Effective Condition Number Bounds for Convex Regularization

Amexlunxen, Dennis and Lotz, Martin and Walvin, Jake

2017

MIMS EPrint: **2017.31**

ISSN 1749-9097

EFFECTIVE CONDITION NUMBER BOUNDS FOR CONVEX REGULARIZATION

DENNIS AMELUNXEN, MARTIN LOTZ, AND JAKE WALVIN

ABSTRACT. We derive bounds relating the statistical dimension of linear images of convex cones to Renegar's condition number. Using results from conic integral geometry, we show that the bounds can be made to depend only on a random projection to a lower dimensional space, and can still be effective if the linear maps are ill-conditioned. As an application, we get new bounds for the undersampling phase transition of composite convex regularizers. Key tools in the analysis are Slepian's inequality, interpreted as monotonicity property of moment functionals, and the kinematic formula from integral geometry. The main results are derived in the generalized setting of the biconic homogeneous feasibility problem.

1. INTRODUCTION

A well-established approach to solving linear inverse problems with missing information is by means of convex regularization. In one of its manifestations, this approach amounts to solving the minimization problem

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2 \le \varepsilon, \tag{1.1}$$

where $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ is an underdetermined linear operator and $f(\boldsymbol{x})$ is a suitable proper convex function, informed by the application at hand.

While there are countless algorithms and heuristics to compute or approximate solutions of (1.1) and related problems, the more fundamental question is: when does a solution of (1.1) actually "make sense"? The latter is important because one is usually not interested in a solution of (1.1) per se, but often uses this and related formulations as a proxy for a different, much more intractable problem. The best-known example is the use of the 1-norm to obtain a sparse solution [FR13], but other popular settings are the total variation norm and its variants for signals with sparse gradient, or the nuclear norm of a matrix when aiming at a low-rank solution.

Regularizers often take the form $f(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$ for a linear map $\boldsymbol{D}$, as in the cosparse recovery setting [NDEG13]. In this article we present general bounds relating the performance of (1.1) to properties of g and the conditioning of $\boldsymbol{D}$. Moreover, we show that for the analysis we can replace $\boldsymbol{D}$ with a *random projection* applied to $\boldsymbol{D}$, where the target dimension of this projection is independent of the ambient dimension n and only depends on intrinsic properties of the regularizer g.

Various parameters have emerged in the study of the performance problems such as (1.1). Two of the most fundamental ones depend on the *descent cone* $\mathscr{D}(f,\boldsymbol{x}_0)$ of the function f at $\boldsymbol{x}_0$, defined as the cone spanned by all directions in which f decreases. These parameters are

- the statistical dimension $\delta(f,\boldsymbol{x}_0) := \delta(\mathscr{D}(f,\boldsymbol{x}_0))$, or equivalently the squared Gaussian width, of the descent cone $\mathscr{D}(f,\boldsymbol{x}_0)$ of f at $\boldsymbol{x}_0$, which determines the admissible amount of undersampling m in (1.1) in the noiseless case ($\varepsilon = 0$), in order to uniquely recover a solution $\boldsymbol{x}_0$[1];
- Renegar's condition number $\mathscr{R}_C(\boldsymbol{A})$ of $\boldsymbol{A}$ with respect to the descent cone $C = \mathscr{D}(f,\boldsymbol{x}_0)$ of f at a point $\boldsymbol{x}_0$, which bounds the recovery error $\|\boldsymbol{x}-\boldsymbol{x}_0\|_2$ of a solution $\boldsymbol{x}$ of (1.1).

Date: July 5, 2017.

[1]Strictly speaking, this is a result for *random* measurement matrices and holds with high probability.

Precise definitions of these quantities are given in subsequent sections. Renegar's condition number has originally been introduced to study the complexity of interior point methods [Ren95], and has recently been linked directly to the performance of algorithms for compressed sensing [RBd15]. The statistical dimension, on the other hand, has featured as a proxy to the squared Gaussian width in [Sto09, CRPW12] and as the main parameter determining phase transitions in convex optimization [ALMT14]. It also features in the error analysis of the generalized LASSO problem [OTH13] and as the minimax mean squared error (MSE) of proximal denoising [DJM13, OH16].

Unfortunately, computing or even estimating these parameters is a notoriously difficult problem for all but a few examples. For the popular case $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$, an effective method of computing $\delta(f, \boldsymbol{x}_0)$ was developed by Stojnic [Sto09], and subsequently generalized in [CRPW12], see also [ALMT14, Recipe 4.1]. In many practical settings the regularizer f has the form $f(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$ for a matrix $\boldsymbol{D}$, such as in the cosparse or analysis ℓ_1 setting where $f(\boldsymbol{x}) = \|\boldsymbol{D}\boldsymbol{x}\|_1$. Even when it is possible to accurately estimate the statistical dimension (and thus, the permissible undersampling) for a function g, the method may fail for a composite function $g(\boldsymbol{D}\boldsymbol{x})$, due to a lack of certain separability properties [ZXCL16]. If $f(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$ with invertible $\boldsymbol{D}$, then it is known (see Section 5) that the descent cone of f at $\boldsymbol{x}_0$ is given by $\mathscr{D}(f, \boldsymbol{x}_0) = \boldsymbol{D}^{-1}\mathscr{D}(g, \boldsymbol{D}\boldsymbol{x}_0)$. The statistical dimension and Renegar condition number associated to these regularizer can therefore be characterized as that of a linear image of the descent cone of g, and the problem becomes one of analyzing the behavior of the statistical dimension of a cone under linear transformations.

1.1. **Main results.** In this article we derive a characterization of Renegar's condition number associated to a cone as a measure of how much the statistical dimension can change under linear images of a cone. The first result in this direction is Theorem A. The upper bound in Equation (1.3) features implicitly in [KRZ15], explicitly in [DH17], and appears to be folklore.

Theorem A. *Let $C \subseteq \mathbb{R}^n$ be a closed convex cone, and $\delta(C)$ the statistical dimension of C. Then for $\boldsymbol{A} \in \mathbb{R}^{\ell \times n}$,*

$$\delta(\boldsymbol{A}C) \le \mathscr{R}_C(\boldsymbol{A})^2 \cdot \delta(C), \tag{1.2}$$

where $\mathscr{R}_C(\boldsymbol{A})$ is Renegar's condition number associated to the matrix $\boldsymbol{A}$ and the cone C. If $\ell = n$ and $\kappa(\boldsymbol{A})$ denotes the matrix condition number of $\boldsymbol{A}$, then

$$\frac{\delta(C)}{\kappa(\boldsymbol{A})^2} \le \delta(\boldsymbol{A}C) \le \kappa(\boldsymbol{A})^2 \cdot \delta(C). \tag{1.3}$$

Theorem A translates into a bound for the statistical dimension of convex regularizers.

Corollary 1.1. *Let $f(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$, where g is a proper convex function and let $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ be non-singular. Then*

$$\delta(f, \boldsymbol{x}_0) \le \mathscr{R}_{\mathscr{D}(g, \boldsymbol{D}\boldsymbol{x}_0)}\left(\boldsymbol{D}^{-1}\right) \cdot \delta(g, \boldsymbol{D}\boldsymbol{x}_0).$$

In particular,

$$\frac{\delta(g, \boldsymbol{D}\boldsymbol{x}_0)}{\kappa(\boldsymbol{D})^2} \le \delta(f, \boldsymbol{x}_0) \le \kappa(\boldsymbol{D})^2 \cdot \delta(g, \boldsymbol{D}\boldsymbol{x}_0).$$

It is not uncommon for the condition number to be large. For example, in the case of the finite difference matrix (see Example 1.4) it is of order $\Omega(n)$, making the condition bounds trivial. While Renegar's condition number, defined by restricting the smallest singular value to a cone, can improve the bound, computing this condition number is not always practical. Using polarity (4.7), we get the following version of the bound that ensures that the right-hand side is always bounded by n.

Proposition 1.2. *Let $C \subseteq \mathbb{R}^n$ be a closed convex cone, and $\delta(C)$ the statistical dimension of C. Then for non-singular $\boldsymbol{A} \in \mathbb{R}^{n\times n}$,*

$$\delta(\boldsymbol{A}C) \le \kappa(\boldsymbol{A})^{-2} \cdot \delta(C) + \left(1 - \kappa(\boldsymbol{A})^{-2}\right) \cdot n.$$

If $f(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$, where g is a proper convex function and $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ is non-singular, then

$$\delta(f,\boldsymbol{x}_0) \le \kappa(\boldsymbol{D})^{-2}\cdot\delta(g,\boldsymbol{D}\boldsymbol{x}_0) + \left(1-\kappa(\boldsymbol{D})^{-2}\right)\cdot n. \tag{1.4}$$

One can interpret the upper bounds in Proposition 1.2 as interpolating between the statistical dimension of C and the ambient dimension n. While Proposition 1.2 ensures that the upper bound does not become completely trivial, when $\boldsymbol{D}$ is ill-conditioned it still does not give satisfactory results. Using methods from conic integral geometry, we derive a "preconditioned" version of Theorem A. The idea is based on the philosophy that a randomly oriented convex cone C ought to behave roughly like a linear subspace of dimension $\delta(C)$, see Section 1.5. In that sense, the statistical dimension of a cone C should be approximately invariant under projecting C to a subspace of dimension close to $\delta(C)$. In fact, in Section 4.6 we will see that for $n \ge m \gtrapprox \delta(C)$, we have

$$\mathbb{E}_{\boldsymbol{Q}}\left[\delta(\boldsymbol{P}\boldsymbol{Q}C)\right] \approx \delta(C),$$

where $\boldsymbol{P}$ is the projection on the the first m coordinates and where the expectation is with respect to a random orthogonal matrix $\boldsymbol{Q}$, distributed according to the normalized Haar measure on the orthogonal group. From this it follows that the condition bounds should ideally depend not on the conditioning of $\boldsymbol{D}$ itself, but on a generic projection of $\boldsymbol{D}$ to linear subspace of dimension of order $\delta(C)$.

For $m \le n$ define

$$\overline{\kappa}_m^2(\boldsymbol{A}) := \mathbb{E}_{\boldsymbol{Q}}[\kappa(\boldsymbol{P}\boldsymbol{Q}\boldsymbol{A})^2].$$

Theorem B. *Let $C \subseteq \mathbb{R}^n$ be a closed convex cone and $\boldsymbol{A}\colon \mathbb{R}^n \to \mathbb{R}^n$ a non-singular linear map. Let $\eta \in (0,1)$ and assume that $m \ge \delta(C) + 2\sqrt{\log(2/\eta)m}$. Then*

$$\delta(\boldsymbol{A}C) \le \mathbb{E}_{\boldsymbol{Q}}\left[\mathscr{R}_C(\boldsymbol{P}\boldsymbol{Q}\boldsymbol{A})^2\right]\cdot\delta(C) + (n-m)\eta.$$

For the matrix condition number,

$$\delta(\boldsymbol{A}C) \le \overline{\kappa}_m^2(\boldsymbol{A})\cdot\delta(C) + (n-m)\eta. \tag{1.5}$$

As a consequence of Theorem B we get the following preconditioned version of Proposition 1.2.

Corollary 1.3. *If $f(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$, where g is a proper convex function and $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ is non-singular. Let $\eta \in (0,1)$ and assume that $m \ge \delta(C) + 2\sqrt{\log(2/\eta)m}$. Then*

$$\delta(f,\boldsymbol{x}_0) \le \mathbb{E}_{\boldsymbol{Q}}\left[\mathscr{R}_{\mathscr{D}(g,\boldsymbol{D}\boldsymbol{x}_0)}(\boldsymbol{P}\boldsymbol{Q}\boldsymbol{D})^2\right]\cdot\delta(g,\boldsymbol{D}\boldsymbol{x}_0) + (n-m)\eta \tag{1.6}$$

and

$$\delta(f,\boldsymbol{x}_0) \le \kappa_m^2(\boldsymbol{D})\cdot\delta(g,\boldsymbol{D}\boldsymbol{x}_0) + (n-m)\eta \tag{1.7}$$

Example 1.4. As an application, consider the matrix

$$\boldsymbol{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

The regularizer $f(\boldsymbol{x}) = \|\boldsymbol{D}\boldsymbol{x}\|_1$ is a one-dimensional version of a total variation regularizer, and is used to promote gradient sparsity. The standard method [ALMT14, Recipe 4.1] for computing the statistical dimension of the descent cone of f is not applicable here, as this regularizer is not separable [ZXCL16]. The standard condition number bound Theorem A is also not applicable, as it is known that the condition number satisfies $\kappa(\boldsymbol{D}) \ge \frac{2(n+1)}{\pi}$.

Using the preconditioned bounds we can determine the optimal m that minimizes the right-hand side of (1.7). For example, if $C = \mathscr{D}(g,\boldsymbol{D}\boldsymbol{x}_0)$ is a cone in $\mathbb{R}^{400}$ with $\delta(C) = 20$, when the following plot shows the best upper bound in (1.7) for various values of $m \ge \delta(C)$.

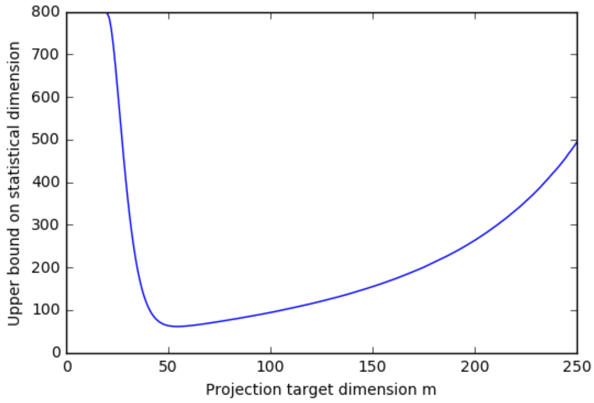


FIGURE 1. The condition number of $\boldsymbol{PD}$ reduces when projecting to a lower dimension m. However, the error bound in (1.7) limits how close to $\delta(C)$ the target dimension m can become.

Note that it is physically not possible, nor do we aim to, locate the precise phase transition of $f(\boldsymbol{x}) = \|\boldsymbol{Dx}\|_1$ in terms of that of the 1-norm, since the statistical dimension $\delta(f, \boldsymbol{x}_0)$ does not only depend on the sparsity pattern of $\boldsymbol{Dx}_0$, but also on the location of the support.

1.2. **Scope and limits of reduction.** The condition bounds in Theorem B naturally lead to the question of how to compute or bound the condition number of a random projection of a matrix,

$$\kappa(\boldsymbol{PQA}),$$

where $\boldsymbol{Q} \in O(n)$ is a random orthogonal matrix, and $\boldsymbol{P}$ a matrix selecting m rows of $\boldsymbol{Q}$. Using the multiplicative inequality $\kappa(\boldsymbol{PQA}) \leq \kappa(\boldsymbol{PQ})\kappa(\boldsymbol{A})$ (which holds if $\boldsymbol{A}$ is invertible), we get the reassurance that the condition number does not increase.

If $m = \lfloor \rho n \rfloor$ with $\rho \in (0,1)$, then in many cases the condition number $\kappa(\boldsymbol{PQA})$ remains bounded with high probability as $n \to \infty$. There are various ways to derive this fact. First of all, note that by the invariance under transposition we can equivalently study $\kappa(\boldsymbol{AQ})$, with $\boldsymbol{Q} \in \mathbb{R}^{n\times m}$ uniformly distributed on the Stiefel manifold. If $\boldsymbol{G} \sim N(\boldsymbol{0}, \boldsymbol{1})$ is a Gaussian random $n \times m$ matrix, then $\boldsymbol{Q} = \boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{G})^{-1/2}$ is uniformly distributed on the Stiefel manifold, so that $\kappa(\boldsymbol{AQ})$ has the same distribution as $\kappa(\boldsymbol{AG}(\boldsymbol{G}^T\boldsymbol{G})^{-1/2})$. We can then bound

$$\kappa\left(\boldsymbol{AG}(\boldsymbol{G}^T\boldsymbol{G})^{-1/2}\right) \leq \kappa\left(\boldsymbol{AG}\right)\kappa\left((\boldsymbol{G}^T\boldsymbol{G})^{-1/2}\right) = \kappa(\boldsymbol{AG})\kappa(\boldsymbol{G}),$$

transforming the problem into one in which the orthogonal matrix is replaced with a Gaussian one. The are various ways to tackle the resulting problem. One would be to appeal to the Hanson-Wright inequality [RV+13, Eft17], or more directly, the Bernstein inequality. Another bound follows from Gordon's inequality. More precisely, by a direct application of Gordon's inequality (Theorem B.1) as in [FR13, Theorem 9.21] one can show that for m small enough (so that the denominator is non-negative),

$$\kappa(\boldsymbol{AG}) \leq \frac{\|\boldsymbol{\sigma}\|_2 + \sqrt{m}\|\boldsymbol{\sigma}\|_\infty}{\|\boldsymbol{\sigma}\|_2 - \sqrt{m}\|\boldsymbol{\sigma}\|_\infty} \tag{1.8}$$

with high probability, where $\boldsymbol{\sigma}$ is the vector of singular values of $\boldsymbol{A}$. It would be interesting to characterize those matrices $\boldsymbol{A}$ for which $\kappa(\boldsymbol{PQA}) \approx 1$ using a kind of restricted isometry property, as for example in [ORS15]. We leave a detailed discussion of the probability distribution of $\kappa(\boldsymbol{PQA})$ and its remifications for another occasion, and instead consider a special case.

Example 1.5. Consider again the matrix $\boldsymbol{D}$ from Example 1.4. For $\rho \in \{0.2, 0.4, 0.6, 0.8\}$ and n ranging from 1 to 400, $m = \lfloor \rho n \rfloor$, we plot the average condition number $\kappa(\boldsymbol{DG})$, where $\boldsymbol{G} \in \mathbb{R}^{n\times m}$ is a Gaussian random matrix. As n increases, this condition number appears to converge to a constant value. Let's compare this with (1.8). The singular values of $\boldsymbol{D}$ are given by

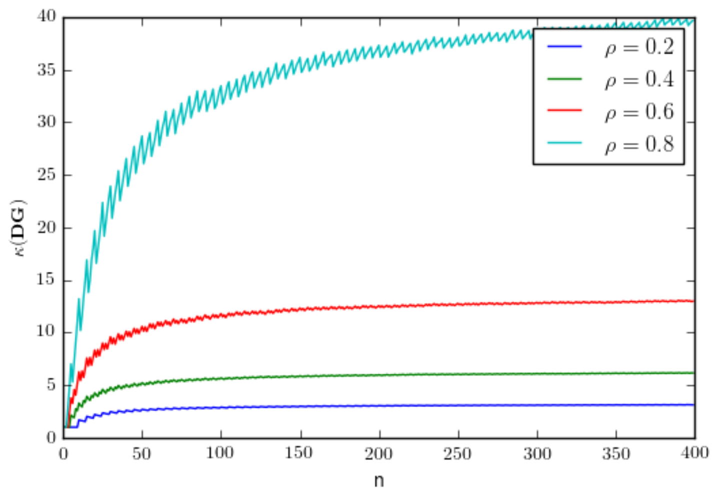


FIGURE 2. Condition number $\kappa(\boldsymbol{DG})$ for the matrix $\boldsymbol{D}$ from Example 1.4. $\boldsymbol{P}$ is the projection to the first $m = \lfloor \rho n \rfloor$ coordinates.

$$\sigma_k(\boldsymbol{D}) = \sqrt{2\left(1-\cos\left(\frac{k\pi}{n+1}\right)\right)},$$

from which we get $\|\boldsymbol{\sigma}\|_\infty \leq 2$. Using the trigonometric identity

$$\sum_{k=1}^{n} \cos(k\alpha) = \frac{\sin\left((n+\frac{1}{2})\alpha\right)}{2\sin(\frac{\alpha}{2})} - \frac{1}{2},$$

and the fact that $\sin((n+\frac{1}{2})\pi/(n+1)) = \sin(\pi/2(n+1))$, we get

$$\|\boldsymbol{\sigma}\|_2 = \sqrt{2n}.$$

Setting $m = \rho n$, the condition number thus concentrates on a value bounded by

$$\frac{\sqrt{2n}+2\sqrt{m}}{\sqrt{2n}-2\sqrt{m}} = \left(\frac{1+\sqrt{2\rho}}{1-\sqrt{2\rho}}\right)^2,$$

which is sensible if $\rho < 1/2$.

1.2.1. *A note on distributions.* The results presented are based in integral geometry, and as such depend crucially on $\boldsymbol{Q}$ being uniformly distributed in the orthogonal group with the Haar measure. By known universality results [OT15], the results are likely to carry over to other distributions. In the context of this paper, however, we are neither interested in actually preconditioning the matrices involved, nor are we using them as a model for observation or measurement matrices as is common in compressive sensing. The randomization here is merely a technical tool to improve bounds based on the condition number, and the question of whether this is a "realistic" distribution is of no concern.

