A New Perspective on Minimum-Norm Interpolation Under Gaussian Covariates

Gil Kur* ETH Zürich

Zong Shang CREST-ENSAE

Paul Simanjuntak Texas A&M

Guillaume Lecué ESSEC

Reese Pathak** Cornell & UC Berkeley

Abstract

Minimum-Norm Interpolators (MNI) in overparameterized linear models have gained attention as a tractable framework for studying interpolation phenomena that resemble empirical observations in neural networks. Most prior work on these interpolators either exploits closed-form solutions when available or relies heavily on Gaussian comparison results, such as the convex Gaussian Min-Max Theorem (CGMT). In this paper, we introduce a new perspective on MNI under isotropic Gaussian covariates by leveraging tools from high-dimensional geometry. First, we obtain a "localized" bound on the MNI's shrinkage of the original ground truth that occurs under isotropic Gaussian covariates when the norm is in an isotropic position. Then, we prove a sharp bound on the Mean Squared Error (MSE) of the ℓ_1-MNI, as obtained by Wang et al. [2022] via a geometric proof, which avoids invoking the CGMT and instead relies on the work of Fleury [2012] on Gaussian polytopes.[1]

1 INTRODUCTION

Recent experiments with neural networks have revealed counterintuitive statistical phenomena: models that interpolate the training data can still generalize well [Nakkiran et al., 2021]. In overparameterized regimes, it has been repeatedly observed that models achieving zero training error—even on noisy data—may nevertheless exhibit good out-of-sample performance [Zhang et al., 2021]. Motivated by these empirical observations, a substantial body of theoretical work has studied this phenomenon, often referred to as *harmless interpolation* or *benign overfitting.*

In the context of regression—the focus of this paper—an *interpolating* estimator is one that fits the training data exactly. However, such an estimator is generally not unique; in linear models, for instance, the interpolation constraints determine the estimate only up to the null space of the design matrix. It is well known that the particular interpolating solution can drastically affect generalization performance [Donhauser et al., 2022]. A common selection mechanism is the *minimum-norm interpolator* (MNI), which selects, among all interpolators, the one with the smallest norm.

Another motivation for studying MNIs comes from the observation that many first-order methods exhibit an "implicit bias": in overparameterized settings, and under suitable initialization, they converge to certain minimum-norm solutions [Gunasekar et al., 2018, Oravkin and Rebeschini, 2021, Shamir, 2022].

Formally, in this work, we consider the standard linear regression model with isotropic Gaussian covariates. We observe:

$$\mathcal{D} := \{(\mathbf{x}_i, y_i)\}_{i=1}^n, \qquad \mathbf{y} \;=\; \mathbf{X}w_\star + \boldsymbol{\xi},$$

where $\mathbf{X} \in \mathbb{R}^{n\times d}$ is the design matrix whose rows are i.i.d. isotropic Gaussian vectors $\mathsf{N}(0, I_d)$. We focus on the overparameterized setting, where $d > n$. Unless stated otherwise, the noise vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ is isotropic Gaussian and independent of $\mathbf{X}$. We write $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_d]$ for its columns.

The MNI, in the case of linear models, is defined via

$$\widehat{w}_n \;:=\; \underset{w\in\mathbb{R}^d}{\operatorname{argmin}} \Big\{ \|w\| \;:\; \mathbf{X}w = \mathbf{y} \Big\}, \qquad (1)$$

[1] This is a **preliminary version** of an extended paper titled "Minimum Norm Interpolation via The Local Theory of Banach Spaces: The Role of Gaussianity" that we will post in ArXiv for the results and their proof in their full generality.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

where $\|\cdot\|: \mathbb{R}^d \to \mathbb{R}_+$ denotes an arbitrary norm.

Inner-product norms. When the norm $\|\cdot\|$ is induced by an inner product, one can derive closed-form expressions for $\widehat{w}_n$. Prominent examples include the ℓ_2-MNI in linear regression and the minimum-norm interpolator with respect to a reproducing kernel Hilbert space (RKHS) norm. In the minimum-ℓ_2 setting, these formulas enable precise analyses in the proportional asymptotic regime where $d, n \to \infty$ with $d/n \to \gamma \in (0, \infty)$; see, for example, Ghorbani et al. [2021], Hastie et al. [2022], Mei and Montanari [2022]. They also facilitate non-asymptotic risk bounds for minimum-ℓ_2 and minimum-Hilbert-norm interpolators: a series of works [Bartlett et al., 2020, Tsigler and Bartlett, 2023, Lecué and Shang, 2022, Chinot et al., 2020, Muthukumar et al., 2020b, Zhou et al., 2023] characterizes the risk in terms of the spectral decay of the feature covariance (for linear regression) or of the associated integral operator (for kernel regression).

Arbitrary norms. In general, MNIs do *not* admit closed-form expressions when the norm in (1) is not induced by an inner product. In this setting, considerably less is known about their statistical behavior. Koehler et al. [2021] provided a first general analysis using local uniform convergence for Gaussian covariates, via the convex Gaussian min–max theorem (CGMT) [Thrampoulidis et al., 2015], obtaining non-asymptotic bounds on the prediction error of the MNI without any inner-product structure. Building on this approach, Donhauser et al. [2022], Wang et al. [2022] studied MNI for ℓ_p norms with $p \in [1, 2]$ in linear models with *isotropic* Gaussian covariates and derived sharp rates for its error. These analyses rely crucially on CGMT. More recently, Kur and Bizeul [2026], Kur et al. [2024] extended Donhauser et al. [2022] to sub-Gaussian covariates using an approach inspired by the geometry of 2-uniformly convex norms (cf. Klartag and Milman [2008]).

1.1 Positions

The notion of a *position* is classical in high-dimensional convex geometry: given a convex body $K \subset \mathbb{R}^d$ and a task at hand, one seeks a linear map $T \in \mathrm{GL}(d)$ so that the image TK is more "regular" with respect to a chosen functional. Equivalently, there exists a linear map T such that these inequalities hold with TK in place of K.

Milman's M-position. First, we denote by K° to be the polar/dual body of K, and B_d denotes the Euclidean unit ball in $\mathbb{R}^d$. We say that K is in M-Position Milman [1986] when the maximum of

$$\mathcal{N}(K, B_d), \mathcal{N}(B_d, K), \mathcal{N}(K^\circ, B_d), \mathcal{N}(B_d, K^\circ),$$

is upper bounded by C^d, for some constant $C \geq 0$, that is independent of K and n, where $\mathcal{N}(A, B)$ denotes the minimal number of translates of A required to cover B. Note that this position is far from being unique.

Isotropic Position. K is in isotropic position if the uniform measure over K has zero mean and identity covariance. Recent progress (e.g., Klartag and Lehec [2025], Bizeul and Klartag [2025]) shows that isotropic convex bodies are, up to universal constants, already close to M-position.

The unit balls of commonly used norms—such as the ℓ_p^d norms and the top-k norms—satisfy all of the standard positions in geometric analysis, up to scaling. Hence, by studying these norms alone, one arrives at a misleading conclusion: that the notion of position is somehow irrelevant. Indeed, we emphasize that **for general, arbitrary norms**, the situation is much more subtle—-see Giannopoulos and Milman [2000] and the monographs Pisier [1989], Artstein-Avidan et al. [2015].

To be clear, throughout this paper, we refer to a norm $\|\cdot\|$ lying in some position if $c_{\|\cdot\|}K$ lies in this position for some normalization constant $c_{\|\cdot\|} > 0$ and where K denotes the unit ball of $\|\cdot\|$.

1.2 The ℓ_1-MNI

The ℓ_1 MNI, also known as basis pursuit (BP), is:

$$\widehat{w}_n := \operatorname*{argmin}_{w \in \mathbb{R}^d} \Big\{ \|w\|_1 \ : \ \mathbf{X}w = \mathbf{y} \Big\}. \qquad (2)$$

It is well-known that it generalizes well in the noiseless case but is highly sensitive to noise Candes [2008], Donoho and Elad [2006]. For noisy data, upper bounds on the prediction error of order σ^2 have been derived for isotropic Gaussian features Koehler et al. [2021], Ju et al. [2020], Wojtaszczyk [2010], subexponential features Foucart [2014], and even heavy-tailed features Chinot et al. [2020], Krahmer et al. [2018]. In the isotropic Gaussian setting and with adversarial noise, Chinot et al. [2020] established a lower bound of order $\sigma^2/\log(d/n)$ (see also Chatterji and Long [2021], Muthukumar et al. [2020a]). They conjectured that the ℓ_1-MNI is inconsistent in the non-adversarial case. This conjecture was recently refuted by Wang et al. [2022]. These authors used the convex Gaussian min–max theorem (CGMT) to establish the sharp rate:

$$(1 + o(1))\frac{\sigma^2}{\log(d/n)}.$$

1.3 Setting

In this work, we present a new geometric perspective on MNI in linear models with *isotropic* Gaussian covariates. Our approach leverages tools from high-dimensional geometry to obtain sharp error rates and does not rely on the convex Gaussian min–max theorem (CGMT) or any Gaussian comparison inequalities. Our goal is to understand the mean-squared error (MSE) of $\widehat{w}_n$ that is, in the case of isotropic covariates,

$$\mathbb{E}_{\mathcal{D}}\big[\|\widehat{w}_n - w_\star\|_2^2\big].$$

We study the following decomposition of the MSE, also used in Kur et al. [2024], which preset is a **sum** of three terms, E_1, E_2, E_3, defined as follows:

$$E_1 := \mathbb{E}\big\|\mathcal{P}_{w_\star}\big(\mathbb{E}_{\boldsymbol{\xi}}[\widehat{w}_n \mid \mathbf{X}]\big) - w_\star\big\|_2^2$$

where $\mathcal{P}_{w_\star}$ denotes the orthogonal projection onto the span of $w_\star \in \mathbb{R}^d$. This term intuitively measures how much in ℓ_2-distance the MNI shrank from the $w_\star$. Then, we define

$$E_2 := \|\mathcal{P}_{(w_\star)^\perp}\mathbb{E}_{\boldsymbol{\xi}}[\widehat{w}_n \mid \mathbf{X}])\big\|_2^2$$

where $\mathcal{P}_{(w_\star)^\perp}$ denotes the projection to the orthogonal complement $(w_\star)^\perp$. And finally,

$$E_3 := \mathbb{E}_{\mathbf{X}}\mathrm{Var}_{\boldsymbol{\xi}}(\widehat{w}_n \mid \mathbf{X}) = \mathbb{E}\|\widehat{w}_n - \mathbb{E}_{\boldsymbol{\xi}}[\widehat{w}_n \mid \mathbf{X}]\|_2^2,$$

that is the expected conditional variance.

Overview of our contributions.

(i) Theorem 1 characterizes the term E_1 in a localized way under isotropic Gaussian covariates, for any MNI, such that its unit ball is in isotropic position. This yields the first *localization* principle for E_1 analogous to results for constrained least squares estimators (cf. Chatterjee [2014], Bartlett et al. [2005], Bartlett and Mendelson [2002]). The characterization depends on the mean norm of sections of the convex body K in the direction $w_\star$, defined rigorously below.

(ii) Theorem 2 establishes a sharp risk bound for the ℓ_1-MNI without invoking Gaussian comparison inequalities or the CGMT. The approach draws on tools from high-dimensional geometry, probability, and super-concentration, in particular on the geometry of symmetric Gaussian polytopesFleury [2012]. We note that the connection between Gaussian polytopes and BP was already present in statistics Donoho and Tanner [2009, 2010], Efron [1965], but was not utilized in more recent works on the benign overfitting.

Organization of this paper. Section 2 introduces notation and preliminaries. Section 3 states our main results. Section 4 provides a detailed discussion. The remaining parts appear in the Appendix below.

2 PRELIMINARIES

Notation

For functions $f, g\colon \mathcal{Z} \to \mathbb{R}_+$, we write $f \lesssim g$ if there exists a constant $C > 0$ such that $f(z) \le C\,g(z)$ for all $z \in \mathcal{Z}$. We write $f \gtrsim g$ if $g \lesssim f$, and $f \asymp g$ if both $f \lesssim g$ and $g \lesssim f$ hold. We also use standard Landau notation $O(\cdot)$, $\Omega(\cdot)$, $\Theta(\cdot)$, and write $f \ll g$ (equivalently, $g \gg f$) to mean $f/g \to 0$. For sets E, F in an additive group, the Minkowski sum is $E + F := \{e + f : e \in E,\ f \in F\}$; for a singleton $E = \{e\}$, we write $e + F := \{e\} + F$.

For any $d \ge 1$ and $x \in \mathbb{R}^d$, we write $\|x\|_p := \big(\sum_{i=1}^d |x_i|^p\big)^{1/p}$ for $p \in [1, \infty)$ and $\|x\|_\infty := \max_{1 \le i \le d} |x_i|$, and B_p^d denotes its unit ball.

For a convex body $K \subset \mathbb{R}^d$ containing the origin, its (Minkowski) gauge is

$$\|x\|_K := \inf\{\lambda \ge 0 : x \in \lambda K\}.$$

For any $m \in \mathbb{N}$ and measurable $A \subset \mathbb{R}^m$, we write $|A|$ for its Lebesgue measure (volume) and $|\partial A|$ for the $(m-1)$-dimensional surface measure of its boundary. For a square matrix M, we may write $|M|$ for its determinant. Finally, symbols $C, C_1, c, c_1, \ldots > 0$ denote absolute constants whose values may change from line to line.

The *mean norm* of a convex set $K \subset \mathbb{R}^d$ is defined as

$$M(K) := \mathbb{E}\|\boldsymbol{\xi}\|_K$$

here $\boldsymbol{\xi} \sim U(\mathbb{S}^{d-1})$ is uniform on the unit sphere in $\mathbb{R}^d$. For every convex body $L \subset \mathbb{R}^m$ (for some $m \ge n+1$), define

$$M_n(L) \;:=\; \mathbb{E}_{\mathbf{X}}\, M(\mathbf{X}L) \;=\; \mathbb{E}_{\mathbf{X},\boldsymbol{\xi}}\, \|\boldsymbol{\xi}\|_{\mathbf{X}L},$$

here $\boldsymbol{\xi} \sim U(\mathbb{S}^{n-1})$ and $\mathbf{X}L \subset \mathbb{R}^n$ is the Gaussian image of L under a random matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ with i.i.d. entries $\mathsf{N}(0,1)$. For every $w \in K$ and $t \ge 0$, define the affine *sections* of K by

$$K(w,t) \;:=\; \mathcal{P}_{w^\perp}\big(\big(K \cap \big(\tfrac{(1-t)w}{M_n(K)} + w^\perp\big)\big) - \tfrac{(1-t)w}{M_n(K)}\big).$$

The corresponding mean gauge is denoted by

$$M_n(w,t) \;:=\; M_n\big(K(w,t)\big) \;=\; \mathbb{E}\,\|\boldsymbol{\xi}\|_{\mathbf{X}K(w,t)}.$$

Finally, $r(K)$ and $R(K)$ be the inner and outer radius of a set K, and $M^*(K)$ is the mean width of a set K, that is

$$M^*(K) = M(K^\circ) = \mathbb{E}_{u \sim U(\mathbb{S}^{d-1})} \sup_{w \in K} \langle u, w \rangle.$$

3 MAIN RESULTS

3.1 On the shrinkage of the MNI

Recall that $K \subset \mathbb{R}^d$ is the unit ball of $(\mathbb{R}^d, \|\cdot\|)$.

Assumption 1. *The following holds:*

- $M_n(K) \gg 1$ *and* $\|w_\star\| \asymp \|w_\star\|_2 \asymp 1$
- K *is in isotropic position.*

In words, noise is "harder" to interpolate than the pure signal, and the ℓ_2 norm of the signal is comparable to its norm in terms of $\|\cdot\|$. Hence, the unit ball must be inflated to interpolate pure noise.

Assumption 2. *With probability of at least* $1 - d^{-3}$ *it holds*

$$d_\star^2 := \|\mathcal{P}_{(w_\star)^\perp}(\widehat{w}_n)\|_2^2 \lesssim \log(d)^C \cdot (E_2 + E_3)$$

for some universal constant $C \geq 0$.

For example, this assumption holds for ℓ_p-norm (with an absolute constant) when $p \in [1, 2]$. Clearly, here we implicitly assume a position to satisfy this assumption. For example, if K is a very thin ellipse with unbalanced eigenvalues, then its corresponding MNI would not satisfy this assumption, as in this case, the random variable $\|\mathcal{P}_{(w_\star)^\perp}(\widehat{w}_n)\|_2$ is heavy-tailed.

To state our first result, denote by $M_n(t) := M_n(w_\star, t)$ and define the **localization radius** $t_\star \equiv t_\star(K, w_\star)$ as

$$\operatorname*{argmax}_{t \geq 0} \left\{ M_n(t) \; - \; t^2 M_n(t) \, \|w_\star\|_2 / 2 \right\},$$

We also define the **offset radius** $\widetilde{t}_\star > 0$ as the minimal $t > 0$ so that

$$M_n(t) \; - \; t^2 M_n(t) \|w_\star\|_2 + \tilde{O}(d_\star / \sqrt{n}) \leq 0.$$

Note that the following relation always holds $t_\star \lesssim \widetilde{t}_\star$.

