelated.

In our view the most relevant link between the results presented in this paper and the conjectured strong polynomiality of linear programming (and the closely associated linear feasibility problem treated here) consists not in bounds on the moments of $\log \mathscr{C}(A)$ but in the exponential decay of its distribution tails observed in Corollary 1. This fact explains why algorithmic experiments have a tendency to strengthen the intuition that the conjecture of strong polynomiality be true: $\mathcal{O}(e^t)$ simulations are needed to observe an event in which $\log \mathscr{C}(A) > t$ (and even in that case it is not guaranteed that an algorithm is necessarily slow at solving the problem). Thus, it is impossible to observe the really bad cases in random experiments. In contrast, in algorithms whose complexity depends polynomially on $\mathscr{C}(A)$, much fewer simulations reveal cases with extreme running times, because it takes $\mathcal{O}(t)$ experiments to detect an event in which $\mathscr{C}(A) > t$. Of course, this argument is again subject to the criticism that the exponential decay of $\mathbf{P}[\mathscr{C}(A) > t]$ might be a particularity of the chosen input distribution for the data of A. However, our analysis of more general distributions shows that exponential decay rates are a common feature of a very general family of distributions.

These observations suggest a notion of "empirical strong polynomiality": if linear programming is not strongly polynomial time solvable, then this fact cannot be observed in random experiments because the relevant events are exponentially rare. This is in our view the correct interpretation of our results and the main message of this paper.

3 Basic Notions and Notation

If there exists $x \in \mathbb{R}^m$, $x \neq 0$, such that $Ax \leq 0$ we say that A is *feasible*. Otherwise, we say that A is *infeasible*. Also, if there exists a vector x such that $Ax < 0$ componentwise, then we say that A is *strictly feasible*. If A is feasible but not strictly feasible then we say that A is *ill-posed*.

In the sequel we consider $\arccos(t)$ as a function from $[-1, 1]$ into $[0, \pi]$. In this region, 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, $\arccos(\frac{a_i \cdot x}{\|a_i\|\|x\|})$.

Let $\theta(A,x) = \min_{1\le i\le n} \theta_i(A,x)$ and $\overline{x}$ be any vector in $\mathbb{R}^m \setminus \{0\}$, s.t. $\theta(A) = \theta(A,\overline{x}) = \sup_{x\in\mathbb{R}^m} \theta(A,x)$. The condition number $\mathscr{C}(A)$ is defined as

$$\mathscr{C}(A) = |\cos(\theta(A))|^{-1}. \tag{6}$$

In the case where the system $Ax \le 0$ is feasible, $\mathscr{C}(A)$ is the same as Goffin's condition number ν [20, 21], which he developed to analyse step length rules that guarantee finite convergence of the relaxation method applied to feasible systems of linear inequalities. $\mathscr{C}(A)$ is also closely related to Agmon's condition number λ [2] with which it coincides in some cases, but again λ is only defined for feasible systems. Thus, $\mathscr{C}(A)$ is more general because it is defined both for feasible and infeasible systems.

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 $\theta(A) < \pi/2$. It is also easy to show, using a compactness argument, that a vector $\overline{x}$ such as used in the definition of $\mathscr{C}(A)$ exists. Note that since $\mathscr{C}(A)$ is defined purely in terms of angles between vectors, $\mathscr{C}$ is invariant under positive rescaling of the rows of A. Hence, we may assume without loss of generality that all rows of A have been rescaled to unit length. Our analysis then reduces to geometry on the unit sphere.

Let S^{m-1} denote the unit sphere in $\mathbb{R}^m$. For $p \in \mathrm{S}^{m-1}$ and $\rho \in [0,\pi]$ we denote by $\mathtt{cap}(p,\rho)$ the circular cap with centre in p and angular radius ρ, that is,

$$\mathtt{cap}(p,\rho) = \left\{x \in \mathrm{S}^{m-1} : x\cdot p \ge \cos\rho\right\}.$$

By $\partial\mathtt{cap}(p,\rho)$ and $\mathrm{Int}\,(\mathtt{cap}(p,\rho))$ we denote the boundary and interior of $\mathtt{cap}(p,\rho)$, respectively in the standard topology on S, that is,

$$\begin{aligned}\partial\mathtt{cap}(p,\rho) &= \left\{x \in \mathrm{S}^{m-1} : x\cdot p = \cos\rho\right\},\\ \mathrm{Int}\,(\mathtt{cap}(p,\rho)) &= \{x \in \mathrm{S}^{m-1} : x\cdot p > \cos\rho\}.\end{aligned}$$

The following result provides geometric insight about the condition number $\mathscr{C}(A)$.

Proposition 1 *It is true that* $0 < \theta(A) \le \rho \le \pi$ *if and only if* $\bigcup_{i=1}^n \mathtt{cap}(a_i,\rho) = \mathrm{S}^{m-1}$.

Proof.

$$\begin{aligned}\theta(A) \le \rho \quad &\Leftrightarrow \quad \max_{x\in\mathbb{R}^m\setminus\{0\}} \theta(A,x) \le \rho\\ &\Leftrightarrow \quad \forall x \in \mathrm{S}^{m-1}, \min_{1\le i\le n} \theta_i(A,x) \le \rho\\ &\Leftrightarrow \quad \forall x \in \mathrm{S}^{m-1}, \exists i \in \{1,2,\ldots,n\}, x \in \mathtt{cap}(a_i,\rho)\\ &\Leftrightarrow \quad \bigcup_{i=1}^n \mathtt{cap}(a_i,\rho) = \mathrm{S}^{m-1}.\end{aligned}$$

□

4 Characterising Extremal Circular Caps

In this section A is a real $n\times m$ matrix with unit row vectors a_i. A *largest circular cap* excluding rows of A (in short, a LCP of A) is a cap $\mathtt{cap}(p^*,\rho^*)$ corresponding to a maximiser p^* and its objective function value $\rho^* = \rho(p^*)$ for the optimisation problem

$$\max_{p\in\mathrm{S}^{m-1}} \rho(p), \tag{7}$$

where $\rho(p) = \min_i \arccos(p \cdot a_i)$. A *smallest circular cap* containing all of the a_i $(i = 1, \ldots, n)$ (in short, a SCP of A) is a complement of a LCP of A.

The following result explains why the notions of LCP and SCP are important in the analysis of the condition number $\mathscr{C}(A)$.

Proposition 2 *Let $\mathtt{cap}(p, \rho)$ be a LCP of A. Then $x^* = p$ maximises the function $\theta(A, x)$ and it is true that $\theta(A) = \theta(A, p) = \rho$, that is, $\mathscr{C}(A) = 1/|\cos \rho|$.*

Proof. Let $\overline{x} \in \mathrm{S}^{m-1}$ be a maximiser of $\theta(A, x)$. Then $\theta(A, \overline{x}) = \theta(A)$ and $\mathrm{Int}\,(\mathtt{cap}(\overline{x}, \theta(A)))$ contains none of the rows of A. This shows that $\theta(A) \leq \rho$. On the other hand, since $a_i \in \mathtt{cap}(-p, \pi - \rho)$ for all i, we have

$$\rho \leq \min_{1 \leq i \leq n} \arccos(p \cdot a_i) = \theta(A, p) \leq \theta(A).$$

□

It is worth investigating the properties of LCPs and SCPs a bit further. A simple compactness argument shows that LCPs and SCPs exist for any A. If A is infeasible then these caps are not unique in general. To visualise this fact, it is helpful to consider the limiting case where a countable set of points densely fills what is left of the unit sphere after two circular caps of equal radius but different midpoints have been removed.

If A is strictly feasible however, there exists a unique LCP, respectively SCP, as the following convexity argument shows: it is easy to see that the strict feasibility of A implies that $p^* \cdot a_i > 0$ for all i and $0 \leq \varrho^* < \pi$ for any SCP $\mathtt{cap}(p^*, \varrho^*)$. It suffices therefore to argue that the minimisation problem

$$\begin{aligned} &\min_{p \in \mathrm{S}^{m-1}} \varrho(p) \\ &\text{s.t. } p \cdot a_i > 0 \quad (i = 1, \ldots, n) \end{aligned} \tag{8}$$

has a unique local minimiser, where $\varrho(p) = \pi - \rho(-p) = \max_i \arccos(p \cdot a_i)$. Suppose $p_1 \neq p_2$ are two distinct local minimisers of (8) such that $\varrho_2 := \varrho(p_2) \geq \varrho_1 := \varrho(p_1)$. Then

$$p_j \cdot a_i \geq \cos \varrho_j > 0 \quad (i = 1, \ldots, n; j = 1, 2). \tag{9}$$

For $\lambda \in (0, 1)$ let $p(\lambda) := \lambda p_1 + (1 - \lambda) p_2$ and $\widehat{p}(\lambda) := p(\lambda)/\|p(\lambda)\|$. Then $p_1 \neq p_2$ implies that $\|p(\lambda)\| < 1$ and

$$\widehat{p}(\lambda) \cdot a_i > p(\lambda) \cdot a_i \overset{(9)}{\geq} \lambda \cos \varrho_1 + (1 - \lambda) \cos \varrho_2 \quad (i = 1, \ldots, n). \tag{10}$$

It follows from (10) that p_2 is not a local minimiser of (8) if $\varrho_2 > \varrho_1$, that is, it must be true that $\varrho_1 = \varrho_2$. But then, for all $\lambda \in (0, 1)$, (10) shows that $\mathtt{cap}\big(\widehat{p}(\lambda), \varrho(\lambda)\big)$ contains all a_i $(i = 1, \ldots, n)$, where $\varrho(\lambda) = \arccos\big(\min_i \widehat{p}(\lambda) \cdot a_i\big) < \varrho_1$, contradicting the local optimality of p_1 and p_2. This shows that the SCP and, by extension, the LCP of A are unique, as claimed.

Next we investigate the idea of blocking sets. Let $\mathtt{cap}(p, \rho)$ be a LCP of A. We say that

$$S = \big\{i : a_i \in \partial \mathtt{cap}(p, \rho)\big\} \tag{11}$$

is the *blocking set* of $\mathtt{cap}(p, \rho)$. The blocking set corresponds to the vectors that locally keep the LCP from having a larger radius.[2] In fact, S is the active set in the following equivalent reformulation of (7):

$$\begin{aligned} &\max_{p \in \mathrm{S}^{m-1}} \rho \\ &\text{s.t. } p \cdot a_i \leq \cos \rho \quad (i = 1, \ldots, n). \end{aligned}$$

[2] We say "locally" in the sense that the largest cap not containing any of the a_i and centred at a point q has radius smaller than ρ for all q in a small neighbourhood around p. Note, however, that the blocking set does not prevent a point q far away from p from being the centre of an even larger cap not containing any of the a_i. This idea of a blocking set thus explains why $\mathtt{cap}(p, \rho)$ is a *local* minimiser of problem (8). Of course, a LCP is defined as a *global* minimiser of this problem. Thus, the existence of a blocking set is a local optimality condition. As always in nonlinear optimisation, useful global optimisation criteria don't really exist.

By straightforward extension we also speak of the blocking set of a SCP.

If A is strictly feasible then the blocking set of the (unique) LCP can have any cardinality $\geq \min(2, n)$, as can easily be illustrated by the example of three points on S^2 that lie on a single grand circle and within a sufficiently small angle of one another.

The situation is rather different if A is infeasible and $n \geq m$. In this case all blocking sets have cardinality $\geq m$. In fact, let $\mathtt{cap}(p, \rho)$ be a LCP of an infeasible A, that is, $\theta(A) = \rho < \pi/2$, let S be the blocking set of $\mathtt{cap}(p, \rho)$ and suppose that $|S| < m$. Then there exists a vector $u \in \mathrm{S}^{m-1} \cap \mathrm{Span}\,(S)^{\perp}$ such that $u \cdot p \geq 0$. Let $p_\delta = p + \delta u$. For $\delta > 0$ we have $\|p_\delta\|^2 = 1 + \delta^2 + 2\delta u \cdot p > 1$. Let $\widehat{p}_\delta = p_\delta / \|p_\delta\|$ and $\rho_\delta = \arccos\bigl(\cos\rho / \|p_\delta\|\bigr)$. Then $\rho_\delta > \rho$. For all $i \notin S$ and $\delta > 0$ sufficiently small, $a_i \notin \mathtt{cap}(\widehat{p}_\delta, \rho_\delta)$, since $a_i \notin \mathtt{cap}(p, \rho)$ and $\mathtt{cap}(\widehat{p}_\delta, \rho_\delta)$ varies continuously as a function of δ. Moreover, for $i \in S$ we have $p_\delta \cdot a_i = p \cdot a_i + \delta u \cdot a_i = p \cdot a_i = \cos\rho$. Therefore $\widehat{p}_\delta \cdot a_i = \cos\bigl(\rho / \|p_\delta\|\bigr) = \cos\rho_\delta$, and this shows that $a_i \in \partial\mathtt{cap}(p_\delta, \rho_\delta)$ for all $i \in S$. We conclude that $\mathrm{Int}\,(\mathtt{cap}(\widehat{p}_\delta, \rho_\delta))$ contains none of the a_i, and since $\rho_\delta > \rho = \theta(A)$ this is a contradiction. Therefore, our assumption was wrong and $|S| \geq m$.

Let us summarise what we have found so far.

Proposition 3 *Let $A \in \mathbb{R}^{n \times m}$ have unit rows. Then*

(i) If A is strictly feasible there exists a unique LCP and SCP of A but the cardinality of the blocking set is arbitrary.

(ii) If A is infeasible then there exist LCPs and SCPs of A which are not unique in general, but their blocking sets always have cardinality $\geq m$.

Proof. See arguments above. □

Let us further explore the link between blocking sets and extremal circular caps. We consider the set of index sets of cardinality m, $\mathcal{P}_m = \{S \subset \{1, \ldots, n\} : |S| = m\}$. If $S \in \mathcal{P}_m$ and $A \in \mathbb{R}^{n \times m}$ we denote by A_S the $m \times m$ matrix obtained by removing all rows from A with index not in S. Let $e = (1 \ldots 1)^{\mathrm{T}} \in \mathbb{R}^m$. If A is nonsingular, the vectors

$$u_S = A_S^{-1} e \quad \text{and} \quad \widehat{u}_S = \frac{u_S}{\|u_S\|} \tag{12}$$

are well defined.

