Gaussian Phase Transitions and Conic Intrinsic Volumes: Steining the Steiner Formula

Larry Goldstein[1], Ivan Nourdin[2], and Giovanni Peccati[2]

[1]University of Southern California
[2]University of Luxembourg

November 23, 2014

Abstract

Intrinsic volumes of convex sets are natural geometric quantities that also play important roles in applications, such as linear inverse problems with convex constraints, and constrained statistical inference. It is a well-known fact that, given a closed convex cone $C \subset \mathbb{R}^d$, its conic intrinsic volumes determine a probability measure on the finite set $\{0, 1, ...d\}$, customarily denoted by $\mathcal{L}(V_C)$. The aim of the present paper is to provide a Berry-Esseen bound for the normal approximation of $\mathcal{L}(V_C)$, implying a general quantitative central limit theorem (CLT) for sequences of (correctly normalised) discrete probability measures of the type $\mathcal{L}(V_{C_n})$, $n \geq 1$. This bound shows that, in the high-dimensional limit, most conic intrinsic volumes encountered in applications can be approximated by a suitable Gaussian distribution. Our approach is based on a variety of techniques, namely: (1) Steiner formulae for closed convex cones, (2) Stein's method and second order Poincaré inequality, (3) concentration estimates, and (4) Fourier analysis. Our results explicitly connect the sharp phase transitions, observed in many regularised linear inverse problems with convex constraints, with the asymptotic Gaussian fluctuations of the intrinsic volumes of the associated descent cones. In particular, our findings complete and further illuminate the recent breakthrough discoveries by Amelunxen, Lotz, McCoy and Tropp (2014) and McCoy and Tropp (2014) about the concentration of conic intrinsic volumes and its connection with threshold phenomena. As an additional outgrowth of our work we develop total variation bounds for normal approximations of the lengths of projections of Gaussian vectors on closed convex sets.

1 Introduction

1.1 Overview

Every closed convex cone $C \subset \mathbb{R}^d$ can be associated with a random variable V_C, with support on $\{0, \ldots, d\}$ whose distribution $\mathcal{L}(V_C)$ coincides with the so-called *conic intrinsic*

[0]MSC 2010 subject classifications: Primary 60D05, 60F05, 62F30
[0]Key words and phrases: stochastic geometry, convex relaxation

volumes of C. The distribution $\mathcal{L}(V_C)$ is a natural object that summarizes key information about the geometry of C, and is important in applications, ranging from compressed sensing to constrained statistical inference. In particular, for a closed convex cone C the mean $\delta_C = EV_C$ (which is customarily called the *statistical dimension* of C) measures in some sense the 'effective' dimension of C, and generalises the classical notion of dimension for linear subspaces. As proved in the groundbreaking papers Amelunxen, Lotz, McCoy and Tropp [3] and by McCoy and Tropp [32] (see also Section 1.4 below for a more detailed discussion of this point), in the case of the so-called *descent cones* arising in convex optimisation, the concentration of the distribution of V_C around δ_C explains with striking precision threshold phenomena exhibited by the probability of success in linear inverse problems with convex constraints.

Our principal aim in this paper is to produce a Berry-Esseen bound for $\mathcal{L}(V_C)$ leading to minimal conditions on a sequence of closed convex cones $\{C_n\}_{n\geq 1}$, ensuring that the sequence

$$\frac{V_{C_n} - EV_{C_n}}{\sqrt{\mathrm{Var}(V_{C_n})}}, \quad n \geq 1,$$

converges in distribution towards a standard Gaussian $\mathcal{N}(0,1)$ random variable. The bounds in our main findings depend only on the mean and the variance of the random variables V_{C_n}, and are summarized in Part 2 of Theorem 1.1 below.

As explained in the sections to follow, the strategy for achieving our goals consists in using the elegant *Master Steiner formula* from McCoy and Tropp [32], in order to connect random variables of the type V_C to objects with the form $\|\Pi_C(\mathbf{g})\|^2$, where $\mathbf{g}$ is a standard Gaussian vector, Π_C is the metric projection onto C, and $\|\cdot\|$ stands for the Euclidean norm. Shifting from V_C to $\|\Pi_C(\mathbf{g})\|^2$ allows one to unleash the full power of some recently developed techniques for normal approximations, based on the interaction between Stein's method (see [17]) and variational analysis on a Gaussian space (see [34]). In particular, our main tool will be the so-called *second order Poincaré inequality* developed in [14, 35]. In Section 4, we will also use techniques from Fourier analysis in order to compute explicit Berry-Esseen bounds.

As discussed below, our findings represent a significant extension of the results of [3, 32], where the concentration of $\mathcal{L}(V_C)$ around δ_C was first studied by means of tools from Gaussian analysis, as well as by exploiting the connection between intrinsic volumes and metric projections. Explicit applications to regularised linear inverse problems are described in detail in Section 1.4 below.

We will now quickly present some basic facts of conic geometry that are relevant for our analysis. Our main theoretical contributions are discussed in Section 1.3, whereas connections with applications are described in Section 1.4 and Section 1.5.

1.2 Elements of conic geometry

The reader is referred to the classical references [36, 37], as well as to [3, 32], for any unexplained notion or result related to convex analysis.

Distance from a convex set and metric projections. Fix an integer $d \geq 1$. Throughout the paper, we shall denote by $\langle \mathbf{x}, \mathbf{y} \rangle$ and $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle$, respectively, the standard inner product and squared Euclidean norm in $\mathbb{R}^d$. Given a closed convex set $C \subset \mathbb{R}^d$, we define the *distance* between a point $\mathbf{x}$ and C as

$$d(\mathbf{x}, C) := \inf_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|. \tag{1}$$

By the strict convexity of the mapping $\mathbf{x} \mapsto \|\mathbf{x}\|^2$, the infimum is attained at a unique vector, called the *metric projection* of $\mathbf{x}$ onto C, which we denote by $\Pi_C(\mathbf{x})$.

Convex cones and polar cones. A set $C \subset \mathbb{R}^d$ is a *convex cone* if $a\mathbf{x} + b\mathbf{y} \in C$ whenever $\mathbf{x}$ and $\mathbf{y}$ are in C and a and b are positive reals. The *polar cone* C^0 of a cone C is given by

$$C^0 = \left\{ \mathbf{y} \in \mathbb{R}^d : \langle \mathbf{y}, \mathbf{x} \rangle \le 0, \forall \mathbf{x} \in C \right\}. \tag{2}$$

It is easy to verify that the polar cone of a closed convex cone is again a closed convex cone. By virtue e.g. of [32, formula (7.2)], any vector $\mathbf{x} \in \mathbb{R}^d$ may be written as:

$$\mathbf{x} = \Pi_C(\mathbf{x}) + \Pi_{C^0}(\mathbf{x}) \quad \text{with } \Pi_C(\mathbf{x}) \perp \Pi_{C^0}(\mathbf{x}), \tag{3}$$

where the orthogonality relation is in the sense of the inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^d$. A quick computation shows also that, for every closed convex cone C and every $\mathbf{x} \in \mathbb{R}^d$,

$$\|\Pi_C(\mathbf{x})\| = \sup_{\mathbf{y} \in C : \|\mathbf{y}\| \le 1} \langle \mathbf{x}, \mathbf{y} \rangle. \tag{4}$$

Steiner formulae and intrinsic volumes. Letting B^d and S^{d-1} denote, respectively, the unit ball and unit sphere in $\mathbb{R}^d$, the classical *Steiner formula* for the Euclidean expansion of a compact convex set K states that

$$\mathrm{Vol}(K + \lambda B_d) = \sum_{j=0}^{d} \lambda^{d-j} \mathrm{Vol}(B^{d-j}) \mathcal{V}_j \quad \text{for all } \lambda \ge 0,$$

where addition on the left-hand side indicates the Minkowski sum of sets, and the numbers $\mathcal{V}_j, j = 0, \ldots, d$ on the right, called *Euclidean intrinsic volumes*, depend only on K. The Euclidean intrinsic volumes numerically encode key geometric properties of K, for instance, $\mathcal{V}_d$ is the volume, $2\mathcal{V}_{d-1}$ the surface area, and $\mathcal{V}_0$ the Euler characteristic of K. See e.g. [1, p. 142], [29, Chapter 7] and [43, p. 600] for standard proofs.

An 'angular' Steiner formula was developed in [2, 26, 39], and expresses the size of an angular expansion of a closed convex cone C as follows:

$$P\left\{ d^2(\boldsymbol{\theta}, C) \le \lambda \right\} = \sum_{j=0}^{d} \beta_{j,d}(\lambda) v_j, \tag{5}$$

where $\boldsymbol{\theta}$ is a random variable uniformly distributed on S^{d-1}, the coefficients

$$\beta_{j,d}(\lambda) = P[B(d-j, d) \le \lambda]$$

(where each $B(d-j, d)$ has the Beta distribution with parameters $(d-j)/2$ and $d/2$) do not depend on C, and the *conic intrinsic volumes* $v_0, \ldots, v_d$ are determined by C only, and can be shown to be nonnegative and sum to one. As a consequence, we may associate to the conic intrinsic volumes of C an integer-valued random variable V, whose probability distribution $\mathcal{L}(V)$ is given by

$$P(V = j) = v_j, \quad \text{for } j = 0, \dots, d. \tag{6}$$

When the dependence of any quantities on the cone needs to be emphasized, we will write V_C for V and $v_j(C)$ for v_j, $j = 0, \dots, d$. As shown in [32], relation (5) can be seen as a consequence of a general result, known as *Master Steiner formula* and stated formally in Theorem 3.2 below. Such a result implies that, writing $\mathbf{g} \sim \mathcal{N}(0, I_d)$ for a standard d-dimensional Gaussian vector, the squared norms $\|\Pi_C(\mathbf{g})\|^2$ and $\|\Pi_{C^0}(\mathbf{g})\|^2$ behave like two independent chi-squared random variables with a random number V_C and $d-V_C$, respectively, of degrees of freedom: in symbols,

$$(\|\Pi_C(\mathbf{g})\|^2, \|\Pi_{C^0}(\mathbf{g})\|^2) \sim (\chi^2_{V_C}, \chi^2_{d-V_C}). \tag{7}$$

In particular, equation (7) is consistent with the well-known relation $v_j(C) = v_{d-j}(C^0)$ ($j = 0, ..., d$), that is: the distribution of the random variable V_{C^0}, associated with the polar cone C^0 via its intrinsic volumes, satisfies the relation

$$V_{C^0} \overset{\text{Law}}{=} d - V_C, \tag{8}$$

where, here and in what follows, $\overset{\text{Law}}{=}$ indicates equality in distribution. To conclude, we notice that partial versions of (7) (only involving $\|\Pi_C(\mathbf{g})\|^2$) were already known in the literature prior to [32], in particular in the context of constrained statistical inference — see e.g. [19, 40, 41], as well as [42, Chapter 3].

Statistical dimensions. As for Euclidean intrinsic volumes, the distribution of V_C encodes key geometric properties of C. For instance, the mean $\delta_C := E[V_C] = E\|\Pi_C(\mathbf{g})\|^2$, generalizes the notion of dimension. In particular, if L_k is a linear subspace of $\mathbb{R}^d$ of dimension k, and hence a closed convex cone, then $v_j(L_k)$ is one when $j = k$ and zero otherwise, and therefore $\delta(L_k) = k$. The parameter δ_C is often called the *statistical dimension* of C. We observe that, in view of (4), the statistical dimension δ_C is tightly related to the so-called *Gaussian width* of a convex cone

$$w_C := E\left(\sup_{\mathbf{y}\in C:\|\mathbf{y}\|\le 1} \langle \mathbf{g}, \mathbf{y}\rangle\right),$$

where $\mathbf{g} \sim \mathcal{N}(0, I_d)$. The notion of Gaussian width plays an important role in many key results of compressed sensing (see e.g. [38]). Standard arguments yield that $w_C^2 \le \delta_C \le w_C^2+1$ (see [3, Proposition 10.2]). One situation where the statistical dimension is particularly simple to calculate is when C is self dual, that is, when $C = -C^0$. In this case, $\delta_C = d/2$ by (8). The nonnegative orthant, the second-order cone, and the cone of positive-semidefinite matrices are all self dual; see [32] for definitions and further explanations.

Polyhedral cones. We recall that a *polyhedral cone* C is one that can be expressed as the intersection of a finite number of halfspaces, that is, one for which there exists an integer N and vectors $\mathbf{u}_1, \dots, \mathbf{u}_N$ in $\mathbb{R}^d$ such that

$$C = \bigcap_{i=1}^{N} \{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{u}_i, \mathbf{x}\rangle \ge 0\}.$$

For polyhedral cones the probabilities $v_j, j = 0, \dots, d$ can be connected to the behavior of the projection $\Pi_C(\mathbf{g})$ of a standard Gaussian variable $\mathbf{g} \sim \mathcal{N}(0, I_d)$ onto C. Indeed, in this case we have the representation

$$v_j = P\left(\Pi_C(\mathbf{g}) \text{ lies in the relative interior of a } j\text{-dimensional face of } C\right) \tag{9}$$

(see e.g. [3, 32]).

1.3 Main theoretical contributions

The main result of the present paper is the following general central limit theorem (CLT), involving the intrinsic volume distributions of a sequence of closed convex cones with increasing statistical dimensions.

Theorem 1.1 *Let $\{d_n : n \ge 1\}$ be a sequence of non-negative integers and let $\{C_n \subset \mathbb{R}^{d_n} : n \ge 1\}$ be a collection of non-empty closed convex cones such that $\delta_{C_n} \to \infty$, and write $\tau_{C_n}^2 = \mathrm{Var}(V_{C_n})$, $n \ge 1$. For every n, let $\mathbf{g}_n \sim \mathcal{N}(0, I_{d_n})$ and write $\sigma_{C_n}^2 = \mathrm{Var}(\|\Pi_{C_n}(\mathbf{g}_n)\|^2)$, $n \ge 1$. Then, the following holds.*

1. *One has that $2\delta_{C_n} \le \sigma_{C_n}^2 \le 4\delta_{C_n}$ for every n and, as $n \to \infty$, the sequence*
$$\frac{\|\Pi_{C_n}(\mathbf{g}_n)\|^2 - \delta_{C_n}}{\sigma_{C_n}}, \quad n \ge 1,$$
converges in distribution to a standard Gaussian random variable $N \sim \mathcal{N}(0,1)$.

2. *If, in addition, $\liminf_{n\to\infty} \tau_{C_n}^2/\delta_{C_n} > 0$, then, as $n \to \infty$, the sequence*
$$\frac{V_{C_n} - \delta_{C_n}}{\tau_{C_n}}, \quad n \ge 1,$$
also converges in distribution to $N \sim \mathcal{N}(0,1)$, and moreover one has the Berry-Esseen estimate
$$\sup_{u\in\mathbb{R}} \left| P\left[\frac{V_{C_n} - \delta_{C_n}}{\tau_{C_n}} \le u\right] - P[N \le u]\right| = O\left(\frac{1}{\sqrt{\log \delta_{C_n}}}\right). \tag{10}$$

Part 1 follows from Corollary 3.1. Part 2 is a consequence of Theorem 5.1 below that provides a Berry-Esseen bound, with small explicit constants, for the normal approximation of V_C and for any closed convex cone C, in terms of δ_C, σ_C^2 and τ_C^2. In particular, if C is a closed convex cone such that $\tau_C > 0$, then we will prove in Theorem 5.1 and Remark 5.1 that, writing $\alpha := \tau_C^2/\delta_C$, for $\delta_C \ge 8$,

$$\sup_{u\in\mathbb{R}} \left| P\left[\frac{V_C - \delta_C}{\tau_C} \le u\right] - P[N \le u]\right| \le h(\delta_C) + \frac{48}{\sqrt{\alpha \log^+(\alpha\sqrt{2}\delta_C)}}, \tag{11}$$

where

$$h(\delta) = \frac{1}{72}\left(\frac{\log \delta}{\delta^{3/16}}\right)^{5/2}. \tag{12}$$

Remark 1.1 Observe that, if one considers the sequence $\{C_d\}_{d\geq 1}$ consisting of the non-negative orthants of $\mathbb{R}^d$, then V_{C_d} follows a binomial distribution with parameters $(1/2, d)$ (in particular, $\delta_{C_d} = d/2$). It follows that, in this case, the supremum on the left-hand side of (10) converges to zero at a speed of the order $O(d^{-1/2})$, from which we conclude that the rate supplied by (10) is, in general, not optimal.

As anticipated, our strategy for proving Theorem 1.1 (exception made for the Berry-Esseen bound (10)) is to connect the distributions of $\|\Pi_{C_n}(\mathbf{g}_n)\|^2$ and V_{C_n} via the Master Steiner formula (7), and then to study the normal approximation of the squared norm of $\Pi_{C_n}(\mathbf{g}_n)$ by means of Stein's method, as well as of general variational techniques on a Gaussian space (see [17, 34]). As illustrated in the Appendix contained in Section 5 below, Stein's method proceeds by manipulating a characterizing equation for a target distribution (in this case the normal), typically through couplings or integration by parts. Hence, we justify the title of this work by the heavy use that our application of Stein's method makes of relation (7), generalizing the angular Steiner formula (5). As mentioned above, our main tool will be a form of the second order Poincaré inequalities studied in [14, 35].

Remark 1.2 A crucial point one needs to address when applying Part 2 of Theorem 1.1 is that, in order to check the assumption $\liminf_{n\to\infty} \tau^2_{C_n}/\delta_{C_n} > 0$, one has to produce an effective lower bound on the sequence of conic variances $\tau^2_{C_n}$, $n \geq 1$. This issue is dealt with in Section 4, where we will prove new upper and lower bounds for conic variances, by using an improved version of the Poincaré inequality (see Theorem 6.2), as well as a representation of the covariance of smooth functionals of Gaussian fields in terms of the Ornstein-Uhlenbeck semigroup, as stated in formula (96) below. In particular, our main findings of Section 4 (see Theorem 4.1) will indicate that, in many crucial examples, the sequence $n \mapsto \tau^2_{C_n}$ eventually satisfies a relation of the type

$$c\|E[\Pi_{C_n}(\mathbf{g})]\|^2 \leq \tau^2_{C_n} \leq 2\|E[\Pi_{C_n}(\mathbf{g})]\|^2,$$

where $c \in (0, 2)$ does not depend on n. In view of Jensen inequality, this conclusion strictly improves the estimate $\tau^2_{C_n} \leq 2\delta_{C_n}$ that one can derive e.g. from [32, Theorem 4.5].