1.3. **Conically restricted operators.** Theorem A is derived as a consequence of a much more general result in the setting of linear maps restricted to convex cones, and using a generalization of Renegar's condition number. The classical condition number of a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is the ratio of the operator norm and the smallest singular value. Using the notation

$$\|\boldsymbol{A}\| := \max_{\boldsymbol{x}\in S^{n-1}} \|\boldsymbol{A}\boldsymbol{x}\|, \qquad \sigma(\boldsymbol{A}) := \min_{\boldsymbol{x}\in S^{n-1}} \|\boldsymbol{A}\boldsymbol{x}\|,$$

the smallest singular value of $\boldsymbol{A}$ is given by $\max\{\sigma(\boldsymbol{A}), \sigma(\boldsymbol{A}^T)\}$, so that the classical condition number is given by

$$\kappa(\boldsymbol{A}) = \min\left\{\frac{\|\boldsymbol{A}\|}{\sigma(\boldsymbol{A})}, \frac{\|\boldsymbol{A}\|}{\sigma(\boldsymbol{A}^T)}\right\}.$$

With a view towards the convex feasibility problem, cf. Section 2.2, we introduce the following generalization: Let $C \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ be closed convex cones, and let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. We define the *restriction* of the linear operator $\boldsymbol{A}$ to C and D by

$$\boldsymbol{A}_{C\to D} \colon C \to D, \qquad \boldsymbol{A}_{C\to D}(\boldsymbol{x}) := \boldsymbol{\Pi}_D(\boldsymbol{A}\boldsymbol{x}), \tag{1.9}$$

where $\boldsymbol{\Pi}_D \colon \mathbb{R}^n \to D$ denotes the orthogonal projection, i.e., $\boldsymbol{\Pi}_D(\boldsymbol{y}) = \arg\min\{\|\boldsymbol{y}-\boldsymbol{z}\| : \boldsymbol{z} \in D\}$. Accordingly, we define restricted versions of the norm and the singular value:

$$\|\boldsymbol{A}\|_{C\to D} := \max_{\boldsymbol{x}\in C\cap S^{n-1}} \|\boldsymbol{A}_{C\to D}(\boldsymbol{x})\|, \qquad \sigma_{C\to D}(\boldsymbol{A}) := \min_{\boldsymbol{x}\in C\cap S^{n-1}} \|\boldsymbol{A}_{C\to D}(\boldsymbol{x})\|. \tag{1.10}$$

In Section 2 we will give a geometric interpretation of these quantities and describe how they appear in applications.

(A generalization of) Renegar's condition number is then defined as

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

and $\mathscr{R}_C(\boldsymbol{A}) := \mathscr{R}_{C,\mathbb{R}^m}(\boldsymbol{A})$. Besides the above mentioned application, Renegar's condition number has originally been used to estimate the running time of interior point algorithms that solve the convex feasibility problem (in the case $D = \mathbb{R}^m$). An average-case analysis of this condition number has been given in [AB13].

Renegar's condition number also features prominently in the context of linear inverse problems of the form (1.1). More precisely, consider the problem of recovering an unknown signal $\boldsymbol{x}_0 \in \mathbb{R}^n$ from noisy observations $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}$, with $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\|\boldsymbol{w}\| \le \varepsilon\|\boldsymbol{A}\|$, by solving the optimization problem

$$\text{minimize} \quad \|\boldsymbol{x}\|_* \quad \text{subject to} \quad \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\| \le \varepsilon\|\boldsymbol{A}\|,$$

where $\|.\|_*$ is a convex function, usually a suitably chosen norm (a typical example is when $\boldsymbol{x}_0$ is sparse and $\|\boldsymbol{x}\|_* = \|\boldsymbol{x}\|_1$). If $\hat{\boldsymbol{x}}$ is a solution of the above problem and $C = \mathscr{D}(\|.\|_*, \boldsymbol{x}_0)$ denotes the cone of descent directions of $\|.\|_*$, then the error $\|\hat{\boldsymbol{x}}-\boldsymbol{x}_0\|$ is easily seen to be bounded by

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}_0\| \le 2\varepsilon\cdot\mathscr{R}_C(\boldsymbol{A}).$$

Similar bounds can be derived for demixing problems using the biconic version of this condition number, see [MT13a] for a related approach.

1.4. **Moment functionals and the inequalities of Slepian and Gordon.** The moment functionals evaluated on a convex body K describe the moments of $h_K(\boldsymbol{g})$, where $h_K(\boldsymbol{x}) = \max_{\boldsymbol{z}\in K}\langle \boldsymbol{z}, \boldsymbol{x}\rangle$ is the support function of K (see Section 2.3) and $\boldsymbol{g}$ is a Gaussian vector,

$$\mu_f(K) = \mathbb{E}\left[f(h_K(\boldsymbol{g}))\right].$$

The prime example is the Gaussian width $w(K)$ of a set K, when $f(x) = x$, or the statistical dimension $\delta(C)$ of a cone C, when $K = C \cap B^n$ (B^n the closed unit ball in $\mathbb{R}^n$) and $f(x) = x^2$,

$$w(K) := \mathbb{E}\left[h_K(\boldsymbol{g})\right], \quad \delta(C) := \mathbb{E}\left[h_{C\cap B^n}(\boldsymbol{g})^2\right] = \mathbb{E}\left[\|\boldsymbol{\Pi}_C(\boldsymbol{g})\|^2\right].$$

Besides these, the expected value of the restricted norm,

$$\mathbb{E}\big[f(\|\boldsymbol{G}\|_{C\to D})\big],$$

can also be interpreted as a moment functional of a suitable tensor product construction associated to the cones C and D, see Section 3.3.

A standard way to bound the restricted norm and the restricted smallest singular value for Gaussian operators is by means of Slepian's and Gordon's inequalities [Gor85, LT91, DS01], see also Appendix B for a derivation. Slepian's lemma and its extensions can be interpretated as monotonicity properties of moment functionals with respect to contractions; direct consequences of these monotonicity properties are the bounds (1.11) and (1.12) in Theorem 1.6.

Theorem 1.6. *Let $C\subseteq\mathbb{R}^n$, $D\subseteq\mathbb{R}^m$ closed convex cones, and let $\boldsymbol{G}\in\mathbb{R}^{m\times n}$ be a Gaussian matrix and $\boldsymbol{g}\in\mathbb{R}^n$, $\boldsymbol{g}'\in\mathbb{R}^m$ independent Gaussian vectors. Then for $f\colon\mathbb{R}\to\mathbb{R}$ monotonically increasing and convex,*

$$\mathbb{E}\big[f(\|\boldsymbol{G}\|_{C\to D})\big]\le\mathbb{E}\big[f\big(\|\boldsymbol{\Pi}_D(\boldsymbol{g}')\|+\|\boldsymbol{\Pi}_C(\boldsymbol{g})\|\big)\big]. \tag{1.11}$$

If $\gamma\in\mathbb{R}$ denotes a standard Gaussian variable, which is independent of $\boldsymbol{G}$, then for every monotonically increasing f,

$$\mathbb{E}\big[f(\|\boldsymbol{G}\|_{C\to D}+\gamma)\big]\le\mathbb{E}\big[f\big(\|\boldsymbol{\Pi}_D(\boldsymbol{g}')\|+\|\boldsymbol{\Pi}_C(\boldsymbol{g})\|\big)\big], \tag{1.12}$$

$$\mathbb{E}\big[f(\sigma_{C\to D}(\boldsymbol{G})+\gamma)\big]\ge\mathbb{E}\big[f\big(\|\boldsymbol{\Pi}_D(\boldsymbol{g}')\|-\|\boldsymbol{\Pi}_C(\boldsymbol{g})\|\big)\big]. \tag{1.13}$$

The moment functionals generalize to support functions of *convex bundles*, which leads to the inequality (1.13). This inequality will not be used in this generality in this article, and we refer to the extended notes [AL14] for a proof and applications.

As a consequence of (1.11) we obtain a very general "master condition bound", from which Theorem A follows as a special case.

Theorem C. *Let $C\subseteq\mathbb{R}^n$ and $D\subseteq\mathbb{R}^m$ closed convex cones, $\boldsymbol{T}\in\mathbb{R}^{\ell\times n}$ and $\boldsymbol{U}\in\mathbb{R}^{p\times m}$. Then for $r\ge1$,*

$$\mathbb{E}\big[\|\tilde{\boldsymbol{G}}\|^r_{\boldsymbol{T}C\to\boldsymbol{U}D}\big]\le\mathscr{R}_C(\boldsymbol{T})^r\,\mathscr{R}_D(\boldsymbol{U})^r\;\mathbb{E}\big[\|\boldsymbol{G}\|^r_{C\to D}\big], \tag{1.14}$$

where $\tilde{\boldsymbol{G}}\in\mathbb{R}^{p\times\ell}$ and $\boldsymbol{G}\in\mathbb{R}^{m\times n}$ Gaussian matrices.

The idea of using Slepian's lemma to obtain condition number estimates for the Gaussian width of linear images of convex cones was suggested to us by Mike McCoy. We will see that (1.14) may fail for $r<1$. This will also show that (1.11) may fail if f is not convex.

1.5. **Conic integral geometry.** The theory of conic integral geometry centers around the *intrinsic volumes* $v_0(C),\dots,v_n(C)$, which are assigned to every closed convex cone $C\subseteq\mathbb{R}^n$. They form a discrete probability distribution on $\{0,\dots,n\}$ that captures statistical properties of the cone C. For example, the (conic) Steiner formula (Section 4.2) describes the Gaussian measure of a neighborhood of C, while the kinematic formulas (Section 4.4) describe the exact intersection probabilities of randomly oriented cones. Instead of giving the exact definition of these quantities (see Section 4.1), we remark that the moment generating function of this discrete probability distribution coincides, after a simple variable transformation, with the moment generating function of $\|\Pi_C(\boldsymbol{g})\|$, where $\boldsymbol{g}$ denotes as usual a standard Gaussian vector in $\mathbb{R}^n$ and Π_C denotes the orthogonal projection on C, as shown by McCoy and Tropp [MT13b]. Moreover, the expectation of the discrete probability distribution given by the intrinsic volumes, which is called the *statistical dimension* of C, coincides with the expectation of the squared projected length $\mathbb{E}\big[\|\Pi_C(\boldsymbol{g})\|^2\big]$: If X_C is a discrete random variable with $\mathbb{P}\{X_C=k\}=v_k(C)$, then

$$\delta(C):=\mathbb{E}[X_C]=\mathbb{E}\big[\|\Pi_C(\boldsymbol{g})\|^2\big].$$

A key insight related to this observation is that the intrinsic volumes *concentrate* sharply around their mean [MT14, ALMT14], see Section 4.1. A direct consequence of this concentration result is that randomly oriented convex cones exhibit an intersection behaviour reminiscent of that of linear subspaces, with the dimension replaced by the statistical dimension:

Theorem 1.7. *Let $\eta \in (0,1)$ and let C and D be convex cones in $\mathbb{R}^n$. Then for $\boldsymbol{Q} \in O(n)$ uniformly at random,*

$$\delta(C)+\delta(D) \le n - a_\eta\sqrt{n} \implies \mathbb{P}\{C \cap \boldsymbol{Q}D \ne \{\mathbf{0}\}\} \le \eta;$$
$$\delta(C)+\delta(D) \ge n + a_\eta\sqrt{n} \implies \mathbb{P}\{C \cap \boldsymbol{Q}D \ne \{\mathbf{0}\}\} \ge 1-\eta,$$

with $a_\eta := 2\sqrt{\log(2/\eta)}$.

Here, by uniformly at random in the orthogonal group $O(n)$, we mean distributed according to the normalized Haar measure. In Section 4.6 we take the intuition of convex cones behaving like linear subspaces further, by showing that random projections of convex cones to subspaces of dimension larger than the statistical dimension approximately preserve their statistical dimension.

Proposition 1.8. *Let $C \subset \mathbb{R}^n$ be a closed convex cone and let $\boldsymbol{T} \in \mathbb{R}^{m\times n}$. Let $\eta \in (0,1)$ and assume that $m \ge \delta(C) + 2\sqrt{\log(2/\eta)m}$. Then*

$$\delta(C) - (n-m)\eta \le \mathbb{E}_{\boldsymbol{Q}}[\delta(\boldsymbol{TQ}C)] \le \delta(C).$$

This observation follows from Crofton's Formula in the guise of (4.11) and holds the key to Theorem B.

1.6. **Organisation of the paper.** In Section 2 we introduce the setting of conically restricted linear operators, the biconic feasibility problem, and Renegar's condition number is some detail. The characterization of this condition number in the generality presented here is new and of independent interest. Section 3 introduces moment functionals and derives the first two identities of Theorem 1.6 and, as a corollary, Theorem C. Also, Theorem A and Proposition 1.2 are derived here. In Section 4 we change the scene and give a brief overview of conic integral geometry, culminating in a proof of Theorem B. Finally, in Section 5 we translate the results to the setting of convex regularizers. Appendix A presents some more details on the biconic feasibility problem, while Appendix B presents a general version of Gordon's inequality. While this version is more general than what is needed in this paper, it may be of independent interest.

1.7. **Acknowledgments.** We thank Mike McCoy and Joel Tropp for fruitful discussions on integral geometry, and in particular for suggesting the TQC Lemma, and Armin Eftekhari for helpful discussions on random projections.

2. Conically restricted linear operators

In this section we discuss the restriction of a linear operator to closed convex cones. Our focus will not be on the restriction itself (1.9), but rather on the restricted norm and the restricted (smallest) singular value (1.10), culminating in a discussion of Renegar's condition number. In Section 2.1 we derive general properties of these quantities and compare them to the unrestricted versions, in Section 2.2 we establish a relation to the generalized homogeneous feasibility problem, and in Section 2.3 we derive a convex geometric perspective.

2.1. **Restricted norm and restricted singular value.** Before discussing conically restricted operators, we record the following simple but useful lemma, which generalizes the relation $\ker \boldsymbol{A} = (\operatorname{im} \boldsymbol{A}^T)^\perp$.

Lemma 2.1. *Let $D \subseteq \mathbb{R}^m$ be a closed convex cone. Then the polar cone is the inverse image of the origin under the projection map, $D^\circ := \{\boldsymbol{z} \in \mathbb{R}^m : \langle \boldsymbol{y}, \boldsymbol{z}\rangle \le 0$ for all $\boldsymbol{y} \in D\} = \boldsymbol{\Pi}_D^{-1}(\boldsymbol{0})$. Furthermore, if $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, then*

$$\boldsymbol{A}^{-1}(D^\circ) = \left(\boldsymbol{A}^T D\right)^\circ, \tag{2.1}$$

where $\boldsymbol{A}^{-1}(D^\circ) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} \in D^\circ\}$ denotes the inverse image of D° under $\boldsymbol{A}$.

Proof. For the first claim, note that $\|\boldsymbol{\Pi}_D(\boldsymbol{z})\| = \max_{\boldsymbol{y}\in D\cap B^m}\langle \boldsymbol{z}, \boldsymbol{y}\rangle$, and $\max_{\boldsymbol{y}\in D\cap B^m}\langle \boldsymbol{z}, \boldsymbol{y}\rangle = 0$ is equivalent to $\langle \boldsymbol{z}, \boldsymbol{y}\rangle \le 0$ for all $\boldsymbol{y} \in D$, i.e., $\boldsymbol{z} \in D^\circ$.

For (2.1), let $\boldsymbol{x} \in \boldsymbol{A}^{-1}(D^\circ)$ and $\boldsymbol{y} \in D$. Then $\langle \boldsymbol{x}, \boldsymbol{A}^T\boldsymbol{y}\rangle = \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}\rangle \le 0$, as $\boldsymbol{A}\boldsymbol{x} \in D^\circ$. Therefore, $\boldsymbol{A}^{-1}(D^\circ) \subseteq (\boldsymbol{A}^T D)^\circ$. On the other hand, if $\boldsymbol{v} \in (\boldsymbol{A}^T D)^\circ$ and $\boldsymbol{y} \in D$, then $\langle \boldsymbol{A}\boldsymbol{v}, \boldsymbol{y}\rangle = \langle \boldsymbol{v}, \boldsymbol{A}^T\boldsymbol{y}\rangle \le 0$, so that $\boldsymbol{A}\boldsymbol{v} \in D^\circ$ and hence, $(\boldsymbol{A}^T D)^\circ \subseteq \boldsymbol{A}^{-1}(D^\circ)$. □

Recall from (1.10) that for $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^m$ closed convex cones, the restricted norm and singular value of $\boldsymbol{A}$ are defined by $\|\boldsymbol{A}\|_{C\to D} := \max\{\|\boldsymbol{A}_{C\to D}(\boldsymbol{x})\| : \boldsymbol{x} \in C \cap S^{n-1}\}$ and $\sigma_{C\to D}(\boldsymbol{A}) := \min\{\|\boldsymbol{A}_{C\to D}(\boldsymbol{x})\| : \boldsymbol{x} \in C\cap S^{n-1}\}$, respectively, where $\boldsymbol{A}_{C\to D}(\boldsymbol{x}) = \boldsymbol{\Pi}_D(\boldsymbol{A}\boldsymbol{x})$. The following proposition provides geometric conditions for the vanishing of the restricted norm or singular value.

Proposition 2.2. *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^m$ be closed convex cones. Then the restricted norm vanishes, $\|\boldsymbol{A}\|_{C\to D} = 0$, if and only if $C \subseteq (\boldsymbol{A}^T D)^\circ$. Furthermore, the restricted singular value vanishes, $\sigma_{C\to D}(\boldsymbol{A}) = 0$, if and only if $C \cap (\boldsymbol{A}^T D)^\circ \neq \{\boldsymbol{0}\}$, which is equivalent to $\boldsymbol{A}C \cap D^\circ \neq \{\boldsymbol{0}\}$ or $\ker \boldsymbol{A} \cap C \neq \{\boldsymbol{0}\}$.*

Proof. Using Lemma 2.1 we have $\boldsymbol{\Pi}_D(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0}$ if and only if $\boldsymbol{A}\boldsymbol{x} \in D^\circ$. This shows $\|\boldsymbol{A}\|_{C\to D} = 0$ if and only if $\boldsymbol{A}\boldsymbol{x} \in D^\circ$ for all $\boldsymbol{x} \in C\cap S^{n-1}$, or equivalently, $C \subseteq \boldsymbol{A}^{-1}(D^\circ) = (\boldsymbol{A}^T D)^\circ$ by (2.1). The claim about the restricted singular value follows similarly: $\sigma_{C\to D}(\boldsymbol{A}) = 0$ if and only if $\boldsymbol{A}\boldsymbol{x} \in D^\circ$ for some $\boldsymbol{x} \in C \cap S^{n-1}$, or equivalently, $C \cap \boldsymbol{A}^{-1}(D^\circ) \neq \{\boldsymbol{0}\}$. If $\boldsymbol{x} \in C\cap \boldsymbol{A}^{-1}(D^\circ)\setminus\{\boldsymbol{0}\}$, then either $\boldsymbol{A}\boldsymbol{x}$ is nonzero or $\boldsymbol{x}$ lies in the kernel of $\boldsymbol{A}$, which shows the second characterization. □

It is easily seen that the restricted norm is symmetric $\|\boldsymbol{A}\|_{C\to D} = \|\boldsymbol{A}^T\|_{D\to C}$,

$$\|\boldsymbol{A}\|_{C\to D} = \max_{\boldsymbol{x}\in C\cap B^m}\max_{\boldsymbol{y}\in D\cap B^n}\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}\rangle = \max_{\boldsymbol{y}\in D\cap B^n}\max_{\boldsymbol{x}\in C\cap B^m}\langle \boldsymbol{A}^T\boldsymbol{y}, \boldsymbol{x}\rangle = \|\boldsymbol{A}^T\|_{D\to C}. \tag{2.2}$$

Such a relation does not hold in general for the restricted singular value. In fact, in Section 2.2 we will see that, unless $C = D = \mathbb{R}^n$, the minimum of $\sigma_{C\to D}(\boldsymbol{A})$ and $\sigma_{D\to C}(-\boldsymbol{A}^T)$ is always zero, if C and D have nonempty interior, cf. (2.5). And if C or D is a linear subspace then $\sigma_{D\to C}(-\boldsymbol{A}^T) = \sigma_{D\to C}(\boldsymbol{A}^T)$.

Remark 2.3. In the case $C = \mathbb{R}^n$, $D = \mathbb{R}^m$, with $m \ge n$, one can characterize the smallest singular value of $\boldsymbol{A}$ as the inverse of the norm of the (Moore-Penrose) *pseudoinverse* of $\boldsymbol{A}$:

$$\sigma(\boldsymbol{A}) = \|\boldsymbol{A}^\dagger\|^{-1}.$$

Such a characterization does *not* hold in general for the restricted singular value, i.e., in general one cannot write $\sigma_{C\to D}(\boldsymbol{A})$ as $\|\boldsymbol{A}^\dagger\|_{D\to C}^{-1}$. Consider for example the case $D = \mathbb{R}^m$ and C a circular cone of angle α around some center $\boldsymbol{p} \in S^{n-1}$. Both cones have nonempty interior, but letting α go to zero, it is readily seen that $\sigma_{C\to D}(\boldsymbol{A})$ tends to $\|\boldsymbol{A}\boldsymbol{p}\|$, while $\|\boldsymbol{A}^\dagger\|_{D\to C}$ tends to $\|\boldsymbol{p}^T\boldsymbol{A}^\dagger\|$, which is in general not equal to $\|\boldsymbol{A}\boldsymbol{p}\|^{-1}$, unless $\boldsymbol{A}^T\boldsymbol{A} = \mathbf{I}_n$.

2.2. **The biconic feasibility problem.** The convex feasibility problem in the setting with two nonzero closed convex cones $C \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ is given as:

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

Using Lemma 2.1 and Proposition 2.2 we obtain the following characterizations of the primal feasible matrices $\mathscr{P}(C,D) := \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \text{(P) is feasible}\}$,

$$\mathscr{P}(C,D) \overset{(2.1)}{=} \left\{\boldsymbol{A} \in \mathbb{R}^{m\times n} : C \cap \left(\boldsymbol{A}^T D\right)^\circ \neq \{\boldsymbol{0}\}\right\} \overset{\text{[Prop. 2.2]}}{=} \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \sigma_{C\to D}(\boldsymbol{A}) = 0\}. \tag{2.3}$$

By symmetry, we obtain for the dual feasible matrices $\mathscr{D}(C,D) := \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \text{(D) is feasible}\}$,

$$\mathscr{D}(C,D) = \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : D \cap (-\boldsymbol{A}C)^\circ \neq \{\boldsymbol{0}\}\} = \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \sigma_{D\to C}(-\boldsymbol{A}^T) = 0\}. \tag{2.4}$$

In fact, we will see that $\sigma_{C\to D}(\boldsymbol{A})$ and $\sigma_{D\to C}(-\boldsymbol{A}^T)$ can be characterized as the distances to $\mathscr{P}(C,D)$ and $\mathscr{D}(C,D)$, respectively. We defer the proofs for this section to Appendix A.

