Average-case complexity without the black swans

Dennis Amelunxen   Martin Lotz

October 3, 2018

Abstract

We introduce the concept of weak average-case analysis as an attempt to achieve theoretical complexity results that are closer to practical experience than those resulting from traditional approaches. This concept is accepted in other areas such as non-asymptotic random matrix theory and compressive sensing, and has a particularly convincing interpretation in the most common situation encountered for condition numbers, where it amounts to replacing a null set of ill-posed inputs by a "numerical null set". We illustrate the usefulness of these notions by considering three settings: (1) condition numbers that are inversely proportional to a distance of a homogeneous algebraic set of ill-posed inputs; (2) the running time of power iteration for computing a leading eigenvector of a Hermitian matrix; (3) Renegar's condition number for conic optimisation.

1   Introduction

Depending on context and tradition, a computational problem can mean something practical that begs to be solved as efficiently as possible, or a mathematical object in its own right, to be analysed, classified, and understood. In the first sense, the aim is to develop methods that work well on problems of interest, while in the second, complexity-theoretic sense, algorithms are merely devices used to show that a problem can be solved within certain resource constraints, e.g., in a certain complexity class or with a running time bounded by some function of the input size. Needless to say, complexity-theoretic results are often only weakly correlated with practical experience; a typical example is the simplex method. This is particularly true for numerical problems, where often a condition number serves as a proxy to computational complexity. In this note we aim at shortening the gap between complexity results and practical experience in a theoretically sound way.

In situations where worst-case analysis is meaningless or overly pessimistic, an established practise in complexity theory is to endow the space of inputs with a probability measure and then analyze random variables of interest, like the condition number, induced by this measure. A major point of discussion is the explanatory power of the random model, which needs to satisfy some assumptions to be within reach of a theoretical analysis, in comparison with "typical problems" that are encountered in practice. In this article we do not address the issue of the accuracy of the chosen probability model for the input, but rather a different point of discussion which has received little attention so far. Because even in the most extreme (and unlikely) case that the chosen probability model is 100% accurate, one may *experience* a behavior that is not predicted by the traditional method of analysis. This discrepancy is due to "black swans"; inputs that dominate the theoretical analysis, but which are at the same time extremely rare,

1

so that they practically never show up. For a very concrete example, the expected condition number of a random quadratic Gaussian matrix is infinite, yet most matrices are well-conditioned. On a side note, this discrepancy might also be the reason for a claim by Goldstine and von Neumann [41, p.14] (also pointed out by Edelman and Rao [21]) that "for a random matrix of order n the expectation value of l [the condition number] has been shown to be about n."

In our approach we allow to discard a small subset of the input space. This more liberal attitude towards accounting for all inputs reflects modern practice better, as seen in convex relaxation methods such as compressive sensing, where it is entirely acceptable that an algorithm may even fail on exponentially small sets. From a numerical point of view, we are not able to distinguish between null sets and very small sets due to round-off errors, so the latter may be considered as "numerical null sets". As we will see, disregarding such a practically invisible sets in the analysis can lead to dramatically improved bounds.

Definition 1.1. For $k \in \mathbb{N}$ let (M_k, μ_k) be a probability space and let $T_k \colon M_k \to \mathbb{R}$ be a μ_k-measurable function. We say that the family $\{T_k\}$ has a *weak expectation* of $O(f(k))$ if there exists a family of sets of *exceptional inputs*, $E_k \subseteq M_k$, such that $\mu_k(E_k) = e^{-\Omega(k)}$ and the conditional expectation, conditioned on the nonexceptional inputs, $\mathbb{E}[T_k(\boldsymbol{x}) \mid \boldsymbol{x} \notin E_k]$ is bounded by $f(k)$.

Accordingly, we will speak of *weak average-case complexity* or *weak smoothed complexity* of algorithms or condition numbers. We use the term 'weak' because every traditional complexity analysis without exceptional inputs is in particular a corresponding weak analysis; when it comes to the informative value of the analysis one may in fact argue that the weak analysis is "stronger" than the traditional one. Of course, as with every other use of the O-calculus, one has to be careful about the involved constants, but it seems natural, if not unavoidable, to make the above definition independent of these constants. We illustrate this analysis on three types of problems:

1. Condition numbers inversely proportional to a distance to a homogeneous algebraic set of ill-posed inputs;

2. The running time of power iteration for computing a leading eigenvector of a Hermitian matrix;

3. Renegar's condition number for conic optimisation.

In all three cases the considered expectations over the whole input space are infinite; taking out exponentially small subsets instead yield polynomial conditional expectations, and in the third case even a constant conditional expectation.

1.1 Conic condition numbers

Many condition numbers in numerical analysis come in form of an inverse normalized distance to a set of ill-posed inputs,

$$\mathcal{C}(\boldsymbol{x}) \approx \frac{\|\boldsymbol{x}\|}{\mathrm{dist}(\boldsymbol{x}, \Sigma)},$$

where Σ is invariant under nonnegative scaling. The prototypical example is the matrix condition number, where Σ is the set of singular matrices, but also condition numbers

for eigenvalue computation and for systems of polynomial equations fall into this framework; see [10] for some examples and [9] for a general discussion. Assuming such a geometric characterisation, then in the case where Σ is well-approximated by a set of real codimension 1, the correct asymptotic order of the tail bounds for large values of t is

$$\text{Prob}\{\mathcal{C}(\boldsymbol{x}) \geq t\} \approx t^{-1} \operatorname{vol} \Sigma,$$

where $\operatorname{vol} \Sigma$ is some natural measure of a projective version of Σ. A direct consequence is

$$\mathbb{E}[\mathcal{C}(\boldsymbol{x})] = \infty.$$

This certainly does not reflect the experience that the condition for random inputs is usually seen as unproblematic. An explanation for this codimension conundrum may be found in the simple illustration in Figure 1: although Σ does have codimension one in both pictures, it influences in the second case only a small part of the input space in this way; the predominant part of inputs perceives Σ as "smaller".

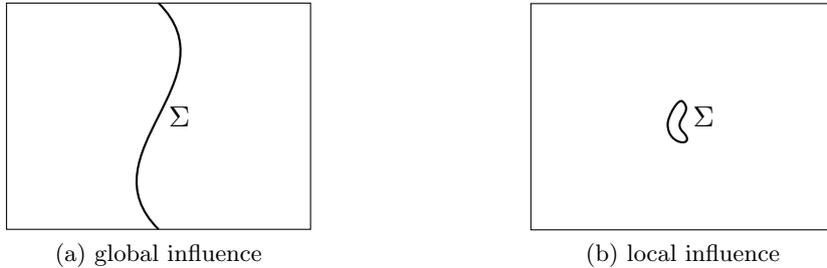(a) global influence　　　　　　　(b) local influence

Figure 1: Illustration of the possible influences of ill-posed inputs.

The result of such a situation is that the small part around the set of ill-posed inputs *dominates the complexity analysis*. And this part of local points around Σ may in fact be *exponentially small*.

Remark 1.2. If one is only interested in the log of the condition, then this is not a big issue, and one may derive again general bounds for the expected condition of a slightly perturbed input, see [10] and [9, Ch. 22].

We adopt the setup of [10]: let $\Sigma \subseteq \mathbb{R}^{n+1}$ be invariant under positive scaling and let the condition number $\mathcal{C} \colon S^n \to \mathbb{R}$ be *defined* by $\mathcal{C}(\boldsymbol{x}) = \|\boldsymbol{x}\|/\operatorname{dist}(\boldsymbol{x}, \Sigma)$, where $\operatorname{dist}(\boldsymbol{x}, \Sigma) = \inf_{\boldsymbol{y} \in \Sigma}\{\|\boldsymbol{x} - \boldsymbol{y}\|\}$. Such a condition number is called a *conic condition number*. For the smoothed analysis setting we denote for a point $\boldsymbol{z} \in S^n$ and $\sigma > 0$ the ball, i.e., the spherical cap, in S^n of radius σ around $\boldsymbol{z}$ by $B(\boldsymbol{z}, \sigma)$.