Lemma 1 *Let $S \in \mathcal{P}_m$ be such that A_S is nonsingular, and let $p \in \mathrm{S}^{m-1}$, $\rho \in [0, \pi/2)$ and $S \in \mathcal{P}_m$ be such that $a_i \in \partial\mathtt{cap}(p, \rho)$ for $(i \in S)$. Then $p = \widehat{u}_S$ and $\cos\rho = \|u_S\|^{-1}$.*

Proof. If $a_i \in \partial\mathtt{cap}(p, \rho)$ then $a_i \cdot p = \cos\rho$. Therefore, $A_S p = (\cos\rho)e$ and $p = (\cos\rho)A_S^{-1}e = (\cos\rho)u_S$. Since $\|p\| = 1$, we have $|\cos\rho| = \|u_S\|^{-1}$ and, using $\rho < \pi/2$, $\cos\rho = \|u_S\|^{-1}$. We conclude that $p = \widehat{u}_S$. □

Blocking sets are the key tool that allow us to gain information about the distribution of $\mathscr{C}(A)$ when the unit rows of A are random vectors with known distribution. The fact that the blocking set S may have cardinality $|S| < m$ for strictly feasible A is an obstacle to this analysis that requires further attention. The following two lemmas allow to overcome this problem in the analysis of upper and lower bounds on the tails of $\mathscr{C}(A)$ respectively.

Lemma 2 *Let $A \in \mathbb{R}^{n \times m}$ have $n \geq m$ unit rows, and let $\mathtt{cap}(p, \rho)$ with $\rho < \pi/2$ contain all rows of A. Then there exist $p' \in \mathrm{S}^{m-1}$ and $\rho \leq \rho' < \frac{\pi}{2}$ such that $\mathtt{cap}(p', \rho')$ also contains all rows of A and at least m of them lie on the boundary $\partial\mathtt{cap}(p', \rho')$.*

PROOF. Let $S = \{i : a_i \in \partial\mathtt{cap}(p,r)\}$, and let us assume that $|S| < m$. Then there exists a vector $u \in \mathrm{S}^{m-1} \cap \mathrm{Span}\,(S)^\perp$, where we set $\mathrm{Span}\,(S)^\perp = \mathbb{R}^m$ if $S = \emptyset$, such that there exists an index $i \notin S$ with $a_i \cdot u \neq 0$. Without loss of generality we may assume that $p \cdot u \geq 0$. Let i^* be a minimiser of $\min_{i\notin S} \left|\frac{\cos\rho - a_i\cdot p}{a_i\cdot u}\right|$ and

$$\delta = \frac{\cos\rho - a_{i^*}\cdot p}{a_{i^*}\cdot u}.$$

Let finally $p_\delta = p + \delta u$. Then $p\cdot u \geq 0$ implies that $\|p_\delta\| \geq 1$ and hence, $\widehat{p}_\delta = \|p_\delta\|^{-1} p_\delta$ and $\rho_\delta = \arccos\big(\|p_\delta\|^{-1}\cos\rho\big) \in [\rho, \frac{\pi}{2})$ are well defined. The proof of Proposition 3 shows that $a_i \in \partial\mathtt{cap}(\widehat{p}_\delta, \rho_\delta)$ for all $i \in S$. Moreover, the definition of δ shows that $a_{i*}\cdot p_\delta = \cos\rho$ and hence that $a_{i*}\cdot\widehat{p}_\delta = \cos\rho_\delta$. This shows that $a_{i^*} \in \partial\mathtt{cap}(\widehat{p}_\delta,\rho_\delta)$ and thus, $|S'| > |S|$, where

$$S' = \{i : a_i \in \partial\mathtt{cap}(\widehat{p}_\delta,\rho_\delta)\}. \tag{13}$$

Note also that if $a_i \notin \mathtt{cap}(\widehat{p}_\delta,\rho_\delta)$ then $i \notin S$ and $a_i\cdot\widehat{p}_\delta < \cos\rho_\delta$, that is, $a_i\cdot p_\delta < \cos\rho$ which contradicts the choice of i^*. Therefore, $\mathtt{cap}(\widehat{p}_\delta,\rho_\delta)$ contains all rows of A. Using this construction recursively, we eventually arrive at the desired p' and ρ'. □

Lemma 3 *Let $a_1,\dots,a_m$ be linearly independent elements of $\partial\mathtt{cap}(p,\rho) \subset \mathrm{S}^{m-1}$ with $\rho \in [0,\pi/2)$. Let $\pi_p : \mathbb{R}^m \to \mathrm{Span}\,(p)^\perp$ be the orthogonal projection along p. Then $\mathtt{cap}(p,\rho)$ is the SCP of $A = [a_1 \dots a_m]^{\mathrm{T}}$ if and only if $0 \in \mathrm{conv}(\pi_p a_1,\dots,\pi_p a_m)$.*

PROOF. Let us first remark that is trivial to see that $\rho < \pi/2$ implies that

$$0 \in \mathrm{conv}(\pi_p a_1,\dots,\pi_p a_m) \Leftrightarrow p \in \mathrm{cone}(a_1,\dots,a_m). \tag{14}$$

Note also that the linear independence of the a_i implies that A is strictly feasible and the SCP of A is unique.

We will now show the "only if" part of the lemma. Suppose $0 \notin \mathrm{conv}(\pi_p a_1,\cdots,\pi_p a_m)$. Because of (14), Farkas' lemma implies that there exists $q \in \mathrm{S}^{m-1}$ such that $q\cdot a_i > 0$ $(i = 1,\dots,m)$ and $q\cdot p < 0$. For $\delta > 0$ let

$$p_\delta = p + \delta q, \quad \widehat{p}_\delta = \frac{p_\delta}{\|p_\delta\|} \quad \text{and} \quad \rho_\delta = \arccos\min_i\big(a_i\cdot\widehat{p}_\delta\big). \tag{15}$$

Then $\|p_\delta\|^{-1} = 1 - \delta p\cdot q + \mathcal{O}(\delta^2)$ and

$$a_i\cdot\widehat{p}_\delta \geq (\cos\rho + \delta a_i\cdot q)(1 - \delta p\cdot q + \mathcal{O}(\delta^2)) > \cos\rho \qquad (i = 1,\dots,m)$$

for all $\delta > 0$ small enough. Therefore, $\mathtt{cap}(\widehat{p}_\delta,\rho_\delta)$ contains all a_i and $\rho_\delta < \rho$ for $0 < \delta \ll 1$, showing that $\mathtt{cap}(p,\rho)$ is not the SCP of A.

It remains to show the "if" part of the lemma. Because of (14) we may assume that $p = \sum_{i=1}^m \lambda_i a_i$ for some $\lambda_1,\dots,\lambda_m \geq 0$. Let us assume that $\mathtt{cap}(p,\rho)$ is not the SCP of A. Then by a construction similar to (10) there exists a direction $d \in \mathrm{Span}\,(p)^\perp$ such that $\mathtt{cap}(\widehat{p}_\delta,\rho_\delta)$ contains all a_i and $\rho_\delta < \rho$ for $0 < \delta \ll 1$, where $\widehat{p}_\delta$ and ρ_δ are defined as in (15). That is to say,

$$a_i\cdot\widehat{p}_\delta = \frac{a_i\cdot p + \delta a_i\cdot d}{\sqrt{1+\delta^2}} > \cos\rho = a_i\cdot p$$

for $(i = 1,\dots,m)$ and $0 < \delta \ll 1$, which shows that $a_i\cdot d > 0$ for all i. But then $0 = p\cdot d = \sum_{i=1}^m \lambda_i a_i\cdot d > 0$, which shows that our assumption was wrong and $\mathtt{cap}(p,\rho)$ is indeed the SCP of A. □

5 The Input Distribution

It is well known that if A is a Gaussian matrix with rows $a_1, \ldots, a_n$ then

$$\frac{a_1}{\|a_1\|}, \ldots, \frac{a_n}{\|a_n\|}, \|a_1\|^2, \ldots, \|a_n\|^2$$

are independent random vectors and variables, the first n of which have uniform distribution on the sphere, $a_i/\|a_i\| \sim \mathscr{U}(\mathrm{S}^{m-1})$, and the last n of which are χ^2_m distributed on $\mathbb{R}_+$. Recall that $\mathscr{C}(A)$ is invariant under row scaling of A. The distribution of $\mathscr{C}(A)$ under Gaussian input is therefore the same as under input matrices with i.i.d. $\mathscr{U}(\mathrm{S}^{m-1})$ rows.

In subsequent sections we will therefore assume that

$$A_i : \Omega \to \mathrm{S}^{m-1} \qquad (i = 1, \ldots, n)$$

are i.i.d. random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $A_i \sim \mathscr{U}(\mathrm{S}^{m-1})$, and that A is the random $n \times m$ matrix

$$A = [A_1 \ldots A_n]^{\mathrm{T}}.$$

We say that A is a *uniform random matrix*. For convenience we shall assume that $m \leq n$, though the case $m < n$ is not difficult to derive from the results we will develop.

We adhere to the usual practice in probability theory and say that a property holds *almost surely* (or for almost all $\omega \in \Omega$) if it holds with probability 1. An event that occurs with probability zero is also called a null-set.

Remark 1 Using Proposition 3 (ii) and Lemma 2 it is trivial to show that if A is a uniform random matrix then almost surely A is not ill-posed. Likewise, the sets S from (11) and S' from (13) are bases of $\mathbb{R}^m$ almost surely. Finally, any m i.i.d. uniformly drawn vectors from S^{m-1} are a basis of $\mathbb{R}^m$ almost surely. Therefore, when A is a uniform random matrix then for all $S \in \mathcal{P}_m$ the matrix A_S is almost surely nonsingular and the random vectors u_S and $\widehat{u}_S$ are defined for almost all $\omega \in \Omega$.

We denote the uniform probability measure on S^{m-1} by ν_{n-1}. It is well known that the $m-1$-dimensional volume of $\mathtt{cap}(p, \rho)$ equals the integral $I_{m-2}(\rho) = \int_0^r \sin^{m-2} x \, dx$ times the volume of the unit sphere in $\mathbb{R}^{m-1}$. Therefore,

$$\mathbf{P}\left[A_i \in \mathtt{cap}(p, \rho)\right] = \nu_{m-1}(\mathtt{cap}(p, \rho)) = \frac{I_{m-2}(\rho)}{I_{m-2}(\pi)}. \tag{16}$$

It is trivial to check by induction that

$$I_m(\pi) \geq \frac{2}{\sqrt{m}}. \tag{17}$$

6 Upper Tail Bounds

The goal of this section is to derive an upper bound on the tail probability $\mathbf{P}[\mathscr{C}(A) \geq t]$ when A is a uniform random matrix.

Let A be a random uniform matrix. For $S \in \mathcal{P}_m$ and $t \geq 1$ we consider the following events:

$$\begin{aligned}
\mathcal{N}_t(S) &= \{\omega \in \Omega : \|u_S(\omega)\| \geq t\}, \\
\mathcal{B}_t^{\text{out}}(S) &= \{\omega \in \Omega : A_i(\omega) \notin \mathtt{cap}\,(\widehat{u}_S(\omega), \arccos(1/t))\, \forall i \notin S\}, \\
\mathcal{B}_{\pi/2}^{\text{in}}(S) &= \{\omega \in \Omega : A_i(\omega) \in \mathtt{cap}\,(\widehat{u}_S(\omega), \pi/2)\, \forall i \notin S\}, \\
\mathcal{A}_t^{\text{if}} &= \big\{\omega \in \Omega : \big(A(\omega) \text{ infeasible }\big) \wedge \big(\mathscr{C}(A(\omega)) \geq t\big)\big\}, \\
\mathcal{A}_t^{\text{sf}} &= \big\{\omega \in \Omega : \big(A(\omega) \text{ strictly feasible }\big) \wedge \big(\mathscr{C}(A(\omega)) \geq t\big)\big\}, \\
\mathcal{S}_t(S) &= \mathcal{N}_t(S) \cap \mathcal{B}_t^{\text{out}}, \\
\mathcal{S}_{t,\pi/2}(S) &= \mathcal{N}_t(S) \cap \mathcal{B}_{\pi/2}^{\text{in}}.
\end{aligned}$$

The following lemma is a key tool in our analysis.

Lemma 4 *Let A be a uniform random matrix and $t \in [1, \infty)$. Then $A_t^{\text{if}} \setminus \big(\bigcup_{S \in \mathcal{P}_m} \mathcal{S}_t(S)\big)$ and $A_t^{\text{sf}} \setminus \big(\bigcup_{S \in \mathcal{P}_m} \mathcal{S}_{t,\pi/2}(S)\big)$ are null-sets.*

Proof. If $A = A(\omega)$ is infeasible and $\mathscr{C}(A) = 1/\cos\theta(A) \geq t$ then

$$\arccos(1/t) \leq \theta(A) < \pi/2. \tag{18}$$

Let $\mathtt{cap}\big(p, \theta(A)\big)$ be a LCP of A and S the corresponding blocking set. By Proposition 3 and Remark 1, we have $|S| = m$ for almost all $\omega \in \Omega$. Lemma 1 implies that $p(A) = \widehat{u}_S$ for these ω and $\arccos(1/\|u_S\|) = \theta(A)$, which together with (18) shows that $\|u_S\| \geq t$ and

$$\mathtt{cap}\,(\widehat{u}_S, \arccos(1/t)) \subseteq \mathtt{cap}\,(\widehat{u}_S, \arccos(1/\|u_S\|))$$

does not contain any of the A_i ($i \notin S$). This shows that $A_t^{\text{if}} \setminus \big(\bigcup_{S \in \mathcal{P}_m} \mathcal{S}_t(S)\big)$ is a null-set.