In the following proposition we collect some general properties of $\mathscr{P}(C,D)$ and $\mathscr{D}(C,D)$.

Proposition 2.4. *Let $C \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ be closed convex cones with nonempty interior. Then*

(1) $\mathscr{P}(C,D)$ and $\mathscr{D}(C,D)$ are closed;

(2) the union of these sets is given by

$$\mathscr{P}(C,D) \cup \mathscr{D}(C,D) = \begin{cases} \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \det \boldsymbol{A} = 0\} & \text{if } C = D = \mathbb{R}^n \\ \mathbb{R}^{m\times n} & \text{else;} \end{cases}$$

(3) the intersection of these sets is nonempty but has zero (Lebesgue) volume, i.e.,

$$\mathbb{P}\big\{\boldsymbol{G} \in \mathscr{P}(C,D) \cap \mathscr{D}(C,D)\big\} = 0,$$

where $\boldsymbol{G} \in \mathbb{R}^{m\times n}$ Gaussian.

Note that from (2) and the characterizations (2.3) and (2.4) of $\mathscr{P}(C,D)$ and $\mathscr{D}(C,D)$, respectively, we obtain for every $\boldsymbol{A} \in \mathbb{R}^{m\times n}$: $\min\{\sigma_{C\to D}(\boldsymbol{A}), \sigma_{D\to C}(-\boldsymbol{A}^T)\} = 0$ or, equivalently,

$$\max\big\{\sigma_{C\to D}(\boldsymbol{A}), \sigma_{D\to C}(-\boldsymbol{A}^T)\big\} = \sigma_{C\to D}(\boldsymbol{A}) + \sigma_{D\to C}(-\boldsymbol{A}^T), \tag{2.5}$$

unless $C = D = \mathbb{R}^n$.

In the following we simplify the notation by writing $\mathscr{P}, \mathscr{D}$ instead of $\mathscr{P}(C,D), \mathscr{D}(C,D)$. For the announced interpretation of the restricted singular value as distance to $\mathscr{P}, \mathscr{D}$ we introduce the following notation: for $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ define

$$\operatorname{dist}(\boldsymbol{A}, \mathscr{P}) := \min\{\|\boldsymbol{\Delta}\| : \boldsymbol{A} + \boldsymbol{\Delta} \in \mathscr{P}\}, \qquad \operatorname{dist}(\boldsymbol{A}, \mathscr{D}) := \min\{\|\boldsymbol{\Delta}\| : \boldsymbol{A} + \boldsymbol{\Delta} \in \mathscr{D}\},$$

where as usual, the norm considered is the operator norm. The proof of the following proposition, given in Appendix A, follows along the lines of similar derivations in the case with a cone and a linear subspace [BF09].

Proposition 2.5. *Let $C \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ nonzero closed convex cones with nonempty interior. Then*

$$\operatorname{dist}(\boldsymbol{A}, \mathscr{P}) = \sigma_{C\to D}(\boldsymbol{A}), \qquad \operatorname{dist}(\boldsymbol{A}, \mathscr{D}) = \sigma_{D\to C}(-\boldsymbol{A}^T).$$

We finish this section by considering the intersection of $\mathscr{P}$ and $\mathscr{D}$, which we denote by

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

or simply Σ when the cones are clear from context. This set is usually referred to as the set of *ill-posed inputs*. As shown in Proposition 2.4, the set of ill-posed inputs, assuming $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^m$ each have nonempty interior, is a nonempty zero volume set. In the special case $C = \mathbb{R}^n$, $D = \mathbb{R}^m$,

$$\Sigma(\mathbb{R}^n, \mathbb{R}^m) = \{\text{rank deficient matrices in } \mathbb{R}^{m\times n}\}.$$

From (2.5) and Proposition 2.5 we obtain, if $(C,D) \neq (\mathbb{R}^n, \mathbb{R}^n)$,

$$\operatorname{dist}(\boldsymbol{A}, \Sigma) = \max\big\{\operatorname{dist}(\boldsymbol{A}, \mathscr{P}), \operatorname{dist}(\boldsymbol{A}, \mathscr{D})\big\} = \operatorname{dist}(\boldsymbol{A}, \mathscr{P}) + \operatorname{dist}(\boldsymbol{A}, \mathscr{D}).$$

The inverse distance to ill-posedness forms the heart of Renegar's condition number [Ren94, Ren95]. We denote

$$\mathscr{R}_{C,D}(\boldsymbol{A}) := \frac{\|\boldsymbol{A}\|}{\operatorname{dist}(\boldsymbol{A}, \Sigma(C,D))} = \min\left\{\frac{\|\boldsymbol{A}\|}{\sigma_{C\to D}(\boldsymbol{A})}, \frac{\|\boldsymbol{A}\|}{\sigma_{D\to C}(-\boldsymbol{A}^T)}\right\}. \tag{2.6}$$

Furthermore, we abbreviate the special case $D=\mathbb{R}^m$, which corresponds to the classical feasibility problem, by the notation

$$\mathscr{R}_C(\boldsymbol{A}) := \mathscr{R}_{C,\mathbb{R}^m}(\boldsymbol{A}). \tag{2.7}$$

Note that the usual matrix condition number is recovered in the case $C=\mathbb{R}^n$, $D=\mathbb{R}^m$,

$$\mathscr{R}_{\mathbb{R}^n}(\boldsymbol{A}) = \mathscr{R}_{\mathbb{R}^n,\mathbb{R}^m}(\boldsymbol{A}) = \kappa(\boldsymbol{A}).$$

Another simple but useful property is the symmetry $\mathscr{R}_{C,D}(\boldsymbol{A}) = \mathscr{R}_{D,C}(-\boldsymbol{A}^T)$. Finally, note that the restricted singular value has the following monotonicity properties

$$C\subseteq C' \Rightarrow \sigma_{C\to D}(\boldsymbol{A}) \geq \sigma_{C'\to D}(\boldsymbol{A}), \qquad D\subseteq D' \Rightarrow \sigma_{C\to D}(\boldsymbol{A}) \leq \sigma_{C\to D'}(\boldsymbol{A}).$$

This indicates that not necessarily $\mathscr{R}_C(\boldsymbol{A}) \leq \mathscr{R}_{C'}(\boldsymbol{A})$ if $C\subseteq C'$. But in the case $C'=\mathbb{R}^n$ and $m\geq n$ this inequality does hold, which we formulate in the following lemma.

Lemma 2.6. *Let $C\subseteq\mathbb{R}^n$ closed convex cone with nonempty interior and $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ with $m\geq n$. Then*

$$\mathscr{R}_C(\boldsymbol{A}) \leq \kappa(\boldsymbol{A}). \tag{2.8}$$

Proof. In the case $C=\mathbb{R}^n$ we have $\mathscr{R}_{\mathbb{R}^n}(\boldsymbol{A}) = \kappa(\boldsymbol{A})$. If $C\neq\mathbb{R}^n$ then $\boldsymbol{A}C\neq\mathbb{R}^m$, as $m\geq n$. It follows that $\mathbb{R}^m\cap(-\boldsymbol{A}C)^\circ\neq\{\boldsymbol{0}\}$, and thus $\sigma_{\mathbb{R}^m\to C}(-\boldsymbol{A}^T)=0$, cf. (2.4). Hence,

$$\mathscr{R}_C(\boldsymbol{A}) = \frac{\|\boldsymbol{A}\|}{\sigma_{C\to\mathbb{R}^m}(\boldsymbol{A})} \leq \frac{\|\boldsymbol{A}\|}{\sigma_{\mathbb{R}^n\to\mathbb{R}^m}(\boldsymbol{A})} = \kappa(\boldsymbol{A}). \qquad \square$$

The interesting case for convex optimizations is in fact where C is some self-dual cone like the nonnegative orthant or the cone of nonnegative definite matrices, and $m<n$. For these cases Renegar's condition number has found applications in the complexity analysis of convex optimization, see [BC13] for a discussion and references. For example, [VRPH07] provides an analysis of the running time of an interior-point algorithm for the convex feasibility problem in terms of this condition number. In [RBd15], Renegar's condition number is studied in the context of compressed sensing. In Section 3.4 we use $\mathscr{R}_C(\boldsymbol{A})$ for upper bounds of some important moment functionals.

2.3. **Support functions.** The restricted norm and the restricted singular value can be interpreted in terms of the support function of convex bodies. Recall that a set $K\subset\mathbb{R}^n$ is a convex body if K is nonempty, compact, and convex. The support function of a convex body $K\subset\mathbb{R}^n$ is given by

$$h_K\colon\mathbb{R}^n\to\mathbb{R}, \qquad h_K(\boldsymbol{x}) := \max_{\boldsymbol{z}\in K}\langle\boldsymbol{x},\boldsymbol{z}\rangle.$$

If $C\subseteq\mathbb{R}^n$ is a closed convex cone, and if $K=C\cap B^n$, where $B^n:=\{\boldsymbol{x}\in\mathbb{R}^n : \|\boldsymbol{x}\|\leq 1\}$, denotes the corresponding cone stub, then one readily verifies that

$$\|\Pi_C(\boldsymbol{x})\| = h_K(\boldsymbol{x}), \qquad \text{for all } \boldsymbol{x}\in\mathbb{R}^n. \tag{2.9}$$

If $K=\operatorname{conv}(C\cap S^{n-1})$ then one still has $\|\Pi_C(\boldsymbol{x})\| = h_K(\boldsymbol{x})$ for all $\boldsymbol{x}\notin\operatorname{int}(C^\circ)$, but in general one only gets an inequality:

$$\|\Pi_C(\boldsymbol{x})\| \geq h_K(\boldsymbol{x}), \qquad \text{for all } \boldsymbol{x}\in\mathbb{R}^n. \tag{2.10}$$

The restricted norm is related to the following construction: for convex bodies $K\subset\mathbb{R}^n$, $K'\subset\mathbb{R}^m$ define the *(convex) tensor product*

$$K\hat{\otimes}K' := \operatorname{conv}\{\boldsymbol{x}\otimes\boldsymbol{y} : \boldsymbol{x}\in K, \boldsymbol{y}\in K'\}\subset\mathbb{R}^{nm},$$

where $\otimes$ denotes the Kronecker product $\boldsymbol{x}\otimes\boldsymbol{y} = (x_1y_1,\dots,x_1y_n,x_2y_2,\dots,x_ny_m)$, which is a concrete model for the classical tensor product. An application of Carathéodory's theorem [Bar02, (2.4)] shows that $K\hat{\otimes}K'$ is again a convex body. If $\operatorname{vec}\colon\mathbb{R}^{m\times n}\to\mathbb{R}^{mn}$ denotes the function that concatenates the columns of the matrices, then

$$\langle\operatorname{vec}(\boldsymbol{A}),\boldsymbol{x}\otimes\boldsymbol{y}\rangle = \langle\boldsymbol{A}\boldsymbol{x},\boldsymbol{y}\rangle. \tag{2.11}$$

Lemma 2.7. *Let $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^m$ be closed convex cones, and let $K = C \cap B^n$ and $K' = D \cap B^m$ denote the corresponding cone stubs. Then*

$$\|\boldsymbol{A}\|_{C\to D} = h_{K \hat{\otimes} K'}(\mathrm{vec}(\boldsymbol{A})). \tag{2.12}$$

Proof. This follows by direct calculation:

$$\begin{aligned}\|\boldsymbol{A}\|_{C\to D} &= \max_{\boldsymbol{x}\in C\cap S^{n-1}} \|\Pi_D(\boldsymbol{A}\boldsymbol{x})\| = \max_{\boldsymbol{x}\in K}\|\Pi_D(\boldsymbol{A}\boldsymbol{x})\| \overset{(2.9)}{=} \max_{\boldsymbol{x}\in K,\boldsymbol{y}\in K'} \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}\rangle \overset{(2.11)}{=} \max_{\boldsymbol{x}\in K,\boldsymbol{y}\in K'} \langle \mathrm{vec}(\boldsymbol{A}), \boldsymbol{x}\otimes\boldsymbol{y}\rangle \\ &= h_{K\hat{\otimes}K'}(\mathrm{vec}(\boldsymbol{A})).\end{aligned}$$

Note that instead of the cone stub $C \cap B^n$ we could have taken $K = \mathrm{conv}(C \cap S^{n-1})$. ☐

Example 2.8. Choosing $C = \mathbb{R}^n$, $D = \mathbb{R}^m$, shows that the operator norm is given by the support function of $B^n \hat{\otimes} B^m$. As the dual of the operator norm is given by the Schatten-1 matrix norm $\|.\|_*$, which returns the sum of the singular values of a matrix, we obtain

$$B^n \hat{\otimes} B^m = \{\mathrm{vec}(\boldsymbol{A}) : \boldsymbol{A} \in \mathbb{R}^{n\times m}, \|\boldsymbol{A}\|_* \le 1\}. \tag{2.13}$$

In particular, the convex tensor product of two cone stubs is not necessarily a cone stub.

The restricted singular value has a similar description as the restricted norm in (2.12), for details on this construction we refer to [AL14].

3. Moment functionals of convex bodies

In this section we consider the moments of the random variable $h_K(\boldsymbol{g})$, where K is a convex body and $\boldsymbol{g}$ is a standard Gaussian vector of appropriate dimension. As shown in Section 2.3 this includes as a special case the moments of the restricted norm $\|\boldsymbol{G}\|_{C\to D}$, where $\boldsymbol{G}$ is a Gaussian matrix. In Section 3.1 we introduce the concept of moment functionals. In Section 3.2 we present Slepian's Lemma and an extension to higher moments as monotonicity properties of moment functionals. Section 3.3 and Section 3.4 describe applications of the extended Slepian's Lemma.

3.1. Introduction of moment functionals.

Recall that $\mathscr{K}(\mathbb{R}^n)$ denotes the set of convex bodies in $\mathbb{R}^n$; additionally, we define $\mathscr{K} := \bigcup_n \mathscr{K}(\mathbb{R}^n)$. In the following let $\boldsymbol{g}$ denote a standard Gaussian vector of appropriate dimension.

Definition 3.1. Let $f : \mathbb{R} \to \mathbb{R}$ be Borel measurable. The *f-moment functional* is defined by

$$\mu_f : \mathscr{K} \to \mathbb{R}, \qquad \mu_f(K) := \mathbb{E}\big[f(h_K(\boldsymbol{g}))\big]. \tag{3.1}$$

An important special case of a moment functional is the *Gaussian width* obtained by choosing $f = \mathrm{id}$. We denote this special functional by

$$w(K) := \mu_{\mathrm{id}}(K) = \mathbb{E}\big[h_K(\boldsymbol{g})\big]. \tag{3.2}$$

Another special case is the constant function $f \equiv 1$, which is in fact an emergence of the *Euler characteristic* $\mu_1(K) = \chi(K) = 1$.

The moment functionals are orthogonal invariant, $\mu_f(\boldsymbol{Q}K) = \mu_f(K)$ if $K \in \mathscr{K}(\mathbb{R}^n)$, $\boldsymbol{Q} \in O(n)$, and monotonic: if $f(x) \ge f(y)$ for all $x \ge y$, then so $\mu_f(K) \ge \mu_f(K')$ for all $K \supseteq K'$.

3.2. Contraction inequalities.

We have seen that for monotonically increasing $f : \mathbb{R} \to \mathbb{R}$, the functional μ_f is monotonically increasing under inclusion. Slepian's Lemma generalizes this monotonicity by weakening the inclusion assumption.

Definition 3.2. For a convex body $K \in \mathscr{K}$ we say that $M \subseteq K$ generates K if $K = \overline{\mathrm{conv}}(M)$ (the closure of the convex hull). For $K_1, K_2 \in \mathscr{K}$ we say that K_2 is a *contraction* of K_1 if there exists a 1-Lipschitz surjection $\varphi : M_1 \to M_2$ between generating sets M_1, M_2 of K_1, K_2. If additionally $\|\varphi(\boldsymbol{x})\| = \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in M_1$ then K_2 is a *norm-preserved contraction* of K_1.

If $\boldsymbol{0} \in K_1 \cap K_2$ we say that K_2 is a *$\boldsymbol{0}$-contraction* of K_1 if there exists a 1-Lipschitz surjection $\varphi : M_1 \to M_2$ such that additionally $\|\varphi(\boldsymbol{x})\| \le \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in M_1$.

The following proposition is a convex geometric formulation of (the generalized) Slepian's Inequality, which is a special case of Gordon's Theorem B.1.

Proposition 3.3. *Let* $K_1, K_2 \in \mathcal{K}$.

(1) If K_2 *is a contraction of* K_1*, then* $w(K_2) \le w(K_1)$.

(2) If K_2 *is a norm-preserved contraction of* K_1 *and* $f\colon \mathbb{R} \to \mathbb{R}$ *monotonically increasing, then* $\mu_f(K_2) \le \mu_f(K_1)$.

(3) If $\mathbf{0} \in K_1 \cap K_2$ *and* K_2 *is a* $\mathbf{0}$*-contraction of* K_1 *and* $f\colon \mathbb{R}_+ \to \mathbb{R}$ *monotonically increasing and convex, then* $\mu_f(K_2) \le \mu_f(K_1)$.

In the following sections we will see that the convexity assumption on f may not be dropped.

Proof. We will prove statement (3), statements (1) and (2) are proved similarly and follow from known versions of Slepian's inequality.

Let M_1, M_2 be generators of K_1, K_2, respectively, such that the properties of a $\mathbf{0}$-contraction in Definition 3.2 are satisfied with a corresponding set of map 1-Lipschitz map φ and the additional property $\|\varphi(\boldsymbol{x})\| \le \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in M_1$.

By a continuity argument we may assume that M_1 and M_2 are finite sets. Furthermore, by embedding the sets in a high-dimensional space, we may assume without loss of generality that each of M_1, M_2 has at most N elements,

$$M_1 = \{\boldsymbol{x}_j : 1 \le j \le N\}, \qquad\qquad M_2 = \{\boldsymbol{y}_j : 1 \le j \le N\},$$

allowing repetitions on the $\boldsymbol{x}_j$ and $\boldsymbol{y}_j$ as well. Define centered Gaussian random variables X_j, Y_j, $1 \le i \le m$, $1 \le j \le N$, via

$$X_j := \langle \boldsymbol{x}_j, \boldsymbol{g} \rangle, \qquad\qquad Y_j := \langle \boldsymbol{y}_j, \boldsymbol{g} \rangle,$$

where $\boldsymbol{g} \in \mathbb{R}^N$ is a standard Gaussian vector. The properties of the maps $\varphi_{\boldsymbol{x}}$ imply that

$$\begin{aligned} \mathbb{E}|X_j - X_\ell|^2 &\ge \mathbb{E}|Y_j - Y_\ell|^2, && \text{for all } j, \ell, \\ \mathbb{E}|X_j|^2 &\ge \mathbb{E}|Y_j|^2, && \text{for all } j. \end{aligned}$$

Since $\mathbf{0} \in K_1 \cap K_2$, we also have $\mu_f(K_1) = \mathbb{E}\max_j f_+(X_j)$ and $\mu_f(K_2) = \mathbb{E}\max_j f_+(Y_j)$, with $f_+(x) = f(x)$ for $x \ge 0$ and 0 else. Applying Theorem B.2 with $m = 1$ in the degenerate case $X_0 := Y_0 := 0$ thus yields

$$\mu_f(K_2) = \mathbb{E}\max_j f_+(Y_j) \le \mathbb{E}\max_j f_+(X_j) = \mu_f(K_1).$$

This completes the proof. □

3.3. **Moments of the restricted norm of a Gaussian matrix.** In this section we compare the moment functionals of the convex tensor product $K \hat{\otimes} K'$ with those of the direct product $K \times K'$. As a corollary we obtain the upper bounds (1.11) and (1.12) in Theorem 1.6.

Recall from (2.12) that the norm restricted to cones $C \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ can be expressed through the support function of the tensor product $K \hat{\otimes} K'$, where $K = C \cap B^n$ and $K' = D \cap B^m$, via $\|\boldsymbol{A}\|_{C \to D} = h_{K \hat{\otimes} K'}(\mathrm{vec}(\boldsymbol{A}))$. This implies that the moments of the restricted norm of a Gaussian matrix are given by the moment functional of the tensor product $K \hat{\otimes} K'$,

$$\mu_f(K \hat{\otimes} K') = \mathbb{E}\left[f(\|\boldsymbol{G}\|_{C \to D})\right], \tag{3.3}$$

where $\boldsymbol{G} \in \mathbb{R}^{m \times n}$ is a (standard) Gaussian matrix. On the other hand, the property $h_{K \times K'}(\boldsymbol{v}, \boldsymbol{v}') = h_K(\boldsymbol{v}) + h_{K'}(\boldsymbol{v}')$ implies that

$$\mu_f(K \times K') = \mathbb{E}\left[f(\|\boldsymbol{\Pi}_C(\boldsymbol{g})\| + \|\boldsymbol{\Pi}_D(\boldsymbol{g})\|)\right]. \tag{3.4}$$

To compare the tensor product with the direct product we consider the function $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} \otimes \boldsymbol{y}$. While this map is not necessarily a contraction , we will see that for cone stubs it is a $\mathbf{0}$-contraction, and that a slight modification gives a norm-preserved contraction. This leads to the proof of Theorem 1.6, which we restate here for convenience.

