

On nearly radial marginals of high-dimensional probability measures

Bo'az Klartag*

Abstract

Suppose that μ is an absolutely continuous probability measure on \mathbb{R}^n , for large n . Then μ has low-dimensional marginals that are approximately spherically-symmetric. More precisely, if $n \geq (C/\varepsilon)^{C^d}$, then there exist d -dimensional marginals of μ that are ε -far from being spherically-symmetric, in an appropriate sense. Here $C > 0$ is a universal constant.

1 Introduction

The purpose of this paper is to clarify a seven line paragraph by Gromov [11, Section 1.2.F]. We are interested in projections of high-dimensional probability measures. Not all probability measures on \mathbb{R}^n , for large n , are truly n -dimensional. For instance, a measure supported on an atom or two should not be considered high-dimensional. Roughly speaking, we think of a probability measure on a linear space as decently high-dimensional if any subspace of bounded dimension contains only a small fraction of the total mass.

Definition 1.1 *Let μ be a Borel probability measure on \mathbb{R}^n and $\varepsilon > 0$. We say that μ is “decently high-dimensional with parameter ε ”, or “ ε -decent” in short, if for any linear subspace $E \subseteq \mathbb{R}^n$,*

$$\mu(E) \leq \varepsilon \dim(E). \quad (1)$$

We say that μ is decent if it is ε -decent for $\varepsilon = 1/n$, the minimal possible value of ε .

Clearly, all absolutely continuous probability measures on \mathbb{R}^n are decent, as are many discrete measures. Note that a decent measure μ necessarily satisfies $\mu(\{0\}) = 0$, however, this feature should not be taken too seriously. A measure μ is “weakly ε -decent” if (1) holds for all subspaces $E \subseteq \mathbb{R}^n$ except $E = \{0\}$. For a measure μ on a measurable space Ω and a measurable map $T : \Omega \rightarrow \Omega'$, we write $T_*(\mu)$ for the push-forward of μ under T , i.e.,

$$T_*(\mu)(A) = \mu(T^{-1}(A))$$

for all measurable sets $A \subseteq \Omega'$. When μ is a probability measure on \mathbb{R}^n and $T : \mathbb{R}^n \rightarrow \mathbb{R}^\ell$ is a linear map with $\ell < n$, we say that $T_*(\mu)$ is a marginal of μ , or a measure projection

*Supported in part by the Israel Science Foundation and by a Marie Curie Reintegration Grant from the Commission of the European Communities.

of μ . The classical Dvoretzky theorem asserts that appropriate geometric projections of any high-dimensional convex body are approximately Euclidean balls (see Milman [23] and references therein). The analogous statement for probability measures should perhaps be the following (see Gromov [11]): Appropriate measure projections of any decent high-dimensional probability measure are approximately spherically-symmetric. When can we say that a probability measure μ on \mathbb{R}^d is approximately radially-symmetric?

We need some notation. Let μ be a finite measure on a measurable space Ω . For a subset $A \subseteq \Omega$ with $\mu(A) > 0$ we write $\mu|_A$ for the conditioning of μ on A , i.e.,

$$\mu|_A(B) = \mu(A \cap B) / \mu(A)$$

for any measurable set $B \subseteq \Omega$. Write S^{d-1} for the unit sphere centered at the origin in \mathbb{R}^d . The uniform probability measure on the sphere S^{d-1} is denoted by σ_{d-1} . For two probability measures μ and ν on the sphere S^{d-1} and $1 \leq p < \infty$ we write $W_p(\mu, \nu)$ for the L^p Monge-Kantorovich transportation distance between μ and ν in the sphere S^{d-1} endowed with the geodesic distance (see, e.g., [33] or Section 2 below). The metrics W_p are all equivalent (we have $W_1 \leq W_p \leq \pi W_1^{1/p}$) and they metrize weak convergence of probability measures. For an interval $J \subset (0, \infty)$ we consider the spherical shell $S(J) = \{x \in \mathbb{R}^d; |x| \in J\}$, where $|\cdot|$ is the standard Euclidean norm in \mathbb{R}^d . The radial projection in \mathbb{R}^d is the map $\mathcal{R}(x) = x/|x|$. An interval is either open, closed or half-open and half-closed.

Definition 1.2 (Gromov [12]) *Let μ be a Borel probability measure on \mathbb{R}^d and let $\varepsilon > 0$. We say that μ is “ ε -radial” if for any interval $J \subset (0, \infty)$ with $\mu(S(J)) \geq \varepsilon$, we have*

$$W_1(\mathcal{R}_*(\mu|_{S(J)}), \sigma_{d-1}) \leq \varepsilon.$$

That is, when we condition μ on any spherical shell that contains at least an ε -fraction of the mass, and then project radially to the sphere, we obtain an approximation to the uniform probability measure on the sphere in the transportation-metric sense.