We obtain normal approximation results for random variables that are more general than $\|\Pi_C(\mathbf{g})\|^2$. To this end, fix a closed convex cone $C \subset \mathbb{R}^d$ and $\boldsymbol{\mu} \in \mathbb{R}^d$, and introduce the shorthand notation:

$$F = \|\boldsymbol{\mu} - \Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2 - m, \quad \text{with} \quad m = E[\|\boldsymbol{\mu} - \Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2] \quad \text{and} \quad \sigma^2 = \mathrm{Var}(F). \tag{13}$$

Then, we prove in Theorem 3.1 that

$$d_{TV}(F, N) \leq \frac{16}{\sigma^2}\left\{\sqrt{m}(1 + 2\|\boldsymbol{\mu}\|) + 3\|\boldsymbol{\mu}\|^2 + \|\boldsymbol{\mu}\|\right\}, \tag{14}$$

where $N \sim \mathcal{N}(0, \sigma^2)$ and d_{TV} stand for the total variation distance, defined in (28), between the distribution of two random variables. In the fundamental case $\boldsymbol{\mu} = \mathbf{0}$, Proposition 3.1 shows that the previous estimate implies the simple relation

$$d_{TV}\left(\|\Pi_C(\mathbf{g})\|^2 - \delta_C,\, N\right) \leq \frac{8}{\sqrt{\delta_C}}, \tag{15}$$

where $N \sim \mathcal{N}(0, \sigma_C^2)$. Relation (15) reinforces our intuition that the statistical dimension δ_C encodes a crucial amount of information about the distributions of $\|\Pi_C(\mathbf{g})\|^2$ and, therefore, about V_C, via (7).

It does not seem possible to directly combine the powerful inequality (15) with (7) in order to deduce an explicit Berry-Esseen bound such as (10). This estimate is obtained in Section 5, by means of Fourier theoretical arguments of a completely different nature.

Remark 1.3 We stress that the crucial idea that one can study a random variable of the type V_C, by applying techniques of Gaussian analysis to the associated squared norm $\|\Pi_C(\mathbf{g})\|^2$, originates from the path-breaking references [3, 32], where this connection is exploited in order to obtain explicit concentration estimates via the entropy method, see [7] and [30].

As stated in the Introduction, we will now show that our results can be used to exactly characterise phase transitions in regularised inverse problems with convex constraints.

1.4 Applications to exact recovery of structured unknowns

1.4.1 General framework

In what follows, we give a summary of how the conic intrinsic volume distribution plays a role in convex optimization for the recovery of structured unknowns and refer the reader e.g. to the excellent discussions in [3, 10, 13, 32] for more detailed information.

In certain high dimension recovery problems some small number of observations may be taken on an unknown high dimensional vector or matrix $\mathbf{x}_0$, thus determining that the unknown lies in the feasible set $\mathcal{F}$ of all elements consistent with what has been observed. As $\mathcal{F}$ may be large, the recovery of $\mathbf{x}_0$ is not possible without additional assumptions, such as that the unknown possesses some additional structure such as being sparse, or of low rank. As searching $\mathcal{F}$ for elements possessing the given structure can be computationally expensive, one instead may consider a convex optimization problem of finding $\mathbf{x} \in \mathcal{F}$ that minimizes $f(\mathbf{x})$ for some proper convex function[1] that promotes the structure desired.

The analysis of such an optimization procedure leads one naturally to the study of the *descent cone* $\mathcal{D}(f, \mathbf{x})$ of f at the point $\mathbf{x}$, given by

$$\mathcal{D}(f, \mathbf{x}) = \{\mathbf{y} : \exists \tau > 0 \text{ such that } f(\mathbf{x} + \tau \mathbf{y}) \leq f(\mathbf{x})\}.$$

That is, $\mathcal{D}(f, \mathbf{x})$ is the conic hull of all directions that do not increase f near $\mathbf{x}$. The proof of Part 1 of Theorem 1.2 below – included here for completeness – reflects the general result, that in the case where $\mathcal{F}$ is a subspace, the convex optimization just described successfully recovers the unknown $\mathbf{x}_0$ if and only if

$$\mathcal{F} \cap (\mathbf{x}_0 + \mathcal{D}(f, \mathbf{x}_0)) = \{\mathbf{x}_0\} \tag{16}$$

(see Section 4 of [38] and Proposition 2.1 [13], and Fact 2.8 of [3]).

[1] a convex function having at least one finite value and never taking the value $-\infty$

The work [13] provides a systematic way according to which an appropriate convex function f may be chosen to promote a given structure. When an unknown vector, or matrix, is expressed as a linear combination

$$\mathbf{x}_0 = c_1\mathbf{a_1} + \cdots c_k\mathbf{a}_k \tag{17}$$

for $c_i \geq 0$, $a_i \in \mathcal{A}$ a set of building blocks or atoms of vectors or matrices, and k small, then one minimizes

$$f(\mathbf{x}) = \inf\{t > 0 : \mathbf{x} \in t\mathrm{conv}(\mathcal{A})\}, \tag{18}$$

over the feasible set, where $\mathrm{conv}(\mathcal{A})$ is the convex hull of $\mathcal{A}$.

1.4.2 Recovery of sparse vectors via ℓ_1 norm minimization

We now consider the underdetermined linear inverse problem of recovering a sparse vector $\mathbf{x}_0 \in \mathbb{R}^d$ from the observation of $\mathbf{z} = A\mathbf{x}_0$, where for $m < d$ the known matrix $A \in \mathbb{R}^{m\times d}$ has independent entries each with the standard normal $\mathcal{N}(0,1)$ distribution. We say the vector $\mathbf{x}_0$ is s-sparse if it has exactly s nonzero components; the value of s is typically much smaller than d. As a sparse vector is a linear combination of a small number of standard basis vectors, the prescription (18) leads us to find a feasible vector that minimizes the ℓ_1 norm, denoted by $\|\cdot\|_1$. It is a well-known fact that such a linear inverse problem displays a *sharp phase transition* (sometimes called a *threshold phenomenon*): heuristically, this means that, for every value of d, there exists a very narrow band $[m_1, m_2]$ (that depends on d and on the sparsity level of $\mathbf{x}_0$) such that the probability of recovering $\mathbf{x}_0$ exactly is negligible for $m < m_1$, and overwhelming for $m > m_2$. Understanding such a phase transition (and, more generally, threshold phenomena in randomised linear inverse problems) has been the object of formidable efforts by many researchers during the last decade, ranging from the seminal contributions by Candès, Romberg and Tao [11, 12], Donoho [20, 21] and Donoho and Tanner [22], to the works of Rudelson and Vershynin [38] and Ameluxen *et al.* [3] (see [10, Section 3], and the references therein, for a vivid description of the dense history of the subject). In particular, reference [3] contains the first proof of the fundamental fact that the above described threshold phenomenon can be explained by the Gaussian concentration of the intrinsic volumes of the descent cone of the ℓ_1 norm at $\mathbf{x}_0$ around its statistical dimension. In what follows, we shall further refine such a finding by showing that, for large values of d, the phase transition for the exact recovery of $\mathbf{x}_0$ has an almost exact Gaussian nature, following from the general quantitative CLTs for conic intrinsic volumes stated at Point 2 of Theorem 1.1.

The next statement provides finite sample estimates, valid in any dimension. Note that we use the symbol $\lfloor a \rfloor$ to indicate the integer part of a real number a.

Theorem 1.2 (Finite sample) *Let $\mathbf{x}_0 \in \mathbb{R}^d$ and let C be the descent cone of the ℓ_1 norm $\|\cdot\|_1$ at $\mathbf{x}_0$. Further, let V be the random variable defined by (6), set $\delta = E[V]$ to be the statistical dimension of C, and $\tau^2 = \mathrm{Var}(V)$. Let $T_{\delta,\tau}$ be the set of real numbers t such that the number of observations*

$$m_t := \lfloor \delta + t\tau \rfloor$$

lies between 1 and d. Fix $t \in T_{\delta,\tau}$. Let $A_t \in \mathbb{R}^{m_t \times d}$ have independent entries, each with the standard normal $\mathcal{N}(0,1)$ distribution and let $\mathcal{F}_t = \{\mathbf{x} \in \mathbb{R}^d : A_t\mathbf{x} = A_t\mathbf{x}_0\}$. Consider the convex program

$$(\mathbf{CP}_t): \quad \min \|\mathbf{x}\|_1 \quad \textit{subject to } \mathbf{x} \in \mathcal{F}_t.$$

Then, for $\delta \geq 8$ one has the estimate

$$\sup_{t \in T_{\delta,\tau}} \left| P\{\mathbf{x}_0 \textit{ is the unique solution of } (\mathbf{CP}_t)\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du \right| \leq h(\delta) + \frac{48}{\sqrt{\alpha \log^+(\alpha\sqrt{2}\delta)}} + \frac{1}{\sqrt{2\pi\tau^2}}, \tag{19}$$

where $\alpha := \tau^2/\delta$, and $h(\delta)$ given by (12).

Remark 1.4 1. The estimate (19) implies that, for a fixed d and up to a uniform explicit error, the mapping

$$t \mapsto P\{\mathbf{x}_0 \text{ is the unique solution of } (\mathbf{CP}_t)\},$$

(expressing the probability of recovery as a function of m_t) can be approximated by the standard Gaussian distribution function $t \mapsto \Phi(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du$, thus demonstrating the Gaussian nature of the threshold phenomena described above. To better understand this point, fix a small $\alpha \in (0,1)$, and let y_α be such that $\Phi(y_\alpha) = 1 - \alpha$. Then, standard computations imply that (up to the uniform error appearing in (19)) the probability

$$P\{\mathbf{x}_0 \text{ is the unique solution of } (\mathbf{CP}_{y_\alpha})\}$$

is bounded from below by $1 - \alpha$, whereas $P\{\mathbf{x}_0 \text{ is the unique solution of } (\mathbf{CP}_{-y_\alpha})\}$ is bounded from above by α. Using the explicit expressions $m_{-y_\alpha} = \lfloor \delta - y_\alpha \tau \rfloor$ and $m_{y_\alpha} = \lfloor \delta + y_\alpha \tau \rfloor$, one therefore sees that the transition from a negligible to an overwhelming probability of exact reconstruction takes place within a band of approximate length $2y_\alpha\tau \leq 2y_\alpha\sqrt{2\delta}$, centered at δ. In particular, if $\delta \to \infty$, then the length of such a band becomes negligible with respect to δ, thus accounting for the sharpness of the phase transition. Sufficient conditions, ensuring that $\alpha = \tau^2/\delta$ is bounded away from zero when $\delta \to \infty$, are given in Theorem 1.3.

2. Define the mapping $\psi : [0,1] \to [0,1]$ as

$$\psi(\rho) := \inf_{\gamma \geq 0} \left\{ \rho(1+\gamma^2) + (1-\rho)E[(|N| - \gamma)_+^2] \right\}, \tag{20}$$

where $N \sim \mathcal{N}(0,1)$. The following estimate is taken from [3, Proposition 4.5]: under the notation and assumptions of Theorem 1.2, if $\mathbf{x_0}$ is s-sparse, then

$$\psi(s/d) - \frac{2}{\sqrt{sd}} \leq \frac{\delta}{d} \leq \psi(s/d). \tag{21}$$

Moreover, as shown in [13, Proposition 3.10] one has the upper bound $\delta \leq 2s\log(d/s) + 5s/4$, an estimate which is consistent with the classical computations contained in [21].

Proof of Theorem 1.2. We divide the proof into three steps.

Step 1. We first show that $\mathbf{x}_0$ is the unique solution of $(\mathbf{CP}_t)$ if and only if $C\cap\text{Null}(A_t) = \{\mathbf{0}\}$. Indeed, assume that $\mathbf{x}_0$ is the unique solution of $(\mathbf{CP}_t)$ and let $\mathbf{y} \in C\cap\text{Null}(A_t)$. Since $\mathbf{y} \in C$, there exists $\tau > 0$ such that $\mathbf{x} := \mathbf{x}_0 + \tau\mathbf{y}$ satisfies $\|\mathbf{x}\|_1 \leq \|\mathbf{x}_0\|_1$. Since $\mathbf{y} \in \text{Null}(A)$ one has $\mathbf{x} \in \mathcal{F}_t$. As $\mathbf{x}$ is feasible the inequality $\|\mathbf{x}\|_1 < \|\mathbf{x}_0\|_1$ would contradict the assumption that $\mathbf{x}_0$ solves $(\mathbf{CP}_t)$. On the other hand, the equality $\|\mathbf{x}\|_1 = \|\mathbf{x}_0\|_1$ would contradict the assumption that $\mathbf{x}_0$ solves $(\mathbf{CP}_t)$ uniquely if $\mathbf{x} \neq \mathbf{x}_0$. Hence $\mathbf{y} = 0$, so $C \cap \text{Null}(A_t) = \{\mathbf{0}\}$. Now assume that $C\cap\text{Null}(A_t) = \{\mathbf{0}\}$ and let $\mathbf{x}$ denote any solution of $(\mathbf{CP}_t)$ (note that such an $\mathbf{x}$ necessarily exists). Set $\mathbf{y} = \mathbf{x} - \mathbf{x}_0$. Of course, $\mathbf{y} \in \text{Null}(A_t)$. Moreover, by definition of $\mathbf{x}$ and that $\mathbf{x}_0 \in \mathcal{F}_t$ one has $\|\mathbf{x}\|_1 = \|\mathbf{x}_0 + \mathbf{y}\|_1 \leq \|\mathbf{x}_0\|_1$, implying in turn that $\mathbf{y} \in C$. Hence, $\mathbf{y} = \mathbf{0}$ and $\mathbf{x} = \mathbf{x}_0$, showing that $\mathbf{x}_0$ is the unique solution to $(\mathbf{CP}_t)$.

Step 2. We show[2] that $\text{Null}(A_t) \overset{\text{Law}}{=} Q(\mathbb{R}^{d-m_t} \times \{\mathbf{0}\})$ for Q a uniformly random $d \times d$ orthogonal matrix. Both $\text{Null}(A_t)$ and $Q(\mathbb{R}^{d-m_t} \times \{\mathbf{0}\})$ belong almost surely to the Grassmannian $G_{d-m_t}(\mathbb{R}^d)$, the set of all $(d - m_t)$-dimensional subspaces of $\mathbb{R}^d$. Defining the distance between two subspaces as the Hausdorff distance between the unit balls of those subspaces makes $G_{d-m_t}(\mathbb{R}^d)$ into a compact metric space. The metric is invariant under the action of the orthogonal group $O(d)$, and the action is transitive on $G_{d-m_t}(\mathbb{R}^d)$. Therefore, there exists a unique probability measure on $G_{d-m_t}(\mathbb{R}^d)$ that is invariant under the action of the orthogonal group. The law of the matrix A, having independent standard Gaussian entries, is orthogonaly invariant. Therefore, $P(\text{Null}(A_t) \in X) = P(\text{Null}(A_t) \in R(X))$ for any $R \in O(d)$ and any measurable subset $X \subset G_{d-m_t}(\mathbb{R}^d)$. On the other hand, it is clear that one also has $P(Q(\mathbb{R}^{d-m_t} \times \{\mathbf{0}\}) \in X) = P(Q(\mathbb{R}^{d-m_t} \times \{\mathbf{0}\}) \in R(X))$ for any $R \in O(d)$ and any measurable subset $X \subset G_{d-m_t}(\mathbb{R}^d)$. Therefore, the claim follows by uniqueness of the probability measure on $G_{d-m_t}(\mathbb{R}^d)$ invariant under the action of $O(d)$.

Step 3. Combining Steps 1 and 2 we find

$$P(\mathbf{x}_0 \text{ is the unique solution of } (\mathbf{CP}_t)) = P(C \cap Q(\mathbb{R}^{d-m_t} \times \{\mathbf{0}\}) = \{\mathbf{0}\}),$$

where Q is a uniformly random orthogonal matrix. On the other hand, with $\overline{C}$ denoting the closure of C,

$$P(C \cap Q(\mathbb{R}^{d-m_t} \times \{\mathbf{0}\}) = \{\mathbf{0}\}) = P(\overline{C} \cap Q(\mathbb{R}^{d-m_t} \times \{\mathbf{0}\}) = \{\mathbf{0}\}).$$

As a result of this subtle point, that follows from the discussion of touching probabilities located in [43, pp. 258–259], we may and will assume in the rest of the proof that C is closed. By the Crofton formula (see [3, formula (5.10)])

$$P(C \cap Q(\mathbb{R}^{d-m_t} \times \{\mathbf{0}\}) = \{\mathbf{0}\}) = 1 - 2h_{m_t+1}(C) \quad \text{where} \quad h_k(C) = \sum_{j=k, j-k \,\text{even}}^{d} v_j(C). \tag{22}$$

Combining (22) with the interlacing relation stated in [3, Proposition 5.9], that states

$$P(V \leq m_t - 1) \leq 1 - 2h_{m_t+1}(C) \leq P(V \leq m_t) \tag{23}$$

[2] This is a well-known result: we provide a proof for the sake of completeness.

yields

$$P(V \le m_t - 1) \le P\{\mathbf{x}_0 \text{ is the unique minimizer of } (\mathbf{CP}_t)\} \le P(V \le m_t).$$

But,

$$\begin{aligned} P(V \le m_t - 1) &= P\big(V \le \lfloor \delta + t\tau \rfloor - 1\big) \\ &\ge P\big(V \le \delta + t\tau - 1\big) = P\left(\frac{V-\delta}{\tau} \le t - \frac{1}{\tau}\right) \end{aligned}$$

and

$$\begin{aligned} P(V \le m_t) &= P\big(V \le \lfloor \delta + t\sqrt{\tau} \rfloor\big) \\ &\le P\big(V \le \delta + t\tau + 1\big) = P\left(\frac{V-\delta}{\tau} \le t + \frac{1}{\tau}\right). \end{aligned}$$

The conclusion now follows from (11), as well as from the fact that the standard Gaussian density on $\mathbb{R}$ is bounded by $(2\pi)^{-1/2}$. □

The next result provides natural sufficient conditions, in order for a sequence of linear inverse problems to display exact Gaussian fluctuations in the high-dimensional limit.

