

LIVING ON THE EDGE: A GEOMETRIC THEORY OF PHASE TRANSITIONS IN CONVEX OPTIMIZATION

DENNIS AMELUNXEN, MARTIN LOTZ, MICHAEL B. MCCOY, AND JOEL A. TROPP

ABSTRACT. Recent empirical research indicates that many convex optimization problems with random constraints exhibit a phase transition as the number of constraints increases. For example, this phenomenon emerges in the ℓ_1 minimization method for identifying a sparse vector from random linear samples. Indeed, this approach succeeds with high probability when the number of samples exceeds a threshold that depends on the sparsity level; otherwise, it fails with high probability.

This paper provides the first rigorous analysis that explains why phase transitions are ubiquitous in random convex optimization problems. It also describes tools for making reliable predictions about the quantitative aspects of the transition, including the location and the width of the transition region. These techniques apply to regularized linear inverse problems with random measurements, to demixing problems under a random incoherence model, and also to cone programs with random affine constraints.

These applications depend on foundational research in conic geometry. This paper introduces a new summary parameter, called the statistical dimension, that canonically extends the dimension of a linear subspace to the class of convex cones. The main technical result demonstrates that the sequence of conic intrinsic volumes of a convex cone concentrates sharply near the statistical dimension. This fact leads to an approximate version of the conic kinematic formula that gives bounds on the probability that a randomly oriented cone shares a ray with a fixed cone.

1. MOTIVATION AND CONTRIBUTIONS

A *phase transition* is a sharp change in the character of a computational problem as its parameters vary. Recent work indicates that phase transitions emerge in a variety of random convex optimization problems from mathematical signal processing and computational statistics; for example, see [DT09b, Sto09, RFP10, CSPW11, MT12, DGM13]. This paper develops geometric tools that allow us to identify the location of these phase transitions using geometric invariants associated with the mathematical program. Our analysis provides the first fully rigorous account of transition phenomena in random linear inverse problems, random demixing problems, and random cone programs.

1.1. Vignette: Compressed sensing. To illustrate our goals, we discuss the compressed sensing problem, a familiar example where a phase transition is plainly visible in numerical experiments [DT09b]. Let $\mathbf{x}_0 \in \mathbb{R}^d$ be an unknown vector with s nonzero entries. Suppose we have access to a vector $\mathbf{z}_0 \in \mathbb{R}^m$ consisting of random linear samples of \mathbf{x}_0 , where the number m of samples is smaller than the ambient dimension d . More precisely, assume that $\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0$ where \mathbf{A} is an $m \times d$ matrix with independent standard normal entries. The aim is to reconstruct \mathbf{x}_0 given the measurement vector \mathbf{z}_0 , the measurement matrix \mathbf{A} , and the prior knowledge that \mathbf{x}_0 is sparse. One may regard this formulation as a toy model for understanding when it is possible to solve an underdetermined linear inverse problem with a structured unknown.

A well-established approach [CDS01, CT06, Don06a] to the compressed sensing problem is the method of ℓ_1 minimization:

$$\text{minimize } \|\mathbf{x}\|_1 \quad \text{subject to } \mathbf{z}_0 = \mathbf{A}\mathbf{x}. \quad (1.1)$$

We say that the convex program (1.1) *succeeds* at solving the compressed sensing problem when it has a unique optimal point $\hat{\mathbf{x}}$ that equals the true unknown \mathbf{x}_0 ; otherwise, it *fails*.

Figure 1.1 depicts the results of a computer experiment designed to estimate the empirical probability that (1.1) succeeds (with respect to the randomness in \mathbf{A}) as the sparsity level s and the number m of samples range from zero to the ambient dimension d . The plot evinces that, for a given sparsity level s , the ℓ_1

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE MAR 2013		2. REPORT TYPE		3. DATES COVERED 00-00-2013 to 00-00-2013	
4. TITLE AND SUBTITLE Living on the Edge: A Geometric Theory of Phase Transitions in Convex Optimization				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) California Institute of Technology, Department of Mechanical Engineering, Pasadena, CA, 91125				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Recent empirical research indicates that many convex optimization problems with random constraints exhibit a phase transition as the number of constraints increases. For example, this phenomenon emerges in the '1 minimization method for identifying a sparse vector from random linear samples. Indeed, this approach succeeds with high probability when the number of samples exceeds a threshold that depends on the sparsity level; otherwise it fails with high probability. This paper provides the first rigorous analysis that explains why phase transitions are ubiquitous in random convex optimization problems. It also describes tools for making reliable predictions about the quantitative aspects of the transition, including the location and the width of the transition region. These techniques apply to regularized linear inverse problems with random measurements, to demixing problems under a random incoherence model and also to cone programs with random affine constraints. These applications depend on foundational research in conic geometry. This paper introduces a new summary parameter, called the statistical dimension, that canonically extends the dimension of a linear subspace to the class of convex cones. The main technical result demonstrates that the sequence of conic intrinsic volumes of a convex cone concentrates sharply near the statistical dimension. This fact leads to an approximate version of the conic kinematic formula that gives bounds on the probability that a randomly oriented cone shares a ray with a fixed cone.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 46	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

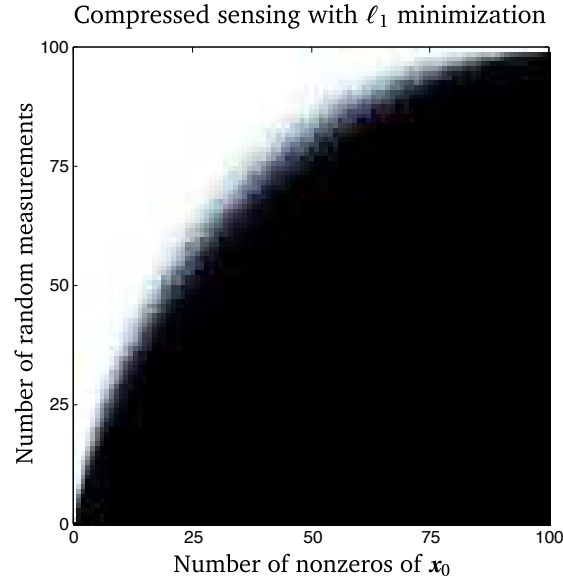


FIGURE 1.1: **Empirical phase transition in compressed sensing.** The colormap indicates the empirical probability that the ℓ_1 minimization problem (1.1) successfully recovers a sparse vector $\mathbf{x}_0 \in \mathbb{R}^{100}$ from the vector $\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0$ of random linear measurements, where \mathbf{A} is a standard normal matrix. The probability of success increases with brightness from certain failure (black) to certain success (white).

minimization technique (1.1) almost always succeeds when we have an adequate number m of samples, while it almost always fails when we have fewer samples. See Appendix A for the experimental details.

Figure 1.1 raises several interesting questions about the performance of the ℓ_1 minimization method for solving the compressed sensing problem:

- **What is the probability of success?** For a given pair (s, m) of parameters, can we estimate the probability that (1.1) succeeds or fails?
- **Does a phase transition exist?** Is there a simple curve $m = \psi(s)$ that separates the parameter space into regions where (1.1) succeeds or fails most of the time?
- **Where is the edge of the phase transition?** Can we find a formula for the location of this threshold between success and failure?
- **How wide is the transition region?** For a given sparsity level s and ambient dimension d , how big is the range of m where the probability of success and failure are comparable?
- **Why does the transition exist?** Is there a geometric explanation for the phase transition in compressed sensing? Can we export this reasoning to understand other problems?

In Section 10, we summarize a large corpus of research that has attempted to address these questions. Unfortunately, the current results are fragmentary, even for the vanilla compressed sensing problem. This work provides a detailed answer to each of the questions we have posed.

1.2. Contributions. We approach phase transition phenomena by studying the intrinsic geometric properties of convex cones. This attack is appropriate for the compressed sensing problem because we can express the optimality condition for (1.1) in terms of a descent cone of the ℓ_1 norm. Our techniques apply more broadly because convex cones play a central role in convex optimization. Let us summarize the main contributions of this work.

- We introduce a new summary parameter for convex cones, which we call the *statistical dimension*. This quantity canonically extends the linear dimension of a subspace to the class of convex cones. (See Definition 2.1, Proposition 4.1, Proposition 5.11, and Proposition 5.12.)

- We prove that a regularized linear inverse problem with random measurements, such as the compressed sensing problem, must exhibit a phase transition as the number of random measurements increases. The location and width of the transition are controlled by the statistical dimension of a cone associated with the regularizer and the unknown. (Theorem II, Theorem 7.1, and Proposition 9.1.)
- The literature [MT12] describes convex programming methods for decomposing a superposition of two structured, randomly oriented vectors into its constituents. These methods exhibit a phase transition whose properties depend on the total statistical dimension of two convex cones. (Theorem III and Theorem 7.1.)
- A cone program with random affine constraints displays a phase transition as the number of constraints increases. We can predict the transition using the statistical dimension of the cone. (Theorem 8.1.)
- In Section 4, we explain how to compute the statistical dimension for several important families of cones that arise in convex optimization and mathematical signal processing. These calculations are not substantially novel, but we provide the first rigorous proof that they are sharp in high dimensions. (Propositions 4.2, 4.3, 4.7, 4.8, and 4.9 and Theorem 4.5.)

These applied results are supported by foundational research in conic geometry. Our theoretical analysis identifies a new geometric phenomenon that governs all the phase transitions mentioned above.

- We prove that the sequence of conic intrinsic volumes of a general convex cone concentrates at the statistical dimension of the cone. This result provides a new family of nontrivial inequalities among the conic intrinsic volumes. (Theorem 6.1 and Section 6.1.)
- The concentration result implies an approximate version of the kinematic formula for cones. This result uses the statistical dimension to bound the probability that a randomly oriented cone shares a ray with a fixed cone. (Theorem I and Theorem 7.1.)

As an added bonus, we provide the final ingredient needed to resolve a series of conjectures [DMM09a, DJM11, DGM13] about the coincidence between the minimax risk of denoising and the location of phase transitions in linear inverse problems. Indeed, Oymak & Hassibi [OH12] have recently shown that the minimax risk is equivalent with the statistical dimension of an appropriate cone, and our results prove that the phase transition occurs at precisely this spot. See Section 10.4 for further details.

1.3. Roadmap. Section 2 explains the connection between conic geometry and phase transitions in two applications. We analyze both problems, and we showcase computer experiments that confirm the accuracy of our predictions. Section 3 introduces notation and background material. In Section 4, we explain how to compute the statistical dimension for several important families of cones, and we prove that the calculations are sharp. Section 5 summarizes the central concepts from conic geometric probability, including the definition of conic intrinsic volumes. Sections 6 and 7 form the technical kernel of the paper. Here, we establish that the sequence of conic intrinsic volumes concentrates at the statistical dimension, and we derive an approximate kinematic formula as a consequence. Afterward, we return to applications. Section 8 shows that cone programs with random affine constraints exhibit a phase transition. Section 9 uses our theory to prove that a certain linear inverse problem is bound to fail. Finally, we canvass the related work in Section 10.

The paper contains five substantial appendices, which house many of the technical details. Appendix A provides information about the computer experiments. Appendix B justifies the methods for computing the statistical dimension of a descent cone. Appendix C completes several statistical dimension calculations. Appendix D contains most of the analysis required to show that the sequence of conic intrinsic volumes concentrates. Finally, Appendix E establishes a connection between the statistical dimension and another geometric quantity called the Gaussian width.

2. CONIC GEOMETRY AND PHASE TRANSITIONS

The existence of phase transitions depends on a striking new fact about the geometry of cones. Indeed, our entire approach can be summarized in the following precept:

Each convex cone C admits a “dimension” parameter $\delta(C)$. For certain purposes, the cone behaves like a linear subspace with dimension $[\delta(C)]$.

This section introduces the quantity $\delta(C)$, which we call the *statistical dimension* of the cone. Then we state an approximate kinematic formula, which uses the statistical dimension to express the probability that

two randomly oriented cones intersect nontrivially. This theorem parallels familiar statements about the intersection of two randomly oriented subspaces. Afterward, we apply the approximate kinematic formula to explain why phase transitions arise in random instances of regularized linear inverse problems (such as compressed sensing) and in a related class of demixing problems.

2.1. The statistical dimension of a convex cone. For certain purposes, the dimension of a linear subspace carries all pertinent information about the subspace. In particular, we can resolve questions in stochastic geometry, such as the probability that two random subspaces have a nontrivial intersection, as soon as we know the dimensions of the subspaces.

Consider the more general problem of determining the probability that a randomly oriented subspace shares a ray with a fixed convex cone. Prima facie, it seems impossible that we could answer this question without detailed information about the cone. Nonetheless, there is a single parameter that encapsulates all the relevant geometry of the cone.

Definition 2.1 (Statistical dimension). The *statistical dimension* $\delta(C)$ of a closed convex cone $C \subset \mathbb{R}^d$ is defined as

$$\delta(C) := \mathbb{E} [\|\Pi_C(\mathbf{g})\|^2], \quad (2.1)$$

where $\mathbf{g} \in \mathbb{R}^d$ is a standard normal vector, $\|\cdot\|$ is the Euclidean norm, and Π_C denotes the Euclidean projection onto the cone C :

$$\Pi_C(\mathbf{x}) := \arg \min \{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in C\}.$$

We define the statistical dimension of a general convex cone to be the statistical dimension of its closure.

As we show in Section 5.3, the statistical dimension canonically extends the linear dimension of a subspace to the class of convex cones. Let us highlight a few properties that support this claim. First, the statistical dimension of a subspace $L \subset \mathbb{R}^d$ satisfies

$$\delta(L) = \mathbb{E} [\|\Pi_L(\mathbf{g})\|^2] = \dim(L). \quad (2.2)$$

Second, the statistical dimension of a cone $C \subset \mathbb{R}^d$ is rotationally invariant:

$$\delta(\mathbf{U}C) = \delta(C) \quad \text{for each orthogonal matrix } \mathbf{U} \in \mathbb{R}^{d \times d}. \quad (2.3)$$

Third, the statistical dimension increases with the size of the cone in the sense that $C \subset K$ implies that $\delta(C) \leq \delta(K)$. See Proposition 4.1 for details about these and other properties of the statistical dimension.

Section 4 explains how to calculate the statistical dimension for several important families of convex cones. The examples include self-dual cones, circular cones, descent cones of the ℓ_1 norm, and descent cones of the Schatten 1-norm.¹ *In all of these cases, we prove rigorously that our estimates for the statistical dimension are sharp in high dimensions.*

Remark 2.2 (Related concepts). The statistical dimension of a cone is closely related to its Gaussian width, another summary parameter for cones that has been proposed in the literature [RV08, Sto09, CRPW12]. See Section 10.3 for further discussion of Gaussian widths. Recent work of Oymak & Hassibi [OH12] also illuminates a connection between the statistical dimension and the minimax risk for denoising problems; see Section 10.4.

2.2. Kinematics and statistical dimension. Conic integral geometry [SW08, Ch. 6] studies properties of convex cones that are invariant under rotation and reflection. A crowning achievement in this field is the *conic kinematic formula* [SW08, Thm. 6.5.6], which yields the exact probability that two randomly oriented convex cones intersect nontrivially. As recognized in [Ame11, MT12], this result is tailor-made for studying random instances of convex optimization problems. Our main applied result is an approximate kinematic formula, expressed in terms of the statistical dimension.

Theorem I (Approximate kinematic formula). *Fix a tolerance $\eta \in (0, 1)$. Suppose that $C, K \subset \mathbb{R}^d$ are closed convex cones, one of which is not a subspace. Draw an orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{d \times d}$ uniformly at random. Then*

$$\begin{aligned} \delta(C) + \delta(K) \leq d - a_\eta \sqrt{d} &\implies \mathbb{P}\{C \cap \mathbf{Q}K = \{\mathbf{0}\}\} \geq 1 - \eta; \\ \delta(C) + \delta(K) \geq d + a_\eta \sqrt{d} &\implies \mathbb{P}\{C \cap \mathbf{Q}K = \{\mathbf{0}\}\} \leq \eta. \end{aligned}$$

The quantity $a_\eta := 4\sqrt{\log(4/\eta)}$. For example, $a_{0.01} < 10$ and $a_{0.001} < 12$.

¹The Schatten 1-norm equals the sum of the singular values of a matrix.

Theorem I allows us to control the probability that two randomly oriented cones strike as soon as we know their statistical dimensions. Roughly, the cones are likely to share a ray if and only if the total dimension of the two cones exceeds the ambient dimension. This claim is in perfect sympathy with the analogous statement about subspaces. In the next few sections, we explore applications of this pluripotent result.

Section 7 contains the proof of Theorem I, which has two main ingredients. The first piece is the conic kinematic formula, which expresses the exact intersection probability in terms of geometric invariants called *conic intrinsic volumes* [SW08, Sec. 6.5]. Each cone in \mathbb{R}^d has $d + 1$ conic intrinsic volumes, much as a convex set in \mathbb{R}^d has a volume, a surface area, a mean width, and so forth [KR97]. The second component of the proof is a new property of conic intrinsic volumes (Theorem 6.1). This result shows that most intrinsic volumes of a cone C are negligible in size, except for the ones whose index is close to the statistical dimension $\delta(C)$. Combining these facts, we can bound the intersection probability accurately via the statistical dimension.

2.3. Regularized linear inverse problems. Our first application of Theorem I concerns a generalization of the compressed sensing problem. A *linear inverse problem* asks us to infer an unknown vector $\mathbf{x}_0 \in \mathbb{R}^d$ from an observed vector $\mathbf{z}_0 \in \mathbb{R}^m$ of the form

$$\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0, \quad (2.4)$$

where $\mathbf{A} \in \mathbb{R}^{m \times d}$ is a matrix that describes a linear data acquisition process. When the matrix is fat ($m < d$), the inverse problem is underdetermined. As a consequence, we cannot hope to identify \mathbf{x}_0 unless we exploit some prior information about its structure.

To solve the underdetermined linear inverse problem, we consider an approach based on optimization. Suppose that $f: \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ is a proper convex function² that reflects the amount of “structure” in a vector. We can attempt to identify the structured unknown \mathbf{x}_0 in (2.4) by solving a convex optimization problem:

$$\text{minimize } f(\mathbf{x}) \quad \text{subject to } \mathbf{z}_0 = \mathbf{A}\mathbf{x}. \quad (2.5)$$

The function f is called a *regularizer*, and the formulation (2.5) is called a *regularized linear inverse problem*. To illustrate the kinds of regularizers that arise in practice, we highlight two familiar examples.

Example 2.3 (Sparse vectors). When the vector \mathbf{x}_0 is known to be sparse, we can minimize the ℓ_1 norm to look for sparse solutions of the inverse problem. Repeating (1.1), we have the optimization

$$\text{minimize } \|\mathbf{x}\|_1 \quad \text{subject to } \mathbf{z}_0 = \mathbf{A}\mathbf{x}. \quad (2.6)$$

This approach was proposed by Chen et al. [CDS01], motivated by work in geophysics [CM73, SS86].

Example 2.4 (Low-rank matrices). Suppose that \mathbf{X}_0 is a low-rank matrix, and we have acquired a vector of measurements of the form $\mathbf{z}_0 = \mathcal{A}(\mathbf{X}_0)$ where \mathcal{A} is a linear operator. This process is equivalent with (2.4). We can look for low-rank solutions to the linear inverse problem by minimizing the Schatten 1-norm:

$$\text{minimize } \|\mathbf{X}\|_{S_1} \quad \text{subject to } \mathbf{z}_0 = \mathcal{A}(\mathbf{X}). \quad (2.7)$$

This idea was proposed by Recht et al. [RFP10], motivated by work in control theory [MP97, Faz02].

The paper [CRPW12] presents a general framework for constructing a regularizer f that promotes a specified type of structure, as well as many additional examples.

We say that the regularized inverse problem (2.5) *succeeds* at solving (2.4) when the convex program has a unique minimizer $\hat{\mathbf{x}}$ that coincides with the true unknown; that is, $\hat{\mathbf{x}} = \mathbf{x}_0$. To develop conditions for success, we introduce a convex cone associated with the regularizer f and the unknown \mathbf{x}_0 .

Definition 2.5 (Descent cone). The *descent cone* $\mathcal{D}(f, \mathbf{x}_0)$ of a function $f: \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ at a point $\mathbf{x}_0 \in \mathbb{R}^d$ is the conic hull of the perturbations that do not increase f near \mathbf{x}_0 .

$$\mathcal{D}(f, \mathbf{x}) := \bigcup_{\tau > 0} \{\mathbf{y} \in \mathbb{R}^d : f(\mathbf{x} + \tau\mathbf{y}) \leq f(\mathbf{x})\}.$$

Chandrasekaran et al. [CRPW12, Prop. 2.1] characterize when the optimization problem (2.5) succeeds in terms of a descent cone. This result is simply a geometric statement of the primal optimality condition.

Fact 2.6 (Optimality condition for linear inverse problems). *The vector \mathbf{x}_0 is the unique optimal point of the convex program (2.5) if and only if $\mathcal{D}(f, \mathbf{x}_0) \cap \text{null}(\mathbf{A}) = \{\mathbf{0}\}$.*

²The extended real numbers $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. A *proper* convex function has at least one finite value, and it does not take the value $-\infty$.

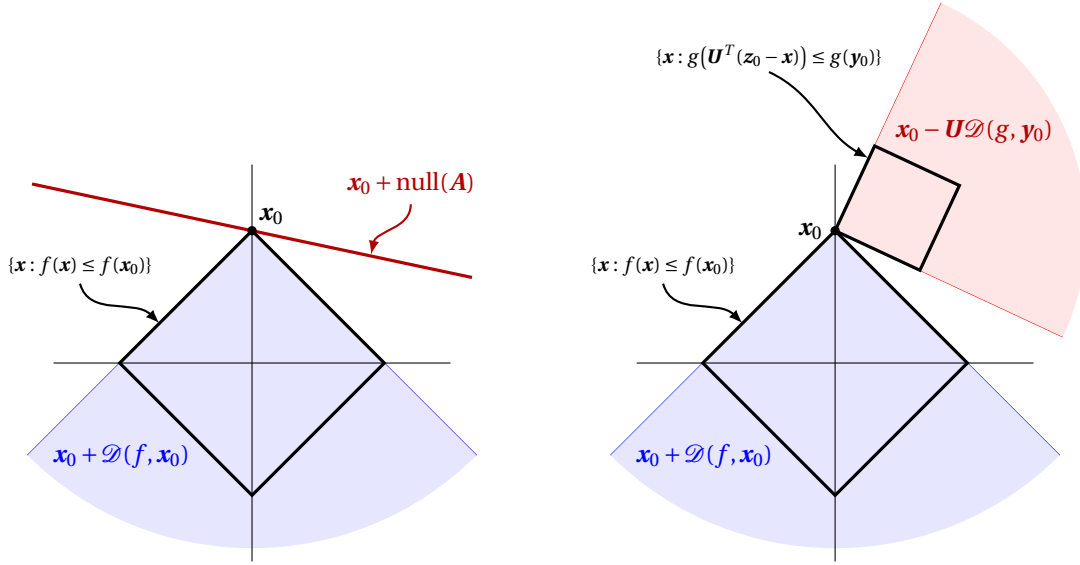


FIGURE 2.1: **Geometry of optimality conditions.** [left] The optimality condition for the regularized linear inverse problem (2.5) states that the descent cone of f at \mathbf{x}_0 is tangent to the null space of A . [right] The optimality condition for the convex demixing method (2.9) states that the descent cone of f at \mathbf{x}_0 has a trivial intersection with a rotated copy of the descent cone of g at \mathbf{y}_0 .

Figure 2.1[left] illustrates the geometry of this optimality condition. Despite its simplicity, this result forges a crucial link between the convex optimization problem (2.5) and the theory of conic integral geometry.

Our goal is to understand the power of convex regularization for solving linear inverse problems, as well as the limitations inherent in this approach. To do so, we consider the case where the measurements are generic. A natural modeling technique is to draw the measurement matrix A at random from the standard normal distribution on $\mathbb{R}^{m \times d}$. For this model, Theorem I allows us to identify a sharp transition in the performance of the regularized problem (2.5).

Theorem II (Phase transitions in linear inverse problems). *Fix a tolerance $\eta \in (0, 1)$. Let $\mathbf{x}_0 \in \mathbb{R}^d$ be a fixed vector. Suppose $A \in \mathbb{R}^{m \times d}$ has independent standard normal entries, and let $\mathbf{z}_0 = A\mathbf{x}_0$. Then*

$$m \geq \delta(\mathcal{D}(f, \mathbf{x}_0)) + a_\eta \sqrt{d} \implies (2.5) \text{ succeeds with probability } \geq 1 - \eta;$$

$$m \leq \delta(\mathcal{D}(f, \mathbf{x}_0)) - a_\eta \sqrt{d} \implies (2.5) \text{ succeeds with probability } \leq \eta.$$

The quantity $a_\eta := 4\sqrt{\log(4/\eta)}$.

Proof. The standard normal distribution on $\mathbb{R}^{m \times d}$ is invariant under rotation, so the null space $L = \text{null}(A)$ is almost surely a uniformly random $(d - m)$ -dimensional subspace of \mathbb{R}^d . According to (2.2), the statistical dimension $\delta(L) = d - m$ almost surely. The result follows immediately when we combine the optimality condition, Fact 2.6, and the kinematic bound, Theorem I. \square

Theorem II proves that we always encounter a phase transition when we use the regularized formulation (2.5) to solve the linear inverse problem (2.4) with random measurements. The transition occurs where the number of measurements equals the statistical dimension of the descent cone: $m = \delta(\mathcal{D}(f, \mathbf{x}_0))$. The shift from failure to success takes place over a range of $O(\sqrt{d})$ measurements.

In Section 4.5, we estimate the statistical dimension of the descent cone of the ℓ_1 norm at a sparse vector of fixed dimension. Owing to symmetries of the ℓ_1 norm, only the sparsity and the ambient dimension play a role in this calculation. When combined with Theorem II, this result yields the exact (asymptotic) location of the phase transition for the ℓ_1 minimization problem (2.6) with random measurements. In Section 4.6, we estimate the statistical dimension of the descent cone of the Schatten 1-norm at a low-rank matrix with

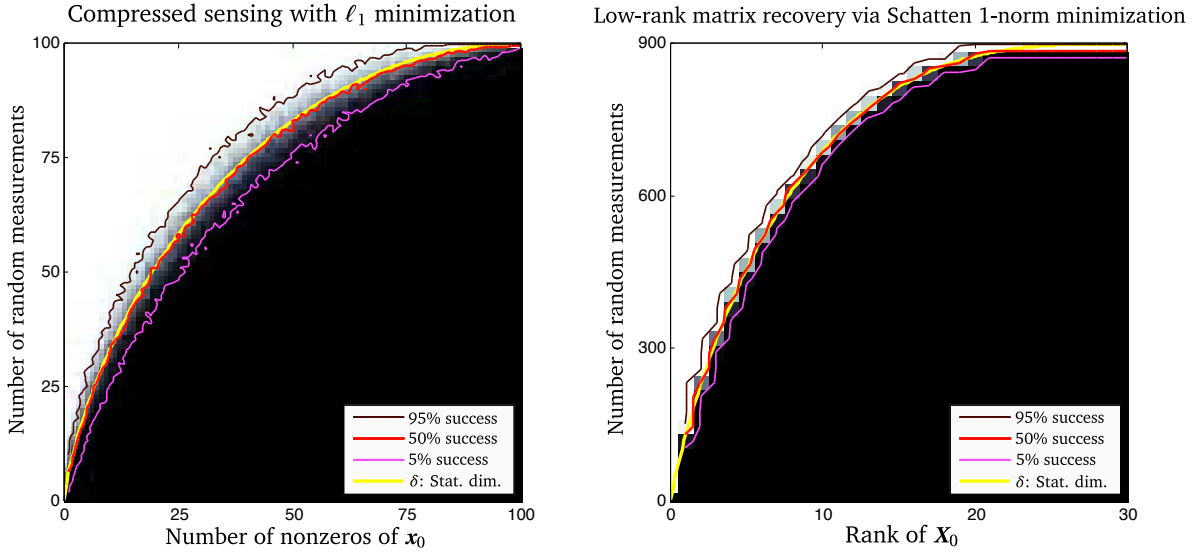


FIGURE 2.2: **Phase transitions for linear inverse problems.** [left] **Recovery of sparse vectors.** The empirical probability that the ℓ_1 minimization problem (2.6) identifies a sparse vector $\mathbf{x}_0 \in \mathbb{R}^{100}$ given random linear measurements $\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0$. [right] **Recovery of low-rank matrices.** The empirical probability that the S_1 minimization problem (2.7) identifies a low-rank matrix $\mathbf{X}_0 \in \mathbb{R}^{30 \times 30}$ given random linear measurements $\mathbf{z}_0 = \mathcal{A}(\mathbf{X}_0)$. In each panel, the colormap indicates the empirical probability of success (black = 0%; white = 100%). The yellow curve marks the theoretical prediction of the phase transition from Theorem II; the red curve traces the empirical phase transition.

fixed dimensions. This calculation gives the exact (asymptotic) location of the phase transition for the S_1 minimization problem (2.7) with random measurements.