Theorem 1.3. *Let $\Sigma \cap S^n \subseteq W$, where $W \neq S^n$ is the zero set in S^n of homogeneous polynomials of degree at most $d \geq 1$ and let $\sigma \in (0, 1]$. If $\operatorname{codim} \Sigma = 1$ and $\boldsymbol{z} \in \Sigma$, then for $\boldsymbol{x} \in B(\boldsymbol{z}, \sigma)$ uniformly at random, $\mathbb{E}[\mathcal{C}(\boldsymbol{x})] = \infty$. On the other hand, regardless of the codimension, for all $\boldsymbol{z} \in S^n$ there exists a set $E_{\boldsymbol{z}} \subseteq S^n$ such that for $\boldsymbol{x} \in B(\boldsymbol{z}, \sigma)$ uniformly at random,*

$$\text{Prob}\{\boldsymbol{x} \in E_{\boldsymbol{z}}\} < e^{-n}, \qquad \mathbb{E}\big[\mathcal{C}(\boldsymbol{x}) \mid \boldsymbol{x} \notin E_{\boldsymbol{z}}\big] < \frac{13dn(n+1)}{(1 - e^{-n})\sigma} = O\Big(\frac{dn^2}{\sigma}\Big).$$

We note that the bounds in Theorem 1.3 are still rather coarse when applied in different settings, the reason being that they are quite general. For example, for a version $\kappa_F(\boldsymbol{A})$ of the matrix condition number (using the Frobenius norm) we get from the above,

$$\mathbb{E}[\kappa_F(\boldsymbol{A})|\boldsymbol{A} \notin E] = O\big(n^4/\sigma\big).$$

Once the derivation is understood, it is straight-forward to apply the weak average-case analysis to more precise, problem specific, bounds, such as the smoothed analysis of the matrix condition number by Wschebor [43].

1.2 Power iteration

Power iteration is a classic method for computing a dominant eigenvector and eigenvalue of a matrix. Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ be a Hermitian matrix (this also works for other matrices, but for simplicity we restrict to this case) with eigenvalues λ_i, ordered according to their absolute values, which are the singular values of the matrix,

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ be eigenvectors corresponding to this ordering of eigenvalues. The eigenvalue λ_1 is called a dominant eigenvalue, while $\boldsymbol{u}_1$ is called a dominant eigenvector. The power iteration generates a sequence of unit vectors $\boldsymbol{p}_k$, $k \geq 0$, by setting

$$\boldsymbol{p}_k = \frac{\boldsymbol{A}\boldsymbol{p}_{k-1}}{\|\boldsymbol{A}\boldsymbol{p}_{k-1}\|} = \frac{\boldsymbol{A}^k \boldsymbol{p}_0}{\|\boldsymbol{A}^k \boldsymbol{p}_0\|}.$$

For a generic starting point, the algorithm converges geometrically with ratio $|\lambda_1|/|\lambda_2|$, see [42, 9.2] or [25, 7.3] for an analysis. As there are degenerate cases in which the algorithm does not converge at all, or very slowly, it is natural to ask about the average-case complexity of power iteration, where the average is taken over both the starting points and the space of input matrices. Such an analysis was carried out by Kostlan [28, 29]. In [28], he analyzes the power method for matrices from the Gaussian orthogonal and unitary ensembles: the expected number of steps is infinite (Thm. 4.3); but in Thm. 4.4 he also proves a result akin to a weak average-case analysis, which he calls 'generalized average' (a notion he attributes to Smale), where he takes out a subset of the input space of measure η. Unfortunately, setting this measure exponentially small, $\eta = e^{-\Omega(n)}$, as required for a weak average-case analysis, only yields an exponential bound on the number of iterations. In [29], Kostlan discusses an iterated squaring method that understandably reduces the complexity of the method, but of course this is impractical for applications. The purpose of our work is to improve Kostlan's analysis by showing that power iteration has polynomial weak average running time.

To quantify the convergence, we use the Fubini-Study metric (a.k.a. angle)

$$d_R(\boldsymbol{x}, \boldsymbol{y}) = \arccos \left(\frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \right),$$

where $\|\cdot\|$ is the norm induced by the Hermitian inner product on $\mathbb{C}^n$. For $0 < \alpha \leq \pi/2$ and $\boldsymbol{x} \in \mathbb{C}^n$, let $\rho_\alpha(\boldsymbol{A}, \boldsymbol{x})$ be the minimum number of iterations of the power method with $\boldsymbol{p}_0 = \boldsymbol{x}$ that bring $\boldsymbol{p}_k$ into an α-neighbourhood of $\boldsymbol{u}_1$,

$$\rho_\alpha(\boldsymbol{A}, \boldsymbol{x}) = \min\{k \mid d_R(\boldsymbol{p}_k, \boldsymbol{u}_1) \leq \alpha\}.$$

Define the expected value

$$\rho_\alpha(\boldsymbol{A}) = \underset{\boldsymbol{x} \sim \mathrm{U}(\mathbb{CP}^{n-1})}{\mathbb{E}} [\rho_\alpha(\boldsymbol{A}, \boldsymbol{x})],$$

where $\mathrm{U}(\mathbb{CP}^{n-1})$ denotes the uniform distribution on complex projective space (the notion $\rho_\alpha(\boldsymbol{A}, \boldsymbol{x})$ is well-defined on $\mathbb{CP}^{n-1}$). In what follows, we assume as random model the Gaussian Unitary Ensemble (GUE). A GUE matrix is defined by $\boldsymbol{H} = \frac{1}{2}(\boldsymbol{G} + \boldsymbol{G}^*)$, where $\boldsymbol{G}$ is a complex Gaussian matrix, whose entries have standard Gaussian real and imaginary parts. In particular, the off-diagonal entries have real and imaginary part with mean 0 and variance $1/2$, while the diagonal entries are real with mean 0 and variance 1.

Theorem 1.4. *Let $\alpha \in (0, \pi/4)$, $n \geq 1$, and $\boldsymbol{H} \in \mathbb{C}^{n \times n}$ a GUE matrix. Then there exists a set $E_n \subset \mathbb{C}^{n \times n}$ of exceptional inputs, such that*

$$\mathrm{Prob}\{\boldsymbol{H} \in E_n\} \leq e^{-n},$$

and

$$\mathbb{E}[\rho_\alpha(\boldsymbol{H}) \mid \boldsymbol{H} \notin E_n] < \frac{p(n)}{1 - e^{-cn}} \log \cot \alpha, \quad p(n) = O(n^2)$$

for some constant $c > 0$. In particular, power iteration runs in weak polynomial time.

The result rests on the analysis of the quotient of the largest two singular values of a GUE matrix. It would be interesting to see to what extent this result generalizes to other random models. The results are likely to carry over to Wigner matrices with sub-Gaussian entries. In this setting, the task would amount to finding good small ball probability bounds on the ratio or sum of extreme eigenvalues. A more challenging, though also more practically relevant, problem would be to study the ratio of singular values for a stochastic model of sparse and structured matrices, such as those that might arise in the discretisation of partial differential equations with random coefficients.

1.3 Renegar's condition number and biconic feasibility

The primal and dual (homogeneous) feasibility problems with reference cone $C \subseteq \mathbb{R}^m$ are the decision problems

$$\exists \boldsymbol{x} \in C \setminus \{\boldsymbol{0}\} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \quad \text{(P)} \qquad \exists \boldsymbol{y} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\} \quad \text{s.t.} \quad -\boldsymbol{A}^T\boldsymbol{y} \in C^\circ, \quad \text{(D)}$$

where $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ and $C^\circ = \{\boldsymbol{z} \in \mathbb{R}^m \mid \langle \boldsymbol{x}, \boldsymbol{z} \rangle \leq 0 \text{ for all } \boldsymbol{x} \in C\}$ denotes the polar cone of C (the problem is to determine whether an input $\boldsymbol{A}$ satisfies (P) or (D), which is almost surely, that is, for generic $\boldsymbol{A}$, a strict alternative). Special cases of interest in conic optimisation are when C is the non-negative orthant (LP), the second-order cone (QP), or the cone of positive semidefinite matrices (SDP). Other cases of interest, that include compressive sensing, are when C is the descent cone of a convex regularizer, in which case (the negation of) (P) is referred to as a *nullspace condition*.

It is obvious that the complexity of numerically solving the conic feasiblity problem does not only depend on the representation of $\boldsymbol{A}$, but also on the intrinsic geometry of the problem. To quantify this, Renegar introduced a notion of condition, $\mathcal{R}_C(\boldsymbol{A})$, that depends on both the geometry and the representation of $\boldsymbol{A}$, and analysed in the context of linear programming the complexity of interior point methods for solving the feasibility problem [32, 33]. For the general conic feasibility problem, it was shown in [39] (see

also [9, 9.4]) that the number of iterations of a barrier method for solving the feasibility problem is bounded by $O(\sqrt{\nu_C} \log(\nu_C \cdot \mathcal{R}_C(\boldsymbol{A})))$, where ν_C is the barrier parameter of C. Renegar's condition number $\mathcal{R}_C(\boldsymbol{A})$ for descent cones of convex regularizers also appears uncredited in the error analysis of robust convex regularisation approaches to solving linear inverse problems [11, 23].

The condition number $\mathcal{R}_C(\boldsymbol{A})$ can become arbitrary large as $\boldsymbol{A}$ approaches the boundary between (P) and (D). A probabilistic analysis of $\mathcal{R}_C(\boldsymbol{A})$ and related quantities has been carried out in various settings; a cone-independent result from [1] gives a bound of order $O(\log(n))$ for the expected logarithm of the related Grassmann condition number of a Gaussian matrix.