If $A = A(\omega)$ is strictly feasible and $\mathscr{C}(A) = 1/|\cos\theta(A)| \geq t$, then $\pi/2 < \theta(A) \leq \arccos(-1/t)$ and hence,

$$\frac{\pi}{2} > \pi - \theta(A) \geq \arccos(1/t) \tag{19}$$

Let $\mathtt{cap}\big(-p(\omega), \pi - \theta(A)\big)$ be the (unique) SCP of A. Lemma 2 applied to this cap shows that there exists $\rho' \in [\pi - \theta(A), \pi/2)$ and $p' \in \mathrm{S}^{m-1}$ such that $\mathtt{cap}(p', \rho')$ contains all rows of A. By Remark 1 $S = \big\{i : A_i(\omega) \in \partial\mathtt{cap}(p', \rho')\big\}$ is of cardinality m for almost all $\omega \in \Omega$. Lemma 1 implies that $p' = \widehat{u}_S$ and $\rho' = \arccos(1/\|u_S\|)$ for these ω. Moreover, $\mathtt{cap}(\widehat{u}_S, \pi/2) \supset \mathtt{cap}(\widehat{u}_S, \rho')$ contains all rows of A and in particular $\{A_i(\omega) : i \notin S\}$. This shows that $A_t^{\text{sf}} \setminus \big(\bigcup_{S \in \mathcal{P}_m} \mathcal{S}_{t,\pi/2}(S)\big)$ is a null-set. □

For shorthand notation we write $S^* = \{1, \ldots, m\} \in \mathcal{P}_m$, $\mathcal{S}_t^* = \mathcal{S}_t(S^*)$ and $\mathcal{S}_{t,\pi/2}^* = \mathcal{S}_{t,\pi/2}(S^*)$ in the sequel. Then, for $t \geq 1$,

$$\begin{aligned}
\mathbf{P}\left[\mathscr{C}(A) \geq t\right] &= \mathbf{P}\left[\mathcal{A}_t^{\text{if}}\right] + \mathbf{P}\left[\mathcal{A}_t^{\text{sf}}\right] + \mathbf{P}\left[\big(A \text{ ill-posed}\big) \wedge \big(\mathscr{C}(A) \geq t\big)\right] \\
&= \mathbf{P}\left[\mathcal{A}_t^{\text{if}} \cap \bigcup_{S \in \mathcal{P}_m} \mathcal{S}_t(S)\right] + \mathbf{P}\left[\mathcal{A}_t^{\text{sf}} \cap \bigcup_{S \in \mathcal{P}_m} S_{t,\pi/2}(S)\right] + 0 \qquad (20) \\
&\leq \sum_{S \in \mathcal{P}_m} \mathbf{P}\left[\mathcal{S}_t(S)\right] + \sum_{S \in \mathcal{P}_m} \mathbf{P}\left[\mathcal{S}_{t,\pi/2}(S)\right] \\
&\leq \binom{n}{m} \left(\mathbf{P}\left[\mathcal{S}_t^*\right] + \mathbf{P}\left[\mathcal{S}_{t,\pi/2}^*\right]\right), \qquad (21)
\end{aligned}$$

where (20) follows from Remark 1 and Lemma 4.

Note that

$$\begin{aligned}
\mathbf{P}\left[\mathcal{S}_t^*\right] &= \mathbf{P}\left[\mathcal{B}_t^{\text{out}}(S^*)\middle\|\mathcal{N}_t(S^*)\right] P\left[\mathcal{N}_t(S^*)\right] \\
&= \left(\int_{\mathrm{S}^{m-1}} \mathbf{P}\left[\mathcal{B}_t^{\text{out}}(S^*)\middle\|\mathcal{N}_t(S^*), \widehat{u}_{S^*} = x\right] \nu_{m-1}(dx)\right) \mathbf{P}\left[\mathcal{N}_t(S^*)\right] \qquad (22) \\
&= \left(\int_{\mathrm{S}^{m-1}} \left(1 - \nu_{m-1}\left(\mathtt{cap}\left(x, \arccos(1/t)\right)\right)\right)^{n-m} \nu_{m-1}(dx)\right) \cdot \mathbf{P}\left[\mathcal{N}_t(S^*)\right] \qquad (23) \\
&= \left(1 - \frac{I_{m-2}\left(\arccos(1/t)\right)}{I_{m-2}(\pi)}\right)^{n-m} \cdot \mathbf{P}\left[\mathcal{N}_t(S^*)\right], \qquad (24)
\end{aligned}$$

where (22) holds because $\mathscr{D}\big(\widehat{u}_{S^*}\big\|\mathcal{N}_t(S^*)\big) \sim \mathscr{U}(\mathrm{S}^{m-1})$, (23) holds because the random vectors A_i ($i \notin S^*$) are independent of A_j ($j \in S^*$), and (24) follows from (16). Here, for a random variable X, $\mathscr{D}(X)$ denotes the distribution of X, and $\mathscr{U}(\mathrm{S}^{m-1})$ denotes the uniform distribution on S^{m-1}. Likewise, a similar argument shows that

$$\mathbf{P}\left[\mathcal{S}_{t,\pi/2}^*\right] = 2^{-(n-m)} \mathbf{P}\left[\mathcal{N}_t(S^*)\right]. \qquad (25)$$

The task remains to determine bounds on $\mathbf{P}\left[\mathcal{N}_t(S^*)\right]$. For $j \in S^*$ let B_j be the unique unit vector in $\mathrm{Span}\left(\{A_i : i \neq j\}\right)^{\perp}$ that complements $\{A_i : i \neq j\}$ to a positively oriented basis of $\mathbb{R}^m$. Note that A_j and B_j are independent random vectors.

Lemma 5 *Let* $\mathcal{C}_t := \left\{\omega \in \Omega : \bigcup_{j \in S^*} \left\{|B_j(\omega) \cdot A_j(\omega)| \leq m/t\right\}\right\}$. *Then* $\mathcal{C}_t \setminus \mathcal{N}_t(S^*)$ *is a nullset for all* $t \geq 1$.

Proof. Almost surely A_{S^*} is nonsingular and then

$$\|u_{S^*}\| \leq \|A_{S^*}^{-1}\| \cdot \|e\| \leq \|A_{S^*}^{-1}\|_F \sqrt{m}. \qquad (26)$$

Together with (26) the inequality $\|u_{S^*}\| \geq t$ implies that $\|A_{S^*}^{-1}\|_F \geq t/\sqrt{m}$. Whenever this holds there exists an index $j \in S^*$ such that $\|A_{\cdot,j}^{-1}\| \geq t/m$, where $A_{\cdot,j}^{-1}$ denotes the j-th column of $A_{S^*}^{-1}$. Now the equation $A_{S^*} A_{S^*}^{-1} = I$ implies that $A_i \cdot A_{\cdot,j}^{-1} = \delta_{ij}$ (the Kronecker symbol), which shows that $A_{\cdot,j}^{-1} / \|A_{\cdot,j}^{-1} = \pm B_j$ and $|B_j \cdot A_j| = 1/\|A_{\cdot,j}^{-1}\| \leq m/t$. This proves the result. □

Lemma 6 *For all* $m \geq 3$, $u \in \mathrm{S}^{m-1}$ *and* $(i = 1, \ldots, n)$ *it is true that*

$$\mathbf{P}\left[|A_i \cdot u| \leq \frac{m}{t}\right] \leq \frac{m^{\frac{3}{2}}}{t}.$$

Proof. Since the statement is trivially true when $m \geq t$, we may assume without loss of generality that $m < t$ and that $\arccos(m/t)$ is well defined. Therefore,

$$|A_i \cdot u| \leq m/t \Leftrightarrow A_i \in \mathtt{cap}\left(u, \arccos(-m/t)\right) \cap \mathtt{cap}\left(-u, \arccos(-m/t)\right).$$

Thus,

$$\begin{aligned}
\mathbf{P}\left[|A_i \cdot u| \leq m/t\right] &= \left(I_{m-2}\left(\arccos(-m/t)\right) - I_{m-2}\left(\arccos(m/t)\right)\right) \left(I_{m-2}(\pi)\right)^{-1} \\
&\leq \frac{2\int_{\arccos(m/t)}^{\pi/2} \sin x \, dx}{\int_0^{\pi} \sin^{m-2} x \, dx} \leq \frac{m\sqrt{m-2}}{t}.
\end{aligned}$$

where the last inequality follows from the fact that $m \geq 3$ and from equation (17). □

Lemma 6 allows us to compute the bound we seek as follows: for $t \geq 1$ and $m \geq 3$ we have

$$\begin{aligned}
\mathbf{P}\left[\mathcal{N}_t(S^*)\right] &\overset{\text{Lem 5}}{\leq} \mathbf{P}\left[\mathcal{C}_t\right] + 0 \\
&\leq \sum_{j \in S^*} \mathbf{P}\left[|B_j \cdot A_j| \leq \frac{m}{t}\right] \\
&= \sum_{j \in S^*} \int_{\mathrm{S}^{m-1}} \mathbf{P}\left[|B_j \cdot A_j| \leq \frac{m}{t} \middle\| B_j = x\right] \nu_{m-1}(dx) \qquad (27) \\
&\leq \frac{m^{\frac{5}{2}}}{t}. \qquad (28)
\end{aligned}$$

where (27) uses the fact that B_j is uniformly distributed on the sphere because the A_i ($i \neq j$) are, and where (28) follows from Lemma 6 and the fact that A_j is independent of B_j. Putting all the pieces together, we can now give an upper bound on the tail decay of $\mathscr{C}(A)$.

Theorem 1 *For all $t \geq 1$, $m \geq 3$ and $n \geq m$ it is true that*

$$\mathbf{P}\left[\mathscr{C}(A) \geq t\right] \leq \binom{n}{m} \cdot 2m^{\frac{5}{2}} \cdot \left(1 - \frac{I_{m-2}\left(\arccos\left(\frac{1}{t}\right)\right)}{I_{m-2}(\pi)}\right)^{n-m} \cdot \frac{1}{t}.$$

Proof. The claim follows immediately from equations (21), (24), (25) and (28) together with the fact that $1 - I_{m-2}(\arccos(1/t))/I_{m-2}(\pi) \geq 1/2$. □

7 Lower Tail Bounds

The goal of this section is to derive lower bounds on the decay rates of $\mathscr{C}(A)$ in Theorem 2. In Section 8 we will see that the combination of Theorems 1 and 2 yields the exact asymptotic decay rates of $\log \mathscr{C}(A)$.

Since $K \cap \mathrm{S}^{m-1}$ is a Borel set for all convex cones $K \subseteq \mathbb{R}^m$, the *angle space* of K

$$\mathfrak{as}(K) = \nu_{m-1}(K \cap \mathrm{S}^{m-1})$$

is well-defined. It follows from the remarks of Section 5 that alternative equivalent definitions are provided by the relations

$$\mathfrak{as}(K) = \mathbf{P}\left[X \in K \cap \mathrm{S}^{m-1}\right] = \mathbf{P}\left[Y \in K\right],$$

where $X \sim \mathscr{U}(\mathrm{S}^{m-1})$ is a uniform random vector on the unit sphere and Y is a multivariate normal random vector on $\mathbb{R}^m$ with covariance matrix $\sigma^2\mathrm{I}$ for any $\sigma^2 > 0$.

Let $A_i \sim \mathscr{U}(\mathrm{S}^{m-1})$ ($i = 1, \ldots, k$) be i.i.d. random vectors, where $k \leq m$, and let $\mathcal{LI}(A_1, \ldots, A_k) \subset \Omega$ be the event that $\{A_1(\omega), \ldots, A_k(\omega)\}$ is a linearly independent set of vectors. Then $\Omega \setminus \mathcal{LI}(A_1, \ldots, A_k)$ is a nullset, and for all $\omega \in \mathcal{LI}(A_1, \ldots, A_k)$ there exists a unique orthogonal basis $\{E_1, \ldots, E_k\}$ of $\mathrm{Span}\left(A_1, \ldots, A_k\right)$ such that $E_i \cdot A_i > 0$ and $\mathrm{Span}\left(E_1, \ldots, E_i\right) = \mathrm{Span}\left(A_1, \ldots, A_i\right)$ for $(i = 1, \ldots, k)$. In fact, the vectors E_i are the column vectors of Q in the thin QR factorisation of the matrix $[A_1 \ldots A_k]$ (i.e., the E_i are obtained by Gram-Schmidt orthogonalisation of the A_i). Let us consider the event

$$\mathcal{C}_{m,k} = \left\{\omega \in \mathcal{LI}(A_1, \ldots, A_k) : \mathrm{cone}\big(A_1, \ldots, A_k\big) \supseteq \mathrm{cone}\big(E_1, \ldots, E_k\big)\right\}. \qquad (29)$$

If $k = m$ we write $\mathcal{C}_m$ instead of $\mathcal{C}_{m,m}$. Note that in this case, $\mathfrak{as}\big(\mathrm{cone}(E_1, \ldots, E_m)\big) = 2^{-m}$. The following lemma shows thus that the angle space of the cone generated by the A_i is not too small with a quantifiable probability.

Lemma 7 *Let $A_1, \dots, A_m$ be i.i.d. random vectors with $A_i \sim \mathscr{U}(\mathrm{S}^{m-1})$. Then*

$$\mathbf{P}\left[\mathcal{C}_m\right] \geq 2^{-\frac{2+m(m-1)}{2}}, \qquad (m \geq 1). \tag{30}$$

PROOF. We proceed by induction over m. For $m = 1$ we have $\mathbf{P}\left[\mathrm{cone}(A_1) \supseteq \mathrm{cone}(E_1)\right] = 1 \geq 2^{-1}$, which shows that (30) holds true in the base case.