Theorem 3.4. *Let $C \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ closed convex cones, and let $\boldsymbol{G} \in \mathbb{R}^{m\times n}$ Gaussian matrix and $\boldsymbol{g} \in \mathbb{R}^n$, $\boldsymbol{g}' \in \mathbb{R}^m$ independent Gaussian vectors. Then for $f\colon \mathbb{R} \to \mathbb{R}$ monotonically increasing and convex,*

$$\mathbb{E}\big[f(\|\boldsymbol{G}\|_{C\to D})\big] \le \mathbb{E}\big[f\big(\|\boldsymbol{\Pi}_C(\boldsymbol{g})\| + \|\boldsymbol{\Pi}_D(\boldsymbol{g}')\|\big)\big]. \tag{3.5}$$

If $\gamma \in \mathbb{R}$ denotes a standard Gaussian variable, which is independent of $\boldsymbol{G}$, then for every monotonically increasing f,

$$\mathbb{E}\big[f(\|\boldsymbol{G}\|_{C\to D} + \gamma)\big] \le \mathbb{E}\big[f\big(\|\boldsymbol{\Pi}_C(\boldsymbol{g})\| + \|\boldsymbol{\Pi}_D(\boldsymbol{g}')\|\big)\big]. \tag{3.6}$$

Proof. Set $K := \operatorname{conv}(C \cap S^{n-1})$, $K_0 := C \cap B^n$, and $K_0' := D \cap B^m$. Note that K_0 is generated by $\{\boldsymbol{0}\} \cup (C \cap S^{n-1})$. In light of the identities (3.3) and (3.4), for the first statement we need to show that

$$\mu_f(K_0 \hat{\otimes} K_0') \le \mu_f(K_0 \times K_0').$$

By Proposition 3.3, this follows once we can show that the map $(\boldsymbol{x},\boldsymbol{y}) \mapsto \boldsymbol{x}\otimes\boldsymbol{y}$ is a $\boldsymbol{0}$-contraction.

To see this, consider the surjective map of generators

$$\varphi\colon \{\boldsymbol{0}\} \cup (C \cap S^{n-1}) \times (D \cap B^m) \to \{\boldsymbol{0}\} \cup (C \cap S^{n-1}) \otimes (D \cap B^m), \quad (\boldsymbol{x},\boldsymbol{y}) \mapsto \boldsymbol{x}\otimes\boldsymbol{y}.$$

This map is clearly satisfies $\|\varphi(\boldsymbol{x},\boldsymbol{y})\| = \|\boldsymbol{x}\otimes\boldsymbol{y}\| \le \|(\boldsymbol{x},\boldsymbol{y})\|$. To show that this is indeed a $\boldsymbol{0}$-contraction, it remains to see that the map is 1-Lipschitz. If $\|\boldsymbol{x}_1\| = \|\boldsymbol{x}_2\| = 1$, then

$$\|(\boldsymbol{x}_1,\boldsymbol{y}_1) - (\boldsymbol{x}_2,\boldsymbol{y}_2)\|^2 - \|\boldsymbol{x}_1\otimes\boldsymbol{y}_1 - \boldsymbol{x}_2\otimes\boldsymbol{y}_2\|^2 = 2\,(1 - \langle \boldsymbol{x}_1,\boldsymbol{x}_2\rangle)\,(1 - \langle \boldsymbol{y}_1,\boldsymbol{y}_2\rangle) \ge 0. \tag{3.7}$$

Furthermore, if $\boldsymbol{x}_1 = \boldsymbol{0}$, we have

$$\|(\boldsymbol{0},\boldsymbol{y}_1) - (\boldsymbol{x}_2,\boldsymbol{y}_2)\|^2 - \|\boldsymbol{0}\otimes\boldsymbol{y}_1 - \boldsymbol{x}_2\otimes\boldsymbol{y}_2\|^2 = \|\boldsymbol{x}_2\|^2 + \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|^2 - \|\boldsymbol{x}_2\|^2\|\boldsymbol{y}_2\|^2 \ge 0,$$

the last inequality being a consequence of $\|\boldsymbol{y}_2\| \le 1$ and $\|\boldsymbol{x}_2\| \in \{0,1\}$. This concludes the first claim.

The second claim follows analogously, where in this case one establishes

$$\mathbb{E}\big[f(\|\boldsymbol{G}\|_{C\to D} + \gamma)\big] = \mu_f\big((K \hat{\otimes} K_0') \times \{1\}\big) \le \mu_f(K \times K_0') = \mathbb{E}\big[f(h_K(\boldsymbol{g}) + \|\boldsymbol{\Pi}_D(\boldsymbol{g}')\|)\big], \tag{3.8}$$

by showing that $K \hat{\otimes} K_0') \times \{1\}$ is a norm-preserved contraction of the direct product $K \times K_0'$. The claim then follows by noting that $h_K(\boldsymbol{g}) \le \|\boldsymbol{\Pi}_C(\boldsymbol{g})\|$, see (2.10). □

The above proof actually shows a slightly stronger bound: from (3.8) and from the symmetry $\|\boldsymbol{G}\|_{C\to D} = \|\boldsymbol{G}^T\|_{C\to D}$, it follows

$$\mathbb{E}\big[f(\|\boldsymbol{G}\|_{C\to D} + \gamma)\big] \le \min\Big\{\mathbb{E}\big[f\big(h_K(\boldsymbol{g}) + \|\boldsymbol{\Pi}_D(\boldsymbol{g}')\|\big)\big], \mathbb{E}\big[f\big(\|\boldsymbol{\Pi}_C(\boldsymbol{g})\| + h_{K'}(\boldsymbol{g}')\big)\big]\Big\},$$

where $K = \operatorname{conv}(C \cap S^{n-1})$, $K' = \operatorname{conv}(D \cap S^{m-1})$.

3.4. **Linear images of cones.** In conic integral geometry the random variable $\|\boldsymbol{\Pi}_C(\boldsymbol{g})\|$, where $C \subseteq \mathbb{R}^n$ a closed convex cone and $\boldsymbol{g} \in \mathbb{R}^n$ a standard Gaussian vector, plays an important role (see Section 4 ahead). In fact, the norm of the projection is a special case of a cone-restricted norm:

$$\|\boldsymbol{\Pi}_C(\boldsymbol{g})\| = \|\boldsymbol{g}\|_{\mathbb{R}_+\to C}, \tag{3.9}$$

where on the right-hand side we interpret $\boldsymbol{g} \in \mathbb{R}^{n\times 1}$ as linear map. Using Theorem 1.6 we will derive estimates for the moments of $\|\tilde{\boldsymbol{G}}\|_{\boldsymbol{T}C\to\boldsymbol{U}D}$, where $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^m$ closed convex cones, $\boldsymbol{T} \in \mathbb{R}^{\ell\times n}$ and $\boldsymbol{U} \in \mathbb{R}^{p\times m}$, and $\tilde{\boldsymbol{G}}$ is a Gaussian $(p\times\ell)$-matrix.

We are now set to prove Theorem C. For convenience we restate it here.

Theorem 3.5. *Let $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^n$ closed convex cones, $\boldsymbol{T} \in \mathbb{R}^{\ell\times n}$ and $\boldsymbol{U} \in \mathbb{R}^{p\times m}$. Then for $r \ge 1$,*

$$\mathbb{E}\big[\|\tilde{\boldsymbol{G}}\|^r_{\boldsymbol{T}C\to\boldsymbol{U}D}\big] \le \mathscr{R}_C(\boldsymbol{T})^r\, \mathscr{R}_D(\boldsymbol{U})^r\, \mathbb{E}\big[\|\boldsymbol{G}\|^r_{C\to D}\big],$$

where $\tilde{\boldsymbol{G}} \in \mathbb{R}^{p\times\ell}$ and $\boldsymbol{G} \in \mathbb{R}^{m\times n}$ Gaussian matrices.

Applying this theorem to the special case (3.9) results in Theorem A, which we restate here as a corollary.

Corollary 3.6. *Let $C \subseteq \mathbb{R}^n$ closed convex cone, and $v_r(C) := \mathbb{E}[\|\boldsymbol{\Pi}_C(\boldsymbol{g})\|^r]$, where $\boldsymbol{g} \in \mathbb{R}^n$ Gaussian. Then for $\boldsymbol{A} \in \mathbb{R}^{\ell \times n}$, and $r \geq 1$,*

$$v_r(\boldsymbol{A}C) \leq \mathscr{R}_C(\boldsymbol{A})^r v_r(C). \tag{3.10}$$

In particular, if $\ell = n$, $\boldsymbol{A}$ is non-singular, and $r = 2$, then

$$\frac{\delta(C)}{\kappa(\boldsymbol{A})^2} \leq \delta(\boldsymbol{A}C) \leq \kappa(\boldsymbol{A})^2 \delta(C). \tag{3.11}$$

Proof. The first bound is just a special case of 3.5. The inequalities (3.11) follow from the inequality $\mathscr{R}_C(\boldsymbol{A}) \leq \kappa(\boldsymbol{A})$, cf. (2.8), and by considering $C = \boldsymbol{A}^{-1}\boldsymbol{A}C$ and using $\kappa(\boldsymbol{A}) = \kappa(\boldsymbol{A}^{-1})$ to obtain the lower bound. □

For the proof of Theorem 3.5 we need two auxiliary results.

Lemma 3.7. *Let $D \subseteq \mathbb{R}^m$ closed convex cone and $\boldsymbol{U} \in \mathbb{R}^{p \times m}$. Then*

$$\boldsymbol{U}D \cap B^p \subseteq \tfrac{1}{\lambda}\boldsymbol{U}(D \cap B^m), \tag{3.12}$$

with $\lambda := \max\{\sigma_{D \to \mathbb{R}^p}(\boldsymbol{U}), \sigma_{\mathbb{R}^p \to D}(-\boldsymbol{U}^T)\}$.

Proof. Let $\lambda_1 := \sigma_{D \to \mathbb{R}^p}(\boldsymbol{U})$, $\lambda_2 := \sigma_{\mathbb{R}^p \to D}(-\boldsymbol{U}^T)$. We will show in two steps that $\boldsymbol{U}D \cap B^p \subseteq \frac{1}{\lambda_1}\boldsymbol{U}(D \cap B^m)$ and $\boldsymbol{U}D \cap B^p \subseteq \frac{1}{\lambda_2}\boldsymbol{U}(D \cap B^m)$.

(1) Since $\boldsymbol{U}D \cap B^p$ as well as $\boldsymbol{U}(D \cap B^m)$ contain the origin, it suffices to show that $\boldsymbol{U}D \cap S^{p-1} \subseteq \frac{1}{\lambda_1}\boldsymbol{U}(D \cap B^m)$. Every element in $\boldsymbol{U}D \cap S^{p-1}$ can be written as $\frac{\boldsymbol{U}\boldsymbol{y}_0}{\|\boldsymbol{U}\boldsymbol{y}_0\|}$ for some $\boldsymbol{y}_0 \in D \cap S^{m-1}$, and since $\sigma_{D \to \mathbb{R}^p}(\boldsymbol{U}) = \min_{\boldsymbol{y} \in D \cap S^{m-1}} \|\boldsymbol{U}\boldsymbol{y}\| \leq \|\boldsymbol{U}\boldsymbol{y}_0\|$, we obtain $\sigma_{D \to \mathbb{R}^p}(\boldsymbol{U}) \frac{\boldsymbol{U}\boldsymbol{y}_0}{\|\boldsymbol{U}\boldsymbol{y}_0\|} \in \operatorname{conv}\{\boldsymbol{0}, \boldsymbol{U}\boldsymbol{y}_0\} \subseteq \boldsymbol{U}(D \cap B^m)$. This shows $\boldsymbol{U}D \cap S^{p-1} \subseteq \frac{1}{\lambda_1}\boldsymbol{U}(D \cap B^m)$.

(2) Recall from (2.4) that $\sigma_{\mathbb{R}^p \to D}(-\boldsymbol{U}^T) > 0$ only if $(\boldsymbol{U}D)^\circ = \{\boldsymbol{0}\}$, i.e., $\boldsymbol{U}D = \mathbb{R}^p$. Observe that

$$\sigma_{\mathbb{R}^p \to D}(-\boldsymbol{U}^T) = \min_{\boldsymbol{z} \in \mathbb{R}^p} \max_{\boldsymbol{y} \in D \cap B^m} \langle \boldsymbol{U}\boldsymbol{y}, \boldsymbol{z} \rangle = \max\{r \geq 0 : rB^p \subseteq \boldsymbol{U}(D \cap B^m)\}.$$

This shows $B^p \subseteq \frac{1}{\lambda_2}\boldsymbol{U}(D \cap B^m)$ and thus finishes the proof. □

Lemma 3.8. *Let K, K' be convex bodies such that $K = \operatorname{conv}(M)$ for some closed set $M \subseteq S^{n-1}$ and $\boldsymbol{0} \in K'$, and let $L := \operatorname{span}(K')$ the linear hull of K'. If $\boldsymbol{T}$ denotes a linear transformation on L, then $K \hat{\otimes} \boldsymbol{T}K'$ is a $\boldsymbol{0}$-contraction of $K \hat{\otimes} \|\boldsymbol{T}\| K'$.*

Proof. Note that the norm of the difference of two rank one matrices can be written as

$$\|\boldsymbol{x}_1 \otimes \boldsymbol{y}_1 - \boldsymbol{x}_2 \otimes \boldsymbol{y}_2\|^2 = \|\boldsymbol{x}_1\|^2\|\boldsymbol{y}_1\|^2 - 2\langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle\langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle + \|\boldsymbol{x}_2\|^2\|\boldsymbol{y}_2\|^2.$$

So for $\|\boldsymbol{x}_1\| = \|\boldsymbol{x}_2\| = 1$,

$$\begin{aligned}
&\|\boldsymbol{x}_1 \otimes \boldsymbol{y}_1 - \boldsymbol{x}_2 \otimes \boldsymbol{y}_2\|^2 - \|\boldsymbol{x}_1 \otimes \boldsymbol{z}_1 - \boldsymbol{x}_2 \otimes \boldsymbol{z}_2\|^2 \\
&= \|\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2 - \|\boldsymbol{z}_1\|^2 - \|\boldsymbol{z}_2\|^2 + 2\langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle\big(\langle \boldsymbol{z}_1, \boldsymbol{z}_2\rangle - \langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle\big) \\
&\geq \begin{cases}
\|\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2 - 2\langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle - \|\boldsymbol{z}_1\|^2 - \|\boldsymbol{z}_2\|^2 + 2\langle \boldsymbol{z}_1, \boldsymbol{z}_2\rangle & \\
= \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|^2 - \|\boldsymbol{z}_1 - \boldsymbol{z}_2\|^2 & \text{if } \langle \boldsymbol{z}_1, \boldsymbol{z}_2\rangle \leq \langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle \\
\|\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2 + 2\langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle - \|\boldsymbol{z}_1\|^2 - \|\boldsymbol{z}_2\|^2 - 2\langle \boldsymbol{z}_1, \boldsymbol{z}_2\rangle & \\
= \|\boldsymbol{y}_1 + \boldsymbol{y}_2\|^2 - \|\boldsymbol{z}_1 + \boldsymbol{z}_2\|^2 & \text{if } \langle \boldsymbol{z}_1, \boldsymbol{z}_2\rangle \geq \langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle.
\end{cases}
\end{aligned}$$

Setting $\varphi\colon M \otimes \|\boldsymbol{T}\|K' \to M \otimes \boldsymbol{T}K'$, $\varphi(\boldsymbol{x} \otimes \|\boldsymbol{T}\|\boldsymbol{x}') := \boldsymbol{x} \otimes \boldsymbol{T}\boldsymbol{x}'$, we have $\|\varphi(\boldsymbol{x} \otimes \|\boldsymbol{T}\|\boldsymbol{x}')\| = \|\boldsymbol{T}\boldsymbol{x}'\| \leq \|\boldsymbol{T}\|\,\|\boldsymbol{x}'\| = \|\boldsymbol{x} \otimes \|\boldsymbol{T}\|\boldsymbol{x}'\|$, and from the above computation, with $\boldsymbol{y}_i = \|\boldsymbol{T}\|\boldsymbol{x}_i'$ and $\boldsymbol{z}_i = \boldsymbol{T}\boldsymbol{x}_i'$, $i = 1, 2$, we obtain either

$$\|\boldsymbol{x}_1 \otimes \|\boldsymbol{T}\|\boldsymbol{x}_1' - \boldsymbol{x}_2 \otimes \|\boldsymbol{T}\|\boldsymbol{x}_2'\|^2 - \|\boldsymbol{x}_1 \otimes \boldsymbol{T}\boldsymbol{x}_1' - \boldsymbol{x}_2 \otimes \boldsymbol{T}\boldsymbol{x}_2'\|^2 \geq \|\boldsymbol{T}\|^2\,\|\boldsymbol{x}_1' - \boldsymbol{x}_2'\|^2 - \|\boldsymbol{T}(\boldsymbol{x}_1' - \boldsymbol{x}_2')\|^2 \geq 0,$$

or

$$\|\boldsymbol{x}_1\otimes\|\boldsymbol{T}\|\boldsymbol{x}_1'-\boldsymbol{x}_2\otimes\|\boldsymbol{T}\|\boldsymbol{x}_2'\|^2-\|\boldsymbol{x}_1\otimes\boldsymbol{T}\boldsymbol{x}_1'-\boldsymbol{x}_2\otimes\boldsymbol{T}\boldsymbol{x}_2'\|^2\ge\|\boldsymbol{T}\|^2\,\|\boldsymbol{x}_1'+\boldsymbol{x}_2'\|^2-\|\boldsymbol{T}(\boldsymbol{x}_1'+\boldsymbol{x}_2')\|^2\ge 0.$$

This shows that $K\hat{\otimes}\boldsymbol{T}K'$ is a $\mathbf{0}$-contraction of $K\hat{\otimes}\|\boldsymbol{T}\|K'$. □

Proof of Theorem 3.5. Assume first that $\ell=n$ and $\boldsymbol{T}=\mathbf{I}_n$. As in the proof of Lemma 3.7, we let $\lambda:=\max\{\sigma_{D\to\mathbb{R}^p}(\boldsymbol{U}),\sigma_{\mathbb{R}^p\to D}(-\boldsymbol{U}^T)\}$, so that

$$\begin{aligned}\mathbb{E}\big[\|\boldsymbol{G}\|_{C\to\boldsymbol{U}D}^r\big]&=\mathbb{E}\Big[\Big(\max_{\boldsymbol{x}\in C\cap S^{n-1}}\max_{\boldsymbol{y}\in\boldsymbol{U}D\cap B^p}\langle\boldsymbol{G}\boldsymbol{x},\boldsymbol{y}\rangle\Big)^r\Big]\\&\le\lambda^{-r}\,\mathbb{E}\Big[\Big(\max_{\boldsymbol{x}\in C\cap B^n}\max_{\boldsymbol{y}\in\boldsymbol{U}(D\cap B^m)}\langle\boldsymbol{G}\boldsymbol{x},\boldsymbol{y}\rangle\Big)^r\Big]=\lambda^{-r}\mu_f(K\hat{\otimes}\boldsymbol{U}K'),\end{aligned}$$

where $f(t):=t^r$, $K:=C\cap B^n$, and $K':=D\cap B^m$. From Lemma 3.8 and Slepian's Inequality (3) in Proposition 3.3 we obtain $\mu_f(K\hat{\otimes}\boldsymbol{U}K')\le\|\boldsymbol{U}\|^r\mu_f(K\hat{\otimes}K')=\|\boldsymbol{U}\|^r\,\mathbb{E}\big[\|\boldsymbol{G}\|_{C\to D}^r\big]$, so that

$$\mathbb{E}\big[\|\boldsymbol{G}\|_{C\to\boldsymbol{U}D}^r\big]\le\left(\frac{\|\boldsymbol{U}\|}{\max\{\sigma_{D\to\mathbb{R}^p}(\boldsymbol{U}),\sigma_{\mathbb{R}^p\to D}(-\boldsymbol{U}^T)\}}\right)^r\mathbb{E}\big[\|\boldsymbol{G}\|_{C\to D}^r\big]=\mathscr{R}_D(\boldsymbol{U})^r\;\mathbb{E}\big[\|\boldsymbol{G}\|_{C\to D}^r\big].$$

This shows the claim for $\ell=n$ and $\boldsymbol{T}=\mathbf{I}_n$. For the general case we use the symmetry of the restricted norm,

$$\begin{aligned}\mathbb{E}\big[\|\boldsymbol{G}\|_{\boldsymbol{T}C\to\boldsymbol{U}D}^r\big]&\le\mathscr{R}_D(\boldsymbol{U})^r\;\mathbb{E}\big[\|\boldsymbol{G}\|_{\boldsymbol{T}C\to D}^r\big]=\mathscr{R}_D(\boldsymbol{U})^r\;\mathbb{E}\big[\|-\boldsymbol{G}^T\|_{D\to\boldsymbol{T}C}^r\big]\\&\le\mathscr{R}_C(\boldsymbol{T})^r\,\mathscr{R}_D(\boldsymbol{U})^r\;\mathbb{E}\big[\|-\boldsymbol{G}^T\|_{D\to C}^r\big]=\mathscr{R}_C(\boldsymbol{T})^r\,\mathscr{R}_D(\boldsymbol{U})^r\;\mathbb{E}\big[\|\boldsymbol{G}\|_{C\to D}^r\big].\end{aligned}$$ □

Note that we can improve the upper bound by using Renegar's condition number. While the above bound can significantly improve on Theorem A, it may still be trivial in cases where the statistical dimension is large with respect to the ambient space. An improvement is given by Proposition 1.2, which we restate and prove here.