The theoretical bounds are somewhat out of sync with the observed performance. In fact, for large m and n and random $\boldsymbol{A}$, it turns out that one of (P) or (D) will hold with overwhelming probability, while the other will hold with negligible probability, rendering the feasibility problem almost trivial; see [3] and the references therein for a discussion of the associated phase transition phenomenon. One setting where this phenomenon has been embraced is compressive sensing [23] and its generalisations. In this framework, the dual feasibility (D) is equivalent to a convex relaxation being successful. From a more practical point of view, popular semidefinite programming solvers such as SDPT3 and Mosek rarely take more than 30 iterations on benchmark problems of various sizes [24, 16]; see also [34] for some experiments in the context of compressive sensing.

For symmetry reasons and to emphasise parallels to the matrix condition case, we consider in our analysis the following more general *biconic* convex feasibility problem with two nonzero closed convex cones $C \subseteq \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$:

$$\exists \boldsymbol{x} \in C \setminus \{\boldsymbol{0}\} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \in D^\circ, \quad \text{(P')} \qquad \exists \boldsymbol{y} \in D \setminus \{\boldsymbol{0}\} \quad \text{s.t.} \quad -\boldsymbol{A}^T\boldsymbol{y} \in C^\circ. \quad \text{(D')}$$

(A generalisation of) Renegar's condition number is then defined by

$$\mathcal{R}_{C,D}(\boldsymbol{A}) = \min \left\{ \frac{\|\boldsymbol{A}\|}{\sigma_{C \to D}(\boldsymbol{A})}, \frac{\|\boldsymbol{A}\|}{\sigma_{D \to C}(-\boldsymbol{A}^T)} \right\}, \tag{1}$$

where the smallest restricted singular value is defined as

$$\sigma_{C \to D}(\boldsymbol{A}) = \min_{\boldsymbol{x} \in C \cap S^{m-1}} \|\Pi_D(\boldsymbol{A}\boldsymbol{x})\|, \qquad \Pi_D(\boldsymbol{y}) = \min\{\|\boldsymbol{y} - \boldsymbol{y}'\| \mid \boldsymbol{y}' \in D\}.$$

Note that in the case $C = \mathbb{R}^m$, $D = \mathbb{R}^n$ we recover the classical matrix condition number

$$\mathcal{R}_{\mathbb{R}^m, \mathbb{R}^n}(\boldsymbol{A}) = \kappa(\boldsymbol{A}) = \|\boldsymbol{A}\| \, \|\boldsymbol{A}^\dagger\|,$$

where $\boldsymbol{A}^\dagger$ is the Moore-Penrose pseudoinverse of $\boldsymbol{A}$. In this unrestricted case a lot is known about the distribution of the condition number for Gaussian matrices but also for more general ensembles [12, 35, 8, 36]; see [9, Notes to Ch. 4] for a concise discussion of this case. We only mention here that for Gaussian matrices it is known that if m_k, n_k are such that $\lim_{k \to \infty} m_k/n_k = \gamma^2$, $0 < \gamma < 1$, then

$$\mathcal{R}_{\mathbb{R}^m, \mathbb{R}^n}(\boldsymbol{G}) = \kappa(\boldsymbol{G}) \to \frac{1+\gamma}{1-\gamma} \quad \text{almost surely.}$$

In particular, in this asymptotic setting the expected condition number is of *constant* order.

If C and D both have nonempty interior then the expectation of Renegar's condition number of Gaussian matrices $\mathcal{R}_{C,D}(\boldsymbol{G})$ is infinity. But if we consider families of cones

such that their "formats" satisfy similar asymptotics as above, then we will show that Renegar's condition number does indeed have constant weak average-case complexity. We will describe next what we mean by similar format, after introducing a notion of dimension of a closed convex cone.

The statistical dimension of a closed convex cone $C \subseteq \mathbb{R}^m$ may be characterized as

$$\delta(C) = \mathop{\mathbb{E}}_{\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_m)} \big[\|\Pi_C(\boldsymbol{g})\|^2 \big].$$

It coincides with the usual dimension if C is the full space, $\delta(\mathbb{R}^m) = m$. We now consider families of cones $C_k \subseteq \mathbb{R}^{m_k}$, $D_k \subseteq \mathbb{R}^{n_k}$, and we assume that we have the asymptotics

$$\lim_{k \to \infty} \frac{\delta(C_k)}{m_k} = \alpha^2, \quad \lim_{k \to \infty} \frac{\delta(D_k)}{n_k} = \beta^2, \quad \lim_{k \to \infty} \frac{m_k}{n_k} = \gamma^2 \tag{2}$$

with $0 < \alpha, \beta, \gamma \leq 1$. A typical example for this situation is $C_k \subseteq \mathbb{R}^{m_k}$ self-dual and $D_k = \mathbb{R}^{n_k}$, in which case $\delta(C_k) = \frac{m_k}{2}$ and $\delta(D_k) = n_k$, so that $\alpha = \frac{1}{\sqrt{2}}$ and $\beta = 1$. We will assume that $\beta \neq \alpha\gamma$, and since $\mathcal{R}_{C,D}(\boldsymbol{A}) = \mathcal{R}_{D,C}(-\boldsymbol{A}^T)$, we may assume without loss of generality

$$\beta > \alpha \gamma. \tag{3}$$

Closely related, but conceptually different, to the statistical dimension is the Gaussian width of a cone, $w(C) = \mathop{\mathbb{E}}_{\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_m)} \big[\|\Pi_C(\boldsymbol{g})\| \big]$ (actually, it is the Gaussian width of the intersection with the unit ball $C \cap B^n$, but we adopt a purely conic point of view here). The squared Gaussian width $w^2(C)$ differs from the statistical dimension by at most one [3, Prop. 10.2], $w^2(C) \leq \delta(C) \leq w^2(C) + 1$. In particular, the asymptotics (2) hold for the squared Gaussian width if and only if they hold for the statistical dimension.

Theorem 1.5. *If $C_k \subseteq \mathbb{R}^{m_k}$, $D_k \subseteq \mathbb{R}^{n_k}$ are families of cones such that the dimensions satisfy the asymptotics (2) with $\beta > \alpha\gamma$, and $n_k \to \infty$ for $k \to \infty$, then there exist exceptional sets E_k with $\mathrm{Prob}\{\boldsymbol{G} \in E_k\} \leq e^{-\Omega(n_k)}$, such that*

$$\mathbb{E}[\mathcal{R}_{C,D}(\boldsymbol{G}) \mid \boldsymbol{G} \notin E_k] < u_k(C, D), \qquad u_k(C_k, D_k) \to \frac{1 + \gamma}{\beta - \alpha\gamma}.$$

In particular, if $n_k = \Omega(k)$, then the weak average-case complexity of $\mathcal{R}_{C,D}$ is constant.

The probabilistic analysis of Renegar's condition number for arbitrary cones has so far been confined to [1] (though there has been considerable work on the linear programming case, see [9] and the references therein). These results were obtained following the classical tail estimate approach. The new approach of allowing to ignore exponentially small input sets, which loosens the requirements for the tail bounds so that for fixed dimensions we do not even require the bounds to go to zero, opens up new ways to directly exploit concentration of measure phenomena. We will do this through the use of Gordon's escape-through-the-mesh approach [26, 17].

1.4   Relation to previous work

Smoothed analysis [37, 7] can also be seen as a way to deal with outliers. In its original Gaussian setting, one interpretation is that smoothed analysis corresponds to a worst-case analysis after Gaussian smoothing of the input space. When perturbing around ill-posed instances, this approach still suffers from the problem that too much weight is

given to practically impossible events. An overview of previous work in the probabilistic analysis of numerical algorithms via condition numbers, that includes smoothed analysis and much more, can be found in [9].

The point of view of weak average-case analysis is not unusual in the setting of non-asymptotic random matrix theory [40] and its applications, such as compressive sensing and related fields [23]. Rather than focussing on precise tail bounds $\text{Prob}\{X > t\}$ that decay to zero for fixed n as $t \to \infty$, and the associated expected values, the emphasis is on showing that a quantity of interest is confined to a certain region with overwhelming probability. In the setting of complexity theory this point of view seems to be new.

We would also like to point out that the idea of removing "bad cases" from a probabilistic analysis has featured in other probabilistic settings [38, 13, 14, 15, 19].