Theorem 1. *There exists a sufficiently large absolute constant* $C > 0$ *such that if* $d \geq Cn$*, then under Assumptions 1–2,*

$$\mathbb{E}\big\|\mathcal{P}_{w_\star}(\widehat{w}_n) - w_\star\big\|_2^2 \lesssim \widetilde{t}_\star^2.$$

Remark 1. Here, we surpass the $O(n^{-1/2})$ barrier appearing in the analysis of [Zhou et al., 2023, Lemma 10] under the consistency of the MNI, and in particular, $E_2 + E_3 \to 0$. ◁

Remark 2. In the well-studied setting of the ℓ_p-MNI for $p \in [1, 2]$, and $w_\star = (1, 0, \ldots, 0)$, one can show that

$$t_\star \approx \operatorname*{argmax}_{t \in [0,1]} \left(M_n(t) - \frac{t^2 M_n(t)}{2} \right),$$

without the deviations parameters, as the sections are homothetic copies ℓ_p ball in $\mathbb{R}^{d-1}$. However, as we only assume that K is isotropic, we carry the offset $d_\star$ and upper-bound it by $\widetilde{t}_\star$. ◁

Remark 3. To upper bound the term E_2, we would require more structure on the norm than being isotropic; we would require the norm to be 2-uniformly convex or at least cotype 2 — we refer to the work of the first author Kur and Bizeul [2026] for more details. ◁

3.2 On the ℓ_1-MNI

Our final result gives a sharp risk bound for the ℓ_1–MNI, defined in (2) via a purely geometric (non–CGMT) argument, that was obtained in Wang et al. [2022].

Theorem 2. *Let* $c \in (0, 1/6)$*, and assume that* $d \in (n \log n^C, \exp(n^c))$*, and assume that* $w_\star$ *is* $O(n/\log(d/n)^5)$*–sparse. Then, with probability of at least* $1 - \exp(-c_1 n \log(d/n)^{-2})$*, it holds that*

$$\|\widehat{w}_n - w_\star\|_2^2 \;=\; \frac{(2 + p_{n,d}) M_{n,d}^2}{n} = \frac{1 + o(1)}{\log(d/n)},$$

where $p_{n,d} \lesssim \log(d/n)^{-2}$*, and* $M_{n,d}$ *is defined in Corollary 1 below..*

We emphasize that our proof relies crucially on Fleury [2012], who showed that a symmetric Gaussian polytope satisfies the KLS conjecture in expectation.

Remark 4. In this version of this work, we prove Theorem 2 for the zero regressor. It is possible to generalize the result to sparse vectors, with additional technical work that brings very few new statistical insights. ◁

4 DISCUSSION

4.1 Theorem 1 in ℓ_p-linear regression

Consider the case of $\|\cdot\| = \|\cdot\|_p$ for $p \in [1, 2]$, and let $w_\star = (1, 0, \ldots, 0)$, which obey Assumptions 1-2, see [Kur and Bizeul, 2026, Cor 2.]. It can be shown Gordon et al. [2007], Paouris et al. [2017] that

$$M_n(B_p^d) \asymp \frac{\sqrt{n}}{d^{1-1/p}} \cdot \sqrt{\max\left\{(p-1), \frac{1}{\log(d/n)}\right\}}.$$

For our theorem to be valid, we must assume that $M_n := M_n(B_p^d) \gg 1$, or equivalently,

$$d \lesssim \tilde{O}(\min\{n^{q/2}, \exp(n^c)\}),$$

where $1/p + 1/q = 1$. As the sections $K(w_\star, t)$ are homothetic B_p^{d-1} balls, we know that

$$M_n(t) - M_n(0) = M_n(1) \cdot (1-(1-t)^p)\left(\frac{1}{M_n}\right)^p \approx M_n \cdot pt \cdot \left(\frac{1}{M_n}\right)^p.$$

Therefore, we need to balance $t^2 \asymp ptM_n^{-p}$. Invoking Theorem 1 yields

$$E_1 \asymp t_\star^2 \asymp \frac{d^{2p-2}}{n^p} \cdot \min\left\{(p-1)^{-1}, \log(d/n)\right\}^p,$$

recovering the result of Donhauser et al. [2022], Wang et al. [2022].

4.2 CGMT and MNI

As we discuss in further detail in the extended version of this paper, our proof is based on studying the symmetric Gaussian polytope. It is defined as

$$P_{n,d} = \mathrm{Conv}\{\pm \mathbf{X}_i\}_{i=1}^d = \mathbf{X}B_1^d \subset \mathbb{R}^n.$$

And we prove that it has a *vanishing* finite volume ratio. Specifically,

$$r_{n,d}B_n \subset P_{n,d} \text{ and } \left(\frac{|P_{n,d}|}{|B_n|}\right)^{1/n} = (1+o_{d/n}(1))r_{n,d},$$

Here, we denote $o_{d/n}(1)$ for a function $f(n,d)$ which vanishes as $n, d/n \to \infty$. By using this volume bound and applying a simple self-improvement argument, our approach yields a sharper estimate on $M_{n,d}$ than was obtained in Wang et al. [2022]. Specifically, our approach establishes the following.

Corollary 1. *The following estimate*

$$M_{n,d} = M_n(B_1^d) = \left(1 + O\left(\frac{1}{\log(d/n)^2}\right)\right) \cdot \sqrt{\frac{n}{\alpha(n,d)}}.$$

Above, we define

$$\alpha(n,d) = 2\log(d/n) - \mathrm{lln}(d/n) - \ln\pi + 2 + o_{d/n}(1).$$

In Wang et al. [2022], they proved a weaker upper bound on $M_{n,d}$ of

$$(1 + O(\log(d/n)^{-2})) \cdot \frac{\sqrt{n}}{t_{n,d}} = \sqrt{\frac{n}{\alpha(n,d) - 2}},$$

where $t_{n,d}$ is defined the solution to

$$\Pr_{g\sim N(0,1)}(|g| \geq t_{n,d}) = \frac{n}{d}.$$

In this paper, they also claimed that a sharp estimate on $M_{n,d}$ implies an upper bound on the MSE. Indeed, this is not an effect of CGMT, as our proof suggest that an exact estimate on $M_{n,d}$ determines $p_{n,d}$ that is defined as the sharpest upper bound of the following:

$$1 \leq \left(\frac{\mathbb{E}\|\boldsymbol{\xi}\|_{P_{n,d}}^{-n}}{(\mathbb{E}\|\boldsymbol{\xi}\|_{P_{n,d}})^{-n}}\right)^{1/n} \leq 1 + p_{n,d}$$

where the LHS follows from Jensen's inequality. We prove that $p_{n,d} \lesssim 1/\log(d/n)^2$, and are not sure if this bound is tight or not. However, we believe that $p_{n,d} \gtrsim 1/(\sqrt{n}\log(d/n))$.

4.3 Theorem 2 for sub-Gaussian covariates

Theorem 2 deeply relies on the result of Fleury [2012], who computed the exact distribution of a facet of a symmetric Gaussian polytope. One may ask if we can extend this result to sub-Gaussian matrices. The key challenge is that Dvoretsky's theorem only holds for rotationally invariant distributions. For completeness, we state it in its sharpest version (cf. Paouris and Valettas [2016], Klartag and Vershynin [2007]):

Theorem. *Assume that $d \geq n+1$. Let $P = \frac{1}{\sqrt{d}}\mathbf{X}$ and consider a convex body $K \subset \mathbb{R}^d$. Then, the following holds with probability of $1 - \exp(-c_1 n)$:*

$$M^*(K)\Big(1 - C_1\sqrt{\tfrac{n}{d}}\underbrace{\sqrt{n \cdot \mathrm{Var}\left(\tfrac{\|u\|_{K^\circ}}{M^*(K)}\right)}}_{(*)}\Big) \leq r(PK)$$

and

$$R(PK) \leq \Big(1 + C_1\sqrt{\tfrac{n}{d}}\underbrace{\frac{R(K)}{M^*(K)}}_{(**)}\Big)M^*(K)$$

The lower and upper inclusion have different behaviors. Indeed, by the Poincare inequality for the uniform measure on the sphere, it holds that $(*) \lesssim (**)$. When the inequality is strict, $\|u\|_{K^\circ}$ is *superconcentrated*, where $u \sim U(\mathbb{S}^{d-1})$.

4.4 Small Ball Probabilities

Definition 1. *The canonical simplex and its centered version are defined via*

$$\triangle := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1\} \text{ and } \triangle_c := \triangle - \frac{\mathbf{1}_n}{n}.$$

Here, we state a useful lemma we proved that allows us to establish a lower bound on the MSE of the ℓ_1-MNI.

Theorem. *The following holds for all $\varepsilon \in (0,1)$:*

$$\Pr_{Z \sim U(\triangle_c)}(\|Z\|_2 \leq (1-\varepsilon)\mathbb{E}\|Z\|_2) \leq \exp(-cn\varepsilon^2).$$

The proof of this lemma follows from standard calculations of MGFs, and noticing that $Z \sim \text{Unif}(\triangle)$, can also be presented as

$$Z \sim \frac{(Z_1, \ldots, Z_n)}{\sum_{i=1}^n Z_i}$$

where $Z_1, \ldots, Z_n \sim \text{Exp}(1)$. **However**, as one can easily show (cf. Paouris [2006]) a **sharp** bound of

$$\Pr_{Z \sim U(\triangle_c)}(\|Z\|_2 \geq (1+\varepsilon)\mathbb{E}\|Z\|_2) \leq \exp(-c\sqrt{n}\varepsilon).$$

Namely, the tails exhibit different behaviors; one is sub-Gaussian, and the other is sub-exponential. Therefore, we would need a different approach for the upper bound the MSE. Finally, we state a useful bound that is known only to specialists; see also [Alonso-Gutiérrez and Prochno, 2015, Corollary 5.1].

Theorem (Lemma 2.3 Paouris [2004])**.** *Let $K \subset \mathbb{R}^n$ be an isotropic convex body with a ℓ_2-**diameter** $C\sqrt{n}$. Then, with probability $1-\exp(-c_1 n)$ over $\theta \sim U(\mathbb{S}^{n-1})$, the marginal $\langle X, \theta \rangle$ satisfies for all $t \in [c_3, C\sqrt{n}]$*

$$\Pr_{X \sim U(K)}\left\{|\langle X, \theta\rangle| \geq t\right\} \leq \exp(-c_2 t^2),$$

where the constants $c_1, c_2, c_3 > 0$ depend on $C \geq 0$.

In words, w.h.p., a random marginal of a bounded isotropic K has a sub-Gaussian tail with parameter $O(1/\sqrt{c_2})$, comparing the worst case that can be sub-exponential.

5 PROOFS

5.1 Proof of Theorem 1

Note for a convex set $L \subset \mathbb{R}^d$ that contains the origin, and $\mathbf{z} \in \mathbb{R}^n$, we have

$$\|\mathbf{z}\|_{\mathbf{X}L} = \|\widehat{w}_n^L(\mathbf{X}, \mathbf{z})\|_L,$$

where $\widehat{w}_n^L$ is as the minimal norm/gauge solution with respect to the origin.

Lemma 1. *[[Kur and Bizeul, 2026, Lemma 1]] Let $P = \frac{1}{\sqrt{d}}\mathbf{X}$, and assume that $d \geq C_3 n$ for large enough $C_3 \geq 0$. Then, with probability of at least $1-\exp(-cn)$ it holds that*

$$1 \lesssim r(PK) \leq M^*(PK) \lesssim \log(d)^3.$$

The proof of this result is based on the recent works of Milman [2015], Klartag and Lehec [2025], Bizeul [2025], Bizeul and Klartag [2025] which concern isotropic convex sets.

Lemma 2. *[Kur and Bizeul, 2026, Cor. 2] Consider a convex set $L \subset \mathbb{R}^{d-1}$ that contains that origin such that*

$$R_{bM^*} = M^*(PK)/r(PK) = \tilde{\Theta}(1)$$

, and with of probability of at least $1-d^{-100}$ over $\mathcal{D}$

$$\|\widehat{w}_n^L(\mathbf{X}, \boldsymbol{\xi})\|_2 \in \widetilde{O}(d_*)B_{d-1}$$

Then, with probability of $1-d^{-98}$, for any fixed $u \in \mathbb{S}^{n-1}$, it holds

$$\left|\frac{\|u\|_{\mathbf{X}L}}{\mathbb{E}\|u\|_{\mathbf{X}L}} - 1\right| = \widetilde{O}\left(\frac{d_*}{\sqrt{n}}\right).$$

Finally, note that u can be arbitrary, as for any fixed rotation matrix U, it holds $U\mathbf{X} \sim \mathbf{X}$. And the sets L that we consider are the shifted sections of the unit ball of $\|\cdot\|_K$ in the direction of $w_\star$ intersected with $\tilde{O}(d_*)$-ℓ_2 ball. Note that we can use the last two lemmas due to Assumptions 1 and 2. The $n^{-1/2}$ scaling in Lemma 2 is due to the fact that w.h.p. $\|\boldsymbol{\xi}\|_2 \approx \sqrt{n}$.

To begin, we assume that $\|w_\star\|_2 = \|w_\star\| = 1$, and let $m_{\boldsymbol{\xi}+\mathbf{w}_\star} = \|\mathbf{w}_\star + \boldsymbol{\xi}\|_n$, and let

$$\tilde{K} = K \cap (d_* B_{d-1} \times \text{span}\{w_\star\})$$

and denote by

$$\widetilde{K}(t) = \left((\widetilde{w}_t + w_\star^\perp) \cap \tilde{K}\right) - \widetilde{w}_t$$

where where $w_t := \frac{(1-t)w_\star}{m_{\boldsymbol{\xi}+\mathbf{w}_\star}}$. Also, denote by

$$K(t) = \left((w_t + w_\star^\perp) \cap \tilde{K}\right) - w_t$$

where $w_t := \frac{(1-t)w_\star}{M_n(0)}$. Denote by $\widetilde{K}_n(t), K_n(t)$ to be the coordinate projection of $\widetilde{K}(t), K^*(t)$ over the design points, i.e. $\widetilde{K}_n(t) := \widetilde{\mathbf{X}}K(t) \subset \mathbb{R}^n$, where $\widetilde{\mathbf{X}} := \mathcal{P}_{w_\star^\perp}(\mathbf{X})$. For every realization of $\mathbf{X}, \boldsymbol{\xi}$ it holds

$$\widehat{w}_n \in \|\boldsymbol{\xi} + t_* \mathbf{w}_\star\|_{\widetilde{K}_n(t_*)} \cdot \widetilde{K}(t_*)$$

and

$$\|\boldsymbol{\xi} + t_* \mathbf{w}_\star\|_{\widetilde{K}_n(t_*)} = \|\boldsymbol{\xi} + \mathbf{w}_*\|_n,$$

for some $t_* \in [0,1]$ that depends both in $\mathbf{X}, \boldsymbol{\xi}$. Also, note that

$$\|\boldsymbol{\xi} + t \mathbf{w}_\star\|_{\widetilde{K}_n(t)} \geq \|\boldsymbol{\xi} + \mathbf{w}_*\|_n, \tag{3}$$

for any $t \in \mathbb{R}$ with equality for $t = t_*$. Note that for $t = 0$, which is our "baseline", there is no shrinkage of the signal. Clearly, $\widetilde{K}(t)$ depends on $\mathbf{X}, \boldsymbol{\xi}$ as $\widetilde{w}_t$ is random variable. As we see below, we can prove our theorem on $K(t)$ instead.

Denote by

$$\alpha(t) = \frac{t^2}{2} - O\left(\sqrt{\frac{\log(d)}{n}} t + t^4\right).$$

The main observation of this theorem is the following: For any fixed $\boldsymbol{\xi}$, we take expectation over $\widetilde{\mathbf{X}}$, and with probability of $1 - d^{-100}$ over $\mathbf{w}_\star$, it holds

$$\begin{aligned} \mathbb{E}\|\boldsymbol{\xi} + t\mathbf{w}_\star\|_{K_n(\mu)} &= \|\boldsymbol{\xi} + t\mathbf{w}_\star\|_2 \cdot \mathbb{E}\|\boldsymbol{\xi}\|_{K_n(\mu)} \\ &= (1 + \alpha(t)) \cdot M_n(\mu). \end{aligned} \tag{4}$$

where we first used that $\mathbf{w}_\star$ is uncorrelated from $K_n(t)$, and under **Gaussianity** is independent, and homogeneity of the gauge function. Then, we used that with probability $1 - d^{-100}$ that

$$\|\boldsymbol{\xi} + t\mathbf{w}_\star\|_2 = 1 + \alpha(t)$$

and Taylor's expansion

$$\sqrt{1 + 2\alpha(t)} = 1 + \alpha(t) + O(\alpha(t)^2).$$

Now, under Assumption 2, we know that for every fixed μ and t in our range it holds

$$\|\boldsymbol{\xi} + t\mathbf{w}_\star\|_{K_n(\mu)} - (1 + \alpha(t)) \cdot M_n(\mu) \lesssim \frac{\log(d)^C d_*}{\sqrt{n}}$$

Therefore, we discretize $t, \mu \in [0,2]$ to cubes with edge length of d^{-100}, and apply Assumption 2 to each μ_i, t_i on this discretization, take a union bound, and note that for $|t_1 - t_2| = o(1)$ and $|\mu_1 - \mu_2| = o(1)$

$$\|\boldsymbol{\xi} + t_1\mathbf{w}_\star\|_{K_n(\mu_1)} - \|\boldsymbol{\xi} + t_2\mathbf{w}_\star\|_{K_n(\mu_2)} \lesssim \tilde{O}(|t_1 - t_2| + |\mu_1 - \mu_2|)M_n,$$

where we used Lemma 1, and therefore the triangle inequality we conclude that

$$\|\boldsymbol{\xi} + t\mathbf{w}_\star\|_{K_n(\mu)} - (1 + \alpha(t)) \cdot M_n(\mu) \lesssim \frac{\log(d)^C d_*}{\sqrt{n}}$$

uniformly for all μ, t. By choosing the proper μ and t such that

$$K_n(\mu) = \tilde{K}_n(t)$$

we can obtain that uniformly it holds

$$\|\boldsymbol{\xi} + t\mathbf{w}_\star\|_{\widetilde{K}_n(t)} - (1 + \alpha(t)) \cdot M_n(t) \lesssim \frac{\log(d)^C d_*}{\sqrt{n}} \tag{5}$$

Hence, we combine (5),(4),(3), and conclude that

$$\frac{\|\boldsymbol{\xi} + \mathbf{w}_\star\|_{\tilde{K}_n(t)}}{m_{\boldsymbol{\xi}+\mathbf{w}_\star}} \leq 1 - F_*(t)$$

where $F_*(t)$ is defined as

$$M_n(\widetilde{K}(t)) - M_n(\widetilde{K}(0)) + \alpha(t) M_n(\widetilde{K}(t)) + \widetilde{O}(\frac{d_*}{\sqrt{n}}).$$

Clearly, $F_*(t)$ must be positive; otherwise, by (3), we contradict the definition of the MNI. Meaning that $\widetilde{t}_*$ is upper bounded by the minimal $t > 0$ such that $F_*(t) = 0$. In order to connect, $K(t)$ and $\widetilde{K}(t)$, we know that

$$|m_{\boldsymbol{\xi}+\mathbf{w}_\star} - M_n(\widetilde{K}(0))| \lesssim \widetilde{t}_*^2 M_n(0),$$

By triangle inequality (and using that $\|w_\star\| = 1$ and $M_n(K) \gg 1$), it holds

$$(1 - \frac{C\widetilde{t}_*^2}{M_n(t)})\widetilde{K}(t) \subset K(t) \subset (1 + \frac{C\widetilde{t}_*^2}{M_n(t)})\widetilde{K}(t)$$

and therefore

$$M_n(K(t)) - M_n(\widetilde{K}(t)) \lesssim \widetilde{t}_*^2$$

and the proof is complete.