Note that this definition is scale-invariant. We think of the dimension n from Definition 1.1 as a very large number, tending to infinity. On the other hand, we usually view the dimension d in Definition 1.2 as being fixed, and typically not very large. The case $d = 1$ of Definition 1.2 corresponds to the measure being approximately even. We are not sure whether Dirac’s measure δ_0 is a good example of an ε -radial measure. An ε -radial measure μ is said to be “proper” if $\mu(\{0\}) = 0$. Our main theorem reads as follows:

Theorem 1.3 *There exists a universal constant $C > 0$ for which the following holds: Let $0 < \varepsilon < 1$ and let d, n be positive integers. Suppose that*

$$n \geq (C/\varepsilon)^{Cd}. \tag{2}$$

Then, for any decent probability measure μ on \mathbb{R}^n , there exists a linear map $T : \mathbb{R}^n \rightarrow \mathbb{R}^d$ such that $T_(\mu)$ is ε -radial proper.*

Furthermore, let $\eta > 0$ be such that $\eta^{-1} \geq (C/\varepsilon)^{Cd}$. Then, for any η -decent probability measure μ on \mathbb{R}^n , there exists a linear map $T : \mathbb{R}^n \rightarrow \mathbb{R}^d$ such that $T_(\mu)$ is ε -radial proper.*

Gromov has a topological proof for the cases $d = 1, 2$ of Theorem 1.3, which does not seem to generalize to higher dimensions [12], [22]. Theorem 1.3 is tight, up to the value of the constant C , as demonstrated by the example where μ is distributed uniformly on n linearly independent vectors: In this case μ is decent, but for any linear map T and an interval J , the discrete measure $\mathcal{R}((T_*\mu)|_{S(J)})$ is composed of at most n atoms. It is not difficult to see that when the support of ν contains no more than $\varepsilon^{-(d-1)}$ points, we have the lower bound $W_1(\nu, \sigma_{d-1}) \geq c\varepsilon$, for a certain universal constant $c > 0$. It is desirable to find the best constant in the exponent in Theorem 1.3, perhaps also with respect to other notions of ε -radial measures.

The conclusion of Theorem 1.3 also holds when the measure μ is assumed to be only “weakly ε -decent”, except that $T_*(\mu)$ is no longer necessarily proper. Another possibility in this context is to allow affine maps in Theorem 1.3 in place of linear maps, and obtain a measure $T_*(\mu)$ which is ε -radial proper. (It is also possible to modify Definition 1.1 slightly, and require that (1) hold for all affine subspaces of dimension at least one. The effect of such a modification is minor, since an ε -decent measure will remain at most 2ε -decent after such a change).

The conclusion of Theorem 1.3 does not necessarily hold for non-decent measures, even when their support spans the entire \mathbb{R}^n : Let e_1, \dots, e_n be linearly independent vectors in \mathbb{R}^n , and consider the probability measure $\mu = (1 - 2^{-n})^{-1} \sum_{i=1}^n 2^{-i} \delta_{e_i}$, where δ_x is Dirac’s unit mass at $x \in \mathbb{R}^n$. Then μ is not decent, and none of the two-dimensional marginals of μ are ε -radial proper, for $\varepsilon = 1/10$.

As in Milman’s proof of Dvoretzky’s theorem (see [23]), Theorem 1.3 will be proved by demonstrating that a random linear map T works with positive probability, once the measure μ is put in the right “position”. That is, we first push-forward μ under an appropriate invertible linear map in \mathbb{R}^n , which is non-random, and only then do we project the resulting probability measure to a random d -dimensional subspace, distributed uniformly in the Grassmannian. The measure μ is in the correct “position” when the covariance matrix of $\mathcal{R}_*\mu$ is proportional to the identity matrix. If we assume that the covariance matrix of μ itself is proportional to the identity, then a random projection will not work, in general, with high probability (compare with Sudakov’s theorem; see [29] or the presentation in Bobkov [4]).

Here is an outline of the proof of Theorem 1.3 and also of the structure of this manuscript: In Section 5 we use the non-degeneracy conditions from Definition 1.1 in order to guarantee the existence of the initial linear transformation that puts μ in the right “position”. Once we know that the covariance matrix of $\mathcal{R}_*\mu$ is approximately a scalar matrix, we prove that the measure μ may be decomposed into many almost-orthogonal ensembles. Each such ensemble is simply a discrete probability measure, uniform on a collection of approximately-orthogonal vectors in \mathbb{R}^n that are not necessarily of the same length. This decomposition, which essentially appeared earlier in the work of Bourgain, Lindenstrauss and Milman [6], is discussed in Section 4. Section 3 is concerned with the analysis of a single ensemble of our decomposition. As it turns out, a random projection works with high probability, and transforms the discrete measure into an almost-radial one. Section 2 contains a preliminary discussion regarding ε -radial measures and the transportation metric. The proof of Theorem 1.3 is completed in Section 6, in which we

also make some related comments and prove the following corollary to Theorem 1.3.

Corollary 1.4 *There exists a sequence $R_n \rightarrow \infty$ with the following property: Let μ be a decent probability measure on \mathbb{R}^n . Then, there exists a non-zero linear functional $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ such that*

$$\mu(\{x; \varphi(x) \geq tM\}) \geq c \exp(-Ct^2) \quad \text{for all } 0 \leq t \leq R_n$$

and

$$\mu(\{x; \varphi(x) \leq -tM\}) \geq c \exp(-Ct^2) \quad \text{for all } 0 \leq t \leq R_n$$

where $M > 0$ is a median, that is,

$$\mu(\{x; |\varphi(x)| \leq M\}) \geq 1/2 \quad \text{and} \quad \mu(\{x; |\varphi(x)| \geq M\}) \geq 1/2 \quad (3)$$

and $c, C > 0$ are universal constants. Moreover, one may take $R_n = c(\log n)^{1/4}$.

In other words, any high-dimensional probability measure has super-gaussian marginals. Furthermore, as is evident from the proof, *most* of the marginals are super-gaussian when the measure is in the right “position”. In the case of independent random variables, Corollary 1.4 essentially goes back to Kolmogorov [20]. See also Nagaev [25].