1.5 Organisation of paper

The remaining sections are devoted to the proofs of the three main results, Theorems 1.3, 1.4 and 1.5. Section 2 deals with the easiest case, condition numbers inversely proportional to a set of ill-posed inputs. Here, a weak average-case analysis follows in a straight-forward way from a simple probabilistic observation on the effect of conditioning out large deviations from an expectation. Section 3 deals with the power iteration. For convenience, we recreate Kostlan's derivation of the number of iterations needed to approach a dominant eigenvector in terms of the ratio of the largest singular values. The probabilistic analysis and proof of Theorem 1.4 then follows from known results on random matrices from the Gaussian unitary ensemble. Finally, Section 4 presents some basic facts about the biconic feasibility problem and Gordon's inequalities (discussed in more depth in [2]), followed by the proof of Theorem 1.5

1.6 Acknowledgments

2 Conic condition numbers

We begin with some basic probabilistic observations about the effect of conditioning on the expectation of a quantity such as a condition number. We also include the weak average-case analysis of conic condition numbers, since it is a trivial consequence of Lemma 2.2. The first observation is that for a weak average-case analysis of a random variable X, while we are allowed to remove any small enough set from the probability space, we do best by removing a set of the form $\{X > t\}$. Therefore, weak average-case analysis of X is equivalent to an average-case analysis of a truncation $X_{\leq t} := X \cdot 1\{X \leq t\}$ for a suitable parameter $t > 0$.

Lemma 2.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and X a random variable that is absolutely continuous with respect to μ. Given $\varepsilon > 0$, let t be such that $\text{Prob}\{X > t\} = \varepsilon$. Then for any measurable $S \subset \Omega$ with $\mu(S) \leq \varepsilon$,*

$$\mathbb{E}\big[X \mid X \leq t\big] \leq \mathbb{E}\big[X \mid \overline{S}\big].$$

Proof. The conditional expectations $\mathbb{E}\big[X \mid X \leq t\big]$ are monotonically increasing in t, which is why we can assume without loss of generality that $\mu(S) = \varepsilon$. Define $\overline{S}_{\leq t} =$

$\{\boldsymbol{\omega} \in \overline{S} \mid X(\boldsymbol{\omega}) \leq t\}$ and $\overline{S}_{>t} = \{\boldsymbol{\omega} \in \overline{S} \mid X(\boldsymbol{\omega}) > t\}$. Then we have the disjoint decompositions $\overline{S} = \overline{S}_{\leq t} \cup \overline{S}_{>t}$ and

$$\{\boldsymbol{\omega} \in \Omega \mid X(\boldsymbol{\omega}) \leq t\} = \overline{S}_{\leq t} \cup \{\boldsymbol{\omega} \in S \mid X(\boldsymbol{\omega}) \leq t\}.$$

Since $\mathrm{Prob}\{X \leq t\} = 1 - \varepsilon = \mu(\overline{S})$, we get

$$\mu(\overline{S}_{>t}) = \mu(\{\boldsymbol{\omega} \in S \mid X(\boldsymbol{\omega}) \leq t\}). \tag{4}$$

For the conditional probabilities we obtain

$$
\begin{aligned}
\mathbb{E}\big[X \mid X \leq t\big] &= \frac{1}{1-\varepsilon}\left(\int_{\overline{S}_{\leq t}} X(\boldsymbol{\omega})d\mu + \int_{\{S \mid X(\boldsymbol{\omega}) \leq t\}} X(\omega)d\mu \right) \\
&\leq \frac{1}{1-\varepsilon}\left(\int_{\overline{S}_{\leq t}} X(\boldsymbol{\omega})d\mu + t\mu(\{S \mid X(\boldsymbol{\omega}) \leq t\}) \right) \\
&\stackrel{(4)}{=} \frac{1}{1-\varepsilon}\left(\int_{\overline{S}_{\leq t}} X(\boldsymbol{\omega})d\mu + t\mu(\overline{S}_{>t}) \right) \\
&\leq \frac{1}{1-\varepsilon} \int_{\overline{S}} X(\boldsymbol{\omega})d\mu = \mathbb{E}\big[X \mid \overline{S}\big]. \qquad \square
\end{aligned}
$$

The next lemma shows how to get finite conditional expectations from tail bounds of order t^{-1}. This will be key for the weak average-case analysis of conic condition numbers and the power iteration method.

Lemma 2.2. *Let X be a non-negative random variable on a probability space $(\Omega, \mathcal{F}, \mu)$ such that for some $a > 0$ and for all $t > a$,*

$$\mathrm{Prob}\{X > t\} \leq \frac{a}{t}. \tag{5}$$

Then, for $t > a$,

$$\mathbb{E}\big[X \mid X \leq t\big] \leq \frac{a}{1 - \frac{a}{t}}\left(1 - \log\left(\frac{a}{t}\right)\right).$$

Proof. The proof is a straight-forward calculation:

$$
\begin{aligned}
\mathbb{E}\big[X \mid X \leq t\big] &= \frac{1}{1 - \mathrm{Prob}\{X > t\}} \int_0^t \mathrm{Prob}\{X > s\}\, ds \\
&\leq \frac{1}{1 - \frac{a}{t}}\left(a + \int_a^t \mathrm{Prob}\{X > s\}\, ds\right) \leq \frac{a}{1 - \frac{a}{t}}(1 + \log(t) - \log(a)). \quad \square
\end{aligned}
$$

Let $\mathcal{C}(\boldsymbol{x})$ be a conic condition number as specified in Theorem 1.3. In [10, Thm. 1.1] it is shown that for this kind of condition numbers,

$$\mathrm{Prob}_{\boldsymbol{x} \in B(\boldsymbol{z},\sigma)}\{\mathcal{C}(\boldsymbol{x}) > t\} < \frac{13dn}{t\sigma} \quad \text{if } t \geq (1 + 2d)(n-1)/\sigma. \tag{6}$$

Combining this powerful result with the simple Lemma 2.2 yields Theorem 1.3.

Proof of Theorem 1.3. Set $a := \frac{13dn}{\sigma}$. Since $a = 13dn/\sigma > (1 + 2d)(n-1)/\sigma$, the tail bound (6) holds in particular for $t \geq a$. Setting $E_n = \{\boldsymbol{x} \mid \mathcal{C}(\boldsymbol{x}) > ae^n\}$, we obtain

$$\mathrm{Prob}_{\boldsymbol{x} \in B(\boldsymbol{z},\sigma)}\{\boldsymbol{x} \in E_n\} < e^{-n}.$$

Therefore,

$$\mathbb{E}_{\boldsymbol{x} \in B(\boldsymbol{z},\sigma)}\big[\mathcal{C}(\boldsymbol{x}) \mid \boldsymbol{x} \notin E_n\big] = \mathbb{E}_{\boldsymbol{x} \in B(\boldsymbol{z},\sigma)}\big[\mathcal{C}(\boldsymbol{x}) \mid \mathcal{C}(\boldsymbol{x}) \leq ae^n\big] \stackrel{\text{Lem. 2.2}}{\leq} \frac{a(n+1)}{1 - e^{-n}} = \frac{13dn(n+1)}{(1 - e^{-n})\sigma}.$$

This completes the proof. $\square$

9

3 Power iteration

We first review a bound on the number of iterations needed to get within a certain distance of the dominant eigenvector. We use the notation from Section 1.2. Recall that for $0 < \alpha \leq \pi/2$ and $\boldsymbol{x} \in \mathbb{C}^n$, $\rho_\alpha(\boldsymbol{A}, \boldsymbol{x})$ denotes the minimum number of iterations that bring $\boldsymbol{A}^k \boldsymbol{x}$ into an α-neighbourhood of $\boldsymbol{u}_1$,

$$\rho_\alpha(\boldsymbol{A}, \boldsymbol{x}) = \min\{k \mid d_R(\boldsymbol{p}_k, \boldsymbol{u}_1) \leq \alpha\},$$

and the expected value is defined as

$$\rho_\alpha(\boldsymbol{A}) = \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathrm{U}(\mathbb{CP}^{n-1})} [\rho_\alpha(\boldsymbol{A}, \boldsymbol{x})],$$

where $\mathrm{U}(\mathbb{CP}^{n-1})$ denotes the uniform distribution on complex projective space. For a linear subspace $L \subseteq \mathbb{C}^n$, write $\boldsymbol{\Pi}_L(\boldsymbol{x}) = \operatorname{argmin}_{\boldsymbol{y} \in L} \|\boldsymbol{y} - \boldsymbol{x}\|$ for the projection of $\boldsymbol{x}$ onto L, and let $\boldsymbol{\Pi}_i(\boldsymbol{x}) := \boldsymbol{\Pi}_{\operatorname{span}\{\boldsymbol{u}_i\}}(\boldsymbol{x})$ be the projection onto the linear subspace spanned by the i-th eigenvector. The following result can be found in a slightly modified form in [28]; the proof is a variation of the well-known analysis of the power method in terms of the ratio of largest singular values, see for example [42, 9.2] or [25, 7.3]. For convenience we provide a complete derivation of the starting point dependent results, since the proof in [28] is rather sketchy.