5.2 Proof Outline of Theorem 2

Notation

For a positive integer $k \geq 1$, we define $[k] := \{1, \ldots, k\}$. For a matrix $A \in \mathbb{R}^{n \times d}$ with columns $\{A_i\}_{i=1}^d \subset \mathbb{R}^n$ and a subset $S \subset [d]$, the matrix $A_S \in \mathbb{R}^{n \times |S|}$ is composed by taking the columns $\{A_i\}_{i \in S}$. Additionally, for a square matrix $B \in \mathbb{R}^{k \times k}$, we denote

$$B^{-\mathsf{T}} = (B^{\mathsf{T}})^{-1} = (B^{-1})^{\mathsf{T}}.$$

Consider a facet of $\mathcal{F}$ of the random polytope $P_{n,d}$. Note that it is defined by $S \subset [d]$ and $|S| = n$ and

$\boldsymbol{\varepsilon} \in \mathbb{R}^d$ such that $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i\in S} \in \{\pm 1\}^n$, such that $\{\varepsilon_i \mathbf{X}_i\}_{i\in S}$ form a facet. We denote the barycenter and the (scaled) normal, respectively, as

$$\mathbf{c}_{\mathcal{F}} := \frac{1}{n}\sum_{i\in S} \varepsilon_i \mathbf{X}_i \text{ and } \mathbf{n}_{\mathcal{F}} = \frac{\mathbf{X}_S^{-\mathsf{T}}\boldsymbol{\varepsilon}}{\|\mathbf{X}_S^{-\mathsf{T}}\boldsymbol{\varepsilon}\|_2^2}$$

Note that in this scaling $n_{\mathcal{F}} \in \mathcal{F}$, and $\mathcal{F}$ is contained in the hyperplane

$$\{v \in \mathbb{R}^n : \langle v, \mathbf{n}_{\mathcal{F}}\rangle = \|\mathbf{n}_{\mathcal{F}}\|_2^2\}.$$

Throughout the argument below, we let $\mathcal{F}$ correspond to the facet (with corresponding n-subset S and sign vector $\boldsymbol{\varepsilon}$) which is selected by $\widehat{w}_n$. For any polytope P, we will use $\mathcal{F}_{n-1}(P)$ to denote its collection of facets (*i.e.,* its $(n-1)$-dimensional faces). We denote by $|\mathcal{F}_{n-1}(P)|$ the number of facets.

5.2.1 Fleury's work on Gaussian polytopes

Lemma 3 (Thm. 3 in Fleury [2012])**.** *Suppose that* $\mathrm{conv}\{\varepsilon_i \mathbf{X}_i\}_{i\in S}$ *forms a facet for some signs* $\varepsilon_i \in \{\pm 1\}$ *and a subset* $S \subset [d]$ *with* $|S| = n$*. The distribution of the matrix* $\mathbf{X}_S$*, formed with the columns* $\varepsilon_i \mathbf{X}_i, i \in S$*, conditional on it being a facet, is distributed as the product* $\mathbf{UA}$*, where* $\mathbf{U}$ *and* $\mathbf{A}$ *are independent,*

$$\mathbf{U} \sim \mathsf{Unif}(O(n)), \quad \mathbf{A} = \begin{bmatrix} \mathbf{Y} \\ T_{n,d}\cdot \mathbf{1}_n^{\mathsf{T}}\end{bmatrix}, \quad \mathbf{Y} \in \mathbb{R}^{(n-1)\times n},$$

where

- $\mathbf{Y}$ *has columns* $\{\mathbf{Y}_i\}_{i=1}^n$ *distributed with density proportional to*

$$|\mathbf{y}\mathbf{y}^{\mathbf{T}}| \cdot \exp\Big(-\sum_{i=1}^n \|\mathbf{y}_i\|_2^2/2\Big),$$

  where $|\cdot|$ *denotes the determinant.*

- $T_{n,d}$ *denotes the distribution of* $\|\mathbf{n}_{\mathcal{F}}\|_2$ *is independent of* $\mathbf{Y}$*, and has the density*

$$p(t) \propto \Pr_{g\sim N(0,1)}(|g| \le t)^{d-n}\exp(-nt^2/2),$$

Note that the distribution of $\mathbf{Y}$ is close to being Gaussian, as the determinant has very light tails; this result is due to Goodman [1963], as we formulate below. The distribution of $T_{n,d}$ is close to the average of the top n-entries of an isotropic Gaussian vector in $\mathbb{R}^d$.

Lemma 4 (Lemma 6 in Fleury [2012])**.** *In the notation Lemma 3, it holds*

$$\frac{1}{n}\sum_{i=1}^n \mathbf{Y}_i \sim \mathsf{N}\left(0, \frac{I_{n-1}}{n}\right) \perp \tilde{\mathbf{Y}},$$

where $\tilde{\mathbf{Y}} := (\mathbf{Y}_2 - \mathbf{Y}_1, \ldots, \mathbf{Y}_n - \mathbf{Y}_1)$*.*

In particular, it implies that $\mathbf{c}_{\mathcal{F}} - \mathbf{n}_{\mathcal{F}}$ is independent of $\mathbf{n}_{\mathcal{F}}$ and from $\tilde{\mathbf{Y}}$. Throughout this work, we refer to the distribution of a facet $\mathcal{F}$ as the **"Fleury's distribution"**.

Lemma 5 (Thm. 4 in Fleury [2012])**.** *: The expected number of facets of* $n-1$ *dimensional number of facets of* $P_{n,d}$ *satisfies*

$$\mathbb{E}\mathcal{F}_{n-1}(P_{n,d}) = \exp\left(2^{-1}n\cdot(\mathrm{lln}(d/n) + \Theta(1))\right). \quad (6)$$

Note that this bound should be compared to the loose upper bound on the number of facets, which is

$$2^n\binom{d}{n} = \exp(n\cdot(\log(d/n) + \Theta(1)))$$

which is a significantly better exponent. First, recall the **conic formula**, which implies that for any polytope $P \subset \mathbb{R}^n$, it holds

$$|P| = n^{-1}\sum_{\mathcal{F}\text{ is a facet}} |\mathcal{F}|\cdot\|\mathbf{n}_{\mathcal{F}}\|_2$$

Now, let

$$\mathcal{E} := \{\mathrm{conv}\{X_1,\ldots,X_n\}\text{ is a facet of } P_{n,d}\}$$

and

$$\mathcal{E}_{\mathcal{S},\varepsilon} := \{\mathrm{conv}\{\varepsilon_1 X_{i_1},\ldots,\varepsilon_n X_{i_n}\}\text{ is a facet of } P_{n,d}\}$$

and note that the following holds for $V = \mathbb{E}|P_{n,d}|$ by linearity of expectation and the conic formula:

$$\begin{aligned} V &= n^{-1}\cdot\mathbb{E}\left[\sum_{\boldsymbol{\varepsilon}\in\{-1,1\}^n, |S|=n, S\subset[d]} 1_{\mathcal{E}_{\mathcal{S},\varepsilon}}\cdot|\mathcal{F}|\cdot\|\mathbf{n}_{\mathcal{F}}\|_2\right] \\ &= n^{-1}2^n\binom{d}{n}\Pr(\mathcal{E})\mathbb{E}\left[|\mathcal{F}|\cdot\|\mathbf{n}_{\mathcal{F}}\|_2|\mathcal{E}\right] \\ &= n^{-1}\mathbb{E}|\mathcal{F}_{n-1}(P_{n,d})|\cdot\mathbb{E}|\mathbf{Y}\mathbf{Y}^{\top}|\cdot|\triangle|\cdot\mathbb{E}T_{n,d}\end{aligned}$$

where we used that $T_{n,d}$ is independent from the facet $\mathcal{F}$. Following the same rational, and using the $\mathbf{c}_{\mathcal{F}}$ is independent from the normal $\mathbf{n}_{\mathcal{F}}$ and from $\widetilde{\mathbf{Y}}$, we obtain the following corollary:

Corollary 2. *Let* $\mathcal{E} := A_1 \cap A_2 \cap A_3$ *where* A_1 *be an event on* $T_{n,d}$*, i.e. the height of the normal; and* A_2 *be an event on the span of* $\tilde{\mathbf{Y}}$*, and* A_3 *an event on the* $c_{\mathcal{F}}$*. Then, it equals to*

$$\mathbb{E}|V_{\mathcal{E}}| = \Pr_{T_{n,d}}(A_1)\Pr_{\tilde{\mathbf{Y}}}(A_2)\Pr_{n^{-1}\sum_{i=1}^n \mathbf{Y}_i}(A_3)\cdot\mathbb{E}|P_{n,d}|.$$

where $|V_{\mathcal{E}}|$ *is the volume of the cones that their facets satisfies* $\mathcal{E}$*.*

5.2.2 Reduction to Fluery's distribution

Throughout this work, we denote by

$$C(n,d) = \frac{n}{\sqrt{\log(d/n)}} \text{ and } E(n,d) = \frac{n}{\log(d/n)^2}.$$

Note that the distributions of $T_{n,d}$, and of $\mathbf{Y}$, and of $\mathbf{c}_{\mathcal{F}} - \mathbf{n}_{\mathcal{F}}$ are sub-Gaussian; and therefore have very light tails. However, the volume of the polytope $P_{n,d}$ has a high variance as observed in Paouris et al. [2019] and references within. This is mainly due to the fact that ℓ_1^d has a small Dvoretzky's dimension. Therefore, due to the decomposition of the volume above, the large variance emerges from the number of facets that suffer. In order to overcome this problem, we show that

$$\exp(-C_1 \cdot C(n,d)) \leq \frac{|P_{n,d}|}{\mathbb{E}|P_{n,d}|} \leq \exp(C_1 \cdot C(n,d))$$

with probability of at least $1 - 2\exp(-C(n,d))$ via Dvoresky's theorem and the KLS property of the canonical simplex, see below for the definitions of these objects.

Then, we would condition on events on $T_{n,d}$ and $\mathcal{F}$ and $\mathbf{c}_{\mathcal{F}}$, with respect to Fluery's distribution, that holds with probability of at least $1 - \exp(-C_1 \cdot C(n,d))$. Therefore, by Corollary 2, and the last equation, the total volume of the facets that do not satisfy is at most $\exp(-C(n,d))|P_{n,d}|$, i.e., most of the facets of $P_{n,d}$, in terms of volume, would satisfy the event.

5.2.3 Canonical Simplex satisfies KLS

An isotropic convex body that satisfies the **KLS** with an absolute constant, if for any 1-Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$ and $\varepsilon \geq 0$:

$$\Pr_{X\sim U(K)}(|F(X) - \mathbb{E}F(X)| \geq \varepsilon\sqrt{n}) \leq \exp(-c\sqrt{n}\varepsilon). \tag{7}$$

We refer to the last equation as the (KLS) condition. It is well known that $n \cdot \triangle_c$ is almost isotropic and satisfies the last equation, see Barthe and Wolff [2009]. In particular, KLS implies the following (see also Paouris [2006], Gromov and Milman [1983] for further details)

- **Thin shell:**
$$\Pr_{X\sim U(K)}\left(\left|\|X\|_2 - \sqrt{n}\right| \leq \varepsilon\sqrt{n}\right) \geq 1-\exp(-c\sqrt{n}\varepsilon).$$
- **Isotropic constant:** For any $\theta \in \mathbb{S}^{n-1}$ it holds
$$\Pr_{X\sim U(K)}\left(|\langle\theta, X\rangle| \leq \varepsilon\sqrt{n}\right) \geq 1 - \exp(-c\sqrt{n}\varepsilon).$$

5.2.4 Proof Outline

First, note that:

$$\|\boldsymbol{\xi}\|_n := \|\boldsymbol{\xi}\|_{P_{n,d}} = \|\widehat{w}_n(\mathbf{X},\boldsymbol{\xi})\|_1,$$

where we used that $P_{n,d} = \mathbf{X}B_1^d$. Furthermore, the ℓ_1-LNI has a special interpretation:

$$\widehat{w}_n(\mathbf{X},\boldsymbol{\xi}) = \|\boldsymbol{\xi}\|_{P_{n,d}} \cdot w(\mathbf{X},\boldsymbol{\xi})$$

where $w(\mathbf{X},\boldsymbol{\xi}) \in \mathbb{R}^d$ such that $\|w(\mathbf{X},\boldsymbol{\xi})\|_0 = n$, and $\|w(\mathbf{X},\boldsymbol{\xi})\|_1 = 1$. Namely, it is the weighting of the convex hull of the n-dimensional facet of the B_1^d (that in particular is a simplex) that gives $\boldsymbol{\xi}$ and whose vertices are $\mathcal{F} := \mathrm{conv}\{\boldsymbol{\varepsilon}_i\mathbf{X}_i\}_{i\in S}$.

Intuitively, we believe that a typical solution of $\widehat{w}_n$ behaves as a random element in $\triangle$. To understand how a random element behaves, recall the definition of $\triangle_c$ above, and that $n{\cdot}\triangle_c$ satisfies the KLS property and, in particular, most of its volume lies in its thin shell. Therefore,

$$\mathbb{E}_{Z\sim\mathtt{Unif}(\triangle_c)}\|Z\|_2 = (1 \pm O(1/\sqrt{n})) \cdot n^{-1/2}.$$

Hence, by Pythagoras's law, we would expect that most of the volume of $P_{n,d}$, has an ℓ_2 length that satisfies

$$\sqrt{\|\boldsymbol{\varepsilon}/n\|_2^2 + \left(\frac{1+o(1)}{\sqrt{n}}\right)^2} \approx \sqrt{\frac{2}{n}}.$$

To further support this intuition, we use Fleury's distribution, which implies that a "typical" facet of $P_{n,d}$ is distributed as $\mathcal{F} = \tilde{G}\triangle$, where $\tilde{G}$ is almost distributed as a $n \times (n-1)$ Gaussian matrix, i.e with i.i.d $N(0,1)$ entries. Hence, by ideas that emerge from the restricted isometry property, we expect that $\sqrt{n}\cdot\mathcal{F}$ is almost isotropic and that the volume of the thin shell of $\sqrt{n}\cdot\mathcal{F}$ maps to the volume of the thin shell of $\triangle_c$. Now, to estimate the ℓ_2 norm of the $\widehat{w}_n$, recall that $\widehat{w}_n$ inflates $P_{n,d}$ by $(1+o(1)) \cdot M_{n,d} \gg 1$ with high probability. Via a simple Dvoretsky's and geometric arguments, we would support this claim, and show that

$$M_{n,d} = \mathbb{E}\|\boldsymbol{\xi}\|_n \approx \sqrt{n/(2\log(d/n))}$$

and therefore, we expect that

$$\|\widehat{w}_n\|_2^2 \approx (M_{n,d} \cdot \sqrt{2/n})^2 \approx \log(d/n)^{-1}.$$

Roughly speaking, in all the remaining steps, we show that $\mathbf{X}\widehat{w}_n$ lies in a typical thin-shell of a **"Fleury's"** facet, and that $\widehat{w}_n$ lies in the thin shell of the n-dimensional facets of the ℓ_1^d which are canonical simplices. The main challenge, is that the $\widehat{w}_n$ induces a different distribution on the polytope facets from Fluery's one, as discussed before.

Acknowledgments

Gil Kur conducted part of this work during his visit to the IDEAL Institute, hosted by Lev Reyzin and supported by NSF ECCS-2217023.

Paul Simanjuntak's research is supported by an NSF award DMS-405441.

Reese Pathak gratefully acknowledges support from the NSF under grant DMS-2503579.

We are grateful to Prof. Grigoris Paouris for the insightful discussions, help, and guidance on many references used in this work. We also thank Pierre Bizeul for the useful discussions on small-ball probabilities, and for allowing us to adapt one of his arguments for this paper.

References

David Alonso-Gutiérrez and Joscha Prochno. On the gaussian behavior of marginals and the mean width of random polytopes. *Proceedings of the American Mathematical Society*, 143(2):821–832, 2015.

Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D Milman. *Asymptotic geometric analysis, Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2015.

Franck Barthe and Paweł Wolff. Remarks on non-interacting conservative spin systems: the case of gamma distributions. *Stochastic processes and their applications*, 119(8):2711–2723, 2009.

Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.