Theorem 1.3 (Asymptotic Gaussian phase transitions) *Let s_n, d_n, $n \ge 1$ be integer-valued sequences diverging to infinity, and assume that $s_n \le d_n$. For every n, let $\mathbf{x}_{n,0} \in \mathbb{R}^{d_n}$ be s_n-sparse, denote by C_n the descent cone of the ℓ_1 norm at $\mathbf{x}_{n,0}$ and write $\delta_n = \delta_{C_n} = E[V_{C_n}]$ and $\tau_n^2 = \tau_{C_n}^2 = \mathrm{Var}(V_{C_n})$. For every real number t, write*

$$m_{n,t} := \begin{cases} 1, & \text{if } \lfloor \delta_n + t\tau_n \rfloor < 1 \\ \lfloor \delta_n + t\tau_n \rfloor, & \text{if } \lfloor \delta_n + t\tau_n \rfloor \in [1, d_n] \\ d_n, & \text{if } \lfloor \delta_n + t\tau_n \rfloor > d_n \end{cases}.$$

For every n, let $A_{n,t} \in \mathbb{R}^{m_{n,t} \times d_n}$ be a random matrix with i.i.d. $\mathcal{N}(0,1)$ entries, let $\mathcal{F}_{n,t} = \{\mathbf{x} \in \mathbb{R}^{d_n} : A_{n,t}\mathbf{x} = A_{n,t}\mathbf{x}_{n,0}\}$, and consider the convex program

$$(\mathbf{CP}_{n,t}): \quad \min \|\mathbf{x}\|_1 \quad \textit{subject to } \mathbf{x} \in \mathcal{F}_{n,t}.$$

Assume that there exists $\rho \in (0,1)$ (independent of n) such that $s_n = \lfloor \rho d_n \rfloor$. Then, as $n \to \infty$, $\liminf_n \tau_n^2/\delta_n > 0$, and

$$P\{\mathbf{x}_0 \textit{ is the unique solution of } (\mathbf{CP}_{n,t})\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du + O\left(\frac{1}{\sqrt{\log \delta_n}}\right),$$

where the implicit constant in the term $O\left(\frac{1}{\sqrt{\log \delta_n}}\right)$ depends uniquely on ρ.

Proof. In view of the estimate (19), the conclusion will follow if we can prove the existence of a finite constant $\alpha(\rho) > 0$, uniquely depending on ρ, such that $\tau_n^2/\delta_n \ge \alpha(\rho)$ for n sufficiently large. The existence of such a $\alpha(\rho)$ is a direct consequence of the results stated in the forthcoming Proposition 4.1.

□

1.4.3 Second example: low-rank matrices

Let the inner product of two $m \times n$ matrices $\mathbf{U}$ and $\mathbf{V}$ be given by

$$\langle \mathbf{U}, \mathbf{V} \rangle = \mathrm{tr}(\mathbf{U}^T \mathbf{V}),$$

and recall that, for $\mathbf{X} \in \mathbb{R}^{m \times n}$, the Schatten 1 (or nuclear) norm is given by

$$\|\mathbf{X}\|_{S_1} = \sum_{i=1}^{\min(m,n)} \sigma_i(\mathbf{X}), \tag{24}$$

where $\sigma_1(\mathbf{X}) \geq \cdots \geq \sigma_{\min(m,n)}(\mathbf{X})$ are the singular values of $\mathbf{X}$. Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times np}$, partition $\mathbf{A}$ as $(\mathbf{A}_1, \ldots, \mathbf{A}_p)$ into blocks of sizes $m \times n$, and let $\mathcal{A}$ be the linear map from $\mathbb{R}^{m \times n}$ to $\mathbb{R}^p$ given by

$$\mathcal{A}(\mathbf{X}) = (\langle \mathbf{X}, \mathbf{A}_1 \rangle, \cdots, \langle \mathbf{X}, \mathbf{A}_p \rangle).$$

Now let $\mathbf{X}_0 \in \mathbb{R}^{m \times n}$ be a low rank matrix, and suppose that one observes

$$\mathbf{z} = \mathcal{A}(\mathbf{X}_0),$$

where the components of $\mathbf{A}$ are independent with distribution $\mathcal{N}(0,1)$. To recover $\mathbf{X}_0$ we consider the convex program

$$\min \|\mathbf{X}\|_{S_1} \quad \text{subject to} \quad \mathbf{X} \in \mathcal{F}, \quad \text{where} \quad \mathcal{F} = \{\mathbf{X} : \mathcal{A}(\mathbf{X}) = \mathbf{z}\}.$$

As $\mathcal{F}$ is the affine space $\mathbf{X}_0 + \mathrm{Null}(\mathcal{A})$, arguing as in the previous section one can show that $\mathbf{X}_0$ is recovered exactly if and only if $C \cap \mathrm{Null}(\mathcal{A}) = \{\mathbf{0}\}$ where $C = \mathcal{D}(\|\cdot\|_{S_1}, \mathbf{X}_0)$, the descent cone of the Schatten 1-norm at $\mathbf{X}_0$.

Furthermore, $\mathrm{Null}(\mathcal{A})$ is a subspace of $\mathbb{R}^{m \times n}$ of dimension $nm - p$, and is rotation invariant in the sense that for any $P \subset \{(i,j) : 1 \leq i \leq m, 1 \leq j \leq n\}$ of size p,

$$\mathrm{Null}(\mathcal{A}) = \mathcal{Q}(S_P)$$

where $\mathcal{Q}$ is a uniformly random orthogonal transformation on $\mathbb{R}^{m \times n}$, and

$$S_P = \{\mathbf{X} \in \mathbb{R}^{m \times n} : X_{ij} = 0 \quad \text{for all } (i,j) \in P\}.$$

Now considering the natural linear mapping between $\mathbb{R}^{m \times n}$ and $\mathbb{R}^{nm}$ that preserves inner product, one may apply the Crofton formula (5.10) and proceed as for the ℓ^1 descent cone as above in Section 1.4.3 to deduce low rank analogues of Theorems 1.2 and 1.3. In particular, for the latter we have the following result. As the Schatten 1-norm of a matrix and its transpose are equal, without loss of generality we assume that all matrices below have at least as many columns as rows.

Theorem 1.4 *For every $k \in \mathbb{N}$, let (n_k, m_k, r_k) be a triple of nonnegative integers depending on k. We assume that $n_k \to \infty$, $m_k/n_k \to \nu \in (0,1]$ and $r_k/m_k \to \rho \in (0,1)$ as $k \to \infty$, and that for every k the matrix $\mathbf{X}(k) \in \mathbb{R}^{m_k \times n_k}$ has rank r_k. Let*

$$C_k = \mathcal{D}(\|\cdot\|_{S_1}, \mathbf{X}(k)), \quad \delta_k = \delta(C_k) \quad \text{and} \quad \tau_k^2 = \mathrm{Var}(V_{C_k})$$

denote the descent cone of the Schatten 1-norm of $\mathbf{X}(k)$*, its statistical dimension, and the the variance of its conic intrinsic volume distribution, respectively. For every real number* t*, write*

$$p_{k,t} := \begin{cases} 1 & \text{if } \lfloor \delta_k + t\tau_k \rfloor < 1 \\ \lfloor \delta_k + t\tau_k \rfloor & \text{if } \lfloor \delta_k + t\tau_k \rfloor \in [1, m_k n_k] \\ m_k n_k & \text{if } \lfloor \delta_k + t\tau_k \rfloor > m_k n_k \end{cases}.$$

For every k*, let* $\mathbf{A}_{k,t} \in \mathbb{R}^{m_k \times n_k p_{k,t}}$ *be a random matrix with i.i.d.* $\mathcal{N}(0,1)$ *entries, let* $\mathcal{F}_{k,t} = \{\mathbf{X} : \mathcal{A}_{k,t}(\mathbf{X}) = \mathcal{A}_{k,t}(\mathbf{X}(k))\}$ *and consider the convex program*

$$(\mathbf{CP}_{k,t}) : \quad \min \|\mathbf{X}\|_{S_1} \quad \text{subject to} \quad \mathbf{X} \in \mathcal{F}_{k,t}.$$

Then, as $k \to \infty$*,* $\liminf \tau_k^2/\delta_k > 0$*, and*

$$P\left\{\mathbf{X}(k) \text{ is the unique solution of } (\mathbf{CP}_{k,t})\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du + O\left(\frac{1}{\sqrt{\log \delta_k}}\right),$$

where the implicit constant in the term $O\left(\frac{1}{\sqrt{\log \delta_k}}\right)$ *depends uniquely on* ν *and* ρ*.*

1.5 Connections with constrained statistical inference

Let $C \subset \mathbb{R}^d$ be a non-trivial closed convex cone, let $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, I_d)$ and fix a vector $\boldsymbol{\mu} \in \mathbb{R}^d$. When $\boldsymbol{\mu}$ is an element of C and $\mathbf{y} = \mathbf{g} + \boldsymbol{\mu}$ is regarded as a d-dimensional sample of observations, then the projection $\Pi_C(\mathbf{g} + \boldsymbol{\mu})$ is the *least square estimator* of $\boldsymbol{\mu}$ *under the convex constraint* C, and the norm $\|\boldsymbol{\mu} - \Pi_C(\mathbf{g} + \boldsymbol{\mu})\|$ measures the distance between this estimator and the true value of the parameter $\boldsymbol{\mu}$; the expectation $E\|\boldsymbol{\mu} - \Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2$ is often referred to as the L^2*-risk* of the least squares estimator.

Properties of least square estimators and associated risks have been the object of vigorous study for several decades; see e.g. [5, 9, 15, 16, 44, 45, 46, 47] for a small sample. Although several results are known about the norm $\|\boldsymbol{\mu} - \Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2$ (for instance, concerning concentration and moment estimates – see [15, 16] for recent developments), to our knowledge no normal approximation result is available for such a random variable, yet. We conjecture that our estimate (14) might represent a significant step in this direction. Note that, in order to make (14) suitable for applications, one would need explicit lower bounds on the variance of $\|\boldsymbol{\mu} - \Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2$ for a general $\boldsymbol{\mu}$, and for the moment such estimates seem to be outside the scope of any available technique: we prefer to think of this problem as a separate issue, and leave it open for future research.

We conclude by observing that, as explained e.g. in [19, 41] and in [42, Chapter 3], the *likelihood ratio test* (LRT) for the hypotheses $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \in C \backslash \{\mathbf{0}\}$ rejects H_0 when the projection $\|\Pi_C(\mathbf{y})\|^2$ of the data $\mathbf{y}$ on C is large. In this case, our results, together with the concentration estimates from [3, 32], provide information on the distribution of the test statistic under the null hypothesis. Similarly, the squared projection length $\|\Pi_{C^0}(\mathbf{y})\|^2$ onto the polar cone C^0 is the LRT statistic for the hypotheses $H_0 : \boldsymbol{\mu} \in C$ versus $H_1 : \boldsymbol{\mu} \in \mathbb{R}^d \backslash C$.

1.6 Plan

The paper is organised as follows. Section 2 deals with normal approximation results for the squared distance between a Gaussian vector and a general closed convex set. Section

3 contains total variation bounds to the normal, and our main CLTs for squared norms of projections onto closed convex cones, as well as for conic intrinsic volumes. In Section 4, we derive new upper and lower bounds on the variance of conic intrinsic volumes. Section 5 is devoted to explicit Berry-Esseen bounds for intrinsic volumes distributions, whereas the Appendix in Section 6 provides a self-contained discussion of Stein's method, Poincaré inequalities and associated estimates on a Gaussian space.

2 Gaussian Projections on Closed Convex Sets: normal approximations and concentration bounds

Let $C \subset \mathbb{R}^d$ be a closed convex set, let $\boldsymbol{\mu} \in \mathbb{R}^d$ and let $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, I_d)$ be a normal vector. In this section, we obtain a total variation bound to the normal, and a concentration inequality, for the centered squared distance between $\mathbf{g}+\boldsymbol{\mu}$ and C, that is, for

$$F = d^2(\mathbf{g}+\boldsymbol{\mu}, C) - E[d^2(\mathbf{g}+\boldsymbol{\mu}, C)], \tag{25}$$

where $d(\mathbf{x}, C)$ is given by (1). We also set $\sigma^2 = \mathrm{Var}(d^2(\mathbf{g}+\boldsymbol{\mu}, C)) = \mathrm{Var}(F)$. It is easy to verify that σ^2 is finite for any non empty closed convex set C, and equals zero if and only if $C = \mathbb{R}^d$. To exclude trivialities, we call a set C *non-trivial* if $\emptyset \subsetneq C \subsetneq \mathbb{R}^d$.

The following two lemmas are the key to our main result Theorem 2.1: their proofs are standard, and are provided for the sake of completeness.

Lemma 2.1 *Let C be a non empty closed convex subset of $\mathbb{R}^d$, and let $\Pi_C(\mathbf{x})$ the metric projection onto C. Then, Π_C and $I_d - \Pi_C$ are 1-Lipschitz continuous, and the Jacobian $\mathrm{Jac}(\Pi_C)(\mathbf{x}) \in \mathbb{R}^{d\times d}$ exists a.e. and satisfies*

$$\|(I_d - \mathrm{Jac}(\Pi_C)(\mathbf{x}))^T\,\mathbf{y}\| \le \|\mathbf{y}\| \quad \textit{for all } \mathbf{y} \in \mathbb{R}^d. \tag{26}$$

Proof: Since Π_C is a projection onto a non-empty closed convex set, by [36, p. 340] (see also B.3 of [3]), we have that

$$\|\Pi_C(\mathbf{v}) - \Pi_C(\mathbf{u})\| \le \|\mathbf{v}-\mathbf{u}\| \quad \text{for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^d,$$

that is, Π_C, and hence $I_d - \Pi_C$, are 1-Lipschitz. Bound (26) now follows by Rademacher's theorem and the fact that, on a Hilbert space, the operator norms of a matrix and that of its transpose are the same. □

Lemma 2.2 *Let C be a non-empty closed convex set $C \subset \mathbb{R}^d$, and let $\Pi_C(\mathbf{x})$ be the metric projection onto C. Then,*

$$\nabla d^2(\mathbf{x}, C) = 2\left(\mathbf{x} - \Pi_C(\mathbf{x})\right), \quad \mathbf{x} \in \mathbb{R}^d. \tag{27}$$

Proof: Fix an arbitrary $\mathbf{x}_0 \in \mathbb{R}^d$, and use the shorthand notation $\mathbf{v}_0 := \mathbf{x_0} - \Pi_C(\mathbf{x_0})$. Writing $\varphi(\mathbf{u}) := d^2(\mathbf{x}_0+\mathbf{u}, C) - d^2(\mathbf{x_0}, C) - 2\langle \mathbf{v}_0, \mathbf{u}\rangle$, relation (27) is equivalent to the statement that the mapping $\mathbf{u} \mapsto \varphi(\mathbf{u})$ is differentiable at $\mathbf{u} = \mathbf{0}$, and $\nabla\varphi(\mathbf{0}) = \mathbf{0}$. To prove this statement, we show the following stronger relation: for every $\mathbf{u} \in \mathbb{R}^d$, one has that $|\varphi(\mathbf{u})| \le \|\mathbf{u}\|^2$. Indeed, the inequality $\varphi(\mathbf{u}) \le \|\mathbf{u}\|^2$ follows from the fact that $d^2(\mathbf{x}_0+\mathbf{u}, C) \le \|\mathbf{u}+\mathbf{v}_0\|^2$ and

$d^2(\mathbf{x}_0, C) = \|\mathbf{v}_0\|^2$. To obtain the relation $\varphi(\mathbf{u}) \geq -\|\mathbf{u}\|^2$, just observe that $\mathbf{u} \mapsto \varphi(\mathbf{u})$ is a convex mapping vanishing at the origin, implying that $\varphi(\mathbf{u}) \geq -\varphi(-\mathbf{u}) \geq -\|-\mathbf{u}\|^2 = -\|\mathbf{u}\|^2$, where the second inequality is a consequence of the estimates deduced in the first part of the proof. This yields the desired conclusion. $\square$

We recall that the *total variation distance* between the laws of two random variables F and G is defined as

$$d_{TV}(F,G) = \sup_A |P(F \in A) - P(G \in A)|, \tag{28}$$

where the supremum runs over all the Borel sets $A \subset \mathbb{R}$. It is clear from the definition that $d_{TV}(F,G)$ is invariant under affine transformations, in the following sense: for any $a, b \in \mathbb{R}$ with $a \neq 0$, one has $d_{TV}(aF+b, aG+b) = d_{TV}(F,G)$. We say that F_n *converges to* F *in total variation* (in symbols, $F_n \overset{TV}{\to} F$) if $d_{TV}(F_n, F) \to 0$ as $n \to \infty$. Note that, if $F_n \overset{TV}{\longrightarrow} F$, then $F_n \overset{\text{Law}}{\longrightarrow} F$, where $\overset{\text{Law}}{\longrightarrow}$ denotes convergence in distribution.

The following statement provides a total variation bound for the normal approximation of the squared distance between a Gaussian vector with arbitrary mean and a closed convex set.