Theorem 3.1 (Kostlan [28]). *Assume $\boldsymbol{x} \in \mathbb{C}^n$ is not orthogonal to the dominant eigenvector $\boldsymbol{u}_1$ of $\boldsymbol{A}$, $|\langle \boldsymbol{x}, \boldsymbol{u}_1 \rangle| < \frac{\pi}{2}$. Then*

$$\rho_\alpha(\boldsymbol{A}, \boldsymbol{x}) \geq \frac{\log \cot(\alpha) + \log \|\boldsymbol{\Pi}_2(\boldsymbol{x})\| - \log \|\boldsymbol{\Pi}_1(\boldsymbol{x})\|}{\log |\lambda_1| - \log |\lambda_2|}$$

and

$$\rho_\alpha(\boldsymbol{A}, \boldsymbol{x}) \leq \frac{\log \cot(\alpha) + \log \|\boldsymbol{\Pi}_{\boldsymbol{u}_1^\perp}(\boldsymbol{x})\| - \log \|\boldsymbol{\Pi}_1(\boldsymbol{x})\|}{\log |\lambda_1| - \log |\lambda_2|}.$$

The expected number of steps is bounded by

$$\frac{\log \cot(\alpha)}{\log |\lambda_1| - \log |\lambda_2|} \leq \rho_\alpha(\boldsymbol{A}) \leq \frac{\frac{1}{2}(\log(n) + 2\frac{n-1}{n}) + \max\{0, \log \cot \alpha\}}{\log |\lambda_1| - \log |\lambda_2|}.$$

Proof. Assume $\|\boldsymbol{x}\| = 1$ and write $\boldsymbol{x} = a_1 \boldsymbol{u}_1 + \cdots + a_n \boldsymbol{u}_n$. Then, in particular, $a_1 = \|\boldsymbol{\Pi}_1(\boldsymbol{x})\|$, $a_2 = \|\boldsymbol{\Pi}_2(\boldsymbol{x})\|$, and $a_2 \boldsymbol{u}_2 + \cdots + a_n \boldsymbol{u}_n = \boldsymbol{\Pi}_{\boldsymbol{u}_1^\perp}(\boldsymbol{x})$. Note that for any $k \geq 1$,

$$\boldsymbol{A}^k \boldsymbol{x} = a_1 \lambda_1^k \boldsymbol{u}_1 + \cdots + a_n \lambda_n^k \boldsymbol{u}_n.$$

Let $k = \rho_\alpha(\boldsymbol{A}, \boldsymbol{x})$ be the smallest integer such that

$$\cos(\alpha) \leq \frac{|\langle \boldsymbol{A}^k \boldsymbol{x}, \boldsymbol{u}_1 \rangle|}{\|\boldsymbol{A}^k \boldsymbol{x}\|} = \frac{|a_1 \lambda_1^k|}{\|\boldsymbol{A}^k \boldsymbol{x}\|} = \frac{\|\boldsymbol{\Pi}_1(\boldsymbol{x})\| |\lambda_1|^k}{\|\boldsymbol{A}^k \boldsymbol{x}\|}. \tag{7}$$

From the identity $\sin^2(\alpha) + \cos^2(\alpha) = 1$ we obtain

$$\sin(\alpha) \geq \frac{\sqrt{a_2^2 |\lambda_2|^{2k} + \cdots + a_n^2 |\lambda_n|^{2k}}}{\|\boldsymbol{A}^k \boldsymbol{x}\|} \geq \frac{|\lambda_2|^k \|\boldsymbol{\Pi}_2(\boldsymbol{x})\|}{\|\boldsymbol{A}^k \boldsymbol{x}\|}. \tag{8}$$

Putting (7) and (8) together, we get

$$\cot(\alpha) \leq \left(\frac{|\lambda_1|}{|\lambda_2|}\right)^k \frac{\|\mathbf{\Pi}_1(\boldsymbol{x})\|}{\|\mathbf{\Pi}_2(\boldsymbol{x})\|} \;\Rightarrow\; k \geq \frac{\log\cot(\alpha) + \log\|\mathbf{\Pi}_2(\boldsymbol{x})\| - \log\|\mathbf{\Pi}_1(\boldsymbol{x})\|}{\log|\lambda_1| - \log|\lambda_2|}.$$

This shows the first inequality. For the second inequality, let $k < \rho_\alpha(\boldsymbol{A}, \boldsymbol{x})$. Then after k iterations we are still outside an α neighbourhood of $\boldsymbol{u}_1$, so that

$$\cos(\alpha) \geq \frac{|\langle \boldsymbol{A}^k\boldsymbol{x}, \boldsymbol{u}_1 \rangle|}{\|\boldsymbol{A}^k\boldsymbol{x}\|} = \frac{|a_1\lambda_1^k|}{\|\boldsymbol{A}^k\boldsymbol{x}\|} = \frac{\|\mathbf{\Pi}_1(\boldsymbol{x})\||\lambda_1|^k}{\|\boldsymbol{A}^k\boldsymbol{x}\|}. \tag{9}$$

Similar as in (8), we get

$$\sin(\alpha) \leq \frac{\sqrt{a_2^2|\lambda_2|^{2k} + \cdots + a_n^2|\lambda_n|^{2k}}}{\|\boldsymbol{A}^k\boldsymbol{x}\|} \leq \frac{|\lambda_2|^k\|\mathbf{\Pi}_{\boldsymbol{u}_1^\perp}(\boldsymbol{x})\|}{\|\boldsymbol{A}^k\boldsymbol{x}\|}. \tag{10}$$

Combining (9) and (10) gives

$$\cot(\alpha) \geq \left(\frac{|\lambda_1|}{|\lambda_2|}\right)^k \frac{\|\mathbf{\Pi}_1(\boldsymbol{x})\|}{\|\mathbf{\Pi}_{\boldsymbol{u}_1^\perp}(\boldsymbol{x})\|} \;\Rightarrow\; k \leq \frac{\log\cot(\alpha) + \log\|\mathbf{\Pi}_{\boldsymbol{u}_1^\perp}(\boldsymbol{x})\| - \log\|\mathbf{\Pi}_1(\boldsymbol{x})\|}{\log|\lambda_1| - \log|\lambda_2|},$$

which establishes the second inequality.

For the expected values of the bounds we refer to Kostlan's work [28]. $\qquad\square$

3.1 Weak average-case analysis of power iteration

Theorem 3.1 implies that the relevant probability in the average-case analysis of power iteration is

$$\mathrm{Prob}\left\{\frac{1}{\log|\lambda_1| - \log|\lambda_2|} > x\right\} = \mathrm{Prob}\left\{\frac{|\lambda_1|}{|\lambda_2|} < e^{\frac{1}{x}}\right\}.$$

Let $\lambda_{\max}$, $\lambda_{2\max}$, $\lambda_{\min}$ and $\lambda_{2\min}$ denote the largest, second largest, smallest and second smallest eigenvalue of a GUE matrix $\boldsymbol{H}$, respectively. Keep in mind that this refers to the eigenvalues themselves, not their absolute values, which underlie the ordering $|\lambda_1| \geq \cdots \geq |\lambda_n|$. Before proceeding, we collect some known results on the distribution of the eigenvalues of a GUE matrix. To ease notation, in what follows, C and c are used for absolute constants which may change from line to line.

1. The probability that all eigenvalues are positive (or all are negative) is exponentially small [18],

$$\mathrm{Prob}\{\lambda_{\min} > 0\} \leq Ce^{-cn^2} \tag{11}$$

   for some constants C and c.

2. The largest eigenvalue of a GUE matrix satisfies

$$\mathrm{Prob}\{\lambda_{\max} \geq \sqrt{n}(2 + \varepsilon)\} \leq Ce^{-cn\varepsilon^{3/2}} \tag{12}$$

   for some constants C and c (e.g., [30, Prop. 2.1], note the slightly different form). We will use this with $\varepsilon = 1$.

3. The smallest gap $\delta_{\min}$ between consecutive eigenvalues satisfies

$$\mathrm{Prob}\{\delta_{\min} \leq \delta/\sqrt{n}\} \leq n\delta^3 + e^{-cn} \tag{13}$$

   for some constant c. This is from [22], as stated in [31, (2)].

11

Proof of Theorem 1.4. Assume $x \geq 1$. Throughout the proof, we will use the fact that $-\lambda_{\min}$ and $\lambda_{\max}$ have the same distribution, as well as $-\lambda_{2\min}$ and $\lambda_{2\max}$. For the relationship between the singular values and the eigenvalues we have the following two cases:

1. The largest and second largest singular values come from the two largest or from the two smallest eigenvalues, that is, $\lambda_1 = \lambda_{\max}$ and $\lambda_2 = \lambda_{2\max}$, or $\lambda_1 = \lambda_{\min}$ and $\lambda_2 = \lambda_{2\min}$.

2. The largest and second largest singular values come from the largest (positive) and smallest (negative) eigenvalues, that is, $\lambda_1 = \lambda_{\max}$ and $\lambda_2 = \lambda_{\min}$, or $\lambda_1 = \lambda_{\min}$ and $\lambda_2 = \lambda_{\max}$.