To underscore these achievements, we have performed some computer experiments to compare the theoretical and empirical phase transitions. Figure 2.2[left] shows the performance of (2.6) for identifying a sparse vector in \mathbb{R}^{100} ; Figure 2.2[right] shows the performance of (2.7) for identifying a low-rank matrix in $\mathbb{R}^{30 \times 30}$. In each case, the colormap indicates the empirical probability of success over the randomness in the measurement operator. The empirical 5%, 50%, and 95% success isoclines are determined from the data. We also draft the theoretical phase transition curve, promised by Theorem II, where the number m of measurements equals the statistical dimension of the appropriate descent cone, which we compute using the formulas from Sections 4.5 and 4.6. See Appendix A for the experimental protocol.

In both examples, the theoretical prediction of Theorem II coincides almost perfectly with the 50% success isocline. Furthermore, the phase transition takes place over a range of $O(\sqrt{d})$ values of m , as promised. Although Theorem II does not explain why the transition region tapers at the bottom-left and top-right corners of each plot, we have established a more detailed version of Theorem I that allows us to predict this phenomenon as well. See the discussion after Theorem 7.1 for more information.

2.4. Demixing problems. In a demixing problem [MT12], we observe a superposition of two structured vectors, and we aim to extract the two constituents from the mixture. More precisely, suppose that we measure a vector $\mathbf{z}_0 \in \mathbb{R}^d$ of the form

$$\mathbf{z}_0 = \mathbf{x}_0 + \mathbf{U}\mathbf{y}_0 \quad (2.8)$$

where $\mathbf{x}_0, \mathbf{y}_0 \in \mathbb{R}^d$ are unknown and $\mathbf{U} \in \mathbb{R}^{d \times d}$ is a known orthogonal matrix. If we wish to identify the pair $(\mathbf{x}_0, \mathbf{y}_0)$, we must assume that each component is structured to reduce the number of degrees of freedom. In addition, if the two types of structure are coherent (i.e., aligned with each other), it may be impossible to disentangle them, so it is expedient to include the matrix \mathbf{U} to model the relative orientation of the two constituent signals.

To solve the demixing problem (2.8), we describe a convex programming technique proposed in [MT12]. Suppose that f and g are proper convex functions on \mathbb{R}^d that promote the structures we expect to find in \mathbf{x}_0

and \mathbf{y}_0 . Then we can frame the convex optimization problem

$$\text{minimize } f(\mathbf{x}) \quad \text{subject to } g(\mathbf{y}) \leq g(\mathbf{y}_0) \quad \text{and} \quad \mathbf{z}_0 = \mathbf{x} + \mathbf{U}\mathbf{y}. \quad (2.9)$$

In other words, we seek structured vectors \mathbf{x} and \mathbf{y} that are consistent with the observation \mathbf{z}_0 . This approach requires the side information $g(\mathbf{y}_0)$, so a Lagrangian formulation is sometimes more natural in practice. Here are two concrete examples of the demixing program (2.9) that are adapted from the literature.

Example 2.7 (Sparse + sparse). Suppose that the first signal \mathbf{x}_0 is sparse in the standard basis, and the second signal $\mathbf{U}\mathbf{y}_0$ is sparse in a known basis \mathbf{U} . In this case, we can use ℓ_1 norms to promote sparsity, which leads to the optimization

$$\text{minimize } \|\mathbf{x}\|_1 \quad \text{subject to } \|\mathbf{y}\|_1 \leq \|\mathbf{y}_0\|_1 \quad \text{and} \quad \mathbf{z}_0 = \mathbf{x} + \mathbf{U}\mathbf{y}. \quad (2.10)$$

This approach for demixing sparse signals is sometimes called *morphological component analysis* [SDC03, SED05, ESQD05, BMS06].

Example 2.8 (Low-rank + sparse). Suppose that we observe $\mathbf{Z}_0 = \mathbf{X}_0 + \mathcal{U}(\mathbf{Y}_0)$ where \mathbf{X}_0 is a low-rank matrix, \mathbf{Y}_0 is a sparse matrix, and \mathcal{U} is a known orthogonal transformation on the space of matrices. We can minimize the Schatten 1-norm to promote low rank, and we can constrain the ℓ_1 norm to promote sparsity. The optimization becomes

$$\text{minimize } \|\mathbf{X}\|_{S_1} \quad \text{subject to } \|\mathbf{Y}\|_1 \leq \|\mathbf{Y}_0\|_1 \quad \text{and} \quad \mathbf{Z}_0 = \mathbf{X} + \mathcal{U}(\mathbf{Y}). \quad (2.11)$$

This demixing problem is called *rank–sparsity decomposition* [CSPW11].

See the paper [MT12] for some additional examples.

We say that the convex program (2.9) for demixing *succeeds* when it has a unique solution $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ that coincides with the true unknown: $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = (\mathbf{x}_0, \mathbf{y}_0)$. As in the case of a linear inverse problem, we can express the primal optimality condition in terms of descent cones [MT12, Lem. 2.3].

Fact 2.9 (Optimality condition for demixing). *The pair $(\mathbf{x}_0, \mathbf{y}_0)$ is the unique optimal point of the convex program (2.9) if and only if $\mathcal{D}(f, \mathbf{x}_0) \cap (-\mathbf{U}\mathcal{D}(g, \mathbf{y}_0)) = \{\mathbf{0}\}$.*

Figure 2.1[right] depicts the geometry of this optimality condition. The parallel with Fact 2.6, the optimality condition for linear inverse problems, is striking. Indeed, the two conditions coalesce when the function g in (2.9) is the indicator of an affine space.

Our goal is to understand the prospects for solving the demixing problem (2.8) with a convex program of the form (2.9). To that end, we use randomness to model the favorable case where the two structures have no interaction with each other. More precisely, we draw the matrix \mathbf{U} uniformly at random from the orthogonal group. Under this assumption, Theorem I delivers a sharp transition in the performance of the optimization problem (2.9).

Theorem III (Phase transitions in demixing). *Fix a tolerance $\eta \in (0, 1)$. Let \mathbf{x}_0 and \mathbf{y}_0 be fixed vectors in \mathbb{R}^d . Draw an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{d \times d}$ uniformly at random, and let $\mathbf{z}_0 = \mathbf{x}_0 + \mathbf{U}\mathbf{y}_0$. Then*

$$\begin{aligned} \delta(\mathcal{D}(f, \mathbf{x}_0)) + \delta(\mathcal{D}(g, \mathbf{y}_0)) &\leq d - a_\eta \sqrt{d} \quad \implies \quad (2.9) \text{ succeeds with probability } \geq 1 - \eta; \\ \delta(\mathcal{D}(f, \mathbf{x}_0)) + \delta(\mathcal{D}(g, \mathbf{y}_0)) &\geq d + a_\eta \sqrt{d} \quad \implies \quad (2.9) \text{ succeeds with probability } \leq \eta. \end{aligned}$$

The quantity $a_\eta := 4\sqrt{\log(4/\eta)}$.

Proof. This theorem follows immediately when we combine the optimality condition, Fact 2.9, with the kinematic bound, Theorem I. We invoke the rotational invariance (2.3) of the statistical dimension to simplify the formulas. \square

Theorem III establishes that there is always a phase transition when we use the convex program (2.9) to solve the demixing problem (2.8) under the isotropic random model for \mathbf{U} . It is accurate to say that the optimization is effective if and only if the total statistical dimension of the two descent cones is smaller than the ambient dimension d .

Our numerical work confirms the analysis in Theorem III. Figure 2.3[left] shows when (2.10) can demix a sparse vector from a vector that is sparse in a random basis for \mathbb{R}^{100} . Figure 2.3[right] shows when (2.11) can demix a low-rank matrix from a matrix that is sparse in a random basis for $\mathbb{R}^{35 \times 35}$. In each case, the experiment provides an empirical estimate for the probability of success with respect to the randomness in

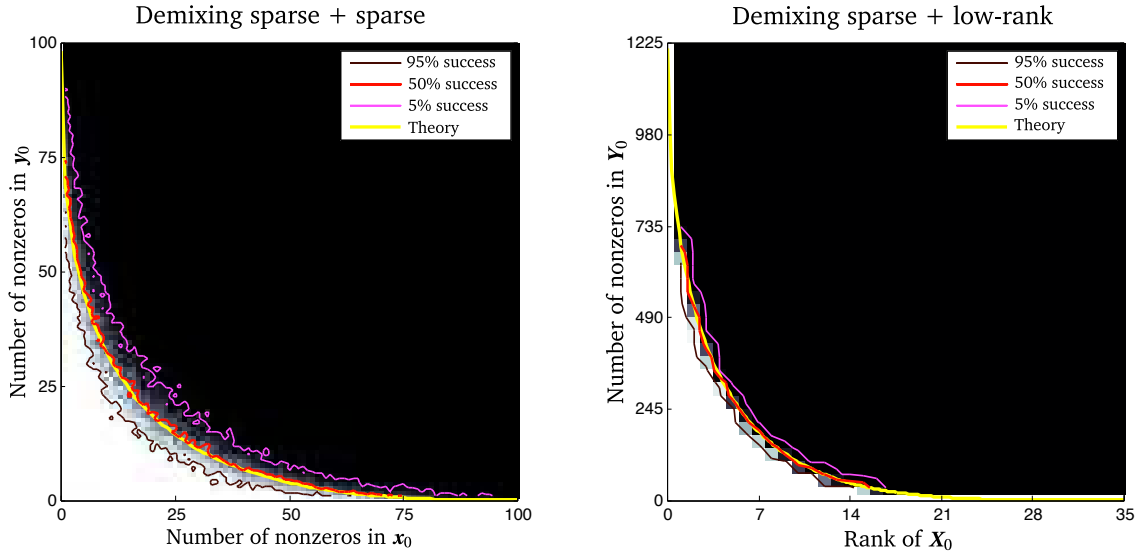


FIGURE 2.3: **Phase transitions for demixing problems.** [left] **Sparse + sparse.** The empirical probability that the convex program (2.10) successfully demixes a vector $\mathbf{x}_0 \in \mathbb{R}^{100}$ that is sparse in the standard basis from a vector $\mathbf{U}\mathbf{y}_0 \in \mathbb{R}^{100}$ that is sparse in the random basis \mathbf{U} . [right] **Low rank + sparse.** The empirical probability that the convex program (2.11) successfully demixes a low-rank matrix $\mathbf{X}_0 \in \mathbb{R}^{35 \times 35}$ from a matrix $\mathcal{U}(\mathbf{Y}_0) \in \mathbb{R}^{35 \times 35}$ that is sparse in the random basis \mathcal{U} . In each panel, the colormap indicates the empirical probability of success (black = 0%; white = 100%). The yellow curve marks the theoretical phase transition predicted by Theorem III. The red curve follows the empirical phase transition.

the measurement operator. We display the 5%, 50%, and 95% success isoclines, determined from the data. We also sketch the theoretical phase transition from Theorem III, which occurs when the total statistical dimension of the relevant cones equals the ambient dimension. The statistical dimension of the descent cones are obtained from the formulas in Sections 4.5 and 4.6. See [MT12, Sec. 6] for the details of the experimental protocol.

Once again, we see that the theoretical curve of Theorem III coincides almost perfectly with the empirical 50% success isocline. The width of the transition region is $O(\sqrt{d})$. Theorem III does not predict the tapering of the transition in the top-left and bottom-right corners, but the discussion after Theorem 7.1 exposes the underlying reason for this phenomenon.

3. NOTATION, CONVENTIONS, AND BACKGROUND

This section contains a short overview of our notation, as well as some important facts from convex geometry that we will use liberally. Some standard references include Rockafellar [Roc70], Hiriart-Urruty & Lemaréchal [HUL93a, HUL93b], and Rockafellar & Wets [RW98].

3.1. Vectors and matrices. We use boldface lowercase letters to denote vectors and boldface capital letters to denote matrices, so \mathbf{x} is a vector and \mathbf{X} is a matrix. We write \mathbf{I} for the identity matrix and $\mathbf{0}$ for the zero vector or matrix; their dimensions are determined by context.

3.2. Euclidean geometry. For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, define the Euclidean inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ and the squared Euclidean norm $\|\mathbf{x}\|^2 := \langle \mathbf{x}, \mathbf{x} \rangle$. The Euclidean unit ball and unit sphere are respectively denoted as

$$B^d := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq 1\} \quad \text{and} \quad S^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}.$$

The group O_d of $d \times d$ orthogonal matrices is defined as

$$O_d := \{\mathbf{U} \in \mathbb{R}^{d \times d} : \mathbf{U}\mathbf{U}^T = \mathbf{I}\}.$$

An *orthogonal basis* for \mathbb{R}^d is an element of O_d .

3.3. Convex cones. Recall that a *convex cone* C is a set that is convex and positive homogeneous: $C = \tau C$ for all $\tau > 0$. A convex cone is *polyhedral* if it can be written as the intersection of a finite number of closed halfspaces. We introduce the family \mathcal{C}_d of all nonempty, closed, convex cones in \mathbb{R}^d .

For a general convex cone $C \subset \mathbb{R}^d$, the *polar cone* $C^\circ \in \mathcal{C}_d$ is the closed convex cone

$$C^\circ := \{\mathbf{u} \in \mathbb{R}^d : \langle \mathbf{u}, \mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{x} \in C\}.$$

The *normal cone* $\mathcal{N}(S, \mathbf{x})$ of a convex set $S \subset \mathbb{R}^d$ at a point $\mathbf{x} \in S$ consists of the outward normals to all hyperplanes that support S at \mathbf{x} , i.e.,

$$\mathcal{N}(S, \mathbf{x}) := \{\mathbf{u} \in \mathbb{R}^d : \langle \mathbf{u}, \mathbf{y} - \mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{y} \in S\}. \quad (3.1)$$

3.4. Representation of descent cones. Let $f: \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$ be a proper convex function. Recall that the *descent cone* of f at a point \mathbf{x} is given by

$$\mathcal{D}(f, \mathbf{x}) := \bigcup_{\tau > 0} \{\mathbf{y} \in \mathbb{R}^d : f(\mathbf{x} + \tau \mathbf{y}) \leq f(\mathbf{x})\}.$$

The polar of a descent cone has some attractive properties that we will exploit later. First, the polar of a descent cone coincides with the normal cone of a sublevel set:

$$\mathcal{D}(f, \mathbf{x})^\circ = \mathcal{N}(S, \mathbf{x}) \text{ where } S = \{\mathbf{y} \in \mathbb{R}^d : f(\mathbf{y}) \leq f(\mathbf{x})\}. \quad (3.2)$$

Next, we introduce the *subdifferential* $\partial f(\mathbf{x})$, which is the closed convex set

$$\partial f(\mathbf{x}) := \{\mathbf{u} \in \mathbb{R}^d : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{u}, \mathbf{y} - \mathbf{x} \rangle \text{ for all } \mathbf{y} \in \mathbb{R}^d\}.$$

In particular, the subdifferential contains the origin $\mathbf{0}$ if and only if \mathbf{x} minimizes f . Assuming that the subdifferential $\partial f(\mathbf{x})$ is nonempty, compact, and does not contain the origin, the result [Roc70, Cor. 23.7.1] provides that

$$\mathcal{D}(f, \mathbf{x})^\circ = \text{cone}(\partial f(\mathbf{x})) := \bigcup_{\tau \geq 0} \tau \cdot \partial f(\mathbf{x}). \quad (3.3)$$

The expression $\tau \cdot \partial f(\mathbf{x})$ represents dilation of the subdifferential by a factor τ . The relation (3.3) offers a powerful tool for computing the statistical dimension of a descent cone. Related identities hold under weaker technical conditions [Roc70, Thm. 23.7].

3.5. Euclidean projections onto sets. Let $S \subset \mathbb{R}^d$ be a closed convex set. The Euclidean distance to the set S is the function

$$\text{dist}(\cdot, S): \mathbb{R}^d \rightarrow \mathbb{R}_+ \text{ where } \text{dist}(\mathbf{x}, S) := \inf\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in S\}.$$

The *Euclidean projection* onto the set S is the map

$$\boldsymbol{\pi}_S: \mathbb{R}^d \rightarrow S \text{ where } \boldsymbol{\pi}_S(\mathbf{x}) := \arg \min\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in S\}.$$

The projection takes a well-defined value because the norm is strictly convex. Let us note some properties of these maps. First, the function $\text{dist}(\cdot, S)$ is convex [Roc70, p. 34]. Next, the maps $\boldsymbol{\pi}_S$ and $\mathbf{I} - \boldsymbol{\pi}_S$ are nonexpansive with respect to the Euclidean norm [Roc70, Thm. 31.5 et seq.]:

$$\|\boldsymbol{\pi}_S(\mathbf{x}) - \boldsymbol{\pi}_S(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\| \text{ and } \|(\mathbf{I} - \boldsymbol{\pi}_S)(\mathbf{x}) - (\mathbf{I} - \boldsymbol{\pi}_S)(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \quad (3.4)$$

As a consequence, the projection $\boldsymbol{\pi}_S$ is continuous, and the distance function is 1-Lipschitz with respect to the Euclidean norm:

$$|\text{dist}(\mathbf{x}, S) - \text{dist}(\mathbf{y}, S)| \leq \|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \quad (3.5)$$

The squared distance is differentiable everywhere, and the derivative satisfies

$$\nabla \text{dist}^2(\mathbf{x}, S) = 2(\mathbf{x} - \boldsymbol{\pi}_S(\mathbf{x})) \text{ for all } \mathbf{x} \in \mathbb{R}^d. \quad (3.6)$$

This point follows from [RW98, Thm. 2.26].

3.6. Euclidean projections onto cones. Let $C \in \mathcal{C}_d$ be a closed convex cone. Recall that the Euclidean projection onto the cone C is the map

$$\Pi_C : \mathbb{R}^d \rightarrow C \quad \text{where} \quad \Pi_C(\mathbf{x}) := \arg \min \{ \|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in C \}.$$

We have used a separate notation for the projection onto a set because the projection onto a cone enjoys a number of additional properties [HUL93a, Sec. III.3.2]. First, the projection onto a cone is nonnegative homogeneous: $\Pi_C(\tau \mathbf{x}) = \tau \Pi_C(\mathbf{x})$ for all $\tau \geq 0$. Next, the cone C induces an orthogonal decomposition of \mathbb{R}^d .

$$\mathbf{x} = \Pi_C(\mathbf{x}) + \Pi_{C^\circ}(\mathbf{x}) \quad \text{and} \quad \langle \Pi_C(\mathbf{x}), \Pi_{C^\circ}(\mathbf{x}) \rangle = 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \quad (3.7)$$

The decomposition (3.7) yields the Pythagorean identity

$$\|\mathbf{x}\|^2 = \|\Pi_C(\mathbf{x})\|^2 + \|\Pi_{C^\circ}(\mathbf{x})\|^2 \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \quad (3.8)$$

It also implies the distance formulas

$$\text{dist}(\mathbf{x}, C) = \|\mathbf{x} - \Pi_C(\mathbf{x})\| = \|\Pi_{C^\circ}(\mathbf{x})\| \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \quad (3.9)$$

The squared norm of the projection has a nice regularity property, which follows from a short argument based on (3.6), (3.7), and (3.9):

$$\nabla \|\Pi_C(\mathbf{x})\|^2 = 2\Pi_C(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \quad (3.10)$$

Finally, the projection map decomposes under Cartesian products. For two cones $C_1 \in \mathcal{C}_{d_1}$ and $C_2 \in \mathcal{C}_{d_2}$ the product $C_1 \times C_2 \in \mathcal{C}_{d_1+d_2}$, and

$$\Pi_{C_1 \times C_2}(\mathbf{x}_1, \mathbf{x}_2) = (\Pi_{C_1}(\mathbf{x}_1), \Pi_{C_2}(\mathbf{x}_2)) \quad \text{for all } \mathbf{x}_1 \in \mathbb{R}^{d_1} \text{ and } \mathbf{x}_2 \in \mathbb{R}^{d_2}. \quad (3.11)$$

The relation (3.11) is easy to check directly.

3.7. Probability. The symbol $\mathbb{P}\{\cdot\}$ denotes the probability of an event, and $\mathbb{E}[\cdot]$ returns the expectation of a random variable. We reserve the letter \mathbf{g} for a standard normal vector, i.e., a vector whose entries are independent normal variables with mean zero and variance one. We reserve the letter $\boldsymbol{\theta}$ for a random vector uniformly distributed on the Euclidean unit sphere. The set O_d of orthogonal matrices is a compact Lie group, so it admits an invariant Haar (i.e., uniform) probability measure. We reserve the letter \mathbf{Q} for a uniformly random element of O_d , and we refer to \mathbf{Q} as a *random orthogonal basis*.

4. CALCULATING THE STATISTICAL DIMENSION

Section 2 demonstrates that we can pinpoint the phase transitions in random linear inverse problems and random demixing problems as soon as we know the statistical dimension of the appropriate descent cones. The goal of this section is to show that we can obtain highly accurate approximation formulas for the statistical dimension of a cone with a modest amount of effort.

Section 4.1 develops some basic properties of the statistical dimension that aid this investigation. The rest of the section explains how to compute the statistical dimension for several families of cones that arise in applications. This discussion includes a recipe for bounding the statistical dimension of a descent cone; Theorem 4.5 ensures that the recipe produces an accurate answer for many problems of interest. We summarize our calculations in Table 4.1 and in Figure 4.1.

4.1. Basic facts about the statistical dimension. The statistical dimension has a number of valuable properties that are readily apparent from Definition 2.1. These facts provide useful tools for making computations, and they strengthen the analogy between the statistical dimension of a cone and the linear dimension of a subspace.

Proposition 4.1 (Properties of statistical dimension). *Let $C \in \mathcal{C}_d$ be a closed convex cone. The statistical dimension obeys the following laws.*

(1) **Gaussian formulation.** *The statistical dimension is defined as*

$$\delta(C) := \mathbb{E} \left[\|\Pi_C(\mathbf{g})\|^2 \right] \quad \text{where } \mathbf{g} \sim \text{NORMAL}(\mathbf{0}, \mathbf{I}_d). \quad (4.1)$$

(2) **Spherical formulation.** *An equivalent definition states that*

$$\delta(C) := d \mathbb{E} \left[\|\Pi_C(\boldsymbol{\theta})\|^2 \right] \quad \text{where } \boldsymbol{\theta} \sim \text{UNIFORM}(S^{d-1}). \quad (4.2)$$

(3) **Rotational invariance.** The statistical dimension does not depend on the orientation of the cone:

$$\delta(\mathbf{U}C) = \delta(C) \quad \text{for each } \mathbf{U} \in \mathbf{O}_d. \quad (4.3)$$

(4) **Subspaces.** For a subspace $L \subset \mathbb{R}^d$, the statistical dimension satisfies $\delta(L) = \dim(L)$.

(5) **Polarity.** The statistical dimension can also be expressed in terms of the polar cone:

$$\delta(C) = \mathbb{E}[\text{dist}^2(\mathbf{g}, C^\circ)]. \quad (4.4)$$

(6) **Totality.** The total statistical dimension of a cone and its polar equals the ambient dimension:

$$\delta(C) + \delta(C^\circ) = d. \quad (4.5)$$

This generalizes the property $\dim(L) + \dim(L^\perp) = d$ for each subspace $L \subset \mathbb{R}^d$.

(7) **Direct products.** For each cone $K \in \mathcal{C}_d$,

$$\delta(C \times K) = \delta(C) + \delta(K). \quad (4.6)$$

In particular, the statistical dimension is invariant under embedding:

$$\delta(C \times \{\mathbf{0}_{d'}\}) = \delta(C).$$

The relation (4.6) generalizes the rule $\dim(L \times M) = \dim(L) + \dim(M)$ for linear subspaces L and M .

(8) **Monotonicity.** For each cone $K \in \mathcal{C}_d$, the inclusion $C \subset K$ implies that $\delta(C) \leq \delta(K)$.

Proof. The Gaussian formulation in (4.1) simply repeats Definition 2.1. To derive the spherical formulation (4.2) from (4.1), we introduce the spherical decomposition $\mathbf{g} = R\boldsymbol{\theta}$, where $R := \|\mathbf{g}\|$ is a chi random variable with d degrees of freedom that is independent from the spherical variable $\boldsymbol{\theta}$. Use nonnegative homogeneity to draw the term R out of the projection and the squared norm, factor the expectation via independence, and note that $\mathbb{E}[R^2] = d$.

The rotational invariance property (4.3) follows immediately from the fact that a standard normal vector is rotationally invariant.

To compute the statistical dimension of a subspace, note that the Euclidean projection of a standard normal vector onto a subspace has the standard normal distribution supported on that subspace, so its expected squared norm equals the dimension of the subspace.

The polar identity (4.4) is a direct consequence of the distance formula (3.9), which implies that $\|\Pi_C(\mathbf{g})\| = \text{dist}(\mathbf{g}, C^\circ)$. Similarly, the totality law (4.5) follows from the Pythagorean identity (3.8).

We obtain the direct product rule (4.6) from the observation (3.11) that projection splits over a direct product, coupled with the fact that projecting a standard normal vector onto each of two orthogonal subspaces results in two independent standard normal vectors.

Finally, we verify the monotonicity law. Polarity reverses inclusion, so $K^\circ \subset C^\circ$. Using the polarity identity (4.4) twice, we obtain

$$\delta(C) = \mathbb{E}[\text{dist}^2(\mathbf{g}, C^\circ)] \leq \mathbb{E}[\text{dist}^2(\mathbf{g}, K^\circ)] = \delta(K).$$

This completes the recitation. \square

The statistical dimension also enjoys some deeper properties. We reserve these results until Section 5.3, which gives us time to introduce additional tools from integral geometry.

4.2. Self-dual cones. We say that a cone C is *self-dual* when $C^\circ = -C$. Self-dual cones are ubiquitous in the theory and practice of convex optimization. Here are three important examples:

- (1) **The nonnegative orthant.** The cone $\mathbb{R}_+^d := \{\mathbf{x} \in \mathbb{R}^d : x_i \geq 0 \text{ for } i = 1, \dots, d\}$ is self-dual.
- (2) **The second-order cone.** The cone $\mathbb{L}^{d+1} := \{(\mathbf{x}, \tau) \in \mathbb{R}^{d+1} : \|\mathbf{x}\| \leq \tau\}$ is self-dual. This example is sometimes called the *Lorentz cone* or the *ice-cream cone*.
- (3) **Symmetric positive-semidefinite matrices.** The cone $\mathbb{S}_+^{n \times n} := \{\mathbf{X} \in \mathbb{R}_{\text{sym}}^{n \times n} : \mathbf{X} \succcurlyeq \mathbf{0}\}$ is self-dual, where the curly inequality denotes the semidefinite order. Note that the linear space $\mathbb{R}_{\text{sym}}^{n \times n}$ of $n \times n$ symmetric matrices has dimension $\frac{1}{2}n(n+1)$.