Theorem 2.1 *Let $C \subset \mathbb{R}^d$ be a non trivial closed convex set, F and σ^2 as in* (25)*, and $N \sim \mathcal{N}(0, \sigma^2)$. Then for $\mathbf{g} \sim \mathcal{N}(0, I_d)$ and $\boldsymbol{\mu} \in \mathbb{R}^d$,*

$$d_{TV}(F,N) \leq \frac{16\sqrt{Ed^2(\mathbf{g}, C - \boldsymbol{\mu})}}{\sigma^2}.$$

Proof: As the translation of a closed convex set is closed and convex, and

$$d^2(\mathbf{g} + \boldsymbol{\mu}, C) = d^2(\mathbf{g}, C - \boldsymbol{\mu})$$

we may replace C by $C - \boldsymbol{\mu}$ and assume (without loss of generality) that $\boldsymbol{\mu} = \mathbf{0}$. Using Lemma 6.2 and Theorem 6.1 in the Appendix we deduce that

$$d_{TV}(F,N) \leq \frac{2}{\sigma^2}\sqrt{\operatorname{Var}\left(\int_0^\infty e^{-t}\langle \nabla F(\mathbf{g}), \widehat{E}(\nabla F(\widehat{\mathbf{g}}_t))\rangle dt\right)}, \tag{29}$$

where

$$\widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1 - e^{-2t}}\widehat{\mathbf{g}},$$

with $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$, and the symbols E and $\widehat{E}$ denote, respectively, expectation with respect to $\mathbf{g}$ and $\widehat{\mathbf{g}}$. Set also $\mathbf{E} = E \otimes \widehat{E}$. Letting $H(\mathbf{g})$ denote the integral inside the variance in (29), by (27) we have

$$H(\mathbf{g}) = 4\int_0^\infty e^{-t}\langle \mathbf{g} - \Pi_C(\mathbf{g}), \widehat{E}[\widehat{\mathbf{g}}_t - \Pi_C(\widehat{\mathbf{g}}_t)]\rangle dt. \tag{30}$$

We bound the variance of $H(\mathbf{g})$ by the Poincaré inequality (see Theorem 6.2 in the Appendix), which states that

$$\operatorname{Var}(H(\mathbf{g})) \leq E\|\nabla H(\mathbf{g})\|^2. \tag{31}$$

Applying the product rule and differentiating under the integral (justified e.g. by a dominated convergence argument), using (30), (27) and Lemma 2.1 we obtain

$$\nabla H(\mathbf{g}) = 4\int_0^\infty e^{-t}\,(I_d - \mathrm{Jac}(\Pi_C)(\mathbf{g}))^T\ \widehat{E}[\widehat{\mathbf{g}}_t - \Pi_C(\widehat{\mathbf{g}}_t)]dt \qquad (32)$$
$$+4\int_0^\infty e^{-t}\widehat{E}[(I_d - \mathrm{Jac}(\Pi_C)(\widehat{\mathbf{g}}_t))^T]\ (\mathbf{g} - \Pi_C(\mathbf{g}))\,dt.$$

The expectation of the squared norm of the first term on the right-hand side of (32) is given by a factor of 16 multiplying

$$\begin{aligned} & E\|\int_0^\infty e^{-t}\,(I_d - \mathrm{Jac}(\Pi_C)(\mathbf{g}))^T\ \widehat{E}[\widehat{\mathbf{g}}_t - \Pi_C(\widehat{\mathbf{g}}_t)]dt\|^2 \\ \leq\ & E\int_0^\infty e^{-t}\|\,(I_d - \mathrm{Jac}(\Pi_C)(\mathbf{g}))^T\ \widehat{E}[\widehat{\mathbf{g}}_t - \Pi_C(\widehat{\mathbf{g}}_t)]\|^2 dt \\ \leq\ & E\int_0^\infty e^{-t}\|\widehat{E}[\widehat{\mathbf{g}}_t - \Pi_C(\widehat{\mathbf{g}}_t)]\|^2 dt \leq \mathbf{E}\int_0^\infty e^{-t}\|\widehat{\mathbf{g}}_t - \Pi_C(\widehat{\mathbf{g}}_t)\|^2 dt \\ =\ & E\int_0^\infty e^{-t}\|\mathbf{g} - \Pi_C(\mathbf{g})\|^2 dt = E\|\mathbf{g} - \Pi_C(\mathbf{g})\|^2 = Ed^2(\mathbf{g}, C), \end{aligned}$$

where we have used the triangle inequality, Lemma 2.1, Jensen's inequality, and the fact that $\widehat{\mathbf{g}}_t$ has the same distribution as $\mathbf{g}$ for all t. Applying a similar chain of inequalities, it is immediate to bound the expectation of the squared norm of the second summand in (32) by the same quantity. Applying (31) together with the inequality $\|\mathbf{x}+\mathbf{y}\|^2 \leq 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$, we therefore deduce that $\mathrm{Var}(H(\mathbf{g}))$ is bounded by $64Ed^2(\mathbf{g}, C)$. Substituting this bound into (29) yields the desired result. □

To conclude the section, we present a concentration bound for random variables of the type (25).

Theorem 2.2 *Let C be a closed convex set, and F given in* (25)*. Then,*

$$Ee^{\xi F} \leq \exp\left(\frac{2\xi^2 Ed^2(\mathbf{g}, C-\boldsymbol{\mu})}{1-2\xi}\right), \quad \text{for all } \xi < 1/2, \qquad (33)$$

and

$$P(F > t) \leq \exp\left(-Ed^2(\mathbf{g}, C-\boldsymbol{\mu})h\left(\frac{t}{2Ed^2(\mathbf{g}, C-\boldsymbol{\mu})}\right)\right) \quad \text{for all } t > 0 \qquad (34)$$

where

$$h(u) = 1 + u - \sqrt{1+2u}.$$

Proof: We reduce to the case $\boldsymbol{\mu} = \mathbf{0}$ as in the proof of Theorem 3.1. The arguments used in the proof of Lemma 4.9 of [32] for convex cones work essentially in the same way for projections on closed convex sets: we shall therefore provide only a quick sketch of the proof, and leave the details to the reader. Similarly to [32], for $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, I_d)$ we set

$$H(\mathbf{g}) = \xi Z \quad \text{for} \quad Z = d^2(\mathbf{g}, C) - Ed^2(\mathbf{g}, C),$$

and, using (27), we deduce that

$$\|\nabla H(\mathbf{g})\|^2 = 4\xi^2\|\mathbf{g} - \Pi_C(\mathbf{g})\|^2 = 4\xi^2 d^2(\mathbf{g}, C) = 4\xi^2 \left(Z + Ed^2(\mathbf{g}, C)\right).$$

Proceeding as in the proof of Lemma 4.9 in [32], with $Ed^2(\mathbf{g}, C)$ here replacing δ_C there, yields the bound (33) on the Laplace transform of F. Using the terminology defined in Section 2.4 of [8], we have therefore shown that F is sub-gamma on the right tail, with variance factor $4Ed^2(\mathbf{g}, C)$ and scale parameter 2. The conclusion now follows by the computations in that same section of [8]. □

Note that the estimate (34) is equivalent to the following bound: for every $t > 0$

$$P\Big(F > \sqrt{8Ed^2(\mathbf{g}, C - \mu)t} + 2t\Big) \le e^{-t}.$$

Remark 2.1 Let C be a closed convex cone. In [32, Lemma 4.9] it is proved that, for every $\xi < \frac{1}{2}$,

$$Ee^{\xi(\|\Pi_C(\mathbf{g})\|^2 - \delta_C)} \le \exp\left(\frac{2\xi^2\delta_C}{1 - 2\xi}\right), \tag{35}$$

where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, I_d)$ and (as before) $\delta_C = E[\|\Pi_C(\mathbf{g})\|^2]$. This estimate can be deduced by applying the general relation (33) to the polar cone C^0 in the case where $\boldsymbol{\mu} = \mathbf{0}$: indeed, by virtue of (3) one has that

$$\|\Pi_C(\mathbf{x})\|^2 = d^2(\mathbf{x}, C^0), \tag{36}$$

so that (35) follows immediately.

3 Steining the Steiner formula: CLTs for conic intrinsic volumes

3.1 Metric projections on cones

The goal of our analysis in this subsection is to demonstrate the following variation of Theorem 2.1.

Theorem 3.1 *Let $C \subset \mathbb{R}^d$ be a non-trivial closed convex cone and let*

$$F = \|\boldsymbol{\mu} - \Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2 - m, \quad \text{with} \quad m = E[\|\boldsymbol{\mu} - \Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2] \quad \text{and} \quad \sigma^2 = \mathrm{Var}(F).$$

Then for every $\boldsymbol{\mu} \in \mathbb{R}^d$,

$$d_{TV}(F, N) \le \frac{16}{\sigma^2}\left\{\sqrt{E\|\Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2} + 2\sqrt{m}\|\boldsymbol{\mu}\| + 3\|\boldsymbol{\mu}\|^2\right\}$$
$$\le \frac{16}{\sigma^2}\left\{\sqrt{m}(1 + 2\|\boldsymbol{\mu}\|) + 3\|\boldsymbol{\mu}\|^2 + \|\boldsymbol{\mu}\|\right\}.$$

Proof: Expanding F we obtain

$$F = \|\boldsymbol{\mu}\|^2 + \|\Pi_C(\mathbf{g}+\boldsymbol{\mu})\|^2 - 2\langle \boldsymbol{\mu}, \Pi_C(\mathbf{g}+\boldsymbol{\mu})\rangle - m.$$

The gradient of the first and last terms above are zero, while

$$\nabla \|\Pi_C(\mathbf{x}+\boldsymbol{\mu})\|^2 = 2\Pi_C(\mathbf{x}+\boldsymbol{\mu}) \quad \text{and} \quad \nabla\langle \boldsymbol{\mu}, \Pi_C(\mathbf{x}+\boldsymbol{\mu})\rangle = \text{Jac}^t(\Pi_C(\mathbf{x}+\boldsymbol{\mu}))\boldsymbol{\mu}.$$

the first equality following from (36) and (27), the second from the definition of the Jacobian, and Lemma 2.1, showing existence. We apply (94), and hence consider

$$G = \int_0^\infty e^{-t}\langle \nabla F(\mathbf{g}), \widehat{E}(\nabla F(\widehat{\mathbf{g}}_t))\rangle dt \quad \text{where} \quad \widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1-e^{-2t}}\widehat{\mathbf{g}},$$

with $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$. As before, we let E and $\widehat{E}$ be expectation with respect to $\mathbf{g}$ and $\widehat{\mathbf{g}}$, respectively, and write $\mathbf{E} = E \otimes \widehat{E}$.

Expanding out the inner product, we obtain

$$\begin{aligned} G \\ = \int_0^\infty e^{-t}\langle 2\Pi_C(\mathbf{g}+\boldsymbol{\mu}) - 2\text{Jac}^t(\Pi_C(\mathbf{g}+\boldsymbol{\mu}))\boldsymbol{\mu}, \widehat{E}\left(2\Pi_C(\widehat{\mathbf{g}}_t+\boldsymbol{\mu}) - 2\text{Jac}^t(\Pi_C(\widehat{\mathbf{g}}_t+\boldsymbol{\mu}))\boldsymbol{\mu}\right)\rangle dt \\ = 4(A_1 - A_2 - A_3 + A_4) \end{aligned}$$

where

$$\begin{aligned} A_1 &= \int_0^\infty e^{-t}\langle \Pi_C(\mathbf{g}+\boldsymbol{\mu}), \widehat{E}\left(\Pi_C(\widehat{\mathbf{g}}_t+\boldsymbol{\mu})\right)\rangle dt \\ A_2 &= \int_0^\infty e^{-t}\langle \Pi_C(\mathbf{g}+\boldsymbol{\mu}), \widehat{E}\left(\text{Jac}^t(\Pi_C(\widehat{\mathbf{g}}_t+\boldsymbol{\mu}))\boldsymbol{\mu}\right)\rangle dt \\ A_3 &= \int_0^\infty e^{-t}\langle \text{Jac}^t(\Pi_C(\mathbf{g}+\boldsymbol{\mu}))\boldsymbol{\mu}, \widehat{E}\left(\Pi_C(\widehat{\mathbf{g}}_t+\boldsymbol{\mu})\right)\rangle dt \quad \text{and} \\ A_4 &= \int_0^\infty e^{-t}\langle \text{Jac}^t(\Pi_C(\mathbf{g}+\boldsymbol{\mu}))\boldsymbol{\mu}, \widehat{E}\left(\text{Jac}^t(\Pi_C(\widehat{\mathbf{g}}_t+\boldsymbol{\mu}))\boldsymbol{\mu}\right)\rangle dt. \end{aligned}$$

Exploiting (94), as well as the fact that $\sigma^2 = E[G] = 4E[A_1 - A_2 - A_3 + A_4]$, we deduce that

$$\begin{aligned} d_{TV}(F,N) &\leq \frac{2}{\sigma^2}E|\sigma^2 - 4\left(A_1 - A_2 - A_3 + A_4\right)| \\ &\leq \frac{8}{\sigma^2}\sum_{i=1}^4 E|A_i - EA_i| \leq \frac{8}{\sigma^2}\left(B_1 + B_2 + B_3 + B_4\right), \end{aligned} \tag{37}$$

where

$$B_1 = \sqrt{\text{Var}(A_1)} \quad \text{and} \quad B_j = 2E|A_j| \quad \text{for } j = 2,3,4.$$

One has that

$$B_j \leq 2E\left(\|\Pi_C(\mathbf{g}+\boldsymbol{\mu})\|^2\right)^{1/2}\|\boldsymbol{\mu}\| \leq 2(\sqrt{m} + \|\mu\|)\|\mu\| \quad \text{for } j = 2,3 \text{ and} \quad B_4 \leq 2\|\boldsymbol{\mu}\|^2,$$

where we have applied the Cauchy-Schwarz and triangle inequality, as well as Lemma 2.1. On the other hand, one can write

$$A_1 = \int_0^\infty e^{-t} \langle \mathbf{g} + \boldsymbol{\mu} - \Pi_{C_0}(\mathbf{g} + \boldsymbol{\mu}), \widehat{E}\,(\widehat{\mathbf{g}}_t + \boldsymbol{\mu} - \Pi_{C_0}(\widehat{\mathbf{g}}_t + \boldsymbol{\mu}))\rangle dt,$$

and exploit exactly the same arguments used after formula (30) (with $\mathbf{g} + \boldsymbol{\mu}$ and $\widehat{\mathbf{g}}_t + \boldsymbol{\mu}$ replacing, respectively, $\mathbf{g}$ and $\widehat{\mathbf{g}}_t$) to deduce

$$B_1^2 = \mathrm{Var}(A_1) \le 4E[\|\mathbf{g} + \boldsymbol{\mu} - \Pi_{C^0}(\mathbf{g} + \boldsymbol{\mu})\|^2] = 4E[\|\Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2],$$

thus yielding the first claim of the theorem. The second follows from observing that

$$\sqrt{E[\|\Pi_C(\mathbf{g} + \boldsymbol{\mu})\|^2]} \le \sqrt{m} + \|\boldsymbol{\mu}\|,$$

where we have applied the triangle inequality with respect to the norm on $\mathbb{R}^d$-valued random vectors defined by the mapping $X \mapsto \sqrt{E\|X\|^2}$. □

3.2 Master Steiner formula and Main CLTs

As anticipated in the Introduction, the aim of this section is to obtain CLTs involving the conic intrinsic volume distributions $\{\mathcal{L}(V_{C_n})\}_{n\ge1}$ (see Section 1.2) associated with a sequence $\{C_n\}_{n\ge1}$ of closed convex cones. The strategy for achieving this goal will consist in connecting the intrinsic volume distribution of a closed convex cone $C \subset \mathbb{R}^d$ to the squared norm of the metric projection of $\mathbf{g} \sim \mathcal{N}(0, I_d)$ onto C.

Our main tool will be the powerful "Master Steiner Formula" stated in [32, Theorem 3.1 and Corollary 3.2]. Throughout the following, we use the symbol χ_j^2 to indicate the chi-squared distribution with j degrees of freedom, $j = 0, 1, 2, \ldots$.

Theorem 3.2 (Master Steiner Formula, see [32]) *Let $C \subset \mathbb{R}^d$ be a non-trivial closed convex cone, denote by C^0 its polar cone, and write $\{v_j : j = 0, ..., d\}$ to indicate the conic intrinsic volumes of C. Then, for every measurable mapping $f : \mathbb{R}_+^2 \to \mathbb{R}$,*

$$Ef(\|\Pi_C(\mathbf{g})\|^2, \|\Pi_{C^0}(\mathbf{g})\|^2) = \sum_{j=0}^{d} E[f(Y_j, Y'_{d-j})]v_j, \tag{38}$$

where $\{Y_j, Y'_j, j = 0, \ldots, d\}$ stands for a collection of independent random variables such that $Y_j, Y'_j \sim \chi_j^2, j = 0, 1 \ldots, d$.

Observe that, somewhat more compactly, we may also express (38) as the mixture relation

$$(\|\Pi_C(\mathbf{g})\|^2, \|\Pi_{C^0}(\mathbf{g})\|^2) \stackrel{\text{Law}}{=} (Y_{V_C}, Y'_{V_{C^0}}) \tag{39}$$

where the integer-valued random variable V_C is independent of $\{Y_j, Y'_j, j = 0, \ldots, d\}$, and $V_{C^0} = d - V_C$. Once combined with (3) and (9), in the case of a polyhedral cone $C \subset \mathbb{R}^d$, relation (39) reinforces the intuition that, given the dimension j of the face of C in which lies the projection $\Pi_C(\mathbf{g})$, the Gaussian vector $\mathbf{g}$ can be written as the sum of two independent

Gaussian elements, with dimension j and $d-j$ respectively, whose squared lengths follow the chi-squared distribution with the same respective degrees of freedom.

Fix a non-trivial closed convex cone $C \subset \mathbb{R}^d$. In order to connect the standardized limiting distributions of $\|\Pi_C(\mathbf{g})\|^2$ and V_C, we use (39) to deduce that

$$\|\Pi_C(\mathbf{g})\|^2 \stackrel{\text{Law}}{=} \sum_{i=1}^{V_C} X_i = W_C + V_C, \quad \text{where} \quad W_C = \sum_{i=1}^{V_C}(X_i - 1), \tag{40}$$

and $\{X_i\}_{i\geq 1}$ denotes a collection of i.i.d. χ_1^2 random variables, independent of V_C. Since $EX_i = 1$, we find $E\|\Pi_C(\mathbf{g})\|^2 = E[V_C]$, and letting G_C denote the squared projection length, we have

$$G_C = \|\Pi_C(\mathbf{g})\|^2 \quad \text{and} \quad \delta_C = E[G_C]. \tag{41}$$

Similarly, applying the conditional (on V_C) variance formula in (40) yields, with $\tau_C^2 := \mathrm{Var}(V_C)$ and $\sigma_C^2 := \mathrm{Var}(G_C)$, that

$$\mathrm{Var}(W_C) = 2\delta_C \quad \text{and} \quad \sigma_C^2 = \tau_C^2 + 2\delta_C, \tag{42}$$

the latter formula recovering Proposition 4.4 of [32]. Standardizing both sides of the first equality in (40) we therefore obtain that

$$\frac{G_C - \delta_C}{\sigma_C} \stackrel{\text{Law}}{=} \frac{\sqrt{2\delta_C}}{\sigma_C}\frac{W_C}{\sqrt{2\delta_C}} + \frac{\tau_C}{\sigma_C}\frac{V_C - \delta_C}{\tau_C}. \tag{43}$$

The following statement, that is partially a consequence of Theorem 3.1, shows that a total variation bound to the normal for the standardized projection can be expressed in terms of the mean δ_C only. We recall that C is self dual when $C^0 = -C$, and that in this case $\delta_C = d/2$ by (8).

Proposition 3.1 *We have that*

$$\tau_C^2 \leq 2\delta_C \quad \textit{and} \quad 2\delta_C \leq \sigma_C^2 \leq 4\delta_C. \tag{44}$$

In addition, with G_C and δ_C as in (41) *and $N \sim \mathcal{N}(0, \sigma_C^2)$, one has that*

$$d_{TV}(G_C - \delta_C, N) \leq \frac{16\sqrt{\delta_C}}{\sigma_C^2} \leq \frac{8}{\sqrt{\delta_C}} \quad \textit{and, if } C \textit{ is self dual, then} \quad d_{TV}(F, N) \leq \frac{8\sqrt{2}}{\sqrt{d}}. \tag{45}$$

Proof: Theorem 4.5 of [32] yields the first bound in (44). The second bound in (44) now follows from the second relation stated in (42). The first inequality in (45) follows from the first inequality of Theorem 3.1 by setting $\boldsymbol{\mu} = \mathbf{0}$, and the remaining claims by the lower bound on σ_C^2 in (44). □

Remark 3.1 The first estimate in (45) can also be directly obtained from Theorem 2.1 by specializing it to the case $\boldsymbol{\mu} = \mathbf{0}$. Indeed, writing C^0 for the dual cone of C, one has that $\|\Pi_C(\mathbf{g})\|^2 = d^2(\mathbf{g}, C^0)$: the conclusion then follows by applying Theorem 2.1 to the random variable $F = d^2(\mathbf{g}, C^0) - Ed^2(\mathbf{g}, C^0)$.