Proposition 3.9. *Let $C\subseteq\mathbb{R}^n$ be a closed convex cone, and $\delta(C)$ the statistical dimension of C. Then for $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ non-singular,*

$$\delta(\boldsymbol{A}C)\le\kappa(\boldsymbol{A})^{-2}\cdot\delta(C)+\big(1-\kappa(\boldsymbol{A})^{-2}\big)\cdot n.$$

Proof. We have

$$\begin{aligned}\delta(\boldsymbol{A}C)&\overset{(1)}{=}n-\delta(\boldsymbol{A}^{-T}C^\circ)\\&\overset{(2)}{\le}n-\kappa(\boldsymbol{A})^{-2}\delta(C^\circ)\\&\overset{(3)}{=}n-\kappa(\boldsymbol{A})^{-2}(n-\delta(C))\\&=\kappa(\boldsymbol{A})^{-2}\cdot\delta(C)+\big(1-\kappa(\boldsymbol{A})^{-2}\big)\cdot n,\end{aligned}$$

where for (1) we used (4.7) and Lemma 2.1, for (2) we used Theorem A, and for (3) we used (4.7) again. □

4. Conic integral geometry

In this section we use integral geometry to develop the tools needed for deriving a preconditioned bound in Theorem B. A comprehensive treatment of integral geometry can be found in [SW08b], while a self-contained treatment in the setting of polyhedral cones, which uses our language, is given in [AL17].

4.1. **Intrinsic volumes.** The theory of conic integral geometry is based on the *intrinsic volumes* $v_0(C),\dots,v_n(C)$ of a closed convex cone $C\subseteq\mathbb{R}^n$. The intrinsic volumes form a discrete probability distribution on $\{0,\dots,n\}$ that capture statistical properties of the cone C. For a polyhedral cone C and $0\le k\le n$, the intrinsic volumes can be defined as

$$v_k(C)=\mathbb{P}\{\boldsymbol{\Pi}_C(\boldsymbol{g}) \text{ lies in relative interior of } k\text{-dimensional face of } C\}.$$

Example 4.1. Let $C=L\subseteq\mathbb{R}^n$ be a linear subspace of dimension i. Then

$$v_k(C)=\begin{cases}1 & \text{if } k=i,\\ 0 & \text{if } k\neq i.\end{cases}$$

Example 4.2. Let $C=\mathbb{R}^n_{\ge 0}$ be the non-negative orthant, i.e., the cone consisting of points with non-negative coordinates. A vector $\boldsymbol{x}$ projects orthogonally to a k-dimensional face of C if and only if exactly k coordinates are non-positive. By symmetry considerations and the invariance of the Gaussian distribution under permutations of the coordinates, it follows that

$$v_k(\mathbb{R}^n_{\ge 0})=\binom{n}{k}2^{-n}.$$

For non-polyhedral closed convex cones, the intrinsic volumes can be defined by polyhedral approximation. To avoid having to explicitly take care of upper summation bounds in many formulas, we use the convention that $v_k(C)=0$ if $C\subseteq\mathbb{R}^n$ and $k>n$ (that this is not just a convention follows from the fact that intrinsic volumes are "intrinsic", i.e., not dependent on the dimension of the space in which C lives).

The following important properties of the intrinsic volumes, which are easily verified in the setting of polyhedral cones, will be used frequently:

(a) **Orthogonal invariance.** For an orthogonal transformation $\boldsymbol{Q}\in O(n)$,

$$v_k(\boldsymbol{Q}C)=v_k(C);$$

(b) **Polarity.**

$$v_k(C)=v_{n-k}(C^\circ);$$

(c) **Product rule.**

$$v_k(C\times D)=\sum_{i+j=k}v_i(C)v_j(D). \tag{4.1}$$

In particular, if $D=L$ is a linear subspace of dimension j, then $v_{k+j}(C\times L)=v_k(C)$.

(d) **Gauss-Bonnet.**

$$\sum_{k=0}^n(-1)^k v_k(C)=\begin{cases}0 & \text{if } C \text{ is not a linear subspace},\\ 1 & \text{else}.\end{cases} \tag{4.2}$$

4.2. **The Steiner formula.** The intrinsic volumes are the essential ingredients in the (generalized) Steiner formula, which in its original formulation gives an expression for the measure of the neighborhood of a convex cone:

$$\mathbb{P}\{\|\Pi_C(\boldsymbol{g})\|\ge r\}=\sum_{i=0}^n v_i(C)\,\mathbb{P}\{\chi_i\ge r\}, \tag{4.3}$$

where $\chi_0=0$ and $\chi_1,\dots,\chi_n$ are independent chi-distributed random variables with χ_i having i degrees of freedom.

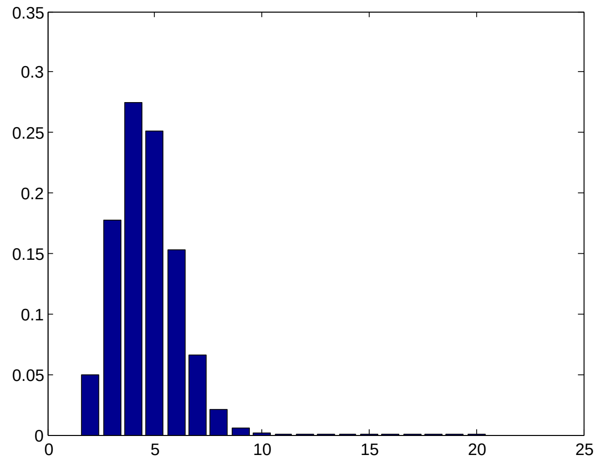


FIGURE 3. Intrinsic volumes of the cone $C = \{\boldsymbol{x} : x_1 \le \cdots \le x_n\}$.

A powerful generalization of the Steiner formula (4.3) was derived in [MT14], which we state for completeness and later reference: if $f: \mathbb{R}^2_+ \to \mathbb{R}$ is a Borel function and $C \subseteq \mathbb{R}^n$ a closed convex cone, then

$$\mathbb{E}\big[f(\|\boldsymbol{\Pi}_C(\boldsymbol{g})\|, \|\boldsymbol{\Pi}_{C^\circ}(\boldsymbol{g})\|)\big] = \sum_{i=0}^{n} v_i(C)\, \mathbb{E}\big[f(\chi_i, \chi'_{n-i})\big], \tag{4.4}$$

where $\chi_0 = \chi'_0 = 0$ and $\chi_1, \dots, \chi_n, \chi'_1, \dots, \chi'_n$ are independent chi-distributed random variables with χ_i and χ'_i having i degrees of freedom.

The Steiner formula allows one to express certain moment functionals as linear combinations of intrinsic volumes.

Example 4.3 (circular cones)**.** In this example we use the Steiner formula to compute the moment functionals for circular cones, and show that the convexity assumption in Proposition 3.3 can not be dropped. Let $\mathrm{Circ}_n(\alpha) = \{\boldsymbol{x} \in \mathbb{R}^n : x_1 \ge \|\boldsymbol{x}\| \cos\alpha\}$ denote the circular cone of radius α around the first coordinate vector. For our purposes it is more convenient to use $\tan\alpha$ instead of α, so we define $C_n(t) := \mathrm{Circ}_n(\arctan(t))$. Consider the linear map $\boldsymbol{T} := \mathrm{diag}(1, s, \dots, s)$ with $s \ge 1$, whose condition number is $\kappa(\boldsymbol{T}) = s$. Then $\boldsymbol{T}C_n(t) = C_n(st)$. Recall the definition $\nu_r(C) = \mathbb{E}[\|\boldsymbol{\Pi}_C(\boldsymbol{g}\|^r]$ for the r-th moment of the projected length. By (3.11) we have

$$\frac{s^r \nu_r(C_n(t))}{\nu_r(C_n(st))} \ge 1 \tag{4.5}$$

for $r \ge 1$. Using the generalized Steiner formula (4.4) we can express $\nu_r(C)$ in terms of the intrinsic volumes of C: for $r > 0$

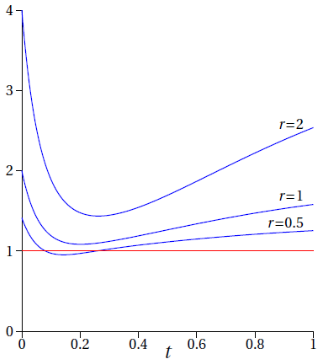
$$\nu_r(C) = \sum_{j=1}^{n} v_j(C)\, \mathbb{E}[\|\boldsymbol{g}_j\|^r] = \sum_{j=1}^{n} v_j(C)\, \frac{2^{r/2}\,\Gamma(\frac{j+r}{2})}{\Gamma(\frac{j}{2})},$$

where $\boldsymbol{g}_j \in \mathbb{R}^j$ denotes a standard Gaussian vector. The intrinsic volumes of the circular cones are given by, cf. [Ame11, Ex. 4.4.8]

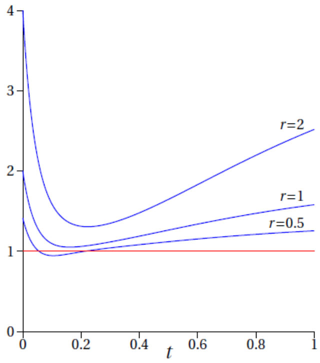
$$v_j(C_n(t)) = \frac{\Gamma(\frac{n}{2})\, t^j}{2\,\Gamma(\frac{j+1}{2})\Gamma(\frac{n-j+1}{2})\,(1+t^2)^{(m-2)/2}}, \quad \text{for } j = 1, \dots, n-1,$$

$$v_n(C_n(t)) = \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi}\,\Gamma(\frac{n-1}{2})} \int_0^t \frac{\tau^{n-2}}{(1+\tau^2)^{n/2}}\, d\tau.$$

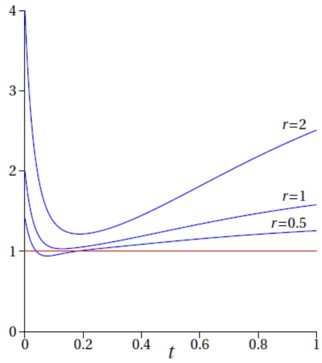
Using these formulas we can compute $\nu_r(D_n(t))$.



(A) $m = 50$



(B) $m = 100$



(C) $m = 200$

FIGURE 4. Plot of the quotient $s^r v_r(C_n(t))/v_r(C_n(st))$, cf. (4.5), with $s = 2$, $m \in \{50, 100, 200\}$, $r \in \{0.5, 1, 2\}$.

Figure 4 shows a plot of the quotient in (4.5) for $s = 2$, $n \in \{50, 100, 200\}$, and $r \in \{0.5, 1, 2\}$. The plot shows that the inequality (4.5) may be violated if $r < 1$, which ultimately shows that the convexity assumptions in Proposition 3.3 may not be dropped. The plots also indicate that the inequality (4.5) is asymptotically sharp for $n \to \infty$. This could be shown with an analysis similar to the one given in [MT13b, Sec. 6.3]; we leave the details to the interested reader.

4.3. **The statistical dimension.** In what follows it will be convenient to work with reparametrizations of the intrinsic volumes, namely the tail and half-tail functionals

$$t_k(C) = \sum_{i \geq 0} v_{k+i}(C), \qquad h_k(C) = 2 \sum_{i \geq 0 \text{ even}} v_{k+i}(C),$$

which are defined for $0 \leq k \leq n$. Adding (or subtracting) the Gauss-Bonnet relation (4.2) to the identity $\sum_{i \geq 0} v_i(C) = 1$, we see that $h_0(C) = h_1(C) = 1$ if C is not a linear subspace, so that the sequences $2v_0(C), 2v_2(C), \ldots$ and $2v_1(C), 2v_3(C), \ldots$ are probability distributions in their own right. Moreover, we have the interleaving property

$$t_{i+1}(C) \leq h_i(C) \leq t_i(C).$$

The intrinsic volumes can be recovered from the half-tail functionals as

$$v_i(C) = \begin{cases} \frac{1}{2}(h_i(C) - h_{i+2}(C)) & \text{for } 0 \leq i \leq n-2, \\ \frac{1}{2} h_i(C) & \text{else.} \end{cases} \tag{4.6}$$

An important summary parameter is the *statistical dimension* of a cone C, defined as the expected value of the intrinsic volumes considered as probability distribution:

$$\delta(C) = \sum_{k=0}^{n} k v_k(C) = \frac{1}{2} h_1(C) + \sum_{i \geq 2} h_i(C).$$

The statistical dimension coincides with the expected squared norm of the projection of a Gaussian vector on the cone, $\delta(C) = \mathbb{E}\left[\|\boldsymbol{\Pi}_C(\boldsymbol{g})\|^2\right]$, and is therefore an instance of a moment function (see Section 3). Moreover, it differs from the squared Gaussian width by at most 1,

$$w^2(C) \leq \delta(C) \leq w^2(C) + 1,$$

see [ALMT14, Proposition 10.2].

The statistical dimension reduces to the usual dimension for linear subspaces, and also extends various properties of the dimension to closed convex cones $C \subseteq \mathbb{R}^n$:

(a) **Orthogonal invariance.** For an orthogonal transformation $\boldsymbol{Q} \in O(n)$,

$$\delta(\boldsymbol{Q}C) = \delta(C);$$

(b) **Complementarity.**

$$\delta(C) + \delta(C^\circ) = n; \tag{4.7}$$

This generalizes the relation $\dim L + \dim L^\perp = n$ for a linear subspace $L \subseteq \mathbb{R}^n$.

(c) **Additivity.**

$$\delta(C \times D) = \delta(C) + \delta(D).$$

(d) **Monotonicity.**

$$\delta(C) \le \delta(D) \text{ if } C \subseteq D.$$

The analogy with linear subspaces will be taken further when discussing concentration of intrinsic volumes, see Section 4.5.

4.4. **The kinematic formulas.** The intrinsic volumes allow to study the properties of random intersections of cones via the *kinematic formulas*. A self-contained proof of these formulas for polyhedral cones is given in [AL17, Section 5]. In what follows, when we say that $\boldsymbol{Q}$ is drawn uniformly at random from the orthogonal group $O(d)$, we mean that it is drawn from the Haar probability measure ν on $O(n)$. This is the unique regular Borel measure on $O(n)$ that is left and right invariant ($\nu(\boldsymbol{Q}A) = \nu(A\boldsymbol{Q}) = \nu(A)$ for $\boldsymbol{Q} \in O(n)$ and a Borel measurable $A \subseteq O(n)$) and satisfies $\nu(O(n)) = 1$. Moreover, for measurable $f\colon O(n) \to \mathbb{R}_+$, we write

$$\mathbb{E}_{\boldsymbol{Q}\in O(n)}[f(\boldsymbol{Q})] := \int_{\boldsymbol{Q}\in O(n)} f(\boldsymbol{Q})\ \nu(\mathrm{d}\boldsymbol{Q})$$

for the integral with respect to the Haar probability measure, and we will occasionally omit the subscript $\boldsymbol{Q} \in O(n)$, or just write $\boldsymbol{Q}$ in the subscript, when there is no ambiguity.

Theorem 4.4 (Kinematic Formula)**.** *Let $C, D \subseteq \mathbb{R}^n$ be polyhedral cones. Then, for $\boldsymbol{Q} \in O(n)$ uniformly at random, and $k > 0$,*

$$\mathbb{E}[v_k(C \cap \boldsymbol{Q}D)] = v_{k+n}(C \times D), \qquad \mathbb{E}[v_0(C \cap \boldsymbol{Q}D)] = t_0(C \times D). \tag{4.8}$$

If $D = L$ is a linear subspace of dimension $n - m$, then

$$\mathbb{E}[v_k(C \cap \boldsymbol{Q}L)] = v_{k+m}(C), \qquad \mathbb{E}[v_0(C \cap \boldsymbol{Q}L)] = \sum_{j=0}^{m} v_j(C). \tag{4.9}$$

Combining Theorem 4.4 with the Gauss-Bonnet relation (4.2) yields the so-called *Crofton formulas*, which we formulate in the following corollary. The intersection probabilities are also know as Grassmann angles in the literature (see [AL17, 2.33] for a discussion and references).

Corollary 4.5. *Let $C, D \subseteq \mathbb{R}^n$ be polyhedral cones such that not both of C and D are linear subspaces, and let $L \subset \mathbb{R}^n$ be a linear subspace of dimension $n - m$. Then, for $\boldsymbol{Q} \in O(n)$ uniformly at random,*

$$\mathbb{P}\{C \cap \boldsymbol{Q}D \ne \boldsymbol{0}\} = h_{n+1}(C \times D), \qquad \mathbb{P}\{C \cap \boldsymbol{Q}L \ne \boldsymbol{0}\} = h_{m+1}(C).$$

Applying the polarity relation $(C \cap D)^\circ = C^\circ + D^\circ$ (see [AL17, Proposition 2.5]) to the kinematic formulas, we obtain a polar version of the kinematic formula, for $k > 0$,

$$\mathbb{E}[v_{n-k}(C + \boldsymbol{Q}D)] = v_{n-k}(C \times D), \qquad \mathbb{E}[v_n(C + \boldsymbol{Q}D)] = t_n(C \times D). \tag{4.10}$$

A convenient consequence of this polar form is a projection formula for intrinsic volumes, due to Glasauer [Gla95]. Let $\boldsymbol{Q} \in O(n)$ uniform at random and $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ a fixed orthogonal projection onto a linear subspace L of dimension m. Then for $0 < k \le m$,

$$\mathbb{E}[v_{m-k}(\boldsymbol{P}\boldsymbol{Q}C)] = v_{m-k}(C), \qquad \mathbb{E}[v_m(\boldsymbol{P}\boldsymbol{Q}C)] = t_m(C). \tag{4.11}$$

As we will see in Section 4.6, this results holds for *any* full rank $\boldsymbol{T} \in \mathbb{R}^{m\times n}$, instead of just for projections $\boldsymbol{P}$.

Remark 4.6. The astute reader may notice that the projection $\boldsymbol{PQ}C$ does not need to be a closed convex cone. For random $\boldsymbol{Q}$, however, the probability of this happening can be shown to be zero.

4.5. **Concentration of measure.** It was shown in [ALMT14] (with a more streamlined and improved derivation in [MT14]), that the intrinsic volumes concentrate sharply around the statistical dimension. For a closed convex cone C, let X_C denote the discrete random variable satisfying

$$\mathbb{P}\{X_C = k\} = v_k(C).$$

The following result is from [MT14].

Theorem 4.7. *Let $\lambda \geq 0$. Then*

$$\mathbb{P}\{|X_C - \delta(C)| \geq \lambda\} \leq 2\exp\left(\frac{-\lambda^2/4}{\min\{\delta(C),\delta(C^\circ)\}+\lambda/3}\right).$$

Roughly speaking, the intrinsic volumes of a convex cone in high dimensions approximate those of a linear subspace of dimension $\delta(C)$. The concentration result 4.7, used in conjunction with the kinematic formula, gives rise to an approximage kinematic formula, which in turn underlies the phase transition results from [ALMT14]. We will only need the following direct consequence of Theorem 4.7.

Corollary 4.8. *Let $\eta \in (0,1)$, let C be a closed convex cone, and let $0 \leq m \leq n$. Then*

$$\delta(C) \leq m - a_\eta\sqrt{m} \implies t_m \leq \eta;$$
$$\delta(C) \geq m + a_\eta\sqrt{m} \implies t_m \geq 1-\eta,$$

with $a_\eta := 2\sqrt{\log(2/\eta)}$.

Applying the above to the statistical dimension, we get the following expression.

Corollary 4.9. *Let $\eta \in (0,1)$ and assume that $m \geq \delta(C) + a_\eta\sqrt{m}$, with $a_\eta = 2\sqrt{\log(2/\eta)}$. Then*

$$\delta(C) - (n-m)\eta \leq \mathbb{E}_{\boldsymbol{Q}}[\delta(\boldsymbol{PQ}C)] \leq \delta(C).$$

Proof. A direct application of the projection formulas (4.11) and the definition of the statistical dimension shows that

$$\mathbb{E}_{\boldsymbol{Q}}[\delta(\boldsymbol{PQ}C)] = \delta(C) - \sum_{k=1}^{n-m} k v_{k+m}(C).$$

The bound then follows by bounding the right-hand side in a straight-forward way and applying Corollary 4.8. □

We conclude this section by proving Theorem B, which we restate for convenience. Recall the notation

$$\overline{\kappa}_m^2(\boldsymbol{A}) := \mathbb{E}_{\boldsymbol{Q}}[\kappa(\boldsymbol{PQA})^2],$$

with $\boldsymbol{P}$ the projection on the first m coordinates.

Theorem 4.10. *Let $C \subseteq \mathbb{R}^n$ be a closed convex cone and $\boldsymbol{A}\colon \mathbb{R}^n \to \mathbb{R}^n$ a non-singular linear map. Let $\eta \in (0,1)$ and assume that $m \geq \delta(C) + 2\sqrt{\log(2/\eta)m}$. Then*

$$\delta(\boldsymbol{A}C) \leq \mathbb{E}_{\boldsymbol{Q}}\left[\mathscr{R}_C(\boldsymbol{PQA})^2\right]\cdot\delta(C) + (n-m)\eta.$$

For the matrix condition number,

$$\delta(\boldsymbol{A}C) \leq \overline{\kappa}_m^2(\boldsymbol{A})\cdot\delta(C) + (n-m)\eta. \tag{4.12}$$

Proof. The upper bound follows from

$$\delta(\boldsymbol{AC}) \le \mathbb{E}_{\boldsymbol{Q}}[\delta(\boldsymbol{PQAC})] + (n-m)\eta \le \mathbb{E}_{\boldsymbol{Q}}\left[\mathscr{R}_C(\boldsymbol{PQA})^2\right]\delta(C) + (n-m)\eta,$$

where we used Theorem A for the second inequality. The upper bound in terms for the matrix condition number follows as in the proof of Theorem A. □

4.6. **The TQC Lemma.** The following generalization of the projection formulas (4.11), first observed by Mike McCoy and Joel Tropp, may at first sight look surprising. While it can be deduced from general integral-geometric considerations (see, for example, [Ame14]), we include a proof because it is illustrative.