For a self-dual cone, the computation of the statistical dimension is particularly simple; cf. [CRPW12, Cor. 3.8]. The first three entries in Table 4.1 follow instantly from this result.

TABLE 4.1: The statistical dimensions of some convex cones.

Cone	Notation	Statistical dimension	Location
The nonnegative orthant	\mathbb{R}_+^d	$\frac{1}{2}d$	Sec. 4.2
The second-order cone	\mathbb{L}^{d+1}	$\frac{1}{2}(d+1)$	Sec. 4.2
Symmetric positive-semidefinite matrices	$\mathbb{S}_+^{n \times n}$	$\frac{1}{4}n(n+1)$	Sec. 4.2
Circular cone in \mathbb{R}^d of angle α	$\text{Circ}_d(\alpha)$	$d \sin^2(\alpha) + O(1)$	Sec. 4.3
Chambers of finite reflection groups acting on \mathbb{R}^d	C_A C_{BC}	$\log(d) + O(1)$ $\frac{1}{2} \log(d) + O(1)$	Sec. 4.7

Proposition 4.2 (Self-dual cones). *Let $C \in \mathcal{C}_d$ be a self-dual cone. The statistical dimension $\delta(C) = \frac{1}{2}d$.*

Proof. Just observe that $\delta(C) = \frac{1}{2}[\delta(C) + \delta(C^\circ)] = \frac{1}{2}d$. The first identity holds because of the self-dual property of the cone and the rotational invariance (4.3) of statistical dimension. The second equality follows from the totality law (4.5). \square

4.3. Circular cones. The circular cone $\text{Circ}_d(\alpha)$ in \mathbb{R}^d with angle $0 \leq \alpha \leq \frac{\pi}{2}$ is defined as

$$\text{Circ}_d(\alpha) := \{\mathbf{x} \in \mathbb{R}^d : x_1 \geq \|\mathbf{x}\| \cos(\alpha)\}.$$

In particular, the cone $\text{Circ}_d(\frac{\pi}{4})$ is isometric to the second-order cone \mathbb{L}^d . Circular cones have numerous applications in optimization; we refer the reader to [BV04, Sec. 4], [BTN01, Sec. 3], and [AG03] for details.

We can obtain an accurate expression for the statistical dimension of a circular cone using trigonometry and basic asymptotic methods.

Proposition 4.3 (Circular cones). *The statistical dimension of a circular cone satisfies*

$$\delta(\text{Circ}_d(\alpha)) = d \sin^2(\alpha) + O(1). \quad (4.7)$$

The error term is approximately equal to $\cos(2\alpha)$. See Figure 4.1[left] for a plot of (4.7).

Turn to Appendix C.1 for the proof, which seems to be original. Even though the formula in Proposition 4.3 is simple, it already gives an accurate approximation in moderate dimensions.

4.4. A recipe for the statistical dimension of a descent cone. Theorems II and III allow us to locate the phase transition in a regularized inverse problem with random data. To apply these results, we must be able to compute the statistical dimension of a descent cone associated with the regularizer. In this section, we describe a method that delivers a superb upper bound for the statistical dimension of a descent cone. This technique is based on an elegant application of polarity. Stojnic [Sto09] developed the basic argument, which Chandrasekaran et al. [CRPW12] subsequently refined. We have undertaken some extra technical work to prove that the upper bound is accurate for the most important examples.

Proposition 4.4 (The statistical dimension of a descent cone). *Let f be a proper convex function. Assume that the subdifferential $\partial f(\mathbf{x})$ is nonempty, compact, and does not contain the origin. Then*

$$\delta(\mathcal{D}(f, \mathbf{x})) \leq \inf_{\tau \geq 0} \mathbb{E} [\text{dist}^2(\mathbf{g}, \tau \cdot \partial f(\mathbf{x}))]. \quad (4.8)$$

The function on the right-hand side of (4.8), namely

$$F: \tau \mapsto \mathbb{E} [\text{dist}^2(\mathbf{g}, \tau \cdot \partial f(\mathbf{x}))] \quad \text{for } \tau \geq 0, \quad (4.9)$$

is strictly convex, continuous at $\tau = 0$, and differentiable for $\tau > 0$. It achieves its minimum at a unique point.

Proof. We use the polarity relation (4.4) to compute the statistical dimension:

$$\delta(\mathcal{D}(f, \mathbf{x})) = \mathbb{E} [\text{dist}^2(\mathbf{g}, \mathcal{D}(f, \mathbf{x})^\circ)] = \mathbb{E} \left[\text{dist}^2 \left(\mathbf{g}, \bigcup_{\tau \geq 0} \tau \cdot \partial f(\mathbf{x}) \right) \right] = \mathbb{E} \inf_{\tau \geq 0} \text{dist}^2(\mathbf{g}, \tau \cdot \partial f(\mathbf{x})).$$

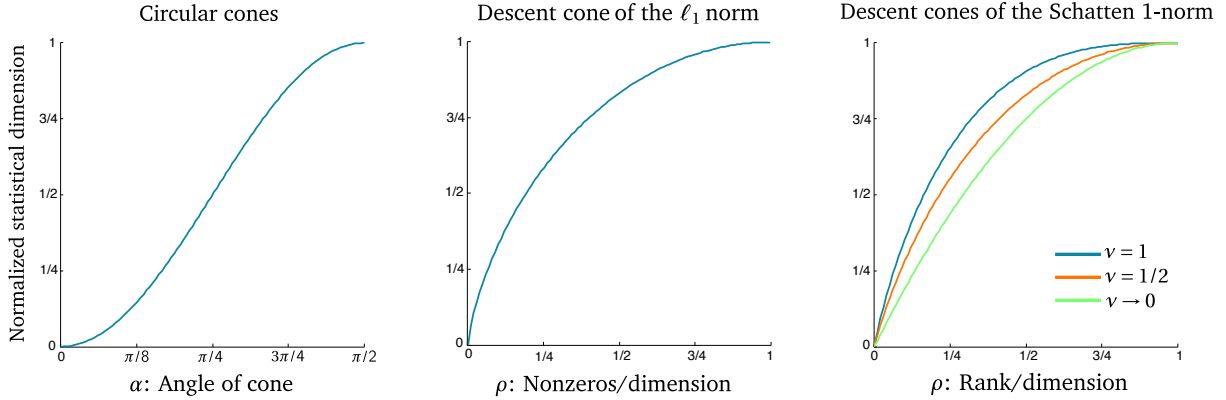


FIGURE 4.1: **Asymptotic statistical dimension computations.** In each panel, we take the dimensional parameters to infinity. **[left] Circular cones.** The plot shows the normalized statistical dimension $\delta(\cdot)/d$ of the circular cone $\text{Circ}_d(\alpha)$. **[center] ℓ_1 descent cones.** The curve traces the normalized statistical dimension $\delta(\cdot)/d$ of the descent cone of the ℓ_1 norm on \mathbb{R}^d at a vector with $[\rho d]$ nonzero entries. **[right] Schatten 1-norm descent cones.** The normalized statistical dimension $\delta(\cdot)/(mn)$ of the descent cone of the S_1 norm on $\mathbb{R}^{m \times n}$ at a matrix with rank $[\rho m]$ for several fixed aspect ratios $\nu = m/n$. As the aspect ratio $\nu \rightarrow 0$, the limiting curve is $\rho \mapsto 2\rho - \rho^2$.

The second identity follows from the fact (3.3) that, under our technical assumptions, the polar of the descent cone is the cone generated by the subdifferential. The third identity holds because the distance to a union is the infimal distance to any one of its members. To reach (4.8), we apply Jensen’s inequality. See Lemma B.2 in Appendix B.1 for the proof of the remaining claims. \square

Proposition 4.4 suggests a method, displayed as Recipe 4.1, for bounding the statistical dimension of a descent cone. In the next two sections, we use this approach to estimate the statistical dimension of the descent cone of the ℓ_1 norm at a sparse vector and the Schatten 1-norm at a low-rank matrix.

The literature contains empirical evidence that Recipe 4.1 leads to very accurate upper bounds. We have obtained an error estimate which shows that the recipe is precise for many important examples.

Theorem 4.5 (Error bound for descent cone recipe). *Let f be a norm on \mathbb{R}^d , and fix a nonzero point \mathbf{x} . Then*

$$\left| \delta(\mathcal{D}(f, \mathbf{x})) - \inf_{\tau \geq 0} F(\tau) \right| \leq \frac{2 \sup \{ \|\mathbf{s}\| : \mathbf{s} \in \partial f(\mathbf{x}) \}}{f(\mathbf{x}/\|\mathbf{x}\|)}, \quad (4.10)$$

where the function F is defined in (4.9).

The proof of Theorem 4.5 is technical in nature, so we defer the details to Appendix B.2. The application of this result requires some care because many different vectors \mathbf{x} can generate the same subdifferential $\partial f(\mathbf{x})$ and hence the same descent cone $\mathcal{D}(f, \mathbf{x})$. From this class of vectors, we ought to select one that maximizes the value $f(\mathbf{x}/\|\mathbf{x}\|)$.

Remark 4.6 (Improved error bounds). We have established several variants of Theorem 4.5. For example, we can study arbitrary convex functions instead of norms if we move to an appropriate asymptotic setting. Unfortunately, we have not identified the optimal form for the error bound (4.10) in the general case. As such, we have chosen to present a simple result that justifies our analysis of phase transitions in ℓ_1 and S_1 minimization problems.

4.5. Descent cones of the ℓ_1 norm. When we wish to solve an inverse problem with a sparse unknown, we often use the ℓ_1 norm as a regularizer; cf. (2.6), (2.10), and (2.11). Our next result summarizes the calculations required to obtain the statistical dimension of the descent cone of the ℓ_1 norm at a sparse vector. When we combine this proposition with Theorems II and III, we obtain the exact location of the phase transition for ℓ_1 regularized inverse problems whose dimension is large.

RECIPE 4.1: The statistical dimension of a descent cone.

Assume that f is a proper convex function on \mathbb{R}^d

Assume that the subdifferential $\partial f(\mathbf{x})$ is nonempty, compact, and does not contain the origin

- (1) Identify the subdifferential $S = \partial f(\mathbf{x})$.
- (2) For each $\tau \geq 0$, compute $F(\tau) = \mathbb{E}[\text{dist}^2(\mathbf{g}, \tau S)]$.
- (3) Find the unique solution, if it exists, to the stationary equation $F'(\tau) = 0$.
- (4) If the stationary equation has a solution τ_* , then $\delta(\mathcal{D}(f, \mathbf{x})) \leq F(\tau_*)$.
- (5) Otherwise, the bound is vacuous: $\delta(\mathcal{D}(f, \mathbf{x})) \leq F(0) = d$.

Proposition 4.7 (Descent cones of the ℓ_1 norm). *Let \mathbf{x} be a vector in \mathbb{R}^d with s nonzero entries. Then the normalized statistical dimension of the descent cone of the ℓ_1 norm at \mathbf{x} satisfies the bounds*

$$\psi(s/d) - \frac{2}{\sqrt{sd}} \leq \frac{\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x}))}{d} \leq \psi(s/d). \quad (4.11)$$

The function $\psi : [0, 1] \rightarrow [0, 1]$ is defined as

$$\psi(\rho) := \inf_{\tau \geq 0} \left\{ \rho(1 + \tau^2) + (1 - \rho) \sqrt{\frac{2}{\pi}} \left[(1 + \tau^2) \int_{\tau}^{\infty} e^{-u^2/2} du - \tau e^{-\tau^2/2} \right] \right\}. \quad (4.12)$$

The infimum in (4.12) is achieved for the unique positive τ that solves the stationary equation

$$\frac{1}{\tau} e^{-\tau^2/2} - \int_{\tau}^{\infty} e^{-u^2/2} du = \sqrt{\frac{\pi}{2}} \cdot \frac{\rho}{1 - \rho}. \quad (4.13)$$

See Figure 4.1[center] for a plot of the function (4.12).

Proposition 4.7 is a direct consequence of Recipe 4.1 and the error bound in Theorem 4.5. See Appendix C.2 for details of the proof; Appendix A.2 explains the numerical aspects.

Let us emphasize the following consequences of Proposition 4.7. When the number s of nonzeros in the vector \mathbf{x} is proportional to the ambient dimension d , the error in the statistical dimension calculation (4.11) is vanishingly small relative to the ambient dimension. When \mathbf{x} is sparser, it is more appropriate to compare the error with the statistical dimension itself. Thus,

$$\left| \frac{\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x})) - d \cdot \psi(s/d)}{\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x}))} \right| \leq \frac{2}{\sqrt{\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x}))}} \quad \text{when } s \geq \sqrt{d} + 1.$$

We have used the observation that $\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x})) \geq s - 1$, which holds because $\mathcal{D}(\|\cdot\|_1, \mathbf{x})$ contains the $(s - 1)$ -dimensional subspace parallel with the face of the ℓ_1 ball containing \mathbf{x} .

Aside from the first inequality in (4.11), the calculations and the resulting formulae in Proposition 4.7 are not substantially novel. Most of the existing analysis concerns the phase transition in compressed sensing, i.e., the ℓ_1 minimization problem (2.6) with Gaussian measurements. In this setting, Donoho [Don06b] and Donoho & Tanner [DT09a] obtained an asymptotic upper bound, equivalent to the upper bound in (4.11), from polytope angle calculations. Stojnic [Sto09] established the same asymptotic upper bound using a precursor of Recipe 4.1; see also Chandrasekaran et al. [CRPW12, App. C]. In addition, there are some heuristic arguments, based on ideas from statistical physics, that lead to the same result, cf. [DMM09a] and [DMM09b, Sec. 17]. Very recently, Bayati et al. [BLM12] have shown that, in the asymptotic setting, the compressed sensing problem undergoes a phase transition at the location predicted by (4.11).

4.6. Descent cones of the Schatten 1-norm. When we wish to solve an inverse problem whose unknown is a low-rank matrix, we often use the Schatten 1-norm as a regularizer, as in (2.7) and (2.11). The following result gives a sharp asymptotic expression for the statistical dimension of the descent cone of the S_1 norm at a low-rank matrix. Together with Theorems II and III, this proposition allows us to identify the exact location of the phase transition for S_1 regularized inverse problems whose dimension is large.

Proposition 4.8 (Descent cones of the S_1 norm). *Consider a sequence $\{\mathbf{X}(r, m, n)\}$ of matrices where $\mathbf{X}(r, m, n)$ has rank r and dimension $m \times n$ with $m \leq n$. Suppose that $r, m, n \rightarrow \infty$ with limiting ratios $r/m \rightarrow \rho \in (0, 1)$ and $m/n \rightarrow \nu \in (0, 1]$. Then*

$$\frac{\delta(\mathcal{D}(\|\cdot\|_{S_1}, \mathbf{X}(r, m, n)))}{mn} \rightarrow \psi(\rho, \nu). \quad (4.14)$$

The function $\psi : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is defined as

$$\psi(\rho, \nu) := \inf_{\tau \geq 0} \left\{ \rho\nu + (1 - \rho\nu) \left[\rho(1 + \tau^2) + (1 - \rho) \int_{a_- \vee \tau}^{a_+} (u - \tau)^2 \cdot \varphi_y(u) du \right] \right\}. \quad (4.15)$$

The quantity $y := (\nu - \rho\nu)/(1 - \rho\nu)$, and the limits of the integral are $a_{\pm} := 1 \pm \sqrt{y}$. The integral kernel φ_y is a probability density supported on $[a_-, a_+]$:

$$\varphi_y(u) := \frac{1}{\pi y u} \sqrt{(u^2 - a_-^2)(a_+^2 - u^2)} \quad \text{for } u \in [a_-, a_+].$$

The optimal value of τ in (4.15) satisfies the stationary equation

$$\int_{a_- \vee \tau}^{a_+} \left(\frac{u}{\tau} - 1 \right) \cdot \varphi_y(u) du = \frac{\rho}{1 - \rho}. \quad (4.16)$$

See Figure 4.1[right] for a visualization of the curve (4.15) as function of ρ for several choices of ν . The operator \vee returns the maximum of two numbers.

See Appendix C.3 for a proof of Proposition 4.8. Appendix A.2 contains details of the numerical calculation.

The literature contains several papers that, in effect, contain loose upper bounds for the statistical dimension of the descent cones of the Schatten 1-norm [RXH11, OKH11]. We single out the work [OH10] of Oymak & Hassibi, which identifies an empirically sharp upper bound via a laborious argument. The approach here is more in the spirit of the upper bound in [CRPW12, App. C], but we have taken extra care to obtain the asymptotically correct estimate.

4.7. Normal cones to permutahedra. We close this section with a more sophisticated example. The (signed) permutahedron generated by a vector $\mathbf{x} \in \mathbb{R}^d$ is the convex hull of all (signed) coordinated permutations of the vector:

$$\mathcal{P}(\mathbf{x}) := \text{conv} \{ \sigma(\mathbf{x}) : \sigma \text{ a coordinate permutation} \} \quad (4.17)$$

$$\mathcal{P}_{\pm}(\mathbf{x}) := \text{conv} \{ \sigma_{\pm}(\mathbf{x}) : \sigma_{\pm} \text{ a signed coordinate permutation} \}. \quad (4.18)$$

(A signed permutation permutes the coordinates of a vector and gives each one an arbitrary sign.) Figure 4.2 displays two signed permutahedra and the normal cone at a vertex.

In this section, we present an exact formula for the statistical dimension of the normal cone of a permutahedron. In Section 9, we use this calculation to study an application in signal processing that was proposed in [CRPW12, p. 812].

Proposition 4.9 (Normal cones of permutahedra). *Suppose that \mathbf{x} has distinct entries. The statistical dimension of the normal cone at a vertex of the (signed) permutahedron generated by \mathbf{x} satisfies*

$$\delta(\mathcal{N}(\mathcal{P}(\mathbf{x}), \mathbf{x})) = H_d \quad \text{and} \quad \delta(\mathcal{N}(\mathcal{P}_{\pm}(\mathbf{x}), \mathbf{x})) = \frac{1}{2} H_d,$$

where $H_d := \sum_{i=1}^d i^{-1}$ is the d th harmonic number.

The proof of Proposition 4.9 appears in Appendix C.4. The argument illustrates some deep connections between conic geometry and classical combinatorics.

5. TOOLS FROM CONIC INTEGRAL GEOMETRY

To prove that the statistical dimension controls the location of phase transitions in random convex optimization problems, we rely on methods from *conic integral geometry*, the field of mathematics concerned with geometric properties of convex cones that remain invariant under rotations and reflections. Here are some of the guiding questions in this area:

- What is the probability that a random unit vector lies at most a specified distance from a fixed cone?
- What is the probability that a randomly rotated cone shares a ray with a fixed cone?

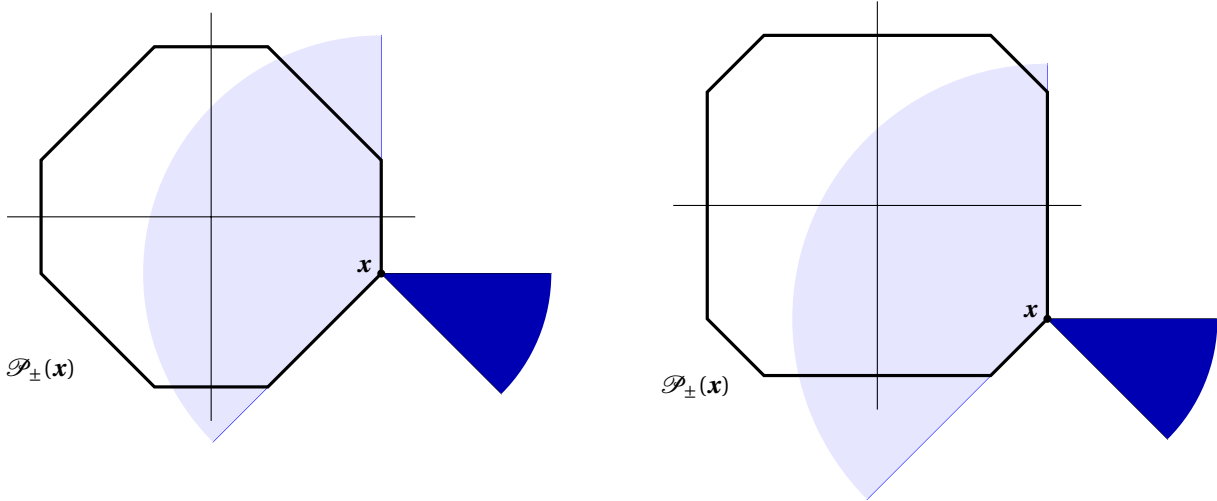


FIGURE 4.2: **Normal cone at the vertex of a permutahedron.** The signed permutahedron $\mathcal{P}_{\pm}(\mathbf{x})$ generated by [left] the vector $\mathbf{x} = (3, -1)$ and [right] the vector $\mathbf{x} = (3, -2)$. In each panel, the darker cone is the normal cone $\mathcal{N}(\mathcal{P}_{\pm}(\mathbf{x}), \mathbf{x})$, and the lighter cone is its polar. Note that the normal cone does not depend on the generator \mathbf{x} provided that the entries of \mathbf{x} are distinct.

The theory of conic integral geometry offers beautiful and precise answers, phrased in terms of a set of geometric invariants called *conic intrinsic volumes*.

In the next section, we introduce the intrinsic volumes of a cone, we compute the intrinsic volumes of some basic cones, and we state the key facts about intrinsic volumes. Section 5.2 contains more advanced formulas from conic integral geometry, which are essential tools for identifying phase transitions. We revisit the statistical dimension in Section 5.3, where we develop some elegant new characterizations.

The material in this section is adapted from the book [SW08, Sec. 6.5] and the dissertation [Ame11]. The foundational research in this area is due to Santaló [San76, Part IV]. Modern treatments depend on the work of Glasauer [Gla95, Gla96]. In these sources, the theory is presented in terms of spherical geometry, rather than in terms of conical geometry. As noted in [AB12], the two approaches are equivalent, but the conic viewpoint provides simpler formulas and has key benefits that are only revealed through deeper structural investigations.

5.1. Conic intrinsic volumes. We begin with the definition of the intrinsic volumes of a convex cone.

Definition 5.1 (Intrinsic volumes). Let $C \in \mathcal{C}_d$ be a polyhedral cone. For each $k = 0, 1, 2, \dots, d$, the k th (conic) intrinsic volume $v_k(C)$ is given by

$$v_k(C) := \mathbb{P}\{\Pi_C(\mathbf{g}) \text{ lies in the relative interior of a } k\text{-dimensional face of } C\}.$$

As usual, \mathbf{g} is a standard normal vector in \mathbb{R}^d . We extend this definition to a general closed convex cone in \mathcal{C}_d by approximating it with a sequence of polyhedral cones.

For polyhedral cones, Definition 5.1 gives an attractive intuition. We can decompose the ambient space into $d + 1$ disjoint regions, where the k th region contains the points whose projection onto the cone lies in the relative interior of a k -dimensional face; the k th intrinsic volume reflects the proportion of the ambient space comprised by the k th region. We see immediately that the sequence of intrinsic volumes forms a probability distribution on $\{0, 1, 2, \dots, d\}$. The definition also delivers insight about several fundamental examples.

Example 5.2 (Linear subspaces). Let $L_j \subset \mathbb{R}^d$ be a j -dimensional subspace. Then L_j is a polyhedral cone with precisely one face, so the map Π_{L_j} projects every point onto this j -dimensional face. Thus,

$$v_k(L_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \quad \text{for } k = 0, 1, 2, \dots, d.$$

Example 5.3 (The nonnegative orthant). The nonnegative orthant \mathbb{R}_+^d is a polyhedral cone, and Table 4.1 lists its statistical dimension as $\delta(\mathbb{R}_+^d) = \frac{1}{2}d$. The projection $\Pi_{\mathbb{R}_+^d}(\mathbf{g})$ lies in the relative interior of a k -dimensional face of the orthant if and only if exactly k coordinates of \mathbf{g} are positive. Each coordinate of \mathbf{g} is positive with probability one-half and negative with probability one-half, and the coordinates are independent. Therefore, the intrinsic volumes of the orthant are given by

$$v_k(\mathbb{R}_+^d) = 2^{-d} \binom{d}{k} \quad \text{for } k = 0, 1, 2, \dots, d.$$

In other words, the intrinsic volumes coincide with the probability density of a $\text{BINOMIAL}(d, \frac{1}{2})$ random variable. From this representation, we learn that the largest intrinsic volume of the orthant occurs at the index $k = \lfloor \frac{1}{2}d \rfloor$, and the large intrinsic volumes concentrate sharply around this index. We will prove that this type of behavior is generic!

For a nonpolyhedral cone $C \in \mathcal{C}_d$, the projection formula in Definition 5.1 breaks down, and we can no longer interpret the k th intrinsic volume in terms of the k -dimensional faces of C . Furthermore, it takes some work to give sense to the phrase ‘‘approximation by polyhedral cones.’’ We refer to the book [SW08, Sec. 6.5] or the thesis [Ame11] for these important details.

In spite of these caveats, the intrinsic volumes of an arbitrary closed convex cone still form a probability distribution. The next result quantifies this statement and several other basic relationships.

Fact 5.4 (Properties of intrinsic volumes). *Let $C \in \mathcal{C}_d$ be a closed convex cone. The intrinsic volumes of the cone obey the following laws.*

(1) **Distribution.** *The intrinsic volumes describe a probability distribution on $\{0, 1, \dots, d\}$:*

$$\sum_{k=0}^d v_k(C) = 1 \quad \text{and} \quad v_k(C) \geq 0 \quad \text{for } k = 0, 1, 2, \dots, d. \quad (5.1)$$

(2) **Polarity.** *The intrinsic volumes reverse under polarity:*

$$v_k(C) = v_{d-k}(C^\circ) \quad \text{for } k = 0, 1, 2, \dots, d. \quad (5.2)$$

(3) **Gauss–Bonnet formula.** *When C is not a subspace,*

$$\sum_{\substack{k=0 \\ k \text{ even}}}^d v_k(C) = \sum_{\substack{k=1 \\ k \text{ odd}}}^d v_k(C) = \frac{1}{2}. \quad (5.3)$$

(4) **Direct products.** *For each closed convex cone $K \in \mathcal{C}_{d'}$,*

$$v_k(C \times K) = \sum_{i+j=k} v_i(C) \cdot v_j(K) \quad \text{for } k = 0, 1, 2, \dots, d + d'. \quad (5.4)$$

The facts (5.1), (5.2), and (5.3) are drawn from [SW08, Sec. 6.5]. The product rule (5.4) appears in [Ame11, Prop. 4.4.13].

Partial sums of intrinsic volumes play a central role in the kinematic theory of convex cones, described below. With an eye toward these developments, we make the following definition.

Definition 5.5 (Tail functionals). *Let $C \in \mathcal{C}_d$ be a closed convex cone. For each $k = 0, 1, 2, \dots, d$, the k th tail functional is given by*

$$t_k(C) := v_k(C) + v_{k+1}(C) + \dots = \sum_{j=k}^d v_j(C). \quad (5.5)$$

The k th half-tail functional is defined as

$$h_k(C) := v_k(C) + v_{k+2}(C) + \dots = \sum_{\substack{j=k \\ j-k \text{ even}}}^d v_j(C). \quad (5.6)$$

We require an interlacing inequality for tail functionals.