Peter L Bartlett, Olivier Bousquet, and Shahar Mendelson. Local Rademacher complexities. *The Annals of Statistics*, 33(4):1497–1537, 2005.

Peter L Bartlett, Philip M Long, Gábor Lugosi, and Alexander Tsigler. Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences*, 117(48):30063–30070, 2020.

Pierre Bizeul. The slicing conjecture via small ball estimates. *arXiv preprint arXiv:2501.06854*, 2025.

Pierre Bizeul and Boaz Klartag. Distances between non-symmetric convex bodies: optimal bounds up to polylog. *arXiv preprint arXiv:2510.20511*, 2025.

Emmanuel J. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9–10):589–592, 2008.

Sourav Chatterjee. A new perspective on least squares under convex constraint. *The Annals of Statistics*, 42(6):2340–2381, 2014.

Niladri S. Chatterji and Philip M. Long. Foolish crowds support benign overfitting. *arXiv preprint arXiv:2110.02941*, 2021.

Geoffrey Chinot, Matthias Löffler, and Sara van de Geer. On the robustness of minimum norm interpolators and regularized empirical risk minimizers. *arXiv preprint arXiv:2012.00807*, 2020.

Konstantin Donhauser, Nicolo Ruggeri, Stefan Stojanovic, and Fanny Yang. Fast rates for noisy interpolation require rethinking the effects of inductive bias. *arXiv preprint arXiv:2203.03597*, 2022.

David Donoho and Jared Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22(1):1–53, 2009.

David L. Donoho and Michael Elad. On the stability of the basis pursuit in the presence of noise. *Signal Processing*, 86(3):511–532, 2006.

David L Donoho and Jared Tanner. Counting the faces of randomly-projected hypercubes and orthants, with applications. *Discrete & computational geometry*, 43(3):522–541, 2010.

Bradley Efron. The convex hull of a random set of points. *Biometrika*, 52(3-4):331–343, 1965.

B Fleury. Poincaré inequality in mean value for gaussian polytopes. *Probability theory and related fields*, 152:141–178, 2012.

Simon Foucart. Stability and robustness of ℓ_1-minimizations with weibull matrices and redundant dictionaries. *Linear Algebra and its Applications*, 441:4–21, 2014.

Behrooz Ghorbani, Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Linearized two-layers neural networks in high dimension. *The Annals of Statistics*, 49(2):1029 – 1054, 2021.

Apostolos A Giannopoulos and Vitali D Milman. Extremal problems and isotropic positions of convex bodies. *Israel Journal of Mathematics*, 117:29–60, 2000.

NR Goodman. The distribution of the determinant of a complex wishart distributed matrix. *The Annals of mathematical statistics*, 34(1):178–180, 1963.

Y. Gordon, A. E. Litvak, S. Mendelson, and A. Pajor. Gaussian averages of interpolated bodies and applications to approximate reconstruction. *J. Approx. Theory*, 149(1):59–73, 2007. ISSN 0021-9045,1096-0430. doi: 10.1016/j.jat.2007.04.007. URL https://doi.org/10.1016/j.jat.2007.04.007.

Mikhael Gromov and Vitali D Milman. A topological application of the isoperimetric inequality. *American Journal of Mathematics*, 105(4):843–854, 1983.

Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. Characterizing implicit bias in terms of optimization geometry. In *International Conference on Machine Learning*, pages 1832–1841. PMLR, 2018.

Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J Tibshirani. Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, 50(2):949–986, 2022.

Peizhong Ju, Xiaojun Lin, and Jia Liu. Overfitting can be harmless for basis pursuit, but only to a degree. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Boaz Klartag and Joseph Lehec. Affirmative resolution of bourgain's slicing problem using guan's bound. *Geometric and Functional Analysis*, pages 1–22, 2025.

Bo'az Klartag and Roman Vershynin. Small ball probability and dvoretzky's theorem. *Israel Journal of Mathematics*, 157(1):193–207, 2007.

Bo'az Klartag and Emanuel Milman. On volume distribution in 2-convex bodies. *Israel Journal of Mathematics*, 164:221–249, 2008.

Frederic Koehler, Lijia Zhou, Danica J Sutherland, and Nathan Srebro. Uniform convergence of interpolators: Gaussian width, norm bounds and benign overfitting. *Advances in Neural Information Processing Systems*, 34:20657–20668, 2021.

Felix Krahmer, Christian Kümmerle, and Holger Rauhut. A quotient property for RIP matrices with heavy-tailed entries and its application to noise-blind compressed sensing. *arXiv preprint arXiv:1806.04261*, 2018.

Gil Kur and Pierre Bizeul. Minimum norm interpolation via the local theory of banach spaces: The role of 2-uniform convexity. *arXiv preprint arXiv:2603.28956*, 2026.

Gil Kur, Pedro Abdalla, Pierre Bizeul, and Fanny Yang. Minimum norm interpolation meets the local theory of banach spaces. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 25726–25754. PMLR, 21–27 Jul 2024.

Guillaume Lecué and Zong Shang. A geometrical viewpoint on the benign overfitting property of the minimum ℓ_2-norm interpolant estimator. *arXiv preprint arXiv:2203.05873*, 2022.

Song Mei and Andrea Montanari. The generalization error of random features regression: Precise asymptotics and the double descent curve. *Communications on Pure and Applied Mathematics*, 75 (4):667–766, 2022.

Emanuel Milman. On the mean-width of isotropic convex bodies and their associated l p-centroid bodies. *International Mathematics Research Notices*, 2015(11):3408–3423, 2015.

Vitali D. Milman. Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espaces normés. *C. R. Acad. Sci. Paris Sér. I Math.*, 302 (1):25–28, 1986. ISSN 0249-6291.

Vidya Muthukumar, Kailas Vodrahalli, Vignesh Subramanian, and Anant Sahai. Harmless interpolation of noisy data in regression. *IEEE Journal on Selected Areas in Information Theory*, 1(1):67–83, 2020a.

Vidya Muthukumar, Kailas Vodrahalli, Vignesh Subramanian, and Anant Sahai. Harmless interpolation of noisy data in regression. *IEEE Journal on Selected Areas in Information Theory*, 1(1):67–83, 2020b.

Preetum Nakkiran, Gal Kaplun, Yamini Bansal, Tristan Yang, Boaz Barak, and Ilya Sutskever. Deep double descent: Where bigger models and more data hurt. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12):124003, 2021.

Eduard Oravkin and Patrick Rebeschini. On optimal interpolation in linear regression. *Advances in Neural Information Processing Systems*, 34:29116–29128, 2021.

G Paouris, P Pivovarov, and P Valettas. Gaussian convex bodies: a nonasymptotic approach. *Journal of Mathematical Sciences*, 238(4):537–559, 2019.

Grigoris Paouris. -estimates for linear functionals on zonoids. In *Geometric Aspects of Functional Analysis: Israel Seminar 2001-2002*, pages 211–222. Springer, 2004.

Grigoris Paouris. Concentration of mass on convex bodies. *Geometric & Functional Analysis GAFA*, 16(5):1021–1049, 2006.

Grigoris Paouris and Petros Valettas. A gaussian small deviation inequality for convex functions. *arXiv preprint arXiv:1611.01723*, 2016.

Grigoris Paouris, Petros Valettas, and Joel Zinn. Random version of dvoretzky's theorem in ℓ_p. *Stochastic Processes and their Applications*, 127(10):3187–3227, 2017.

Gilles Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989. ISBN 0-521-36465-5; 0-521-66635-X. doi: 10.1017/CBO9780511662454. URL https://doi.org/10.1017/CBO9780511662454.

Ohad Shamir. The implicit bias of benign overfitting. In *Conference on Learning Theory*, pages 448–478. PMLR, 2022.

Christos Thrampoulidis, Samet Oymak, and Babak Hassibi. Regularized linear regression: A precise analysis of the estimation error. In *Conference on Learning Theory*, pages 1683–1709. PMLR, 2015.

Alexander Tsigler and Peter L Bartlett. Benign overfitting in ridge regression. *Journal of Machine Learning Research*, 24(123):1–76, 2023.

Guillaume Wang, Konstantin Donhauser, and Fanny Yang. Tight bounds for minimum in lone-norm interpolation of noisy data. In *International Conference on Artificial Intelligence and Statistics*, pages 10572–10602. PMLR, 2022.

P. Wojtaszczyk. Stability and instance optimality for gaussian measurements in compressed sensing. *Foundations of Computational Mathematics*, 10(1):1–13, 2010.

Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

Lijia Zhou, Zhen Dai, Frederic Koehler, and Nathan Srebro. Uniform convergence with square-root lipschitz loss. *arXiv preprint arXiv:2306.13188*, 2023.

Checklist

The checklist follows the references. For each question, choose your answer from the three possible options: Yes, No, Not Applicable. You are encouraged to include a justification to your answer, either by referencing the appropriate section of your paper or providing a brief inline description (1-2 sentences). Please do not modify the questions. Note that the Checklist section does not count towards the page limit. Not including the checklist in the first submission won't result in desk rejection, although in such case we will ask you to upload it during the author response period and include it in camera ready (if accepted).

In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Yes
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. Not Applicable
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. Yes
   (b) Complete proofs of all theoretical results. Yes
   (c) Clear explanations of any assumptions. Yes
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes/No/Not Applicable]
   (b) All the training details (e.g., data splits, hy-

perparameters, how they were chosen).Not Applicable

(c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). Not Applicable

(d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). Not Applicable

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. Not Applicable

   (b) The license information of the assets, if applicable. [Yes/No/Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. Not Applicable

   (d) Information about consent from data providers/curators. Not Applicable

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not Applicable

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. Not Applicable

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not Applicable

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable

A Proof of Theorem 2

A.1 Remaining Preliminaries

We state a lemma that follows from Paouris and Valettas [2016], see also Klartag and Vershynin [2007]. Let $r(A), R(A)$ are the inradius and outradius of a set A. For a matrix A, $\|A\|$ denotes its operator norm. We also use the notation of $\bar{B}_n := \sqrt{n} \cdot B_n$.

Lemma 6 (Boosted Dvoretzky's theorem)**.** *Assume that $d \geq C_1 n$. Then, the following holds with probability of $1 - \exp(-c_1 n)$:*

$$M^*(K)\left(1 - c_2\sqrt{n \operatorname{Var}\left(\frac{\|\boldsymbol{\xi}\|_{K^\circ}}{M^*(K)}\right)}\right) \leq r(\mathbf{X}K) \leq R(\mathbf{X}K) \leq M^*(K)\cdot\left(1 + c_3\frac{R(\sqrt{n}\cdot K)}{M^*(K)}\right), \tag{8}$$

for any convex body $K \subset \mathbb{R}^n$ and universal constants $c_1, c_2, c_3 > 0$.

Note that since $\|\boldsymbol{\xi}\|_{K^\circ}$ is $R(K)$-Lipschitz, the Gaussian Poincaré inequality shows the deviation from $M^\star(K)$ in the upper inclusion is larger than the deviation in the lower inclusion implied by (8). Note that the density of $\mathbf{Y}$ is "almost" a Gaussian matrix up a $|\mathbf{Y}^\top\mathbf{Y}| \cdot |\triangle|$. For this, we use a result of Goodman [1963], showing that

$$\log|\mathbf{Y}^\top\mathbf{Y}| = \sum_{i=2}^{n}\log Z_i \text{ where are independent } Z_i \sim \chi_i^2.$$

The following result follows by computing the moment generating function of the corresponding χ^2 random variables and a Chernoff bound.

Lemma 7. *Let G be an $(n-1)\times n$ Gaussian matrix. Set $S_n := \log|G|$. There is a universal constant $c > 0$ sufficiently large such that the following inequalities hold for $t \geq 0$.*

1. *For the upper tail:*

$$\Pr(S_n - \mathbb{E}S_n \geq t) \leq \exp\left(-\frac{t^2}{c\log n}\right).$$

2. *For the lower tail:*

$$\Pr(S_n - \mathbb{E}S_n \leq -t) \leq \begin{cases}\exp\left(-\frac{t^2}{c\log n}\right) & t < \log n\\ cne^{-t/c} & t \geq \log n.\end{cases}$$

Hence, using the last lemma and the upper bound on the number of facets, we obtain the following one-tailed estimate:

Corollary 3. *With high probability over $P_{n,d}$ it holds for all facets simultaneously*

$$\forall \mathcal{F} \in \mathcal{F}_{n-1}(P_{n,d}) \quad \log|\mathbf{X}_{\mathcal{F}}| - \mathbb{E}S_n \lesssim \sqrt{n\cdot \operatorname{lln}(d/n)\cdot\log(n)}.$$

The next corollary follows from tail integration of the determinant that our proof relies on the relation between G, the Gaussian matrix, and $\mathbf{Y}$ from Fluery's distribution, which is almost Gaussian.

Corollary 4. *Let A be an event G is smaller than n^{-100}. Then, it is smaller than n^{-99} over the probability space of $\mathbf{Y}$.*

In words, it means a rare event in G is also rare in $\mathbf{Y}$, however, it does not mean that those densities are close, for example, in TV-distance. We finally state a lemma regarding the distribution of $T_{n,d}$. To state it, denote by

$$(t^\star_{n,d})^2 = 2\log(d/n) - \operatorname{lln}(d/n) - \ln\pi + o(1)$$

denote the mode of the density associated with the random variable $T_{n,d}$.

Lemma 8. *Let $0 \leq \varepsilon \leq c/\log(d/n)$ and assume that $d \geq 4n \geq \sqrt{\log(d/n)}$. Then,*

$$\Pr\left\{|T_{n,d} - \operatorname{Med}T_{n,d}| \geq \varepsilon\cdot t^\star_{n,d}\right\} \leq \exp\left\{-cn\log(d/n)^2\varepsilon^2\right\},$$

and in particular

$$|t^\star_{n,d} - \operatorname{Med}T_{n,d}| \lesssim \frac{1}{\sqrt{n\log(d/n)}}$$

This proof of the last lemma follows from a direct computation on the distribution of Lemma 3. Next, we would need the following lemma:

Lemma 9 (Lemma 2.3 Paouris [2004])**.** *Let $K \subset \mathbb{R}^n$ be an isotropic convex body with ℓ_2-diameter $C\sqrt{n}$. Then, with probability $1-\exp(-c_1t^2)$ over $\theta \sim U(\mathbb{S}^{n-1})$, the marginal $\langle X, \theta\rangle$ satisfies*

$$\Pr_{X\sim U(K)}\left\{|\langle X,\theta\rangle| \geq t\right\} \leq 2\exp(-c_2t^2),$$

where the constants $c_1, c_2, c_3 > 0$ depend on $C \geq 0$.

In words, $\langle X, \theta\rangle$ has a sub-Gaussian tail with parameter $O(1/\sqrt{c_2})$.

A.1.1 Step I: Global Properties of $P_{n,d}$

First, we now provide a cheap estimate of the volume of the polytope $P_{n,d}$, and also on the mean norm $M_{n,d}$, which is sufficient to carry out the argument for Theorem 2. We consider the truncated polytope and corresponding Gaussian mean width,

$$K(n,d) := B_1^d \cap \tfrac{1}{n}B_\infty^d, \quad \text{and} \quad \widetilde{M}_{n,d} := M^*(K(n,d)).$$

We use the shorthand notation $\bar{P}_{n,d} = M_{n,d}P_{n,d}$.

Lemma 10. *Suppose that $d \gtrsim n$. Then, with probability at least $1-\exp(-cn)$ (over $\mathbf{X}$), we have:*

(i) It holds that

$$\left(1-\frac{C}{\log(d/n)}\right)\cdot \widetilde{M}_{n,d}\cdot B_n \subset P_{n,d}$$

and in particular $M_{n,d} \le (1+\frac{C}{\log(d/n)})\cdot \widetilde{M}_{n,d}^{-1}$.

(ii) the volume satisfies the bounds

$$|P_{n,d}| \le \exp(C(n,d))|\widetilde{M}_{n,d}\cdot B_n| \quad \text{and} \quad \exp(-C_1\cdot C(n,d)) \le \frac{|\bar{P}_{n,d}|}{|\bar{B}_n|} \le \exp(C_1\cdot C(n,d)),$$

and consequently $M_{n,d} \ge (1-\frac{c}{\sqrt{\log(d/n)}})\cdot \widetilde{M}_{n,d}^{-1}$.

The proof of this lemma appears below. Also, note that

$$\left|\frac{\|\mathbf{n}_{\mathcal{F}}\|_2}{\widetilde{M}_{n,d}}-1\right| \lesssim \frac{1}{\sqrt{\log(d/n)}},$$

and same bounds holds for $\mathbb{E}|P_{n,d}|$, i.e.

$$e^{\frac{-Cn}{\log(d/n)}}|\widetilde{M}_{n,d}\cdot B_n| \le \mathbb{E}|P_{n,d}| \le \exp(C(n,d))|\widetilde{M}_{n,d}\cdot B_n| \tag{9}$$

via integrating the tail of $R(P_{n,d})$ with the tail bound of Dvoretzky's theorem.

Here, we provide its sketch: We apply Dvoretzky's theorem on the set $K(n,d)\subset B_1^d$, from which we can obtain a lower bound on the inradius of the projected polytope $P_{n,d}$, which in turn bounds from above the expected norm of the projected body, $M_{n,d}$.