Using the union bound we may consider these cases separately:

$$\mathrm{Prob}\left\{\frac{|\lambda_1|}{|\lambda_2|} < e^{\frac{1}{x}}\right\} \leq \mathrm{Prob}\left\{\frac{|\lambda_{\max}|}{|\lambda_{2\max}|} < e^{\frac{1}{x}}\right\} + \mathrm{Prob}\left\{\frac{|\lambda_{\min}|}{|\lambda_{2\min}|} < e^{\frac{1}{x}}\right\}$$
$$+ \mathrm{Prob}\left\{e^{-\frac{1}{x}} < \frac{|\lambda_{\max}|}{|\lambda_{\min}|} < e^{\frac{1}{x}}\right\}$$

Clearly, the first two terms on the right-hand side are subject to the same analysis.

Adjacent eigenvalues. We consider the case where the largest and second largest singular values are the largest and second largest (positive) eigenvalues of the matrix. Note that for $x \geq 1$ we have $e^{1/x} - 1 < 2/x$, which will be used in the sequel. We will also generically use C and c for constants that may vary within one derivation. We have for $x \geq 1$,

$$\mathrm{Prob}\left\{\frac{|\lambda_{\max}|}{|\lambda_{2\max}|} < e^{\frac{1}{x}}\right\} \leq \mathrm{Prob}\left\{\frac{\lambda_{\max}}{\lambda_{2\max}} < e^{\frac{1}{x}}\right\} = \mathrm{Prob}\left\{\frac{\lambda_{\max} - \lambda_{2\max}}{\lambda_{2\max}} < e^{1/x} - 1\right\}$$
$$\leq \mathrm{Prob}\left\{\frac{\delta_{\min}}{\lambda_{\max}} < \frac{2}{x}\right\} \overset{(12),\varepsilon=1}{\leq} \mathrm{Prob}\left\{\delta_{\min} < \frac{6n}{x\sqrt{n}}\right\} + Ce^{-cn}$$
$$\overset{(13)}{<} 216\frac{n^4}{x^3} + Ce^{-cn}.$$

For $x > 6n^{4/3}$ and n large enough, the bound becomes non-trivial.

Extreme eigenvalues. For the ratio of the eigenvalues at the edge we have for $x \geq 1$,

$$\mathrm{Prob}\left\{e^{-\frac{1}{x}} < \frac{|\lambda_{\max}|}{|\lambda_{\min}|} < e^{\frac{1}{x}}\right\} \leq \mathrm{Prob}\left\{e^{-\frac{1}{x}} < \frac{\lambda_{\max}}{-\lambda_{\min}} < e^{\frac{1}{x}}\right\} + 2\,\mathrm{Prob}\{\lambda_{\min} > 0\}$$
$$\overset{(11)}{\leq} \mathrm{Prob}\left\{\frac{\lambda_{\max}}{-\lambda_{\min}} < e^{\frac{1}{x}}\right\} - \mathrm{Prob}\left\{\frac{\lambda_{\max}}{-\lambda_{\min}} \leq e^{-\frac{1}{x}}\right\} + Ce^{-cn^2}$$
$$= \mathrm{Prob}\left\{\frac{\lambda_{\max}}{-\lambda_{\min}} < e^{\frac{1}{x}}\right\} - \left(1 - \mathrm{Prob}\left\{\frac{\lambda_{\max}}{-\lambda_{\min}} > e^{-\frac{1}{x}}\right\}\right) + Ce^{-cn^2}$$
$$= 2\,\mathrm{Prob}\left\{\frac{\lambda_{\max}}{-\lambda_{\min}} < e^{\frac{1}{x}}\right\} - 1 + Ce^{-cn^2}$$
$$= 2\,\mathrm{Prob}\left\{\frac{\lambda_{\max} + \lambda_{\min}}{-\lambda_{\min}} < e^{\frac{1}{x}} - 1\right\} - 1 + Ce^{-cn^2}$$
$$\leq 2\,\mathrm{Prob}\left\{\frac{\lambda_{\max} + \lambda_{\min}}{-\lambda_{\min}} < \frac{2}{x}\right\} - 1 + Ce^{-cn^2}$$
$$\overset{(12),\varepsilon=1}{\leq} 2\,\mathrm{Prob}\left\{\lambda_{\max} + \lambda_{\min} < \frac{6\sqrt{n}}{x}\right\} - 1 + Ce^{-cn^2}.$$

12

Since $\lambda_{\max}$ and $-\lambda_{\min}$ are equally distributed, the difference $\lambda_{\max} - (-\lambda_{\min})$ is symmetric around the origin, so that we obtain

$$2\operatorname{Prob}\left\{\lambda_{\max} + \lambda_{\min} < \frac{6\sqrt{n}}{x}\right\} - 1 = \operatorname{Prob}\left\{-\frac{6}{x} < \frac{\lambda_{\max}}{\sqrt{n}} + \frac{\lambda_{\min}}{\sqrt{n}} < \frac{6}{x}\right\}.$$

Now consider the normalized variables

$$\tilde{\lambda}_{\max} := n^{2/3}\left(\frac{\lambda_{\max}}{\sqrt{n}} - 2\right), \qquad \tilde{\lambda}_{\min} := n^{2/3}\left(\frac{\lambda_{\max}}{\sqrt{n}} + 2\right).$$

The maximum M of the density of $\tilde{\lambda}_{\max} + \tilde{\lambda}_{\min}$ is bounded, as can be deduced directly from the joint distribution of the eigenvalues of the GUE ensemble. We thus obtain

$$\operatorname{Prob}\left\{-\frac{6n^{2/3}}{x} < \tilde{\lambda}_{\max} + \tilde{\lambda}_{\min} < \frac{6n^{2/3}}{x}\right\} \le \frac{12Mn^{2/3}}{x}.$$

In summary, we have the bound

$$\operatorname{Prob}\left\{e^{-\frac{1}{x}} < \frac{|\lambda_{\max}|}{|\lambda_{\min}|} \le e^{\frac{1}{x}}\right\} \le \frac{12Mn^{2/3}}{x} + Ce^{-cn}. \tag{14}$$

Combined bound. Putting the bounds together, we get

$$\operatorname{Prob}\left\{\frac{1}{\log|\lambda_1| - \log|\lambda_2|} > x\right\} < \frac{12Mn^{2/3}}{x} + \frac{432n^4}{x^3} + Ce^{-cn}$$

for some constant K. This bound is non-trivial for $x > Kn^{4/3}$ for a suitable constant K. Set $x_0 = Kn^{4/3}e^{cn}$ and

$$E_n := \left\{\boldsymbol{H} \,\Big|\, \frac{1}{\log|\lambda_1| - \log|\lambda_2|} > x_0\right\}.$$

Then

$$\operatorname{Prob}\left\{\boldsymbol{H} \in E_n\right\} < Ce^{-cn}$$

for suitable absolute constants C and c. We can apply a variation of the proof of Lemma 2.2 (with an adjustment to take into account the x^{-3} and the exponential term), with $a = Kn^{4/3}$ and $t = x_0$, to get

$$\mathbb{E}\left[(\log|\lambda_1| - \log|\lambda_2|)^{-1} \,\Big|\, E_n\right] < \frac{Kn^2}{1 - e^{-cn}}$$

for some absolute constant K. Incorporating the numerator from the bound on the power iteration gives the desired result. $\qquad\square$

Remark 3.2. With a more careful analysis, the constants in the bounds can be explicitly estimated, but we haven't done so as the first aim was to establish a polynomial bound.

Remark 3.3. There are potential alternative ways to derive the results from this section. One such approach is based on the joint distribution of the singular values, as studied by Edelman and La Croix [20]. There, it is shown that the singular values of a GUE matrix can be characterized as the union of the singular values of two independent bidiagonal matrices with χ-distributed entries. By an interlacing property it appears that, with high probability, the largest two singular values correspond to the largest singular values of each of these matrices. In a different direction, by examining the analysis in [5, 6] (for asymptotic independence of the extreme eigenvalues) and [27] (for the rate of convergence to Tracy-Widom) might also hope to carry out the above analysis via approximation by the Tracy-Widom distribution.

4 Renegar's condition number

Note that we have a complete symmetry between (P') and (D') via the exchange of $\boldsymbol{A}$ by $-\boldsymbol{A}^T$ and by a swap between C and D. We denote the set of primal feasible instances and the set of dual feasible instances by

$$\mathcal{P}(C, D) := \{\boldsymbol{A} \in \mathbb{R}^{n \times m} \mid (\text{P'}) \text{ is feasible}\}, \quad \mathcal{D}(C, D) := \{\boldsymbol{A} \in \mathbb{R}^{n \times m} \mid (\text{D'}) \text{ is feasible}\},$$

and we call

$$\Sigma(C, D) := \mathcal{P}(C, D) \cap \mathcal{D}(C, D)$$

the set of *ill-posed* inputs. Indeed, it can be shown that $\mathcal{P}(C, D)$ and $\mathcal{D}(C, D)$ are both closed, the union of $\mathcal{P}(C, D)$ and $\mathcal{D}(C, D)$ is the whole input space $\mathbb{R}^{n \times m}$ unless $C = D = \mathbb{R}^m$, and the probability that a Gaussian matrix lies in $\Sigma(C, D)$ is zero [2, Sec. 2.2].