In Section 7 we formulate our results in an infinite-dimensional setting. Unless stated otherwise, throughout the text the letters c, C, C', \tilde{c} etc. stand for various positive universal constants, whose value may change from one instance to the next. We usually denote by lower-case $c, \tilde{c}, c', \bar{c}$ etc. positive universal constants that are assumed sufficiently small, and by upper-case $C, \tilde{C}, C', \bar{C}$ etc. sufficiently large universal constants. We write $x \cdot y$ for the usual scalar product of $x, y \in \mathbb{R}^n$.

Acknowledgments. I would like to thank Misha Gromov for his interest in this work and for introducing me to the problem, to Vitali Milman for encouraging me to write this note, to Boris Tsirelson for his explanations regarding measures on infinite-dimensional linear spaces, to Sasha Sodin for reading a preliminary version of this text and to Semyon Alesker, Noga Alon and Apostolos Giannopoulos for related discussions.

2 Transportation distance and empirical distributions

Let (X, ρ) be a metric space and let μ_1, μ_2 be Borel probability measures on X . A *coupling* of μ_1 and μ_2 is a Borel probability measure γ on $X \times X$ whose first marginal is μ_1 and whose second marginal is μ_2 , that is, $(P_1)_* \gamma = \mu_1$ and $(P_2)_* \gamma = \mu_2$ where $P_1(x, y) = x$ and $P_2(x, y) = y$. The L^1 Monge-Kantorovich distance is

$$W_1(\mu_1, \mu_2) = \inf_{\gamma} \int_{X \times X} \rho(x, y) d\gamma(x, y)$$

where the infimum runs over all couplings γ of μ_1 and μ_2 . Then W_1 is a metric, and it satisfies the convexity relation

$$W_1(\lambda\mu_1 + (1 - \lambda)\mu_2, \nu) \leq \lambda W_1(\mu_1, \nu) + (1 - \lambda)W_1(\mu_2, \nu) \quad (4)$$

for any $0 < \lambda < 1$ and probability measures μ_1, μ_2, ν on X . The Kantorovich-Rubinstein duality theorem (see [33, Theorem 1.14]) states that

$$W_1(\mu, \nu) = \sup_{\varphi} \int_X \varphi d[\mu - \nu] \quad (5)$$

where the supremum runs over all 1-Lipschitz functions $\varphi : X \rightarrow \mathbb{R}$ (i.e., $|\varphi(x) - \varphi(y)| \leq \rho(x, y)$ for all $x, y \in X$). We are concerned mostly with the case where the metric space X is the Euclidean sphere S^{n-1} with the metric $\rho(x, y)$ being the geodesic distance in S^{d-1} , i.e., $\cos \rho(x, y) = x \cdot y$. Denote by $\mathbb{M}(S^{d-1})$ the space of Borel probability measures on S^{d-1} , endowed with the weak* topology and the corresponding Borel σ -algebra. Similarly, $\mathbb{M}(\mathbb{R}^d)$ is the space of Borel probability measures on \mathbb{R}^d , endowed with the weak* topology (convergence of integrals of compactly-supported continuous functions) and σ -algebra. A measure here always means a non-negative measure. The total variation distance between two measures μ and ν on a measurable space Ω is

$$d_{TV}(\mu, \nu) = \sup_{A \subseteq \Omega} |\mu(A) - \nu(A)|$$

where the supremum runs over all measurable sets $A \subseteq \Omega$. Clearly, for $\mu, \nu \in \mathbb{M}(S^{d-1})$,

$$W_1(\mu, \nu) \leq \pi d_{TV}(\mu, \nu) \leq \pi. \quad (6)$$

Additionally, $d_{TV}(S_*\mu, S_*\nu) \leq d_{TV}(\mu, \nu)$ for any measures μ, ν and a measurable map S . When S is a λ -Lipschitz map between metric spaces, we obtain the inequality $W_1(S_*\mu, S_*\nu) \leq \lambda W_1(\mu, \nu)$. The following lemma is an obvious consequence of (4) and (6), via Jensen's inequality.

Lemma 2.1 *Let d be a positive integer, $0 \leq \varepsilon < 1$ and $\mu \in \mathbb{M}(S^{d-1})$. Suppose that we are given a "random probability measure" on S^{d-1} . That is, let λ be a probability measure on a measurable space Ω , and suppose that with any $\alpha \in \Omega$ we associate a probability measure $\mu_\alpha \in \mathbb{M}(S^{d-1})$ such that the map $\Omega \ni \alpha \mapsto \mu_\alpha \in \mathbb{M}(S^{d-1})$ is measurable. Assume that*

$$d_{TV} \left(\mu, \int_{\Omega} \mu_\alpha d\lambda(\alpha) \right) \leq \varepsilon.$$

Then,

$$W_1(\mu, \sigma_{d-1}) \leq \int_{\Omega} W_1(\mu_\alpha, \sigma_{d-1}) d\lambda(\alpha) + \pi\varepsilon \leq \sup_{\alpha \in \Omega} W_1(\mu_\alpha, \sigma_{d-1}) + 4\varepsilon.$$

Recall that $\mu|_X$ stands for the conditioning of μ to X .

Lemma 2.2 *Suppose that μ and ν are finite measures on a measurable space Ω and let $\varepsilon > 0$. Let $X \subseteq \Omega$ be such that $\nu(X) > \varepsilon$. Suppose that*

$$|\mu(A) - \nu(A)| \leq \varepsilon \quad \text{for all } A \subseteq X.$$

Then $d_{TV}(\mu|_X, \nu|_X) \leq 2\varepsilon/\nu(X)$.