A.1.2 Step II: Volume reductions

Here, we state few volume reductions techniques that we use throughout our proof. To see this, note that for every fixed Im$\mathbf{X}$, the map

$$\boldsymbol{\xi} \mapsto \frac{\|\boldsymbol{\xi}\|_n}{\mathbb{E}\|\boldsymbol{\xi}\|_n}$$

is $O(\mathbb{E}\|\boldsymbol{\xi}\|_n/\sqrt{n})$-Lipschitz by the ball inclusion under the event of the previous step. Therefore, we can assume that

$$\mathbb{E}|\bar{P}_{n,d}\cap \bar{B}_n| \ge \exp\left(-C\sqrt{n}\right)|\bar{B}_n|,$$

Remark 5. In the extended version via super concentration methods, we would show that

$$\mathbb{E}|\bar{P}_{n,d}\cap \bar{B}_n| \ge \exp\left(-\frac{C\sqrt{n}}{\log(d/n)}\right)|\bar{B}_n|$$

◁

Let A be an event on a facet $\mathcal{F}$ such that under Fleury's distribution, it holds with probability of at most $\exp(-CC(n,d))$. Then, we know by Markov's inequality, that

$$\left|\bigcup_{\mathcal{F}\text{ satisfies } A} M_{n,d}\cdot \mathcal{F}\right| \le \exp(-C_3C(n,d))|\partial \bar{B}_n|,$$

where we used Lemma 7 and its corollaries.

This means that for a set of directions $B\subset \sqrt{n}\cdot S^{n-1}$ lies in the cones whose their facets satisfy the event A, then it means that we have to inflate $P_{n,d}$ by at least $1+1/\sqrt{C/\log(d/n)}$ to contain 0.5 of the measure of B. To see this monotonicity of the volume, and its n-homogeneity, means that we need to inflate by x such that

$$(1+x)^n \ge \exp(CC(n,d))$$

which is equal to $x \ge C\log(d/n)^{-1/2}$, However, this would contradict the ball inclusion and estimates on $M_{n,d}$ from the previous step, as $\bar{B}_n \subset (1+\frac{C_1}{\sqrt{\log(d/n)}})\bar{P}_{n,d}$.

Therefore, we obtained the following corollary:

Corollary 5. *Let A be an event on a facet $\mathcal{F}$ such that under Fleury's distribution, it holds with probability of at most $\exp(-CC(n,d))$. Then, with probability of at least $1-\exp(-C_2C(n,d))$, $\widehat{w}_n$ would lie in a facet that satisfy A^c.*

To prove the lower bound on the MSE, we show that

$$B^-_{n,d}(M_{n,d}) := M_{n,d}\cdot \bigcup_{\triangle\text{ is } n\text{-dimensional facet of } B_1^d} \underbrace{\frac{\boldsymbol{\varepsilon}}{n}+\triangle_c\cap\Bigg(\frac{1-\frac{C}{\log(d/n}}{\sqrt{n}}}_{:=\triangle_-}$$

(where we view $\triangle_c$ as a set in $\mathbb{R}^{n-1}$ and B_{n-1} is the ℓ_2-ball with radius one in $\mathbb{R}^{n-1}$) does not have enough volume, i.e., we show that

$$|\mathbf{X}B^-_{n,d}(M_{n,d})| \le \exp(-c_1C\cdot C(n,d))|\bar{P}_{n,d}| \le \exp(-c_1C_2\cdot C(n,d))|\bar{B}_n$$

which means that with probability of at least $1-\exp(-cC(n,d))$

$$\|\widehat{w}_n\|_2^2 \ge \frac{1-\frac{C}{\log(d/n)^{1/4}}}{\log(d/n)}.$$

To prove the upper bound, we would need to adopt a more refined approach, as will become clearer below. Very roughly speaking, we would show that

$$B^+_{n,d}(M_{n,d}) := M_{n,d}\cdot \bigcup_{\triangle\text{ is } n\text{-dimensional facet of } B_1^d} \underbrace{\frac{\boldsymbol{\varepsilon}}{n}+\triangle_c\cap\Bigg(\frac{1+\frac{C}{\log(d/n}}{\sqrt{n}}}_{:=\triangle_+}$$

$$|\mathbf{X}B^-_{n,d}(M_{n,d})\cap\bar{B}_n| \le \exp(-c_1 C\cdot C(n,d))|\bar{P}_{n,d}| \le \exp(-c_1 C_2 C(n,d))|\bar{B}_n|$$

and as the $\|\boldsymbol{\xi}\|_2 = \sqrt{n}$, provides the desired bound with the same probability. Yet, this argument is much more delicate, as will be seen below.

A.1.3 Step III: Canonical Simplex

First, we introduce the following useful lemma:

Lemma 11 (Small ball probabilities for a simplex)**.** *The following holds for all $\varepsilon \in (0,1)$:*

$$\Pr_{Z\sim U(\triangle_c)}(\|Z\|_2 \le (1-\varepsilon)\mathbb{E}\|Z\|_2) \le \exp(-cn\varepsilon^2),$$

and also the converse holds, i.e., for $\varepsilon \in (0,c)$

$$\Pr_{Z\sim U(\triangle_c)}(\|Z\|_2 \le (1-\varepsilon)\mathbb{E}\|Z\|_2) \ge \exp(-c_2 n\varepsilon^2).$$

The proof of this lemma follows from standard calculations of MGFs, and noticing that $Z \sim \mathrm{Unif}(\triangle)$, can also be represented as

$$Z \sim \frac{(Z_1,\ldots,Z_n)}{\sum_{i=1}^n Z_i}$$

where $Z_1,\ldots,Z_n \sim \mathrm{Exp}(1)$, for completeness it appears below. Surprisingly, it does not appear in the literature. Now, we prove the easy part of the theorem, on the lower bound on the MSE

Lemma 12 (Lower bound on the MSE)**.** *With probability of at least $1 - C_1\exp(C(n,d))$, it holds*

$$\|\widehat{w}_n\|_2^2 \ge \frac{1-\frac{C}{\log(d/n)^{1/4}}}{\log(d/n)}.$$

Proof. Recall that we need to prove that $\widehat{w}_n \notin B^-_{n,d}(M_{n,d})$ as discussed in Section A.1.2, by showing

$$\exp(-c\sqrt{n})\cdot|\partial\bar{B}_n| \le |\partial\bar{P}_{n,d}| \le \exp(c\cdot C(n,d))|\partial\bar{B}_n|.$$

By choosing $\varepsilon \ge C_3\cdot\log(d/n)^{-1/4}$ for some $C_3 \ge 0$ in Lemma 11, we have that

$$|\triangle_-| \le e^{-C\cdot C(n,d)}|\triangle|$$

for some large enough $C \ge 0$. Now, by Lemma 7, it holds

$$\forall\mathcal{F}\in\mathcal{F}_{n-1}(P_{n,d}) \quad |\mathbf{X}_S\triangle_-| \le \exp(-c_1\cdot C(n,d))\cdot\mathbb{E}|\mathcal{F}|,$$

Therefore, we conclude that

$$\left|\mathbf{X}B^-_{n,d}(M_{n,d})\right| \le \exp(-C\cdot C(n,d))|\partial\bar{P}_{n,d}| \le \exp(-C_2\cdot C(n,d))\cdot|\partial\bar{B}_n| \tag{10}$$

where we used the fact that we can set $C \ge 0$ to be large enough. Hence, by Section A.1.2, the claim follows.

□

To see why this approach cannot work for the upper bound side, as one can easily show that (cf. Paouris [2006]) that

$$\Pr_{Z\sim U(\triangle_c)}(\|Z\|_2 \ge (1+\varepsilon)\mathbb{E}\|Z\|_2) \ge \exp(-c\sqrt{n}\varepsilon),$$

Namely, the tails of $\|Z\|_2$ exhibit a different behavior; one is sub-Gaussian, and the other is sub-exponential. Meaning that we would need to use a different argument, as

$$\left|\mathbf{X}B^+_{n,d}(M_{n,d})\right| \ge e^{-c\sqrt{\frac{n}{\log(d/n)}}}|\partial\bar{P}_{n,d}|, \tag{11}$$

i.e., the "outer part" has too much volume, as we cannot infer that

$$e^{-c\sqrt{\frac{n}{\log(d/n)}}}|\partial\bar{P}_{n,d}| \ll |\bar{B}_n|.$$

A.1.4 Step III: Reduction to the thin shell of $\mathcal{F}$

The following follows from the Fubini theorem and Lipschitz concentration:

Lemma 13. *Let G be a $(n-1)\times n$ Gaussian matrix with $N(0,1)$ i.i.d. entries, and let $\delta\in(0,1)$. Then, with probability* ***over*** *G of at least $1-\exp(-c_1 n\delta^2)$, the following determinsic equation holds:*

$$\Pr_{X\sim U(\triangle_c)}(\|GX\|_2 \le 1+\delta/2 \cap \|\sqrt{n}X\|_2 \ge 1+\delta) \lesssim \exp(-cn\delta^2).$$

The proof of this lemma appears below. This lemma means that most of the thin shell of the $\mathcal{F}$ will emerge from thin shell $\triangle_c$, and **only** $\exp(-cn\delta^2)$ of the thin shell volume emerges from the areas that are δ-far from the thin shell of $\triangle$. Namely, we have the desired sub-Gaussian tail, and note that by Corollary 4, this holds for the distribution of Fluery's facet $\tilde{\mathbf{Y}}$.

In the next steps, we prove the following:

Lemma 14. *With probability of at least of $1-\exp(-C\cdot E(n,d))$ over $\boldsymbol{\xi}$,*

$$\|\boldsymbol{\xi} - \|\boldsymbol{\xi}\|_n \boldsymbol{c}_{\mathcal{F}}\|_2 = (1+O(\log(d/n)^{-1}))\cdot M_{n,d}$$

where $E(n,d) = n/\log(d/n)^2$, and $\|\boldsymbol{\xi}\|_n\cdot\mathcal{F}$ contains $\boldsymbol{\xi}$.

Therefore, the theorem follows, by Lemma 13, it holds that with probability of $1-\exp(-c_2 E(n,d))$

that

$$\|\widehat{w}_n-(1+O(1/\log(d/n))\cdot\frac{\|\boldsymbol{\xi}\|_n\mathbf{1}_n}{n}\|_2^2 \leq \frac{\|\boldsymbol{\xi}\|_n^2}{n} = \frac{(1+o(1))}{2\log(d/n)}$$

And by Pythagoras, the claim follows, as

$$\widehat{w}_n - \|\boldsymbol{\xi}\|_n\mathbf{1}_n/n \perp \|\boldsymbol{\xi}\|_n\mathbf{1}_n/n$$

and $\|\boldsymbol{\xi}\|_n \approx M_{n,d}$, and therefore $\|\|\boldsymbol{\xi}\|_n\mathbf{1}_n/n\|_2^2 = (1+o(1))/2\log(d/n)$.

A.1.5 Step IV: On the height of the facets of $P_{n,d}$

In this part, we will need to study the heights of the facets of the random polytope $P_{n,d}$. We remind that $T_{n,d}$ is the distribution $\|\mathbf{n}_{\mathcal{F}}\|_2$, Recall that when a facet $\mathcal{F}$ is drawn from Fleury's distribution (see Lemma 3) it satisfies the following with probability of $1-\exp(-cn\varepsilon^2)$:

$$\|\boldsymbol{c}_{\mathcal{F}} - \boldsymbol{n}_{\mathcal{F}}\|_2 \lesssim \varepsilon \tag{12}$$

Hence, with probability of at least $1-\exp(C\cdot C(n,d))$, it holds

$$\|\boldsymbol{c}_{\mathcal{F}} - \boldsymbol{n}_{\mathcal{F}}\|_2 \lesssim \log(d/n)^{-1/4} \tag{13}$$

We denote this event by $\mathcal{E}_1$.

Consider the linear functional defined via $\langle\theta,\cdot\rangle$, where $\theta := \frac{\mathbf{c}_{\mathcal{F}}-\mathbf{n}_{\mathcal{F}}}{\|\mathbf{c}_{\mathcal{F}}-\mathbf{n}_{\mathcal{F}}\|_2}$, and the isotropic constant of $\mathcal{F}$ implies that most of the volume (surface area) of $\mathcal{F}$ lies in a "thin slab" in the direction of θ. Formally, there exists and event $\mathcal{E}_3$ with probability of $1-\exp(Cn\mathrm{lln}(d/n))$, such that for all facets, it holds

$$\Pr_{Z\sim U(\mathcal{F})}(|\langle Z,\theta\rangle| \lesssim \sqrt{\frac{\mathrm{lln}(d/n)}{n}}) \geq 0.9 \tag{14}$$

and we also recall the thin-shell property

$$\Pr_{Z\sim U(\mathcal{F})}(\|Z\|_2 \in (1-C_1\sqrt{\frac{\mathrm{lln}(d/n)}{n}}, 1+C_1\sqrt{\frac{\mathrm{lln}(d/n)}{n}})) \geq 0.9. \tag{15}$$

Under the event of the last three equations (with $\varepsilon = O(1/\sqrt{n})$) and Lemma 8, we apply Pythagoras's law, and obtain that

$$\Pr_{Z\sim U(\mathcal{F})}(\|z\|_2^2 \in (\mathrm{Med}\, T_{n,d}^2 + 2 + \tilde{O}(1/\sqrt{n})) \geq 0.9$$

As by scaling by $M_{n,d}$, we know that

$$\mathrm{Med}\,\bar{T}_{n,d}^2 + 2M_{n,d}^2 - n \gtrsim -\sqrt{n},$$

where $\bar{T}_{n,d} = M_{n,d}T_{n,d}$. It which follows from the fact that

$$|\bar{P}_{n,d}| \geq \exp(-c\sqrt{n})|\bar{B}_n|$$

Now, we show that

$$\mathrm{Med}\,\bar{T}_{n,d}^2+2M_{n,d}^2-n \lesssim \log(d/n)^{-5/4}\mathrm{Med}\,\bar{T}_{n,d}^2 \asymp \log(d/n)^{-5/4}\cdot n$$

To see this, Apply Lemma 8 with $\delta(n,d) = C_1\log(d/n)^{-5/4}$ and combine it with Corollary 3, that the total of volume of cones that their facets would have normals lower than $\delta(n,d)\cdot M_{n,d}$, is at most

$$\exp(-C_2\cdot C(n,d))\cdot\mathbb{E}|\bar{P}_{n,d}| \lesssim \exp(-C_1\cdot C(n,d))|\bar{B}_n|$$

which contradict our argument in Section A.1.2, as it contradicting the definition of $M_{n,d}$. Now, for $Z\sim U(\bar{P}_{n,d})$, with probability of 0.9, it holds

$$\|Z\|_2 \leq (1+\log(d/n)^{-5/4})\cdot\sqrt{n}$$

by using the conic formula, we conclude that

$$|\bar{P}_{n,d}| \lesssim \exp(n\cdot\delta_{n,d})|\bar{B}_n|,$$

where $\delta_{n,d} \lesssim \log(d/n)^{-5/4}$ with high probability (and the same holds in expectation).

Step IV+: Improving our bounds

Note that we may use the previous step, with the **improved** estimate of

$$|\bar{P}_{n,d}| \lesssim \exp(Cn\cdot\log(d/n)^{-5/4})|\bar{B}_n|$$

rather than the one of

$$|\bar{P}_{n,d}| \lesssim \exp(CC(n,d))|\bar{B}_n|$$

and reiterate. It is not hard to verify that it would give the following estimate:

Lemma 15.

$$\exp(-C\sqrt{n})\cdot|\bar{B}_n| \leq \mathbb{E}|\bar{P}_{n,d}| \lesssim \exp(C\cdot E(n,d))|\bar{B}_n|$$

where $E(n,d) \asymp n/\log(d/n)^2$ as defined above.

which significantly improved Dvoretzky's estimate by a factor of $\log(d/n)^{3/2}$. Note that this self-improvement argument would stop when

$$n\varepsilon \asymp n\log(d/n)^2\varepsilon^2$$

where the left-hand side comes from our volume radius estimate, and right hand side emerges from Lemma 8. Meaning that

$$\varepsilon \asymp \log(d/n)^{-2},$$

and conclude that

$$\mathbb{E}|\bar{P}_{n,d}| \lesssim \exp(Cn/\log(d/n)^2)|\bar{B}_n|.$$

Remark 6. In our extended version of this manuscript, we would discuss whether this bound can be improved further. ◁

Another observation that this argument implies a sharp esimate of

$$M_{n,d} = \big(1 + O(\log(d/n)^{-2})\big) \cdot \sqrt{\frac{n}{(t^*_{n,d})^2 + 2}}$$

where $t^*_{n,d}$ is the mode of $T_{n,d}$. Now, as

$$(t^*_{n,d})^2 = \sqrt{2\log(d/n)} - \frac{\operatorname{lln}(d/n) + \ln\pi}{2\sqrt{2\log(d/n)}} + o\left(\frac{\operatorname{lln}(d/n) + \ln\pi}{2\sqrt{2\log(d/n)}}\right)$$

by taking square

$$(t^*_{n,d})^2 = 2\log(d/n) - \operatorname{lln}(d/n) + \ln\pi + o(1)$$

meaning

$$M_{n,d} = (1 + O(\log(d/n)^{-2})) \cdot \sqrt{\frac{n}{2\log(d/n) - \operatorname{lln}(d/n) - \ln\pi + 2 + o(1)}}$$

A.1.6 Step V: Upper bound on the MSE

We denote by

$$\bar{\boldsymbol{c}}_{\mathcal{F}} = M_{n,d}\boldsymbol{c}_{\mathcal{F}} \text{ and } \bar{\boldsymbol{n}}_{\mathcal{F}} = M_{n,d}\boldsymbol{n}_{\mathcal{F}} \text{ and } \bar{\mathcal{F}} = M_{n,d}\mathcal{F}.$$

Let us combine what we have so far, first we can assume that the normal of facets $\mathcal{F}$ that the $\widehat{w}_n$ lie at (with probability of $1 - \exp(-cE(n,d))$ satisfies

$$\|\bar{\boldsymbol{n}}_{\mathcal{F}}\|_2 - \operatorname{Med}\bar{T}_{n,d} \lesssim \log(d/n)^{-2} \cdot \sqrt{n}$$