Lemma 4.11. *Let $\boldsymbol{T} \in \mathbb{R}^{m\times n}$ be of full rank. Then for $0 \le k < m$,*

$$\mathbb{E}[v_k(\boldsymbol{TQ}C)] = v_k(C), \qquad \mathbb{E}[v_m(\boldsymbol{TQ}C)] = t_m(C) \tag{4.13}$$

Proof. In view of (4.6), it suffices to show (4.13) for the half-tail functionals h_j instead of the intrinsic volumes v_j. Let $L \subset \mathbb{R}^n$ be a linear subspace of dimension $\dim L = k \le m$. From Proposition 2.2 it follows that

$$\boldsymbol{Q}C \cap \boldsymbol{T}^{-1}L \ne \{\boldsymbol{0}\} \iff \boldsymbol{TQ}C \cap L \ne \{\boldsymbol{0}\} \text{ or } \ker \boldsymbol{T} \cap \boldsymbol{Q}C \ne \{\boldsymbol{0}\},$$

where in this case, as before, $\boldsymbol{T}^{-1}L$ denotes the pre-image of L under $\boldsymbol{T}$. Denoting by $\boldsymbol{P}$ the orthogonal projection onto the complement $(\ker \boldsymbol{T})^{\perp}$, we thus get

$$\boldsymbol{PQ}C \cap (\boldsymbol{T}^{-1}L \cap (\ker \boldsymbol{T})^{\perp}) \ne \{\boldsymbol{0}\} \iff \boldsymbol{TQ}C \cap L \ne \{\boldsymbol{0}\},$$

and taking probabilities,

$$\mathbb{P}\left\{\boldsymbol{PQ}C \cap (\boldsymbol{T}^{-1}L \cap (\ker \boldsymbol{T})^{\perp}) \ne \{\boldsymbol{0}\}\right\} = \mathbb{P}\{\boldsymbol{TQ}C \cap L \ne \{\boldsymbol{0}\}\}. \tag{4.14}$$

To compute the probability on the left, let $\boldsymbol{Q}_0$ is a random orthogonal transformation of the space $(\ker \boldsymbol{T})^{\perp}$. Restricting to $(\ker \boldsymbol{T})^{\perp}$ as ambient space,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{Q}}\left\{\boldsymbol{PQ}C \cap (\boldsymbol{T}^{-1}L \cap (\ker \boldsymbol{T})^{\perp}) \ne \{\boldsymbol{0}\}\right\} &= \mathbb{P}_{\boldsymbol{Q}}\left\{\boldsymbol{PQ}C \cap \boldsymbol{Q}_0(\boldsymbol{T}^{-1}L \cap (\ker \boldsymbol{T})^{\perp}) \ne \{\boldsymbol{0}\}\right\}\\
&= \mathbb{E}_{\boldsymbol{Q}_0}\mathbb{P}_{\boldsymbol{Q}}\left\{\boldsymbol{PQ}C \cap \boldsymbol{Q}_0(\boldsymbol{T}^{-1}L \cap (\ker \boldsymbol{T})^{\perp}) \ne \{\boldsymbol{0}\}\right\}\\
&\overset{(1)}{=} \mathbb{E}_{\boldsymbol{Q}}\mathbb{P}_{\boldsymbol{Q}_0}\left\{\boldsymbol{PQ}C \cap \boldsymbol{Q}_0(\boldsymbol{T}^{-1}L \cap (\ker \boldsymbol{T})^{\perp}) \ne \{\boldsymbol{0}\}\right\}\\
&\overset{(2)}{=} \mathbb{E}_{\boldsymbol{Q}}[h_{m-k+1}(\boldsymbol{PQ}C)]
\end{aligned}$$

where for (1) we summoned Fubini on the representation of the probability as expectation of an indicator variable and for (2) the Crofton formula 4.5 with $(\ker \boldsymbol{T})^{\perp}$ as ambient space. A similar argument on the right-hand side of (4.14) shows that

$$\mathbb{P}_{\boldsymbol{Q}}\{\boldsymbol{TQ}C \cap L \ne \{\boldsymbol{0}\}\} = \mathbb{E}_{\boldsymbol{Q}}[h_{m-k+1}(\boldsymbol{TQ}C)].$$

In summary, we have for shown that $\mathbb{E}_{\boldsymbol{Q}}[h_{m-k+1}(\boldsymbol{TQ}C)] = \mathbb{E}_{\boldsymbol{Q}}[h_{m-k+1}(\boldsymbol{PQ}C)]$ for $0 \le k \le m$, and hence also $\mathbb{E}_{\boldsymbol{Q}}[v_i(\boldsymbol{TQ}C)] = \mathbb{E}_{\boldsymbol{Q}}[v_i(\boldsymbol{PQ}C)]$ for $0 \le i \le m$. The claim now follows by applying the projection formula (4.11). □

As with the case where $\boldsymbol{T}$ is a projection, applying the above to the statistical dimension, we get the following expression.

Corollary 4.12. *Let $\eta \in (0,1)$ and assume that $m \ge \delta(C) + a_\eta\sqrt{m}$, with $a_\eta = 2\sqrt{\log(2/\eta)}$. Then under the conditions of Lemma 4.11, we have*

$$\delta(C) - (n-m)\eta \le \mathbb{E}_{\boldsymbol{Q}}[\delta(\boldsymbol{TQ}C)] \le \delta(C) - \eta.$$

It remains to be seen whether the fact that the main preconditionining results can be formulated with an arbitrary matrix $\boldsymbol{T}$, rather than just a projection $\boldsymbol{P}$, can be of use.

5. APPLICATIONS

In this section we apply the results derived for convex cones to the setting of convex regularizers. To give this application some context, we briefly review some of the theory.

5.1. **Convex regularization, subdifferentials and the descent cone.** In practical applications the cones of interest often arise as cones generated by the subgradient of a proper convex function $f\colon \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$.

The exact form of the general convex regularization problem is

$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{5.1}$$

while the noisy form is

$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2 \le \varepsilon. \tag{5.2}$$

Interchanging the role of the function f and the residual, we get the *generalized LASSO*

$$\text{minimize} \quad \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2 \quad \text{subject to} \quad f(\boldsymbol{x}) \le \tau. \tag{5.3}$$

Finally, we have the Lagrangian form,

$$\text{minimize} \quad \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 + \lambda f(\boldsymbol{x}). \tag{5.4}$$

These last three problems are, in fact, equivalent (see [FR13, Chapter 3] for a concise derivation in the case $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$). The practical problem consists in effectively finding the parameters involved.

The first-order optimality condition states that $\hat{\boldsymbol{x}}$ is a unique solution of (5.1) if and only if

$$\exists \boldsymbol{y} \neq \boldsymbol{0}\colon \boldsymbol{A}^T\boldsymbol{y} \in \partial f(\hat{\boldsymbol{x}}), \tag{5.5}$$

where $\partial f(\hat{\boldsymbol{x}})$ denotes the subdifferential of f at $\hat{\boldsymbol{x}}$, i.e., the set

$$\partial f(\hat{\boldsymbol{x}}) = \{\boldsymbol{z} \in \mathbb{R}^n : f(\hat{\boldsymbol{x}}+\boldsymbol{z}) \ge f(\hat{\boldsymbol{x}}) + \langle \boldsymbol{z}, \boldsymbol{x}\rangle\}.$$

If f is differentiable at $\hat{\boldsymbol{x}}$, then of course the subdifferential contains only the gradient of f at $\hat{\boldsymbol{x}}$, and the vector $\boldsymbol{y}$ in (5.5) consists of the Lagrange multipliers.

Example 5.1. If f is a norm, with dual norm f°, then the subdifferential of f at $\hat{\boldsymbol{x}}$ is

$$\partial f(\hat{\boldsymbol{x}}) = \begin{cases} \{\boldsymbol{z} \in \mathbb{R}^n : f^\circ(\boldsymbol{z}) = 1, \langle \boldsymbol{z}, \hat{\boldsymbol{x}}\rangle = f(\hat{\boldsymbol{x}})\} & \hat{\boldsymbol{x}} \neq \boldsymbol{0} \\ \{\boldsymbol{z} \in \mathbb{R}^n : f^\circ(\boldsymbol{z}) \le 1\} & \hat{\boldsymbol{x}} = \boldsymbol{0}. \end{cases}$$

Example 5.2. For the ℓ_1-norm at an s-sparse vector $\hat{\boldsymbol{x}}$,

$$\partial\|\hat{\boldsymbol{x}}\|_1 = \{\boldsymbol{z} \in \mathbb{R}^n : \|\boldsymbol{z}\|_\infty = 1, \langle \boldsymbol{z}, \hat{\boldsymbol{x}}\rangle = \|\hat{\boldsymbol{x}}\|_1\},$$

or more explicitly,

$$\partial\|\hat{\boldsymbol{x}}\|_1 = \{\boldsymbol{z} \in \mathbb{R}^n : z_i = \text{sign}\,(\hat{x}_i) \text{ if } \hat{x}_i \neq 0,\ z_j \in [-1,1] \text{ if } \hat{x}_j = 0\}. \tag{5.6}$$

The descent cone of f at $\hat{\boldsymbol{x}}$ is defined as

$$\mathcal{D}(f, \hat{\boldsymbol{x}}) = \bigcup_{\tau>0} \left\{\boldsymbol{y} \in \mathbb{R}^n : f(\hat{\boldsymbol{x}} + \tau \boldsymbol{y}) \le f(\hat{\boldsymbol{x}})\right\}.$$

The convex cone generated by the subdifferential of f at $\hat{\boldsymbol{x}}$ and is the closure of the polar cone of $\mathcal{D}(f, \hat{\boldsymbol{x}})$,

$$\text{cone}\left(\partial f(\hat{\boldsymbol{x}})\right) = \overline{\mathcal{D}(f, \hat{\boldsymbol{x}})^\circ}, \tag{5.7}$$

Condition (5.5) is therefore equivalent to

$$\ker \boldsymbol{A} \cap \mathcal{D}(f, \hat{\boldsymbol{x}}) = \{\boldsymbol{0}\},$$

namely, that the kernel of $\boldsymbol{A}$ does not intersect the descent cone nontrivially.

For the robust problem (5.2), given a point $\boldsymbol{x}_0$ that satisfies the constraints, we have seen in the introduction that the error satisfies the bound

$$\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\| \leq 2\varepsilon \cdot \sigma_{\mathscr{D}(f,\boldsymbol{x}_0)}(\boldsymbol{A})^{-1},$$

where σ denotes the smallest restricted singular value.

An important class of regularizers are of the form $f(\boldsymbol{x}) := g(\boldsymbol{A}\boldsymbol{x}) + h(\boldsymbol{B}\boldsymbol{x})$, with $\boldsymbol{A}$ and $\boldsymbol{B}$ linear maps. It follows from [Roc70, Theorems 23.8, 23.9] that the subdifferential is

$$\partial f(\boldsymbol{x}) = \boldsymbol{A}^T \partial g(\boldsymbol{A}\boldsymbol{x}) + \boldsymbol{B}^T \partial h(\boldsymbol{B}\boldsymbol{x}).$$

Such composite regularizers include the "cosparse" setting [NDEG13]. For $f(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$ and invertible $\boldsymbol{D}$, combining (5.7) with Lemma 2.1 we get,

$$\mathscr{D}(f,\boldsymbol{x}_0) = \boldsymbol{D}^{-1}\mathscr{D}(g,\boldsymbol{D}\boldsymbol{x}_0). \tag{5.8}$$

Example 5.3. The ℓ_1 norm can be written as

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |\boldsymbol{\Pi}_i(\boldsymbol{x})|,$$

where $\boldsymbol{\Pi}_i(\boldsymbol{x}) = x_i$ is the projection on the i-th component. The subdifferential at $\hat{\boldsymbol{x}}$ is therefore

$$\partial\|\hat{\boldsymbol{x}}\|_1 = \sum_{i=1}^n \boldsymbol{\Pi}_i^T \partial |x_i|.$$

The subdifferential of the absolute value is

$$\partial|x| = \begin{cases} \frac{x}{|x|} & x \neq 0 \\ [-1,1] & x = 0. \end{cases}$$

This leads to the same description of the subdifferential of the ℓ_1 norm as face of a unit hypercube as the one given in (5.6).

Example 5.4 (Finite differences)**.** Let $\boldsymbol{x} \in \mathbb{R}^n$ and let

$$\boldsymbol{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \tag{5.9}$$

be the discrete finite difference matrix. Thus

$$\boldsymbol{D}\boldsymbol{x} = (x_2 - x_1, x_3 - x_2, \dots, x_d - x_{d-1})^T.$$

Define $g(\boldsymbol{x}) := f(\boldsymbol{D}\boldsymbol{x})$. Then for a fixed $\hat{\boldsymbol{x}}$, the subdifferential is given by

$$\partial g(\hat{\boldsymbol{x}}) = \boldsymbol{D}^T \partial f(\boldsymbol{D}\hat{\boldsymbol{x}}).$$

In the special case where f is the ℓ_1-norm and $\boldsymbol{D}\hat{\boldsymbol{x}}$ is s-sparse with support $I \subset [n]$,

$$\partial g(\hat{\boldsymbol{x}}) = \{\boldsymbol{D}^T\boldsymbol{z} : \|\boldsymbol{z}\|_\infty = 1, \langle \boldsymbol{z}, \boldsymbol{D}\hat{\boldsymbol{x}}\rangle = \|\boldsymbol{D}\hat{\boldsymbol{x}}\|_1\}.$$

One can think of such a vector $\hat{\boldsymbol{x}}$ as a signal with sparse gradient.

Example 5.5. (Weighted ℓ_1 norm). Let $\boldsymbol{\omega} \in \mathbb{R}^n$ be a vector of weights and define the weighted ℓ_1-norm

$$\|\boldsymbol{x}\|_{\boldsymbol{\omega},1} = \sum_{j=1}^n \omega_j |x_j|.$$

By extension from the ℓ_1 example, we have

$$\begin{aligned}\partial\|\hat{\boldsymbol{x}}\|_{\boldsymbol{\omega},1} &= \{\boldsymbol{z}\in\mathbb{R}^n : z_i=\omega_i \text{ sign } (\hat{x}_i) \text{ if } \hat{x}_i\neq 0,\ z_j\in[-\omega_j,\omega_j] \text{ if } \hat{x}_j=0\}\\ &= \operatorname{diag}(\boldsymbol{\omega})\ \partial\|\hat{\boldsymbol{x}}\|_1.\end{aligned}$$

This example becomes interesting when considering *weighted* s-sparse vectors, that is, vectors such that

$$\|\boldsymbol{x}\|_{\boldsymbol{\omega},0}=\sum_{x_j\neq 0}\omega_j^2=s.$$

The use of composite regularizers to recover simultaneously structured models was studied in [OJF+15], where it was found that the performance is not better than when using a regularizers adapted to a single structure.

5.2. **Performance bounds in convex regularization.** As mentioned in the introduction, computing the statistical dimension of convex regularizers is in general a difficult problem, with only few cases allowing for closed-form expressions. Using the condition bounds for the statistical dimension of linear images of convex cones, and translating these to the setting of convex regularizers, we get the corresponding statements in Corollary 1.1, which we restate here.

Corollary 5.6. *Let $f(\boldsymbol{x})=g(\boldsymbol{D}\boldsymbol{x})$, where g is a proper convex function and let $\boldsymbol{D}\in\mathbb{R}^{n\times n}$ be non-singular. Then*

$$\delta(f,\boldsymbol{x}_0)\leq \mathscr{R}_{\mathscr{D}(g,\boldsymbol{D}\boldsymbol{x}_0)}\left(\boldsymbol{D}^{-1}\right)\cdot\delta(g,\boldsymbol{D}\boldsymbol{x}_0).$$

In particular,

$$\frac{\delta(g,\boldsymbol{D}\boldsymbol{x}_0)}{\kappa(\boldsymbol{D})^2}\leq\delta(f,\boldsymbol{x}_0)\leq\kappa(\boldsymbol{D})^2\cdot\delta(g,\boldsymbol{D}\boldsymbol{x}_0).$$

Proof. Let $C=\mathscr{D}(g,\boldsymbol{D}\boldsymbol{x}_0)$. Then from (5.8) we get that

$$\delta(f,\boldsymbol{x}_0)=\delta(\boldsymbol{D}^{-1}C).$$

The claims then follows from Theorem A and Proposition 1.2, noting that $\kappa(\boldsymbol{D}^{-1})=\kappa(\boldsymbol{D})$. □

In a similar fashion, we also get the preconditioned bounds in Corollary 1.3.

A popular method [ALMT14, Recipe 4.1], going back to Stojnic [Sto09] and generalized in [CRPW12], is to approximate the statistical dimension of the descent cone $\mathscr{D}(f,\boldsymbol{x}_0)$ by the expected value

$$\inf_{\tau\geq 0}\mathbb{E}[\operatorname{dist}^2(\boldsymbol{g},\tau\cdot\partial f(\boldsymbol{x}))]. \tag{5.10}$$

This approximation, however, does not work for all regularizers f for two reasons: it my not be tight, and computing the quantity may not be feasible. In [ALMT14, Theorem 4.1], the following error bound is derived.

$$0\leq\inf_{\tau\geq 0}\mathbb{E}[\operatorname{dist}^2(\boldsymbol{g},\tau\cdot\partial f(\boldsymbol{x}))]-\delta(f,\boldsymbol{x}_0)\leq\frac{2\sup\{\|\boldsymbol{s}\|:\ \boldsymbol{s}\in\partial f(\boldsymbol{x})\}}{f(\boldsymbol{x}/\|\boldsymbol{x}\|)}. \tag{5.11}$$

In [ZXCL16], this error was analysed and it was shown to be bounded, so that the approximation is asymptotically tight. Using a different route, in [DH17], it was shown that in the case where $f(\boldsymbol{x})=\|\boldsymbol{D}\boldsymbol{x}\|_1$ one has

$$0\leq\inf_{\tau\geq 0}\mathbb{E}[\operatorname{dist}^2(\boldsymbol{g},\tau\cdot\partial f(\boldsymbol{x}))]-\delta(f,\boldsymbol{x}_0)\leq\frac{2\kappa(\boldsymbol{D})}{\sqrt{s(n-1)}}.$$

While this bound is not sharp (the derivation makes use of norm inequalities), it is enlightening as it gives sufficient conditions for the applicability of Bound (5.11) in terms of the condition number of $\boldsymbol{A}$. It remains to be seen whether randomized preconditioning can be incorporated into this bound, and therefore whether this approach can lead to bounds that would rival those derived in [ZXCL16].

References

[AB13] D. Amelunxen and P. Bürgisser. Probabilistic analysis of the Grassmann condition number. *Found. Comput. Math.*, 2013.

[AL14] Dennis Amelunxen and Martin Lotz. Gordon's inequality and condition numbers in conic optimization. *arXiv preprint arXiv:1408.3016*, 2014.

[AL17] Dennis Amelunxen and Martin Lotz. Intrinsic volumes of polyhedral cones: a combinatorial perspective. *Discrete and Computational Geometry (to appear)*, 2017.

[ALMT14] Dennis Amelunxen, Martin Lotz, Michael B. McCoy, and Joel A. Tropp. Living on the edge: phase transitions in convex programs with random data. *Information and Inference*, 2014.

[Ame11] D. Amelunxen. Geometric analysis of the condition of the convex feasibility problem. PhD Thesis, Univ. Paderborn, 2011.

[Ame14] Dennis Amelunxen. Measures on polyhedral cones: characterizations and kinematic formulas. *arXiv preprint arXiv:1412.1569*, 2014.

[Bar02] A. Barvinok. *A course in convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

[BC13] P. Bürgisser and F. Cucker. *Condition: The geometry of numerical algorithms*. Number 349 in Grundlehren der Mathematischen Wissenschaften. Springer Verlag, 2013.

[BF09] A. Belloni and R. M. Freund. A geometric analysis of Renegar's condition number, and its interplay with conic curvature. *Math. Program.*, 119(1, Ser. A):95–107, 2009.

[CRPW12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[DH17] Sajad Daei and Farzan Haddadi. How to exploit prior information in low-complexity models. *CoRR*, abs/1704.05397, 2017.

[DJM13] D. L. Donoho, I. Johnstone, and A. Montanari. Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. *IEEE Trans. Inform. Theory*, 59(6):3396–3433, June 2013.

[DS01] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and banach spaces. *Handbook of the geometry of Banach spaces*, 1:317–366, 2001.

[Eft17] A. Eftekhari. Private communication, 2017.

[FR13] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*, volume 336 of *Applied and Numerical Harmonic Analysis*. Birkhäuser, Basel, 2013.

[Gla95] S. Glasauer. Integralgeometrie konvexer Körper im sphärischen Raum. Thesis, Univ. Freiburg i. Br., 1995.

[Gor85] Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50(4):265–289, 1985.

[Gor87] Y. Gordon. Elliptically contoured distributions. *Probability theory and related fields*, 76(4):429–438, 1987.

[KRZ15] Maryia Kabanava, Holger Rauhut, and Hui Zhang. Robust analysis âĎŞ 1-recovery from gaussian measurements and total variation minimization. *European Journal of Applied Mathematics*, 26(06):917–929, 2015.

[KW12] N. Krislock and H. Wolkowicz. Euclidean distance matrices and applications. In *Handbook on semidefinite, conic and polynomial optimization*, volume 166 of *Internat. Ser. Oper. Res. Management Sci.*, pages 879–914. Springer, New York, 2012.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. A Series of Modern Surveys in Mathematics Series. Springer-Verlag, Berlin, 1991.

[Mau11] A. Maurer. A proof of Slepian's inequality. www.andreas-maurer.eu/Slepian3.pdf, 2011.