Suppose (30) holds true for $m-1$ and let us show that it holds for m. For almost all $\omega \in \Omega$,

$$A_m(\omega) \notin \mathrm{Span}\left(E_1(\omega), \dots, E_{m-1}(\omega)\right). \tag{31}$$

Let us thus assume that (31) holds and let

$$\pi_{E_m} : \mathrm{Span}\left(E_1(\omega), \dots, E_m(\omega)\right) \to \mathrm{Span}\left(E_1(\omega), \dots, E_{m-1}(\omega)\right)$$

denote the orthogonal projection along $E_m(\omega)$. Let

$$\mathcal{D}_m = \left\{\omega \in \Omega : \frac{\pi_{E_m}(-A_m)}{\|\pi_{E_m}(-A_m)\|} \in \mathrm{cone}\left(E_1, \dots, E_{m-1}\right)\right\}.$$

We claim that

$$\mathcal{D}_m \subseteq \left\{E_m \in \mathrm{cone}\big(E_1, \dots, E_{m-1}, A_m\big)\right\}. \tag{32}$$

In fact,

$$\frac{\pi_{E_m}(-A_m)}{\|\pi_{E_m}(-A_m)\|} \in \mathrm{cone}\left(E_1, \dots, E_{m-1}\right) \Leftrightarrow \pi_{E_m}(-A_m) \in \mathrm{cone}\left(E_1, \dots, E_{m-1}\right),$$

except on a nullset, and $\pi_{E_m}(-A_m) \in \mathrm{cone}\left(E_1, \dots, E_{m-1}\right)$ implies that there exist $\mu_i \geq 0$ $(i = 1, \dots, m-1)$ and $\mu_m > 0$ such that

$$A_m = \mu_m E_m + \pi_{E_m} A_m = \mu_m E_m - \sum_{i=1}^{m-1} \mu_i E_i.$$

Hence, $E_m = \mu_m^{-1}\big(A_m + \sum_{i=1}^{m-1} \mu_i E_i\big)$, which proves (32). Clearly (32) implies that

$$\mathcal{D}_m \cap \mathcal{C}_{m-1} \subseteq \mathcal{C}_m. \tag{33}$$

Let $\mathrm{G}_{m-1,m}$ be the Grassmannian of 1-codimensional linear subspaces of $\mathbb{R}^m$. $\mathrm{G}_{m-1,m}$ is a compact manifold with a transitive group action defined by the orthogonal group O_m. The $G_{m-1,m}$-valued random variable

$$G : \omega \mapsto \mathrm{Span}\left(A_1(\omega), \dots, A_{m-1}(\omega)\right) \in \mathrm{G}_{m-1,m} \tag{34}$$

has uniform distribution $\nu_{G_{m-1,m}}$, that is, $\nu_{G_{m-1,m}}$ is the unique probability measure on $G_{m-1,m}$ that is invariant under the group action of O_m. In fact, this follows trivially from the spatial symmetry of the joint distribution of the A_i $(i = 1, \dots, m-1)$. It follows likewise from this symmetry that for all $g \in \mathrm{G}_{m-1,m}$ the random vectors

$$\frac{\pi_{E_m}(-A_m)}{\|\pi_{E_m}(-A_m)\|}, A_1, \dots, A_{m-1}$$

are independent random variables when conditioned on the event $\big\{\omega \in \Omega : G(\omega) = g\big\}$, with uniform conditional distributions

$$\mathscr{D}\left(\frac{\pi_{E_m}(-A_m)}{\|\pi_{E_m}(-A_m)\|}\Big\|G(\omega) = g\right), \mathscr{D}\left(A_i\Big\|G(\omega) = g\right) \sim \mathscr{U}\left(\mathrm{S}^{m-1} \cap g\right) \equiv \mathscr{U}\left(\mathrm{S}^{m-2}\right). \tag{35}$$

These facts, (33) and the induction hypothesis finally imply

$$\begin{aligned}
\mathbf{P}\left[\mathcal{C}_m\right] &\geq \mathbf{P}\left[\mathcal{D}_m \cap \mathcal{C}_{m-1}\right] \\
&= \int_{G_{m-1,m}} \Big(\int_{y \in \otimes^{m-1} \mathrm{S}^{m-1} \cap g} \mathbf{P}\left[\mathcal{D}_m \middle\| (A_1, \ldots, A_{m-1}) = y, G = g\right] \otimes^{m-1} \nu_{m-2}(dy) \Big) \\
&\qquad \cdot \mathbf{P}\left[\mathcal{C}_{m-1} \middle\| G = g\right] \nu_{G_{m-1,m}}(dg) \\
&\geq \int_{G_{m-1,m}} \int_{y \in \otimes^{m-1} \mathrm{S}^{m-1} \cap g} 2^{-(m-1)} \otimes^{m-1} \nu_{m-2}(dy) \cdot \; 2^{-\frac{2+(m-1)(m-2)}{2}} \nu_{G_{m-1,m}}(dg) \\
&= 2^{-\frac{2+m(m-1)}{2}}.
\end{aligned}$$

□

The next lemma shows that the angle space defined on a 1-codimensional hyperplane does not change too much under an orthogonal projection into a nearby hyperplane.

Lemma 8 *Let $p_1, p_2 \in \mathrm{S}^{m-1}$, let us denote the angle space defined on $p_i^\perp = \{x \in \mathbb{R}^m : p_i \cdot x = 0\}$ by $\mathfrak{as}_i$ $(i = 1, 2)$, let $\pi_{p_2^\perp}$ be the orthogonal projection of $\mathbb{R}^m$ into $p_2^\perp$ along p_2 and let $\pi = \pi_{p_2^\perp}|_{p_1^\perp}$ be its restriction to $p_1^\perp$. Let finally K be a convex cone in $p_1^\perp$. Then*

$$\mathfrak{as}_2(\pi_2 K) \geq \mathfrak{as}_1(K)|p_1 \cdot p_2|.$$

Proof. If $p_1 \cdot p_2 = 0$ then the bound is trivial. Therefore, w.l.o.g. $p_1 \cdot p_2 \neq 0$ and then π is a vector space isomorphism between $p_1^\perp$ and $p_2^\perp$. Let $\{e_1, \ldots, e_{m-2}\}$ be an orthonormal basis of $p_1^\perp \cap p_2^\perp$, let $e_j^{(i)} = e_j$ $(j = 1, \ldots, m-2)$ and let $e_{m-1}^{(i)}$ be chosen so that $\{e_1^{(i)}, \ldots, e_{m-1}^{(i)}\}$ is an orthonormal basis of $p_i^\perp$ for $(i = 1, 2)$. Then $|e_{m-1}^{(1)} \cdot e_{m-1}^{(2)}| = |p_1 \cdot p_2|$. Let us express vectors in $p_1^\perp$ in terms of coordinates $y \in \mathbb{R}^{m-1}$ defined by linear combinations $\sum_{j=1}^{m-1} y_j e_j^{(1)}$. Likewise, let z be the coordinate system defined on $p_2^\perp$ by $\{e_1^{(2)}, \ldots, e_{m-1}^{(2)}\}$. Then π expressed in terms of y–z coordinates is the matrix

$$\pi = \begin{pmatrix} \mathrm{I} & 0 \\ 0 & e_{m-1}^{(2)} \cdot e_{m-1}^{(1)} \end{pmatrix} = \begin{pmatrix} \mathrm{I} & 0 \\ 0 & \pm|p_1 \cdot p_2| \end{pmatrix}.$$

Now let Z be a multivariate standard normal random vector on $p_2^\perp$, that is, Z has the density function

$$f_Z(z) = (2\pi)^{-\frac{(m-1)}{2}} \exp\left(-\frac{\sum_{i=1}^{m-1} z_i^2}{2} \right).$$

Then

$$\mathfrak{as}_2(\pi K) = \mathbf{P}\left[Z \in \pi(K)\right] = \mathbf{P}\left[Y \in K\right], \tag{36}$$

where $Y = \pi^{-1} Z$ has density

$$\begin{aligned}
f_Y(y) &= f_Z(z(y)) \left| \det\left(\frac{\partial z_i}{\partial y_j} \right) \right| \\
&= (2\pi)^{-\frac{m-1}{2}} \exp\left(-\frac{1}{2} \left(\sum_{i=1}^{m-2} y_i^2 + y_{m-1}^2 |p_1 \cdot p_2|^2 \right) \right) |p_1 \cdot p_2| \\
&\geq (2\pi)^{-\frac{m-1}{2}} \exp\left(-\frac{1}{2} \sum_{j=1}^{m-1} y_i^2 \right) |p_1 \cdot p_2|,
\end{aligned}$$

where the last inequality holds because $|p_1 \cdot p_2| < 1$. Therefore,

$$\mathbf{P}\left[Y \in K\right] \geq |p_1 \cdot p_2| \int_K (2\pi)^{-\frac{m-1}{2}} \exp\left(-\frac{\sum_{i=1}^{m-1} y_i^2}{2} \right) dy = |p_1 \cdot p_2| \mathfrak{as}_1(K).$$

Together with (36) this proves the lemma. □

The combination of Lemmas 3, 7 and 8 now allows us to derive lower bounds on the tail probabilities of $\mathscr{C}(A)$.

Theorem 2 *Let A be a uniform random $n \times m$ matrix where $m \geq 2$ and $n \geq m$. Then there exists a constant $c(m) > 0$ that depends only on m such that for all $t \geq 1/\cos(\pi/4)$ it is true that*

$$\mathbf{P}\left[\mathscr{C}(A) \geq t\right] \geq c(m) \cdot \left(\frac{I_{m-2}\left(\arccos\frac{1}{t}\right)}{I_{m-2}(\pi)}\right)^{n-m} \cdot \frac{1}{t}.$$

Proof. It follows from (35), the definition of G in Lemma 7 and the claim of the same result that

$$\begin{aligned}\mathbf{P}\left[\mathcal{C}_{m,k-1}\right] &= \int_{G_{m-1,m}} \mathbf{P}\left[\mathcal{C}_{m,k-1} \| G = g\right] \nu_{G_{m-1,m}}(dg) \\ &\geq 2^{-\frac{2+(m-1)(m-2)}{2}} \int_{G_{m-1,m}} \nu_{G_{m-1,m}}(dg) \\ &= 2^{-\frac{2+(m-1)(m-2)}{2}}. \end{aligned} \tag{37}$$

Since $\{A_1, \ldots, A_m\}$ and $\{E_1, \ldots, E_{m-1}, A_m\}$ are linearly independent sets for all $\omega \in \mathcal{C}_m$, it follows from Proposition 3 (i) that the vectors $\{A_1, \ldots, A_m\}$ define a unique SCP $\mathtt{cap}\big(P_A(\omega), R_A(\omega)\big)$, and likewise there exists a unique SCP $\mathtt{cap}\big(P_E(\omega), R_E(\omega)\big)$ corresponding to the set of vectors $\{E_1, \ldots, E_{m-1}, A_m\}$. Moreover, for all $\omega \in \mathcal{C}_{m,k-1}$ we have $\mathtt{cap}\big(P_E, R_E\big) \subseteq \mathtt{cap}\big(P_A, R_A\big)$, and hence,

$$0 \leq R_E(\omega) \leq R_A(\omega) \leq \frac{\pi}{2} \qquad \forall \omega \in \mathcal{C}_{m,k-1}. \tag{38}$$

Inequalities (37) and (38) imply that for all $t \geq 1$,

$$\begin{aligned}\mathbf{P}\left[R_A \geq \arccos 1/t\right] &\geq \mathbf{P}\left[R_A \geq \arccos 1/t \| \mathcal{C}_{m,k-1}\right] \mathbf{P}\left[\mathcal{C}_{m,k-1}\right] \\ &\geq \mathbf{P}\left[R_E \geq \arccos 1/t \| \mathcal{C}_{m,k-1}\right] \cdot 2^{-\frac{2+(m-1)(m-2)}{2}}. \end{aligned} \tag{39}$$

Let $\{e_1, \ldots, e_m\}$ be the canonical basis of $\mathbb{R}^m$. Then $\{e_1, \ldots, e_{m-1}, A_m\}$ is linearly independent almost surely, and it follows from Proposition 3 (i) that there exists a unique SCP $\mathtt{cap}(P, R)$ that corresponds to these vectors. Since A_m is independent of the event $\mathcal{C}_{m,k-1}$, which is defined entirely in terms of $A_1, \ldots, A_{m-1}$, and since the invariance of $\mathscr{D}(A_i)$ under the action of the orthogonal group O_m on S^{m-1} implies that $[E_1, \ldots, E_{m-1}]$ is uniformly distributed on the Stiefel manifold V_{m-1} of $m \times (m-1)$ matrices with orthonormal columns, we have

$$\mathbf{P}\left[R_E \geq \arccos 1/t \middle\| \mathcal{C}_{m,k-1}\right] = \mathbf{P}\left[R \geq \arccos 1/t\right]. \tag{40}$$

We will consider the unit vectors

$$\overline{e} = \frac{1}{\sqrt{m-1}} \sum_{i=1}^{m-1} e_i \quad \text{and} \quad p_\vartheta = \cos\vartheta \cdot e_m + \sin\vartheta \cdot \overline{e} \quad \text{for} \quad \vartheta \in [-\pi/2, \pi/2].$$

Then a random angle $\Theta : \Omega \to [-\pi/2, \pi/2)$ is defined almost everywhere by the condition $A_m(\omega) \in p^\perp_{\Theta(\omega)} + e_1$. In fact,

$$\Theta = \begin{cases} \arctan \frac{\sqrt{m-1} A_m \cdot e_m}{1 - \sum_{i=1}^{m-1} A_m \cdot e_i} & \text{if } \sum_{i=1}^{m-1} A_m \cdot e_i \neq 1, \\ -\frac{\pi}{2} & \text{if } \sum_{i=1}^{m-1} A_m \cdot e_i = 1 \text{ and } A_m \cdot e_m \neq 0, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Note that $p^\perp_\vartheta + e_1 = p^\perp_\vartheta + e_i$ $(i = 2, \ldots, m-1)$ for all $\vartheta \in [-\pi/2, \pi/2)$, that is, the definition of Θ is symmetric with respect to the e_i.

It is easy to see that Θ has a continuous density function $f_\Theta > 0$ such that $f_\Theta(-\vartheta) = f_\Theta(\vartheta)$ for all $\vartheta \in (-\pi/2, \pi/2)$, and there exists a constant $c_\Theta > 0$ such that $f_\Theta(\vartheta) \geq c_\Theta$ for all ϑ in the compact set $[-\pi/4, \pi/4]$.