Proposition 5.6 (Interlacing). *For each closed convex cone $C \in \mathcal{C}_d$ that is not a linear subspace,*

$$2h_k(C) \geq t_k(C) \geq 2h_{k+1}(C) \quad \text{for each } k = 0, 1, 2, \dots, d-1.$$

We establish Proposition 5.6 in Appendix D.

5.2. The formulas of conic geometry. We continue with a selection of more sophisticated results from conic integral geometry. These formulas provide detailed answers, expressed in terms of conic intrinsic volumes, to the geometric questions posed at the beginning of Section 5. For this discussion, we introduce a family of geometric functions that also plays a key role in our analysis.

Definition 5.7 (Tropic functions). Let L_k be a k -dimensional subspace of \mathbb{R}^d . Define

$$I_k^d(\varepsilon) := \mathbb{P}\{\|\Pi_{L_k}(\boldsymbol{\theta})\|^2 \geq \varepsilon\} \quad \text{for } \varepsilon \in [0, 1], \quad (5.7)$$

where $\boldsymbol{\theta}$ is uniformly distributed on the unit sphere in \mathbb{R}^d .

Basic geometric reasoning reveals that $I_k^d(\varepsilon)$ is the proportion of points on the sphere S^{d-1} that lie within an angle $\arccos(\sqrt{\varepsilon})$ of the subspace L_k . Our terminology derives from the approximate geographical fact that the tropics lie within a fixed angle ($23^\circ 26'$) of the equator; the usual term *regularized incomplete beta function* is longer and less evocative.

The core fact in conic integral geometry is the *spherical Steiner formula* [Her43, All48, San50], which describes the fraction of points on the sphere that lie at most a fixed angle from a closed convex cone.

Fact 5.8 (Spherical Steiner formula). *Let $C \in \mathcal{C}_d$ be a closed convex cone. For each $\varepsilon \in [0, 1]$,*

$$\mathbb{P}\{\|\Pi_C(\boldsymbol{\theta})\|^2 \geq \varepsilon\} = \sum_{k=0}^d v_k(C) I_k^d(\varepsilon). \quad (5.8)$$

The spherical Steiner formula often serves as the *definition* of conic intrinsic volumes; it can also be derived from the definition here. For a modern proof of Fact 5.8 in the spirit of this work, see [SW08, Thm 6.5.1].

The second essential result is the *conic kinematic formula*, which provides an exact expression for the probability that a randomly oriented cone strikes a fixed cone.

Fact 5.9 (Conic kinematic formula). *Let $C, K \in \mathcal{C}_d$ be closed convex cones, and assume that C is not a subspace. Then*

$$\mathbb{P}\{C \cap \mathbf{Q}K \neq \{\mathbf{0}\}\} = 2h_{d+1}(C \times K). \quad (5.9)$$

For a linear subspace $L_{d-m} \subset \mathbb{R}^d$ with dimension $d - m$, this expression reduces to the Crofton formula

$$\mathbb{P}\{C \cap \mathbf{Q}L_{d-m} \neq \{\mathbf{0}\}\} = 2h_{m+1}(C). \quad (5.10)$$

See [SW08, p. 261] for a proof of Fact 5.9. In Section 7, we use deep properties of the intrinsic volumes to produce an approximate version of the conic kinematic formula, which ultimately delivers detailed information about phase transitions.

Remark 5.10 (Extended kinematic formula). By induction, the kinematic formula generalizes to a family $C, K_1, \dots, K_r \in \mathcal{C}_d$ of closed convex cones where C is not a subspace:

$$\mathbb{P}\{C \cap \mathbf{Q}_1 K_1 \cap \dots \cap \mathbf{Q}_r K_r \neq \{\mathbf{0}\}\} = 2h_{rd+1}(C \times K_1 \times \dots \times K_r). \quad (5.11)$$

Each matrix \mathbf{Q}_i is an independent, random orthogonal basis. This result can be used to analyze demixing problems with more than two constituents.

5.3. Revisiting the statistical dimension. The machinery of conic integral geometry offers several new insights about the statistical dimension, and it underscores the analogy with the dimension of a linear subspace. The next proposition demonstrates that the statistical dimension of a cone is the mean of the random variable on $\{0, 1, 2, \dots, d\}$ whose distribution is given by the intrinsic volumes of the cone. This fact motivated us to select the terminology “statistical” dimension.

Proposition 5.11 (Statistical dimension as mean intrinsic volume). *For each closed convex cone $C \in \mathcal{C}_d$,*

$$\delta(C) = \sum_{k=1}^d k v_k(C). \quad (5.12)$$

Proof. The spherical formulation (4.2) of statistical dimension shows that

$$\delta(C) = d \mathbb{E} [\|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2] = d \int_0^1 \mathbb{P}\{\|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 \geq \varepsilon\} d\varepsilon.$$

We have used integration by parts to express the expectation as an integral of tail probabilities. The Steiner formula (5.8) and the definition (5.7) of the tropic function allow us to write the probability as a sum:

$$\delta(C) = d \sum_{k=1}^d \nu_k(C) \int_0^1 \mathbb{P}\{\|\mathbf{\Pi}_{L_k}(\boldsymbol{\theta})\|^2 \geq \varepsilon\} d\varepsilon = \sum_{k=1}^d \nu_k(C) (d \mathbb{E} [\|\mathbf{\Pi}_{L_k}(\boldsymbol{\theta})\|^2]),$$

where L_k is an arbitrary k -dimensional subspace. A second application of (4.2) shows that the parenthesis is the statistical dimension of L_k , and Proposition 4.1(4) provides that $\delta(L_k) = k$. \square

Proposition 5.11 has a significant consequence. Each intrinsic volume ν_k is a *valuation* on \mathcal{C}_d , so the statistical dimension is also a valuation on \mathcal{C}_d . Equivalently, (i) the statistical dimension of the trivial cone $\delta(\{\mathbf{0}\}) = 0$, and (ii) if $C, K \in \mathcal{C}_d$ and $C \cup K \in \mathcal{C}_d$, then we have the inclusion–exclusion rule

$$\delta(C \cup K) = \delta(C) + \delta(K) - \delta(C \cap K).$$

This property is analogous with the inclusion–exclusion law for the dimension of a subspace.

The long-standing *spherical Hadwiger conjecture* posits that each continuous, rotation-invariant valuation on \mathcal{C}_d can be written as a linear combination of the conic intrinsic volumes. If this conjecture holds, then every such valuation is determined by its values on linear subspaces. Under this surmise, we obtain a fundamental characterization of the statistical dimension.

Proposition 5.12 (Statistical dimension is canonical). *If the spherical Hadwiger conjecture holds, then the statistical dimension δ is the unique continuous, rotation-invariant valuation on \mathcal{C}_d that satisfies $\delta(L) = \dim(L)$ for each subspace $L \subset \mathbb{R}^d$.*

A weaker version of the spherical Hadwiger conjecture does hold for geometric parameters known as *curvature measures*, which provides a rigorous claim that the statistical dimension is canonical under an additional technical assumption of “localizability”; see [SW08, p. 254 and Thm. 6.5.4]. For a discussion of the spherical Hadwiger conjecture, see the works [McM93, p. 976], [KR97, Sec. 11.5], and [SW08, p. 263]. The conjecture currently stands open for $d \geq 4$.

6. INTRINSIC VOLUMES CONCENTRATE NEAR THE STATISTICAL DIMENSION

The main technical result in this paper describes a deep new property of conic intrinsic volumes. The intrinsic volumes of a cone concentrate near the statistical dimension of the cone on a scale determined by the statistical dimension.

Theorem 6.1 (Concentration of intrinsic volumes). *Let C be a closed convex cone. Define the transition width*

$$\omega(C) := \sqrt{\delta(C) \wedge \delta(C^\circ)},$$

and introduce the function

$$p_C(\lambda) := 4 \exp\left(\frac{-\lambda^2/8}{\omega^2(C) + \lambda}\right) \quad \text{for } \lambda \geq 0. \quad (6.1)$$

Then

$$k_- \leq \delta(C) - \lambda + 1 \quad \implies \quad t_{k_-}(C) \geq 1 - p_C(\lambda); \quad (6.2)$$

$$k_+ \geq \delta(C) + \lambda \quad \implies \quad t_{k_+}(C) \leq p_C(\lambda). \quad (6.3)$$

The tail functional t_k is defined in (5.5). The operator \wedge returns the minimum of two numbers.

In other words, the sequence $\{t_k(C) : k = 0, 1, 2, \dots, d\}$ of tail functionals drops from one to zero near the statistical dimension $\delta(C)$, and the transition occurs over a range of $O(\omega(C))$ indices. Owing to the fact (5.1) that the intrinsic volumes form a probability distribution, we must conclude that the intrinsic volumes $\nu_k(C)$ are all negligible in size, except for those whose index k is close to the statistical dimension $\delta(C)$. We learn that the intrinsic volumes of a convex cone C with statistical dimension $\delta(C)$ are qualitatively similar to the intrinsic volumes of a subspace with dimension $\lfloor \delta(C) \rfloor$.

Theorem 6.1 contains additional information about the rate at which the tail functionals of a cone transit from one to zero. To extract this information, it helps to note the weaker inequality

$$p_C(\lambda) \leq \begin{cases} 4e^{-\lambda^2/(16\omega^2(C))}, & 0 \leq \lambda \leq \omega^2(C) \\ 4e^{-\lambda/16}, & \lambda \geq \omega^2(C). \end{cases} \quad (6.4)$$

We see that (6.2) and (6.3) are vacuous until $\lambda \approx 4\omega(C)$. As λ increases, the function $p_C(\lambda)$ decays like the tail of a Gaussian random variable with standard deviation $\leq 2\sqrt{2}\omega(C)$. When λ reaches $\omega^2(C)$, the decay slows to match the tail of an exponential random variable with mean ≤ 16 . In particular, the behavior of the tail functionals depends on the intrinsic properties of the cone, rather than the ambient dimension.

In Section 6.1, we outline some connections between Theorem 6.1 and classical inequalities from Euclidean integral geometry. In Section 6.2, we argue that the result is nearly optimal. Afterward, in Section 6.3 and 6.4, we summarize the intuition behind the proof of Theorem 6.1, and we follow up with the technical details. Later, Sections 7–9 highlight applications in conic geometry, optimization theory, and signal processing.

6.1. Parallels with Euclidean integral geometry. Motivated by analogies between conic and Euclidean integral geometry, Schneider & Weil [SW08, p. 263] ask what relationships hold among the conic intrinsic volumes v_k of a convex cone. To the best of our knowledge, the literature contains only two nontrivial results [GHS02]: Among all cones in \mathcal{C}_d with a fixed value of v_d , a circular cone minimizes v_{d-1} and maximizes v_0 . This fact is a consequence of spherical isoperimetry.

Our work provides a rich family of inequalities relating conic intrinsic volumes and tail functionals. For any convex cone C , Theorem 6.1 implies that

$$v_k(C) \leq s_k(C) := \exp\left(\frac{-(k - \delta(C))^2}{\omega^2(C) + |k - \delta(C)|}\right) \quad \text{for each } k = 0, 1, 2, \dots, d.$$

This bound also relies on the trivial inequalities $v_k(C) \leq t_k(C)$ and $v_k(C) \leq 1 - t_{k+1}(C)$. Observe that the sequence $\{s_k(C) : k = 0, 1, 2, \dots, d\}$ of upper bounds is strictly log-concave:

$$s_k^2(C) > s_{k-1}(C) \cdot s_{k+1}(C) \quad \text{for } k = 1, 2, 3, \dots, d-1.$$

This point follows from the fact that $u \mapsto u^2/(\omega^2(C) + |u|)$ is strictly convex for every nontrivial convex cone C . (Indeed, the Gauss–Bonnet formula (5.3) implies that $\frac{1}{2} \leq \delta(C) \leq d - \frac{1}{2}$, which ensures that $\omega^2(C) \geq \frac{1}{2}$.) We see that the sequence of conic intrinsic volumes is dominated by a log-concave sequence; cf. the conjecture in [Ame11, Conj. 4.4.16].

The Euclidean intrinsic volumes of a convex body also form a log-concave sequence, which is a corollary of the Alexandrov–Fenchel inequalities. Log-concavity of the Euclidean intrinsic volumes has many fundamental consequences, including the usual isoperimetric inequality for convex bodies, the Brunn–Minkowski inequality, and the Urysohn inequality [Sch93, Chap. 6]. Theorem 6.1 delivers a system of inequalities for conic intrinsic volumes that parallel these deep classical results.

6.2. Optimality of Theorem 6.1. By considering circular cones, we discover that Theorem 6.1 is nearly optimal. Let us proceed with a heuristic discussion that conveys the key ideas. Consider the cone $C = \text{Circ}_d(\alpha)$. For simplicity, we assume that $d = 2(n+1)$ for a large integer n , and we abbreviate $q = \sin^2(\alpha)$. Proposition 4.3 shows that the statistical dimension $\delta(C) \approx d \sin^2(\alpha) \approx 2nq$ and $\delta(C^\circ) \approx 2n(1-q)$. According to [Ame11, Ex. 4.4.8], the odd intrinsic volumes of C satisfy

$$v_{2k+1}(C) = \frac{1}{2} \binom{n}{k} q^k (1-q)^{n-k} \quad \text{for } k = 0, 1, 2, \dots, n.$$

In other words, $2v_{2k+1}(C) = \mathbb{P}\{X = k\}$ where $X \sim \text{BINOMIAL}(n, q)$. This observation invites us to study these cones using probabilistic methods.

First, we approximate the width of the region over which the tail functionals of C change from one to zero. The interlacing result, Proposition 5.6, indicates that

$$t_{2k}(C) \approx 2h_{2k+1}(C) = 2 \sum_{i=k}^n v_{2i+1}(C) = \mathbb{P}\{X \geq k\}.$$

Under appropriate assumptions on q , k , and n , the Central Limit Theorem implies

$$\mathbb{P}\{X \geq k\} = \mathbb{P}\left\{\frac{X - nq}{\sqrt{nq(1-q)}} \geq \frac{k - nq}{\sqrt{nq(1-q)}}\right\} \approx \mathbb{P}\left\{Z \geq \frac{k - nq}{\sqrt{nq(1-q)}}\right\}$$

where Z is a standard normal variable. For the index $2k = 2nq + \lambda \approx \delta(C) + \lambda$, we see that

$$t_{2k}(C) \approx \mathbb{P}\left\{Z \geq \frac{\lambda}{2\sqrt{nq(1-q)}}\right\} \approx \exp\left(\frac{-\lambda^2/4}{2nq(1-q)}\right) \approx \exp\left(\frac{-\lambda^2/4}{((1-q)\delta(C)) \wedge (q\delta(C^\circ))}\right) \quad \text{for } \lambda \gg 0.$$

The second approximation follows from the normal tail estimate $\mathbb{P}\{Z \geq z\} \approx e^{-z^2/2}$. Similarly, for the index $2k = 2nq - \lambda \approx \delta(C) - \lambda$, we have

$$t_{2k}(C) \approx 1 - \exp\left(\frac{-\lambda^2/4}{((1-q)\delta(C)) \wedge (q\delta(C^\circ))}\right) \quad \text{for } \lambda \gg 0.$$

In other words, the width of the transition region for the tail functionals of the circular cone C really does have order $\omega(C)$. Furthermore, considering the case where q is close to zero, we discover that the constant in the exponent of (6.1) lies within a factor two of the best possible.

To argue that p_C cannot have subgaussian decay when λ is large relative to the statistical dimension $\delta(C)$, we consider a Poisson limit of the binomial variable. Suppose that $q = b/n$ for a constant b , so the statistical dimension $\delta(C) \approx 2b$. For $2k = 2b(1 + \lambda) \approx (1 + \lambda)\delta(C)$,

$$t_{2k}(C) \approx \mathbb{P}\{X \geq k\} = \mathbb{P}\{X \geq b(1 + \lambda)\} \approx e^{\lambda - (1 + \lambda)\log(1 + \lambda)}.$$

The tail estimate follows, for example, by applying Cramér's Theorem [DZ10, Thm. 2.2.3] to the binomial random variable X . In other words, for a very small circular cone, the tail functionals decay only a little faster than subexponential when the tail index k is a multiple of the statistical dimension.

6.3. Heuristic proof of Theorem 6.1. The basic ideas behind the argument are easy to summarize, but the details demand some effort. Let $C \in \mathcal{C}_d$ be a closed convex cone. Recall the Steiner formula (5.8):

$$\mathbb{P}\{\|\Pi_C(\boldsymbol{\theta})\|^2 \geq \varepsilon\} = \sum_{k=0}^d v_k(C) I_k^d(\varepsilon), \quad (6.5)$$

where $\boldsymbol{\theta}$ is uniformly distributed on the sphere S^{d-1} and the tropic function I_k^d is defined in (5.7).

Concentration of measure on the sphere implies that the random variable $\|\Pi_C(\boldsymbol{\theta})\|^2$ is typically very close to its expected value $\delta(C)/d$, determined by (4.2). Thus, the left-hand side of (6.5) is very close to one when $\varepsilon d < \delta(C)$ and very close to zero when $\varepsilon d > \delta(C)$.

As for the right-hand side of (6.5), recall that the tropic function $I_k^d(\varepsilon)$ is the proportion of points on the sphere within a distance of $\sqrt{1 - \varepsilon}$ from a fixed k -dimensional subspace. Once again, concentration of measure ensures that $I_k^d(\varepsilon)$ is close to zero when $k < \varepsilon d$ and close to one when $k > \varepsilon d$. Therefore, the sum on the right-hand side of (6.5) is approximately equal to the tail functional $t_{\varepsilon d}(C)$.

Combining these two observations, we conclude that the sequence $\{t_k(C) : k = 0, 1, 2, \dots, d\}$ of tail functionals makes a sharp transition from one to zero when $k \approx \delta(C)$. It remains to make this reasoning rigorous and to determine the range of k over which the transition takes place.

6.4. Proof of Theorem 6.1. Let $C \in \mathcal{C}_d$ be a closed convex cone, and define $\varepsilon := k_+/d$. The first part of the argument requires a technical lemma that we prove in Appendix D. This result quantifies how much of the sphere in \mathbb{R}^d lies within an angle $\arccos(k/d)$ of a k -dimensional subspace.

Lemma 6.2 (The tropics). *For all integers $0 \leq k \leq d$, the tropic function $I_k^d(k/d) \geq 0.3$.*

We begin by expressing the tail functional $t_{k_+}(C)$ in terms of the probability that a spherical variable lies near the cone C .

$$\begin{aligned} t_{k_+}(C) &= \sum_{k=k_+}^d v_k(C) [I_k^d(\varepsilon) + (1 - I_k^d(\varepsilon))] \\ &\leq \sum_{k=0}^d v_k(C) I_k^d(\varepsilon) + (1 - I_{k_+}^d(\varepsilon)) \sum_{k=k_+}^d v_k(C) \\ &\leq \mathbb{P}\{\|\Pi_C(\boldsymbol{\theta})\|^2 \geq \varepsilon\} + 0.7 t_{k_+}(C). \end{aligned} \quad (6.6)$$

The first identity is the definition (5.5) of the tail function. To reach the second inequality, we inspect the definition (5.7) to see that $I_k^d(\varepsilon)$ is a decreasing function of k when the other parameters are fixed. In the third line, we invoke the Steiner formula (5.8) to rewrite the first sum. The bound for the second sum follows and the definition (5.5) of the tail functional and from Lemma 6.2 seeing that $\varepsilon = k_+/d$.

Rearranging (6.6), we obtain the bound

$$t_{k_+}(C) \leq 4\mathbb{P}\{d \|\Pi_C(\boldsymbol{\theta})\|^2 \geq d\varepsilon\} \leq 4\mathbb{P}\{d \|\Pi_C(\boldsymbol{\theta})\|^2 \geq \delta(C) + \lambda\}. \quad (6.7)$$

The last inequality depends on the fact that $\varepsilon = k_+/d$ and the definition (6.3) of k_+ . In other words, the tail functional is dominated by the probability that a random point on the sphere is close to the cone.

To estimate the probability in (6.7), we need a tail bound for the squared norm of the projection of a spherical variable onto a cone. This result is encapsulated in the following lemma. The approach is more or less standard, so we defer the details to Appendix D.

Lemma 6.3 (Tail bound for conic projections). *For each closed convex cone $C \in \mathcal{C}_d$,*

$$\mathbb{P}\{d \|\Pi_C(\boldsymbol{\theta})\|^2 \geq \delta(C) + \lambda\} \leq \exp\left(\frac{-\lambda^2/8}{\omega^2(C) + \lambda}\right) \quad \text{for } \lambda \geq 0. \quad (6.8)$$

Introducing (6.8) into (6.7), we reach the upper bound (6.3) on the tail functional.

To develop the lower bound (6.2) on the tail functional $t_{k_-}(C)$, we use a polarity argument. Note that

$$t_{k_-}(C) = \sum_{k=k_-}^d v_k(C) = \sum_{k=0}^{d-k_-} v_k(C^\circ) = 1 - t_{d-k_-+1}(C^\circ). \quad (6.9)$$

The first identity is the definition (5.5) of the tail functional $t_{k_-}(C)$. The second relation holds because of the fact (5.2) that polarity reverses intrinsic volumes, and the last part relies on (5.5) and the property (5.1) that the intrinsic volumes sum to one. Owing to the totality law (4.5) and the definition (6.2) of k_- ,

$$d - k_- + 1 = \delta(C^\circ) + \delta(C) - k_- + 1 \geq \delta(C^\circ) + \lambda.$$

Therefore, we may apply (6.3) to obtain an upper bound on the tail functional $t_{d-k_-+1}(C^\circ)$. Substitute this bound into (6.9) to establish the lower bound on the tail functional $t_{k_-}(C)$ stated in (6.2).

7. APPROXIMATE KINEMATIC BOUNDS

We are now prepared to establish an approximate version of the conic kinematic formula, expressed in terms of the statistical dimension. Most of the applied theorems in this paper ultimately depend on this result. The proof combines the exact kinematic formula (5.9) with the concentration of intrinsic volumes, guaranteed by Theorem 6.1.

Theorem 7.1 (Approximate kinematics). *Assume that $\lambda \geq 0$. Let $C \subset \mathbb{R}^d$ be a convex cone that is not a subspace. For a $(d-m)$ -dimensional subspace L_{d-m} , it holds that*

$$\begin{aligned} m \geq \delta(C) + \lambda &\implies \mathbb{P}\{C \cap \mathbf{Q}L_{d-m} \neq \{\mathbf{0}\}\} \leq p_C(\lambda); \\ m \leq \delta(C) - \lambda &\implies \mathbb{P}\{C \cap \mathbf{Q}L_{d-m} \neq \{\mathbf{0}\}\} \geq 1 - p_C(\lambda). \end{aligned} \quad (7.1)$$

For an arbitrary convex cone $K \subset \mathbb{R}^d$, it holds that

$$\begin{aligned} \delta(C) + \delta(K) \leq d - 2\lambda &\implies \mathbb{P}\{C \cap \mathbf{Q}K \neq \{\mathbf{0}\}\} \leq p_C(\lambda) + p_K(\lambda); \\ \delta(C) + \delta(K) \geq d + 2\lambda &\implies \mathbb{P}\{C \cap \mathbf{Q}K \neq \{\mathbf{0}\}\} \geq 1 - (p_C(\lambda) + p_K(\lambda)). \end{aligned} \quad (7.2)$$

The functions p_C and p_K are defined by the expression (6.1).

Theorem 7.1 has an attractive interpretation. The first statement (7.1) shows that a randomly oriented subspace with codimension m is unlikely to share a ray with a fixed cone C , provided that the codimension m is larger than the statistical dimension $\delta(C)$ of the cone. When the codimension m is smaller than the statistical dimension $\delta(C)$, the subspace and the cone are likely to share a ray.

The transition in behavior expressed in (7.1) takes place when the codimension m of the subspace changes by about $\omega(C) = \sqrt{\delta(C) \wedge \delta(C^\circ)}$. This point explains why the empirical success curves taper in the corners of the graphs in Figure 2.2. Indeed, on the bottom-left side of each panel, the relevant descent cone is small; on the top-right side of each panel, the descent cone is large, so its polar is small. In these regimes, the result (7.1) shows that the phase transition must occur over a narrow range of codimensions.

The second statement (7.2) provides analogous results for the probability that a randomly oriented cone shares a ray with a fixed cone. This event is unlikely when the total statistical dimension of the two cones is smaller than the ambient dimension; it is likely to occur when the total statistical dimension exceeds the ambient dimension.

For the case of two cones, it is harder to analyze the size of the transition region. Since the probability bounds in (7.2) are controlled by the sum $p_C(\lambda) + p_K(\lambda)$, we can only be certain that the probability estimate depends on the larger of the two quantities. It follows that the width of the transition does not exceed the larger of $\omega(C) = \sqrt{\delta(C) \wedge \delta(C^\circ)}$ and $\omega(K) = \sqrt{\delta(K) \wedge \delta(K^\circ)}$. This observation is sufficient to explain why the empirical success curves taper at the top-left and bottom-right of the graphs in Figure 2.3. Indeed, these are the regions where one of the descent cones is small and the polar of the other descent cone is small.

In the next subsection, we proceed with the proof of Theorem 7.1. Afterward, we derive Theorem I from a similar, but slightly easier argument.

7.1. Proof of Theorem 7.1. We may assume that C and K are both closed because $\mathbb{P}\{C \cap \mathbf{Q}K \neq \{\mathbf{0}\}\} = \mathbb{P}\{\overline{C} \cap \overline{\mathbf{Q}K} \neq \{\mathbf{0}\}\}$, where the overline denotes closure. This is a subtle point that follows from the discussion of touching probabilities located in [SW08, pp. 258–259].

Let us begin with the first set (7.1) of results, concerning the probability that a randomly oriented subspace strikes a fixed cone. Consider the first implication, which operates when $m \geq \delta(C) + \lambda$. The Crofton formula (5.10) shows that

$$\mathbb{P}\{C \cap \mathbf{Q}L \neq \{\mathbf{0}\}\} = 2 h_{m+1}(C) \leq t_m(C),$$

where the inequality depends on the interlacing result, Proposition 5.6. The concentration of intrinsic volumes, Theorem 6.1, demonstrates that the tail functional satisfies the bound

$$t_m(C) \leq p_C(\lambda) \quad \text{when} \quad m \geq \delta(C) + \lambda.$$

This completes the first bound. The second result, which holds when $m \leq \delta(C) - \lambda$, follows from a parallel argument.

The conic kinematic formula is required for the second set (7.2) of results, which concern the probability that a randomly oriented cone strikes a fixed cone. Consider the situation where $\delta(C) + \delta(K) \leq d - \lambda$. The kinematic formula (5.9) yields

$$\mathbb{P}\{C \cap \mathbf{Q}K \neq \{\mathbf{0}\}\} = 2 h_{d+1}(C \times K) \leq t_d(C \times K), \quad (7.3)$$

where the inequality follows from Proposition 5.6.

We rely on a simple lemma to bound the tail functional of the product in terms of the individual tail functionals. The proof appears in Appendix D.

Lemma 7.2 (Tail functionals of a product). *Let C and K be closed convex cones. Then*

$$t_{[\delta(C)+\delta(K)+2\lambda]}(C \times K) \leq t_{[\delta(C)+\lambda]}(C) + t_{[\delta(K)+\lambda]}(K).$$

Since the tail functionals are weakly decreasing, our assumption that $\delta(C) + \delta(K) \leq d - \lambda$ implies that

$$t_d(C \times K) \leq t_{[\delta(C)+\delta(K)+2\lambda]}(C \times K) \leq t_{[\delta(C)+\lambda]}(C) + t_{[\delta(K)+\lambda]}(K).$$

Theorem 6.1 delivers an upper bound of $p_C(\lambda) + p_K(\lambda)$ for the right-hand side. Introduce these bounds into the probability inequality (7.3) to complete the proof of the first statement in (7.2). The second result follows from an analogous argument.