As for the relation of this generalized feasibility problem to conically restricted linear operators, observe that $\boldsymbol{A} \in \mathcal{P}(C, D)$ if and only if $\sigma_{C \to D}(\boldsymbol{A}) = 0$, and $\boldsymbol{A} \in \mathcal{D}(C, D)$ if and only if $\sigma_{D \to C}(-\boldsymbol{A}^T) = 0$. Moreover, we have [2, Sec. 2.2]

$$\text{dist}(\boldsymbol{A}, \mathcal{P}) = \sigma_{C \to D}(\boldsymbol{A}), \qquad \text{dist}(\boldsymbol{A}, \mathcal{D}) = \sigma_{D \to C}(-\boldsymbol{A}^T), \qquad (15)$$

where dist denotes the Euclidean distance.

The following theorem is the consequence of a classic result by Gordon [26]. For various derivations, see [17, 23]. The form stated here, for a map between two cones, is from [2]. Here, we use the Gaussian width $w(C) = \underset{\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_m)}{\mathbb{E}} \big[\|\Pi_C(\boldsymbol{g})\|\big]$.

Theorem 4.1. *Let $C \subseteq \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$ closed convex cones, and let $\boldsymbol{G} \in \mathbb{R}^{n \times m}$ be a Gaussian matrix. Then for $\lambda \geq 0$,*

$$\text{Prob}\big\{\|\boldsymbol{G}\|_{C \to D} \geq w(C) + w(D) + \lambda\big\} \leq e^{-\frac{\lambda^2}{2}},$$
$$\text{Prob}\big\{\sigma_{C \to D}(\boldsymbol{G}) \leq w(D) - w(C) - \lambda\big\} \leq e^{-\frac{\lambda^2}{2}}. \qquad (16)$$

Before getting to the proof of the asymptotic statement in Theorem 1.5 we use Gordon's estimate (16) for the following finite-dimensional bound.

Proposition 4.2. *Let $C \subseteq \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$ closed convex cones with $w(D) - w(C) > 2\sqrt{2}$, and let $\boldsymbol{G} \in \mathbb{R}^{n \times m}$ be a Gaussian matrix. Then there exists a set $E \subseteq \mathbb{R}^{n \times m}$ with $\text{Prob}\{\boldsymbol{G} \in E\} < \varepsilon := 2e^{-(w(D) - w(C))^2/8}$, such that*

$$\mathbb{E}[\mathcal{R}_{C,D}(\boldsymbol{G}) \mid \boldsymbol{G} \notin E] < \frac{1}{1 - \varepsilon}\left(\frac{w(\mathbb{R}^n) + w(\mathbb{R}^m)}{w(D) - w(C)} + 4 \int_{\frac{b}{a}}^{\frac{a+b}{2a}} \frac{e^{-\frac{(as-b)^2}{2}}}{(1-s)^2} \, ds\right), \qquad (17)$$

where $a = \frac{1}{2}\big(w(\mathbb{R}^m) + w(\mathbb{R}^n) + w(D) - w(C)\big)$, $b = \frac{1}{2}\big(w(\mathbb{R}^m) + w(\mathbb{R}^n) - (w(D) - w(C))\big)$.

Proof. To ease the notation, set $w_m = w(\mathbb{R}^m)$, $w_n = w(\mathbb{R}^n)$. Note that $a + b = w_m + w_n$ and $a - b = w(D) - w(C)$. Equations (16) and the union bound give for $0 \leq \lambda < w(D) - w(C)$,

$$\text{Prob}\left\{\frac{\|\boldsymbol{G}\|}{\sigma_{C \to D}(\boldsymbol{G})} \geq \frac{w_m + w_n + \lambda}{w(D) - w(C) - \lambda}\right\} \leq 2e^{-\frac{\lambda^2}{2}}.$$

We introduce $t = \frac{w_m + w_n + \lambda}{w(D) - w(C) - \lambda}$, for which $\frac{t-1}{t+1} = \frac{b+\lambda}{a}$, so that we can rewrite the bound as

$$\text{Prob}\left\{\frac{\|\boldsymbol{G}\|}{\sigma_{C \to D}(\boldsymbol{G})} \geq t\right\} \leq 2\exp\left(-\frac{1}{2}\left(a\,\frac{t-1}{t+1} - b\right)^2\right).$$

This holds for $\frac{t-1}{t+1} \geq \frac{b}{a}$, or equivalently, $t \geq \frac{a+b}{a-b} = \frac{w_m + w_n}{w(D) - w(C)}$. In particular, for $t_\varepsilon := 2\frac{a+b}{a-b} + 1$, that is, $\frac{t_\varepsilon - 1}{t_\varepsilon + 1} = \frac{a+b}{2a}$, we have

$$\text{Prob}\left\{\frac{\|\boldsymbol{G}\|}{\sigma_{C \to D}(\boldsymbol{G})} \geq t_\varepsilon\right\} \leq 2\exp\left(-\frac{(a-b)^2}{8}\right) = \varepsilon.$$

Defining the exceptional set $E := \{\boldsymbol{G} \mid \frac{\|\boldsymbol{G}\|}{\sigma_{C \to D}(\boldsymbol{G})} \geq t_\varepsilon\}$, we have, in particular, that $\sigma_{C \to D}(\boldsymbol{G}) > 0$ for $\boldsymbol{G} \notin E$. By (1) and (15) this implies that the condition number of $\boldsymbol{G} \notin E$ is given by $\mathcal{R}_{C,D}(\boldsymbol{G}) = \frac{\|\boldsymbol{G}\|}{\sigma_{C \to D}(\boldsymbol{G})}$. For the conditional expectation we thus obtain

$$\mathbb{E}[\mathcal{R}_{C,D}(\boldsymbol{G}) \mid \boldsymbol{G} \notin E] \leq \frac{1}{1-\varepsilon} \int_0^{t_\varepsilon} \text{Prob}\left\{\mathcal{R}_{C,D}(\boldsymbol{G}) \geq t \text{ and } \boldsymbol{G} \notin E\right\} dt$$

$$\leq \frac{1}{1-\varepsilon}\left(\frac{a+b}{a-b} + 2\int_{\frac{a+b}{a-b}}^{t_\varepsilon} \exp\left(-\frac{1}{2}\left(a\,\frac{t-1}{t+1} - b\right)^2\right) dt\right).$$

Substituting $s = \frac{t-1}{t+1}$, the expression reads as

$$\mathbb{E}[\mathcal{R}_{C,D}(\boldsymbol{G}) \mid \boldsymbol{G} \notin E] \leq \frac{1}{1-\varepsilon}\left(\frac{a+b}{a-b} + 4\int_{\frac{b}{a}}^{\frac{a+b}{2a}} \frac{e^{-\frac{(as-b)^2}{2}}}{(1-s)^2} ds\right),$$

which was to be shown. $\square$

For the proof of Theorem 1.5 it remains to estimate the integral in (17). Since the boundaries of integration are both in the interval $(0,1)$ we could get an easy bound by computing the maximum of the integrand and bounding the integral accordingly. Instead, we will use Laplace's method [4] to show that the integral goes to zero in the assumed asymptotic setting.

Proof of Theorem 1.5. Recall that $C_k \subseteq \mathbb{R}^{m_k}$, $D_k \subseteq \mathbb{R}^{n_k}$ are families of cones such that the dimensions satisfy $\lim_{k \to \infty} \delta(C_k)/m_k = \alpha^2$, $\lim_{k \to \infty} \delta(D_k)/n_k = \beta^2$, and $\lim_{k \to \infty} m_k/n_k = \gamma^2$, with $0 < \alpha, \beta, \gamma \leq 1$ and $\beta > \alpha\gamma$. To see that Proposition 4.2, applied to these sequences of cones, does indeed give the claimed asymptotics in Theorem 1.5, note first that the size of the exceptional set is bounded by $e^{-\Omega(n_k)}$, as

$$\frac{(w(D) - w(C))^2}{n_k} = \frac{w(D)^2}{n_k} - 2\frac{w(D)w(C)}{n_k} + \frac{w(C)^2}{n_k}$$

$$\to \beta^2 - 2\alpha\beta\gamma + \alpha\gamma = \beta(\beta - \alpha\gamma) + \alpha\gamma(1-\beta) > 0,$$

where we used $w(C)^2 \leq \delta(C) \leq w(C)^2 + 1$ and the assumption $n_k \to \infty$ for $k \to \infty$. So the size of the exceptional set is small enough.