[MT13a] M. B. McCoy and J. A. Tropp. The achievable performance of convex demixing. arXiv:1309.7478v1 [cs.IT], 2013.

[MT13b] M. B. McCoy and J. A. Tropp. From Steiner formulas for cones to concentration of intrinsic volumes. arXiv:1308.5265v1 [math.MG], 2013.

[MT14] Michael B McCoy and Joel A Tropp. From steiner formulas for cones to concentration of intrinsic volumes. *Discrete & Computational Geometry*, 51(4):926–963, 2014.

[NDEG13] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Appl. Comput. Harmon. Anal.*, 34(1):30–56, 2013.

[OH16] Samet Oymak and Babak Hassibi. Sharp mse bounds for proximal denoising. *Foundations of Computational Mathematics*, 16(4):965–1029, 2016.

[OJF+15] Samet Oymak, Amin Jalali, Maryam Fazel, Yonina C Eldar, and Babak Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Transactions on Information Theory*, 61(5):2886–2908, 2015.

[ORS15] Samet Oymak, Benjamin Recht, and Mahdi Soltanolkotabi. Isometric sketching of any set via the restricted isometry property. *arXiv preprint arXiv:1506.03521*, 2015.

[OT15] Samet Oymak and Joel A Tropp. Universality laws for randomized dimension reduction, with applications. *arXiv preprint arXiv:1511.09433*, 2015.
[OTH13] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi. The squared-error of generalized lasso: A precise analysis. In *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pages 1002–1009. IEEE, 2013.
[RBd15] Vincent Roulet, Nicolas Boumal, and Alexandre d'Aspremont. Renegar's condition number and compressed sensing performance. *arXiv preprint arXiv:1506.03295*, 2015.
[Ren94] J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994.
[Ren95] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995.
[Roc70] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
[RV+13] Mark Rudelson, Roman Vershynin, et al. Hanson-wright inequality and sub-gaussian concentration. *Electron. Commun. Probab*, 18(82):1–9, 2013.
[Sto09] Mihailo Stojnic. Various thresholds for ℓ_1-optimization in compressed sensing. *preprint*, 2009. arXiv:0907.3666.
[SW08a] R. Schneider and W. Weil. *Stochastic and integral geometry*. Probability and its Applications (New York). Springer-Verlag, Berlin, 2008.
[SW08b] Rolf Schneider and Wolfgang Weil. *Stochastic and Integral Geometry*. Springer series in statistics: Probability and its applications. Springer, 2008.
[VRPH07] J. C. Vera, J. C. Rivera, J. Peña, and Yao Hui. A primal-dual symmetric relaxation for homogeneous conic systems. *J. Complexity*, 23(2):245–261, 2007.
[ZXCL16] Bingwen Zhang, Weiyu Xu, Jian-Feng Cai, and Lifeng Lai. Precise phase transition of total variation minimization. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 4518–4522. IEEE, 2016.

Appendix A. The biconic feasibility problem - proofs

In this appendix we provide the proofs for Section 2.2. Recall that for $C \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ closed convex cones, the biconic feasibility problem is given by

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

and the sets of primal feasible and dual feasible instances can be characterized by

$$\mathscr{P}(C,D) = \big\{\boldsymbol{A} \in \mathbb{R}^{m\times n} : C \cap \big(\boldsymbol{A}^T D\big)^\circ \neq \{\mathbf{0}\}\big\} = \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \sigma_{C\to D}(\boldsymbol{A}) = 0\},$$
$$\mathscr{D}(C,D) = \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : D \cap (-\boldsymbol{A}C)^\circ \neq \{\mathbf{0}\}\} = \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \sigma_{D\to C}(-\boldsymbol{A}^T) = 0\},$$

respectively, cf. (2.3)/(2.4). The proof of Proposition 2.4 uses the following generalization of Farkas' Lemma.

Lemma A.1. *Let $C, \tilde{C} \subseteq \mathbb{R}^n$ be closed convex cones with* $\operatorname{int}(C) \neq \emptyset$. *Then*

$$\operatorname{int}(C) \cap \tilde{C} = \emptyset \iff C^\circ \cap (-\tilde{C}^\circ) \neq \{\mathbf{0}\}. \tag{A.1}$$

Proof. If $\operatorname{int}(C) \cap \tilde{C} = \emptyset$, then there exists a separating hyperplane $H = \boldsymbol{v}^\perp$, $\boldsymbol{v} \neq \mathbf{0}$, so that $\langle \boldsymbol{v}, \boldsymbol{x}\rangle \leq 0$ for all $\boldsymbol{x} \in C$ and $\langle \boldsymbol{v}, \boldsymbol{y}\rangle \geq 0$ for all $\boldsymbol{y} \in \tilde{C}$. But this means $\boldsymbol{v} \in C^\circ \cap (-\tilde{C}^\circ)$. On the other hand, if $\boldsymbol{x} \in \operatorname{int}(C) \cap \tilde{C}$ then only in the case $C = \mathbb{R}^n$, for which the claim is trivial, can $\boldsymbol{x} = \mathbf{0}$. If $\boldsymbol{x} \neq \mathbf{0}$, then $C^\circ \setminus \{\mathbf{0}\}$ lies in the open half-space $\{\boldsymbol{v} : \langle \boldsymbol{v}, \boldsymbol{x}\rangle < 0\}$ and $-\tilde{C}^\circ$ lies in the closed half-space $\{\boldsymbol{v} : \langle \boldsymbol{v}, \boldsymbol{x}\rangle \geq 0\}$, and thus $C^\circ \cap (-\tilde{C}^\circ) = \{\mathbf{0}\}$. □

For the proof of the third claim in Proposition 2.4 we also need the following well-known convex geometric lemma; a proof can be found, for example, in [SW08a, proof of Thm. 6.5.6]. We say that two cones $C, D \subseteq \mathbb{R}^n$, with $\operatorname{int}(C) \neq \emptyset$, *touch* if $C \cap D \neq \{\mathbf{0}\}$ but $\operatorname{int}(C) \cap D = \emptyset$.

Lemma A.2. *Let $C, D \subseteq \mathbb{R}^n$ closed convex cones with* $\operatorname{int}(C) \neq \emptyset$. *If $\boldsymbol{Q} \in O(n)$ uniformly at random, then the randomly rotated cone $\boldsymbol{Q}D$ almost surely does not touch C.*

Proof of Proposition 2.4. (1) The sets $\mathscr{P}(C,D)$ and $\mathscr{D}(C,D)$ are closed as they are preimages of the closed set $\{0\}$ under continuous functions, c.f. (2.3)/(2.4). Indeed, for any $\boldsymbol{x}$, the function $\boldsymbol{A} \mapsto \|\boldsymbol{\Pi}_D(\boldsymbol{Ax})\|$ is continuous, and as a minimum of such functions over the compact set $C \cap S^{m-1}$, it follows that $\sigma_{C\to D}(\boldsymbol{A})$ is continuous. Hence, $\mathscr{P}(C,D) = \{\boldsymbol{A} \in \mathbb{R}^{n\times m} : \sigma_{C\to D}(\boldsymbol{A}) = 0\}$ is closed. The same argument applies to $\mathscr{D}(C,D)$.

(2) For the claim about the union of the sets $\mathscr{P}(C,D)$ and $\mathscr{D}(C,D)$ we first consider the case $C \neq \mathbb{R}^n$, so that $\boldsymbol{0} \notin \operatorname{int}(C)$. Using the generalized Farkas' Lemma A.1, we obtain

$$\boldsymbol{A} \notin \mathscr{P}(C,D) \iff C \cap \left(\boldsymbol{A}^T D\right)^\circ = \{\boldsymbol{0}\} \Rightarrow \operatorname{int}(C) \cap \left(\boldsymbol{A}^T D\right)^\circ = \emptyset \overset{(A.1)}{\Longrightarrow} C^\circ \cap (-\boldsymbol{A}^T D) \neq \{\boldsymbol{0}\} \Rightarrow \boldsymbol{A} \in \mathscr{D}(C,D).$$

This shows $\mathscr{P}(C,D) \cup \mathscr{D}(C,D) = \mathbb{R}^{n\times m}$. For $D \neq \mathbb{R}^n$ the argument is the same. For $C = \mathbb{R}^n$ and $D = \mathbb{R}^m$:

$$\mathscr{P}(\mathbb{R}^n,\mathbb{R}^m) = \left\{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \ker \boldsymbol{A} \neq \{\boldsymbol{0}\}\right\} = \begin{cases} \{\text{rank deficient matrices}\} & \text{if } n \le m \\ \mathbb{R}^{m\times n} & \text{if } n > m, \end{cases}$$

$$\mathscr{D}(\mathbb{R}^n,\mathbb{R}^m) = \left\{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \ker \boldsymbol{A}^T \neq \{\boldsymbol{0}\}\right\} = \begin{cases} \mathbb{R}^{m\times n} & \text{if } n < m \\ \{\text{rank deficient matrices}\} & \text{if } n \ge m. \end{cases}$$

In particular,this shows $\mathscr{P}(\mathbb{R}^n,\mathbb{R}^n) \cup \mathscr{D}(\mathbb{R}^n,\mathbb{R}^n) = \{\text{rank deficient matrices}\}$.

(3) If $(C,D) = (\mathbb{R}^n,\mathbb{R}^m)$ then by the characterization above $\Sigma(\mathbb{R}^n,\mathbb{R}^m)$ consists of the rank deficient matrices, which is a nonempty set. If $(C,D) \neq (\mathbb{R}^n,\mathbb{R}^n)$, then the union of the closed sets $\mathscr{P}(C,D)$ and $\mathscr{D}(C,D)$ equals $\mathbb{R}^{m\times n}$, which is an irreducible topological space, so that their intersection $\Sigma(C,D) = \mathscr{P}(C,D) \cap \mathscr{D}(C,D)$ must be nonempty.

As for the claim about the Lebesgue measure of $\Sigma(C,D)$, we may use the symmetry between (P) and (D) to assume without loss of generality $m \le n$. If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has full rank, then $\boldsymbol{A}C$ has nonempty interior and from Proposition 2.2 and Farkas' Lemma,

$$\sigma_{C\to D}(\boldsymbol{A}) = 0 \iff C \cap (\boldsymbol{A}^T D)^\circ \neq \{\boldsymbol{0}\} \iff \boldsymbol{A}C \cap D^\circ \neq \{\boldsymbol{0}\} \text{ or } \ker \boldsymbol{A} \cap C \neq \{\boldsymbol{0}\},$$

$$\sigma_{D\to C}(-\boldsymbol{A}^T) = 0 \iff D \cap (-\boldsymbol{A}C)^\circ \neq \{\boldsymbol{0}\} \overset{(A.1)}{\iff} D^\circ \cap \operatorname{int}(\boldsymbol{A}C) = \emptyset.$$

Note that if $\boldsymbol{Ax} = \boldsymbol{0}$ for some $\boldsymbol{x} \in \operatorname{int}(C)$, then $\boldsymbol{A}$, being a continuous surjection, maps an open neighborhood of $\boldsymbol{x}$ to an open neighborhood of the origin, so that $\boldsymbol{A}C = \mathbb{R}^m$. Hence, $D \cap (-\boldsymbol{A}C)^\circ \neq \{\boldsymbol{0}\}$ implies $\ker \boldsymbol{A} \cap \operatorname{int}(C) = \emptyset$, since otherwise $\boldsymbol{A}C = \mathbb{R}^m$, i.e., $(\boldsymbol{A}C)^\circ = \{\boldsymbol{0}\}$.

If $\boldsymbol{A} \in \Sigma(C,D)$, i.e., $\sigma_{C\to D}(\boldsymbol{A}) = \sigma_{D\to C}(-\boldsymbol{A}^T) = 0$, and if $\boldsymbol{A}$ has full rank, then $\boldsymbol{A}C \cap D^\circ \neq \{\boldsymbol{0}\}$ implies that D° touches $\boldsymbol{A}C$, while $\ker \boldsymbol{A} \cap C \neq \{\boldsymbol{0}\}$ implies that $\ker \boldsymbol{A}$ touches C. Hence, if $\boldsymbol{A} = \boldsymbol{G}$ Gaussian, then $\boldsymbol{G}$ has almost surely full rank, and Lemma A.2 implies that both touching events have zero probability, so that almost surely $\boldsymbol{G} \notin \Sigma(C,D)$. □

We next provide the proof for the characterization of the restricted singular values as distances to the primal and dual feasible sets. From now on we use again the short-hand notation $\mathscr{P} := \mathscr{P}(C,D)$ and $\mathscr{D} := \mathscr{D}(C,D)$.

Proof of Proposition 2.5. By symmetry, it suffices to show that $\operatorname{dist}(\boldsymbol{A},\mathscr{P}) = \sigma_{C\to D}(\boldsymbol{A})$. If $\boldsymbol{A} \in \mathscr{P}$ then $\operatorname{dist}(\boldsymbol{A},\mathscr{P}) = 0 = \sigma_{C\to D}(\boldsymbol{A})$, so assume that $\boldsymbol{A} \notin \mathscr{P}$. Let $\boldsymbol{\Delta A} \in \mathbb{R}^{m\times n}$ such that $\boldsymbol{A} + \boldsymbol{\Delta A} \in \mathscr{P}$ and $\operatorname{dist}(\boldsymbol{A},\mathscr{P}) = \|\boldsymbol{\Delta A}\|$. Since $\boldsymbol{A} + \boldsymbol{\Delta A} \in \mathscr{P}$, there exists $\boldsymbol{x}_0 \in C \cap S^{n-1}$ such that $\boldsymbol{w}_0 := (\boldsymbol{A} + \boldsymbol{\Delta A})\boldsymbol{x}_0 \in D^\circ$. For all $\boldsymbol{y} \in D$

$$0 \ge \langle \boldsymbol{w}_0, \boldsymbol{y}\rangle = \langle (\boldsymbol{A} + \boldsymbol{\Delta A})\boldsymbol{x}_0, \boldsymbol{y}\rangle = \langle \boldsymbol{Ax}_0, \boldsymbol{y}\rangle - \langle -\boldsymbol{\Delta A x}_0, \boldsymbol{y}\rangle.$$

If $\boldsymbol{y}_0 \in B^m \cap D$ is such that $\|\Pi_D(\boldsymbol{Ax}_0)\| = \langle \boldsymbol{Ax}_0, \boldsymbol{y}_0\rangle$, then

$$\begin{aligned}\operatorname{dist}(\boldsymbol{A},\mathscr{P}) = \|\boldsymbol{\Delta A}\| \ge \|\boldsymbol{\Delta A x}_0\| \ge \|\Pi_D(-\boldsymbol{\Delta A x}_0)\| &= \max_{\boldsymbol{y} \in B^m \cap D} \langle -\boldsymbol{\Delta A x}_0, \boldsymbol{y}\rangle \\ &\ge \langle -\boldsymbol{\Delta A x}_0, \boldsymbol{y}_0\rangle \ge \langle \boldsymbol{Ax}_0, \boldsymbol{y}_0\rangle = \|\Pi_D(\boldsymbol{Ax}_0)\| \ge \min_{\boldsymbol{x} \in C \cap S^{n-1}} \|\Pi_D(\boldsymbol{Ax})\| = \sigma_{C\to D}(\boldsymbol{A}).\end{aligned}$$

For the reverse inequality $\operatorname{dist}(\boldsymbol{A},\mathscr{P}) \le \sigma_{C\to D}(\boldsymbol{A})$ we need to construct a perturbation $\boldsymbol{\Delta A}$ such that $\boldsymbol{A}+\boldsymbol{\Delta A} \in \mathscr{P}$ and $\|\boldsymbol{\Delta A}\| \le \sigma_{C\to D}(\boldsymbol{A})$. Let $\boldsymbol{x}_0 \in C \cap S^{n-1}$ and $\boldsymbol{y}_0 \in D \cap B^m$ such that

$$\sigma_{C\to D}(\boldsymbol{A}) = \min_{\boldsymbol{x}\in C\cap S^{n-1}} \max_{\boldsymbol{y}\in D\cap B^m} \langle \boldsymbol{A}\boldsymbol{x},\boldsymbol{y}\rangle = \langle \boldsymbol{A}\boldsymbol{x}_0,\boldsymbol{y}_0\rangle.$$

Since $\boldsymbol{A}\notin\mathscr{P}$ we have $\sigma_{C\to D}(\boldsymbol{A}) > 0$, which implies $\|\boldsymbol{y}_0\| = 1$, i.e., $\boldsymbol{y}_0 \in D\cap S^{m-1}$. We define

$$\boldsymbol{\Delta A} := -\boldsymbol{y}_0\boldsymbol{y}_0^T\boldsymbol{A}.$$

Note that

$$\|\boldsymbol{\Delta A}\| = \|\boldsymbol{A}^T\boldsymbol{y}_0\| \le \langle \boldsymbol{A}^T\boldsymbol{y}_0,\boldsymbol{x}_0\rangle = \sigma_{C\to D}(\boldsymbol{A}).$$

Furthermore,

$$(\boldsymbol{A}+\boldsymbol{\Delta A})\boldsymbol{x}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{y}_0\boldsymbol{y}_0^T\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \langle \boldsymbol{A}\boldsymbol{x}_0,\boldsymbol{y}_0\rangle\boldsymbol{y}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \Pi_D(\boldsymbol{A}\boldsymbol{x}_0) = \Pi_{D^\circ}(\boldsymbol{A}\boldsymbol{x}_0).$$

So $\boldsymbol{x}_0 \in C\setminus\{\boldsymbol{0}\}$ and $(\boldsymbol{A}+\boldsymbol{\Delta A})\boldsymbol{x}_0 \in D^\circ$, which shows that $\boldsymbol{A}+\boldsymbol{\Delta A}\in\mathscr{P}$, and hence $\operatorname{dist}(\boldsymbol{A},\mathscr{P}) \le \|\boldsymbol{\Delta A}\| \le \sigma_{C\to D}(\boldsymbol{A})$. □

APPENDIX B. A NEW VARIANT OF GORDON'S COMPARISON THEOREM

Underlying some of our analysis is a new variant of Slepian' inequality, which is a special case of Gordon's comparison theorem. For completeness we first recall the familiar version of Gordon's inequality [Gor85], see also [Gor87] and [FR13, Chapter 8] for a simplified derivation.

Theorem B.1 (Gordon). *Let X_{ij}, Y_{ij}, $1\le i\le m$, $1\le j\le n$, be centered Gaussian random variables, and assume that*

$$\begin{aligned}
\mathbb{E}|X_{ij}-X_{k\ell}|^2 &\le \mathbb{E}|Y_{ij}-Y_{k\ell}|^2, && \text{for all } i\neq k \text{ and } j,\ell,\\
\mathbb{E}|X_{ij}-X_{i\ell}|^2 &\ge \mathbb{E}|Y_{ij}-Y_{i\ell}|^2, && \text{for all } i,j,\ell.
\end{aligned}$$

Then $\mathbb{E}\min_i\max_j X_{ij} \ge \mathbb{E}\min_i\max_j Y_{ij}$. If additionally

$$\mathbb{E}X_{ij}^2 = \mathbb{E}Y_{ij}^2, \qquad \text{for all } i,j,$$

then for any monotonically increasing function $f:\mathbb{R}\to\mathbb{R}$,

$$\mathbb{E}\min_i\max_j f(X_{ij}) \ge \mathbb{E}\min_i\max_j f(Y_{ij}).$$

Slepian's lemma is obtained by setting $m=1$ in Gordon's theorem. The following theorem (in the degenerate case of $X_0 = Y_0 = 0$) lies somewhere in the middle between the two cases treated by Gordon's theorem.

Theorem B.2. *Let X_0, Y_0, X_{ij}, Y_{ij}, $1\le i\le m$, $1\le j\le n$, be centered Gaussian random variables, and assume that*

$$\begin{aligned}
\mathbb{E}|X_{ij}-X_{k\ell}|^2 &\le \mathbb{E}|Y_{ij}-Y_{k\ell}|^2, && \text{for all } i\neq k \text{ and } j,\ell,\\
\mathbb{E}|X_{ij}-X_{i\ell}|^2 &\ge \mathbb{E}|Y_{ij}-Y_{i\ell}|^2, && \text{for all } i,j,\ell,\\
\mathbb{E}|X_{ij}-X_0|^2 &\ge \mathbb{E}|Y_{ij}-Y_0|^2, && \text{for all } i,j.
\end{aligned}$$

Then for any monotonically increasing convex function $f:\mathbb{R}_+\to\mathbb{R}$,

$$\mathbb{E}\min_i\max_j f_+(X_{ij}-X_0) \ge \mathbb{E}\min_i\max_j f_+(Y_{ij}-Y_0), \tag{B.1}$$

where $f_+(x) := f(x)$, if $x\ge 0$, and $f_+(x) := f(0)$, if $x\le 0$.

In Section 3.4 we provide an example, which shows that (B.1) may fail if f is not convex. The proof we present is based on a geometric reduction from Maurer [Mau11], cf. Lemma B.5.