We now consider normal limits for the conic intrinsic volumes. Explicit Berry-Esseen bounds will be presented in Theorem 5.1.

Theorem 3.3 *Let $\{d_n : n \geq 1\}$ be a sequence of non-negative integers and let $\{C_n \subset \mathbb{R}^{d_n} : n \geq 1\}$ be a collection of non-trivial closed convex cones such that $\delta_{C_n} \to \infty$. For notational simplicity, write δ_n, σ_n, τ_n, etc., instead of δ_{C_n}, σ_{C_n}, τ_{C_n}, etc., respectively. Then,*

1.
$$d_{TV}\left(\frac{W_n}{\sqrt{2\delta_n}}, N\right) \leq \frac{2\sigma_n}{\delta_n}, \qquad \text{for all } n \geq 1, \tag{46}$$

   where $N \sim \mathcal{N}(0,1)$, and

$$\frac{W_n}{\sqrt{2\delta_n}} \xrightarrow{TV} \mathcal{N}(0,1), \qquad \text{as } n \to \infty.$$

2. *The two random variables $\frac{W_n}{\sqrt{2\delta_n}}$ and $\frac{V_n - \delta_n}{\tau_n}$ are asymptotically independent in the following sense: if $\{n_k : k \geq 1\}$ is a subsequence diverging to infinity and*

$$\frac{V_{n_k} - \delta_{n_k}}{\tau_{n_k}}, \qquad k \geq 1, \tag{47}$$

   converges in distribution to some random variable Z, then

$$\left(\frac{W_{n_k}}{\sqrt{2\delta_{n_k}}}, \frac{V_{n_k} - \delta_{n_k}}{\tau_{n_k}}\right) \xrightarrow{\text{Law}} (N, Z),$$

   where N has the $\mathcal{N}(0,1)$ distribution and is stochastically independent of Z.

3. *If*

$$\frac{V_n - \delta_n}{\tau_n} \xrightarrow{\text{Law}} \mathcal{N}(0,1), \qquad \text{as } n \to \infty, \tag{48}$$

   then

$$\frac{G_n - \delta_n}{\sigma_n} \xrightarrow{\text{Law}} \mathcal{N}(0,1), \qquad \text{as } n \to \infty, \tag{49}$$

   and the converse implication holds if $\liminf_{n\to\infty} \tau_n^2/\delta_n > 0$.

Remark 3.2 Proposition 3.1 shows that, if $\delta_n \to \infty$, then (49) holds and, provided

$$\liminf \tau_n^2/\delta_n > 0,$$

relation (48) also takes place by virtue of Part 3 of Theorem 3.3. This chain of implications, which is one of the main achievements of the present paper, corresponds to the statement of Theorem 1.1 in the Introduction (exception made for the Berry-Esseen bound). Results analogous to Part 3 of Theorem 3.3 (involving general mixtures of independent χ^2 random variables) can be found in Dykstra [25].

Proof of Theorem 3.3: Throughout the proof, and when there is no risk of confusion, we drop the subscript n for readability.
(Point 1) By [31], a variable X with a $\Gamma(\alpha, \lambda)$ distribution satisfies

$$E[Xf'(X) + (\alpha - \lambda X)f(X)] = 0$$

for all locally absolutely continuous functions f for which these expectations exist. Hence, since conditionally on V, W has a centered chi-squared distribution with V degrees of freedom, one verifies immediately that, for every Lipschitz mapping $\phi : \mathbb{R} \to \mathbb{R}$,

$$E\left[\frac{W}{\sqrt{2\delta}}\phi\left(\frac{W}{\sqrt{2\delta}}\right)\right] = \frac{1}{\delta}E\left[(W+V)\phi'\left(\frac{W}{\sqrt{2\delta}}\right)\right].$$

Stein's inequality (89) in the Appendix therefore yields that

$$d_{TV}\left(\frac{W}{\sqrt{2\delta}}, N\right) \le \frac{2}{\delta}E|W+V-\delta| \le \frac{2}{\delta}\sqrt{2\delta+\tau^2} = \frac{2\sigma}{\delta} \le \frac{4}{\sqrt{\delta}} \to 0$$

using (44) together with the fact that $\delta \to \infty$ by assumption.
(Point 2) Let η, ξ be arbitrary real numbers. Using that the conditional distribution $\mathcal{L}(W|V)$ corresponds to a centered chi-squared distribution with V degrees of freedom, we have

$$E[e^{i\eta W}|V] = \frac{e^{-i\eta V}}{(1-2i\eta)^{V/2}} = \exp(-V(i\eta + (1/2)\log(1-2i\eta)).$$

Conditioning on V, we obtain the following expression for the joint characteristic function of $W/\sqrt{2\delta}$ and $(V-\delta)/\tau$:

$$\begin{aligned}\psi(\eta,\xi) := E\left[e^{i\eta\frac{W}{\sqrt{2\delta}} + i\xi\frac{V-\delta}{\tau}}\right] &= E[e^{-V(i\eta/\sqrt{2\delta}+(1/2)\log(1-2i\eta/\sqrt{2\delta}))+i\xi\frac{V-\delta}{\tau}}] \\ &= e^{\delta\left[-i\eta/\sqrt{2\delta}-\frac{1}{2}\log(1-2i\eta/\sqrt{2\delta})\right]} \times E\left[e^{\frac{V-\delta}{\tau}\left(i\xi - i\eta\tau/\sqrt{2\delta}-\frac{\tau}{2}\log(1-2i\eta/\sqrt{2\delta})\right)}\right]. \quad (50)\end{aligned}$$

As $\delta \to \infty$, one has clearly that

$$\delta\left[-i\eta/\sqrt{2\delta} - \frac{1}{2}\log(1-2i\eta/\sqrt{2\delta})\right] \to -\eta^2/2.$$

Moreover, since $\tau/\delta \le \sqrt{2/\delta} \to 0$ by (44), we obtain as well that

$$i\xi - i\eta\tau/\sqrt{2\delta} - \tau/2\log(1-2i\eta/\sqrt{2\delta}) \to i\xi.$$

Hence, letting ψ_Z be the characteristic function of the limiting distribution Z of the sequence in (47), we infer that

$$\psi(\eta,\xi) \to e^{-\eta^2/2}\psi_Z(\xi),$$

thus yielding the desired conclusion.
(Point 3) For both implications it is sufficient to show that, for every subsequence n_k, $k \ge 1$, of $1, 2, 3, \ldots$, there exists a further subsequence n_{k_l}, $l \ge 1$, along which the claimed distributional convergence holds. By (44), $0 \le \liminf \tau^2/\delta \le \limsup \tau^2/\delta \le 2$, so for every $n_k, k \ge 0$

there exists a further subsequence $n_{k_l}, l \geq 1$, along which τ^2/δ converges to a limit, say r, in $[0,2]$. Hence, along $n_{k_l}, l \geq 1$, we obtain

$$\sqrt{2\delta}/\sigma = \sqrt{2\delta/(2\delta+\tau^2)} \to \sqrt{\frac{2}{2+r}} \quad \text{and} \quad \tau/\sigma \to \sqrt{\frac{r}{2+r}}.$$

Assume first that (48) is satisfied. Then, according to (43) and Point 2 in the statement, one has that $\frac{G-\delta}{\sigma}$ converges in distribution along n_{k_l}, $l \geq 1$, to $\sqrt{\frac{2}{2+r}}N + \sqrt{\frac{r}{2+r}}Z$, where N and Z are two independent $\mathcal{N}(0,1)$ random variables, and we conclude that (49) holds along n_{k_l}, $l \geq 1$. Now assume that (49) is satisfied and that $\liminf_{n\to\infty} \tau_n^2/\delta_n > 0$; in this case, we may assume that τ^2/δ converges to $r \in (0,2]$ along n_{k_l}. Observe that, by virtue of boundedness in L^2, the family $\{\frac{V-\delta}{\tau}\}$ is tight. Consider now a further subsequence of n_{k_l} along which $\frac{V-\delta}{\tau}$ converges in distribution to, say, Z. According to Point 2 we know that the elements of the limiting pair (N,Z) are independent, and by (49) the sum $\sqrt{\frac{2}{2+r}}N + \sqrt{\frac{r}{2+r}}Z$ is normal. By Cramér's theorem we conclude that both N and Z are normally distributed, yielding the desired conclusion.

□

As Table 1 below shows, Theorem 1.1 yields a central limit theorem for G_n and V_n for the most common examples of convex cones that appear in practice. The last two rows refer to C_A and C_{BC}, chambers of finite reflection groups acting on $\mathbb{R}^d$, which are the normal cones to the permutahedon, and signed permutahedron, respectively. For further definitions and properties, see e.g. [3, 32] and the references therein.

Cone	**Ambient**	δ	τ^2
Orthant	$\mathbb{R}^d$	$\frac{1}{2}d$	$\frac{1}{4}d$
Real Positive Semi-Definite Cone	$\mathbb{R}^{n^2}$	$\frac{1}{4}n(n+1)$	$\simeq \left(\frac{4}{\pi^2}-\frac{1}{4}\right)n^2$
Circ_α	$\mathbb{R}^d$	$d\sin^2\alpha$	$\frac{1}{2}(d-2)\sin^2(2\alpha)$
C_A	$\mathbb{R}^d$	$\sum_{k=1}^d k^{-1}$	$\sum_{k=1}^d k^{-1}(1-k^{-1})$
C_{BC}	$\mathbb{R}^d$	$\sum_{k=1}^d \frac{1}{2}k^{-1}$	$\sum_{k=1}^d \frac{1}{2}k^{-1}(1-\frac{1}{2}k^{-1})$

Table 1: *Some common cones*

Remark 3.3 The first three lines of Table 1 are taken from Table 6.1 of [32]. The means for the permutathedron and signed permutahedron are from Section D.4. of [3]. The expressions for the variances τ^2 associated with the permutathedron and signed permutahedron can be deduced as follows. Let

$$q(s) = \sum_{k=0}^d v_k s^k,$$

be the probability generating function of the distribution of $V = V_{C_d}$. We have

$$q'(1) = EV \quad \text{and} \quad q''(1) = EV(V-1)$$

so in particular,

$$\mathrm{Var}(V) = q'(1) + q''(1) - q'(1)^2 = q'(1) + \log q(s)''|_{s=1}.$$

For the permutahedron, one can use Theorem 3 of [23, Theorem 3] (see also the first line of Table 10 of [18]) to deduce that

$$q(s) = \frac{1}{d!}\prod_{k=1}^{d}(s+k-1) \quad \text{so that} \quad \log q(s) = -\log d! + \sum_{k=1}^{d}\log(s+k-1).$$

Hence,

$$EV = q'(1) = \log q(s)'|_{s=1} = \left(\sum_{k=1}^{d}\frac{1}{s+k-1}\right)_{s=1} = \sum_{k=1}^{d}\frac{1}{k},$$

and

$$\mathrm{Var}(V) \;=\; q'(1) \,+\, \log q(s)''|_{s=1} \;=\; q'(1) \,-\, \left(\sum_{k=1}^{d}\frac{1}{(s+k-1)^2}\right)_{s=1} \;=\; \sum_{k=1}^{d}\left(\frac{1}{k}-\frac{1}{k^2}\right).$$

The calculation for the signed permutahedron is the analogous, but now one has to use [6, formula (3)]; see also the second line of Table 10 of [18].

We conclude the section with a statement showing that the rate of convergence appearing in (46) is often optimal. Also, by suitably adapting the techniques introduced in [33], one can deduce precise information about the local asymptotic behaviour of the difference $P[W_n/\sqrt{2\delta_n} \le x] - P[N \le x]$, where $x \in \mathbb{R}$ and $N \sim \mathcal{N}(0,1)$.

Proposition 3.2 *Let the notation and assumptions of Theorem 3.3 prevail, and assume further that $\tau_n^2/\delta_n \to r$ for some $r \ge 0$, as $n \to \infty$. Then, for every $x \in \mathbb{R}$ one has that, as $n \to \infty$,*

$$\frac{\delta_n}{\sigma_n}\left(P\left[\frac{W_n}{\sqrt{2\delta_n}} \le x\right] - P[N \le x]\right) \longrightarrow -\sqrt{\frac{2}{18+9r}}(x^2-1)\frac{e^{-x^2/2}}{\sqrt{2\pi}}. \tag{51}$$

As a consequence, there exists a constant $c \in (0,1)$ (independent of n) such that, for all n sufficiently large,

$$c\frac{\sigma_n}{\delta_n} \le d_{Kol}\left(\frac{W_n}{\sqrt{2\delta_n}}, N\right) \le d_{TV}\left(\frac{W_n}{\sqrt{2\delta_n}}, N\right). \tag{52}$$

Proof. Fix $x \in \mathbb{R}$. It suffices to show that, for every sequence $n_k, k \ge 1$ diverging to infinity, there exists a subsequence $n_{k_l}, l \ge 1$ along which the convergence (51) takes place. Let then $n_k \to \infty$ be an arbitrary divergent sequence. By L^2-boundedness, the collection of the laws of the random variables $\frac{V_{n_k}-\delta_{n_k}}{\tau_{n_k}}$, $k \ge 1$ is tight, and therefore there exists a subsequence n_{k_l} such that $\frac{V_{n_{k_l}}-\delta_{n_{k_l}}}{\tau_{n_{k_l}}}$ converges in distribution to some random variable Z. Exploiting again L^2-boundedness, which additionally implies uniform integrability, one sees immediately that Z is necessarily centered. Now let $\phi_x = \phi_h$ denote the solution (90) to the Stein equation

(88) for the indicator test function $h = \mathbf{1}_{(-\infty,x]}$. By (2.8) of [17], ϕ_x is Lipschitz, so as in part 1 of the proof of Theorem (3.3), we have

$$E\left[\frac{W_n}{\sqrt{2\delta_n}}\phi\left(\frac{W_n}{\sqrt{2\delta_n}}\right)\right] = \frac{1}{\delta_n}E\left[(W_n+V_n)\phi'\left(\frac{W_n}{\sqrt{2\delta_n}}\right)\right].$$

Hence, by (88), we obtain

$$\begin{aligned} P\left[\frac{W_n}{\sqrt{2\delta_n}} \le x\right] - P[N \le x] &= E\left[\phi_x'\left(\frac{W_n}{\sqrt{2\delta_n}}\right) - \frac{W_n}{\sqrt{2\delta_n}}\phi_x\left(\frac{W_n}{\sqrt{2\delta_n}}\right)\right] \\ &= \frac{1}{\delta_n}E\left[\phi_x'\left(\frac{W_n}{\sqrt{2\delta_n}}\right)(\delta_n - W_n - V_n)\right]. \end{aligned}$$

Dividing both sides by σ_n/δ_n, one obtains

$$\frac{\delta_n}{\sigma_n}\left(P\left[\frac{W_n}{\sqrt{2\delta_n}} \le x\right] - P[N \le x]\right) = E\left[\phi_x'\left(\frac{W_n}{\sqrt{2\delta_n}}\right)\left(-\frac{\tau_n}{\sigma_n}\frac{V_n-\delta_n}{\tau_n} - \frac{\sqrt{2\delta_n}}{\sigma_n}\frac{W_n}{\sqrt{2\delta_n}}\right)\right].$$

In view of Parts 1 and 2 of Theorem 3.3, of formula (42), and of the fact that Z is centered, one has, along the subsequence n_{k_l}, that

$$\frac{\delta_n}{\sigma_n}\left(P\left[\frac{W_n}{\sqrt{2\delta_n}} \le x\right] - P[N \le x]\right) \to -\sqrt{\frac{2}{2+r}}E[\phi_x'(N)N],$$

where $N \sim \mathcal{N}(0,1)$. We can now use e.g. [33, formula (2.20)] to deduce that, for every real x,

$$E[\phi_x'(N)N] = \frac{(x^2-1)}{3} \times \frac{e^{-x^2/2}}{\sqrt{2\pi}},$$

from which the desired conclusion follows at once.

□

In the next section, we shall prove general upper and lower bounds for the variance of conic intrinsic volumes. In particular, these results will apply to two fundamental examples that are *not* covered by the estimates contained in Table 1, and that are key in convex recovery of sparse vectors and low rank matrices: the descent cone of the ℓ_1 norm, and of the Schatten 1-norm.

4 Bounds on the variance of conic intrinsic volumes

4.1 Upper and lower bounds

Fix $d \ge 1$, let $C \subset \mathbb{R}^d$ be a closed convex cone, and let $V = V_C$ be the integer-valued random variable associated with C via relation (6). As before, we will denote by $\mathbf{g} \sim \mathcal{N}(0, I_d)$ a d-dimensional standard Gaussian random vector. The following statement provides useful new upper and lower bounds on the variance of V_C.

Theorem 4.1 *Define*

$$v := \|E[\Pi_C(\mathbf{g})]\|^2 \quad \textit{and} \quad b := \sqrt{d\delta_C/2}, \tag{53}$$

where δ_C is the statistical dimension of C. Then, one has the following estimates:

$$\frac{\min(v^2, 4b^2)}{b} \leq \mathrm{Var}(V_C) \leq 2v. \tag{54}$$

Remark 4.1 (a) In view of the orthogonal decomposition (3) and of the fact that $\mathbf{g}$ is a centered Gaussian vector, one has that

$$v = -\langle E[\Pi_C(\mathbf{g})], E[\Pi_{C^0}(\mathbf{g})]\rangle = \|E[\Pi_{C^0}(\mathbf{g})]\|^2, \tag{55}$$

where C^0 is the polar of C. Moreover, since the mapping $x \mapsto \min(x^2, 4b^2)$ is increasing on $\mathbb{R}_+$, one has also that $\mathrm{Var}(V_C) \geq \min(x^2, 4b^2)/b$, for every $0 \leq x < v$.

(b) An elementary consequence of (54) is the intuitive fact that a closed convex cone C is a subspace if and only if $v = 0$, that is, if and only if $\Pi_C(\mathbf{g})$ is a centered random vector.

In order to prove Theorem 4.1, we need the following auxiliary result.