Now, using Fluery's distribution, we know that

$$\|\bar{\boldsymbol{n}}_{\mathcal{F}} - \bar{\boldsymbol{c}}_{\mathcal{F}}\|_2 \lesssim (1 + O(\log(d/n)^{-1})) \cdot M_{n,d}$$

Now, as $\|\boldsymbol{\xi}\|_2 = \sqrt{n}$, we know that

$$\|\boldsymbol{\xi} - \bar{\boldsymbol{c}}_{\mathcal{F}}\|_2 \le (1 + O(\log(d/n)^{-1})) \cdot M_{n,d}$$

Now, we can apply Lemma 9 with $K = \mathcal{F} \cap C_1 B_{n-1}$ and $\varepsilon \asymp \log(d/n)^{-1}$, and it holds that

$$\Pr_{X \sim U(\bar{\mathcal{F}})}\left\{|\langle X, \theta\rangle| \le \frac{M_{n,d}}{\log(d/n)}\right\} \ge 1 - \exp(-c_2 E(n,d))$$

where $\theta = \frac{\boldsymbol{c}_{\mathcal{F}} - \boldsymbol{n}_{\mathcal{F}}}{\|\boldsymbol{c}_{\mathcal{F}} - \boldsymbol{n}_{\mathcal{F}}\|_2}$. Note that Flurey's distribution implies that θ is drawn uniformly from the sphere. We conclude that

$$\|\boldsymbol{\xi} - \bar{\boldsymbol{c}}_{\mathcal{F}}\|_2 = (1 + O(\log(d/n)^{-2}))\sqrt{n}$$

with probability of at least $1 - \exp(-cE(n,d))$. And as explained in Section A.1.4, the claim follows, as in those facets

$$\|\boldsymbol{\xi} - \|\boldsymbol{\xi}\|_n \boldsymbol{c}_{\mathcal{F}}\|_2 = (1 + O(\log(d/n)^{-1})) \cdot M_{n,d}$$

B Missing parts

Proof of Lemma 11. Let $Z_1, \dots, Z_n \overset{\text{i.i.d.}}{\sim} \operatorname{Exp}(1)$, $S = \sum_{i=1}^n Z_i$, and define

$$X = \frac{Z}{S} \in \triangle, \qquad Y = \sqrt{n(n+1)}\Big(X - \tfrac{\mathbf{1}}{n}\Big).$$

Then X is uniform on the simplex, and the rescaled Y is isotropic in $\mathbb{R}^n$:

$$\mathbb{E}Y = 0, \qquad \mathbb{E}\|Y\|_2^2 = n - 1.$$

We want to bound, for $\delta \in (0,1)$,

$$p(\delta) = \Pr\big(\|Y\|_2^2 \le (1-\delta)(n-1)\big).$$

We have

$$\|Y\|_2^2 = \frac{n+1}{\overline{Z}^2}\left(\frac{1}{n}\sum_{i=1}^n Z_i^2 - \overline{Z}^2\right), \qquad \overline{Z} = \frac{S}{n}.$$

Equivalently,

$$\|Y\|_2^2 = (n+1)\left(\frac{U}{\overline{Z}^2} - 1\right), \quad \text{where} \quad U = \frac{1}{n}\sum_{i=1}^n Z_i^2.$$

Thus, for $\delta \in (0,1)$,

$$\begin{aligned}\Pr\big(\|Y\|_2^2 \le (1-\delta)(n-1)\big) &= \Pr\left(\frac{U}{\overline{Z}^2} \le 1 + \frac{(1-\delta)(n-1)}{n+1}\right) \\ &\le \Pr\left(\frac{U}{\overline{Z}^2} \le 2 - \delta\right).\end{aligned}$$

We control the numerator and denominator separately. Set the event

$$F := \{\overline{Z} \le 1 + \delta/8\}.$$

On F we have

$$\Pr\left(\frac{U}{\overline{Z}^2} \le 2 - \delta \,\middle|\, F\right) \le \Pr\left(U \le \frac{(1+\delta/8)^2}{2-\delta}\right) \le \Pr\left(U \le 2 - \frac{\delta}{4}\right).$$

Thus we conclude that

$$p(\delta) \le \Pr(F^c) + \Pr\left(U \le 2 - \frac{\delta}{4}\right). \tag{16}$$

Bounding $\Pr(F^c)$. For $Z \sim \operatorname{Exp}(1)$, its mgf is

$$M_Z(\lambda) = \mathbb{E}e^{\lambda Z} = \frac{1}{1-\lambda}, \qquad \lambda < 1.$$

Recall $\overline{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$. Then, for $t > 0$,

$$\Pr(\overline{Z} \ge 1 + t) \le \exp\big(-n\,[\,t - \ln(1+t)\,]\big).$$

In particular, for $0 < t \le 1$,

$$\Pr(\overline{Z} \ge 1 + t) \le \exp\left(-\frac{nt^2}{6}\right),$$

where we used $t - \ln(1+t) \ge t^2/6$ for $t \in [0,1]$. Thus,

$$\Pr(F^c) = \Pr\left(\overline{Z} \ge 1 + \frac{\delta}{8}\right) \le \exp\left(-\frac{n\delta^2}{384}\right).$$

Small ball for U. The moments of the exponential are $\mathbb{E}[Z^k] = \Gamma(1+k) = k!$. Let $W = Z^2$. For any $\lambda > 0$, apply the exponential trick with exponent λn:

$$\Pr\left(U \le 2 - \frac{\delta}{4}\right) = \Pr\left(e^{-\lambda n U} \ge e^{-\lambda n(2-\delta/4)}\right)$$
$$\le e^{\lambda n(2-\delta/4)}\, \mathbb{E}\left[e^{-\lambda \sum_{i=1}^n W_i}\right] = e^{2\lambda n}\, e^{-\lambda\delta n/4} \left(\mathbb{E}e^{-\lambda W}\right)^n.$$

Using $1 - x \le e^{-x} \le 1 - x + \frac{x^2}{2}$ for $x \ge 0$, we get

$$\left(\mathbb{E}e^{-\lambda W}\right)^n \le \left(1 - \lambda\mathbb{E}W + \frac{\lambda^2}{2}\mathbb{E}W^2\right)^n \le \left(1 - 2\lambda + 12\lambda^2\right)^n \le \exp\left(-2\lambda n + 12\lambda^2 n\right),$$

since $\mathbb{E}W = \mathbb{E}[Z^2] = 2$ and $\mathbb{E}[W^2] = \mathbb{E}[Z^4] = 24$. Therefore,

$$\Pr\left(U \le 2 - \frac{\delta}{4}\right) \le \exp\left(12\lambda^2 n - \frac{1}{4}\lambda\delta n\right).$$

Choosing $\lambda = \delta/96$ yields

$$\Pr\left(U \le 2 - \frac{\delta}{4}\right) \le \exp\left(-\frac{\delta^2 n}{768}\right).$$

Conclusion. Combining with (16),

$$p(\delta) \le \exp\left(-\frac{n\delta^2}{384}\right) + \exp\left(-\frac{\delta^2 n}{768}\right) \le 2\,\exp\left(-\frac{\delta^2 n}{768}\right).$$

Equivalently, for all $\delta > 0$,

$$\Pr\left(\|Y\|_2^2 \le (1-\delta)\, n\right) \le 2\,\exp\left(-\frac{\delta^2 n}{768}\right).$$

□

B.0.1 Proof of Lemma 13

Without loss of generality, we consider the isotropic simplex scaled by $1/\sqrt{n}$, which we denote here by $\triangle$, and G is a Gaussian matrix scaled by $1/\sqrt{n}$. For some $\delta \in (0,1)$, we can see that

$$\begin{aligned}\mathbb{E}_G \mathrm{Unif}_{X\sim\mathrm{Vol}(\triangle)}(\|GX\|_2 \le 1 + \frac{\delta}{2}, \|X\|_2 \ge 1) &= \int_{(\mathbb{R}^n)^{n-1}} \int_{\triangle} 1_{\|GX\|_2 \le 1+\delta/2, \|X\|_2 \ge 1}\, dX\, dG \\ &= \int_{\triangle} \int_{(\mathbb{R}^n)^{n-1}} 1_{\|GX\|_2 \le 1+\delta/2, \|X\|_2 \ge 1}\, dG\, dx \\ &= \int_{\triangle} \Pr_G(\|GX\|_2 \le 1 + \delta/2, \|X\|_2 \ge 1)\, dX \\ &\le \exp(-c_1 n\delta^2),\end{aligned}$$

where we used Fubini's, and that for any $x \in \mathbb{S}^{n-1}$, $\|Gx\|_2$ it holds

$$\Pr_G\{\|Gx\|_2 \le 1 - \delta/2\} \le \exp(-c\delta^2 n).$$

By Markov's inequality (on the matrix G), the proof is complete.

B.0.2 Proof of Lemma 10

Claim (i): We begin by controlling the (normalized) variance of the map $\boldsymbol{\xi} \mapsto \|\boldsymbol{\xi}\|_{K(n,d)^\circ}$, where $\boldsymbol{\xi} \sim \gamma_d \equiv \mathsf{N}(0, I_d)$; we do this below in Lemma 16. Then, by the boosted form of Dvoretzky's theorem (Lemma 6) applied to $K(n,d) \subset B_1^d$, we have

$$\begin{aligned} r(P_{n,d}) &\ge r(\mathbf{X}K(n,d)) \\ &\ge M^*(K(n,d))\left[1 - c\sqrt{n \cdot \mathrm{Var}_{\gamma_d}\left(\frac{\|\boldsymbol{\xi}\|_{K(n,d)^\circ}}{M^*(K(n,d))}\right)}\right] \ge M^*(K(n,d))\Big(1-\end{aligned}$$

The first part of Lemma 10 now follows by an integration argument. Namely, let $\mathcal{E}$ denote the event on which the inequality (17) holds. Then, for $d \ge 2n$, we have

$$M_{n,d} \le \mathbb{E}\frac{1}{r(P_{n,d})}\mathbf{1}_{\mathcal{E}} + \sqrt{\mathbb{E}_{\mathbf{X},\boldsymbol{\xi}}[\|\boldsymbol{\xi}\|^2_{\mathbf{X}B_1^d}]}\sqrt{\mathbb{P}(\mathcal{E}^c)} \le \frac{(1 + \frac{C}{\log(d/n)})}{\widetilde{M}_{n,d}} + c\,\exp$$

where above we simply recognize $\widetilde{M}_{n,d} = M^*(K(n,d))$.

Lemma 16. *Suppose $d \ge 2n$. Then, it holds for $K(n,d) = B_1^d \cap \frac{1}{n}B_\infty^d$ that for $\boldsymbol{\xi} \sim \gamma_d \equiv N(0, I_d)$,*

$$\mathrm{Var}_{\gamma_d}\left(\frac{\|\boldsymbol{\xi}\|_{K(n,d)^\circ}}{M^*(K(n,d))}\right) \lesssim \frac{1}{n\log^2(d/n)}.$$

Proof. Let $f(\boldsymbol{\xi}) = \|\boldsymbol{\xi}\|_{K(n,d)^\circ}$; we also denote by $I_{n,d} \subset [d]$ the (random) subset of n largest coordinates (by magnitude) of $\boldsymbol{\xi}$. It is easy to see that, with probability one,

$$f(\boldsymbol{\xi}) = \|\boldsymbol{\xi}\|_{K(n,d)^\circ} = \frac{1}{n}\sum_{i \in I_{n,d}} |\xi_i|.$$

In particular, we have $\partial_i f = \frac{\mathrm{sign}(\boldsymbol{\xi}_i)}{n}\mathbf{1}\{i \in I_{n,d}\}$, for any $i \in [d]$. It follows that

$$\|\partial_i f\|^2_{L^2(\gamma_d)} = \frac{1}{n^2}\cdot\frac{n}{d} = \frac{1}{nd}, \quad \text{while} \quad \|\partial_i f\|_{L^1(\gamma_d)} = \frac{1}{n}\cdot\frac{n}{d} = \frac{1}{d}.$$

Hence, Talagrand's $L^1 - L^2$ inequality, we have

$$\mathrm{Var}_{\gamma_d}(f(\boldsymbol{\xi})) \lesssim d\frac{\|\partial_1 f\|^2_{L^2(\gamma_d)}}{\log\left(e\frac{\|\partial_1 f\|_{L^2(\gamma_d)}}{\|\partial_1 f\|_{L^1(\gamma_d)}}\right)} = \frac{1}{n\log(e\sqrt{d/n})} \lesssim \frac{1}{n\log(ed/n)}. \tag{18a}$$

On the other hand, writing $\boldsymbol{\xi}_i^*$ for the order statistics (sorted by decreasing magnitude), we have

$$\mathbb{E}f(\boldsymbol{\xi}) = M^*(K(n,d)) = \frac{1}{n}\mathbb{E}\sum_{i=1}^n \boldsymbol{\xi}_i^* \asymp \sqrt{\log(d/n)}, \tag{18b}$$

where the last relation follows from standard estimates for the order statistics of a Gaussian random vector (*e.g.*, see [Gordon et al., 2007, Lemma 3.1]). Combining relations (18a) and (18b), we obtain the result. □

Claim (ii): By integrating with polar coordinates and using Jensen's inequality, for any centrally symmetri convex body $K \subset \mathbb{R}^n$, it holds that

$$\left(\frac{|K|}{|B_2^n|}\right)^{1/n} = \left(\int_{\mathbb{S}^{n-1}} \|\theta\|_K^{-n}\, d\sigma_{n-1}(\theta)\right)^{1/n} \geq \frac{1}{M(K)}.$$

Therefore we have

$$M_{n,d} \geq (1-\exp(-cn))\cdot\exp(-C(n,d)/n)\cdot\frac{1}{\widetilde{M}_{n,d}} \geq \left(1-\frac{c}{\sqrt{\log(d/n)}}\right)\frac{1}{\widetilde{M}_{n,d}}.$$

Define the event

$$\mathcal{E}_1 := \left\{ \frac{\|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\|_{\mathrm{op}}}{\sqrt{n}} \leq 4 + \sqrt{2\log\tfrac{ed}{n}} \right.$$

$$\left. \text{and} \quad \frac{\|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\|_{\mathrm{HS}}}{n} \leq 1+\frac{\sqrt{2\log\frac{ed}{n}}+2}{\sqrt{n}}, \text{ for all } S\subset[d], |S|=n, \boldsymbol{\varepsilon}\in\{-1,1\}^n \right\}$$

Lemma 17. *The event $\mathcal{E}_1$ holds with probability at least $1-2\exp(-2n)$.*

Proof. Fix $t>0$. We claim that the following tail bounds hold:

$$\Pr\left\{\exists S,\boldsymbol{\varepsilon} : \|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\|_{\mathrm{op}} \geq (2+\sqrt{2\log\tfrac{ed}{n}}+t)\sqrt{n}\right\} \leq \exp(-nt^2/2), \quad \text{and} \tag{19a}$$

$$\Pr\left\{\exists S,\boldsymbol{\varepsilon} : \frac{\|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\|_{\mathrm{HS}}}{n} \geq 1+\frac{\sqrt{2\log\frac{ed}{n}}+t}{\sqrt{n}}\right\} \leq \exp(-nt^2/2). \tag{19b}$$

To obtain the claimed bounds, we begin by noticing that for any fixed n-subset $S\subset[d], |S|=n$, it holds that

$$\|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\|_{\mathrm{op}} = \|\mathbf{X}_{S,\boldsymbol{\varepsilon}'}\|_{\mathrm{op}} \quad \text{and} \quad \|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\|_{\mathrm{HS}} = \|\mathbf{X}_{S,\boldsymbol{\varepsilon}'}\|_{\mathrm{HS}}$$

for all $\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}'\in\{-1,1\}^n$.

Now, inequality (19a) follows by the standard Davdison-Szarek tail bound for operator norms of Gaussian random matrices, and then applying a union bound over all $\binom{d}{n} \leq (\frac{ed}{n})^n$ many n-subsets $S\subset[d]$. Inequality (19b) follows by a union bound over all n-subsets applied to the the Borell-TIS inequality; we also used Jensen's inequality to estimate $\mathbb{E}\,\|\mathbf{X}_{S,\varepsilon}\|_{\mathrm{HS}} \leq n$. Finally, taking $t=2$ and union bound over the events underlying inequalities (19) yields the claim. □

Proof.