Note that one can parameterise the sphere S^{m-1} by $\mathrm{S}^{m-2} \times [-\pi/2, \pi/2)$ via

$$\left(\sum_{i=1}^{m-1} \lambda_i e_i, \vartheta\right) \mapsto \left(\sum_{j=1}^{m} \mu_j e_j\right) \cdot \cos\vartheta + \frac{1}{\sqrt{m-1}} p_\vartheta,$$

for $\sum_{i=1}^{m-1} \lambda_i e_i \in \mathrm{S}^{m-2} \subset \mathrm{Span}\,(e_1, \ldots, e_{m-1})$, that is, $\sum_{i=1}^{m-1} \lambda_i^2 = 1$, and where

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = \begin{pmatrix} \frac{\cos\vartheta+(m-2)}{m-1} & \frac{\cos(\vartheta)-1}{m-1} & \cdots & \frac{\cos(\vartheta)-1}{m-1} & \frac{\sin\vartheta}{\sqrt{m-1}} \\ \frac{\cos(\vartheta)-1}{m-1} & \frac{\cos\vartheta+(m-2)}{m-1} & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\cos(\vartheta)-1}{m-1} & \frac{\cos(\vartheta)-1}{m-1} & \cdots & \frac{\cos\vartheta+(m-2)}{m-1} & \frac{\sin\vartheta}{\sqrt{m-1}} \\ -\frac{\sin\vartheta}{\sqrt{m-1}} & -\frac{\sin\vartheta}{\sqrt{m-1}} & \cdots & -\frac{\sin\vartheta}{\sqrt{m-1}} & \cos\vartheta \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{m-1} \\ 0 \end{pmatrix}.$$

Note also that the matrix appearing in the display is orthogonal with last column corresponding to p_ϑ. Thus, the chosen parameterisation corresponds to tilting the unit sphere $\mathrm{S}^{m-2} \subset \mathrm{Span}\,(e_1, \ldots, e_{m-1})$ by an angle ϑ about the affine hull $\mathrm{aff}(e_1, \ldots, e_{m-1})$ and shrinking it by $\cos\vartheta$ to fit the radius of the sphere cut out of S^{m-1} by the tilted plane.

This parameterisation defines a conditional distribution $\mathscr{D}(A_m \| \Theta = \vartheta)$ on S^{m-2} with continuous Radon-Nikodym derivative $f_{A_m\|\Theta}$ with respect to ν_{m-2}. Moreover, $f_{A_m\|\Theta}(x\|\vartheta) = 0$ if and only if $x \in \mathrm{aff}(e_1, \ldots, e_{m-1})$. Therefore, there exists a constant $c_A > 0$ such that

$$f_{A_m\|\Theta}(x\|\vartheta) \geq c_A \tag{41}$$

for all (x, ϑ) in the compact set $\left\{x \in \mathrm{S}^{m-1} : x \cdot \overline{e} \leq 0\right\} \times [-\pi/4, \pi/4]$.

Now Lemma 3 shows that for all $t \geq 1/\cos(\pi/4)$,

$$\begin{aligned}
\mathbf{P}\left[R \geq \arccos(1/t)\right] &= \mathbf{P}\big[\big(\Theta \in [\arccos 1/t - \pi/2, \pi/2 - \arccos 1/t]\big) \\
&\qquad \wedge \big(-\pi_{p_\Theta^\perp} A_m \in \mathrm{cone}(\pi_{p_\Theta^\perp} e_1, \ldots, \pi_{p_\Theta^\perp} e_{m-1})\big)\big)\big] \\
&= \int_{\arccos\frac{1}{t}-\frac{\pi}{2}}^{\frac{\pi}{2}-\arccos\frac{1}{t}} \int_{\pi_{p_\Theta^\perp}(-\mathrm{cone}(e_1,\ldots,e_{m-1}))} f_{A_m\|\Theta}(x\|\vartheta) f_\Theta(\vartheta) \nu_{m-2}(dx) d\vartheta \\
&\overset{(41),\text{Lem } 8}{\geq} \int_{\arccos\frac{1}{t}-\frac{\pi}{2}}^{\frac{\pi}{2}-\arccos\frac{1}{t}} c_A \cdot 2^{-(m-1)} |\cos\vartheta| \cdot f_\Theta(\vartheta) d\vartheta \\
&\geq c_A c_\Theta \cdot \left(2 \int_0^{\frac{\pi}{2}-\arccos\frac{1}{t}} \cos\vartheta d\vartheta\right) \cdot 2^{-(m-1)} \\
&= c_A c_\Theta \cdot 2^{-(m-2)} \cdot \frac{1}{t}.
\end{aligned}$$

Therefore, (39) and (40) imply that

$$\mathbf{P}\left[R_A \geq \arccos 1/t\right] \geq 2^{-\left(1+\frac{(m-1)(m-2)}{2}\right)} \cdot c_A c_\Theta \cdot 2^{-(m-2)} \cdot \frac{1}{t}. \tag{42}$$

Finally, if $A_{m+1}, \ldots, A_n \in \mathtt{cap}(P_A, R_A)$ then $\mathtt{cap}(P_A, R_A)$ is the SCP of $\{A_1, \ldots, A_n\}$ and it follows from the remarks of Section 4 that $\mathscr{C}(A) = |\cos R_A|^{-1}$. Therefore,

$$\begin{aligned}
\mathbf{P}\left[\mathscr{C}(A) \geq t\right] &\geq \mathbf{P}\left[(A_{m+1}, \ldots, A_n \in \mathtt{cap}(P_A, R_A)) \wedge (R_A \geq \arccos 1/t)\right] \\
&\geq \left(\frac{I_{m-2}\left(\arccos\frac{1}{t}\right)}{I_{m-2}(\pi)}\right)^{n-m} \cdot 2^{-\left(1+\frac{(m-1)(m-2)}{2}\right)} \cdot c_A c_\Theta \cdot 2^{-(m-2)} \cdot \frac{1}{t}.
\end{aligned}$$

Since c_A and c_Θ depend only on m, this proves the claim of the theorem. □

8 Exact Tail Decay Rates

The decay rates of $\mathbf{P}[\mathscr{C}(A) \geq t]$ developed in Sections 6 and 7 give an estimate on the rarity of a large backward error, high instability or long running times for some algorithms applied to a random linear feasibility problem drawn from Gaussian data. Moreover, as mentioned in the introduction, the best upper bounds on the running time of modern linear programming or linear feasibility solvers applied to real input data are polynomial in the problem dimension and $\log \mathscr{C}(A)$. We are therefore also interested in estimates on probability tails

$$\mathbf{P}[\log \mathscr{C}(A) \geq t] \tag{43}$$

for $t \gg 1$.

Theorem 1 implies that

$$\mathbf{P}[\log \mathscr{C}(A) \geq t] = \mathbf{P}[\mathscr{C}(A) \geq \mathrm{e}^t] \leq \binom{n}{m} 2m^{\frac{5}{2}} \left(1 - \frac{I_{m-2}(\arccos(\mathrm{e}^{-t}))}{I_{m-2}(\pi)}\right)^{n-m} \cdot \mathrm{e}^{-t}. \tag{44}$$

On the other hand, Theorem 2 shows

$$\mathbf{P}[\log \mathscr{C}(A) \geq t] \geq c(m) \left(\frac{I_{m-2}(\arccos(\mathrm{e}^{-t}))}{I_{m-2}(\pi)}\right)^{n-m} \cdot \mathrm{e}^{-t}.$$

Since $I_{m-2}(\arccos \mathrm{e}^{-t})/I_{m-2}(\pi)$ increases monotonically to $1/2$ for $t \to \infty$, these formulas show that the exponential decay rate of (43) is exactly -1.

Corollary 1 *If A is a random uniform $n \times m$ matrix then*

$$\lim_{t \to \infty} \frac{\log \mathbf{P}[\log \mathscr{C}(A) \geq t]}{t} = -1.$$

Proof. The proof is immediate from the arguments above. □

Thus, although the multiplicative constant in (44) is too large, the formula captures the correct qualitative behaviour of the tails of $\log \mathscr{C}(A)$ and the best possible upper bound on (43) must be of the form

$$\mathbf{P}[\log \mathscr{C}((A) \geq t] \leq c(m, n) \cdot \mathrm{e}^{-t} \tag{45}$$

for some constant $c(m, n)$ that depends on m and n.

The exponential decay of $\mathbf{P}[\log \mathscr{C}((A) \geq t]$ shows that the linear feasibility problem, and by extension linear programming, is "empirically strongly polynomial". See Section 2 for further comments on this important point.

9 Moment Estimates

The probabilistic analysis of linear programming is primarily concerned with the average running time of LP algorithms on random input data. Because complexity bounds for interior-point methods are polynomial in $\log \mathscr{C}(A)$ (see the introduction), upper bounds on the expectation, the variance and higher moments of the running time are easily derived from upper bounds on the corresponding moments of $\log \mathscr{C}(A)$.

Since $\mathbf{E}[X] = \int_0^\infty \mathbf{P}[X > x]dx$ for any random variable X that takes only nonnegative values, the estimate (44) can be used to derive upper bounds on all moments of $\log \mathscr{C}(A)$. Indeed, (45) shows that for all $\gamma > 0$,

$$\int_0^\infty \mathbf{P}\left[(\log \mathscr{C}(A))^\gamma \geq t\right] dt \leq \int_0^\infty c(n,m)\mathrm{e}^{-t^{\frac{1}{\gamma}}}\, dt = c(n,m)\Gamma(\gamma+1) < \infty,$$

that is, all moments of $\mathscr{C}(A)$ are finite. To turn this into a quantitative estimate, we consider the function $\varphi : \mathbb{R}_+ \to (\frac{1}{2}, 1]$ defined as follows:

$$\varphi(t) = 1 - \frac{I_{m-2}(\arccos(\mathrm{e}^{-t}))}{I_{m-2}(\pi)}.$$

Note that φ is continuous, and strictly decreasing with $\varphi(0) = 1$ and $\lim_{t\to\infty} \varphi(t) = \dfrac{1}{2}$. Let us define

$$f(m,n) = \varphi^{-1}\left((1/2)^{1/\sqrt{n}}\right).$$

Since $f(m,n) > 0$, the following result follows.

Corollary 2 *Let A be a uniform random $n \times m$ matrix with $n \geq m \geq 3$. Then for all $\gamma \in \mathbb{R}_+$ the γ-th moment of $\log \mathscr{C}(A)$ is bounded by*

$$\mathbf{E}\left[(\log \mathscr{C}(A))^\gamma\right] \leq f(m,n)^\gamma + \binom{n}{m} 2m^{\frac{5}{2}} 2^{-\frac{n-m}{\sqrt{n}}} \Gamma(\gamma+1).$$

Proof. Using (44), we find

$$\begin{aligned}
\mathbf{E}\left[(\log \mathscr{C}(A))^\gamma\right] &= \int_0^\infty \mathbf{P}\left[(\log \mathscr{C}(A))^\gamma > t\right] dt \\
&\leq f(m,n)^\gamma + \int_{f(m,n)^\gamma}^\infty \binom{n}{m} 2m^{\frac{5}{2}} \left(1 - \frac{I_{m-2}\big(\arccos(\mathrm{e}^{-f(m,n)})\big)}{I_{m-2}(\pi)}\right)^{n-m} \cdot \mathrm{e}^{-t^{\frac{1}{\gamma}}}\, dt \\
&\leq f(m,n)^\gamma + \binom{n}{m} 2m^{\frac{5}{2}} 2^{-\frac{n-m}{\sqrt{n}}} \int_{f(m,n)^\gamma}^\infty \mathrm{e}^{-t^{\frac{1}{\gamma}}}\, dt \\
&\leq f(m,n)^\gamma + \binom{n}{m} 2m^{\frac{5}{2}} 2^{-\frac{n-m}{\sqrt{n}}} \Gamma(\gamma+1).
\end{aligned}$$

□

In Section 10 we will see that the bounds of Corollary 2 are particularly useful for understanding the behaviour of $\mathscr{C}(A)$ when $n \gg m$. Note however that these bounds grow exponentially in m. One of the major objectives of the probabilistic analysis of linear programming is to show that the expected running times of particular families of algorithms are bounded by a polynomial of the dimension of the input data. Such results are often interpreted in the light of "average strongly polynomiality" of linear programming.

Does the exponential growth of the estimates from Corollary 2 mean that Theorem 1 fails to lead to "average strong polynomiality" results when used to bound the complexity of interior-point algorithms for example? Not in the least! The exponential growth of the estimates from Corollary 2 is purely a consequence of our definition of the cut-off point $f(m,n)$, which we chose so as to converge to zero as n tends to infinity to fit the purposes of the limit theorems of Section 10. Giving up on this condition one can easily derive linear bounds:

Lemma 9 *Let $\{X_{m,n} : (m,n) \in \mathbb{N} \times \mathbb{N}\}$ be a set of random variables and $\gamma \geq 1$ a real number. Furthermore, let $p(m,n)$ and $t(m,n)$ be functions of m and n such that*

$$\mathbf{P}[X_{m,n} > t] \leq \mathrm{e}^{p(m,n)-t^{\frac{1}{\gamma}}} \qquad \forall\, t \geq t(m,n).$$

Then

$$\mathbf{E}[X_{m,n}] \leq \max\big(p(m,n)^\gamma, t(m,n)\big) + \Gamma(\gamma+1)2^{\gamma-1}.$$

Proof.