7.2. Proof of Theorem I. The simplified kinematic bound of Theorem I involves an argument similar with the proof of Theorem 7.1. First, assume that $\delta(C) + \delta(K) \leq d - \lambda$. The product rule (4.6) for cones states that $\delta(C \times K) = \delta(C) + \delta(K)$, so the implication (6.3) in Theorem 6.1 yields

$$t_d(C \times K) \leq p_{C \times K}(\lambda) \leq 4 \exp\left(\frac{-\lambda^2/8}{d+\lambda}\right) \leq 4e^{-\lambda^2/(16d)} \quad \text{for } 0 \leq \lambda \leq d. \quad (7.4)$$

The second relation holds because the totality rule (4.5) ensures that $\delta(C \times K) \leq d$ or $\delta((C \times K)^\circ) \leq d$. Substitute the inequality (7.4) into the kinematic bound (7.3). Then make the change of variables $\lambda \mapsto a_\eta \sqrt{d}$, where $a_\eta := 4\sqrt{\log(4/\eta)}$, to obtain the estimate

$$\mathbb{P}\{C \cap \mathbf{Q}K \neq \{\mathbf{0}\}\} \leq \eta.$$

This establishes the first part of Theorem I. The argument for the second part is cut from the same pattern.

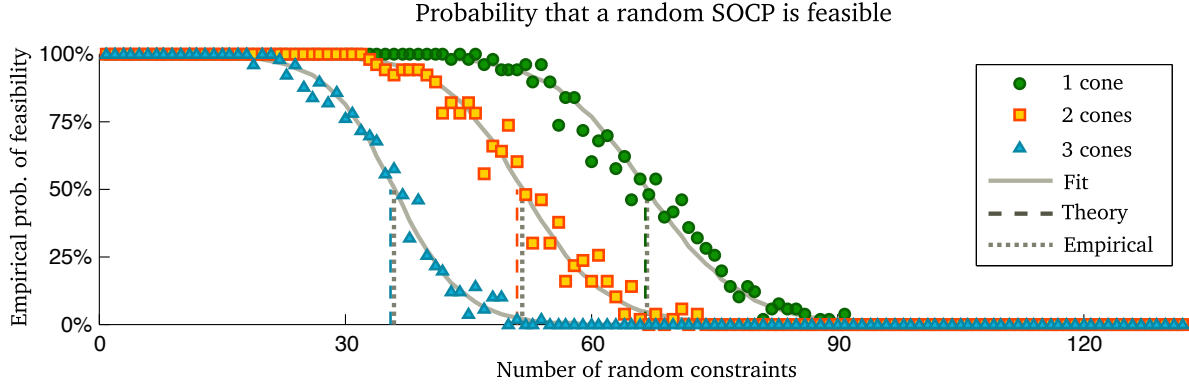


FIGURE 8.1: **Phase transitions in random cone problems.** For each cone C_i in (8.2), we plot the empirical probability that the cone program (8.1) with random affine constraints is feasible. The solid gray curve traces the logistic fit to the data, and the finely dashed line is the empirical 50% success threshold, computed from the regression model. The coarsely dashed line marks the statistical dimension $\delta(C_i)$, which is the theoretical estimate for the location of the phase transition.

8. APPLICATION: CONE PROGRAMS WITH RANDOM CONSTRAINTS

The concentration of intrinsic volumes has far-reaching consequences for the theory of optimization. This section describes a new type of phase transition phenomenon that appears in a cone program with random affine constraints. We begin with a theoretical result, and then we exhibit some numerical examples that confirm the analysis.

8.1. Cone programs. A *cone program* is a convex optimization problem with the following structure:

$$\text{minimize } \langle \mathbf{u}, \mathbf{x} \rangle \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \quad \text{and} \quad \mathbf{x} \in C, \quad (8.1)$$

where $C \in \mathcal{C}_d$ is a closed convex cone. The decision variable $\mathbf{x} \in \mathbb{R}^d$, and the problem data consists of a vector $\mathbf{u} \in \mathbb{R}^d$, a matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$, and another vector $\mathbf{b} \in \mathbb{R}^m$. This formalism includes several fundamental classes of convex programs:

- (1) **Linear programs.** If $C = \mathbb{R}_+^d$, then (8.1) reduces to a linear program in standard form.
- (2) **Second-order cone programs.** If $C = \mathbb{L}^{d+1}$, then (8.1) is a type of second-order cone program.
- (3) **Semidefinite programs.** When $C = \mathbb{S}_+^{n \times n}$, we recover the class of (real) semidefinite programs.

In addition to their flexibility and modeling power, cone programs enjoy effective algorithms and a crisp theory. We refer to [BTN01] for further details.

The cone program (8.1) can exhibit several interesting behaviors. Let us remind the reader of the terminology. A point \mathbf{x} that satisfies the constraints $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \in C$ is called a *feasible point*, and the cone program is *infeasible* when no feasible point exists. The cone program is *unbounded* when there exists a sequence $\{\mathbf{x}_k\}$ of feasible points with the property $\langle \mathbf{u}, \mathbf{x}_k \rangle \rightarrow -\infty$.

Our theory allows us analyze the properties of a random cone program. It turns out that the number m of affine constraints controls whether the cone program is infeasible or unbounded.

Theorem 8.1 (Phase transitions in cone programming). *Let $C \in \mathcal{C}_d$ be a closed convex cone. Consider the cone program (8.1) where the vector $\mathbf{b} \neq \mathbf{0}$ is fixed while the vector $\mathbf{u} \in \mathbb{R}^d$ and the matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ have independent standard normal entries. Then*

$$\begin{aligned} m \leq \delta(C) - \lambda &\implies (8.1) \text{ is unbounded with probability } \geq 1 - p_C(\lambda); \\ m \geq \delta(C) + \lambda &\implies (8.1) \text{ is infeasible with probability } \geq 1 - p_C(\lambda). \end{aligned}$$

The function p_C is defined by the expression (6.1).

TABLE 8.1: **Empirical versus theoretical phase transitions.** For each cone C_i listed in (8.2), we compare the theoretical location of the phase transition, equal to the statistical dimension $\delta(C_i)$, with the empirical location, computed from the logistic regression model in Figure 8.1. The last column lists the errors, relative to the dimension $d = 396$ of the problem.

Cone	$\delta(C_i)$: Theoretical	μ_i : Empirical	$ \delta(C_i) - \mu_i /d$: Error
C_1	66.67	66.88	0.054%
C_2	51.00	51.70	0.177%
C_3	35.50	36.06	0.141%

Proof. Amelunxen & Bürgisser [AB12, Thm. 1.3] have shown that the intrinsic volumes of the cone C control the properties of the random cone program (8.1):

$$\mathbb{P}\{\text{(8.1) is infeasible}\} = 1 - t_m(C);$$

$$\mathbb{P}\{\text{(8.1) has a unique minimizer}\} = v_m(C);$$

$$\mathbb{P}\{\text{(8.1) is unbounded}\} = t_{m+1}(C).$$

We apply Theorem 6.1 to see that the tail functional $t_{m+1}(C)$ is extremely close to one when the number m of constraints is smaller than the statistical dimension $\delta(C)$. Likewise, $t_m(C)$ is extremely close to zero when the number m of constraints is larger than the statistical dimension. We omit the details, which are analogous with the proof of Theorem 7.1. \square

8.2. A numerical example. We have conducted a computer experiment to compare the predictions of Theorem 8.1 with the empirical behavior of a generic cone program. For this purpose, we study some random second-order cone programs. In each case, the ambient dimension $d = 396$, and we consider three options for the cone C in (8.1):

$$C_1 := \text{Circ}_d(\alpha_1); \tag{8.2a}$$

$$C_2 := \text{Circ}_{d/2}(\alpha_2) \times \text{Circ}_{d/2}(\alpha_2); \tag{8.2b}$$

$$C_3 := \text{Circ}_{d/3}(\alpha_3) \times \text{Circ}_{d/3}(\alpha_3) \times \text{Circ}_{d/3}(\alpha_3). \tag{8.2c}$$

The angles satisfy $\tan^2(\alpha_1) = \frac{1}{5}$ and $\tan^2(\alpha_2) = \frac{1}{7}$ and $\tan^2(\alpha_3) = \frac{1}{11}$. Using the product rule (4.6) and the integral expression (C.1) for the statistical dimension of a circular cone, numerical quadrature yields

$$\delta(C_1) \approx 66.67; \quad \delta(C_2) \approx 51.00; \quad \delta(C_3) \approx 35.50.$$

Theorem 8.1 indicates that a cone program (8.1) with the cone C_i and generic constraints is likely to be feasible when the number m of affine constraints is smaller than $\delta(C_i)$; it is likely to be infeasible when the number m of affine constraints is larger than $\delta(C_i)$.

We can test this prediction numerically. For each $i = 1, 2, 3$ and each $m \in \{1, 2, 3, \dots, \lfloor \frac{1}{3}d \rfloor\}$, we perform the following steps 50 times:

- (1) Independently draw a standard normal matrix $A \in \mathbb{R}^{m \times d}$ and standard normal $\mathbf{u} \in \mathbb{R}^d$ and $\mathbf{b} \in \mathbb{R}^m$.
- (2) Use the MATLAB package CVX to solve the cone program (8.1) with $C = C_i$.
- (3) Report failure if CVX declares the cone program infeasible.

For each $i = 1, 2, 3$, Figure 8.1 displays the empirical success probability, along with a logistic fit (Appendix A.3). We also mark the theoretical estimate for the location of the phase transition, which is equal to the statistical dimension $\delta(C_i)$. Table 8.1 reports the discrepancy between the theoretical and empirical behaviors.

9. APPLICATION: VECTORS FROM LISTS?

This section describes a situation where our results prove that a particular linear inverse problem does *not* provide an effective way to recover a structured vector. Indeed, a significant contribution of our theory, which has no parallel in the current literature, is that we can obtain negative results as well as positive results.

In [CRPW12, Sec. 2.2], Chandrasekaran et al. propose a method for recovering a vector from an unordered list of its entries, along with some linear measurements. Here is one way to frame this problem. Suppose

that $\mathbf{x}_0 \in \mathbb{R}^d$ is an unknown vector. We are given the vector $\mathbf{y}_0 = \mathbf{x}_0^\dagger$, whose entries list the components of \mathbf{x}_0 in weakly decreasing order. We also collect data $\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0$ where \mathbf{A} is an $m \times d$ matrix. To identify \mathbf{x}_0 , we must solve a structured linear inverse problem.

To solve this problem, Chandrasekaran et al. propose to use a convex regularizer f that exploits the information in the vector \mathbf{y}_0 . They consider the Minkowski gauge of the permutahedron generated by \mathbf{y}_0

$$\|\mathbf{x}\|_{\mathcal{P}(\mathbf{y}_0)} := \inf\{\tau > 0 : \mathbf{x} \in \tau \mathcal{P}(\mathbf{y}_0)\},$$

and they frame the regularized linear inverse problem

$$\text{minimize } \|\mathbf{x}\|_{\mathcal{P}(\mathbf{y}_0)} \quad \text{subject to } \mathbf{z}_0 = \mathbf{A}\mathbf{x}. \quad (9.1)$$

It is natural to ask how many linear samples we need to be able to solve this inverse problem reliably. Our theory allows us to answer this question decisively when the measurements are random.

Proposition 9.1 (Vectors from lists?). *Let $\mathbf{x}_0 \in \mathbb{R}^d$ be a fixed vector with distinct entries. Suppose we are given the data $\mathbf{y}_0 = \mathbf{x}_0^\dagger$ and $\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0$, where the matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ has standard normal entries. In the range $0 \leq \lambda < \sqrt{H_d}$, it holds that*

$$\begin{aligned} m \leq d - H_d - \lambda\sqrt{H_d} &\implies (9.1) \text{ succeeds with probability } \leq 4e^{-\lambda^2/16}, \\ m \geq d - H_d + \lambda\sqrt{H_d} &\implies (9.1) \text{ succeeds with probability } \geq 1 - 4e^{-\lambda^2/16}. \end{aligned}$$

The d th harmonic number H_d satisfies $\log d < H_d < 1 + \log d$.

Proposition 9.1 yields the depressing assessment that we need a near-complete set of linear measurements to resolve our uncertainty about the ordering of the vector. Nevertheless, we do not need *all* of the measurements. It would be interesting to understand how much the situation improves for vectors with many duplicated entries.

Proof. This result follows from Fact 2.6 and the kinematic bound (7.1) in Theorem 7.1 as soon as we compute the statistical dimension of the descent cone of the regularizer $\|\cdot\|_{\mathcal{P}(\mathbf{y}_0)}$ at the point \mathbf{x}_0 . By construction, the unit ball of $\|\cdot\|_{\mathcal{P}(\mathbf{y}_0)}$ coincides with the permutahedron $\mathcal{P}(\mathbf{y}_0)$, which equals $\mathcal{P}(\mathbf{x}_0)$ by permutation invariance. Therefore,

$$\mathcal{D}(\|\cdot\|_{\mathcal{P}(\mathbf{y}_0)}, \mathbf{x}_0) = \mathcal{D}(\|\cdot\|_{\mathcal{P}(\mathbf{x}_0)}, \mathbf{x}_0) = \mathcal{N}(\mathcal{P}(\mathbf{x}_0), \mathbf{x}_0)^\circ.$$

The second identity follows from (3.1). See Figure 4.2 for an illustration of the corresponding facts about signed permutahedra. To compute the statistical dimension, we apply the totality law (4.5) to see that

$$\delta(\mathcal{D}(\|\cdot\|_{\mathcal{P}(\mathbf{y}_0)}, \mathbf{x}_0)) = d - \delta(\mathcal{N}(\mathcal{P}(\mathbf{x}_0), \mathbf{x}_0)) = d - H_d,$$

where the second relation follows from Proposition 4.9. Apply the kinematic result (7.1) for subspaces, and invoke (6.4) to simplify the error bound $p_C(\lambda)$. \square

Remark 9.2 (Signed vectors). The same negative results hold for the problem of reconstructing a general vector from an unordered list of the *magnitudes* of its entries, along with some linear measurements. In this case, the appropriate regularizer is the Minkowski gauge of the signed permutahedron. We can use Proposition 4.9 to compute the statistical dimension of the descent cone. For a d -dimensional vector with distinct entries, we need about $d - \frac{1}{2}H_d$ random measurements to succeed reliably.

9.1. A numerical example. We present a computer experiment that confirms our pessimistic analysis. Fix the ambient dimension $d = 100$. Set $\mathbf{x}_0 = (1, 2, \dots, 100)$ and $\mathbf{y}_0 = \mathbf{x}_0^\dagger$. For each $m = 85, 86, \dots, 100$, we repeat the following procedure 50 times:

- (1) Draw a matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ with independent standard normal entries, and form $\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0$.
- (2) Use the MATLAB package CVX to solve the linear inverse problem (9.1).
- (3) Declare success if the solution $\hat{\mathbf{x}}$ satisfies $\|\hat{\mathbf{x}} - \mathbf{x}_0\| \leq 10^{-5}$.

Figure 9.1 displays the outcome of this experiment. As usual, the phase transition predicted at the statistical dimension $d - H_d$ is very close to the empirical 50% mark, which we obtain by performing a logistic regression of the data (see Appendix A.3).

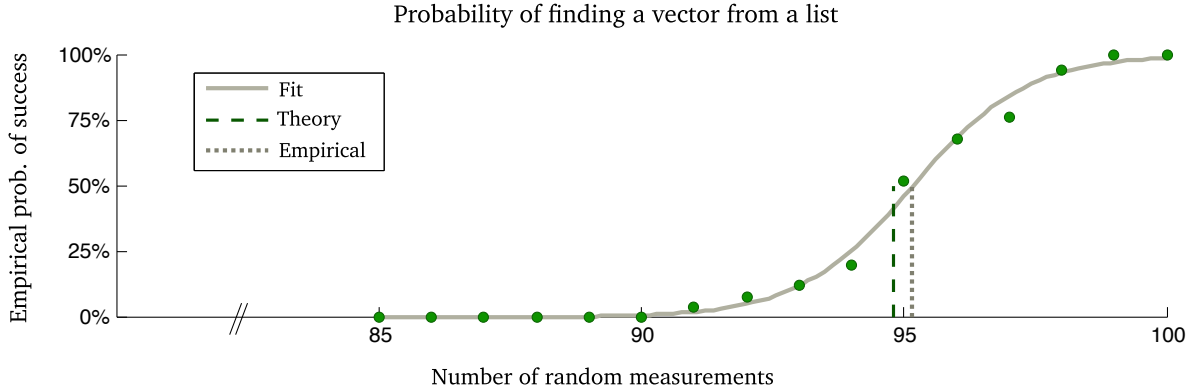


FIGURE 9.1: **Vectors from lists?** The empirical probability that the convex program (9.1) correctly identifies a vector \mathbf{x}_0 in \mathbb{R}^{100} with distinct entries, provided an unordered list \mathbf{y}_0 of the entries of \mathbf{x}_0 and m random linear measurements $\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0$. The solid gray curve marks the logistic fit to the data. The midpoint of the logistic curve $\mu = 95.17$ (finely dashed line), while the theory predicts a phase transition at the statistical dimension $\delta = 94.81$ (coarsely dashed line). The error relative to the dimension $|\mu - \delta|/d = 0.36\%$.

10. RELATED WORK

To conclude the body of the paper, we place our work in the context of the literature on geometric analysis of random convex optimization problems. We trace four lines of thought on this subject. The first draws from the theory of polytope angles; the second involves conic integral geometry; the third is based on comparison inequalities for Gaussian processes; and the last makes a connection with statistical decision theory. Our results have some overlap with earlier work, but our discovery that the sequence of conic intrinsic volumes concentrates at the statistical dimension allows us to resolve several subtle but important questions that have remained open until now.

10.1. Asymptotic polytope-angle computations for inverse problems. The theory of polytope angles dates to the work of Schläfli in the 1850s [Sch50b]. In pioneering research, Vershik & Sporyshev [VS86] applied these ideas to analyze random convex optimization problems. They were able to estimate the average number of steps that the simplex algorithm requires to solve a linear program with random constraints as the number of decision variables tends to infinity. This research inspired further theoretical work on the neighborliness of random polytopes [VS92, AS92, BH99]. More recently, Donoho [Don06b] and Donoho & Tanner [DT05, DT09a, DT10a, DT10b] have used similar ideas to study specific regularized linear inverse problems with random data. The papers [XH11, KXAH11] contain some additional work in this direction. Let us offer a short, qualitative summary of this research.

Donoho [Don06b] analyzed the performance of the convex program (1.1) for solving the compressed sensing problem described in Section 1. In the asymptotic regime where the number s of nonzeros is proportional to the ambient dimension d , he obtained a lower bound $m \geq \psi(s)$ on the number m of Gaussian measurements required for the optimization to succeed (the *weak* transition). Numerical experiments [DT09b] suggest that this bound is sharp, but the theoretical analysis in [Don06b] falls short of establishing that a phase transition actually exists and identifying its location rigorously. Finite-dimensional results with a similar flavor appear in [DT10b].

Donoho [Don06b] also established an asymptotic lower bound on the number m of random measurements required to recover *all* s -sparse vectors in \mathbb{R}^d with high probability (the *strong* threshold). Using different methods, Stojnic [Sto09] has improved this bound for some values of the sparsity s . These bounds are not subject to numerical interrogation, so we do not have reliable evidence about what actually happens. Indeed, it remains an open question to prove that a strong phase transition exists and to identify its exact location in the regime where the sparsity is proportional to the ambient dimension.

Donoho & Tanner [DT09a] have also made a careful study of the behavior of the convex program (1.1) in the asymptotic regime where the sparsity $s \ll d$. In this case, they succeeded in proving that weak and strong thresholds exist, and they obtained exact formulas for the thresholds. More precisely, at the computed

thresholds, they show that the probability of success jumps from one to $1 - \varepsilon$, where ε is positive. Although these results do not ensure that certain failure awaits on the other side of the threshold curve, they do establish that the behavior changes.

Donoho & Tanner [DT05, DT09a] provide similar results for the problem of recovering a sparse nonnegative vector by solving the ℓ_1 minimization problem (1.1) with an additional nonnegativity constraint. Once again, they obtain lower bounds on the number of Gaussian measurements required for weak and strong success. These bounds are sharp in the regime where the ultrasparse regime $s = o(\log d)$. The earlier work of Vershik & Sporyshev [VS86] contains results equivalent with the weak transition estimate from [DT09a].

Other authors have used polytope angle calculations to develop theory for related ℓ_1 minimization problems. For example, Khajehnejad et al. [KXAH11] provide an analysis of the performance of weighted ℓ_1 regularizers. Xu & Hassibi [XH11] obtain lower bounds for the number of measurements required for *stable* recovery of a sparse vector via ℓ_1 minimization.

Finally, let us mention that Donoho & Tanner [DT10a] have obtained a more complete theory about the location of the phase transition for a regularized linear inverse problem where the ℓ_∞ norm is used as a regularizer. Their results, formulated in terms of projections of the hypercube and orthant, are a consequence of a geometric theorem that goes back to Schläfli [Sch50a]; see the discussion [SW08, p. 299].

10.1.1. *Commentary.* The analysis of structured inverse problems by means of polytope angle computations has led to some striking conclusions, but this approach has inherent limitations. First, the method is restricted to polyhedral cones, which means that it is silent about the behavior of many important regularizers, including the Schatten 1-norm. Second, it requires detailed bounds on *all* angles of a given polytope (equivalently, all the intrinsic volumes of the normal cones of the polytope), which means that it is difficult to extend beyond a few highly symmetric examples. For this reason, most of the existing results are asymptotic in nature. Third, because of the intricacy of the calculations, this research has produced few definitive results of the form “the probability of success jumps from one to zero at a specified location.”

We believe that our analysis supersedes most of the research on weak phase transitions for ℓ_1 regularized linear inverse problems that is based on polytope angles. We have shown for the first time that there is a transition from absolute success to absolute failure, and we have characterized the location of the threshold when the sparsity $s \geq \sqrt{d} + 1$. On the other hand, we have not verified that our upper bound for the statistical dimension of the descent cone of the ℓ_1 norm at a sparse vector is sharp when $s \leq \sqrt{d}$. Currently, the paper [DT09a] contains the only authoritative results in the ultrasparse regime $s = o(\log d)$.

It is not hard to extend our analysis to the other settings discussed in this section. Indeed, we can easily study regularized inverse problems involving weighted ℓ_1 norms and ℓ_1 norms with nonnegativity constraints. We can effortlessly rederive phase transitions for ℓ_∞ regularized problems. Bounds for strong transitions are also accessible to our methods. We have omitted all of this material for brevity.

10.2. **Conic intrinsic volumes.** In modern geometry, work on polytope angles has largely been supplanted by research on spherical and conical integral geometry [SW08]. Several authors have independently recognized the power of this approach for analyzing random instances of convex optimization problems.

Amelunxen [Ame11] and Amelunxen & Bürgisser [AB11, AB12] have shown that conic geometry offers an elegant way to perform average-case and smoothed analysis of conic optimization problems. Their work requires detailed computations of conic intrinsic volumes, which can make it challenging to apply to particular cases. We can simplify some of their techniques using the new fact, Theorem 6.1, that intrinsic volumes concentrate at the statistical dimension. Theorem 8.1 is based on their research.

McCoy & Tropp [MT12] have used conic geometry to study the behavior of regularized linear inverse problems with random measurements and regularized demixing problems under a random model. This approach leads to both upper and lower bounds for weak and strong phase transitions in a variety of problems. As with Amelunxen’s work [Ame11], this research depends on detailed computations of conic intrinsic volumes. As a consequence, it was not possible to rigorously locate the phase transition, nor was there any general theory to inform us that phase transitions must exist in general. Combining the ideas from [MT12] with Theorem 7.1, we are able to reach more definitive conclusions.

10.3. **Gordon’s comparison and Gaussian widths.** The work we have discussed so far depends on various flavors of integral geometry. There is a completely different technique for analyzing linear inverse problems with random data that depends on a comparison principle for Gaussian processes, due to Gordon [Gor85].

Gordon [Gor88] explains how to use this comparison to find accurate bounds on the probability that a randomly oriented subspace strikes a subset of the sphere. In particular, his ideas can be used to bound the probability that the null space of a random matrix intersects a descent cone, which results in a version of Theorem II.

Rudelson & Vershynin [RV08] were the first authors to observe that Gordon’s work is relevant to the analysis of the compressed sensing problem. Stojnic [Sto09] refined the method enough that he was able to establish empirically sharp lower bounds for the number m of measurements required for the compressed sensing problem. Oymak & Hassibi [OH10] have also applied these techniques to develop empirically sharp lower bounds on the number of random measurements needed to identify a low-rank matrix using the Schatten 1-norm. Later, Chandrasekaran et al. [CRPW12] showed that a similar approach leads to empirically sharp lower bounds on the number of random measurements required to solve other types of regularized inverse problems.

All of this work depends on a summary parameter for convex cones called the *Gaussian width*. For a closed convex cone $C \in \mathcal{C}_d$, the width is defined as

$$w(C) := \mathbb{E} \sup_{\mathbf{x} \in C \cap S^{d-1}} \langle \mathbf{g}, \mathbf{x} \rangle. \quad (10.1)$$

The Gaussian width has a tight connection with the statistical dimension.

Proposition 10.1 (Statistical dimension & Gaussian width). *Let C be a convex cone. Then*

$$w^2(C) \leq \delta(C) \leq w^2(C) + 1. \quad (10.2)$$

The lower bound in (10.2) is an easy consequence of duality, and the upper bound depends on some concentration arguments. See Appendix E for a short proof.

As a consequence of Proposition 10.1, we can import ideas from the literature on Gaussian widths to obtain accurate computations of the statistical dimension. Conversely, our calculations of the statistical dimension lead to accurate bounds for Gaussian widths. In particular, Theorem 4.5 provides the first proof that previous calculations of Gaussian widths are essentially sharp.

We have a strong preference for the statistical dimension over the Gaussian width. Indeed, the statistical dimension canonically extends the linear dimension to the class of convex cones (Proposition 5.12). The statistical dimension also summarizes the sequence of conic intrinsic volumes (Proposition 5.11). Since intrinsic volumes drive the kinematic formula (5.9) and its generalization (5.11), this connection has many consequences for conic integral geometry.

10.4. Minimax Denoising. Several authors [DMM09a, DJM11, DGM13] have remarked on the power of statistical decision theory to empirically predict the location of the phase transition in a regularized linear inverse problem with random data. For the compressed sensing problem, two recent papers [BM12, BLM12] provide a rigorous explanation for this coincidence. But there is no general theory that illuminates the connection between these two settings. Our work, together with a recent paper of Oymak & Hassibi [OH12], resolves this issue. In short, Oymak & Hassibi show that the minimax risk for denoising is essentially the same as the statistical dimension, while our research proves that a phase transition must occur at the statistical dimension. Let us elaborate.