The boundary of the polytope $P_{n,d}$ can be decomposed facially as follows. For any $x\in\partial P_{n,d}$, there exists some facet $\mathcal{F}\in\mathcal{F}_{n-1}(P_{n,d})$ for which we can write

$$x = \mathbf{c}_{\mathcal{F}} + (x-\mathbf{c}_{\mathcal{F}}) = \mathbf{c}_{\mathcal{F}} + \mathbf{X}_{\mathcal{F}} z, \quad \text{for some } z\in\Delta_c. \tag{20}$$

Thus, by translation invariance of volume:

$$\min_{\mathcal{F}\in\mathcal{F}_{n-1}(P_{n,d})} \frac{|\mathcal{F}\cap(\mathbf{c}_{\mathcal{F}}+rB_2^n)|}{|\mathcal{F}|} \geq \min_{\substack{S\subset[d]\\|S|=n}} \min_{\boldsymbol{\varepsilon}\in\{-1,1\}^n} \frac{|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\Delta_c \cap rB_2^n|}{|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\Delta_c|}, \quad \ldots \tag{21}$$

Using the fact that $z\mapsto\mathbf{X}_{S,\boldsymbol{\varepsilon}}z$ is $\|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\|_{\mathrm{op}}$-Lipschitz, KLS for the centered simplex Δ_c gives

$$\min_{S,\boldsymbol{\varepsilon}} \frac{|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\Delta_c \cap r_{n,d}^{(1)}(\mathbf{X})B_2^n|}{|\mathbf{X}_{S,\boldsymbol{\varepsilon}}\Delta_c|} \geq \frac{1}{2}, \quad \text{for,} \quad r_{n,d}^{(1)}(\mathbf{X}) := \max_{S,\boldsymbol{\varepsilon}}\left\{\mathbb{E}_{z\sim\mathrm{Unif}(\Delta_c} \ldots\right. \tag{22}$$

Additionally define

$$r_{n,d}^{(2)}(\mathbf{X}) := \max_{\mathcal{F}\in\mathcal{F}_{n-1}(P_{n,d})} \|\mathbf{c}_{\mathcal{F}}\|_2, \quad \text{and} \quad r_{n,d}(\mathbf{X}) := r_{n,d}^{(1)}(\mathbf{X})+r_{n,d}^{(2)}(\mathbf{X}). \tag{23}$$

Note by the triangle inequality applied to display (20), we have

$$\partial P_{n,d}\cap r_{n,d}(\mathbf{X})B_2^n \supset \bigcup_{\mathcal{F}\in\mathcal{F}_{n-1}(P_{n,d})} \mathcal{F}\cap(\mathbf{c}_{\mathcal{F}}+r_{n,d}^{(1)}(\mathbf{X})).$$

Therefore, combining inequalities (21) and (22), it holds that $|\partial P_{n,d}\cap r_{n,d}(\mathbf{X})B_2^n| \geq \frac{1}{2}|\partial P_{n,d}|$. Hence, it follows that

$$|P_{n,d}| \leq \frac{1}{n}\sum_{\mathcal{F}\in\mathcal{F}_{n-1}(P_{n,d})} \|\mathbf{n}_{\mathcal{F}}\|_2|\mathcal{F}| \leq \frac{1}{n}R(P_{n,d})|\partial P_{n,d}| \leq \frac{R(P_{n,d})}{n} r_{n,d}(\ldots$$

On the event $\mathcal{E}_1$, using that $\mathrm{Cov}(\Delta_c) \preceq \frac{1}{n^2}I_n$, we have

$$r_{n,d}^{(1)}(\mathbf{X}) \leq 1+\frac{4c_1+2+(1+c_1)\sqrt{2\log\frac{ed}{n}}}{\sqrt{n}} \leq 1+c_2\sqrt{\frac{\log\frac{ed}{n}}{n}}$$

Applying Dvoretzky's theorem we also have with probability at least $1-\exp(-c_1 n)$ that

$$r_{n,d}^{(2)}(\mathbf{X}) \leq R(\mathbf{X}K(n,d)) \leq \widetilde{M}_{n,d}\left(1+c_3\frac{1}{\widetilde{M}_{n,d}}\right) \leq \widetilde{M}_{n,d}+c_3.$$

Hence, combining the previous two inequalities, and using that $\widetilde{M}_{n,d} \asymp \sqrt{\log(ed/n)}$, it holds that

$$|P_{n,d}| \leq \exp\left(c_4\frac{n}{\sqrt{\log(ed/n)}}\right)\cdot|\widetilde{M}_{n,d}\cdot B_2^n|.$$

Note that the norm induced by $K(n,d)^\circ$ corresponds to the averaged top n-norm. By, applying $L_1 - L_2$ Talgarnd's inequality to the average of the top n-order statistics of a Gaussian vector in $\mathbb{R}^d$, we obtain that

$$\mathrm{Var}\left(\frac{\|\boldsymbol{\xi}\|_{K(n,d)^\circ}}{\widetilde{M}_{n,d}}\right) \lesssim \frac{1}{\log(d/n)^2}. \tag{24}$$

and also note that $R((\mathbf{X}K(n,d)) \le \widetilde{M}_{n,d} + \Theta(1)$ with high probability over $P_{n,d}$. Note that by Steps I and II and the KLS property of the simplex and Fluery's distribution, we know that $(1 - \exp(-C\tilde{c}_{n,d})) \cdot |\mathcal{F}_{n-1}(P_{n,d})|$ of the facets of $P_{n,d}$ satisfies

$$|\tilde{G}_S \triangle_c \cap \left(1 + \frac{C}{\sqrt{n}}\right) \cdot \sqrt{\mathrm{Tr}(G_S^T G_S)} \cdot \mathbb{E}\|Z\|_2 \cdot B_{n-1}| \ge |\tilde{G}_S \triangle_c \cap \left(1 + \frac{C_1}{\sqrt{n}}\right) \cdot B_{n-1}| \ge 0.9 \cdot |\tilde{G}_S \triangle_c|,$$

where we used that $\|\tilde{G}_S\| \lesssim \|\tilde{G}_S/\sqrt{n}\|_F = (1 + O(1/\sqrt{n})) \cdot \sqrt{n}$ with probability of at last $1 - \exp(-cn)$. Hence, it holds that

$$|P_{n,d}| \lesssim \left(\max_{\mathcal{F} \in \mathcal{F}_{n-1}(P_{n,d})} \|\mathbf{X}_{\mathcal{S}}\boldsymbol{\varepsilon}\|_2 + 1 + \frac{C}{\sqrt{n}}\right)^n |B_n| \lesssim \exp(Cn/\sqrt{\log(d/n)})|\widetilde{M}_{n,d} \cdot B_n| \tag{25}$$

where the last inequality follows from the upper bound on $R(P_{n,d})$ as by this inclusion, we can control all the barycenters ℓ_2 norms, i.e . $\|\mathbf{c}_{\mathcal{F}}\|_2$, uniformly with probability of at least $1 - \exp(-cn)$. We conclude that

$$|P_{n,d}| \le \exp(\frac{Cn}{\sqrt{\log(d/n)}}) \cdot |\widetilde{M}_{n,d} B_n|.$$

Finally, by the lower and upper inclusions, it holds with probability of $1 - \exp(-cn)$ that

$$\exp(-cn/\sqrt{\log(d/n)})|\bar{B}_n| \le |\bar{P}_{n,d}| \le \exp(cn/\sqrt{\log(d/n)})|\bar{B}_n|,$$

and proof is complete. □

B.0.3 Proof of Lemma 8

We write, for $t \ge 0$, $F(t) := \Pr\{|g| \le t\}$, where $g \sim \mathsf{N}(0,1)$. We write the density of $T_{n,d}$ as

$$f_{n,d}(t) = \frac{1}{Z_{n,d}} F(t)^{d-n} \mathrm{e}^{-nt^2/2} = \frac{1}{Z_{n,d}} \mathrm{e}^{-V_{n,d}(t)}.$$

Here, we set

$$V_{n,d}(t) := (d-n)\log\frac{1}{F(t)} + \frac{n}{2}t^2.$$

Throughout, we use the following shorthand notation:

$$L := \log(d/n), \qquad t^\star := t^\star_{n,d}, \qquad m := \mathrm{Med}\, T_{n,d}.$$

We observe that $f_{n,d}$ is log-concave, as $V_{n,d}$ is convex. Thus, the mode $t^\star$ is the unique minimizer of $V_{n,d}$. The proof of the tail bound is based on several elementary lemmas controlling $V_{n,d}$ and its curvature near the mode.

Lemma 18. *Let $V = V_{n,d}$ and $t^\star = t^\star_{n,d}$. Then:*

(i) V'' is nonincreasing on $\mathbb{R}_+$;

(ii) for every $s \le t^\star$,

$$V(s) - V(t^\star) \ge \frac{V''(t^\star)}{2}(t^\star - s)^2;$$

(iii) for every $s \ge t^\star$,

$$V(s) - V(t^\star) \le \frac{V''(t^\star)}{2}(s - t^\star)^2.$$

Lemma 19. *Fix $\varepsilon \in (0,1)$. Then*

$$\Pr\left\{T_{n,d} \le (1-\varepsilon)t^\star_{n,d}\right\} \le \frac{6}{5\varepsilon t^\star_{n,d}\sqrt{V''_{n,d}(t^\star_{n,d})}} \exp\left\{-\frac{V''_{n,d}(t^\star_{n,d})}{2}\varepsilon^2(t^\star_{n,d})$$

Lemma 20. *Suppose that $d \ge 4n$. Then there exist absolute constants $c, C > 0$ such that*

$$\sqrt{L} \le t^\star_{n,d} \le C\sqrt{L} \qquad \text{and} \qquad cnL \le V''_{n,d}(t^\star_{n,d}) \le CnL.$$

Lemma 21. *For every fixed $K > 0$, there exist constants $c_K, C_K > 0$ such that if*

$$c_K nL \le V''_{n,d}(t) \le C_K nL \quad \text{for all}\, t \in \left[t^\star_{n,d},\, t^\star_{n,d} + \frac{K}{\sqrt{L}}\right].$$

Lemma 22. *For every fixed $K > 0$, there exist constants $c_K, C_K > 0$ such that for every $r \in (0, \frac{K}{\sqrt{L}})$, we have*

$$\Pr\left\{T_{n,d} \ge t^\star_{n,d} + r\right\} \le \frac{C_K}{r\sqrt{nL}} \exp\left\{-c_K nLr^2\right\}.$$

Lemma 23. *The median satisfies $\mathrm{Med}\, T_{n,d} \ge t^\star_{n,d}$.*

Lemma 24. *There exists an absolute constant $c > 0$ such that $f_{n,d}(t^\star_{n,d}) \ge c\sqrt{nL}$.*

We first show how these lemmas imply the proposition.

Proof of Lemma 8. Fix $\varepsilon > 0$; denote $r := \varepsilon t^\star$. By Lemma 20, we have $t^\star \le C\sqrt{L}$, and hence, for a sufficiently small $c > 0$, $\varepsilon \le \frac{c}{L}$ implies $r \le \frac{K_0}{\sqrt{L}}$, for an absolute constant $K_0 > 0$. We first locate the median. Fix $A > 0$. Lemma 22 with $K = A$ yields

$$\Pr\left\{T_{n,d} \geq t^\star + \frac{A}{\sqrt{nL}}\right\} \leq \frac{C_A}{A}\exp(-c_A A^2).$$

Evidently, we can chose A so large that the right-hand side is strictly smaller than 1/2. By Lemma 23, we have $m \geq t^\star$, and hence for a sufficiently large $C_0 > 0$, we have

$$t^\star \leq m \leq t^\star + \frac{C_0}{\sqrt{nL}},$$

thereby furnishing the second claim of Lemma 8.

We next prove the tail bound around the median. Throughout, we denote by $r_0 = \frac{C_0}{\sqrt{nL}}$. We proceed in two cases.

Case 1: $r \geq 2r_0$. Since $m \leq t^\star + r_0 \leq t^\star + r/2$, we have

$$\{T_{n,d} \leq m - r\} \subseteq \{T_{n,d} \leq t^\star - r/2\}. \tag{26a}$$

Also, since $m \geq t^\star$,

$$\{T_{n,d} \geq m + r\} \subseteq \{T_{n,d} \geq t^\star + r\}. \tag{26b}$$

Hence, by Lemma 19 and Lemma 20 for the lower tail, and Lemma 22 for the upper tail (recalling that $r \leq \frac{K_0}{\sqrt{L}}$) we have for sufficiently large $C > 0$ and sufficiently small $c > 0$ that:

$$\max\Big\{\Pr\{T_{n,d} \geq t^\star + r\}, \Pr\{T_{n,d} \leq t^\star - r/2\}\Big\} \leq \frac{C}{r\sqrt{nL}}\exp(-cnLr^2).$$

Hence, a union bound and the inclusions Eqs. (26a) and (26b) yield

$$\Pr\{|T_{n,d} - m| \geq r\} \leq \frac{C'}{r\sqrt{nL}}\exp(-cnLr^2).$$

Since $r\sqrt{nL} \geq 2C_0$, the prefactor is bounded by a universal constant. Hence, after decreasing c if necessary, we obtain the desired inequality in this case:

$$\Pr\{|T_{n,d} - m| \geq r\} \leq \exp(-c'nLr^2).$$

Case 2: $0 < r < 2r_0$. In this case, we have $m + r \leq t^\star + 3r_0$. For every $u \in [m, m + r]$, we have $u \geq t^\star$, and Lemma 18(iii) together with Lemma 20 implies

$$V(u) - V(t^\star) \leq \frac{V''(t^\star)}{2}(u - t^\star)^2 \leq C'.$$

Above, C' depends only on C_0. Therefore, Lemma 24 yields

$$f_{n,d}(u) = f_{n,d}(t^\star)\exp(-(V(u) - V(t^\star))) \geq c'\sqrt{nL} \qquad \text{for all } u \in [m, r$$

Again, c' only depends on C_0. Consequently,

$$\Pr\{|T_{n,d} - m| < r\} \geq \Pr\{m \leq T_{n,d} \leq m + r\} \geq c'r\sqrt{nL}.$$

Hence, using $1 - u \leq \mathrm{e}^{-u}$, we have

$$\Pr\{|T_{n,d} - m| \geq r\} \leq 1 - c'r\sqrt{nL} \leq \exp(-c'r\sqrt{nL}).$$

Now, in this case, we have $r\sqrt{nL} \leq 2C_0$; equivalently $r\sqrt{nL} \geq \frac{1}{2C_0}nLr^2$, which yields

$$\Pr\{|T_{n,d} - m| \geq r\} \leq \exp(-c''nLr^2).$$

Combining the two cases proves that

$$\Pr\{|T_{n,d} - m| \geq r\} \leq \exp(-c'''nLr^2).$$

Recalling that $r = \varepsilon t^\star$ and using Lemma 20—-namely that $(t^\star)^2 \geq L$—we conclude that

$$\Pr\left\{|T_{n,d} - \operatorname{Med} T_{n,d}| \geq \varepsilon t^\star_{n,d}\right\} \leq \exp\{-c'''nL^2\varepsilon^2\},$$

as needed. □

B.0.4 Proof of Lemma 18

We use the shorthand $V = V_{n,d}$ and $t^\star = t^\star_{n,d}$. Define

$$R(t) := \frac{F'(t)}{F(t)} = \frac{2\phi(t)}{F(t)} \qquad \text{and} \qquad S(t) := R(t)^2 + tR(t),$$

so that

$$R'(t) = -S(t), \qquad V'(t) = nt - (d-n)R(t), \qquad V''(t) = n + (d-n)S($$

A direct computation gives

$$S'(t) = 2R(t)R'(t) + R(t) + tR'(t) = R(t)\big(1 - t^2 - 3tR(t) - 2R(t)^2\big).$$

Since $R(t) \geq 0$, it is enough to show that

$$H(t) := 1 - t^2 - 3tR(t) - 2R(t)^2 \leq 0 \qquad \text{for all } t \geq 0.$$

This is immediate when $t \geq 1$. If $0 \leq t < 1$, then

$$F(t) = 2\int_0^t \phi(s)\,\mathrm{d}s \leq 2t\phi(0),$$

which implies

$$R(t) = \frac{2\phi(t)}{F(t)} \geq \frac{\mathrm{e}^{-t^2/2}}{t}.$$

Therefore,

$$H(t) \le 1 - 3tR(t) \le 1 - 3\mathrm{e}^{-t^2/2} \le 1 - \frac{3}{\sqrt{\mathrm{e}}} < 0.$$

Hence $S'(t) \le 0$, so $V'''(t) = (d-n)S'(t) \le 0$ on $\mathbb{R}_+$, proving item (i).

If $s \le t^\star$, then $V'(t^\star) = 0$ and the fundamental theorem of calculus gives

$$V(s) - V(t^\star) = \int_s^{t^\star}\int_x^{t^\star} V''(y)\,\mathrm{d}y\,\mathrm{d}x = \int_s^{t^\star}(x-s)V''(x)\,\mathrm{d}x.$$

By item (i), we have $V''(x) \ge V''(t^\star)$ for $s \le x \le t^\star$, and therefore

$$V(s) - V(t^\star) \ge \frac{V''(t^\star)}{2}(t^\star - s)^2,$$

which proves item (ii).

Similarly, if $s \ge t^\star$, then

$$V(s) - V(t^\star) = \int_{t^\star}^{s}\int_{t^\star}^{x} V''(y)\,\mathrm{d}y\,\mathrm{d}x = \int_{t^\star}^{s}(s-x)V''(x)\,\mathrm{d}x.$$

Again by item (i), now $V''(x) \le V''(t^\star)$ for $x \ge t^\star$, and hence

$$V(s) - V(t^\star) \le \frac{V''(t^\star)}{2}(s - t^\star)^2,$$

proving item (iii).

B.0.5 Proof of Lemma 19

We write, for short,

$$T = T_{n,d}, \qquad Z = Z_{n,d}, \qquad V = V_{n,d}, \qquad t^\star = t^\star_{n,d}, \qquad t^- = (1-\varepsilon)t^\star.$$

We claim that

$$Z \overset{(a)}{\ge} \frac{5}{6\sqrt{V''(t^\star)}}\mathrm{e}^{-V(t^\star)} \qquad \text{and} \qquad \int_0^{t^-} \mathrm{e}^{-V(t)}\,\mathrm{d}t \overset{(b)}{\le} \frac{\mathrm{e}^{-V(t^\star)}}{\varepsilon V''(t^\star)t^\star}\exp\left\{-\frac{V''(t^\star)}{2}\varepsilon^2(t^\star)^2\right\}. \tag{27}$$

Assuming (27), we obtain

$$\Pr\{T \le t^-\} = \frac{1}{Z}\int_0^{t^-} \mathrm{e}^{-V(t)}\,\mathrm{d}t \le \frac{6}{5\varepsilon t^\star\sqrt{V''(t^\star)}}\exp\left\{-\frac{V''(t^\star)}{2}\varepsilon^2(t^\star)^2\right\},$$

as required.