Proof: For any $A \subseteq X$,

$$|\nu|_X(A) - \mu|_X(A)| = \left| \frac{\nu(A) - \mu(A)}{\nu(X)} + \frac{\mu(A)}{\mu(X)} \cdot \frac{\mu(X) - \nu(X)}{\nu(X)} \right| \leq \frac{\varepsilon}{\nu(X)} + \frac{\varepsilon}{\nu(X)} = \frac{2\varepsilon}{\nu(X)},$$

since $\mu(A) \leq \mu(X)$. \square

Next we describe a few simple properties of ε -radial measures.

Lemma 2.3 *Let d be a positive integer and $0 < \varepsilon < 1/2$. Let μ, ν be Borel probability measures on \mathbb{R}^d . Additionally, assume that we are given a “random probability measure” on \mathbb{R}^d . That is, let λ be a probability measure on a measurable space Ω . Suppose that with any $\alpha \in \Omega$ we associate a measure $\mu_\alpha \in \mathbb{M}(\mathbb{R}^d)$ such that the map $\Omega \ni \alpha \mapsto \mu_\alpha \in \mathbb{M}(\mathbb{R}^d)$ is measurable.*

- (a) *Suppose that μ is ε -radial and that $d_{TV}(\mu, \nu) \leq \varepsilon^2$. Then ν is 5ε -radial.*
- (b) *Suppose that μ_α is ε -radial for any $\alpha \in \Omega$. Assume that $\mu = \int_\Omega \mu_\alpha d\lambda(\alpha)$. Then μ is $4\sqrt{\varepsilon}$ -radial.*
- (c) *Suppose that $A \subseteq \Omega$ satisfies $\lambda(A) \geq 1 - \varepsilon$, and μ_α is ε -radial for any $\alpha \in A$. Assume that $\mu = \int_\Omega \mu_\alpha d\lambda(\alpha)$. Then μ is $20\sqrt{\varepsilon}$ -radial.*

Proof: Begin with (a). Let $J \subset (0, \infty)$ be an interval with $\nu(S(J)) \geq 2\varepsilon$, where $S(J) = \{x \in \mathbb{R}^d; |x| \in J\}$ is a spherical shell. Denote $\nu_J = \nu|_{S(J)}$ and $\mu_J = \mu|_{S(J)}$. Since $d_{TV}(\mu, \nu) \leq \varepsilon^2$, we may apply Lemma 2.2 with ε^2 and $X = S(J)$. We conclude that $d_{TV}(\mu_J, \nu_J) \leq 2(\varepsilon^2)/2\varepsilon = \varepsilon$. Consequently,

$$d_{TV}(\mathcal{R}_*(\mu_J), \mathcal{R}_*(\nu_J)) \leq \varepsilon. \quad (7)$$

Since μ is ε -radial and $\mu(S(J)) \geq 2\varepsilon - \varepsilon^2 \geq \varepsilon$, then $W_1(\mathcal{R}_*(\mu_J), \sigma_{d-1}) \leq \varepsilon$ according to Definition 1.2. From (6), (7) and the triangle inequality, $W_1(\mathcal{R}_*(\nu_J), \sigma_{d-1}) \leq (\pi+1)\varepsilon \leq 5\varepsilon$. This completes the proof of (a).

We move to the proof of (b). Let $J \subset (0, \infty)$ be an interval with $\mu(S(J)) \geq 4\sqrt{\varepsilon}$. Let $X = \{\alpha \in \Omega; \mu_\alpha(S(J)) \geq \varepsilon\}$. Denote $\nu = \int_X \mu_\alpha d\lambda(\alpha)$, a finite Borel measure on \mathbb{R}^n . Then for any $A \subseteq S(J)$,

$$|\mu(A) - \nu(A)| = \int_{\Omega \setminus X} \mu_\alpha(A) d\lambda(\alpha) \leq \int_{\Omega \setminus X} \mu_\alpha(S(J)) d\lambda(\alpha) \leq \varepsilon \lambda(\Omega \setminus X) \leq \varepsilon. \quad (8)$$

Denote $\mu_J = \mu|_{S(J)}$ and $\nu_J = \nu|_{S(J)}$. From (8) and Lemma 2.2,

$$d_{TV}(\mu_J, \nu_J) \leq 2\varepsilon/(4\sqrt{\varepsilon}) = \sqrt{\varepsilon}/2. \quad (9)$$

Note that $\nu_J = \int_X \mu_\alpha|_{S(J)} d\lambda'(\alpha)$ where λ' is a certain probability measure on X . Since μ_α is ε -radial and $\mu_\alpha(S(J)) \geq \varepsilon$ for $\alpha \in X$, then from Definition 1.2,

$$W_1(\mathcal{R}_*(\mu_\alpha|_{S(J)}), \sigma_{d-1}) \leq \varepsilon \quad \text{for } \alpha \in X. \quad (10)$$

We have $\mathcal{R}_*(\nu_J) = \int_X \mathcal{R}_*(\mu_\alpha|_{S(J)}) d\lambda'(\alpha)$. Thus, (10) and Lemma 2.1 yield that $W_1(\mathcal{R}_*(\nu_J), \sigma_{d-1}) \leq \varepsilon$. Combining the last inequality with (6) and (9), we see that

$$W_1(\mathcal{R}_*(\mu_J), \sigma_{d-1}) \leq 4d_{TV}(\mathcal{R}_*(\mu_J), \mathcal{R}_*(\nu_J)) + W_1(\mathcal{R}_*(\nu_J), \sigma_{d-1}) \leq 2\sqrt{\varepsilon} + \varepsilon \leq 4\sqrt{\varepsilon}.$$

Since $\mu_J = \mu|_{S(J)}$, and $J \subset (0, \infty)$ is an arbitrary interval with $\mu(S(J)) \geq 4\sqrt{\varepsilon}$, the assertion (b) is proven.