Lemma 4.1 (Steiner form of the conic variance) *For any closed convex cone C,*

$$\mathrm{Var}(V_C) = -\mathrm{Cov}(\|\Pi_C(\mathbf{g})\|^2, \|\Pi_{C^0}(\mathbf{g})\|^2).$$

Proof: From the Master Steiner Formula (38), we deduce that

$$\mathrm{Cov}(\|\Pi_C(\mathbf{g})\|^2, \|\Pi_{C^0}(\mathbf{g})\|^2) = \sum_{j=0}^{d} E[Y_j Y'_{d-j}]v_j - \delta_C(d-\delta_C) = \sum_{j=0}^{d} j(d-j)v_j - \delta_C(d-\delta_C),$$

and the conclusion follows from straightforward simplifications. □

Proof of Theorem 4.1. (Upper bound) Using (42), one has that $\mathrm{Var}(V_C) = \mathrm{Var}(\|\Pi_C(\mathbf{g})\|^2) - 2\delta_C$. Now we apply Lemma 6.2 and Theorem 6.2 in the Appendix to the mapping $F(\mathbf{g}) = \|\Pi_C(\mathbf{g})\|^2 = d^2(\mathbf{g}, C^0)$, to obtain that

$$\mathrm{Var}(\|\Pi_C(\mathbf{g})\|^2) \leq \frac{1}{2}E[\|\nabla F(\mathbf{g})\|^2] + \frac{1}{2}\|E[\nabla F(\mathbf{g})]\|^2 = 2\delta_C + 2v,$$

where we have used the fact that $\nabla\|\Pi_C(\mathbf{g})\|^2 = 2\Pi_C(\mathbf{g})$, following from (36) and (27). (Lower bound) For every $t > 0$, define $\hat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1-e^{-2t}}\hat{\mathbf{g}}$, where $\hat{\mathbf{g}}$ is an independent copy of $\mathbf{g}$. The crucial step is to apply relation (96) in the Appendix to the random variables $F(\mathbf{g}) = \|\Pi_C(\mathbf{g})\|^2$ and $G(\mathbf{g}) = \|\Pi_{C^0}(\mathbf{g})\|^2$, obtaining that, for any $a \geq 0$,

$$\mathrm{Cov}(\|\Pi_C(\mathbf{g})\|^2, \|\Pi_{C^0}(\mathbf{g})\|^2) = 4\mathbf{E}\int_0^\infty e^{-t}\langle \Pi_C(\mathbf{g}), \Pi_{C^0}(\widehat{\mathbf{g}}_t)\rangle dt \leq 4\mathbf{E}\int_a^\infty e^{-t}\langle \Pi_C(\mathbf{g}), \Pi_{C^0}(\widehat{\mathbf{g}}_t)\rangle dt,$$

where we have used the definition of the polar cone C^0 as that set that has non-positive inner product with all elements of C, and $\mathbf{E}$ indicates expectation over $\mathbf{g}$ and $\hat{\mathbf{g}}$. Now write

$$\langle \Pi_C(\mathbf{g}), \Pi_{C^0}(\widehat{\mathbf{g}}_t)\rangle = \langle \Pi_C(\mathbf{g}), \Pi_{C^0}(\widehat{\mathbf{g}})\rangle + \langle \Pi_C(\mathbf{g}), \Pi_{C^0}(\widehat{\mathbf{g}}_t) - \Pi_{C^0}(\widehat{\mathbf{g}})\rangle. \tag{56}$$

For the second term, using the fact that the projection $\Pi_{C^0}(\mathbf{x})$ is 1-Lipschitz,

$$|\mathbf{E}\langle \Pi_C(\mathbf{g}), \Pi_{C^0}(\widehat{\mathbf{g}}_t) - \Pi_{C^0}(\widehat{\mathbf{g}})\rangle| \le \mathbf{E}\left(\|\Pi_C(\mathbf{g})\| \, \|\Pi_{C^0}(\widehat{\mathbf{g}}_t) - \Pi_{C^0}(\widehat{\mathbf{g}})\|\right)$$
$$\le \mathbf{E}\left(\|\Pi_C(\mathbf{g})\| \, \|\widehat{\mathbf{g}}_t - \widehat{\mathbf{g}}\|\right) \le \sqrt{\delta(C)\mathbf{E}\|\widehat{\mathbf{g}}_t - \widehat{\mathbf{g}}\|^2} \le \sqrt{2d\delta(C)}e^{-t} = 2be^{-t},$$

as

$$\mathbf{E}\|\widehat{\mathbf{g}}_t - \widehat{\mathbf{g}}\|^2 = \mathbf{E}\|e^{-t}\mathbf{g} + (\sqrt{1-e^{-2t}} - 1)\widehat{\mathbf{g}}\|^2 = 2\left(1 - \sqrt{1-e^{-2t}}\right) d \le 2e^{-2t}d.$$

Now use Lemma 4.1: multiplying (56) by e^{-t}, integrating over $[a, \infty)$ and taking expectation yields

$$-\mathrm{Var}(V_C) \le 4\mathbf{E}\int_a^\infty e^{-t}\langle \Pi_C(\mathbf{g}), \Pi_{C^0}(\widehat{\mathbf{g}}_t)\rangle dt \le 4e^{-a}(-v + be^{-a}),$$

showing that, for every $y \in [0,1]$,

$$\mathrm{Var}(V_C) \ge 4y(v - by).$$

The claim now follows by maximizing the mapping $y \mapsto 4y(v - by)$ on $[0,1]$. □

In the next two sections, we shall apply the variance bounds of Theorem (4.1) to the descent cones of the ℓ_1 and Schatten-1 norms.

4.2 The descent cone of the ℓ_1 norm at a sparse vector

The next result provides the key for completing the proof of Theorem 1.3. In the body of the proofs in this subsection and the next, given two positive sequences a_n, b_n, $n \ge 1$, we shall use the notation $a_n \approx b_n$ to indicate that $a_n/b_n \to 1$, as $n \to \infty$.

Proposition 4.1 *Let the assumptions and notation of Theorem 1.3 prevail (in particular, $s_n = \lfloor \rho d_n \rfloor$ for a fixed $\rho \in (0,1)$). Then,*

$$\liminf_n \frac{\tau_n^2}{\delta_n} \ge \sqrt{2}\min\left\{ 2\sqrt{\frac{1}{\psi(\rho)}} \; ; \; \frac{\rho^2\gamma(\rho)^4}{\psi(\rho)^{3/2}} \right\} > 0, \tag{57}$$

where $\psi(\rho)$ is defined in (20) *and $\gamma = \gamma(\rho) > 0$ is the unique solution to the stationary equation*

$$\sqrt{\frac{2}{\pi}}\int_\gamma^\infty \left(\frac{u}{\gamma} - 1\right) e^{-u^2/2} du = \frac{\rho}{1-\rho}.$$

Proof. Since the ℓ_1 norm is invariant with respect to signed permutations, we can assume – without loss of generality – that the sparse vector $\mathbf{x}_{n,0}$ has the form $(x_{n,1}, ..., x_{n,s_n}, 0, ..., 0)$, $x_{n,j} > 0$. Also, by virtue of the estimate (21), one has that $\delta_n \approx s_n\psi(\rho)/\rho$. Now write

$$v_n := \|E[\Pi_{C_n}(\mathbf{g}_n)]\|^2 = \|E[\Pi_{C_n^0}(\mathbf{g}_n)]\|^2, \quad n \ge 1$$

where we have used (55), and: (i) C_n is the descent cone of the ℓ_1 norm at $\mathbf{x}_{n,0}$, (ii) C_n^0 is the polar cone of C_n, and (iii) $\mathbf{g}_n = (g_1, ..., g_{d_n})$ stands for a d_n-dimensional standard centered Gaussian vector.

Using the lower bound in (54) together with some routine simplifications, it is easily seen that relation (57) is established if one can show that

$$\liminf_n \frac{v_n}{s_n} \geq \gamma(\rho)^2. \tag{58}$$

To accomplish this task, we first reason as in [3, Section B.1] to deduce that, for every n, the polar cone C_n^0 has the form $\bigcup_{\gamma\geq 0}\gamma\cdot\partial\|\mathbf{x}_{n,0}\|_1$, where $\partial\|\mathbf{x}_{n,0}\|_1$ denotes the subdifferential of the ℓ^1 norm at $\mathbf{x}_{n,0}$, that collection of vectors $\mathbf{z} = (z_1, ..., z_{d_n}) \in \mathbb{R}^{d_n}$ such that $z_1 = \cdots = z_{s_n} = 1$ and $|z_j| \leq 1$, for every $j = s_n + 1, ..., d_n$. As a consequence, for every n, the projection $\Pi_{C_n^0}(\mathbf{g}_n)$ has the form

$$\Pi_{C_n^0}(\mathbf{g}) = (\gamma_{\rho,n}, ..., \gamma_{\rho,n}, \star, ..., \star),$$

where the symbol '$\star$' stands for entries whose exact values are immaterial for our discussion, and $\gamma_{\rho,n} > 0$ is defined as the unique random point minimising the mapping $\gamma \mapsto F_{n,\rho}(\gamma) := \sum_{i=1}^{s_n}(g_i-\gamma)^2 + \sum_{i=s_n+1}^{d_n}(|g_i|-\gamma)_+^2$ over $\mathbb{R}_+$. This shows that $v_n \geq s_n E[\gamma_{\rho,n}]^2$: as a consequence, in order to prove that (58) holds it suffices to check that

$$\liminf_n E[\gamma_{\rho,n}] \geq \gamma(\rho). \tag{59}$$

The key point is now that $\gamma_{\rho,n}$ is (trivially) the unique minimiser of the *normalised* mapping $\gamma \mapsto \frac{1}{d_n}F_{n,\rho}(\gamma)$, and also that, in view of the strong law of large numbers, for every $\gamma \geq 0$,

$$\frac{1}{d_n}F_{n,\rho}(\gamma) \longrightarrow H_\rho(\gamma) := \left\{\rho(1+\gamma^2) + (1-\rho)E[(|N|-\gamma)_+^2]\right\}, \quad \text{as } n \to \infty, \tag{60}$$

with probability 1.

The function $\gamma \mapsto H_\rho(\gamma)$ is minimised at the unique point $\gamma = \gamma(\rho) > 0$ given in the statement, and $F_{n,\rho}(\gamma)$ is convex by (1) of Lemma C.1 of [3]. Fix $\omega \in \Omega$ and $0 < \varepsilon < \gamma(\rho)$, and set

$$D_\varepsilon = \min_{u\in\{\pm 1\}}[H_\rho(\gamma(\rho)+\varepsilon u) - H_\rho(\gamma(\rho))].$$

Since $\gamma(\rho)$ is the unique minimizer of H_ρ, one has $D_\varepsilon > 0$. From (60) we deduce the existence of $n_0(\omega)$ large enough such that $n \geq n_0(\omega)$ implies

$$2\max_{v\in\{0,\pm 1\}}\left|\frac{1}{d_n}F_{n,\rho}(\gamma(\rho)+\varepsilon u) - H_\rho(\gamma(\rho)+\varepsilon u)\right| < D_\varepsilon,$$

implying in turn, by Lemma 6.3, that

$$|\gamma_{\rho,n} - \gamma(\rho)| \leq \varepsilon.$$

That is, with probability 1,

$$\gamma_{\rho,n} \longrightarrow \gamma(\rho) \quad \text{as } n \to \infty.$$

Relation (59) now follows from a standard application of Fatou's Lemma, and the proof of (57) is therefore achieved. □

4.3 The descent cone of the Schatten 1-norm at a low rank matrix

In this section we provide lower bounds on the conic variances of the descent cones of the Schatten 1-norm (see definition (24)) for a sequence of low rank matrices.

For every $k \in \mathbb{N}$, let (n, m, r) be a triple of nonnegative integers depending on k. We drop explicit dependence of n, m and r on k for notational ease, and continue to take $m \leq n$ without loss of generality. We assume that $n \to \infty, m/n \to \nu \in (0, 1]$ and $r/m \to \rho \in (0, 1)$ as $k \to \infty$, and that for every k the matrix $\mathbf{X}(k) \in \mathbb{R}^{m \times n}$ has rank r. Let

$$C_k = \mathcal{D}(\| \cdot \|_{S_1}, \mathbf{X}(k)), \quad \delta_k = \delta(C_k) \quad \text{and} \quad \tau_k^2 = \mathrm{Var}(V_{C_k})$$

denote the descent cone of the Schatten 1-norm of $\mathbf{X}(k)$, its statistical dimension, and the the variance of its conic intrinsic volume distribution, respectively. Proposition 4.7 of [3] provides that

$$\lim_{k \to \infty} \frac{\delta_k}{nm} = \psi(\rho, \nu), \tag{61}$$

where $\psi : [0, 1]^2 \to [0, 1]$ is given by

$$\psi(\rho, \nu) = \inf_{\gamma \geq 0} \eta(\gamma) \quad \text{with}$$

$$\eta(\gamma) = \left\{ \rho\nu + (1 - \rho\nu) \left[\rho(1 + \gamma^2) + (1 - \rho) \int_{a_-}^{a_+} (u - \gamma)_+^2 \phi_y(u) du \right] \right\}, \tag{62}$$

and $y = (\nu - \rho\nu)/(1 - \rho\nu)$, $a_{\pm} = 1 \pm \sqrt{y}$, and

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

The infimum of $\eta(\gamma)$ over $[0, \infty)$ is attained at the solution $\gamma(\nu, \rho)$ to

$$\int_{a_- \vee \gamma}^{a_+} \left(\frac{u}{\gamma} - 1 \right) \phi_y(u) du = \frac{\rho}{1 - \rho}.$$

It is not difficult to verify that $\gamma(\nu, \rho) > 0$ for all $\nu \in (0, 1], \rho \in (0, 1)$.

Proposition 4.2 *For the sequence of matrices $\mathbf{X}(k), k \in \mathbb{N}$,*

$$\liminf_{k \to \infty} \frac{\tau_k^2}{\delta_k} \geq \min \left(\frac{\sqrt{2}[\rho(1 - \nu\rho)\gamma(\nu, \rho)]^2}{\psi(\rho, \nu)^{3/2}}, \frac{2^{3/2}}{\sqrt{\psi(\rho, \nu)}} \right). \tag{63}$$

Proof. By (D.8) of [3], the subdifferential of the Schatten 1-norm at $\mathbf{X}(k)$ is given by

$$\partial \|\mathbf{X}(k)\|_{S_1} = \left\{ \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & \mathbf{W} \end{bmatrix} \in \mathbb{R}^{m \times n} : \sigma_1(W) \leq 1 \right\}, \tag{64}$$

and it generates the polar C^0 of the descent cone, see Corollary 23.7.1 of [36]. Closely following the proof of Proposition 4.7 of [3], and in particular the application of the Hoffman-Wielandt Theorem, see [28], Corollary 7.3.8] for the second equality below, taking $\mathbf{G}$ to be an $m \times n$ matrix with independent $\mathcal{N}(0,1)$ entries, we have

$$\begin{aligned} \operatorname{dist}(\mathbf{G}, \gamma \cdot \partial \|\mathbf{X}(k)\|_{S_1})^2 &= \left\| \begin{bmatrix} \mathbf{G}_{11} - \gamma I_r & \mathbf{G}_{12} \\ \mathbf{G}_{21} & 0 \end{bmatrix} \right\|_F^2 + \inf_{\sigma_1(\mathbf{W}) \le 1} \|\mathbf{G}_{22} - \gamma \mathbf{W}\|_F^2 \\ &= \left\| \begin{bmatrix} \mathbf{G}_{11} - \gamma I_r & \mathbf{G}_{12} \\ \mathbf{G}_{21} & 0 \end{bmatrix} \right\|_F^2 + \sum_{i=1}^{m-r} (\sigma_i(\mathbf{G}_{22}) - \gamma)_+^2 \,, \end{aligned} \tag{65}$$

with $\|\cdot\|_F$ denoting the Frobenius norm and where $\mathbf{G}$ is partitioned into the 2×2 block matrix $(\mathbf{G}_{ij})_{1 \le i,j \le 2}$ formed by grouping successive rows of sizes r and $m-r$, and successive columns of sizes r and $n-r$. Hence, we obtain

$$\Pi_{C_k^0}(\mathbf{G}) = \begin{bmatrix} \gamma_k I_r & 0 \\ 0 & \gamma_k W^* \end{bmatrix} \tag{66}$$

for some matrix W^* with largest singular value at most 1, and γ_k the minimizer of the map $\gamma \to \operatorname{dist}(\mathbf{G}, \gamma \cdot \partial \|\mathbf{X}(k)\|_{S_1})^2$ given by (65). As the subdifferential (64) is a nonempty, compact, convex subset of $\mathbb{R}^{m \times n}$ that does not contain the origin, Lemma C.1 of [3] guarantees that the map is convex.

By [4], Theorem 3.6,

$$\frac{1}{nm} \operatorname{dist}^2(\mathbf{G}, \gamma\sqrt{n-r} \cdot \partial \|\mathbf{X}\|_{S_1}) \to_{\text{a.s.}} \eta(\gamma),$$

where $\eta(\gamma)$ is given in (62). Reasoning as in Section 4.2 (that is, using Lemma 6.3 followed by Fatou's lemma), we obtain

$$\frac{\gamma_k}{\sqrt{n-r}} = \operatorname{argmin} \left(\operatorname{dist}^2(\mathbf{G}, \gamma\sqrt{n-r} \cdot \partial \|\mathbf{X}\|_{S_1}) \right) \to_{\text{a.s.}} \gamma(\nu, \rho)$$
$$\text{and} \quad \liminf_{k \to \infty} \frac{E[\gamma_k]}{\sqrt{n-r}} \ge \gamma(\nu, \rho). \tag{67}$$

We now invoke Theorem 4.1, and make use of b) of Remark 4.1, to compute a variance lower bound in terms of

$$v_k = \|E[\Pi_{C_k^0}(\mathbf{G})]\|_F^2.$$

The two terms in the minimum in (54) give rise to the corresponding terms in (63). By (66),

$$\|\Pi_{C^0}(\mathbf{G})\|_F \ge \sqrt{r}\gamma_k.$$

Squaring, taking expectation, and applying (67), we find

$$\liminf_{k \to \infty} \frac{v_k}{nm} \ge \liminf_{k \to \infty} \frac{r\gamma_k^2}{nm} = \rho(1 - \nu\rho)\gamma(\nu, \rho). \tag{68}$$

Letting $b_k = \sqrt{\delta_k nm/2}$, since (61) provides that $\delta_k \approx nm\psi(\rho, \nu)$, we obtain

$$\liminf_{k \to \infty} \frac{v_k^2}{\delta_k b_k} = \liminf_{k \to \infty} \frac{\sqrt{2} v_k^2}{(nm)^2 \psi(\rho, \nu)^{3/2}}.$$

Applying (68) now yields the first term in (63). Next, as

$$\liminf_{k \to \infty} \frac{4b_k}{\delta_k} = \liminf_{k \to \infty} 2^{3/2} \sqrt{\frac{nm}{\delta_k}},$$

applying (61) now yields the second term in (63), completing the proof. □

5 Bound to the normal for V_C

Fix a non-trivial convex cone $C \subset \mathbb{R}^d$, and denote by δ_C and τ_C, respectively, the mean and variance of its intrinsic conic distribution. The main result of the present section is Theorem 5.1, providing a bound on the L^∞ norm

$$\eta = \|F - \Phi\|_\infty = \sup_{u \in \mathbb{R}} |F(u) - \Phi(u)| \tag{69}$$

of the difference between the distribution function $F(u)$ of $(V_C - \delta_C)/\tau_C$ and $\Phi(u) = P[N \leq u]$, where $N \sim \mathcal{N}(0,1)$. In the following, we set $\log^+ x = \max(\log x, 0)$.