In the following we fix a monotonically increasing convex function $f\colon \mathbb{R}_+ \to \mathbb{R}$, which is differentiable on $(0,\infty)$ and satisfies $\lim_{x\to 0+} f'(x) = 0$. The extension $f_+\colon \mathbb{R}\to\mathbb{R}$, with $f_+(x) := f(x)$, if $x \geq 0$, and $f_+(x) := f(0)$, if $x \leq 0$, is thus monotonically increasing, convex, and differentiable on $\mathbb{R}$. On the Euclidean space $\mathbb{R}\times\mathbb{R}^{m\times n}$, whose elements we denote by $\boldsymbol{x} = (x_0, x_{11}, \dots, x_{mn})$, we define $F\colon \mathbb{R}\times\mathbb{R}^{m\times n} \to \mathbb{R}$ by

$$F(\boldsymbol{x}) := \min_i \max_j f_+(x_{ij} - x_0). \tag{B.2}$$

This function is differentiable almost everywhere. More precisely, it is differentiable if

$$\min_i \max_j x_{ij} < x_0 \quad \text{or} \quad \left|\left\{(k,\ell) : x_{k\ell} = \min_i \max_j \max\{x_{ij}, x_0\} > x_0\right\}\right| = 1.$$

In the first case $\nabla F(\boldsymbol{x}) = 0$. In the second case $\nabla F(\boldsymbol{x})$ is zero except for the (k,ℓ)th entry, $x_{k\ell} = \min_i \max_j \max\{x_{ij}, x_0\}$ $(> x_0)$, which is given by $f'(x_{k\ell} - x_0)$. So, if $\boldsymbol{x}(t)$ is a differentiable curve through $\boldsymbol{x}$ with $\boldsymbol{x}(0) = \boldsymbol{x}$, $\dot{\boldsymbol{x}} := \dot{\boldsymbol{x}}(0)$, then

$$\tfrac{d}{dt} F(\boldsymbol{x}(t))|_{t=0} = \langle \nabla F(\boldsymbol{x}), \dot{\boldsymbol{x}}\rangle = \dot{x}_{k\ell} f'(x_{k\ell} - x_0). \tag{B.3}$$

Lemma B.3. *Let X_0 and X_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, be centered Gaussian random variables such that their joint covariance matrix has full rank. Fix $1 \leq k_0, k \leq m$ and $1 \leq \ell_0, \ell \leq n$ with $(k_0,\ell_0) \neq (k,\ell)$, and let Y, Z be Gaussians, defined in one of the two following ways:*

(1) $X_{k_0\ell_0} = Y + Z$ with Z independent of Y, X_0, X_{ij}, for all $(i,j) \neq (k_0,\ell_0)$,
(2) $X_0 = Y + Z$ with Z independent of Y, X_{ij}, for all (i,j).

If $\boldsymbol{X}(t)$ is defined by

$$X_0(t) := X_0, \qquad X_{ij}(t) := X_{ij}, \quad \text{for } (i,j) \neq (k,\ell), \qquad X_{k\ell}(t) := X_{k\ell} + tZ, \tag{B.4}$$

then

$$\tfrac{d}{dt}\mathbb{E}\big[F(\boldsymbol{X}(t))\big]|_{t=0} \begin{cases} \leq 0 & \text{if } Y,Z \text{ defined as in (1) and } k = k_0, \text{ or } Y,Z \text{ defined as in (2)} \\ \geq 0 & \text{if } Y,Z \text{ defined as in (1) and } k \neq k_0. \end{cases}$$

Proof. We distinguish between the cases (1) and (2).

(1) Let $X_{k_0\ell_0} = Y + Z$ as described above. We define $\boldsymbol{X}^+(t), \boldsymbol{X}^-(t)$ by

$$X_0^+(t) := X_0^-(t) := X_0, \qquad X_{ij}^+(t) := X_{ij}^-(t) := X_{ij}, \quad \text{if } (i,j) \notin \{(k,\ell),(k_0,\ell_0)\},$$

$$X_{k_0\ell_0}^+(t) := Y + |Z|, \qquad X_{k_0\ell_0}^-(t) := Y - |Z|, \qquad X_{k\ell}^+(t) := X_{k\ell} + t|Z|, \qquad X_{k\ell}^-(t) := X_{k\ell} - t|Z|,$$

and denote $X_{ij}^+ := X_{ij}^+(0)$ and $X_{ij}^- := X_{ij}^-(0)$. Since Z is independent of Y, X_0, X_{ij}, for $(i,j) \neq (k_0,\ell_0)$, we have

$$\mathbb{E}\big[F(\boldsymbol{X}(t))\big] = \tfrac{1}{2}\mathbb{E}\big[F(\boldsymbol{X}^+(t)) + F(\boldsymbol{X}^-(t))\big].$$

To simplify the notation, we set

$$\dot{\boldsymbol{X}}^+ = \tfrac{d}{dt}\boldsymbol{X}^+(t)|_{t=0}, \qquad \dot{\boldsymbol{X}}^- = \tfrac{d}{dt}\boldsymbol{X}^-(t)|_{t=0}.$$

A standard argument involving Lebesgue's dominated convergence theorem shows that

$$\tfrac{d}{dt}\mathbb{E}\big[F(\boldsymbol{X}^+(t))\big] = \mathbb{E}\big[\tfrac{d}{dt}F(\boldsymbol{X}^+(t))\big], \qquad \tfrac{d}{dt}\mathbb{E}\big[F(\boldsymbol{X}^-(t))\big] = \mathbb{E}\big[\tfrac{d}{dt}F(\boldsymbol{X}^-(t))\big],$$

so that it is enough to show that almost surely

$$\tfrac{d}{dt}F(\boldsymbol{X}^+(t)) + \tfrac{d}{dt}F(\boldsymbol{X}^-(t)) = \langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle + \langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle \begin{cases} \leq 0 & \text{if } k = k_0 \\ \geq 0 & \text{if } k \neq k_0. \end{cases} \tag{B.5}$$

Note that $X^+_{k\ell} = X_{k\ell}$ and $X^+_0 = X_0$. By (B.3), almost surely

$$\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle = \begin{cases} \dot{X}^+_{k\ell} f'(X_{k\ell} - X_0) & \text{if } X_{k\ell} = \min_i \max_j \max\{X^+_{ij}, X_0\} > X_0 \text{ and} \\ & \quad X^+_{k'\ell'} \neq \min_i \max_j \max\{X^+_{ij}, X_0\} \text{ for all } (k',\ell') \neq (k,\ell) \\ 0 & \text{else (almost surely),} \end{cases}$$

and similarly for $\langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle$. Note that if $X_{k\ell} = \min_i \max_j \max\{X^+_{ij}, X_0\} > X_0$, then almost surely $X^+_{k'\ell'} \neq \min_i \max_j \max\{X^+_{ij}, X_0\}$ for all $(k',\ell') \neq (k,\ell)$, so we may skip this additional condition. Since $\dot{X}^+_{k\ell} = |Z|$ and $\dot{X}^-_{k\ell} = -|Z|$ and by the monotonicity of f, we have $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle \geq 0$ and $\langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle \leq 0$.

If $k = k_0$ then $X_{k\ell} = \min_i \max_j \max\{X^+_{ij}, X_0\} > X_0$ implies $X_{k\ell} = \min_i \max_j \max\{X^-_{ij}, X_0\} > X_0$, and in this case $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle + \langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle = 0$. Since this is the only case in which $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle$ is nonzero (with positive probability), we have almost surely $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle + \langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle \leq 0$.

If $k \neq k_0$ then $X_{k\ell} = \min_i \max_j \max\{X^-_{ij}, X_0\} > X_0$ implies $X_{k\ell} = \min_i \max_j \max\{X^+_{ij}, X_0\} > X_0$, and in this case $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle + \langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle = 0$. Since this is the only case in which $\langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle$ is nonzero (with positive probability), we have almost surely $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle + \langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle = 0$.

This settles the first case.

(2) Let $X_0 = Y + Z$ as described above. We define $\boldsymbol{X}^+(t), \boldsymbol{X}^-(t)$ by

$$X^+_{ij}(t) := X^-_{ij}(t) := X_{ij}, \quad \text{if } (i,j) \neq (k,\ell),$$

$$X^+_0(t) := Y + |Z|, \qquad X^-_0(t) := Y - |Z|, \qquad X^+_{k\ell}(t) := X_{k\ell} + t|Z|, \qquad X^-_{k\ell}(t) := X_{k\ell} - t|Z|.$$

Again, from the independence assumption on Z we obtain $\mathbb{E}[F(\boldsymbol{X}(t))] = \frac{1}{2}\mathbb{E}[F(\boldsymbol{X}^+(t)) + F(\boldsymbol{X}^-(t))]$, and it suffices to show that almost surely

$$\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle + \langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle \leq 0 \tag{B.6}$$

Note that $X^+_{ij} = X_{ij}$ for all (i,j). By (B.3),

$$\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle = \begin{cases} \dot{X}^+_{k\ell} f'(X_{k\ell} - X^+_0) & \text{if } X_{k\ell} = \min_i \max_j \max\{X_{ij}, X^+_0\} > X^+_0 \text{ and} \\ & \quad X_{k'\ell'} \neq \min_i \max_j \max\{X_{ij}, X^+_0\} \text{ for all } (k',\ell') \neq (k,\ell) \\ 0 & \text{else (almost surely),} \end{cases}$$

and similarly for $\langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle$. As in the first case, we can skip the additional uniqueness condition, which is almost surely satisfied, and again, $\dot{X}^+_{k\ell} = |Z|$ and $\dot{X}^-_{k\ell} = -|Z|$ and the monotonicity of f imply $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle \geq 0$ and $\langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle \leq 0$.

Now, $X_{k\ell} = \min_i \max_j \max\{X_{ij}, X^+_0\} > X^+_0$ implies $X_{k\ell} = \min_i \max_j \max\{X_{ij}, X^-_0\} > X^-_0$, and in this case $X_{k\ell} - X^+_0 \leq X_{k\ell} - X^-_0$. By convexity of f it follows that $f'(X_{k\ell} - X^+_0) \leq f'(X_{k\ell} - X^-_0)$, and thus $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle + \langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle \leq 0$. Since this is the only case in which $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle$ is nonzero, we have almost surely $\langle \nabla F(\boldsymbol{X}^+), \dot{\boldsymbol{X}}^+\rangle + \langle \nabla F(\boldsymbol{X}^-), \dot{\boldsymbol{X}}^-\rangle \leq 0$. This settles the second case. □

It remains to show that Lemma B.3 indeed implies Theorem B.2. We deduce this from general geometric arguments as used in [Mau11]. We reproduce these arguments in the following for convenience of the reader, except for Lemma B.4, which is a copy of [Mau11, Lem. 4] and follows from well-known properties of Euclidean distance matrices, cf. [KW12] for a recent survey on this theory.

In the following let $\mathscr{E}$ denote d-dimensional Euclidean space $\mathbb{R}^d$. Recall that a function $\Phi\colon \mathscr{E}^k \to \mathbb{R}$ is Euclidean motion invariant if

$$\Phi(\boldsymbol{x}_1 + \boldsymbol{y}, \ldots, \boldsymbol{x}_k + \boldsymbol{y}) = \Phi(\boldsymbol{Q}\boldsymbol{x}_1, \ldots, \boldsymbol{Q}\boldsymbol{x}_k) = \Phi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$$

for all $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k,\boldsymbol{y}\in\mathscr{E}$ and $\boldsymbol{Q}\in O(\mathscr{E})$. We identify the general linear group on $\mathscr{E}$ with the set of bases of $\mathscr{E}$:

$$\mathrm{GL}(\mathscr{E})=\{(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d):\boldsymbol{x}_i \text{ linear independent}\}\subset\mathscr{E}^d.$$

Let $D\colon\mathscr{E}^d\to\mathbb{R}^{\binom{d}{2}}$ be defined by

$$D(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d):=\big(\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2\big)_{i<j},\tag{B.7}$$

and denote $\Delta:=D(\mathscr{E}^d)$ and $\Delta_0:=D(\mathrm{GL}(\mathscr{E}))$.

Lemma B.4. *Let $\mathscr{E},D,\Delta,\Delta_0$ be defined as above. The sets Δ,Δ_0 are convex, Δ_0 is open, and Δ is the closure of Δ_0. Furthermore, any Euclidean motion invariant function $\Phi\colon\mathscr{E}\to\mathbb{R}$ factorizes uniquely over D,*

$$\Phi=\varphi\circ D,\qquad \varphi\colon\mathbb{R}^{\binom{d}{2}}\to\mathbb{R}.$$

Additionally, if Φ is continuous, then so is φ, and if Φ is differentiable on $\mathrm{GL}(\mathscr{E})$, then φ is differentiable on Δ_0.

The following lemma describes the proof strategy as demonstrated in [Mau11].

Lemma B.5. *Let $\mathscr{E},D,\Delta,\Delta_0$ be defined as above and let $\Phi\colon\mathscr{E}^d\to\mathbb{R}$ be a continuous Euclidean motion invariant function, which is differentiable on $\mathrm{GL}(\mathscr{E})$. Assume that for some symmetric sign matrix $\boldsymbol{S}\in\{1,-1\}^{d\times d}$ the following holds: for every basis $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d)\in\mathrm{GL}(\mathscr{E})$ and all i_0,j_0 there exists a curve $(\boldsymbol{x}_1(t),\dots,\boldsymbol{x}_d(t))$, $\boldsymbol{x}_i(0)=\boldsymbol{x}_i$, such that*

$$s_{ij}\,\tfrac{d}{dt}\|\boldsymbol{x}_i(t)-\boldsymbol{x}_j(t)\|^2|_{t=0}\begin{cases}<0 & \textit{if } (i,j)=(i_0,j_0)\\ =0 & \textit{else,}\end{cases}\qquad \textit{and}\qquad \tfrac{d}{dt}\Phi(\boldsymbol{x}_1(t),\dots,\boldsymbol{x}_d(t))|_{t=0}\le 0.$$

Then for all $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d),(\boldsymbol{y}_1,\dots,\boldsymbol{y}_d)\in\mathscr{E}^d$ satisfying $s_{ij}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2\ge s_{ij}\|\boldsymbol{y}_i-\boldsymbol{y}_j\|^2$ for all $i<j$, we have

$$\Phi(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d)\ge\Phi(\boldsymbol{y}_1,\dots,\boldsymbol{y}_d).\tag{B.8}$$

Proof. Using the decomposition $\Phi=\varphi\circ D$, we can paraphrase the claim in terms of φ. For this, we define

$$\big\{(a_{ij})_{i<j}\in\mathbb{R}^{\binom{d}{2}}: s_{ij}a_{ij}\le 0 \text{ for all } i<j\big\}=:C_{\boldsymbol{S}},$$

which is an isometric image of the nonnegative orthant $\mathbb{R}_+^{\binom{d}{2}}$. The claim of the lemma is that for all $(a_{ij}),(b_{ij})\in\Delta$ with $(b_{ij}-a_{ij})\in C_{\boldsymbol{S}}$ we have $\varphi(a_{ij})\ge\varphi(b_{ij})$.

By continuity of Φ it suffices to show the claim (B.8) for bases $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d),(\boldsymbol{y}_1,\dots,\boldsymbol{y}_d)\in\mathrm{GL}(\mathscr{E})$. In terms of φ the claim can then be restated by saying that for any point $(a_{ij})\in\Delta_0$ the derivative of φ is nonpositive in any direction $(v_{ij})\in C_{\boldsymbol{S}}$. By linearity of the derivative of φ and by convexity of $C_{\boldsymbol{S}}$, it suffices to show the monotonicity of φ in the extreme directions of the cone $C_{\boldsymbol{S}}$. Choosing such an extreme direction (v_{ij}) with $v_{ij}<0$ if $(i,j)=(i_0,j_0)$ and $v_{ij}=0$ if $(i,j)\ne(k,\ell)$, and letting the curve $\boldsymbol{X}(t)=(\boldsymbol{x}_1(t),\dots,\boldsymbol{x}_d(t))$ be such that $\frac{d}{dt}D(\boldsymbol{X}(t))|_{t=0}=(v_{ij})$, we obtain

$$\nabla_{(a_{ij})}\varphi(v_{ij})=\tfrac{d}{dt}\Phi(\boldsymbol{x}_1(t),\dots,\boldsymbol{x}_d(t))|_{t=0}\le 0$$

by assumption. This shows the monotonicity of φ in direction $C_{\boldsymbol{S}}$ and thus proves the claim. □

Proof of Theorem B.2. By continuity we may assume that f is differentiable on $(0,\infty)$ and satisfies $\lim_{x\to 0+}f'(x)=0$. We consider the Euclidean space $\mathscr{E}=\mathbb{R}\times\mathbb{R}^{m\times n}$, and define

$$\Phi\colon\mathscr{E}^{1+mn}\to\mathbb{R},\qquad \Phi(\boldsymbol{x}_0,\boldsymbol{x}_{11},\dots,\boldsymbol{x}_{mn}):=\mathbb{E}\big[\min_i\max_j f_+(\langle\boldsymbol{x}_{ij}-\boldsymbol{x}_0,\boldsymbol{g}\rangle)\big],$$

where $\boldsymbol{g}$ is a standard Gaussian vector in $\mathscr{E}$. The map Φ is Euclidean motion invariant, continuous, and differentiable on $\mathrm{GL}(\mathscr{E})$. Setting $X_0=\langle\boldsymbol{x}_0,\boldsymbol{g}\rangle$ and $X_{ij}=\langle\boldsymbol{x}_{ij},\boldsymbol{g}\rangle$, we have

$$\mathbb{E}|X_{ij}-X_{k\ell}|^2=\|\boldsymbol{x}_{ij}-\boldsymbol{x}_{k\ell}\|^2,\qquad \mathbb{E}|X_{ij}-X_0|^2=\|\boldsymbol{x}_{ij}-\boldsymbol{x}_0\|^2,$$

and we can reformulate the claim of Theorem B.2 in terms of Φ:

If $(\boldsymbol{x}_0, \boldsymbol{x}_{11}, \dots, \boldsymbol{x}_{mn}), (\boldsymbol{y}_0, \boldsymbol{y}_{11}, \dots, \boldsymbol{y}_{mn}) \in \mathscr{E}^{1+mn}$ satisfy

$$\|\boldsymbol{x}_{ij} - \boldsymbol{x}_{k\ell}\|^2 \le \|\boldsymbol{y}_{ij} - \boldsymbol{y}_{k\ell}\|^2 \text{ if } i \neq k, \qquad \|\boldsymbol{x}_{ij} - \boldsymbol{x}_{i\ell}\|^2 \ge \|\boldsymbol{y}_{ij} - \boldsymbol{y}_{i\ell}\|^2, \qquad \|\boldsymbol{x}_{ij} - \boldsymbol{x}_0\|^2 \ge \|\boldsymbol{y}_{ij} - \boldsymbol{y}_0\|^2,$$

then $\Phi(\boldsymbol{x}_0, \boldsymbol{x}_{11}, \dots, \boldsymbol{x}_{mn}) \ge \Phi(\boldsymbol{y}_0, \boldsymbol{y}_{11}, \dots, \boldsymbol{y}_{mn})$.

By Lemma B.5 we obtain a different condition that we need to verify, and in the remainder of the proof we will show that Lemma B.3 is exactly this condition. We restrict to the presentation of case (1), the second case follows analogously.

The decomposition $X_{k_0\ell_0} = Y + Z$ corresponds to the decomposition $\boldsymbol{x}_{k_0\ell_0} = \boldsymbol{y} + \boldsymbol{z}$ with $\boldsymbol{y}$ the orthogonal projection of $\boldsymbol{x}_{k_0\ell_0}$ on the linear span of $\boldsymbol{x}_0$ and $\boldsymbol{x}_{ij}$, $(i,j) \neq (k_0, \ell_0)$. Note that $\boldsymbol{z} \neq \boldsymbol{0}$. The curve $\boldsymbol{X}(t)$ defined in (B.4) corresponds to the curve $(\boldsymbol{x}_0(t), \boldsymbol{x}_{11}(t), \dots, \boldsymbol{x}_{mn}(t))$ in $\mathscr{E}^{1+mn}$ given by

$$\boldsymbol{x}_0(t) = \boldsymbol{x}_0, \qquad \boldsymbol{x}_{ij}(t) = \boldsymbol{x}_{ij}, \quad \text{for } (i,j) \neq (k,\ell), \qquad x_{k\ell}(t) = x_{k\ell} + t\boldsymbol{z}.$$

We obtain

$$\begin{aligned}
\|\boldsymbol{x}_{k\ell}(t) - \boldsymbol{x}_0(t)\|^2 &= \|\boldsymbol{x}_{k\ell} + t\boldsymbol{z} - \boldsymbol{x}_0\|^2 = \|\boldsymbol{x}_{k\ell} - \boldsymbol{x}_0\|^2 + t^2\|\boldsymbol{z}\|^2, \\
\|\boldsymbol{x}_{k\ell}(t) - \boldsymbol{x}_{ij}(t)\|^2 &= \|\boldsymbol{x}_{k\ell} + t\boldsymbol{z} - \boldsymbol{x}_{ij}\|^2 = \|\boldsymbol{x}_{k\ell} - \boldsymbol{x}_{ij}\|^2 + t^2\|\boldsymbol{z}\|^2, \quad \text{if } (i,j) \notin \{(k,\ell), (k_0,\ell_0)\}, \\
\|\boldsymbol{x}_{k\ell}(t) - \boldsymbol{x}_{k_0\ell_0}(t)\|^2 &= \|\boldsymbol{x}_{k\ell} + t\boldsymbol{z} - \boldsymbol{y} - \boldsymbol{z}\|^2 = \|\boldsymbol{x}_{k\ell} - \boldsymbol{y}\|^2 + (t-1)^2\|\boldsymbol{z}\|^2,
\end{aligned}$$

and thus

$$\begin{aligned}
\tfrac{d}{dt}\|\boldsymbol{x}_{ij}(t) - \boldsymbol{x}_0(t)\|^2|_{t=0} &= 0, \\
\tfrac{d}{dt}\|\boldsymbol{x}_{ij}(t) - \boldsymbol{x}_{i'j'}(t)\|^2|_{t=0} &= 0, \quad \text{if } \{(i,j),(i',j')\} \neq \{(k,\ell),(k_0,\ell_0)\}, \\
\tfrac{d}{dt}\|\boldsymbol{x}_{k\ell}(t) - \boldsymbol{x}_{k_0\ell_0}(t)\|^2|_{t=0} &= -2\|\boldsymbol{z}\|^2.
\end{aligned}$$

Hence, Lemma B.3 shows exactly the condition described in Lemma B.5, which finishes the proof. □