To prove (c), denote $\nu = \int_A \mu_\alpha d\lambda|_A(\alpha)$, a probability measure on \mathbb{R}^d . Then ν is $4\sqrt{\varepsilon}$ -radial, according to (b). Furthermore, clearly $d_{TV}(\lambda|_A, \lambda) = 1 - \lambda(A) \leq \varepsilon$, and hence $d_{TV}(\mu, \nu) \leq \varepsilon \leq (4\sqrt{\varepsilon})^2$. According to part (a), the measure μ is $20\sqrt{\varepsilon}$ -radial, and (c) is proven. \square

Probability measures are the protagonists of this text. Some of our constructions of probability measures are probabilistic in nature. To avoid confusion, we try to distinguish sharply between the measures themselves, and the randomness used in their construction. Whenever we have objects that are declared random (for instance, random vectors in S^{d-1}), all statements containing probability estimates or using the symbol \mathbb{P} refer to these random objects and only to them.

The crude bound in the following lemma is certainly a standard application of the so-called ‘‘empirical distribution method’’ (see, e.g., [5]). We were not able to find it in the literature, so a proof is provided. Recall that δ_x stands for the Dirac unit mass at the point x .

Lemma 2.4 *Let d, N be positive integers, and let X_1, \dots, X_N be independent random vectors, distributed uniformly on S^{d-1} . Denote $\mu = N^{-1} \sum_{i=1}^N \delta_{X_i}$. Then, with probability greater than $1 - C \exp(-c\sqrt{N})$ of selecting X_1, \dots, X_N ,*

$$W_1(\mu, \sigma_{d-1}) \leq \frac{C}{N^{c/d}}$$

where $C > 1$ and $0 < c < 1$ are universal constants.

Proof: Denote by \mathcal{F} the class of all 1-Lipschitz functions $\varphi : S^{d-1} \rightarrow \mathbb{R}$ such that $\int \varphi d\sigma_{d-1} = 0$. Note that $\sup |\varphi| \leq \pi$ for any $\varphi \in \mathcal{F}$. According to (5),

$$W_1(\mu, \sigma_{d-1}) = \sup_{\varphi \in \mathcal{F}} \int_{S^{d-1}} \varphi d\mu. \quad (11)$$

Let $\varepsilon > 0$ be a parameter to be specified later on. A subset $\mathcal{N} \subset S^{d-1}$ is an ε -net if for any $x \in S^{d-1}$ there exists $y \in \mathcal{N}$ with $\rho(x, y) \leq \varepsilon$. Let \mathcal{N} be an ε -net of cardinality $\#(\mathcal{N}) \leq (C/\varepsilon)^d$ (see, e.g., [26, Lemma 4.16]). For $\varphi \in \mathcal{F}$ denote

$$\tilde{\varphi}(x) = \min_{y \in \mathcal{N}} (\varepsilon \lceil \varphi(y)/\varepsilon \rceil + \rho(x, y)),$$

where $\lceil a \rceil$ is the minimal integer that is not smaller than a . Then $\tilde{\varphi}$ is a 1-Lipschitz function, as a minimum of 1-Lipschitz functions. It is easily verified that $\varphi \leq \tilde{\varphi} \leq$

$\varphi + 3\varepsilon$. Denote $\varphi^\circ(x) = \tilde{\varphi}(x) - \int \tilde{\varphi}(y) d\sigma_{d-1}(y)$. Then $\varphi^\circ \in \mathcal{F}$ for any $\varphi \in \mathcal{F}$, and $\sup |\varphi^\circ - \varphi| \leq 3\varepsilon$. Hence,

$$W_1(\mu, \sigma_{d-1}) = \sup_{\varphi \in \mathcal{F}} \int_{S^{d-1}} \varphi d\mu \leq 3\varepsilon + \sup_{\varphi \in \mathcal{F}} \int_{S^{d-1}} \varphi^\circ d\mu = 3\varepsilon + \sup_{\varphi \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N \varphi^\circ(X_i). \quad (12)$$

Denote $\tilde{\mathcal{F}} = \{\tilde{\varphi}; \varphi \in \mathcal{F}\}$ and $\mathcal{F}^\circ = \{\varphi^\circ; \varphi \in \mathcal{F}\}$. These sets are finite. In fact, as each $\varphi \in \tilde{\mathcal{F}}$ is determined by the restriction $\varphi|_{\mathcal{N}}$, we have

$$\#(\mathcal{F}^\circ) \leq \#(\tilde{\mathcal{F}}) \leq \left(\frac{2\pi}{\varepsilon} + 1\right)^{\#(\mathcal{N})} \leq \exp((C/\varepsilon)^{2d}). \quad (13)$$