$$\begin{aligned}
\mathbf{E}[X] &= \int_0^\infty \mathbf{P}[X > t]\, dt \\
&\leq \int_0^{\max(p(m,n)^\gamma, t(m,n))} 1 dt + \int_{\max(p(m,n)^\gamma, t(m,n))}^\infty \exp\left\{p(m,n) - t^{\frac{1}{\gamma}}\right\} dt \\
&\leq \max\big(p(m,n)^\gamma, t(m,n)\big) + \int_0^\infty \exp\left\{-t^{\frac{1}{\gamma}} 2^{1-\frac{1}{\gamma}}\right\} dt \qquad (46) \\
&= \max\big(p(m,n)^\gamma, t(m,n)\big) + \Gamma(\gamma+1)2^{\gamma-1},
\end{aligned}$$

where (46) holds true because we claim that

$$p(m,n) - t^{\frac{1}{\gamma}} 2^{1-\frac{1}{\gamma}} \leq -\left(t - \max\big(p(m,n)^\gamma, t(m,n)\big)\right)^{\frac{1}{\gamma}}$$

for all $t \geq \max\big(p(m,n)^\gamma, t(m,n)\big)$. In fact, $x \mapsto x^{\frac{1}{\gamma}}$ is a concave function, since $\gamma \geq 1$. Therefore,

$$\begin{aligned}
&\frac{1}{2}\left(p(m,n)^\gamma\right)^{\frac{1}{\gamma}} + \frac{1}{2}\left(t - \max\big(p(m,n)^\gamma, t(m,n)\big)\right)^{\frac{1}{\gamma}} \\
&\qquad \leq 2^{-\frac{1}{\gamma}}\left(p(m,n)^\gamma + t - \max\big(p(m,n)^\gamma, t(m,n)\big)\right)^{\frac{1}{\gamma}} \\
&\qquad \leq 2^{-\frac{1}{\gamma}} t^{\frac{1}{\gamma}},
\end{aligned}$$

which shows that

$$p(m,n) + \left(t - \max\big(p(m,n)^\gamma, t(m,n)\big)\right)^{\frac{1}{\gamma}} \leq 2^{1-\frac{1}{\gamma}} t^{\frac{1}{\gamma}}$$

and proves our claim. □

Corollary 3 *Let A be a uniform random $n \times m$ matrix where $n \geq m \geq 3$, and let $\gamma \geq 1$ be a real number. Then*

$$\mathbf{E}[(\log \mathscr{C}(A))^\gamma] \leq \left(m \log n + \frac{5}{2}\log m + \log 2\right)^\gamma + \Gamma(\gamma+1)2^{\gamma-1}.$$

In particular,

$$\begin{aligned}
\mathbf{E}[\log \mathscr{C}(A)] &\leq m \log n + \frac{5}{2}\log m + \log 2 + 1, \\
\mathbf{VAR}(\log \mathscr{C}(A)) &\leq \left(m \log n + \frac{5}{2}\log m + \log 2\right)^2 + 4.
\end{aligned}$$

Proof. Equation (44) shows that for all $t \geq t(m,n) = 1$,

$$\begin{aligned}
\mathbf{P}[(\log \mathscr{C}(A))^\gamma > t] &= \mathbf{P}\left[\log \mathscr{C}(A) > t^{\frac{1}{\gamma}}\right] \\
&\leq \binom{n}{m} 2m^{\frac{5}{2}} \left(1 - \frac{I_{m-2}\big(\arccos(\exp\{-t^{\frac{1}{\gamma}}\})\big)}{I_{m-2}(\pi)}\right)^{n-m} \cdot \mathrm{e}^{-t^{\frac{1}{\gamma}}} \\
&\leq n^m 2m^{\frac{5}{2}} \mathrm{e}^{-t^{\frac{1}{\gamma}}} \\
&\leq \mathrm{e}^{m\log n + \frac{5}{2}\log m + \log 2 - t^{\frac{1}{\gamma}}}.
\end{aligned}$$

The first claim now follows from Lemma 9. Finally, since

$$\mathbf{VAR}(\log \mathscr{C}(A)) = \mathbf{E}[(\log \mathscr{C}(A))^2] - \mathbf{E}[\log \mathscr{C}(A)]^2 \leq \mathbf{E}[(\log \mathscr{C}(A))^2],$$

the last two claims are special cases of the first claim. □

Corollary 3 recovers the main result in [11]. However, it still does not fully exhaust the potential power of Theorem 1 and Lemma 9: indeed, in Section 11 we will further strengthen Corollary 3 and show that $\mathbf{E}[\log \mathscr{C}(A)]$ is asymptotically bounded by $m \log 2$ for arbitrary $n \geq m$, and by $\mathcal{O}(m^\gamma)$ for any $\gamma > 0$ when $n \geq 5m$.

Remark 2 Let us briefly remark here that the main reason for the appearance of the binomial term $\binom{n}{m}$ in the bound of Theorem 1 is a lack of proper understanding of $\mathscr{D}(A_i \| P = p)$, where P is the centre of an LCP of $A = [A_1, \ldots, A_n]^{\mathrm{T}}$. We suspect that knowledge of this conditional distribution would make it possible to replace $\binom{n}{m}$ by a polynomial term in m and n. If this hunch were true then the bound of Corollary 3 would of course become logarithmic in m and n.

Let us finally investigate the moments of $\mathscr{C}(A)$ itself, which is interesting in its own right for reasons mentioned in the introduction.

Corollary 4 *Let A be a uniform random $n \times m$ matrix where $n \geq m \geq 3$. Then*

$$\mathbf{E}[\mathscr{C}(A)^\gamma] \begin{cases} = +\infty & \text{if } \gamma \geq 1, \\ \leq 1 + \binom{n}{m} 2m^{\frac{5}{2}} \frac{\gamma}{1-\gamma} & \text{if } \gamma \in (0,1). \end{cases}$$

Proof. Theorem 2 shows that for all $t \geq (\cos(\pi/4))^{-\gamma}$,

$$\mathbf{P}\left[\mathscr{C}(A)^\gamma > t\right] \geq c(m) \left(\frac{I_{m-2}\left(\arccos t^{-\frac{1}{\gamma}}\right)}{I_{m-2}(\pi)} \right)^{n-m} t^{-\frac{1}{\gamma}}.$$

Therefore,

$$\begin{aligned} \mathbf{E}\left[\mathscr{C}(A)^\gamma\right] &= \int_0^\infty \mathbf{P}\left[\mathscr{C}(A)^\gamma > t\right] dt \\ &\geq \int_{\left(\cos(\frac{\pi}{4})\right)^{-\gamma}}^\infty c(m) 2^{m-n} t^{-\frac{1}{\gamma}} dt = \infty \qquad \forall \gamma \geq 1. \end{aligned}$$

On the other hand, using Theorem 1, we find that for $\gamma \in (0,1)$,

$$\begin{aligned} \mathbf{E}\left[\mathscr{C}(A)^\gamma\right] &= 1 + \int_1^\infty \mathbf{P}\left[\mathscr{C}(A)^\gamma > t\right] dt \\ &\leq 1 + \binom{n}{m} 2m^{\frac{5}{2}} \int_1^\infty t^{-\frac{1}{\gamma}} dt \\ &= 1 + \binom{n}{m} 2m^{\frac{5}{2}} \frac{\gamma}{1-\gamma}. \end{aligned}$$

□

Note that the moment bound of Corollary 4 grows exponentially in n for $\gamma < 1$. This bound does not reflect the correct limiting behaviour, since we will show in Corollary 7 below that $\lim_{n\to\infty} \mathbf{E}\left[\mathscr{C}(A)^\gamma\right] = 1$ occurs for $\gamma < 1$ and m fixed.

10 Limit Theorems for $n \gg m$

In this section we investigate the behaviour of $\mathscr{C}(A)$ and $\log \mathscr{C}(A)$ in the situation where $n \gg m$. For example, we will show that for fixed m,

$$\mathscr{C}(A) \overset{n\to\infty}{\longrightarrow} 1 \tag{47}$$

with probability 1, and

$$\mathbf{E}\left[\log \mathscr{C}(A)\right] \overset{n\to\infty}{\longrightarrow} 0. \tag{48}$$

Intuitively it is clear that this behaviour should be observed: whenever the system $Ax \leq 0, x \neq 0$ contains a very large numbers of random constraints, the system should be infeasible and this infeasibility should be easy to detect algorithmically. Our results confirm this intuition.

In the results below $m \geq 3$ is a fixed dimension, $(A_i)_{\mathbb{N}}$ denotes a sequence of i.i.d. random vectors with uniform distribution on the sphere, $\mathscr{D}(A_i) \sim \mathscr{U}(\mathrm{S}^{m-1})$, and $(A^{[n]})_{\mathbb{N}}$ is the sequence of random matrices $A^{[n]} = [A_1, \ldots, A_n]^{\mathrm{T}}$.

Theorem 3 *Let $(A_i)_{\mathbb{N}}$ and $(A^{[n]})_{\mathbb{N}}$ be as defined above. Then*

$$\mathbf{P}\left[\lim_{n\to\infty} \mathscr{C}(A^{[n]}) = 1\right] = 1.$$

Proof. For all $\omega \in \Omega$ and $n \in \mathbb{N}$ let $\mathtt{cap}(P_n(\omega), R_n(\omega))$ be a LCP of $A^{[n]}$. By virtue of Proposition 2 it suffices to prove that

$$\mathbf{P}\left[\lim_{n\to\infty} R_n = 0\right] = 1. \tag{49}$$

Let $\rho \in (0, \pi/2)$ be a fixed radius. Since S^{m-1} is compact, there exists a finite set of vectors $\{p_1, \ldots, p_k\} \subset \mathrm{S}^{m-1}$ such that $\bigcup_{i=1}^{k} \mathtt{cap}(p_i, \rho/2) = \mathrm{S}^{m-1}$. By

$$\mathcal{CE}_{i,n} = \{\omega \in \Omega : A_1, \ldots, A_n \notin \mathtt{cap}(p_i, \rho/2)\}$$

let us denote the event that the i-th cap does not contain any of the n first vectors of $(A_i)_{\mathbb{N}}$. Then

$$\mathbf{P}\left[\mathcal{CE}_{i,n}\right] = \left(\frac{I_{m-2}(\pi - \rho/2)}{I_{m-2}(\pi)}\right)^n,$$

and hence,

$$\mathbf{P}\left[\bigcup_{i=1}^{k} \mathcal{CE}_{i,n}\right] \leq \sum_{i=1}^{k} \mathbf{P}[\mathcal{CE}_{i,n}] = k \cdot \left(\frac{I_{m-2}(\pi - \rho/2)}{I_{m-2}(\pi)}\right)^n. \tag{50}$$

We now claim that

$$\{\omega \in \Omega : R_n \geq \rho\} \subseteq \bigcup_{i=1}^{k} \mathcal{CE}_{i,n}. \tag{51}$$

In fact, if $\omega \in \left(\bigcup_{i=1}^{k} \mathcal{CE}_{i,n}\right)^c$, the complement of $\bigcup_{i=1}^{k} \mathcal{CE}_{i,n}$, then there exist indices $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, n\}$ such that $P_n(\omega) \in \mathtt{cap}(p_i, \rho/2)$ and $A_j(\omega) \in \mathtt{cap}(p_i, \rho/2)$. Using the triangular inequality on the sphere we find $R_n \leq \arccos\big(P_n(\omega) \cdot A_j(\omega)\big) < 2 \cdot \rho/2$. This shows

$$\{\omega \in \Omega : R_n < \rho\} \supseteq \left(\bigcup_{i=1}^{k} \mathcal{CE}_{i,n}\right)^c,$$

which is equivalent to our claim.

Now (50) and (51) show

$$\mathbf{P}\left[R_n \geq \rho\right] \leq k \cdot \left(\frac{I_{m-2}(\pi - \rho/2)}{I_{m-2}(\pi)}\right)^n \overset{n\to\infty}{\longrightarrow} 0.$$

Finally, since $n_1 \leq n_2$ implies $R_{n_1} \geq R_{n_2}$, we have

$$0 \leq \mathbf{P}\left[\lim_{n\to\infty} R_n > \rho\right] \leq \lim_{n\to\infty} \mathbf{P}\left[R_n > \rho\right] = 0.$$

Since this is true for all $\rho \in (0, \pi/2)$, (49) follows. □

Corollary 5 *Let $(A^{[n]})_{\mathbb{N}}$ be as above. Then*

i) $\mathscr{C}(A^{[n]}) \overset{n\to\infty}{\longrightarrow_P} 1$,

ii) $\mathscr{C}(A^{[n]}) \overset{n\to\infty}{\Longrightarrow} 1$,

iii) $\mathbf{P}\left[\lim_{n\to\infty} \log \mathscr{C}(A^{[n]}) = 0\right] = 1$,

iv) $\log \mathscr{C}(A^{[n]}) \overset{n\to\infty}{\longrightarrow_P} 0$,

v) $\log \mathscr{C}(A^{[n]}) \overset{n\to\infty}{\Longrightarrow} 0$,

where $\longrightarrow_P$ denotes convergence in probability and $\Rightarrow$ denotes weak convergence.

Proof. These are all standard consequences of Theorem 3, see for example Theorem 25.2 in [4]. □

Using (44) and Corollary 5 one can analyse the asymptotic behaviour of $\mathbf{E}[\mathscr{C}(A)]$ and $\mathbf{E}[\log \mathscr{C}(A)]$, see Corollary 7 below. In the case of $\log \mathscr{C}(A)$ for example, one can show that $\lim_{n\to\infty} \mathbf{E}[\log \mathscr{C}(A^{[n]})] = 0$, using Skorhohod's theorem. Remarkably, the estimates of Corollary 2 are strong enough to yield this result directly, without resort to Theorem 3.

Corollary 6 *Let m be fixed and $(A_i)_{\mathbb{N}}$ and $(A^{[n]})_{\mathbb{N}}$ defined as above. Then*

$$\lim_{n\to\infty} \mathbf{E}\left[(\log \mathscr{C}(A^{[n]}))^{\gamma}\right] = 0, \qquad \forall \gamma > 0.$$

In particular,

i) $\lim_{n\to\infty} \mathbf{E}\big[\log \mathscr{C}(A^{[n]})\big] = 0$ *and*

ii) $\lim_{n\to\infty} \mathbf{VAR}\left(\log \mathscr{C}(A^{[n]})\right) = 0.$

Proof. Since $2^{-(n-m)/\sqrt{n}}$ is exponentially decreasing in n and

$$\binom{n}{m} \leq n^m$$

increases only polynomially in n, the second term in the estimate of Corollary 2 tends to zero as n tends to infinity. In addition, $f(m,n)$ tends to zero as n tends to infinity by definition of $f(m,n)$. This proves the displayed formula and as a particular case part *i)*. Part *ii)* follows from the display and the fact that $\mathbf{VAR}(\log \mathscr{C}(A)) \leq \mathbf{E}[(\log \mathscr{C}(A))^2]$. □

Let us now analyse the asymptotic behaviour of $\mathbf{E}\left[\mathscr{C}(A)\right]$. Recall from Corollary 4 that $\mathbf{E}[\mathscr{C}(A)^{\gamma}]$ is finite if and only if $\gamma < 1$.