A classical problem in statistics is to estimate a target vector \mathbf{x}_0 given an observation of the form $\mathbf{z}_0 = \mathbf{x}_0 + \sigma \mathbf{g}$ where \mathbf{g} is a standard normal vector and σ is an unknown variance parameter. When the unknown vector \mathbf{x}_0 has specified properties (e.g., sparsity), we can often construct a convex regularizer f that promotes this type of structure [CRPW12]. A natural estimation procedure is to solve the convex optimization problem

$$\hat{\mathbf{x}}_\gamma := \arg \min_{\mathbf{x} \in \mathbb{R}^d} \gamma f(\mathbf{x}) + \frac{1}{2} \|\mathbf{z}_0 - \mathbf{x}\|^2. \quad (10.3)$$

The regularization parameter $\gamma > 0$ negotiates a tradeoff between the structural penalty and the data fidelity term. One way to assess the performance of the estimator (10.3) is the *minimax MSE risk*,³ defined as

$$R_{\text{mm}}(\mathbf{x}_0) := \sup_{\sigma > 0} \inf_{\gamma > 0} \frac{1}{\sigma^2} \mathbb{E} [\|\hat{\mathbf{x}}_\gamma - \mathbf{x}_0\|^2].$$

³The usual definition of the minimax risk involves an additional supremum over a class of distributions on the target \mathbf{x}_0 . In many applications, the symmetries in the regularizer f allow a straightforward reduction to the case of a fixed target \mathbf{x}_0 .

In other words, the risk identifies the relative mean-square error for the best choice of tuning parameter γ at the worst choice of the noise variance σ^2 .

The papers [DMM09a, DJM11, DGM13] examine several regularizers f where the minimax risk empirically predicts the performance of the linear inverse problem (2.5) with a Gaussian measurement matrix A . The authors of this research propound a conjecture that may be expressed as follows.

Conjecture 10.2 (Minimax risk predicts phase transitions). *Suppose that $A \in \mathbb{R}^{m \times d}$ is a matrix with independent standard normal entries, and let $f: \mathbb{R}^d \rightarrow \mathbb{R}$ be a convex function. Then*

$$\begin{aligned} m \geq R_{\text{mm}}(\mathbf{x}_0) + o(d) &\implies (2.5) \text{ succeeds with probability } 1 - o(1); \\ m \leq R_{\text{mm}}(\mathbf{x}_0) + o(d) &\implies (2.5) \text{ succeeds with probability } o(1). \end{aligned}$$

The order notation here should be interpreted heuristically.

To the best of our knowledge, this claim has been established rigorously only for the ℓ_1 norm [BLM12, Thm. 8] in the asymptotic setting. The paper [BLM12] also includes analysis for a wider class of matrices.

Together, our paper and the recent paper [OH12] settle Conjecture 10.2 in the nonasymptotic setting for many regularizers of interest. Indeed, Oymak & Hassibi [OH12] prove that

$$|R_{\text{mm}}(\mathbf{x}_0) - \delta(\mathcal{D}(f, \mathbf{x}_0))| = O(\sqrt{d}).$$

Their result holds under mild conditions on the regularizer f that suffice to address most of the phase transitions conjectured in the literature. Our result, Theorem II, demonstrates that the phase transition in the linear inverse problem (2.5) with a standard normal matrix $A \in \mathbb{R}^{m \times d}$ occurs when

$$|m - \delta(\mathcal{D}(f, \mathbf{x}_0))| = O(\sqrt{d}).$$

Combining these two results, we conclude that, in some generality, the minimax risk coincides with the location of the phase transition in a regularized linear inverse problem with random measurements.

APPENDIX A. COMPUTER EXPERIMENTS

We confirm the predictions of our theoretical analysis by performing computer experiments. This appendix contains some of the details of our numerical work. All experiments were performed using the CVX package [GB13] for MATLAB with the default settings in place.

A.1. Linear inverse problems with random measurements. This section describes the two experiments from Section 2.3 that illustrate the empirical phase transition in compressed sensing via ℓ_1 minimization and in low-rank matrix recovery via Schatten 1-norm minimization.

In the compressed sensing example, we fix the ambient dimension $d = 100$. For each $m = 1, 2, 3, \dots, d-1$ and each $s = 1, 2, 3, \dots, d-1$, we repeat the following procedure 50 times:

- (1) Construct a vector $\mathbf{x}_0 \in \mathbb{R}^d$ with s nonzero entries. The locations of the nonzero entries are selected at random; each nonzero equals ± 1 with equal probability.
- (2) Draw a standard normal matrix $A \in \mathbb{R}^{m \times d}$, and form $\mathbf{z}_0 = A\mathbf{x}_0$.
- (3) Solve (2.6) to obtain an optimal point $\hat{\mathbf{x}}$.
- (4) Declare success if $\|\hat{\mathbf{x}} - \mathbf{x}_0\| \leq 10^{-5}$.

All random variables are drawn independently in each step and at each iteration. Figures 1.1 and 2.2[left] show the empirical probability of success for this procedure.

We take a similar approach in the low-rank matrix recovery problem. Fix $n = 30$, and consider square $n \times n$ matrices. For each rank $r = 1, 2, \dots, n$ and each $m = 1, 29, 58, 87, \dots, n^2$, we repeat the following procedure 50 times:

- (1) If $r \geq \lceil \sqrt{m} \rceil + 1$, declare failure because the number of degrees of freedom in an $n \times n$ rank- r matrix exceeds the number m of measurements.
- (2) Draw a rank- r matrix $X_0 = Q_1 Q_2^T$, where Q_1 and Q_2 are independent $n \times r$ matrices with orthonormal columns, drawn uniformly from an appropriate Stiefel manifold [Mez07].
- (3) Draw a standard normal matrix $A \in \mathbb{R}^{m \times n^2}$, and define $\mathcal{A}(X) := A \cdot \text{vec}(X)$, where the vectorization operator stacks the columns of a matrix. Form the vector of measurements $\mathbf{z}_0 = \mathcal{A}(X_0)$.
- (4) Solve (2.7) to obtain an optimal point \hat{X} .
- (5) Declare success if $\|\hat{X} - X_0\|_F \leq 10^{-5}$.

As before, all random variables are chosen independently. Readers interested in reproducing this experiment should be aware that this procedure required nearly one month to execute on a desktop workstation. Figure 2.2[right] displays the results of this experiment.

A.2. Statistical dimension curves. The formulas (4.11) and (4.14) for the statistical dimension of the descent cones of the ℓ_1 norm and the Schatten 1-norm do not have a closed form representation. Nevertheless, we can evaluate these expressions using simple numerical methods. Indeed, in each case, we solve the stationary equation (4.13) and (4.16) using the rootfinding procedure `fzero`, which works well because the left-hand side of each equation is a monotone function of τ . To evaluate the integral in (4.11), we use the command `erfc`. To evaluate the integral in (4.14), we use the quadrature function `quadgk`.

We have encountered some numerical stability problems evaluating (4.11) when the proportional sparsity $\rho = s/d$ is close to zero or one. Similarly, there are sometimes difficulties with (4.14) when the proportional rank $\rho = r/m$ or the aspect ratio $\nu = m/n$ are close to zero or one. Nevertheless, relatively simple code based on this approach is usually reliable.

A.3. Logistic regression. Several of the experiments involve fitting the logistic function

$$\ell(x) := \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

to the data, where $\beta_0, \beta_1 \in \mathbb{R}$ are parameters. We use the command `glmfit` to accomplish this task. The center $\mu := -\beta_0/\beta_1$ of the logistic function is the point such that $\ell(\mu) = \frac{1}{2}$.

APPENDIX B. THEORETICAL RESULTS ON DESCENT CONES

This appendix contains the theoretical analysis that permits us to calculate the statistical dimension of descent cones. In particular, we complete the proof of Proposition 4.4, and we establish Theorem 4.5. The material in this appendix has some overlap with independent work due to Oymak & Hassibi [OH12].

B.1. The expected distance to the subdifferential. To complete the proof of Proposition 4.4, we must show that the function $F : \tau \mapsto \mathbb{E}[\text{dist}^2(\mathbf{g}, \tau \cdot \partial f(\mathbf{x}))]$ exhibits a number of analytic and geometric properties. The hypotheses of the proposition ensure that $\partial f(\mathbf{x})$ is a nonempty, compact, convex set that does not contain the origin. For clarity, we establish an abstract result that only depends on the distinguished properties of the subdifferential. Let us begin with a lemma about a related, but simpler, function.

Lemma B.1 (Distance to a dilated set). *Let S be a nonempty, compact, convex subset of \mathbb{R}^d that does not contain the origin. In particular, there are numbers that satisfy $b \leq \|\mathbf{s}\| \leq B$ for all $\mathbf{s} \in S$. Fix a point $\mathbf{u} \in \mathbb{R}^d$, and define the function*

$$F_{\mathbf{u}} : \tau \mapsto \text{dist}^2(\mathbf{u}, \tau S) \quad \text{for } \tau \geq 0. \quad (\text{B.1})$$

The following properties hold.

- (1) The function $F_{\mathbf{u}}$ is convex.
- (2) The function satisfies the lower bound

$$F_{\mathbf{u}}(\tau) \geq (\tau b - \|\mathbf{u}\|)^2 \quad \text{for all } \tau \geq \|\mathbf{u}\|/b. \quad (\text{B.2})$$

In particular, $F_{\mathbf{u}}$ attains its minimum value in the interval $[0, 2\tau/b]$.

- (3) The function $F_{\mathbf{u}}$ is continuously differentiable, and the derivative takes the form

$$F'_{\mathbf{u}}(\tau) = -\frac{2}{\tau} \langle \mathbf{u} - \boldsymbol{\pi}_{\tau S}(\mathbf{u}), \boldsymbol{\pi}_{\tau S}(\mathbf{u}) \rangle \quad \text{for } \tau > 0. \quad (\text{B.3})$$

The right derivative $F'_{\mathbf{u}}(0)$ exists, and $F'_{\mathbf{u}}(0) = \lim_{\tau \downarrow 0} F'_{\mathbf{u}}(\tau)$.

- (4) The derivative admits the bound

$$|F'_{\mathbf{u}}(\tau)| \leq 2B(\|\mathbf{u}\| + \tau B) \quad \text{for all } \tau \geq 0. \quad (\text{B.4})$$

- (5) Furthermore, the map $\mathbf{u} \mapsto F'_{\mathbf{u}}(\tau)$ is Lipschitz for each $\tau \geq 0$:

$$|F'_{\mathbf{u}}(\tau) - F'_{\mathbf{y}}(\tau)| \leq 2B\|\mathbf{u} - \mathbf{y}\| \quad \text{for all } \mathbf{u}, \mathbf{y} \in \mathbb{R}^d. \quad (\text{B.5})$$

Proof. These claims all take some work. Along the way, we also need to establish some auxiliary results to justify the main points.

Convexity. For $\tau > 0$, convexity follows from the representation

$$F_{\mathbf{u}}(\tau) = \left[\inf_{\mathbf{s} \in S} \|\mathbf{u} - \tau \mathbf{s}\| \right]^2 = \left[\tau \cdot \inf_{\mathbf{s} \in S} \|\mathbf{u}/\tau - \mathbf{s}\| \right]^2 = [\tau \cdot \text{dist}(\mathbf{u}/\tau, S)]^2. \quad (\text{B.6})$$

By way of justification, the distance to a closed convex set is a convex function [Roc70, p. 34], the perspective transformation [HUL93a, Sec. IV.2.2] of a convex function is convex, and the square of a nonnegative convex function is convex by a direct calculation.

Continuity. The representation (B.6) shows that the function $F_{\mathbf{u}}$ is continuous for $\tau > 0$ because the distance to a convex set is a Lipschitz function, as stated in (3.5). To obtain continuity at $\tau = 0$, simply note that

$$|F_{\mathbf{u}}(\varepsilon) - F_{\mathbf{u}}(0)| = \left| \|\mathbf{u} - \boldsymbol{\pi}_{\varepsilon S}(\mathbf{u})\|^2 - \|\mathbf{u}\|^2 \right| \leq 2|\langle \mathbf{u}, \boldsymbol{\pi}_{\varepsilon S}(\mathbf{u}) \rangle| + \|\boldsymbol{\pi}_{\varepsilon S}(\mathbf{u})\|^2 \leq 2\|\mathbf{u}\|(\varepsilon B) + (\varepsilon B)^2 \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Indeed, each point in εS is bounded in norm by εB , so the projection $\boldsymbol{\pi}_{\varepsilon S}(\mathbf{u})$ admits the same bound. Continuity implies that $F_{\mathbf{u}}$ is convex on the entire range $\tau \geq 0$.

Attainment of minimum. Assume that $\tau \geq \|\mathbf{u}\|/b$. Then

$$\text{dist}(\mathbf{u}, \tau S) = \inf_{\mathbf{s} \in S} \|\tau \mathbf{s} - \mathbf{u}\| \geq \inf_{\mathbf{s} \in S} [\tau \|\mathbf{s}\| - \|\mathbf{u}\|] \geq \tau b - \|\mathbf{u}\| \geq 0.$$

Square this relation to reach (B.2). It follows that $F_{\mathbf{u}}(\tau) > F_{\mathbf{u}}(0) = \|\mathbf{u}\|^2$ for all $\tau > 2\|\mathbf{u}\|/b$. Therefore, any minimizer of $F_{\mathbf{u}}$ must occur in the compact interval $[0, 2\|\mathbf{u}\|/b]$. Since $F_{\mathbf{u}}$ is continuous, it attains its minimal value in this range.

Differentiability. We obtain the derivative from a direct calculation:

$$\begin{aligned} F'_{\mathbf{u}}(\tau) &= \frac{d}{d\tau} [\tau^2 \text{dist}^2(\mathbf{u}/\tau, S)] = 2\tau \cdot \text{dist}^2(\mathbf{u}/\tau, S) + \tau^2 \langle 2((\mathbf{u}/\tau) - \boldsymbol{\pi}_S(\mathbf{u}/\tau)), -\mathbf{u}/\tau^2 \rangle \\ &= \frac{2}{\tau} [\text{dist}^2(\mathbf{u}, \tau S) - \langle \mathbf{u} - \boldsymbol{\pi}_{\tau S}(\mathbf{u}), \mathbf{u} \rangle] = -\frac{2}{\tau} \langle \mathbf{u} - \boldsymbol{\pi}_{\tau S}(\mathbf{u}), \boldsymbol{\pi}_{\tau S}(\mathbf{u}) \rangle \end{aligned}$$

The first relation follows from (B.6). The second relies on the formula (3.6) for the derivative of the squared distance. To obtain the fourth relation, we express the squared distance as $\|\mathbf{u} - \boldsymbol{\pi}_{\tau S}(\mathbf{u})\|^2$.

Right derivative at zero. The right derivative $F'_{\mathbf{u}}(0)$ exists, and the limit formula holds because $F_{\mathbf{u}}$ is a proper convex function that is continuous on $[0, \infty)$ and differentiable on $(0, \infty)$; see [Roc70, Thm. 24.1].

Continuity of the derivative. The expression (B.3) already implies that $F'_{\mathbf{u}}$ is continuous for $\tau > 0$ because the projection onto a convex set is continuous [RW98, Thm. 2.26]. Continuity of the derivative at zero follows from the limit formula for the right derivative at zero.

Bound for the derivative. Given the formula (B.3), it is easy to control the derivative when $\tau > 0$:

$$|F'_{\mathbf{u}}(\tau)| \leq \frac{2}{\tau} \|\mathbf{u} - \boldsymbol{\pi}_{\tau S}(\mathbf{u})\| \|\boldsymbol{\pi}_{\tau S}(\mathbf{u})\| \leq \frac{2}{\tau} (\|\mathbf{u}\| + \tau B)(\tau B) = 2B(\|\mathbf{u}\| + \tau B).$$

We obtain the estimate for $\tau = 0$ by taking the limit.

Lipschitz property. We obtain the Lipschitz bound (B.5) from (B.3) after some effort. Fix $\tau > 0$. The optimality condition [HUL93a, Thm. III.3.1.1] for a projection onto a closed convex set implies that

$$\langle \mathbf{y} - \boldsymbol{\pi}_{\tau S}(\mathbf{y}), \boldsymbol{\pi}_{\tau S}(\mathbf{y}) \rangle \geq \langle \mathbf{y} - \boldsymbol{\pi}_{\tau S}(\mathbf{y}), \boldsymbol{\pi}_{\tau S}(\mathbf{u}) \rangle \quad \text{for all } \mathbf{u}, \mathbf{y} \in \mathbb{R}^d.$$

As a consequence,

$$\begin{aligned} \langle \mathbf{u} - \boldsymbol{\pi}_{\tau S}(\mathbf{u}), \boldsymbol{\pi}_{\tau S}(\mathbf{u}) \rangle - \langle \mathbf{y} - \boldsymbol{\pi}_{\tau S}(\mathbf{y}), \boldsymbol{\pi}_{\tau S}(\mathbf{y}) \rangle &\leq \langle (\mathbf{u} - \boldsymbol{\pi}_{\tau S}(\mathbf{u})) - (\mathbf{y} - \boldsymbol{\pi}_{\tau S}(\mathbf{y})), \boldsymbol{\pi}_{\tau S}(\mathbf{u}) \rangle \\ &\leq \|(\mathbf{I} - \boldsymbol{\pi}_{\tau S})(\mathbf{u}) - (\mathbf{I} - \boldsymbol{\pi}_{\tau S})(\mathbf{y})\| \|\boldsymbol{\pi}_{\tau S}(\mathbf{u})\| \leq \|\mathbf{u} - \mathbf{y}\| \cdot (\tau B). \end{aligned}$$

The last relation relies on the fact (3.4) that the map $\mathbf{I} - \boldsymbol{\pi}_{\tau S}$ is nonexpansive. Reversing the roles of \mathbf{u} and \mathbf{y} in the last calculation, we see that

$$|\langle \mathbf{u} - \boldsymbol{\pi}_{\tau S}(\mathbf{u}), \boldsymbol{\pi}_{\tau S}(\mathbf{u}) \rangle - \langle \mathbf{y} - \boldsymbol{\pi}_{\tau S}(\mathbf{y}), \boldsymbol{\pi}_{\tau S}(\mathbf{y}) \rangle| \leq (\tau B) \cdot \|\mathbf{u} - \mathbf{y}\|.$$

Combining this estimate with the expression (B.3) for the derivative, we reach

$$|F'_{\mathbf{u}}(\tau) - F'_{\mathbf{y}}(\tau)| \leq 2B \cdot \|\mathbf{u} - \mathbf{y}\|.$$

For $\tau = 0$, the result follows when we take the limit as $\tau \downarrow 0$. \square

With this result at hand, we are prepared to prove a lemma that confirms the remaining claims from Proposition 4.4.

Lemma B.2 (Expected distance to a dilated set). *Let S be a nonempty, compact, convex subset of \mathbb{R}^d that does not contain the origin. Then the function*

$$F(\tau) := \mathbb{E}[\text{dist}^2(\mathbf{g}, \tau S)] = \mathbb{E}[F_{\mathbf{g}}(\tau)] \quad \text{for } \tau \geq 0$$

is strictly convex, continuous at $\tau = 0$, and differentiable for $\tau > 0$. It attains its minimum at a unique point. Furthermore,

$$F'(\tau) = \mathbb{E}[F'_{\mathbf{g}}(\tau)] \quad \text{for all } \tau \geq 0. \quad (\text{B.7})$$

For $\tau = 0$, we interpret $F'(\tau)$ as a right derivative.

Proof. These properties will follow from Lemma B.1, and we continue using the notation from this result.

Convexity. The function F is convex for $\tau \geq 0$ because it is an average of functions of the form $F_{\mathbf{g}}$, each of which is convex.

Strict convexity. We argue by contradiction. Since F is convex, if F were not strictly convex, its graph would contain a linear segment. More precisely, there would be numbers $0 \leq \rho < \tau$ and $\eta \in (0, 1)$ for which

$$\mathbb{E}[F_{\mathbf{g}}((\eta\rho + (1-\eta)\tau)S)] = \mathbb{E}[\eta \cdot F_{\mathbf{g}}(\rho) + (1-\eta) \cdot F_{\mathbf{g}}(\tau)]. \quad (\text{B.8})$$

The convexity of $F_{\mathbf{g}}$ ensures that, for each \mathbf{g} , the bracket on the right-hand side is no smaller than the bracket on the left-hand side. Therefore, the relation (B.8) holds if and only if the two brackets are equal almost surely with respect to the Gaussian measure. But note that

$$\begin{aligned} F_0(\eta\rho + (1-\eta)\tau) &= \text{dist}^2(\mathbf{0}, (\eta\rho + (1-\eta)\tau)S) = (\eta\rho + (1-\eta)\tau)^2 \cdot \inf_{\mathbf{s} \in S} \|\mathbf{s}\|^2 \\ &< (\eta\rho^2 + (1-\eta)\tau^2) \cdot \inf_{\mathbf{s} \in S} \|\mathbf{s}\|^2 = \eta \cdot \text{dist}^2(\mathbf{0}, \rho S) + (1-\eta) \cdot \text{dist}^2(\mathbf{0}, \tau S) = \eta \cdot F_0(\rho) + (1-\eta) \cdot F_0(\tau). \end{aligned}$$

The strict inequality depends on the strict convexity of the square, together with the fact that the infimum is strictly positive. On account of (3.5), the squared distance to a convex set is a continuous function, so there is an open ball around the origin where the same relation holds. That is, for some $\varepsilon > 0$,

$$F_{\mathbf{u}}(\eta\rho + (1-\eta)\tau) < \eta \cdot F_{\mathbf{u}}(\rho) + (1-\eta) \cdot F_{\mathbf{u}}(\tau) \quad \text{when } \|\mathbf{u}\| < \varepsilon.$$

This statement contravenes (B.8).

Continuity at zero. Imitating the continuity argument in Lemma B.1, we find that

$$F(\varepsilon) - F(0) = \mathbb{E}[F_{\mathbf{g}}(\varepsilon) - F_{\mathbf{g}}(0)] \leq \mathbb{E}[2\|\mathbf{g}\| \|\pi_{\varepsilon S}(\mathbf{g})\| + \|\pi_{\varepsilon S}(\mathbf{g})\|^2] \leq 2\sqrt{d} \cdot (\varepsilon B) + (\varepsilon B)^2 \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

This is all that is required.

Differentiability. This point follows from a routine application of the Dominated Convergence Theorem. Indeed, for every $\tau \geq 0$, the function $F(\tau) = \mathbb{E}[F_{\mathbf{g}}(\tau)]$ takes a finite value, and Lemma B.1 establishes that $F'_{\mathbf{g}}$ is continuously differentiable. For each compact interval I , the bound (B.4) ensures that

$$\mathbb{E} \sup_{\tau \in I} |F'_{\mathbf{g}}(\tau)| \leq \mathbb{E} \sup_{\tau \in I} [2B(\|\mathbf{g}\| + \tau B)] \leq 2B\sqrt{d} + 2B^2 \left[\sup_{\tau \in I} \tau^2 \right] < \infty.$$

The convergence theorem now implies that $F'(\tau) = \frac{d}{d\tau} \mathbb{E}[F_{\mathbf{g}}(\tau)] = \mathbb{E}[F'_{\mathbf{g}}(\tau)]$ for all $\tau \geq 0$.

Attainment of minimum. The median of the random variable $\|\mathbf{g}\|$ does not exceed \sqrt{d} . Therefore, when $\tau b \geq \sqrt{d}$, we have

$$F(\tau) \geq \mathbb{E}[F_{\mathbf{g}}(\tau) \mid \|\mathbf{g}\| \leq \sqrt{d}] \cdot \mathbb{P}\{\|\mathbf{g}\| \leq \sqrt{d}\} \geq \frac{1}{2} \mathbb{E}[(\tau b - \|\mathbf{g}\|)^2 \mid \|\mathbf{g}\| \leq \sqrt{d}] \geq \frac{1}{2}(\tau b - \sqrt{d})^2.$$

The first inequality follows from the law of total expectation, and the second depends on (B.2). In particular, $F(\tau) > F(0) = d$ when $\tau > 2b^{-1}\sqrt{d}$. Thus, any minimizer of F must occur in the compact interval $[0, 2b^{-1}\sqrt{d}]$. Since F is strictly convex and continuous, it attains its minimum at a unique point. \square

B.2. Error bound for descent cone calculations. In this section, we prove Theorem 4.5, which provides an error bound for Proposition 4.4. We require a standard result concerning the variance of a Lipschitz function of a standard normal vector.

Fact B.3 (Variance of a Lipschitz function). *Let $H : \mathbb{R}^d \rightarrow \mathbb{R}$ be a function that is Lipschitz with respect to the Euclidean norm:*

$$|H(\mathbf{u}) - H(\mathbf{y})| \leq M \cdot \|\mathbf{u} - \mathbf{y}\| \quad \text{for all } \mathbf{u}, \mathbf{y} \in \mathbb{R}^d.$$

Then

$$\text{Var}(H(\mathbf{g})) \leq M^2 \tag{B.9}$$

where \mathbf{g} is a standard normal vector.

Fact B.3 is a consequence of the Gaussian Poincaré inequality; see [Bog98, Thm. 1.6.4] or [Led01, p. 49].

Proof of Theorem 4.5. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a norm, and fix a nonzero point $\mathbf{x} \in \mathbb{R}^d$. According to [HUL93a, Ex. VI.3.1], the subdifferential of the norm satisfies

$$S = \partial f(\mathbf{x}) = \{\mathbf{s} \in \mathbb{R}^d : \langle \mathbf{s}, \mathbf{x} \rangle = f(\mathbf{x}) \quad \text{and} \quad f^\circ(\mathbf{s}) = 1\}, \tag{B.10}$$

where f° is the norm dual to f . Thus, S is nonempty, compact, convex, and it does not contain the origin.

As in Lemmas B.1 and B.2, we introduce the functions

$$F_{\mathbf{u}} : \tau \mapsto \text{dist}^2(\mathbf{u}, \tau S) \quad \text{and} \quad F : \tau \mapsto \mathbb{E}[F_{\mathbf{g}}(\tau)],$$

where \mathbf{g} is a standard normal vector. Proposition 4.4 provides the upper bound $\mathbb{E} \inf_{\tau \geq 0} F_{\mathbf{g}}(\tau) \leq \inf_{\tau \geq 0} F(\tau)$. Our objective is to develop a reverse inequality.

We establish the result by linearizing each function $F_{\mathbf{g}}$ around a suitable point. Lemma B.2 shows that the function F attains its minimum at a unique location, so we may define

$$\tau_\star := \arg \min_{\tau \geq 0} F(\tau).$$

Similarly, for each $\mathbf{u} \in \mathbb{R}^d$, Lemma B.1 shows that $F_{\mathbf{u}}$ attains its minimum at some point $\tau_{\mathbf{u}} \geq 0$. For the moment, it does not matter how we select this minimizer. Since $F_{\mathbf{u}}$ is convex and differentiable, we can bound its minimum value below using the tangent at τ_\star . That is,

$$\inf_{\tau \geq 0} F_{\mathbf{u}}(\tau) = F_{\mathbf{u}}(\tau_{\mathbf{u}}) \geq F_{\mathbf{u}}(\tau_\star) + (\tau_{\mathbf{u}} - \tau_\star) \cdot F'_{\mathbf{u}}(\tau_\star).$$

Should $\tau_\star = 0$, we interpret $F'_{\mathbf{u}}(\tau_\star)$ as a right derivative. Replacing \mathbf{u} by the random vector \mathbf{g} and taking the expectation, we reach

$$\begin{aligned} \mathbb{E} \left[\inf_{\tau \geq 0} F_{\mathbf{g}}(\tau) \right] &\geq \mathbb{E}[F_{\mathbf{g}}(\tau_\star)] + \mathbb{E}[(\tau_{\mathbf{g}} - \tau_\star) \cdot F'_{\mathbf{g}}(\tau_\star)] \\ &= F(\tau_\star) + \mathbb{E}[(\tau_{\mathbf{g}} - \mathbb{E}[\tau_{\mathbf{g}}]) \cdot (F'_{\mathbf{g}}(\tau_\star) - \mathbb{E}[F'_{\mathbf{g}}(\tau_\star)])] + \mathbb{E}[\tau_{\mathbf{g}} - \tau_\star] \cdot \mathbb{E}[F'_{\mathbf{g}}(\tau_\star)] \\ &\geq \inf_{\tau \geq 0} F(\tau) - [\text{Var}(\tau_{\mathbf{g}}) \cdot \text{Var}(F'_{\mathbf{g}}(\tau_\star))]^{1/2} + \mathbb{E}[\tau_{\mathbf{g}} - \tau_\star] \cdot F'(\tau_\star). \end{aligned} \tag{B.11}$$

The second inequality depends on Cauchy–Schwarz, and we have invoked (B.7) to identify the derivative of F . From here, we obtain the conclusion (4.10) as soon as we estimate the two error terms. The advantage of this formulation is that the Lipschitz properties of the random variables allow us to control their variances.