Hence, all that remains to show is that the integral in the bound (17) goes to zero for $k \to \infty$. More precisely, if we define as in Proposition 4.2,

$$a_k = \frac{w(\mathbb{R}^{m_k}) + w(\mathbb{R}^{n_k})}{2} + \frac{w(D_k) - w(C_k)}{2}, \qquad b_k = \frac{w(\mathbb{R}^{m_k}) + w(\mathbb{R}^{n_k})}{2} - \frac{w(D_k) - w(C_k)}{2},$$

then the decisive term from (17) is given by

$$R_k := \int_{\frac{b_k}{a_k}}^{\frac{a_k+b_k}{2a_k}} g(s) \exp\left(n_k f_k(s)\right) ds,$$

where $g(s) = (1-s)^{-2}$ and $f_k(s) = -\frac{(a_k s - b_k)^2}{2n_k}$, which we need to show that it goes to zero.

The lower integral bound is not a problem so that we use the trivial lower bound $b_k/a_k \geq 0$. For the upper integral bound we have the asymptotics

$$\frac{a_k + b_k}{2a_k} = \frac{w(\mathbb{R}^{m_k}) + w(\mathbb{R}^{n_k})}{w(\mathbb{R}^{m_k}) + w(\mathbb{R}^{n_k}) + w(D_k) - w(C_k)} \to u := \frac{1}{1 + \frac{\beta + \alpha\gamma}{1+\gamma}} < 1,$$

and for the functions f_k we notice that

$$f_k(s) = -\frac{\left(\left(\frac{w(\mathbb{R}^{m_k})}{n_k} + \frac{w(\mathbb{R}^{n_k})}{n_k}\right)(s-1) + \left(\frac{w(D_k)}{n_k} - \frac{w(C_k)}{n_k}\right)(s+1)\right)^2}{8}$$

$$\to f(s) := -\frac{((\gamma+1)(s-1) + (\beta - \alpha\gamma)(s+1))^2}{8}.$$

So if we define

$$R := \int_0^u g(s)\, e^{nf(s)}\, ds,$$

we see that all that remains is to show that $R \to 0$ for $n \to \infty$. Now, $f(s)$ can be written in the form

$$f(s) = -\frac{(cs-d)^2}{8}, \qquad c = \gamma + 1 + \beta - \alpha\gamma, \quad d = \gamma + 1 - (\beta - \alpha\gamma),$$

and we can estimate $R \leq g(u) \int_0^u e^{-n(cs-d)^2/8}\, ds$. Since

$$0 < \frac{d}{c} = \frac{\gamma + 1 - (\beta - \alpha\gamma)}{\gamma + 1 + \beta - \alpha\gamma} < \frac{\gamma + 1}{\gamma + 1 + \beta - \alpha\gamma} = u,$$

Laplace's method yields $\int_0^u e^{-n(cs-d)^2/8}\, ds = O(n^{-1/2})$. This shows $R \to 0$ for $n \to \infty$ and thus finishes the proof. $\qquad\square$

References

[1] D. Amelunxen and P. Bürgisser. Probabilistic analysis of the Grassmann condition number. *Found. Comput. Math.*, 15(1):3–51, 2015.

[2] D. Amelunxen and M. Lotz. Gordon's inequality and condition numbers in conic optimization. arXiv:1408.3016 [math.PR].

[3] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: phase transitions in convex programs with random data. *Inf. Inference*, 3(3):224–294, 2014.

[4] C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Advanced Mathematical Methods for Scientists and Engineers. Springer, 1999.

[5] P. Bianchi, M. Debbah, and J. Najim. Asymptotic independence in the spectrum of the Gaussian unitary ensemble. *Electron. Commun. Probab*, 15:376–395, 2010.

[6] F. Bornemann. Asymptotic independence of the extreme eigenvalues of Gaussian unitary ensemble. *Journal of Mathematical Physics*, 51(2):023514, 2010.

[7] P. Bürgisser. Smoothed analysis of condition numbers. In *Proceedings of the International Congress of Mathematicians, Hyderabad*, volume IV, pages 2609–2633. World Scientific, 2010.

[8] P. Bürgisser and F. Cucker. Smoothed analysis of Moore–Penrose Inversion. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2769–2783, 2010.

[9] P. Bürgisser and F. Cucker. *Condition: The geometry of numerical algorithms*. Number 349 in Grundlehren der Mathematischen Wissenschaften. Springer Verlag, 2013.

[10] P. Bürgisser, F. Cucker, and M. Lotz. The probability that a slightly perturbed numerical analysis problem is difficult. *Math. Comp.*, 77(263):1559–1583, 2008.

[11] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[12] Z. Chen and J. J. Dongarra. Condition numbers of Gaussian random matrices. *SIAM J. Matrix Anal. Appl.*, 27(3):603–620 (electronic), 2005.

[13] A. Coja-Oghlan. The asymptotic k-sat threshold. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 804–813. ACM, 2014.

[14] A. Coja-Oghlan and D. Vilenchik. Chasing the k-colorability threshold. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 380–389. IEEE, 2013.

[15] A. Coja-Oghlan and L. Zdeborová. The condensation transition in random hypergraph 2-coloring. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 241–250. SIAM, 2012.

[16] J. Dahl. Semidefinite optimization using MOSEK. ISMP, Berlin, 2012.

[17] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and Banach spaces. *Handbook of the geometry of Banach spaces*, 1:317–366, 2001.

[18] D. S. Dean and S. N. Majumdar. Extreme value statistics of eigenvalues of Gaussian random matrices. *Phys. Rev. E (3)*, 77(4):041108, 12, 2008.

[19] V. Diekert, A. G. Myasnikov, and A. Weiß. Amenability of Schreier graphs and strongly generic algorithms for the conjugacy problem. arXiv:1501.05579 [math.GR].

[20] A. Edelman and M. La Croix. The singular values of the GUE (less is more). arXiv:1410.7065 [math.PR].

[21] A. Edelman and N. R. Rao. Random matrix theory. *Acta Numerica*, 14:233–297, 2005.

[22] L. Erdős, B. Schlein, and H.-T. Yau. Wegner estimate and level repulsion for Wigner random matrices. *International Mathematics Research Notices*, 2010(3):436–479, 2010.

[23] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*, volume 336 of *Applied and Numerical Harmonic Analysis*. Birkhäuser, Basel, 2013.

[24] R. M. Freund, F. Ordóñez, and K.-C. Toh. Behavioral measures and their correlation with IPM iteration counts on semi-definite programming problems. *Mathematical programming*, 109(2-3):445–475, 2007.

[25] G. Golub and C. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[26] Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pages 84–106. Springer, Berlin, 1988.

[27] I. M. Johnstone and Z. Ma. Fast approach to the Tracy-Widom law at the edge of GOE and GUE. *The Annals of Applied Arobability: an official journal of the Institute of Mathematical Statistics*, 22(5):1962, 2012.

[28] E. Kostlan. Complexity theory of numerical linear algebra. *J. Comput. Appl. Math.*, 22(2-3):219–230, 1988. Special issue on emerging paradigms in applied mathematical modelling.

[29] E. Kostlan. Statistical complexity of dominant eigenvector calculation. *J. Complexity*, 7(4):371–379, 1991.

[30] M. Ledoux. Deviation inequalities on largest eigenvalues. In *Geometric aspects of functional analysis*, pages 167–219. Springer, 2007.

[31] H. Nguyen, T. Tao, and V. Vu. Random matrices: tail bounds for gaps between eigenvalues. arXiv:1504.00396 [math.PR].

[32] J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994.

[33] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Programming*, 70(3, Ser. A):279–351, 1995.

[34] V. Roulet, N. Boumal, and A. d'Aspremont. Renegar's condition number and compressed sensing performance. arXiv:1506.03295 [math.PR].

[35] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739, 2009.

[36] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians, Hyderabad*, volume III, pages 1576–1602. World Scientific, 2010.

[37] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms and heuristics: progress and open questions. In *Foundations of computational mathematics, Santander 2005*, volume 331 of *London Math. Soc. Lecture Note Ser.*, pages 274–342. Cambridge Univ. Press, Cambridge, 2006.

[38] R. Venkataramanan and S. Tatikonda. The Gaussian rate-distortion function of sparse regression codes with optimal encoding. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 2844–2848. IEEE, 2014.

[39] J. C. Vera, J. C. Rivera, J. Peña, and Y. Hui. A primal-dual symmetric relaxation for homogeneous conic systems. *J. Complexity*, 23(2):245–261, 2007.

[40] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. C. Eldar and G. Kutyniok, editors, *Compressed sensing*, pages xii+544. Cambridge University Press, Cambridge, 2012. Theory and applications.

[41] J. von Neumann and A. H. Taub. *John von Neumann Collected Works: Volume V - Design of Computers, Theory of Automata and Numerical Analysis.*

[42] J. H. Wilkinson. *The algebraic eigenvalue problem.* Clarendon Press, Oxford, 1965.

[43] M. Wschebor. Smoothed analysis of $\kappa(a)$. *J. of Complexity*, 20:97–107, 2004.