Corollary 7 *Let m be fixed and $(A_i)_{\mathbb{N}}$ and $(A^{[n]})_{\mathbb{N}}$ defined as above. Then*

$$\lim_{n\to\infty} \mathbf{E}\big[(\mathscr{C}(A^{[n]}))^\gamma\big] = 1, \qquad \forall \gamma \in [0,1).$$

PROOF. The result is trivial for $\gamma = 0$. Let us therefore assume that $\gamma \in (0,1)$. Let $t_0 \in \mathbb{R}_+$ be large enough so that

$$1 - \frac{I_{m-2}\left(\arccos t^{-\frac{1}{\gamma}}\right)}{I_{m-2}(\pi)} \leq \frac{3}{4}, \qquad \forall t \geq t_0.$$

Since $\mathscr{C}(A^{[n]}) \to_P 1$ by Corollary 5, for all $\epsilon > 0$ there exists a number $n_\epsilon \in \mathbb{N}$ such that

$$\mathbf{P}\left[\mathscr{C}(A^{[n]}) > (1+\epsilon)^{\frac{1}{\gamma}}\right] < \epsilon, \qquad \forall n \geq n_\epsilon. \tag{52}$$

Therefore, for $n \geq n_\epsilon$,

$$\begin{aligned}
\mathbf{E}\left[\left(\mathscr{C}(A^{[n]})\right)^\gamma\right] &= \int_0^\infty \mathbf{P}\left[\left(\mathscr{C}(A^{[n]})\right)^\gamma > t\right] dt \\
&\leq \int_0^{(1+\epsilon)^{\frac{1}{\gamma}}} 1dt + \int_{(1+\epsilon)^{\frac{1}{\gamma}}}^{t_0} \mathbf{P}\left[\mathscr{C}(A^{[n]}) > (1+\epsilon)^{\frac{1}{\gamma}}\right] dt + \int_{t_0}^\infty \mathbf{P}\left[\mathscr{C}(A^{[n]}) > t^{\frac{1}{\gamma}}\right] dt \\
&\overset{\text{Thm1,(52)}}{\leq} (1+\epsilon)^{\frac{1}{\gamma}} + \epsilon\left(t_0 - (1+\epsilon)^{\frac{1}{\gamma}}\right) + \binom{n}{m} 2m^{\frac{5}{2}} \left(\frac{3}{4}\right)^{n-m} \int_{t_0}^\infty t^{-\frac{1}{\gamma}} dt \\
&= (1+\epsilon)^{\frac{1}{\gamma}} + \epsilon\left(t_0 - (1+\epsilon)^{\frac{1}{\gamma}}\right) + 2n^m m^{\frac{5}{2}} \left(\frac{3}{4}\right)^{n-m} \frac{t_0^{1-\frac{1}{\gamma}}}{\frac{1}{\gamma} - 1}.
\end{aligned}$$

Taking limits as $n \to \infty$ and observing that $\epsilon > 0$ was arbitrary, the claim follows. □

11 Limit Theorems for $m \gg 1$

In this section we investigate the behaviour of $\log \mathscr{C}(A)$ in the situation where $m \gg 1$. We will see that $\limsup_{m\to\infty} \mathbf{E}[\log \mathscr{C}(A)]/m \leq \log 2$, and we point out why we suspect that the correct value of this limit is zero.

Let $(\rho_m)_{\mathbb{N}} \subset (0, \pi/2]$ be a sequence such that $\lim_{m\to\infty} \rho_m = \pi/2$, and let $(A^{[m]})_{\mathbb{N}}$ be a sequence of random vectors such that $A^{[m]} \sim \mathscr{U}(\mathrm{S}^{m-1})$. It can be shown that if ρ_m converges to $\pi/2$ at an algebraic rate as a function of m, then

$$\lim_{m\to\infty} \mathbf{P}\left[A^{[m]} \in \mathtt{cap}(p, \rho_m)\right] = 0.$$

This effect is a special case of the so-called *concentration of measure phenomenon*, see e.g. [26]. The phenomenon is remarkable, because it implies that after fixing an equator by choosing an arbitrary grand circle on a high dimensional sphere, one will observe that a counter-intuitively high proportion of uniformly drawn sample points from that sphere lie in a very narrow neighbourhood around that equator. This phenomenon affects the analysis of the distribution tails of $\mathscr{C}(A)$ for large m.

However, for the purposes of this analysis it suffices to know that for any fixed real exponent $\gamma > 0$,

$$\lim_{m\to\infty} \frac{I_{m-2}\left(\arccos \mathrm{e}^{-m^\gamma}\right)}{I_{m-2}(\pi)} = \frac{1}{2}. \tag{53}$$

An elementary proof of this fact can be found in Lemma 10 of Appendix A.

Corollary 8 *For all (m,n) such that $m \leq n$ let $A^{[m,n]}$ be a uniform random $n \times m$ matrix. Then*

$$\limsup_{m\to\infty}\left(\sup_{n\geq m}\frac{\mathbf{E}\left[\log\mathscr{C}(A^{[m,n]})\right]}{m}\right)\leq\log 2.$$

PROOF. Let $\epsilon > 0$ be a small real number and X a binomially distributed random variable $X \sim \mathbf{Bin}(m+k,(1-\epsilon)/2)$. If Φ denotes the cumulative distribution function of the standard normal distribution, then the central limit theorem shows that

$$\begin{aligned}
&\binom{m+k}{m}\left(\frac{1+\epsilon}{2}\right)^k=\left(\frac{2}{1-\epsilon}\right)^m\mathbf{P}\left[X=m\right]\\
&=\left(\frac{2}{1-\epsilon}\right)^m\mathbf{P}\left[\frac{X-\frac{(m+k)(1-\epsilon)}{2}}{\frac{\sqrt{(m+k)(1-\epsilon^2)}}{2}}\in\left[\frac{m-k-1+\epsilon(m+k)}{\sqrt{(m+k)(1-\epsilon^2)}},\frac{m-k+1+\epsilon(m+k)}{\sqrt{(m+k)(1-\epsilon^2)}}\right)\right]\\
&\approx\left(\frac{2}{1-\epsilon}\right)^m\left(\Phi\left(\frac{m-k+1+\epsilon(m+k)}{\sqrt{(m+k)(1-\epsilon^2)}}\right)-\Phi\left(\frac{m-k-1+\epsilon(m+k)}{\sqrt{(m+k)(1-\epsilon^2)}}\right)\right) \qquad (54)\\
&<\left(\frac{2}{1-\epsilon}\right)^m\cdot\frac{2}{\sqrt{(m+k)(1-\epsilon^2)}}\cdot\\
&\quad\cdot\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\min\left(\frac{(m-k+1+\epsilon(m+k))^2}{(m+k)(1-\epsilon^2)},\frac{(m-k-1+\epsilon(m+k))^2}{(m+k)(1-\epsilon^2)}\right)\right\}.
\end{aligned}$$

The approximate equality (54) becomes asymptotically exact. Therefore, there exists a number $m_\epsilon^1 \in \mathbb{N}$ such that for all $m \geq m_\epsilon^1$ we have

$$\begin{aligned}
&\binom{m+k}{m}\left(\frac{1+\epsilon}{2}\right)^k\leq\exp\left\{m\log\frac{2}{1-\epsilon}+\frac{1}{2}\log\frac{2}{\pi}\right.\\
&\left.-\frac{1}{2}\log(m+k)-\frac{1}{2}\min\left(\frac{(m-k+1+\epsilon(m+k))^2}{(m+k)(1-\epsilon^2)},\frac{(m-k-1+\epsilon(m+k))^2}{(m+k)(1-\epsilon^2)}\right)\right\}.
\end{aligned} \qquad (55)$$

Moreover, (53) shows that there exists a number $m_\epsilon^2 \in \mathbb{N}$ such that for all $m \geq m_\epsilon^2$ we have

$$\frac{I_{m-2}\left(\arccos\mathrm{e}^{-\sqrt{m}}\right)}{I_{m-2}(\pi)}\leq\frac{1+\epsilon}{2}. \qquad (56)$$

Equations (44), (55) and (56) show that for all $m \geq \max(m_\epsilon^1, m_\epsilon^2)$, $n = m+k \geq m$ and $t \geq \sqrt{m}$,

$$\begin{aligned}
\mathbf{P}\left[\log\mathscr{C}(A^{[m,n]})\geq t\right]&\leq\binom{m+k}{m}2m^{\frac{5}{2}}\left(\frac{1+\epsilon}{2}\right)^k\mathrm{e}^{-t}\\
&\leq\exp\left\{m\log\frac{2}{1-\epsilon}+\frac{3}{2}\log 2+\frac{5}{2}\log m-t\right\}.
\end{aligned}$$

Lemma 9 implies that for the same parameters,

$$\mathbf{E}\left[\log\mathscr{C}(A^{[m,n]})\right]\leq\max\left(m\log\frac{2}{1-\epsilon}+\frac{3}{2}\log 2+\frac{5}{2}\log m,\sqrt{m}\right)+1.$$

Since ϵ was arbitrary, the claim follows. □

The asymptotic linearity in m of the bound on $\mathbf{E}[\log\mathscr{C}(A)]$ derived in the above corollary is due to the appearance of a binomial term in (44). This is largely an artifact of our specific analysis, and if the hunch of Remark 2 is true, then the asymptotic behaviour for $n \geq m \gg 1$ is given by

$$\mathbf{E}[\log\mathscr{C}(A)]=\mathcal{O}(m^\gamma) \qquad (57)$$

for any arbitrarily small real exponent $\gamma > 0$. However, it will not be logarithmically small in m, because of the concentration of measure phenomenon. Thus, if the hunch is true, then (57) describes the asymptotic behaviour for arbitrary $n \geq m$ and $m \gg 1$. On the other hand, when $n \geq 5m$, we can actually prove that (57) holds true.

Corollary 9 *Let $\{A^{[m,n]}\}$ be the set of random matrices defined in Corollary 8. Then for any real exponent $\gamma > 0$,*

$$\limsup_{m\to\infty} \left(\sup_{n\geq 5m} \frac{\mathbf{E}\left[\log \mathscr{C}(A^{[m,n]})\right]}{m^\gamma} \right) = 0,$$

that is, $\mathbf{E}[\log \mathscr{C}(A^{[m,n]})]$ grows more slowly than any algebraic function of m when $n \geq 5m$.

Proof. We need to consider (55) again. For $k \geq 4m$ and ϵ small enough we have

$$m \log \frac{2}{1-\epsilon} < \frac{1}{2} \min \left(\frac{(m-k+1+\epsilon(m+k))^2}{(m+k)(1-\epsilon^2)}, \frac{(m-k-1+\epsilon(m+k))^2}{(m+k)(1-\epsilon^2)} \right),$$

and then

$$\binom{m+k}{m} \left(\frac{1+\epsilon}{2}\right)^k \leq \exp\left\{ \frac{1}{2} \log \frac{2}{\pi} \right\} \tag{58}$$

for all $m \geq m_\epsilon^1$. Moreover, (53) shows that there exists a number m_ϵ^3 such that for all $m \geq m_\epsilon^3$

$$\frac{I_{m-2}\left(\arccos \mathrm{e}^{-m^{\frac{\gamma}{2}}}\right)}{I_{m-2}(\pi)} \leq \frac{1+\epsilon}{2}. \tag{59}$$

Equations (44), (56), (58) and (59) together imply that

$$\mathbf{P}\left[\log \mathscr{C}(A^{[m,n]}) \geq t\right] \leq \exp\left\{ \frac{3}{2}\log 2 - \frac{1}{2}\log \pi + \frac{5}{2}\log m - t \right\}$$

for all $n \geq 5m$, $m \geq m_\epsilon^{\frac{\gamma}{2}} = \max(m_\epsilon^1, m_\epsilon^3)$ and $t \geq m^{\frac{\gamma}{2}}$. Finally, applying Lemma 9, we get

$$\mathbf{E}\left[\log \mathscr{C}(A^{[m,n]})\right] \leq 1 + \max\left(\frac{3}{2}\log 2 - \frac{1}{2}\log \pi + \frac{5}{2}\log m, m^{\frac{\gamma}{2}} \right).$$

Dividing by m^γ and taking limits, the result follows. □

12 A Final Remark About the Case $n < m$

The development in the previous sections assumes $n \geq m$. The case $n < m$ has been dealt with in [11], where it is proved that

$$\mathbf{E}[\log \mathscr{C}(A)] \leq \frac{5}{2}\log n + 2.$$

References

[1] I. Adler and N. Megiddo. A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *Journal of the ACM*, 32:871–895, 1985.

[2] S. Agmon. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:382–392, 1954.

[3] K. Anstreicher, F.A. Potra, and Y. Ye. Probabilistic analysis of an infeasible-interior-point algorithm for linear programming. *Mathematics of Operations Research*, 24:176–192, 1999.

[4] P. Billingsley. *Probability and measure*. John Wiley & Sons, 3 edition, 1995.

[5] H.D. Block and S.A. Levin. On the boundedness of an iterative procedure for solving a system of linear inequalities. *Proc. Amer. Math. Soc.*, 26:229–235, 1970.

[6] K.H. Borgwardt. *Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der linearen Optimierung*. PhD thesis, Universitt Kaiserslautern, 1977.

[7] K.H. Borgwardt. The average number of pivot steps required by the simplex–method is polynomial. *Zeitschrift fr Operations Research*, 7:157–177, 1982.

[8] 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.

[9] K.H. Borgwardt. *The Simplex Method – A Probabilistic Analysis*. Springer Verlag, 1987.

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

[11] D. Cheung and F. Cucker. Probabilistic analysis of condition numbers for linear programming. *Journal of Optimization Theory and Applications*, 114:55–67, 2002.

[12] D. Cheung, F. Cucker, and Ye. Y. Linear programming and condition numbers under the real number computation model. In Ph. Ciarlet and F. Cucker, editors, *Handbook of Numerical Analysis*, volume XI, pages 141–207. North-Holland, 2003.

[13] F. Cucker. Approximate zeros and condition numbers. *J. of Complexity*, 15:214–226, 1999.

[14] 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.

[15] F. Cucker and S. Smale. Complexity estimates depending on condition and round-off error. *Journal of the ACM*, 46:113–184, 1999.

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

[17] R.G. Downey and M.R. Fellows. *Parameterized Complexity*. Springer-Verlag, 1999.

[18] R.M. Freund and J.R. Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM Journal on Optimization*, 10:155–176, 1999.