Fix $\varphi^\circ \in \mathcal{F}^\circ$. Then φ° is a 1-Lipschitz function on the sphere S^{d-1} with $\int \varphi^\circ d\sigma_{d-1} = 0$. According to Lévy's lemma (see Milman and Schechtman [24, Section 2 and Appendix V]), for any $i = 1, \dots, N$,

$$\mathbb{P}(|\varphi^\circ(X_i)| \geq t) \leq C \exp(-ct^2 d) \quad \forall t \geq 0,$$

where \mathbb{P} refers, of course, to the probability of choosing the random vectors X_1, \dots, X_N . From Bernstein's inequality (see, e.g., Bourgain, Lindenstrauss and Milman [6, Proposition 1]),

$$\mathbb{P}\left(\left|\frac{1}{N} \sum_{i=1}^N \varphi^\circ(X_i)\right| \geq t\right) \leq C' \exp(-c't^2 Nd) \quad \forall t \geq 0. \quad (14)$$

Set $t = \varepsilon$ in (14). From (13) and (14),

$$\mathbb{P}\left(\sup_{\varphi^\circ \in \mathcal{F}^\circ} \left|\frac{1}{N} \sum_{i=1}^N \varphi^\circ(X_i)\right| \geq \varepsilon\right) \leq C' \exp((C/\varepsilon)^{2d} - c'\varepsilon^2 Nd). \quad (15)$$

We now select $\varepsilon = CN^{-1/(2d+2)}$, for a sufficiently large universal constant $C > 0$. Substitute the value of ε in (15) and apply (12), to deduce that

$$W_1(\mu, \sigma_{d-1}) \leq 3\varepsilon + \sup_{\varphi^\circ \in \mathcal{F}^\circ} \frac{1}{N} \sum_{i=1}^N \varphi^\circ(X_i) \leq 4\varepsilon \leq C' N^{-1/(2d+2)},$$

with probability greater than $1 - C' \exp(-c'N^{d/(d+1)})$ of selecting X_1, \dots, X_N . \square

Remark. The discrepancy of $\mu \in \mathbb{M}(S^{d-1})$ is defined as

$$D(\mu) = \sup_B |\mu(B) - \sigma_{d-1}(B)|$$

where the supremum runs over all geodesic balls $B \subseteq S^{d-1}$. A result analogous to Lemma 2.4 for discrepancy appears in Beck and Chen [3, Section 7.4]. It is possible to adapt our technique to suit the discrepancy metric, and some of its variants, in place of W_1 . In fact, the only properties of the metric W_1 that are used in our proof are Lemma 2.4 and (4) and (6) above. Our method, of course, works for the W_p metrics as long as $1 \leq p < \infty$, but it does not seem to apply for the W_∞ metric. The W_∞ metric induces a topology that is much stronger than weak convergence, and it is not even weaker than convergence in norm.

3 Isotropic Gaussians

A centered Gaussian random vector in \mathbb{R}^d is a random vector whose density is proportional to $x \mapsto \exp(-Mx \cdot x)$ for a positive definite matrix M . A centered Gaussian random vector is said to be isotropic if M is a scalar matrix. It is called standard if $M = Id/2$, where Id is the identity matrix. Recall that \mathcal{R} stands for radial projection.

Lemma 3.1 *Let d, N be positive integers and let Z_1, \dots, Z_N be independent isotropic Gaussian random vectors in \mathbb{R}^d . Denote $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{Z_i}$.*

Then, with probability greater than $1 - C \exp(-cN^{1/4})$ of selecting the Z_i 's, the measure μ is δ -radial, for $\delta = CN^{-c/d}$. Here, $C > 1$ and $0 < c < 1$ are universal constants.

Proof: Set $\varepsilon = 5/N^{1/4}$. We may assume that $\varepsilon \leq 1/10$, as otherwise the conclusion of the lemma is obvious for a suitable choice of universal constants $c, C > 0$. The central observation is that the radii $|Z_1|, \dots, |Z_N|$ are independent of the angular parts $\mathcal{R}(Z_1), \dots, \mathcal{R}(Z_N)$, and that the random vectors $\mathcal{R}(Z_1), \dots, \mathcal{R}(Z_N)$ are independent random vectors that are distributed uniformly on S^{d-1} .

With probability one, none of the $|Z_i|$'s are zero, and there are no $i \neq j$ with $|Z_i| = |Z_j|$. We condition on the values $|Z_1|, \dots, |Z_N|$, which are assumed to be distinct and non-zero. For an interval $J \subset (0, \infty)$ write

$$Z(J) = \{i; |Z_i| \in J\} \quad \text{and} \quad w(J) = \#(Z(J)).$$

Denote $k = \lceil 1/\varepsilon^2 \rceil$, and let $J_1, \dots, J_k \subset (0, \infty)$ be disjoint intervals such that

$$w(J_i) = \lfloor Ni/k \rfloor - \lfloor N(i-1)/k \rfloor \quad \text{for } i = 1, \dots, k.$$

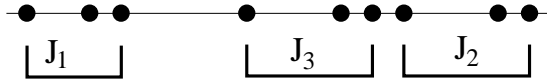


Figure 1

Since $\varepsilon^2 N \geq 2$, then $\varepsilon^2 N/2 \leq w(J_i) \leq 2\varepsilon^2 N$ for any i . For an interval $J \subset (0, \infty)$ with $w(J) \neq 0$, denote

$$\mu_J = \frac{1}{w(J)} \sum_{j \in Z(J)} \delta_{\mathcal{R}(Z_j)}.$$

Fix $i = 1, \dots, k$. We abbreviate $\mu_i = \mu_{J_i}$. Since $\{\mathcal{R}(Z_j)\}_{j \in Z(J_i)}$ is a collection of $w(J_i)$ independent random vectors, distributed uniformly on the sphere, then Lemma 2.4 applies, and yields,

$$\mathbb{P} \left(W_1(\mu_i, \sigma_{d-1}) \leq \frac{C}{w(J_i)^{c/d}} \right) \geq 1 - C \exp(-c\sqrt{w(J_i)}).$$