Lemma 5.1 *Let $\psi_F(t)$ and $\psi_G(t)$ denote the characteristic functions of a mean-zero distribution with variance 1 and the standard normal distribution $\mathcal{N}(0,1)$, respectively. If*

$$\sup_{|t| \leq L} |\psi_F(t) - \psi_G(t)| \leq B \tag{70}$$

for some positive real numbers L and B, then

$$\eta \leq B \log^+(L) + \frac{4}{L}. \tag{71}$$

Proof: The result holds trivially for $L < 1$, so assume $L \geq 1$. Let $h_L(x)$ be the 'smoothing' density function

$$h_L(x) = \frac{1 - \cos Lx}{\pi L x^2},$$

corresponding to the distribution function $H_L(x)$, let $\Delta(x) = F(x) - G(x)$, and let

$$\Delta_L = \Delta * H_L \quad \text{and} \quad \eta_L = \sup |\Delta_L(x)|.$$

By Lemma 3.4.10 and the proof of Lemma 3.4.11 of [24] we have

$$\eta \leq 2\eta_L + \frac{24}{\sqrt{2}\pi^{3/2} L} \quad \text{and} \quad \eta_L \leq \frac{1}{2\pi} \int_{|t| \leq L} |\psi_F(t) - \psi_G(t)| \frac{dt}{|t|}. \tag{72}$$

As $\psi_F(t)$ is a characteristic function of a mean-zero distribution with variance 1, it is straightforward to prove that

$$|\psi_F(t) - 1| \leq \frac{t^2}{2},$$

so

$$|\psi_F(t) - \psi_G(t)| = |(\psi_F(t) - 1) - (\psi_G(t) - 1)| \leq t^2.$$

Hence for all $\epsilon \in (0, L]$

$$\int_{|t| \leq \epsilon} |\psi_F(t) - \psi_G(t)| \frac{dt}{|t|} \leq \int_{|t| \leq \epsilon} |t| = \epsilon^2. \tag{73}$$

By (70),

$$\int_{\epsilon<|t|\le L} |\psi_F(t)-\psi_G(t)|\frac{dt}{|t|} \le 2B\log(L/\epsilon). \tag{74}$$

Hence, by (73), (74) and (72),

$$\eta \le \frac{1}{\pi}\left(\epsilon^2 + 2B\log(L/\epsilon) + \frac{24}{\sqrt{2\pi}L}\right).$$

As $L \ge 1$ we may choose $\epsilon = L^{-1/2}$. The conclusion now follows. □

Lemma 5.2 *Let $\tau \ge 0$ and $\delta > 0$ satisfy $\tau^2 \le 2\delta$. Then, the quantity*

$$L = \sqrt{\frac{\tau^2}{144\delta}\log^+\left(\frac{\tau^3}{\delta}\right)} \quad \textit{satisfies} \quad L \le \tau/8. \tag{75}$$

Proof: Consider the function on $[0,\infty)$ given by

$$f(x) = 2\sqrt{2}x - e^{\frac{9x^2}{4}}, \quad \text{with derivative} \quad f'(x) = 2\sqrt{2} - \frac{9x}{2}e^{\frac{9x^2}{4}}.$$

Clearly, $f'(x)$ is positive at zero and decreases strictly to $-\infty$ as $x \to \infty$. Hence $f(x)$ has a global maximum value on $[0,\infty)$ achieved at the unique solution x_0 to the equation

$$xe^{\frac{9x^2}{4}} = \frac{4\sqrt{2}}{9}.$$

Note that

$$f(x_0) = 2\sqrt{2}x_0 - e^{\frac{9x_0^2}{4}} = \frac{2\sqrt{2}}{9x_0}\left(9x_0^2 - \frac{9}{2\sqrt{2}}x_0 e^{\frac{9x_0^2}{4}}\right) = g(x_0) \quad \text{where} \quad g(x) = \frac{2\sqrt{2}}{9x}\left(9x^2-2\right)$$

and that

$$f'\left(\frac{\sqrt{2}}{3}\right) = \frac{\sqrt{2}}{2}\left(4-3\sqrt{e}\right) < 0.$$

Hence $x_0 \le \sqrt{2}/3$, and since $g(x)$ is increasing in $[0,\infty)$, we have $f(x_0) = g(x_0) \le g(\sqrt{2}/3) = 0$. As $f(x_0)$ is the global maximum of $f(x)$ on $[0,\infty)$ we conclude that

$$2\sqrt{2}x \le e^{\frac{9x^2}{4}}. \tag{76}$$

Using $\tau^2 \le 2\delta$ and (76) we obtain

$$\tau^3 \le 2\sqrt{2}\delta^{3/2} \le \delta e^{\frac{9\delta}{4}} \quad \text{implying} \quad \log\left(\frac{\tau^3}{\delta}\right) \le \frac{9\delta}{4}.$$

The final inequality holds with log replaced by $\log^+$ since the right hand side is always non negative. The inequality so obtained provides an upper bound on L in (75) that verifies the claim. □

In the following theorem, for notational simplicity we will write δ, τ and σ instead of δ_C, τ_C and σ_C respectively, and also set $a \vee b = \max\{a,b\}$.

Theorem 5.1 *The L^∞ norm η given in* (69) *satisfies*

$$\eta \le \frac{1}{108}\left(\frac{\tau}{\delta^3} \vee \frac{1}{\delta^{8/3}}\right)^{\frac{3}{16}} \left(\log^+\left(\frac{\tau^3}{\delta}\right)\right)^{\frac{3}{2}} \log^+\left(\frac{\tau^2}{144\delta}\log^+\left(\frac{\tau^3}{\delta}\right)\right) + 48\sqrt{\frac{\delta}{\tau^2 \log^+\left(\frac{\tau^3}{\delta}\right)}}. \quad (77)$$

Remark 5.1 The estimate (11) follows immediately from (77) and the following inequalities, valid for $\delta \ge 8$:

$$\begin{aligned}
\left(\frac{\tau}{\delta^3} \vee \frac{1}{\delta^{8/3}}\right)^{\frac{3}{16}} &\le \frac{\sqrt{2}}{\delta^{15/32}}, \\
\left(\log^+\left(\frac{\tau^3}{\delta}\right)\right)^{\frac{3}{2}} &\le (\log 2\sqrt{2\delta})^{3/2} \le (\log\delta)^{3/2}, \\
\log^+\left(\frac{\tau^2}{144\delta}\log^+\left(\frac{\tau^3}{\delta}\right)\right) &\le \log(\log\delta) \le \log\delta.
\end{aligned}$$

The above relations all follow from the bound $\tau \le \sqrt{2\delta}$ stated in (44).

Remark 5.2 When considering a sequence of cones such that $\liminf \tau^2/\delta > 0$, the right-hand side of the bound (77) behaves like $O\left(1/\sqrt{\log\delta}\right)$, thus yielding the Berry-Esseen estimate stated in Part 2 of Theorem 1.1. However, one should note that the bound (77) covers in principle a larger spectrum of asymptotic behaviors in the parameters τ^2 and δ: in particular, in order for the right-hand side of (77) to converge to zero, it is not necessary that the ratio τ^2/δ is bounded away from zero.

Proof: We show Lemma 5.1 may be applied with L as in (75) and

$$B = 32L^3 e^{\frac{9L^2\delta}{\tau^2}}\frac{\delta}{\tau^3}. \quad (78)$$

Let $t \in \mathbb{R}$ satisfy $|t| \le L$. As was done in [32] for the Laplace transform, the implication (40) of the Steiner formula (39) can be applied to show that the relationship

$$Ee^{itV} = Ee^{\xi_{it}G} \quad \text{with} \quad \xi_t = \frac{1}{2}\left(1 - e^{-2t}\right) \quad (79)$$

holds between the characteristic functions of $V = V_C$ and $G = \|\Pi_C(\mathbf{g})\|^2$. Replacing t by t/τ and multiplying by $e^{-it\delta/\tau}$ in (79) yields the following expression for the standardized characteristic function of V,

$$Ee^{it\left(\frac{V-\delta}{\tau}\right)} = Ee^{\xi_{it/\tau}G}e^{-\frac{it\delta}{\tau}}. \quad (80)$$

Comparing the characteristic function of the standardized V to that of the standard normal, identity (80) and the triangle inequality yield

$$\begin{aligned}
|Ee^{it\left(\frac{V-\delta}{\tau}\right)} - e^{-t^2/2}| &= |Ee^{\xi_{it/\tau}G}e^{-\frac{it\delta}{\tau}} - e^{-t^2/2}| \\
&\le |Ee^{\xi_{it/\tau}G}\left(e^{-\frac{it\delta}{\tau}} - e^{\left(\frac{t^2}{\tau^2} - \xi_{it/\tau}\right)\delta}\right)| + e^{\frac{t^2\delta}{\tau^2}}|Ee^{\xi_{it/\tau}(G-\delta)} - Ee^{it/\tau(G-\delta)}| \\
&\quad + e^{\frac{t^2\delta}{\tau^2}}|Ee^{it/\tau(G-\delta)} - e^{-\sigma^2t^2/2\tau^2}|. \quad (81)
\end{aligned}$$

For the final term we have used (42), which shows that $2\delta - \sigma^2 = -\tau^2$. For the first two terms we will make use of the inequality

$$|e^{(a+bi)g} - e^{cig}| \le (|b-c| + |a|)\, e^{|ga|}|g|, \tag{82}$$

valid for all $a, b, c, g \in \mathbb{R}$, which follows immediately by substitution from

$$\begin{aligned} |e^{a+bi} - e^{ci}| &= |e^{a+bi} - e^{a+ci} + e^{a+ci} - e^{ci}| \\ &\le e^a|b-c| + |e^a - 1| \\ &\le e^a|b-c| + e^{|a|} - 1 \\ &\le e^{|a|}\left(|b-c| + |a|\right). \end{aligned}$$

Now using (79), implying $|Ee^{\xi_{it/\tau}G}| = |Ee^{(it/\tau)V}| \le 1$, we bound the first term in (81) by

$$|Ee^{\xi_{it/\tau}G}|\, |e^{-\frac{it\delta}{\tau}} - e^{\left(\frac{t^2}{\tau^2} - \xi_{it/\tau}\right)\delta}| \le |e^{-\frac{it\delta}{\tau}} - e^{\left(\frac{t^2}{\tau^2} - \xi_{it/\tau}\right)\delta}| = |e^{ci} - e^{a+bi}|,$$

where we have set

$$a = \frac{t^2\delta}{\tau^2} - \frac{1}{2}\left(1 - \cos\left(2t/\tau\right)\right)\delta, \quad b = -\frac{1}{2}\sin(2t/\tau)\delta, \quad \text{and} \quad c = -\frac{t\delta}{\tau},$$

which satisfy

$$|a| \le \frac{2|t|^3\delta}{3\tau^3} \quad \text{and} \quad |b-c| \le \frac{2|t|^3\delta}{3\tau^3}.$$

By (44) of Corollary 3.1 we have $\tau^2 \le 2\delta$, and in particular we may apply Lemma 5.2 to yield $|t| \le L \le \tau/8$. Now (82) with $g = 1$ shows that the first term is bounded by

$$\frac{4|t|^3\delta}{3\tau^3} e^{\frac{2t^2\delta}{\tau^2}}. \tag{83}$$

Now we write the second term as

$$e^{\frac{t^2\delta}{\tau^2}} E|e^{\xi_{it/\tau}(G-\delta)} - e^{it/\tau(G-\delta)}| = e^{\frac{t^2\delta}{\tau^2}} E|e^{(a+bi)g} - e^{cig}|, \tag{84}$$

where

$$a = \frac{1}{2}\left(1 - \cos(2t/\tau)\right), \quad b = \frac{1}{2}\sin(2t/\tau), \quad c = t/\tau \quad \text{and} \quad g = G - \delta,$$

for which

$$|a| \le \min\left(\frac{|t|}{\tau}, \frac{t^2}{\tau^2}\right) \quad \text{and} \quad |b-c| \le \frac{t^2}{\tau^2}.$$

Applying (82) and the Cauchy Schwarz inequality we may bound (84) as

$$e^{\frac{t^2\delta}{\tau^2}} \frac{2t^2}{\tau^2} E\left(e^{\frac{|t|}{\tau}|G-\delta|}|G-\delta|\right) \le \frac{2\sigma t^2}{\tau^2} e^{\frac{t^2\delta}{\tau^2}} \sqrt{Ee^{\frac{2|t|}{\tau}|G-\delta|}}. \tag{85}$$

Recalling that $\|\Pi_C(\mathbf{x})\|^2 = d^2(\mathbf{x}, C^0)$, invoking Theorem 2.2 for the polar cone C^0 and $\boldsymbol{\mu} = \mathbf{0}$, for $0 \le \xi < 1/2$ inequality (33) yields

$$\begin{aligned} Ee^{\xi|G-\delta|} &= Ee^{\xi(G-\delta)}\mathbf{1}(G-\delta \ge 0) + Ee^{-\xi(G-\delta)}\mathbf{1}(G-\delta < 0) \\ &\le Ee^{\xi(G-\delta)} + Ee^{-\xi(G-\delta)} \\ &\le \exp\left(\frac{2\xi^2\delta}{1-2\xi}\right) + \exp\left(\frac{2\xi^2\delta}{1+2\xi}\right) \le 2\exp\left(\frac{2\xi^2\delta}{1-2\xi}\right). \end{aligned}$$

Thus, applying this bound with $\xi = 2|t|/\tau$, where $\xi < 1/2$ by virtue of $|t| \le \tau/8$ we obtain a bound on (85), and hence on the second term of (81), of the form

$$\frac{2\sigma t^2}{\tau^2} e^{\frac{t^2\delta}{\tau^2}} \sqrt{2\exp\left(\frac{8t^2\delta}{\tau^2(1-4|t|/\tau)}\right)} \le 2\sqrt{2}\frac{\sigma t^2}{\tau^2} e^{\frac{9t^2\delta}{\tau^2}}. \tag{86}$$

For the final term, as the function $e^{it/\tau}$ has modulus 1, Theorem 2.1 yields

$$e^{\frac{t^2\delta}{\tau^2}} |Ee^{it/\tau(G-\delta)} - e^{-\sigma^2 t^2/2\tau^2}| \le 16\frac{\sqrt{\delta_C}}{\sigma^2} e^{\frac{t^2\delta}{\tau^2}}. \tag{87}$$

Combining the three terms (83), (86) and (87), for $|t| \le L$ we obtain

$$\frac{4|t|^3\delta}{3\tau^3} e^{\frac{2t^2\delta}{\tau^2}} + 2\sqrt{2}\frac{\sigma t^2}{\tau^2} e^{\frac{9t^2\delta}{\tau^2}} + 16\frac{\sqrt{\delta_C}}{\sigma^2} e^{\frac{t^2\delta}{\tau^2}} \le \left(\frac{4L^3\delta}{3\tau^3} + 2\sqrt{2}\frac{\sigma L^2}{\tau^2} + 16\frac{\sqrt{\delta_C}}{\sigma^2}\right) e^{\frac{9L^2\delta}{\tau^2}}.$$

From the bounds (44) in Corollary 3.1, we have

$$\frac{4L^3\delta}{3\tau^3} + \frac{2\sqrt{2}L^2\sigma}{\tau^2} + \frac{16\sqrt{\delta_C}}{\sigma^2} \le \left(\frac{4L^3}{3} + 8L^2 + 16\sqrt{2}\right)\frac{\delta}{\tau^3}.$$

As the bound (71) holds for $L < 1$, we may assume $L \ge 1$, in which case B as in (78) satisfies (70) when ψ_F and ψ_G are the characteristic functions of $(V-\delta)/\tau$ and the standard normal, respectively. Invoking Lemma 5.1, the proof is completed by specializing (71) to yield (77) for the given values of L and B. □

6 Appendix

6.1 A total variation bound

Here, we prove the total variation bound (29) used in the proof of Theorem 2.1. We begin with a standard lemma based on Stein's method (see [34]), involving the solution ϕ_h to the Stein equation

$$\phi_h'(x) - x\phi_h(x) = h(x) - E[h(N)] \tag{88}$$

for $N \sim \mathcal{N}(0,1)$ and a given test functions h.

Lemma 6.1 *If $E[F] = 0$ and $E[F^2] = 1$, then*

$$d_{TV}(F, N) \le \sup_{\phi} |E[\phi'(F)] - E[F\phi(F)]|, \tag{89}$$

where $N \sim \mathcal{N}(0,1)$ and the supremum runs over all C^1 functions $\phi : \mathbb{R} \to \mathbb{R}$ with $\|\phi'\|_\infty \le 2$.

Proof: For a given $h \in C^0$ taking values in $[0,1]$, by e.g. (2.5) of [17], the unique bounded solution $\phi_h(x)$ to the Stein equation (88) is given by

$$\phi_h(x) = e^{x^2/2} \int_{-\infty}^{x} e^{-u^2/2}(h(u) - E[h(N)])du = -e^{x^2/2} \int_{x}^{\infty} e^{-u^2/2}(h(u) - E[h(N)])du, \quad (90)$$

where the second equality holds since

$$\int_{\mathbb{R}} e^{-u^2/2}(h(u) - E[h(N)])du = \sqrt{2\pi} E[h(N) - E[h(N)]] = 0.$$

One can easily check that ϕ_h is C^1. Using the first equality in (90) for $x < 0$, and the second one for $x > 0$ one obtains that $|x\phi_h(x)| \le e^{x^2/2} \int_{|x|}^{\infty} ue^{-u^2/2} = 1$. We deduce that $|\phi_h'|_\infty \le 2$. Recall that the total variation distance $d_{TV}(F,G)$ (as defined in (28)) may also be represented as the supremum over all measurable functions h taking values in $[0,1]$. Using this fact, together with Lusin's theorem, relation (88) and the properties of the solution ϕ_h, we infer that

$$\begin{aligned} d_{TV}(F,N) &= \sup_{h:\mathbb{R}\to[0,1]} |E[h(F)] - E[h(N)]| \\ &= \sup_{h:\mathbb{R}\to[0,1],\, h\in C^0} |E[h(F)] - E[h(N)]| \le \sup_{\phi} |E[\phi'(F)] - E[F\phi(F)]|, \end{aligned}$$

as claimed. □

To make the paper as self-contained as possible, we will also prove the total variation bound (29) that was applied in the proof of Theorem 2.1; this result is given, at a slightly lesser level of generality, as Lemma 5.3 in [14].