[19] R.M. Freund and J.R. Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Math. Program.*, 86:225–260, 1999.

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

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

[22] 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*, pages 19–24. Springer Verlag, 1998.

[23] 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. of Complexity*, 18:833–910, 2002.

[24] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[25] L.G. Khachijan. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR*, 244:1093–1096, 1979. (In Russian, English translation in *Soviet Math. Dokl.*, 20:191–194, 1979.).

[26] M. Ledoux. *The Concentration of Measure Phenomenon*. Math. Surveys and Monographs, Volume 89, AMS, 2001.

[27] N. Megiddo. Improved asymptotic analysis of the average number of steps performed by the self-dual simplex algorithm. *Math. Program.*, 35:140–172, 1986.

[28] M. Minoux. Probabilistic bounds on one step objective/potential function improvement in Karmarkar's algorithm. *RAIRO Rech. Opér.*, 28:329–355, 1994.

[29] T. Motzkin and I.Y. Schönberg. The relaxation method for linear inequalitites. *Canadian Journal of Mathematics*, 6:393–404, 1954.

[30] J.K. Reid. On the method of conjugate gradients for the solution of large sparse linear equations. In J.K. Reid, editor, *Large Sparse Sets of Linear Equations*, pages 231–254. Academic Press, 1971.

[31] J. Renegar. Is it possible to know a problem instance is ill-posed? *J. of Complexity*, 10:1–56, 1994.

[32] J. Renegar. Some perturbation theory for linear programming. *Math. Program.*, 65:73–91, 1994.

[33] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM Journal on Optimization*, 5:506–524, 1995.

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

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

[36] M. Shub and S. Smale. Complexity of Bézout's theorem I: geometric aspects. *Journal of the Amer. Math. Soc.*, 6:459–501, 1993.

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

[38] S. Smale. Mathematical problems for the next century. *Mathematical Intelligencer*, 20:7–15, 1998.

[39] M.J. Todd. Polynomial expected behavior of a pivoting algorithm for linear complementarity and linear programming problems. *Math. Program.*, 35:173–192, 1986.

[40] M.J. Todd. Probabilistic models for linear programming. *Mathematics of Operations Research*, 16:671–693, 1991. Erratum, 23:767–768.

[41] 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.

[42] L.N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, 1997.

[43] S.A. Vavasis and Y. Ye. Condition numbers for polyhedra with real number data. *Oper. Res. Lett.*, 17:209–214, 1995.

[44] S.A. Vavasis and Y. Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Math. Program.*, 74:79–120, 1996.

[45] J.R. Vera. On the complexity of linear programming under finite precision arithmetic. *Math. Program.*, 80:91–123, 1998.

Appendix A: A Concentration of Measure Inequality

Let $X \sim \mathscr{U}(\mathrm{S}^{m-1})$ and $p \in \mathrm{S}^{m-1}$. Note that for $\rho < \pi/2$,

$$\begin{aligned}\mathbf{P}\left[X \in \mathtt{cap}(p,\rho)\right] &= \frac{I_{m-2}(\rho)}{I_{m-2}(\pi)} \\ &= \frac{1}{2}\frac{\int_0^\rho \sin^{m-2}\tau d\tau}{\int_0^\rho \sin^{m-2}\tau d\tau + \int_\rho^{\frac{\pi}{2}} \sin^{m-2}\tau d\tau} \\ &= \frac{1}{2}\frac{\int_0^\rho \sin^{m-2}\tau d\tau}{\int_\rho^{\frac{\pi}{2}} \sin^{m-2}\tau d\tau}\left(1 + \frac{\int_0^\rho \sin^{m-2}\tau d\tau}{\int_\rho^{\frac{\pi}{2}} \sin^{m-2}\tau d\tau}\right)^{-1} \\ &= \frac{1}{2}\left(\xi - \xi^2 + \xi^3 - \dots\right) = \frac{1}{2}\xi + \mathcal{O}(\xi^2),\end{aligned} \tag{60}$$

where the last line holds if

$$\xi := \frac{\int_0^\rho \sin^{m-2}\tau d\tau}{\int_\rho^{\frac{\pi}{2}} \sin^{m-2}\tau d\tau} < 1.$$

Now, for $\rho \ll \pi/2$ we have $\xi = \mathcal{O}(\rho^{m-1})$, and hence, $\mathbf{P}[X \in \mathtt{cap}(p,\rho)] = \mathcal{O}(\rho^{m-1})$ as one would expect. Likewise, one expects intuitively that if

$$\frac{\frac{\pi}{2} - \rho}{\frac{\pi}{2}} \ll 1 \tag{61}$$

then

$$\mathbf{P}\left[X \in \mathtt{cap}(p,\rho)\right] = \frac{1}{2} - \mathcal{O}\left(\frac{\frac{\pi}{2} - \rho}{\frac{\pi}{2}}\right), \tag{62}$$

and this is indeed the case. However, it is somewhat surprising that for large m, the expression

$$\frac{\frac{\pi}{2} - \rho}{\frac{\pi}{2}} \tag{63}$$

has to be extremely small indeed before the order (62) is observed. In fact, if (63) decreases to zero at an algebraic rate as a function of the dimension m, then $\mathbf{P}\left[X \in \mathtt{cap}(p,\rho)\right]$ converges to zero. We are not going to prove this property here, although an elementary proof can be given along the lines of Lemma 10 below, but we remark that this is a special case of a type of properties of high-dimensional probability distributions that are jointly referred to as the *concentration of measure phenomenon*. See e.g. [26] for a good account of this theory. The purpose of this appendix is in some sense to get around the adverse effects of the concentration of measure phenomenon and to show that if the expression (63) is exponentially small in terms of m then (62) is asymptotically observed. In fact, we are going to prove a slightly weaker result which is sufficient for the purposes of the analysis of Section 11.

Lemma 10 *Let $\gamma > 0$ be a constant, $p \in \mathrm{S}^{m-1}$ and $X \sim \mathscr{U}(\mathrm{S}^{m-1})$. Then*

$$\lim_{m\to\infty} \mathbf{P}\left[X \in \mathtt{cap}\left(p, \arccos \mathrm{e}^{-m^\gamma}\right)\right] = \frac{1}{2}.$$

Proof. Let $\theta \in (0, 1/2)$ and let $m_\theta \in \mathbb{N}$ be such that

$$\theta < \frac{1 - \frac{1}{m_\theta - 3}}{2}.$$

Then, for $m \geq m_\theta$ equation (60) implies that

$$\mathbf{P}[X \in \mathtt{cap}(p,\rho)] < \theta \Leftrightarrow \int_0^\rho \sin^{m-2}\tau d\tau < 2\theta \int_0^{\frac{\pi}{2}} \sin^{m-2}\tau d\tau. \tag{64}$$

It is easy to show by induction and partial integration that

$$\int_\rho^{\frac{\pi}{2}} \sin^m \tau d\tau = \begin{cases} \frac{1}{m}(\cos\rho)\left[\sin^{m-1}\rho + \sum_{k=0}^{\frac{m-3}{2}} \frac{(m-1)(m-3)\ldots(m-2k-1)}{(m-2)(m-4)\ldots(m-2k-2)} \sin^{m-2k-3}\rho\right] \\ \qquad \text{if } m \text{ is odd,} \\ \frac{1}{m}(\cos\rho)\left[\sin^{m-1}\rho + \sum_{k=0}^{\frac{m}{2}-2} \frac{(m-1)(m-3)\ldots(m-2k-1)}{(m-2)(m-4)\ldots(m-2k-2)} \sin^{m-2k-3}\rho\right] \\ \quad + \frac{(m-1)(m-3)\ldots3\cdot1}{m(m-2)\ldots2}(\frac{\pi}{2}-\rho) \text{ if } m \text{ is even.} \end{cases} \tag{65}$$

In particular,

$$\int_0^{\frac{\pi}{2}} \sin^m \tau d\tau = \begin{cases} \frac{(m-1)(m-3)\ldots4\cdot2}{m(m-2)\ldots3\cdot1} & \text{if } m \text{ is odd,} \\ \frac{(m-1)(m-3)\ldots3\cdot1}{m(m-2)\ldots2}\cdot\frac{\pi}{2} & \text{if } m \text{ is even.} \end{cases} \tag{66}$$

It follows from (65) and (66) that

$$\begin{aligned} \int_\rho^{\frac{\pi}{2}} \sin^m \tau d\tau &< \begin{cases} \frac{(m-1)(m-3)\ldots2}{m(m-2)\ldots1}\cdot\frac{1-sin^{m+1}\rho}{\cos\rho} = \frac{1-\sin^{m+1}\rho}{\cos\rho}\int_0^{\frac{\pi}{2}} \sin^m \tau d\tau \\ \qquad \text{if } m \text{ is odd,} \\ \frac{(m-1)(m-3)\ldots3}{m(m-2)\ldots2}\left[(\tan\rho)(1-\sin^m\rho)+\frac{\pi}{2}-\rho\right] \\ \quad = \left[\frac{2\sin\rho(1-\sin^m\rho)}{\pi\cos\rho}+1-\frac{2\rho}{\pi}\right]\cdot\int_0^{\frac{\pi}{2}}\sin^m\tau d\tau \\ \qquad \text{if } m \text{ is even.} \end{cases} \\ &< \frac{(1-\sin^{m+1}\rho)+2(1-\sin^2\rho)}{\cos\rho}\cdot\left(\int_0^{\frac{\pi}{2}}\sin^m\tau d\tau\right) \qquad \text{(in both cases),} \end{aligned}$$

where the last inequality holds at least for $\pi/4 \le \rho < \pi/2$. Therefore,

$$\begin{aligned} \int_0^\rho \sin^m \tau d\tau &= \int_0^{\frac{\pi}{2}} \sin^m \tau d\tau - \int_\rho^{\frac{\pi}{2}} \sin^m \tau d\tau \\ &> \left[1 - \frac{(1-\sin^{m+1}\rho)+2(1-\sin^2\rho)}{\cos\rho}\right]\cdot\left(\int_0^{\frac{\pi}{2}}\sin^m\tau d\tau\right). \end{aligned} \tag{67}$$

Note that the combination of (64) and (67) implies that for $\pi/4 \le \rho < \pi/2$,

$$2\theta < 1 - \frac{(1-\sin^{m+1}\rho)+2(1-\sin^2\rho)}{\cos\rho} \Rightarrow \mathbf{P}\left[X \in \mathtt{cap}(p,\rho)\right] \ge \theta. \tag{68}$$

Now let the sequence $(\rho_m)_{\mathbb{N}}$ be defined by

$$\rho_m = \arccos \mathrm{e}^{-m^\gamma},$$

and note that

$$1 - \frac{(1-\sin^{m+1}\rho_m)+2(1-\sin^2\rho_m)}{\cos\rho_m} = 1 - \mathrm{e}^{m^\gamma(1-2(m+1))} - 2\mathrm{e}^{-m^\gamma} \overset{m\to\infty}{\longrightarrow} 1.$$

Therefore, for $m \ge m_\theta$ large enough, $\rho_m \in (\pi/4, \pi/2)$ and the condition on the left hand side of (68) is satisfied. This shows that

$$\lim_{m\to\infty} \mathbf{P}\left[X \in \mathtt{cap}\left(p, \arccos \mathrm{e}^{-m^\gamma}\right)\right] \ge \theta,$$

and since this is true for any $\theta \in (0, 1/2)$, this proves that $\lim_{m\to\infty} \mathbf{P}\left[X \in \mathtt{cap}\left(p, \arccos \mathrm{e}^{-m^\gamma}\right)\right] \ge \frac{1}{2}$. Moreover, the inequality $\lim_{m\to\infty} \mathbf{P}\left[X \in \mathtt{cap}\left(p, \arccos \mathrm{e}^{-m^\gamma}\right)\right] \le \frac{1}{2}$ is trivial, and the result follows. □

13 Appendix B: Complexity of the Relaxation Method

The purpose of this appendix is to make a convincing argument that the complexity of relaxation methods for the solution of the linear system $Ax \leq 0,\ x \neq 0$ is proportional to $\mathscr{C}(A)^2$. In a sense, this fact is in the general knowledge of researchers familiar with both the relaxation method and condition numbers, as conversations with Marina Epelman, Rob Freund and Dan Spielman confirmed. Moreover, this fact has been implicitly stated in the relaxation method literature for decades, and all it takes to make it explicit is to translate well-known results from the relaxation method literature into the language of condition numbers. For lack of an explicit reference, let us give such an example here.

The algorithm we consider is the so-called perceptron algorithm [35]. When applied to solving a strictly feasible system $Ax < 0$, where $A \in \mathbb{R}^{n\times m}$ with unit row vectors, this algorithm starts from an initial point $x_0 \in \mathbb{R}^m$ and constructs an iterative sequence of points $(x_i)_{\mathbb{N}} \subset \mathbb{R}^m$ as follows: if $Ax_i < 0$ then x_i is a solution and the algorithm stops. Otherwise, a row vector $a^{[i]}$ from A is chosen so that $a^{[i]}x_i \geq 0$, and the next point is computed by $x_{i+1} = x_i - a_i$.

The usual convergence analysis then proceeds as follows, see e.g. [5]: let $w \in \mathbb{R}^m$ be a solution of $Ax < 0$, let $\alpha = \max_{1\leq j\leq n} a_j w$, where a_j is the j-th row vector of A, and let $w^* = w/|\alpha|$. It is then easy to show that if the algorithm has not stopped before or during iteration i, that is if $x_{i+1} \neq x_i$, then

$$\|x_{i+1} - w^*\|^2 \leq \|x_i - w^*\|^2 - 1.$$

This shows that at most $\|x_0 - w^*\|^2$ iterations can take place. For simplicity, we can choose x_0 to be the origin, so that the algorithm has complexity $\mathcal{O}(\|w^*\|^2)$.

Now note that $\alpha = \cos\theta(A, w)$, where we use the notation of Section 3. Without loss of generality we may assume that w is a unit vector, so that $\|w^*\| = |\cos\theta(A, w)|^{-1}$. In order to minimise the complexity estimate, we need to choose $w = \overline{x}$, so that we find that the algorithm terminates after at most $\mathscr{C}(A)^2$ iterations.