We now let i vary. Since $w(J_i)$ has the order of magnitude of $\varepsilon^2 N$, then,

$$\mathbb{P} \left(\max_{i=1, \dots, k} W_1(\mu_i, \sigma_{d-1}) \leq \frac{C}{(\varepsilon^2 N)^{c/d}} \right) \geq 1 - Ck \exp(-c\varepsilon\sqrt{N}). \quad (16)$$

Write \mathcal{I} for the collection of all non-empty intervals in $(0, \infty)$. Fix an interval J with $w(J) \geq N\varepsilon$. Let $J_{i_1}, \dots, J_{i_\ell}$ be all the intervals among the J_i 's that are contained in J . Then $J_{i_1} \cup \dots \cup J_{i_\ell}$ covers all but at most $4\varepsilon^2 N$ of the $|Z_i|$'s that belong to the interval J . Therefore,

$$d_{TV} \left(\mu_J, \sum_{j=1}^{\ell} \lambda_j \mu_{i_j} \right) \leq \frac{4\varepsilon^2 N}{N\varepsilon} = 4\varepsilon$$

where $\lambda_1, \dots, \lambda_\ell$ are appropriate non-negative coefficients that add to one. According to Lemma 2.1,

$$W_1(\mu_J, \sigma_{d-1}) \leq \max_{i=1, \dots, k} W_1(\mu_i, \sigma_{d-1}) + 20\varepsilon \quad \text{for all } J \in \mathcal{I} \text{ with } w(J) \geq N\varepsilon.$$

We thus conclude from (16) that with probability at least $1 - Ck \exp(-c\varepsilon\sqrt{N})$,

$$W_1(\mu_J, \sigma_{d-1}) \leq \frac{C}{(\varepsilon^2 N)^{c/d}} + 20\varepsilon \quad \text{for all } J \in \mathcal{I} \text{ with } w(J) \geq N\varepsilon. \quad (17)$$

The latter probability bound is valid under the conditioning on $|Z_1|, \dots, |Z_N|$, and it holds for all possible values of $|Z_1|, \dots, |Z_N|$, up to measure zero. Hence the aforementioned probability bound for (17) also holds with no conditioning at all. Recall that we write $S(J) = \{x \in \mathbb{R}^n; |x| \in J\}$, and note that $\mu_J = \mathcal{R}_*(\mu|_{S(J)})$ and $w(J) = N\mu(S(J))$. Since $\varepsilon\sqrt{N} \geq N^{1/4}$, then (17) translates as follows: With probability greater than $1 - C \exp(-cN^{1/4})$ of selecting Z_1, \dots, Z_N ,

$$W_1(\mathcal{R}_*(\mu|_{S(J)}), \sigma_{d-1}) \leq CN^{-c/d} + C\varepsilon \quad \text{for all } J \in \mathcal{I} \text{ with } \mu(S(J)) \geq \varepsilon.$$

This means that μ is $C(N^{-c/d} + \varepsilon)$ -radial with probability greater than $1 - C \exp(-cN^{1/4})$. Since $N^{-c/d} + \varepsilon \leq C'N^{-c'/d}$, the lemma is proven. \square

Lemma 3.2 *Let k be a positive integer. For an invertible $k \times k$ matrix A , write γ_A for the probability measure on \mathbb{R}^k whose density is proportional to $x \mapsto \exp(-|Ax|^2/2)$. Then, for any $k \times k$ invertible matrices A and B ,*

$$d_{TV}(\gamma_A, \gamma_B) \leq Ck \|BA^{-1} - Id\|$$

where Id is the identity matrix, $\|\cdot\|$ is the operator norm, and $C > 0$ is a universal constant.

Proof: Let X be a standard Gaussian random vector in \mathbb{R}^k . Then $\gamma_A(U) = \mathbb{P}(A^{-1}X \in U)$ for any measurable set $U \subseteq \mathbb{R}^k$. Therefore,

$$d_{TV}(\gamma_A, \gamma_B) = \sup_{U \subseteq \mathbb{R}^k} |\mathbb{P}(X \in U) - \mathbb{P}(AB^{-1}X \in U)| = d_{TV}(\gamma_{BA^{-1}}, \gamma_{Id}).$$

Denote $M = BA^{-1}$, write $\gamma = \gamma_{Id}$ and set $\varepsilon = \|M - Id\| = \sup_{|x|=1} |Mx - x|$. We write $\varphi_M(x) = (\det M)(2\pi)^{-d/2} \exp(-|Mx|^2/2)$ for the density of γ_M and similarly φ stands for the density of γ . We may assume that $\varepsilon < 1/2$, as otherwise the conclusion

of the lemma is trivial. Then $||Mx|^2 - |x|^2| \leq 3\varepsilon|x|^2$ for any $x \in \mathbb{R}^k$, and also $(1 + 2\varepsilon)^{-k} \leq \det M \leq (1 + \varepsilon)^k$. Therefore,

$$\begin{aligned} & |\varphi(x) - \varphi_M(x)| \\ &= \varphi(x) \left| 1 - (\det M) \exp\left(\frac{|x|^2 - |Mx|^2}{2}\right) \right| \leq \varphi(x) [(1 + 2\varepsilon)^k \exp(3\varepsilon|x|^2) - 1] \end{aligned}$$

for any $x \in \mathbb{R}^k$. Consequently,

$$d_{TV}(\gamma, \gamma_M) = \frac{1}{2} \int_{\mathbb{R}^k} |\varphi(x) - \varphi_M(x)| dx \leq (1 + 2\varepsilon)^k \int_{\mathbb{R}^k} \exp(3\varepsilon|x|^2) \varphi(x) dx - 1.$$