First, let us demonstrate that the last term on the right-hand side of (B.11) is nonnegative. Abbreviate $e_1 := \mathbb{E}[\tau_{\mathbf{g}} - \tau_\star] \cdot F'(\tau_\star)$. There are two possibilities to consider. When $\tau_\star > 0$, the derivative $F'(\tau_\star) = 0$ because τ_\star minimizes F . Thus, $e_1 = 0$. On the other hand, when $\tau_\star = 0$, it must be the case that the right derivative $F'(\tau_\star) \geq 0$, or else the minimum of the convex function F would occur at a strictly positive value. Since $\tau_{\mathbf{g}} \geq 0$, we see that the quantity $e_1 \geq 0$.

To compute the variance of $\tau_{\mathbf{g}}$, we need to devise a consistent method for selecting a minimizer $\tau_{\mathbf{u}}$ of $F_{\mathbf{u}}$. Introduce the closed convex cone $K := \text{cone}(S)$, and notice that

$$\inf_{\tau \geq 0} F_{\mathbf{u}}(\tau) = \inf_{\tau \geq 0} \text{dist}^2(\mathbf{u}, \tau S) = \text{dist}^2(\mathbf{u}, K).$$

In other words, the minimum distance to one of the sets τS is attained at the point $\Pi_K(\mathbf{u})$. As such, it is natural to pick a minimizer $\tau_{\mathbf{u}}$ of $F_{\mathbf{u}}$ according to the rule

$$\tau_{\mathbf{u}} := \inf \{ \tau \geq 0 : \Pi_K(\mathbf{u}) \in \tau S \} = \frac{\langle \Pi_K(\mathbf{u}), \mathbf{x} \rangle}{f(\mathbf{x})}. \tag{B.12}$$

The latter identity follows from the expression (B.10) for the subdifferential S . In light of (B.12),

$$|\tau_{\mathbf{u}} - \tau_{\mathbf{y}}| = \frac{1}{f(\mathbf{x})} |\langle \mathbf{\Pi}_K(\mathbf{u}) - \mathbf{\Pi}_K(\mathbf{y}), \mathbf{x} \rangle| \leq \frac{\|\mathbf{x}\|}{f(\mathbf{x})} \cdot \|\mathbf{\Pi}_K(\mathbf{u}) - \mathbf{\Pi}_K(\mathbf{y})\| \leq \frac{\|\mathbf{x}\|}{f(\mathbf{x})} \cdot \|\mathbf{u} - \mathbf{y}\|.$$

We have used the fact (3.4) that the projection onto a closed convex set is nonexpansive. Fact B.3 delivers

$$[\text{Var}(\tau_{\mathbf{g}})]^{1/2} \leq \frac{\|\mathbf{x}\|}{f(\mathbf{x})} \quad (\text{B.13})$$

Finally, let us turn to the remaining variance term in (B.11). We have already computed the Lipschitz bound we need for the analysis. Indeed, the inequality (B.5) states that

$$|F'_{\mathbf{u}}(\tau_{\star}) - F'_{\mathbf{y}}(\tau_{\star})| \leq \left[2 \sup_{\mathbf{s} \in S} \|\mathbf{s}\| \right] \cdot \|\mathbf{u} - \mathbf{y}\|.$$

Another invocation of Fact B.3 delivers the estimate

$$[\text{Var}(F'_{\mathbf{g}}(\tau_{\star}))]^{1/2} \leq 2 \sup_{\mathbf{s} \in S} \|\mathbf{s}\|. \quad (\text{B.14})$$

To complete the proof, we combine the inequalities (B.11), (B.13), (B.14), and the fact that $e_1 \geq 0$. This is the advertised result (4.10). \square

APPENDIX C. STATISTICAL DIMENSION CALCULATIONS

This appendix contains the details of the calculations of the statistical dimension for several families of convex cones: circular cones, ℓ_1 descent cones, and Schatten 1-norm descent cones.

C.1. Circular cones. First, we approximate the statistical dimension of a circular cone.

Proof of Proposition 4.3. We begin with an exact integral expression for the statistical dimension of the circular cone $C = \text{Circ}_d(\alpha)$. The spherical formulation (4.2) of the statistical dimension asks us to average the squared norm of the projection of a random unit vector $\boldsymbol{\theta}$ onto the cone. Introduce the angle $\beta := \beta(\boldsymbol{\theta}) := \arccos(\theta_1)$ between $\boldsymbol{\theta}$ and the first standard basis vector $(1, 0, \dots, 0)$. Elementary trigonometry shows that the squared norm of the projection of $\boldsymbol{\theta}$ onto the cone C admits the expression

$$F(\beta) := \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 = \begin{cases} 1, & 0 \leq \beta < \alpha, \\ \cos^2(\beta - \alpha), & \alpha \leq \beta < \frac{\pi}{2} + \alpha, \\ 0, & \frac{\pi}{2} + \alpha \leq \beta \leq \pi. \end{cases}$$

To obtain the exact statistical dimension $\delta(C)$ from (4.2), we integrate $F(\varphi)$ in polar coordinates in the usual way (cf. [SW08, Lem. 6.5.1]):

$$\delta(C) = d \cdot \frac{\Gamma(\frac{1}{2}d)}{\sqrt{\pi}\Gamma(\frac{1}{2}(d-1))} \int_0^\pi \sin^{d-2}(\beta) F(\beta) d\beta. \quad (\text{C.1})$$

We can approximate the integral by a routine application of Laplace's method [AF03, Lem. 6.2.3]:

$$\int_0^\pi \sin^{d-2}(\beta) F(\beta) d\beta = \sqrt{\frac{2\pi}{d}} F\left(\frac{\pi}{2}\right) + O(d^{-3/2}).$$

To simplify the ratio of gamma functions, recall Gautschi's inequality [OLBC10, Sec. 5.6.4]:

$$\sqrt{d-2} < \frac{\sqrt{2}\Gamma(\frac{1}{2}d)}{\Gamma(\frac{1}{2}(d-1))} < \sqrt{d}.$$

Combine the last three displays to reach the expression (4.7).

To obtain the more refined estimate $\cos(2\alpha)$ for the error term, one may use the fact that the intrinsic volumes of a circular cone satisfy

$$v_k(\text{Circ}_d(\alpha)) = \frac{1}{2} \binom{\frac{1}{2}(d-2)}{\frac{1}{2}(k-1)} \sin^{k-1}(\alpha) \cos^{d-k-1}(\alpha) \quad \text{for } k = 1, \dots, d-1. \quad (\text{C.2})$$

This formula is drawn from [Ame11, Ex. 4.4.8]. We are using the analytic extension to define the binomial coefficient. The easiest way to study this sequence is to observe the close connection with the density of a

binomial random variable and to apply the interlacing result, Proposition 5.6. In the interest of brevity, we omit the details. \square

C.2. Descent cones of the ℓ_1 norm. In this section, we show how to use Recipe 4.1 and Theorem 4.5 to compute the statistical dimension of the descent cone of the ℓ_1 norm at a sparse vector. This is a warmup for the more difficult, but entirely similar, calculation in the next section.

Proof of Proposition 4.7. Since the ℓ_1 norm is invariant under signed permutations, we may assume that the sparse vector $\mathbf{x} \in \mathbb{R}^d$ takes the form $\mathbf{x} = (x_1, \dots, x_s, 0, \dots, 0)$, where $x_i > 0$. To compute $\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x}))$, we use the subdifferential bound (4.8) for the statistical dimension of a descent cone:

$$\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x})) \leq \inf_{\tau \geq 0} \mathbb{E} [\text{dist}^2(\mathbf{g}, \tau \cdot \partial \|\mathbf{x}\|_1)]. \quad (\text{C.3})$$

Observe that the subdifferential of the ℓ_1 norm at \mathbf{x} has the following structure:

$$\mathbf{u} \in \partial \|\mathbf{x}\|_1 \iff \begin{cases} u_i = 1, & i = 1, \dots, s \\ |u_i| \leq 1, & i = s+1, \dots, d. \end{cases} \quad (\text{C.4})$$

We can compute the distance from a standard normal vector \mathbf{g} to the dilated subdifferential as follows.

$$\text{dist}^2(\mathbf{g}, \tau \cdot \partial \|\mathbf{x}\|_1) = \sum_{i=1}^s (g_i - \tau)^2 + \sum_{i=s+1}^d \text{Pos}^2(|g_i| - \tau),$$

where $\text{Pos}(a) := a \vee 0$ and the operator \vee returns the maximum of two numbers. Indeed, we always suffer an error in the first s components, and we can always reduce the magnitude of the other components by the amount τ . Taking the expectation, we reach

$$\mathbb{E} [\text{dist}^2(\mathbf{g}, \tau \cdot \partial \|\mathbf{x}\|_1)] = s(1 + \tau^2) + (d - s) \sqrt{\frac{2}{\pi}} \int_{\tau}^{\infty} (u - \tau)^2 e^{-u^2/2} du. \quad (\text{C.5})$$

Simplify the integral, and introduce the resulting expression into (C.3). We reach

$$\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x})) \leq \inf_{\tau \geq 0} \left\{ s(1 + \tau^2) + \sqrt{\frac{2}{\pi}} (d - s) \left[(1 + \tau^2) \int_{\tau}^{\infty} e^{-u^2/2} du - \tau e^{-\tau^2/2} \right] \right\}. \quad (\text{C.6})$$

This expression coincides with the upper bound in (4.12), which has been normalized by the ambient dimension d .

Now, we need to invoke the error estimate, Theorem 4.5. An inspection of (C.4) shows that the subdifferential $\partial \|\mathbf{x}\|_1$ depends on the number s of nonzero entries in \mathbf{x} but not on their magnitudes. It follows from (3.3) that, up to isometry, the descent cone $\mathcal{D}(\|\cdot\|_1, \mathbf{x})$ only depends on the sparsity. Therefore, we may as well assume that $\mathbf{x} = (1, \dots, 1, 0, \dots, 0)$. For this vector, $\|\mathbf{x}\|_1 = \sqrt{s}$. Second, the expression (C.4) for the subdifferential shows that $\|\mathbf{u}\| \leq \sqrt{d}$ for every subgradient $\mathbf{u} \in \partial \|\mathbf{x}\|_1$. Therefore, the error in the inequality (C.6) is at most $2\sqrt{d}\tau$. We reach the lower bound in (4.12).

Finally, Lemma B.2 shows that the brace in (C.6) is a strictly convex, differentiable function of τ with a unique minimizer. It can be verified that the minimum does not occur at $\tau = 0$. Therefore, we determine the stationary equation (4.13) by setting the derivative of the brace to zero and simplifying. \square

C.3. Descent cones of the Schatten 1-norm. Now, we present the calculation of the statistical dimension of the descent cone of the Schatten 1-norm at a low-rank matrix. The approach is entirely similar with the argument in Appendix (C.2).

Proof of Proposition 4.8. Our aim is to identify the statistical dimension of the descent cone of the Schatten 1-norm at a fixed low-rank matrix. The argument here parallels the proof of Proposition 4.7, but we use classical results from random matrix theory to obtain the final expression. Our asymptotic theory demonstrates that this simplification still results in a sharp estimate.

We begin with the fixed-dimension setting. Consider an $m \times n$ real matrix \mathbf{X} with rank r . Without loss of generality, we assume that $m \leq n$ and $0 < r \leq m$. The Schatten 1-norm is unitarily invariant, so we can also assume that \mathbf{X} takes the form

$$\mathbf{X} = \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{where} \quad \Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \quad \text{and} \quad \sigma_i > 0 \text{ for } i = 1, \dots, r.$$

The subdifferential bound (4.8) for the statistical dimension of a descent cone states that

$$\delta(\mathcal{D}(\|\cdot\|_{S_1}, \mathbf{X})) \leq \inf_{\tau \geq 0} \mathbb{E} [\text{dist}^2(\mathbf{G}, \tau \cdot \partial \|\mathbf{X}\|_{S_1})], \quad (\text{C.7})$$

where we compute distance with respect to the Frobenius norm. The $m \times n$ matrix \mathbf{G} has independent standard normal entries, and it is partitioned conformally with \mathbf{X} :

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix} \quad \text{where} \quad \mathbf{G}_{11} \text{ is } r \times r \quad \text{and} \quad \mathbf{G}_{22} \text{ is } (m-r) \times (n-r).$$

According to [Wat92, Ex. 2], the subdifferential of the Schatten 1-norm at \mathbf{X} takes the form

$$\partial \|\mathbf{X}\|_{S_1} = \left\{ \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix} : \sigma_1(\mathbf{W}) \leq 1 \right\}, \quad (\text{C.8})$$

where $\sigma_1(\mathbf{W})$ denotes the maximum singular value of \mathbf{W} . It follows that

$$\text{dist}(\mathbf{G}, \tau \cdot \partial \|\mathbf{X}\|_{S_1})^2 = \left\| \begin{bmatrix} \mathbf{G}_{11} - \tau \mathbf{I}_r & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{0} \end{bmatrix} \right\|_{\text{F}}^2 + \inf_{\sigma_1(\mathbf{W}) \leq 1} \|\mathbf{G}_{22} - \tau \mathbf{W}\|_{\text{F}}^2.$$

Using the Hoffman–Wielandt Theorem [HJ90, Cor. 7.3.8], we can derive

$$\inf_{\sigma_1(\mathbf{W}) \leq 1} \|\mathbf{G}_{22} - \tau \mathbf{W}\|_{\text{F}}^2 = \inf_{\sigma_1(\mathbf{W}) \leq 1} \sum_{i=1}^{m-r} (\sigma_i(\mathbf{G}_{22}) - \tau \sigma_i(\mathbf{W}))^2 = \sum_{i=1}^{m-r} \text{Pos}^2(\sigma_i(\mathbf{G}_{22}) - \tau),$$

where $\sigma_i(\cdot)$ is the i th largest singular value. Combining the last two displays and taking the expectation,

$$\mathbb{E} [\text{dist}^2(\mathbf{G}, \tau \cdot \partial \|\mathbf{X}\|_{S_1})] = r(m+n-r+\tau^2) + \mathbb{E} \left[\sum_{i=1}^{m-r} \text{Pos}^2(\sigma_i(\mathbf{G}_{22}) - \tau) \right]. \quad (\text{C.9})$$

Introduce this expression into (C.7):

$$\delta(\mathcal{D}(\|\cdot\|_{S_1}, \mathbf{X})) \leq \inf_{\tau \geq 0} \left\{ r(m+n-r+\tau^2) + \mathbb{E} \left[\sum_{i=1}^{m-r} \text{Pos}^2(\sigma_i(\mathbf{G}_{22}) - \tau) \right] \right\}. \quad (\text{C.10})$$

We reach a nonasymptotic bound on the statistical dimension.

Next, we apply the error bound, Theorem 4.5. The expression (C.8) shows that the subdifferential $\partial \|\mathbf{X}\|_{S_1}$ depends only on the rank of the matrix \mathbf{X} and its dimension, so the descent cone $\mathcal{D}(\|\cdot\|_{S_1}, \mathbf{X})$ has the same invariance. Therefore, we may consider the $m \times n$ rank- r matrix $\mathbf{X} = \mathbf{I}_r \oplus \mathbf{0}$, which verifies $\|\mathbf{X}\|_{S_1} / \|\mathbf{X}\|_{\text{F}} = \sqrt{r}$. Each subgradient $\mathbf{Y} \in \partial \|\mathbf{X}\|_{S_1}$ satisfies the norm bound $\|\mathbf{Y}\|_{\text{F}} \leq \sqrt{m}$. We conclude that the error in (C.10) is no worse than $2\sqrt{m/r}$.

It is challenging to evaluate the formula (C.10) exactly. In principle, we could accomplish this task using the joint singular value density [And84, p. 534] of the Gaussian matrix \mathbf{G}_{22} . Instead, we set up a framework in which we can use classical random matrix theory to obtain a sharp asymptotic result.

Consider an infinite sequence $\{\mathbf{X}(r, m, n)\}$ of matrices, where $\mathbf{X}(r, m, n)$ has rank r and dimension $m \times n$ with $m \leq n$. For simplicity, we assume that the problem parameters $r, m, n \rightarrow \infty$ with constant ratios $r/m =: \rho \in (0, 1)$ and $m/n =: \nu \in (0, 1]$. The general case follows from a continuity argument. After a change of variables $\tau \mapsto \tau\sqrt{n-r}$ and a rescaling, the expression (C.9) leads to

$$\begin{aligned} & \frac{1}{mn} \mathbb{E} [\text{dist}^2(\mathbf{G}, \tau\sqrt{n-r} \cdot \partial \|\mathbf{X}(r, m, n)\|_{S_1})] \\ &= \rho\nu + \rho(1-\rho\nu)(1+\tau^2) + (1-\rho)(1-\rho\nu) \cdot \mathbb{E} \left[\frac{1}{m-r} \sum_{i=1}^{m-r} \text{Pos}^2(\sigma_i(\mathbf{Z}) - \tau) \right]. \end{aligned} \quad (\text{C.11})$$

Here, \mathbf{G} is an $m \times n$ standard normal matrix. The matrix \mathbf{Z} has dimension $(m-r) \times (n-r)$, and its entries are independent $\text{NORMAL}(0, (n-r)^{-1})$ random variables.

Observe that the expectation in (C.11) can be viewed as a spectral function of a Gaussian matrix. We can obtain the limiting value of this expectation from a variant of the Marčenko–Pastur Law [MP67].

Fact C.1 (Spectral functions of a Gaussian matrix). *Fix a continuous function $F: \mathbb{R}_+ \rightarrow \mathbb{R}$. Suppose $p, q \rightarrow \infty$ and $p/q \rightarrow y \in (0, 1]$. Let \mathbf{Z}_{pq} be a $p \times q$ matrix with independent $\text{NORMAL}(0, q^{-1})$ entries. Then*

$$\mathbb{E} \left[\frac{1}{p} \sum_{i=1}^p F(\sigma_i(\mathbf{Z}_{pq})) \right] \rightarrow \int_{a_-}^{a_+} F(u) \cdot \varphi_y(u) \, du.$$

The limits $a_{\pm} := 1 \pm \sqrt{y}$. The kernel φ_y is a probability density supported on $[a_-, a_+]$:

$$\varphi_y(u) := \frac{1}{\pi y u} \sqrt{(u^2 - a_-^2)(a_+^2 - u^2)} \quad \text{for } u \in [a_-, a_+].$$

Fact C.1 is usually stated differently in the literature. The result here follows from the almost sure weak convergence of the empirical spectral density of a sample covariance matrix to the Marčenko–Pastur density [BS10, Thm. 3.6] and the almost sure convergence of the extreme eigenvalues of a sample covariance matrix [BS10, Thm. 5.8], followed by a change of variables in the integral. We omit the uninteresting details of this reduction.

Let us apply Fact C.1 to our problem. The limiting aspect ratio y of the matrix \mathbf{Z} satisfies

$$y = \frac{m-r}{n-r} = \frac{v(1-\rho)}{1-\rho v}.$$

As $r, m, n \rightarrow \infty$, we obtain the limit

$$\mathbb{E} \left[\frac{1}{m-r} \sum_{i=1}^{m-r} \text{Pos}^2(\sigma_i(\mathbf{Z}) - \tau) \right] \rightarrow \int_{a_-}^{a_+} \text{Pos}^2(u - \tau) \cdot \varphi_y(u) \, du.$$

Simplifying the latter integral and introducing it into (C.11), we reach

$$\frac{1}{mn} \mathbb{E} [\text{dist}^2(\mathbf{G}, \tau \sqrt{n-r} \cdot \partial \|\mathbf{X}(r, m, n)\|_{S_1})] \rightarrow \rho v + \rho(1-\rho v)(1+\tau^2) + (1-\rho)(1-\rho v) \int_{a_- \vee \tau}^{a_+} (u-\tau)^2 \cdot \varphi_y(u) \, du.$$

Rescaling the error estimate for (C.10), we see that the error in the normalized statistical dimension is at most $2/(n\sqrt{m\tau})$, which converges to zero as the parameters grow. We obtain the asymptotic result

$$\frac{1}{mn} \delta(\mathcal{D}(\|\cdot\|_{S_1}, \mathbf{X}(r, m, n))) \rightarrow \inf_{\tau \geq 0} \left\{ \rho v + \rho(1-\rho v)(1+\tau^2) + (1-\rho)(1-\rho v) \int_{a_- \vee \tau}^{a_+} (u-\tau)^2 \cdot \varphi_y(u) \, du \right\}.$$

This is the main conclusion (4.14). To obtain the stationary equation (4.16), we differentiate the brace with respect to τ and set the derivative to zero. \square

C.4. Permutahedra and finite reflection groups. In this section, we use a deep connection between conic geometry and classical combinatorics to compute the statistical dimension of the normal cone of a (signed) permutahedron. This computation requires the full power of Proposition 5.11, the characterization of the statistical dimension as the mean intrinsic volume of a cone.

A *finite reflection group* is a finite subgroup \mathcal{G} of the orthogonal group O_d that is generated by reflections across hyperplanes [ST54, Ste59, CM72, BB10]. Each finite reflection group partitions \mathbb{R}^d into a set $\{\mathbf{U}C_{\mathcal{G}} : \mathbf{U} \in \mathcal{G}\}$ of polyhedral cones called *chambers*. The chambers of the infinite families A_{d-1} and BC_d of irreducible finite reflection groups are isometric to the cones

$$C_A := \{\mathbf{x} \in \mathbb{R}^d : x_1 \leq \dots \leq x_d\} \quad \text{and} \quad C_{BC} := \{\mathbf{x} \in \mathbb{R}^d : 0 \leq x_1 \leq \dots \leq x_d\}.$$

It turns out that the chambers C_A and C_{BC} coincide with the normal cones of certain permutahedra.

Fact C.2 (Normal cones of permutahedra). *Suppose that the vector \mathbf{x} has distinct entries. Then the normal cone $\mathcal{N}(\mathcal{P}(\mathbf{x}), \mathbf{x})$ is isometric to C_A and the normal cone $\mathcal{N}(\mathcal{P}_{\pm}(\mathbf{x}), \mathbf{x})$ is isometric to C_{BC} .*

See [HLT11, Sec. 2] for a proof of Fact C.2.

We claim that the statistical dimensions of the chambers C_A and C_{BC} can be expressed as

$$\delta(C_A) = H_d \quad \text{and} \quad \delta(C_{BC}) = \frac{1}{2} H_d, \tag{C.12}$$

where $H_d := \sum_{k=1}^d i^{-1}$ is the d th harmonic number. Proposition 4.9 follows immediately when we combine this statement with Fact C.2.

Let us explain how the theory of finite reflection groups allows us to deduce the expression (C.12) for the statistical dimension of the chambers. First, it follows from [BZ09] and the characterization [SW08, Eq. (6.50)] of intrinsic volumes in terms of polytope angles that

$$v_k(C_{\mathcal{G}}) = \frac{|\mathcal{H}_k|}{|\mathcal{G}|},$$

where \mathcal{H}_k is the subset of \mathcal{G} consisting of matrices with a $(d-k)$ -dimensional null space. Define the generating polynomial of the intrinsic volumes

$$q_{\mathcal{G}}(s) := \sum_{k=0}^d v_k(C_{\mathcal{G}}) s^k = \frac{1}{|\mathcal{G}|} \sum_{k=0}^d |\mathcal{H}_k| s^k.$$

This polynomial is a well-studied object in the theory of finite reflection groups, and it has many applications in conic geometry as well [Ame11, Sec. 4.4]. For our purposes, we only need the relationships

$$q_{\mathcal{G}}(1) = 1 \quad \text{and} \quad \frac{dq_{\mathcal{G}}}{ds}(1) = \sum_{k=1}^d k v_k(C_{\mathcal{G}}) = \delta(C_{\mathcal{G}}).$$

These points follow immediately from (5.1) and Proposition 5.11.

The roots $\{\zeta_k : k = 1, \dots, d\}$ of the polynomial $q_{\mathcal{G}}$ are called the (negative) *exponents* of the reflection group [CM72, Sec. 7.9]. Factoring the generating polynomial, we obtain a concise expression for the statistical dimension:

$$\delta(C_{\mathcal{G}}) = \frac{dq_{\mathcal{G}}}{ds}(1) = \left(\prod_{k=1}^d \frac{1}{1-\zeta_k} \right) \cdot \frac{d}{ds} \prod_{k=1}^d (s-\zeta_k) \Big|_{s=1} = \sum_{k=1}^d \frac{1}{1-\zeta_k}.$$

We can deduce the value of the large parenthesis because of the normalization $q_{\mathcal{G}}(1) = 1$. The exponents associated with the groups A_{d-1} and BC_d are collected in [CM72, Tab. 10], from which it follows immediately that

$$\delta(C_A) = \sum_{k=1}^d \frac{1}{k} \quad \text{and} \quad \delta(C_{BC}) = \frac{1}{2} \sum_{k=1}^d \frac{1}{k}.$$

This completes the proof of the claim (C.12).

APPENDIX D. TECHNICAL LEMMAS FOR CONCENTRATION OF INTRINSIC VOLUMES

This appendix contains the technical results that undergird the proof of the result on concentration of intrinsic volumes, Theorem 6.1, and the approximate kinematic bound, Theorem 7.1.

D.1. Interlacing of tail functionals. First, we establish the interlacing inequality for the tail functionals. This result is a straightforward consequence of the Crofton formula (5.10).

Proof of Proposition 5.6. Let L_{d-k+1} be a linear subspace of dimension $d-k+1$, and let L_{d-k} be a linear subspace of dimension $d-k$ inside L_{d-k+1} . The Crofton formula (5.10) shows that the half-tail functionals are weakly decreasing:

$$2h_{k+1}(C) = \mathbb{P}\{C \cap \mathbf{Q}L_{d-k} \neq \{\mathbf{0}\}\} \leq \mathbb{P}\{C \cap \mathbf{Q}L_{d-k+1} \neq \{\mathbf{0}\}\} = 2h_k(C),$$

where the inequality follows from the containment of the subspaces. We can express the tail functional t_k as the average of the half-tail functionals:

$$\frac{1}{2} t_k(C) = \frac{1}{2} [h_k(C) + h_{k+1}(C)].$$

Therefore, $2h_k(C) \geq t_k(C) \geq 2h_{k+1}(C)$. □

D.2. Bounds for tropic functions. We continue with the proof of Lemma 6.2, which provides a bound on the tropic functions. This argument is based on an approximation formula from the venerable compendium of Abramowitz & Stegun [AS64, Sec. 26.5.21].

Fact D.1 (Approximation of beta distributions). *Let X be a $\text{BETA}(a, b)$ random variable. Assume that $a + b > 8$ and $(a + b - 1)(1 - x) \geq 0.8$. Define the quantity y via the formula*

$$y = \frac{3 \left[w_1 \left(1 - \frac{1}{9b}\right) - w_2 \left(1 - \frac{1}{9a}\right) \right]}{\left[\frac{w_1^2}{b} + \frac{w_2^2}{a} \right]^{1/2}} \quad \text{where} \quad w_1 = (bx)^{1/3} \quad \text{and} \quad w_2 = (a(1-x))^{1/3}.$$

Then

$$\mathbb{P}\{X \leq x\} = \Phi(y) + \varepsilon(x) \quad \text{with} \quad |\varepsilon(x)| \leq 5 \cdot 10^{-3}.$$

The function Φ represents the cumulative distribution of a standard normal random variable.