Given $d \ge 1$, we use the symbol $\mathbb{D}^{1,2}$ to denote the Sobolev class of all mappings $f : \mathbb{R}^d \to \mathbb{R}$ that are in the closure of the set of polynomials $p : \mathbb{R}^d \to \mathbb{R}$ with respect to the norm

$$\|p\|_{1,2} = \left(\int_{\mathbb{R}^d} p(x)^2 d\gamma(x) \right)^{1/2} + \left(\int_{\mathbb{R}^d} \|\nabla p(x)\|^2 d\gamma(x) \right)^{1/2},$$

where γ stands for the standard Gaussian measure on $\mathbb{R}^d$. It is not difficult to show that a sufficient condition in order for f to be a member of $\mathbb{D}^{1,2}$ is that f is of class C^1, with f and its derivatives having subexponential growth at infinity. We stress that, in general, when f is in $\mathbb{D}^{1,2}$ the symbol ∇f has to be interpreted in a weak sense. See e.g. [34, Chapters 1 and 2] for details on these concepts.

Theorem 6.1 *Let $H : \mathbb{R}^d \to \mathbb{R}$ be an element of $\mathbb{D}^{1,2}$. Let $\mathbf{g} \sim \mathcal{N}(0, I_d)$ be a standard Gaussian random vector in $\mathbb{R}^d$. Let $F = H(\mathbf{g})$ and set $m = E[F]$ and $\sigma^2 = \mathrm{Var}(F)$. Further, for $t \ge 0$, set $\widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1-e^{-2t}}\widehat{\mathbf{g}}$, where $\widehat{\mathbf{g}}$ is an independent copy of $\mathbf{g}$. Write $\widehat{E}$ to indicate expectation with respect to $\widehat{\mathbf{g}}$. Then, with $N \sim N(m, \sigma^2)$,*

$$d_{TV}(F,N) \le \frac{2}{\sigma^2} \sqrt{\mathrm{Var}\left(\int_0^\infty e^{-t} \langle \nabla H(\mathbf{g}), \widehat{E}(\nabla H(\widehat{\mathbf{g}}_t)) \rangle dt \right)}. \quad (91)$$

Proof: Without loss of generality, assume that $m = 0$ and $\sigma^2 = 1$. The random vector

$$\mathbf{g}_t = \sqrt{1-e^{-2t}}\mathbf{g} - e^{-t}\widehat{\mathbf{g}} \quad \text{is an independent copy of } \widehat{\mathbf{g}}_t, \text{ and } \quad \mathbf{g} = e^{-t}\widehat{\mathbf{g}}_t + \sqrt{1-e^{-2t}}\mathbf{g}_t. \tag{92}$$

By a standard approximation argument, it is sufficient to show the result for $H \in C^1$, with H and its derivatives having subexponential growth at infinity. Let $\mathbf{E} = E \otimes \widehat{E}$. If $\varphi : \mathbb{R} \to \mathbb{R}$ is C^1, then using the growth conditions imposed on H to carry out the interchange of expectation and integration and the integration-by-parts, one has

$$\begin{aligned}
E[F\varphi(F)] &= E[(H(\mathbf{g}) - H(\widehat{\mathbf{g}}))\,\varphi(H(\mathbf{g}))] = -\int_0^\infty \frac{d}{dt}\mathbf{E}[H(\widehat{\mathbf{g}}_t)\varphi(H(\mathbf{g}))]dt \\
&= \int_0^\infty e^{-t}\mathbf{E}\langle \nabla H(\widehat{\mathbf{g}}_t), \mathbf{g}\rangle \varphi(H(\mathbf{g}))dt - \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\mathbf{E}\langle \nabla H(\widehat{\mathbf{g}}_t), \widehat{\mathbf{g}}\rangle \varphi(H(\mathbf{g}))dt \\
&= \int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-2t}}}\mathbf{E}\langle \nabla H(\widehat{\mathbf{g}}_t), \mathbf{g}_t\rangle \varphi(H(e^{-t}\widehat{\mathbf{g}}_t + \sqrt{1-e^{-2t}}\mathbf{g}_t))dt \\
&= \int_0^\infty e^{-t}\mathbf{E}\langle \nabla H(\widehat{\mathbf{g}}_t), \nabla H(e^{-t}\widehat{\mathbf{g}}_t + \sqrt{1-e^{-2t}}\mathbf{g}_t)\rangle \varphi'(H(e^{-t}\widehat{\mathbf{g}}_t + \sqrt{1-e^{-2t}}\mathbf{g}_t))dt \\
&= E\int_0^\infty e^{-t}\langle \nabla H(\mathbf{g}), \widehat{E}(\nabla H(\widehat{\mathbf{g}}_t))\rangle \varphi'(H(\mathbf{g}))dt. \qquad (93)
\end{aligned}$$

Applying identity (93) to (89) yields

$$d_{TV}(F, N) \le 2E\left|1 - \int_0^\infty e^{-t}\langle \nabla H(\mathbf{g}), \widehat{E}(\nabla H(\widehat{\mathbf{g}}_t))\rangle dt\right|, \tag{94}$$

and for $\varphi(x) = x$ yields

$$\mathrm{Var}(F) = E\int_0^\infty e^{-t}\langle \nabla H(\mathbf{g}), \widehat{E}(\nabla H(\widehat{\mathbf{g}}_t))\rangle dt. \tag{95}$$

As $\mathrm{Var}(F) = 1$ the conclusion (91), with $\sigma^2 = 1$, now follows by applying the Cauchy Schwarz inequality in (94). □

We now prove the following useful fact that was applied in the proofs of Theorem 2.1 and Lemma 4.1.

Lemma 6.2 *Let C be a closed convex subset of $\mathbb{R}^d$. Then, the mapping*

$$\mathbf{x} \mapsto d^2(\mathbf{x}, C)$$

is an element of $\mathbb{D}^{1,2}$.

Proof: It is sufficient to show that $d^2(\cdot, C)$ and its derivative have sub-exponential growth at infinity. To prove this, observe that Lemma 2.1 together with the triangle inequality imply that $d(\cdot, C)$ is 1-Lipschitz, so that $d^2(\mathbf{x}, C) \le 2d^2(\mathbf{0}, C) + 2\|\mathbf{x}\|^2$. To conclude, use (27) in order to deduce that

$$\|\nabla d^2(\mathbf{x}, C)\| = 2d(\mathbf{x}, C) \le 2d(\mathbf{0}, C) + 2\|\mathbf{x}\|.$$

□

A variation of the arguments leading to the proof of (93) (whose details are left to the reader) yield also the following useful result.

Proposition 6.1 *Let $F, G \in \mathbb{D}^{1,2}$, and let the notation adopted in the statement and proof of Theorem 6.1 prevail. Then,*

$$\mathrm{Cov}[F(\mathbf{g})G(\mathbf{g})] = \mathbf{E}\int_0^\infty e^{-t}\langle \nabla F(\mathbf{g}), \nabla G(\widehat{\mathbf{g}}_t)\rangle dt. \tag{96}$$

6.2 An improved Poincaré inequality

The next result refines the classical Poincaré inequality, and plays a pivotal role in Theorems 2.1 and 4.1.

Theorem 6.2 (Improved Poincaré inequality) *Fix $d \geq 1$, let $F \in \mathbb{D}^{1,2}$, and $\mathbf{g} = (g_1, ..., g_d) \sim \mathcal{N}(0, I_d)$. Then,*

$$\mathrm{Var}(F(\mathbf{g})) \leq \frac{1}{2}E[\|\nabla F(\mathbf{g})\|^2] + \frac{1}{2}\|E[\nabla F(\mathbf{g})]\|^2 \leq E[\|\nabla F(\mathbf{g})\|^2].$$

Proof: The quickest way to show the estimate $\mathrm{Var}(F(\mathbf{g})) \leq \frac{1}{2}E[\|\nabla F(\mathbf{g})\|^2] + \frac{1}{2}\|E[\nabla F(\mathbf{g})]\|^2$ is to adopt a spectral approach. To accomplish this task, we shall use some basic results of Gaussian analysis, whose proofs can be found e.g. in [34, Chapter 2]. Recall that, for $k = 0, 1, 2, ...$, the k^{th} *Wiener chaos* associated with $\mathbf{g}$, written C_k, is the subspace spanned by all random variables of the form $\prod_{i=1}^m H_{k_i}(g_{j_i})$, where $\{H_k : k = 0, 1, ...\}$ denotes the collection of Hermite polynomials on the real line, $k_1 + \cdots + k_m = k$, and the indices $k_1, \cdots, k_m$ are pairwise distinct. It is easily checked that Wiener chaoses of different orders are orthogonal in $L^2(\Omega)$, and also that every square-integrable random variable of the type $F(\mathbf{g})$ can be decomposed as an infinite sum of the type $F(\mathbf{g}) = \sum_{k=0}^\infty F_k(\mathbf{g})$, where the series converges in $L^2(\Omega)$ and where, for every k, $F_k(\mathbf{g})$ denotes the projection of $F(\mathbf{g})$ on C_k (in particular, $F_0(\mathbf{g}) = E[F(\mathbf{g})]$). This decomposition yields in particular that

$$\mathrm{Var}(F(\mathbf{g})) = \sum_{k=1}^\infty E[F_k^2(\mathbf{g})].$$

The key point is now that, if $F \in \mathbb{D}^{1,2}$, then one has the additional relations

$$E[\|\nabla F(\mathbf{g})\|^2] = \sum_{k=1}^\infty kE[F_k^2(\mathbf{g})]$$

(see e.g. [34, Exercice 2.7.9]) and

$$E[F_1^2(\mathbf{g})] = \|E[\nabla F(\mathbf{g})]\|^2,$$

the last identity being justified as follows: if F is a smooth mapping, then the projection of $F(\mathbf{g})$ on C_1 is given by

$$F_1(\mathbf{g}) = \sum_{i=1}^d E[F(\mathbf{g})g_i]g_i = \sum_{i=1}^d E\left[\frac{\partial F}{\partial x_i}(\mathbf{g})\right] g_i,$$

and the result for a general $F \in \mathbb{D}^{1,2}$ is deduced by an approximation argument. The previous relations imply therefore that

$$\mathrm{Var}(F(\mathbf{g})) = \sum_{k=1}^{\infty} E[F_k^2(\mathbf{g})] \leq E[F_1^2(\mathbf{g})] + \sum_{k=2}^{\infty} \frac{k}{2} E[F_k^2(\mathbf{g})] = \frac{1}{2}\|E[\nabla F(\mathbf{g})]\|^2 + \frac{1}{2}E[\|\nabla F(\mathbf{g})\|^2].$$

The proof is concluded by observing that, in view of Jensen inequality, $\|E[\nabla F(\mathbf{g})]\|^2 \leq E[\|\nabla F(\mathbf{g})\|^2]$. □

6.3 A bound on the distance to the minimizer of a convex function.

Following an idea introduced by Hjort and Pollard [27], one has the following lemma, providing a bound on the distance to the minimizer of a convex function in terms of another, not necessarily convex, function.

Lemma 6.3 *Suppose $f : [0,\infty) \to \mathbb{R}$ is a convex function, and let $g : [0,\infty) \to \mathbb{R}$ be any function. If x_0 is a minimizer of f, $y_0 \in (0,\infty)$ and $\varepsilon \in (0, y_0)$, then*

$$2 \max_{v \in \{0,\pm 1\}} |g(y_0 + \varepsilon v) - f(y_0 + \varepsilon v)| < \min_{u \in \{\pm 1\}} [g(y_0 + \varepsilon u) - g(y_0)] \tag{97}$$

implies $|x_0 - y_0| \leq \varepsilon$.

Proof. Suppose $a := |x_0 - y_0| > \varepsilon > 0$. Set $u = a^{-1}(x_0 - y_0)$. Then $u \in \{\pm 1\}$, $x_0 = y_0 + au$ and the convexity of f implies

$$(1 - \varepsilon/a) f(y_0) + (\varepsilon/a) f(x_0) \geq f(y_0 + \varepsilon u).$$

Hence

$$\begin{aligned} & \frac{\varepsilon}{a}(f(x_0) - f(y_0)) \geq f(y_0 + \varepsilon u) - f(y_0) \\ = \;& g(y_0 + \varepsilon u) - g(y_0) + [f(y_0 + \varepsilon u) - g(y_0 + \varepsilon u)] + [g(y_0) - f(y_0)] \\ \geq \;& \min_{u \in \{\pm 1\}} [g(y_0 + \varepsilon u) - g(y_0)] - 2 \max_{v \in \{0,\pm 1\}} |g(y_0 + \varepsilon v) - f(y_0 + \varepsilon v)|. \end{aligned}$$

If (97) is satisfied, then $\frac{\varepsilon}{a}(f(x_0) - f(y_0)) > 0$. But this contradicts that x_0 is a minimizer of f. Hence, $|x_0 - y_0| > \varepsilon$ is impossible. □

Acknowledgments: The authors would like to thank John Pike for his assistance with the final two rows of Table 1, and the associated computations.

References

[1] Adler, R. and Taylor, J. (2007). *Random Fields and Geometry*. Springer.

[2] Allendoerfer, C. (1948). Steiner's formulae on a general S^{n+1}. *Bull. Amer. Math. Soc.* **54**, 128-135.

[3] Amelunxen, D., Lotz, M., McCoy, M. B., and Tropp, J. A. (2014). Living on the edge: A geometric theory of phase transitions in convex optimization. *Inform. Inference* **3**, no. 3, pp. 224-294.

[4] Bai, J., and Silverstein, J. W. (2006). *Spectral analysis of large dimensional random matrices*. Springer.

[5] Birgé, L. and Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97**, 113-150.

[6] Blass, A. and Sagan, B. (1998). Characteristic and Ehrhart polynomials. *J. Algebraic Combin.* **7**, 115-126.

[7] Bogachev, V. (1998). Gaussian measures. Volume 62, Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI.

[8] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities: A non asymptotic theory of independence*. Oxford.

[9] Bühlmann, P., and van de Geer, S. (2011). *Statistics for high-dimensional data*. Springer.

[10] Candès, E. J. (2014). Mathematics of sparsity (and few other things). *ICM 2014 Proceedings*, to appear.

[11] Candès, E.J., Romberg, J. and Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, **52**(2), 489-509.

[12] Candès, E.J., Romberg, J. and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, **59**(8), 1207-1223.

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

[14] Chatterjee, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. Probability Theory and Related Fields, 143(1-2), 1-40.

[15] Chatterjee, S. (2014). A new perspective on least squares under convex constraints. *Ann. Stat.*, to appear.

[16] Chatterjee, S., Guntuboyina, A. and Sen, B. (2013). Improved Risk Bounds in Isotonic Regression. arXiv preprint.

[17] Chen, L. H. Y., Goldstein, L., & Shao, Q. M. (2010). *Normal approximation by Stein's method*. Springer.

[18] Coxeter, H.M. and Moser W.J. (1972). *Generators and relations for discrete groups*. Springer.

[19] Davis, K.A. (2012). Constrained Statistical Inference: A Hybrid of Statistical Theory, Projective Geometry and Applied Optimization Techniques. *Progress in Applied Mathematics* **4**(2), 167-181.

[20] Donoho, D.L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, **52**(4), 1289-1306, 2006

[21] Donoho, D. L. (2006). High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete & Computational Geometry*, **35**(4), 617-652.

[22] Donoho, D.L. and Tanner, J. (2009). Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, **22**(1), 1-53.

[23] Drton, M. and C. Klivans, C. (2010). A geometric interpretation of the characteristic polynomial of reflection arrangements. *Proceedings of the American Mathematical Society* **138**(8),2873–2887.

[24] Durrett, R. (2010). *Probability: Theory and Examples (Fourth edition)*. Cambridge University Press.

[25] Dykstra, R. (1991). Asymptotic normality for chi-bar-square distributions. *The Canadian Journal of Statistics* **19**(3), 297-306.

[26] Herglotz, G. (1943). Über die Steinersche Formel für Parallelflächen. *Abh. Math. Sem. Hansischen Univ.*, **15**, 165-177.

[27] Hjort, N.L. and Pollard, D. (1993). Asymptotic for minimisers of convex processes. Preprint, Dept. of Statistics, Yale Univ.

[28] Horn, R and Johnson, C. (1990). *Matrix analysis*. Cambridge University Press, Cambridge, 1990.

[29] Klain, D.A., and Rota, G.-C. (1997). *Introduction to geometric probability*. Cambridge.

[30] Ledoux, M. (2005). The concentration of measure phenomenon. Vol. 89. Amer. Math. Soc.

[31] Luk, M. (1994). *Stein's method for the Gamma distribution and related statistical applications*. Ph.D. dissertation, University of Southern California, Los Angeles, USA.

[32] McCoy, M. B., and Tropp, J. A. (2014). From Steiner Formulas for Cones to Concentration of Intrinsic Volumes. *Discrete Comput. Geom.* **51**, no. 4, pp. 926-963.

[33] Nourdin, I. and Peccati, G. (2010). Stein's method and exact Berry-Esséen asymptotics for functionals of Gaussian fields. *Ann. Probab.* **37**(6), 2231-2261.

[34] Nourdin, I., and Peccati, G. (2012). Normal Approximations with Malliavin Calculus: From Stein's Method to Universality (Vol. 192). Cambridge University Press.

[35] Nourdin, I., Peccati, G., and Reinert, G. (2009). Second order Poincaré inequalities and CLTs on Wiener space. *Journal of Functional Analysis*, **257**(2), 593-609.

[36] Rockafellar, R. T. (1970). *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[37] Rockafellar, R.T. and Wets, R. J.-B. (1998). *Variational analysis*. Springer.

[38] Rudelson, M. and Vershynin, R. (2008). On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.* **61**(8), 1025-1045.

[39] Santaló, L. (1950). On parallel hypersurfaces in the elliptic and hyperbolic n-dimensional space. *Proc. Amer. Math. Soc.* **1**, 325-330.

[40] Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **70**, 597-586.

[41] Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review* **56**(1), 49-62.

[42] Silvapulle, M.J. and Sen, P.K. (2005). *Constrained statistical inference*. Wiley.

[43] Schneider, R. and Weil, W. (2008). *Stochastic and integral geometry*. Springer

[44] Taylor, J. (2013). The geometry of least squares in the 21st century. *Bernoulli* **19**(4), 1449-1464.

[45] van de Geer, S. (1990). Estimating a regression function. *Ann. Statist.* **18**(2), 907-924.

[46] Wang, Y. (1996). The l_2 risk of an isotonic estimate. *Comm. Statist. Theory Methods* **25**, 281-294.

[47] Zhang, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30**(2), 528-555.