To prove (27)(a), we use Lemma 18(iii). For every $\delta > 0$,

$$Z \ge \int_{t^\star}^{t^\star+\delta} \mathrm{e}^{-V(x)}\,\mathrm{d}x \ge \mathrm{e}^{-V(t^\star)}\int_0^{\delta} \mathrm{e}^{-V''(t^\star)x^2/2}\,\mathrm{d}x.$$

Changing variables $u = x\sqrt{V''(t^\star)}$, we get

$$Z \ge \frac{\mathrm{e}^{-V(t^\star)}}{\sqrt{V''(t^\star)}}\int_0^{\delta\sqrt{V''(t^\star)}} \mathrm{e}^{-u^2/2}\,\mathrm{d}u.$$

Choosing $\delta = 1/\sqrt{V''(t^\star)}$ yields

$$Z \ge \frac{\mathrm{e}^{-V(t^\star)}}{\sqrt{V''(t^\star)}}\int_0^1 \mathrm{e}^{-u^2/2}\,\mathrm{d}u \ge \frac{5}{6}\frac{\mathrm{e}^{-V(t^\star)}}{\sqrt{V''(t^\star)}}.$$

For (27)(b), Lemma 18(ii) gives

$$\begin{aligned}\int_0^{t^-} \mathrm{e}^{-V(t)}\,\mathrm{d}t &\le \mathrm{e}^{-V(t^\star)}\int_0^{t^-} \mathrm{e}^{-V''(t^\star)(t^\star-s)^2/2}\,\mathrm{d}s \\ &= \mathrm{e}^{-V(t^\star)}\int_{\varepsilon t^\star}^{t^\star} \mathrm{e}^{-V''(t^\star)x^2/2}\,\mathrm{d}x \\ &\le \mathrm{e}^{-V(t^\star)}\int_{\varepsilon t^\star}^{\infty} \mathrm{e}^{-V''(t^\star)x^2/2}\,\mathrm{d}x.\end{aligned}$$

Using the elementary bound

$$\int_x^\infty \mathrm{e}^{-\beta u^2/2}\,\mathrm{d}u \le \frac{1}{x}\int_x^\infty u\,\mathrm{e}^{-\beta u^2/2}\,\mathrm{d}u = \frac{\mathrm{e}^{-\beta x^2/2}}{\beta x} \qquad (\beta, x > 0),$$

with $\beta = V''(t^\star)$ and $x = \varepsilon t^\star$, we obtain (27)(b).

B.0.6 Proof of Lemma 20

We write $V = V_{n,d}$ and $t^\star = t^\star_{n,d}$. Set

$$t_0 := \sqrt{L}.$$

We first show that $t^\star \ge t_0$. Since V is convex, it suffices to prove that $V'(t_0) \le 0$. Using the identity

$$V'(s) = ns - (d-n)\frac{2\phi(s)}{F(s)} \le ns - 2(d-n)\phi(s),$$

and recalling that $\phi(t_0) = (2\pi)^{-1/2}\mathrm{e}^{-L/2} = (2\pi)^{-1/2}\sqrt{n/d}$, we obtain

$$V'(t_0) \le d\left[\frac{n}{d}\sqrt{\log(d/n)} - \sqrt{\frac{2}{\pi}}\left(1 - \frac{n}{d}\right)\sqrt{\frac{n}{d}}\right].$$

If we set

$$\psi(x) := x\sqrt{\log(1/x)} - \sqrt{\frac{2}{\pi}}(1-x)\sqrt{x},$$

then $V'(t_0) \le d\psi(n/d)$. A direct calculus check shows that $\psi(x) \le 0$ on $[0, 1/4]$. Since $d \ge 4n$, we have $n/d \in (0, 1/4]$, hence $V'(t_0) \le 0$, proving that $t^\star \ge \sqrt{L}$. We next prove the upper bound on $t^\star$. Since $d \ge 4n$, we have $L \ge \log 4 > 1$, and thus $t^\star \ge \sqrt{L} > 1$. Using $V'(t^\star) = 0$, we get

$$nt^\star = (d-n)\frac{2\phi(t^\star)}{F(t^\star)} \le \frac{2(d-n)}{F(1)}\phi(t^\star) \le Cd\,\phi(t^\star).$$

From the form of the Gaussian density, this implies $\mathrm{e}^{-(t^\star)^2/2} \ge c\frac{n}{d}t^\star$; taking logarithms yields $(t^\star)^2 \le$

$2L + C \le CL$. Finally, using again $V'(t^\star) = 0$, we have

$$R(t^\star) = \frac{F'(t^\star)}{F(t^\star)} = \frac{n}{d-n} t^\star.$$

Therefore,

$$V''(t^\star) = n + (d-n)\big(R(t^\star)^2 + t^\star R(t^\star)\big) = n + \frac{n^2}{d-n}(t^\star)^2 + n(t^\star)^2 \le CnL,$$

where we used the fact that the middle term is at most $\frac{n}{3}(t^\star)^2$. On the other hand, the display above also yields

$$V''(t^\star) \ge n(t^\star)^2 \ge nL,$$

as required.

B.0.7 Proof of Lemma 21

Fix $K > 0$, and write $t^\star = t^\star_{n,d}$. Let

$$t \in \Big[t^\star,\, t^\star + \frac{K}{\sqrt{L}}\Big].$$

Since

$$V''(t) = n+(d-n)\big(R(t)^2+tR(t)\big) \ge (d-n)tR(t) = (d-n)\frac{2t\phi(t)}{F(t)} \ge 2(d-n)t\phi(t),$$

it suffices to lower bound $t\phi(t)$ in terms of $t^\star\phi(t^\star)$. Write $t = t^\star + u$ with $0 \le u \le K/\sqrt{L}$. Then

$$\phi(t) = \phi(t^\star)\exp\Big(-t^\star u - \frac{u^2}{2}\Big).$$

By Lemma 20, $t^\star \le C\sqrt{L}$, so

$$\exp\Big(-t^\star u - \frac{u^2}{2}\Big) \ge c_K.$$

Also $t \ge t^\star$, hence

$$t\phi(t) \ge c_K t^\star \phi(t^\star).$$

Using $V'(t^\star) = 0$ and $t^\star \ge 1$, we obtain

$$2(d-n)t^\star\phi(t^\star) = nt^{\star 2}F(t^\star) \ge cnL,$$

because $F(t^\star) \ge F(1) > 0$ and $(t^\star)^2 \ge L$ by Lemma 20. Therefore,

$$V''(t) \ge c_K nL.$$

For the upper bound, note that R is decreasing because $R' = -S \le 0$. Hence

$$R(t) \le R(t^\star) = \frac{n}{d-n}t^\star.$$

Therefore,

$$V''(t) \le n + (d-n)\big(R(t^\star)^2 + tR(t^\star)\big) = n + \frac{n^2}{d-n}(t^\star)^2 + ntt^\star.$$

By Lemma 20, $t^\star \le C\sqrt{L}$ and $t \le t^\star + K/\sqrt{L} \le C_K\sqrt{L}$. Thus $V''(t) \le C_K nL$, as claimed.

B.0.8 Proof of Lemma 22

Fix $K > 0$ and let $0 < r \le \frac{K}{\sqrt{L}}$. We set $m_r := \inf_{0\le u\le r} V''(t^\star_{n,d} + u)$. By Lemma 21, we have $m_r \ge c_K nL$. Moreover, the fundamental theorem of calculus gives us

$$V(t^\star+r)-V(t^\star) = \int_0^r (r-u)V''(t^\star+u)\,\mathrm{d}u \ge \frac{m_r r^2}{2}, \quad \text{and} \quad V'(t^\star+r)$$

Moreover, from the convexity of V, we have the affine lower bound

$$V(t^\star+r+s) \ge V(t^\star+r)+sV'(t^\star+r) \qquad \text{for all } s \ge 0.$$

Therefore,

$$\int_{t^\star+r}^\infty \mathrm{e}^{-V(t)}\,\mathrm{d}t \le \mathrm{e}^{-V(t^\star+r)}\int_0^\infty \mathrm{e}^{-sV'(t^\star+r)}\,\mathrm{d}s \le \frac{\mathrm{e}^{-V(t^\star)}}{m_r r}\exp\Big(-\frac{m_r r}{2}$$

On the other hand, by (27)(a),

$$Z_{n,d} \ge \frac{5}{6\sqrt{V''(t^\star)}}\mathrm{e}^{-V(t^\star)}.$$

Using Lemma 20, namely $V''(t^\star) \le CnL$, we deduce that $Z_{n,d} \ge c\frac{\mathrm{e}^{-V(t^\star)}}{\sqrt{nL}}$. Dividing the two estimates and using $m_r \ge c_K nL$ gives

$$\Pr\{T_{n,d} \ge t^\star_{n,d} + r\} \le \frac{C_K}{r\sqrt{nL}}\exp(-c_K nLr^2),$$

as claimed.

B.0.9 Proof of Lemma 23

Let $f = f_{n,d}$ denote the density of $T_{n,d}$. For $x \in [0, t^\star]$, Lemma 18(i) implies

$$V(t^\star - x) - V(t^\star) = \int_0^x (x-u)V''(t^\star - u)\,\mathrm{d}u \ge \int_0^x (x-u)V''(t^\star +$$

Thus

$$f(t^\star - x) \le f(t^\star + x) \qquad \text{for all } x \in [0, t^\star].$$

Integrating over $x \in [0, t^\star]$ gives

$$\Pr\{T_{n,d} \le t^\star\} = \int_0^{t^\star} f(t^\star-x)\,\mathrm{d}x \le \int_0^{t^\star} f(t^\star+x)\,\mathrm{d}x \le \Pr\{T_{n,d} \ge t^\star\}$$

Hence $\Pr\{T_{n,d} \le t^\star\} \le 1/2$, and therefore every median satisfies $\operatorname{Med} T_{n,d} \ge t^\star$.

B.0.10 Proof of Lemma 24

Write $t^\star = t^\star_{n,d}$ and $V = V_{n,d}$. Decompose the normalizing constant into two parts:

$$Z_{n,d} = I_-+I_+, \qquad I_- := \int_0^{t^\star} \mathrm{e}^{-V(t)}\,\mathrm{d}t, \qquad I_+ := \int_{t^\star}^\infty \mathrm{e}^{-V(t)}\,\mathrm{d}t.$$

By Lemma 18(ii),

$$I_- \leq e^{-V(t^\star)} \int_0^{t^\star} e^{-V''(t^\star)(t^\star - s)^2/2}\, ds \leq C\frac{e^{-V(t^\star)}}{\sqrt{V''(t^\star)}} \leq C\frac{e^{-V(t^\star)}}{\sqrt{nL}},$$

where the last step uses Lemma 20.

Fix $w := \frac{1}{\sqrt{L}}$. We further split I_+ in to two tersm such that $I_+ = I_{+,1} + I_{+,2}$:

$$I_{+,1} := \int_{t^\star}^{t^\star + w} e^{-V(t)}\, dt, \qquad I_{+,2} := \int_{t^\star + w}^{\infty} e^{-V(t)}\, dt.$$

By Lemma 21 with $K = 1$, we have $V''(t) \geq cnL$ on $[t^\star, t^\star + w]$. Hence, for $0 \leq u \leq w$,

$$V(t^\star + u) - V(t^\star) = \int_0^u (u - s)V''(t^\star + s)\, ds \geq \frac{cnL}{2}u^2.$$

Therefore,

$$I_{+,1} \leq e^{-V(t^\star)} \int_0^w e^{-cnLu^2/2}\, du \leq C\frac{e^{-V(t^\star)}}{\sqrt{nL}}.$$

Also,

$$V(t^\star + w) - V(t^\star) \geq cnLw^2 = cn, \qquad V'(t^\star + w) \geq cnLw = cn\sqrt{L}.$$

Using convexity exactly as in the proof of Lemma 22, we obtain

$$I_{+,2} \leq \frac{e^{-V(t^\star + w)}}{V'(t^\star + w)} \leq \frac{e^{-V(t^\star)}}{cn\sqrt{L}} e^{-cn} \leq C\frac{e^{-V(t^\star)}}{\sqrt{nL}}.$$

Combining the bounds on I_-, $I_{+,1}$, and $I_{+,2}$, we conclude that

$$Z_{n,d} \leq C\frac{e^{-V(t^\star)}}{\sqrt{nL}}.$$

Equivalently,

$$f_{n,d}(t^\star) = \frac{e^{-V(t^\star)}}{Z_{n,d}} \geq c\sqrt{nL},$$

which proves the lemma.

B.0.11 Proof of Lemma 7

By Bartlett's decomposition, we can write the determinant $|G|$ as a product of independent χ_k random variables. In particular, we have

$$S_n - \mathbb{E}S_n = \sum_{k=2}^n X_k, \quad \text{where} \quad X_k = \log \chi_k - \mathbb{E}\log \chi_k.$$

Define the cumulant generating function,

$$\psi(\lambda) = \log \mathbb{E}\exp(\lambda(S_n - \mathbb{E}S_n)), \quad \lambda \in \mathbb{R}.$$

We use the following lemma.

Lemma 25. *For $\lambda > -k$, it holds that*

$$\log \mathbb{E}\exp(\lambda X_k) \leq \frac{3}{4}\frac{\lambda^2}{k}.$$

Proof. Using the fact that χ_k^2 is distributed as a Gamma random variate with shape parameter $k/2$ and scale parameter 2, it follows from standard identities for Gamma random variables that

$$\mathbb{E}\chi_k^\lambda = 2^{\lambda/2}\frac{\Gamma(\frac{\lambda+k}{2})}{\Gamma(k/2)}.$$

Above $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\, dt$ and we set $\psi = \Gamma'/\Gamma$. Differentiating the identity above at $\lambda = 0$, we obtain

$$\mathbb{E}\log \chi_k = \frac{1}{2}\log 2 + \frac{1}{2}\psi(k/2).$$

Hence,

$$\begin{aligned}\log \mathbb{E}\exp(\lambda X_k) &= \log\Gamma(\tfrac{k+\lambda}{2}) - \Big[\log\Gamma(\tfrac{k}{2}) + \frac{\lambda}{2}\psi(\tfrac{k}{2})\Big] \\ &\overset{\text{(i)}}{=} \sum_{m=0}^\infty \Big\{\log\Big(1 + \frac{\lambda}{k+2m}\Big) - \frac{\lambda}{k+2m}\Big\} \\ &\overset{\text{(ii)}}{\leq} \frac{\lambda^2}{2}\sum_{m=0}^\infty \frac{1}{(k+2m)^2} \\ &\leq \frac{\lambda^2}{2}\Big(\frac{1}{k^2} + \int_0^\infty \frac{1}{(k+2x)^2}\, dx\Big) \leq \frac{3}{4}\frac{\lambda^2}{k}.\end{aligned}$$

Above, relation (i) is the Taylor expansion of the logarithm of the Gamma function and inequality (ii) used $\log(1+x) - x \leq x^2/2$, which holds for $x > -1$; we applied it term-wise with $x = \frac{\lambda}{k+2m}$. □

Applying it to the sum $S_n - \mathbb{E}S_n$, we have for $\lambda > -1$ that

$$\psi(\lambda) \leq \frac{3}{4}\lambda^2 \sum_{k=2}^n \frac{1}{k} \leq \lambda^2 \log n.$$

Therefore, for $t \geq 0$ and with $\lambda = \frac{t}{2\log n}$, we have

$$\mathbb{P}\Big\{S_n - \mathbb{E}S_n > t\Big\} \leq \exp\left\{-\lambda t + \lambda^2 \log n\right\} = \exp\Big\{-\frac{t^2}{4\log n}\Big\}.$$

Similarly, applying the same argument with $\lambda = -\frac{t}{\log n}$, we have

$$\mathbb{P}\Big\{S_n - \mathbb{E}S_n < -t\Big\} \leq \exp\left\{\lambda t + \lambda^2 \log n\right\} = \exp\Big\{-\frac{t^2}{4\log n}\Big\}, \quad \text{for } t \in$$

For the lower tail and for $t \geq 2\log n$, we require a different argument. In this case, we write

$$2S_n = \log |G|^2 = \sum_{k=2}^n \log \chi_k^2 = Z_2 + Y,$$

where we set $Z_k = \log \chi_k^2$ and $Y = \sum_{k=3}^n Z_k$. Then, using the fact that $\mathbb{P}\{\chi_2^2 \le u\} = 1 - e^{-u/2} \le u/2$, we have

$$\mathbb{P}\{S_n - \mathbb{E}S_n \le -t\} = \mathbb{E}_Y \mathbb{P}\Big\{\chi_2^2 \le \frac{e^{2\mathbb{E}S_n - 2t}}{e^Y} \mid Y\Big\} \le \frac{e^{2\mathbb{E}S_n - 2t}}{2} \mathbb{E}_Y\left[e^{-Y}\right] = \frac{e^{2\mathbb{E}S_n - 2t}}{2} \prod_{k=3}^n \mathbb{E}\frac{1}{\chi_k^2} = \frac{e^{2\mathbb{E}S_n - 2t}}{2(n-2)!}.$$

The final relation used that $\mathbb{E}\chi_k^{-2} = (k-2)^{-1}$ for $k \ge 3$. Note that

$$\mathbb{E}Z_k = 2\mathbb{E}\log \chi_k = \log 2 + \psi\Big(\frac{k}{2}\Big) \le \log(k-1).$$

The final inequality comes from the log-convexity of Γ:

$$\psi\Big(\frac{k}{2}\Big) = (\log \Gamma)'\Big(\frac{k}{2}\Big) \le \log \frac{\Gamma(k/2 + 1/2)}{\Gamma(k/2 - 1/2)} = \log\Big(\frac{k-1}{2}\Big),$$

which holds for integers $k \ge 2$. Consequently,

$$e^{2\mathbb{E}S_n} = \prod_{k=2}^n e^{\mathbb{E}Z_k} \le (n-1)!.$$

Hence, combining the previous displays,

$$\mathbb{P}\{S_n - \mathbb{E}S_n \le -t\} \le \frac{n-1}{2} e^{-2t},$$

for $t \ge 0$, as needed.