Proof of Lemma 6.2. We need to show that the tropic function $I_k^d(k/d) \geq 0.3$ for all integers k and d that satisfy $1 \leq k \leq d-1$. (The cases $k=0$ and $k=d$ are trivial.) To accomplish this goal, we represent the tropic function in terms of a beta random variable, and we apply Fact D.1 to approximate its value.

Let $X \sim \text{BETA}(a, b)$ with shape parameters $a = \frac{1}{2}k$ and $b = \frac{1}{2}(d-k)$. In particular, $\mathbb{E}[X] = k/d$. It is well known [Art02, Sec. 2] that the tropic function can be expressed in terms of this random variable:

$$1 - I_k^d(k/d) = \mathbb{P}\{\|\Pi_{L_k}(\boldsymbol{\theta})\|^2 < k/d\} = \mathbb{P}\{X < \mathbb{E}[X]\}.$$

We must show that this probability is bounded above by 0.7.

First, the mean–median–mode inequality [vdVW93] implies that the mean k/d of X is smaller than the median when $k \geq \frac{1}{2}d$, so the probability is bounded by 0.5 in this regime.

Second, the cases where $a+b \leq 6$ correspond with the situation where $d \leq 12$. We can enumerate the cases where $d \leq 12$ and $0 \leq k \leq d/2$. In each case, we verify numerically that the required probability is less than 0.7.

It is easy to check that for $a = \frac{1}{2}k$, $b = \frac{1}{2}(d-k)$, and $x = k/d$, the inequality $(a+b-1)(1-x) \geq 0.8$ holds when $0 < k < \frac{1}{2}d$ and $d \geq 12$. Instantiating the formula from Fact D.1 and simplifying, we reach

$$y = \frac{\sqrt{2}}{3} \cdot \frac{d-2k}{\sqrt{dk(d-k)}} < \frac{\sqrt{2}}{3}.$$

Indeed, for each d , the extremal choice is $k=1$. The function Φ is increasing, so we conclude that

$$\mathbb{P}\{X \leq k/d\} \leq \Phi\left(\frac{\sqrt{2}}{3}\right) + 5 \cdot 10^{-3} < 0.69.$$

The latter bound results from numerical computation. \square

D.3. The projection of a spherical variable onto a cone. In this section, we establish Lemma 6.3, which controls the probability that a spherical random variable has an unusually large projection on a cone. Although the lemma is framed in terms of a spherical random variable, it is cleaner to derive the result using Gaussian methods. We require an exponential moment inequality [Bog98, Cor. 1.7.9] that ultimately depends on the Gaussian logarithmic Sobolev inequality.

Fact D.2 (Exponential moments for a function of a Gaussian variable). *Suppose that $F: \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies $\mathbb{E}F(\mathbf{g}) = 0$. Assume moreover that F belongs to the Gaussian Sobolev class $H^1(\gamma_d)$, i.e., the squared norm of its gradient is integrable with respect to the Gaussian measure. Then*

$$\mathbb{E}e^{\xi F(\mathbf{g})} \leq \left(\mathbb{E}e^{(\xi/4)\|\nabla F(\mathbf{g})\|^2}\right)^{2\xi/(1-2\xi)} \quad \text{when } 0 < \xi < \frac{1}{2}. \quad (\text{D.1})$$

Using this exponential moment bound, we reach an elegant estimate for the moment generating function of the squared projection of a standard normal vector onto a cone.

Sublemma D.3 (Exponential moment bounds). *Let $K \in \mathcal{C}_d$ be a closed convex cone. Then*

$$\mathbb{E}e^{\xi(\|\Pi_K(\mathbf{g})\|^2 - \delta(K))} \leq \exp\left(\frac{2\xi^2 \delta(K)}{1-4|\xi|}\right) \quad \text{for } -\frac{1}{4} < \xi < \frac{1}{4}. \quad (\text{D.2})$$

Proof. The result is trivial when $\xi = 0$, so we may limit our attention to the case where the parameter ξ is strictly positive or strictly negative. First, suppose that $\xi > 0$. Consider the zero-mean function

$$F(\mathbf{g}) = \|\Pi_K(\mathbf{g})\|^2 - \delta(K) \quad \text{with} \quad \|\nabla F(\mathbf{g})\|^2 = 4\|\Pi_K(\mathbf{g})\|^2. \quad (\text{D.3})$$

The gradient calculation follows from (3.10), and it is easy to see that $F \in H^1(\gamma_d)$ because the projection onto a cone is a contraction. The exponential moment bound (D.1) delivers the estimate

$$\mathbb{E}e^{\xi F(\mathbf{g})} \leq \left(\mathbb{E}e^{\xi\|\Pi_K(\mathbf{g})\|^2}\right)^{2\xi/(1-2\xi)} = \exp\left(\frac{2\xi^2 \delta(K)}{1-2\xi}\right) \cdot \left(\mathbb{E}e^{\xi F(\mathbf{g})}\right)^{2\xi/(1-2\xi)}.$$

The second relation follows when we add and subtract $\xi \delta(K)$ in the exponential function. We have compared the moment generating function of $F(\mathbf{g})$ with itself. Solving the relation, we obtain the inequality

$$\mathbb{E}e^{\xi F(\mathbf{g})} \leq \exp\left(\frac{2\xi^2 \delta(K)}{1-4\xi}\right) \quad \text{for } 0 < \xi < \frac{1}{4}.$$

This is the bound (D.2) for the positive range of parameters.

Now, we turn to the negative range of parameters, which requires a more convoluted argument. To make the analysis clearer, we continue to assume that $\xi > 0$, and we write the negation explicitly. Replacing F by $-F$, the exponential moment bound (D.1) yields

$$\mathbb{E} e^{-\xi F(\mathbf{g})} = \mathbb{E} e^{\xi(-F(\mathbf{g}))} \leq \left(\mathbb{E} e^{\xi \|\mathbf{\Pi}_K(\mathbf{g})\|^2} \right)^{2\xi/(1-2\xi)}. \quad (\text{D.4})$$

This time, we cannot identify a copy of the left-hand side on the right-hand side. Instead, let us run the moment comparison argument directly on the remaining expectation:

$$\mathbb{E} e^{\xi \|\mathbf{\Pi}_K(\mathbf{g})\|^2} = e^{\xi \delta(K)} \cdot \mathbb{E} e^{\xi F(\mathbf{g})} \leq e^{\xi \delta(K)} \left(\mathbb{E} e^{\xi \|\mathbf{\Pi}_K(\mathbf{g})\|^2} \right)^{2\xi/(1-2\xi)}.$$

The last inequality follows from the exponential moment bound (D.1), just as before. Solving this relation, we obtain

$$\mathbb{E} e^{\xi \|\mathbf{\Pi}_K(\mathbf{g})\|^2} \leq \exp\left(\frac{\xi(1-2\xi)\delta(K)}{1-4\xi}\right) \quad \text{for } 0 < \xi < \frac{1}{4}.$$

Introduce the latter inequality into (D.4) to reach

$$\mathbb{E} e^{-\xi F(\mathbf{g})} \leq \exp\left(\frac{2\xi^2 \delta(K)}{1-4\xi}\right).$$

This estimate addresses the remaining part of the parameter range in (D.2). \square

With this result at hand, we can easily prove the tail bound for the projection of a spherical random variable onto a cone.

Proof of Lemma 6.3. For a parameter $\xi > 0$, the Laplace transform method [Bar05, Sec. 2] delivers

$$\mathbb{P}\{d \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 \geq \delta(C) + \lambda\} \leq e^{-\xi\lambda - \xi\delta(C)} \cdot \mathbb{E} e^{\xi d \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2}.$$

Let R be a chi random variable with d degrees of freedom, independent from $\boldsymbol{\theta}$. Using Jensen's inequality, we can bound the expectation:

$$\mathbb{E} e^{\xi d \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2} = \mathbb{E} e^{\xi (\mathbb{E} R^2) \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2} \leq \mathbb{E} e^{\xi \|\mathbf{\Pi}_C(R\boldsymbol{\theta})\|^2} = \mathbb{E} e^{\xi \|\mathbf{\Pi}_C(\mathbf{g})\|^2}.$$

Combining these results, we obtain

$$\mathbb{P}\{d \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 \geq \delta(C) + \lambda\} \leq e^{-\xi\lambda} \cdot \mathbb{E} e^{\xi (\|\mathbf{\Pi}_C(\mathbf{g})\|^2 - \delta(C))}. \quad (\text{D.5})$$

Substitute the inequality for the moment generating function (D.2) with $K = C$ into (D.5) to reach

$$\mathbb{P}\{d \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 \geq \delta(C) + \lambda\} \leq e^{-\xi\lambda} \cdot \exp\left(\frac{2\xi^2 \delta(C)}{1-4\xi}\right) \quad \text{for } 0 < \xi < \frac{1}{4}.$$

Select $\xi = \lambda/(4\delta(C) + 4\lambda)$ to complete the first half of the argument.

The second half of the proof results in an analogous bound with C replaced by C° . Note that

$$\mathbb{P}\{d \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 \geq \delta(C) + \lambda\} = \mathbb{P}\{d (\|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 - 1) + (d - \delta(C)) \geq \lambda\} = \mathbb{P}\{\delta(C^\circ) - d \|\mathbf{\Pi}_{C^\circ}(\boldsymbol{\theta})\|^2 \geq \lambda\}.$$

The second relation follows from the Pythagorean identity (3.8) and the totality law (4.4). Repeating the Laplace transform argument from above, with $\xi > 0$, we obtain the inequality

$$\mathbb{P}\{d \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 \geq \delta(C) + \lambda\} \leq e^{-\xi\lambda} \cdot \mathbb{E} e^{-\xi (\|\mathbf{\Pi}_{C^\circ}(\mathbf{g})\|^2 - \delta(C^\circ))}. \quad (\text{D.6})$$

Introduce the bound (D.2) with $K = C^\circ$ into (D.6) to see that

$$\mathbb{P}\{d \|\mathbf{\Pi}_C(\boldsymbol{\theta})\|^2 \geq \delta(C) + \lambda\} \leq e^{-\xi\lambda} \cdot \exp\left(\frac{2\xi^2 \delta(C^\circ)}{1-4\xi}\right) \quad \text{for } 0 < \xi < \frac{1}{4}.$$

Choose $\xi = \lambda/(4\delta(C^\circ) + 4\lambda)$ to complete the second half of the argument. \square

D.4. Tail functionals of a product. Finally, we argue that the tail functional of a product cone is controlled by the tail functionals of the two summands. This result is very simple, and it can surely be strengthened by taking advantage of the structure of the product.

Proof of Lemma 7.2. Let $C, K \in \mathcal{C}_d$ be closed convex cones. Define independent random variables X and Y that take values in $\{0, 1, \dots, d\}$ and whose distributions are given by the intrinsic volumes of the cones C and K , respectively. That is,

$$\mathbb{P}\{X = k\} = \nu_k(C) \quad \text{and} \quad \mathbb{P}\{Y = k\} = \nu_k(K) \quad \text{for } k = 0, 1, 2, \dots, d.$$

According to the rule (5.4) for the intrinsic volumes of a product cone,

$$\mathbb{P}\{X + Y = k\} = \nu_k(C \times K) \quad \text{for } k = 0, 1, 2, \dots, 2d.$$

By dint of this identity, we can use probabilistic reasoning to bound the tail functionals of the cone $\nu_k(C \times K)$. Indeed, observe that

$$\mathbb{P}\{X + Y \geq \delta(C) + \delta(K) + 2\lambda\} \leq \mathbb{P}\{X \geq \delta(C) + \lambda\} + \mathbb{P}\{Y \geq \delta(K) + \lambda\}.$$

We can rewrite this inequality in terms of tail functionals:

$$t_{[\delta(C) + \delta(K) + 2\lambda]}(C \times K) \leq t_{[\delta(C) + \lambda]}(C) + t_{[\delta(K) + \lambda]}(K).$$

This is the advertised conclusion. \square

APPENDIX E. THE RELATIONSHIP BETWEEN STATISTICAL DIMENSION AND GAUSSIAN WIDTH

This appendix contains a short proof of Proposition 10.1, which states that the Gaussian width of a spherical convex set is comparable with the statistical dimension of the cone generated by the set.

Proof of Proposition 10.1. Let $C \in \mathcal{C}_d$ be a closed convex cone. It is easy to check that the statistical dimension of C dominates its Gaussian width:

$$w^2(C) := [\mathbb{E} \sup_{\mathbf{x} \in C \cap S^d} \langle \mathbf{g}, \mathbf{x} \rangle]^2 \leq [\mathbb{E} \sup_{\mathbf{x} \in C \cap B^d} \langle \mathbf{g}, \mathbf{x} \rangle]^2 = [\mathbb{E} \text{dist}(\mathbf{g}, C^\circ)]^2 \leq \mathbb{E} [\text{dist}^2(\mathbf{g}, C^\circ)] = \delta(C).$$

The first inequality holds because we have enlarged the range of the supremum, and the second relation depends on a standard duality argument [CRPW12, Eqns. (30), (33)]. Afterward, we invoke Jensen's inequality, and we recognize the polar form (4.4) of the statistical dimension.

For the reverse inequality, define the random variable $Z := Z(\mathbf{g}) := \sup_{\mathbf{x} \in C \cap S^{d-1}} \langle \mathbf{g}, \mathbf{x} \rangle$, and note that $w(C) = \mathbb{E} Z$. The function $\mathbf{g} \mapsto Z(\mathbf{g})$ is 1-Lipschitz because the supremum occurs over a subset of the Euclidean unit sphere. Therefore, we can bound the fluctuation of Z as follows.

$$\mathbb{E}[Z^2] - w^2(C) = \mathbb{E}[(Z - \mathbb{E}Z)^2] = \text{Var}(Z) \leq 1. \quad (\text{E.1})$$

The last inequality follows from Fact B.3

As a consequence of (E.1), we obtain the required bound $\delta(C) \leq w^2(C) + 1$ as soon as we verify that $\delta(C) \leq \mathbb{E}[Z^2]$. Since Z^2 is a nonnegative random variable,

$$\mathbb{E}[Z^2] \geq \mathbb{E}[Z^2 \cdot \mathbb{1}_{\mathbb{R}^d \setminus C^\circ}(\mathbf{g})] = \mathbb{E}\left[\left(\sup_{\mathbf{x} \in C \cap S^{d-1}} \langle \mathbf{g}, \mathbf{x} \rangle\right)^2 \cdot \mathbb{1}_{\mathbb{R}^d \setminus C^\circ}(\mathbf{g})\right],$$

where $\mathbb{1}_E$ denotes the indicator of the event E . We claim that the right-hand side of this inequality equals the statistical dimension $\delta(C)$. Indeed, for any $\mathbf{u} \notin C^\circ$, a homogeneity argument delivers $\sup_{\mathbf{x} \in C \cap S^{d-1}} \langle \mathbf{u}, \mathbf{x} \rangle = \sup_{\mathbf{x} \in C \cap B^{d-1}} \langle \mathbf{u}, \mathbf{x} \rangle$. On the other hand, when $\mathbf{u} \in C^\circ$, we have the relation $\sup_{\mathbf{x} \in C \cap B^{d-1}} \langle \mathbf{u}, \mathbf{x} \rangle = 0$. Combine these observations to reach

$$\mathbb{E}[Z^2] \geq \mathbb{E}\left[\left(\sup_{\mathbf{x} \in C \cap S^{d-1}} \langle \mathbf{g}, \mathbf{x} \rangle\right)^2 \cdot \mathbb{1}_{\mathbb{R}^d \setminus C^\circ}(\mathbf{g})\right] = \mathbb{E}\left[\left(\sup_{\mathbf{x} \in C \cap B^{d-1}} \langle \mathbf{g}, \mathbf{x} \rangle\right)^2\right] = \mathbb{E}[\text{dist}^2(\mathbf{g}, C^\circ)].$$

The last relation follows from the same duality argument mentioned above. On account of (4.4), we identify the right-hand side as the statistical dimension $\delta(C)$. \square

ACKNOWLEDGMENTS

DA is with the School of Operations Research and Information Engineering, Cornell University. Research supported by DFG grant AM 386/1-1.

ML is with the School of Mathematics, The University of Manchester. Research supported by Leverhulme Trust grant R41617 and a Seggie Brown Fellowship of the University of Edinburgh.

MBM and JAT are with the Department of Computing and Mathematical Sciences, California Institute of Technology. Research supported by ONR awards N00014-08-1-0883 and N00014-11-1002, AFOSR award FA9550-09-1-0643, and a Sloan Research Fellowship.

The authors wish to thank Babak Hassibi and Samet Oymak for helpful discussions on the connection between phase transitions and minimax risk. Jared Tanner provided detailed information about contemporary research on phase transitions for random linear inverse problems. Venkat Chandrasekaran offered advice on some aspects of convex geometry.

REFERENCES

- [AB11] D. Amelunxen and P. Bürgisser. Probabilistic analysis of the Grassmann condition number. Available at arxiv.org/abs/1112.2603, 2011.
- [AB12] D. Amelunxen and P. Bürgisser. Intrinsic volumes of symmetric cones. Available at arxiv.org/abs/1205.1863, 2012.
- [AF03] M. J. Ablowitz and A. S. Fokas. *Complex variables: Introduction and applications*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2nd edition, 2003.
- [AG03] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program.*, 95(1, Ser. B):3–51, 2003. ISMP 2000, Part 3 (Atlanta, GA).
- [All48] C. B. Allendoerfer. Steiner’s formulae on a general S^{n+1} . *Bull. Amer. Math. Soc.*, 54:128–135, 1948.
- [Ame11] D. Amelunxen. Geometric analysis of the condition of the convex feasibility problem. PhD Thesis, Univ. Paderborn, 2011.
- [And84] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 2nd edition, 1984.
- [Art02] S. Artstein. Proportional concentration phenomena on the sphere. *Israel J. Math.*, 132:337–358, 2002.
- [AS64] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.
- [AS92] F. Affentranger and R. Schneider. Random projections of regular simplices. *Discrete Comput. Geom.*, 7(3):219–226, 1992.
- [Bar05] A. Barvinok. Math 710: Measure concentration. Lecture notes. Available at www.math.lsa.umich.edu/~barvinok/total710.pdf, 2005.
- [BB10] A. V. Borovik and A. Borovik. *Mirrors and reflections*. Universitext. Springer, New York, 2010. The geometry of finite reflection groups.
- [BH99] K. Böröczky, Jr. and M. Henk. Random projections of regular polytopes. *Arch. Math. (Basel)*, 73(6):465–473, December 1999.
- [BLM12] M. Bayati, M. Lelarge, and A. Montanari. Universality in polytope phase transitions and message passing algorithms. Available at arxiv.org/abs/1207.7321, 2012.
- [BM12] M. Bayati and A. Montanari. The LASSO risk for Gaussian matrices. *IEEE Trans. Inform. Theory*, 58(4):1997–2017, 2012.
- [BMS06] J. Bobin, Y. Moudou, and J.-L. Starck. Morphological diversity and source separation. *IEEE Trans. Signal Process.*, 13(7):409–412, 2006.
- [Bog98] V. I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.
- [BS10] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, 2nd edition, 2010.
- [BTN01] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Analysis, algorithms, and engineering applications.
- [BV04] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.
- [BZ09] P. V. Bibikov and V. S. Zhgoon. Angle measures of some cones associated with finite reflection groups. *J. Lie Theory*, 19(4):767–769, 2009.
- [CDS01] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, 2001. Reprinted from *SIAM J. Sci. Comput.* 20 (1998), no. 1, 33–61 (electronic).
- [CM72] H. S. M. Coxeter and W. O. J. Moser. *Generators and relations for discrete groups*. Springer-Verlag, New York, 3rd edition, 1972. *Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 14*.
- [CM73] J. F. Claerbout and F. Muir. Robust modeling of erratic data. *Geophysics*, 38(5):826–844, October 1973.
- [CRPW12] 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.
- [CSPW11] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.*, 21(2):572–596, 2011.
- [CT06] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

- [DGM13] D. L. Donoho, M. Gavish, and A. Montanari. The phase transition of matrix recovery from Gaussian measurements matches the minimax MSE of matrix denoising. Available at arxiv.org/abs/1302.2331, 2013.
- [DJM11] D. L. Donoho, I. Johnstone, and A. Montanari. Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. Available at arxiv.org/abs/1111.1041, 2011.
- [DMM09a] D. L. Donoho, A. Maleki, and A. Montanari. Message-passing algorithms for compressed sensing. *Proc. Nat. Acad. Sci. U.S.A.*, 106(45):18914–18919, 2009.
- [DMM09b] D. L. Donoho, A. Maleki, and A. Montanari. Supporting information to: Message-passing algorithms for compressed sensing. *Proc. Nat. Acad. Sci. U.S.A.*, 2009.
- [Don06a] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [Don06b] D. L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.*, 35(4):617–652, 2006.
- [DT05] D. L. Donoho and J. Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA*, 102(27):9452–9457 (electronic), 2005.
- [DT09a] D. L. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.*, 22(1):1–53, 2009.
- [DT09b] D. L. Donoho and J. Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 367(1906):4273–4293, 2009.
- [DT10a] D. L. Donoho and J. Tanner. Counting the faces of randomly-projected hypercubes and orthants, with applications. *Discrete Comput. Geom.*, 43(3):522–541, 2010.
- [DT10b] D. L. Donoho and J. Tanner. Exponential bounds implying construction of compressed sensing matrices, error-correcting codes, and neighborly polytopes by random sampling. *IEEE Trans. Inform. Theory*, 56(4):2002–2016, 2010.
- [DZ10] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.
- [ESQD05] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Appl. Comput. Harmon. Anal.*, 19(3):340–358, 2005.
- [Faz02] M. Fazel. *Matrix rank minimization with applications*. Phd thesis, Stanford University, 2002.
- [GB13] M. C. Grant and S. P. Boyd. The CVX user’s guide, release 2.0 (beta). User manual. Available at cvxr.com/cvx/doc, 2013.
- [GHS02] F. Gao, D. Hug, and R. Schneider. Intrinsic volumes and polar sets in spherical space. *Math. Notae*, 41:159–176 (2003), 2001/02. Homage to Luis Santaló. Vol. 1 (Spanish).
- [Gla95] S. Glasauer. Integralgeometrie konvexer Körper im sphärischen Raum. PhD Thesis, Univ. Freiburg i. Br., 1995.
- [Gla96] S. Glasauer. Integral geometry of spherically convex bodies. *Diss. Summ. Math.*, 1(1-2):219–226, 1996.
- [Gor85] Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50(4):265–289, 1985.
- [Gor88] 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.
- [Her43] G. Herglotz. Über die Steinersche Formel für Parallellflächen. *Abh. Math. Sem. Hansischen Univ.*, 15:165–177, 1943.
- [HJ90] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.
- [HLT11] C. Hohlweg, C. E. M. C. Lange, and H. Thomas. Permutahedra and generalized associahedra. *Adv. Math.*, 226(1):608–640, 2011.
- [HUL93a] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. I*, volume 305 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Fundamentals.
- [HUL93b] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. II*, volume 306 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Advanced theory and bundle methods.
- [KR97] D. A. Klain and G.-C. Rota. *Introduction to geometric probability*. Lezioni Lincee. [Lincei Lectures]. Cambridge University Press, Cambridge, 1997.
- [KXAH11] A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi. Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models. *IEEE Trans. Signal Process.*, 59(5):1985–2001, 2011.
- [Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [McM93] P. McMullen. Valuations and dissections. In *Handbook of convex geometry, Vol. A, B*, pages 933–988. North-Holland, Amsterdam, 1993.
- [Mez07] F. Mezzadri. How to generate random matrices from the classical compact groups. *Notices Amer. Math. Soc.*, 54(5):592–604, 2007.
- [MP67] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72(114):507–536, 1967.
- [MP97] M. Mesbahi and G. P. Papavassilopoulos. On the rank minimization problem over a positive semidefinite linear matrix inequality. *IEEE Trans. Automat. Control*, 42(2):239–243, 1997.
- [MT12] M. B. McCoy and J. A. Tropp. Sharp recovery bounds for convex deconvolution, with applications. Available at arxiv.org/abs/1205.1580, 2012.
- [OH10] S. Oymak and B. Hassibi. New null space results and recovery thresholds for matrix rank minimization. Available at arxiv.org/abs/1011.6326, 2010.
- [OH12] S. Oymak and B. Hassibi. On the equivalence of minimax risk and the phase transitions of compressed signal recovery. In preparation., 2012.

- [OKH11] S. Oymak, M. A. Khajehnejad, and B. Hassibi. Improved thresholds for rank minimization. In *2011 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5988–5991, 2011.
- [OLBC10] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST handbook of mathematical functions*. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC, 2010.
- [RFP10] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.
- [Roc70] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
- [RV08] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.
- [RW98] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.
- [RXH11] B. Recht, W. Xu, and B. Hassibi. Null space conditions and thresholds for rank minimization. *Math. Program.*, 127(1, Ser. B):175–202, 2011.
- [San50] L. A. Santaló. On parallel hypersurfaces in the elliptic and hyperbolic n -dimensional space. *Proc. Amer. Math. Soc.*, 1:325–330, 1950.
- [San76] L. A. Santaló. *Integral geometry and geometric probability*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1976.
- [Sch50a] L. Schläfli. *Gesammelte mathematische Abhandlungen. Band I*. Verlag Birkhäuser, Basel, 1950.
- [Sch50b] L. Schläfli. *Theorie der Vielfachen Kontinuität (1852), Gesammelte mathematische Abhandlungen. Band I*. Verlag Birkhäuser, Basel, 1950.
- [Sch93] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.
- [SDC03] J.-L. Starck, D. L. Donoho, and E. J. Candès. Astronomical image representation by the curvelet transform. *Astronom. Astrophys.*, 398(2):785–800, 2003.
- [SED05] J.-L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Trans. Image Process.*, 14(10):1570–1582, October 2005.
- [SS86] F. Santosa and W. W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Statist. Comput.*, 7(4):1307–1330, 1986.
- [ST54] G. C. Shephard and J. A. Todd. Finite unitary reflection groups. *Canadian J. Math.*, 6:274–304, 1954.
- [Ste26] J. Steiner. Einige gesetze über die Theilung der Ebene und des Raumes. *J. Reine Angew. Math.*, pages 349–364, 1826.
- [Ste59] R. Steinberg. Finite reflection groups. *Trans. Amer. Math. Soc.*, 91:493–504, 1959.
- [Sto09] M. Stojnic. Various thresholds for ℓ_1 -optimization in compressed sensing. Available at arxiv.org/abs/0907.3666, 2009.
- [SW08] R. Schneider and W. Weil. *Stochastic and Integral Geometry*. Springer series in statistics: Probability and its applications. Springer, 2008.
- [vdVW93] R. van de Ven and N. C. Weber. Bounds for the median of the negative binomial distribution. *Metrika*, 40(3-4):185–189, 1993.
- [VS86] A. M. Vershik and P. V. Sporyshev. An asymptotic estimate for the average number of steps in the parametric simplex method. *USSR Comput. Maths. Math. Phys.*, 26(3):104–113, 1986.
- [VS92] A. M. Vershik and P. V. Sporyshev. Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Selecta Math. Soviet.*, 11(2):181–201, 1992. Selected translations.
- [Wat92] G. A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra Appl.*, 170:33–45, 1992.
- [Win66] R. O. Winder. Partitions of N -space by hyperplanes. *SIAM J. Appl. Math.*, 14:811–818, 1966.
- [XH11] W. Xu and B. Hassibi. Precise stability phase transitions for ℓ_1 minimization: A unified geometric framework. *IEEE Trans. Inform. Theory*, 57(10):6894–6919, October 2011.