However,

$$\int_{\mathbb{R}^k} \exp(3\varepsilon|x|^2) \varphi(x) dx = (2\pi)^{-d/2} \int_{\mathbb{R}^k} \exp(-|\sqrt{1 - 6\varepsilon}|x|^2/2) dx = (1 - 6\varepsilon)^{-k/2}.$$

We deduce that

$$d_{TV}(\gamma, \gamma_M) \leq (1 + 2\varepsilon)^k (1 - 6\varepsilon)^{-k/2} - 1 \leq Ck\varepsilon,$$

under the legitimate assumption that $\varepsilon < c/k$ (otherwise, there is nothing to prove). \square

Consider the map $\mathcal{E} : (\mathbb{R}^d)^N \rightarrow \mathbb{M}(\mathbb{R}^d)$ defined by

$$(\mathbb{R}^d)^N \ni (x_1, \dots, x_N) \xrightarrow{\mathcal{E}} \frac{1}{N} \sum_{i=1}^N \delta_{x_i}.$$

A Borel probability measure α on $(\mathbb{R}^d)^N$ thus induces the Borel probability measure $\mathcal{E}_*\alpha$ on the space $\mathbb{M}(\mathbb{R}^d)$. The next lemma is a small perturbation of Lemma 3.1.

Lemma 3.3 *Let d, N be positive integers, $\varepsilon > 0$. Let X_1, \dots, X_N be independent, standard Gaussian random vectors in \mathbb{R}^d . Let $(a_{ij})_{1 \leq j \leq i \leq N}$ be real numbers, with $a_{ii} \neq 0$ for all i , such that*

$$|a_{ij}| \leq \varepsilon |a_{ii}| \quad \text{for } j < i. \quad (18)$$

Denote $Y_i = \sum_{j \leq i} a_{ij} X_j$ and consider the probability measure $\mu = N^{-1} \sum_{i=1}^N \delta_{Y_i}$.

Then, with probability greater than $1 - C \exp(-cN^{1/4}) - CN^2 d^2 \varepsilon$ of selecting the random vectors X_1, \dots, X_N , the measure μ is δ -radial for $\delta = CN^{-c/d}$. Here, $C > 1$ and $0 < c < 1$ are universal constants.

Proof: Denote $Z_i = a_{ii} X_i$. The Z_i 's are independent, isotropic, Gaussian random vectors in \mathbb{R}^d . Denote by $U \subseteq \mathbb{M}(\mathbb{R}^d)$ the collection of all δ -radial probability measures, where $\delta = CN^{-c/d}$ is the same as in Lemma 3.1. Let α be the probability measure on $(\mathbb{R}^d)^N$ that is the joint distribution of Z_1, \dots, Z_N . According to Lemma 3.1,

$$(\mathcal{E}_*\alpha)(U) \geq 1 - C \exp(-cN^{1/4}).$$

Let β be the probability measure on $(\mathbb{R}^d)^N$ which is the joint distribution of Y_1, \dots, Y_N . To prove the lemma, we need to show that

$$(\mathcal{E}_* \beta)(U) \geq 1 - C \exp(-cN^{1/4}) - C' N^2 d^2 \varepsilon.$$

This would follow, if we could prove that

$$d_{TV}((\mathcal{E}_* \alpha), (\mathcal{E}_* \beta)) \leq d_{TV}(\alpha, \beta) \leq C' N^2 d^2 \varepsilon. \quad (19)$$

Let $k = dN$. Let A be the $k \times k$ matrix that represents the linear operator

$$\mathbb{R}^k = (\mathbb{R}^d)^N \ni (x_1, \dots, x_N) \mapsto (a_{11}x_1, \dots, a_{NN}x_N) \in (\mathbb{R}^d)^N = \mathbb{R}^k.$$

That is, A is a diagonal matrix, and the diagonal of A contains each number a_{ii} exactly d times. Denoting $X = (X_1, \dots, X_N) \in (\mathbb{R}^d)^N = \mathbb{R}^k$ and $Z = (Z_1, \dots, Z_N) \in (\mathbb{R}^d)^N = \mathbb{R}^k$, we see that $Z = AX$. Therefore, in the notation of Lemma 3.2, we have $\alpha = \gamma_{A^{-1}}$.

Similarly, let B be the $k \times k$ matrix that corresponds to the linear operator

$$(x_i)_{i=1, \dots, N} \mapsto \left(\sum_{j \leq i} a_{ij} x_j \right)_{i=1, \dots, N}$$

where x_1, \dots, x_N are vectors in \mathbb{R}^d . Denoting $Y = (Y_1, \dots, Y_N) \in \mathbb{R}^k$ we have $Y = BX$ and consequently $\beta = \gamma_{B^{-1}}$. Condition (18) implies that the off-diagonal elements of $A^{-1}B$ do not exceed ε in absolute value. The diagonal elements of $A^{-1}B$ are all ones. Hence $\|A^{-1}B - Id\| \leq k\varepsilon$, and according to Lemma 3.2,

$$d_{TV}(\alpha, \beta) = d_{TV}(\gamma_{B^{-1}}, \gamma_{A^{-1}}) \leq Ck \|A^{-1}B - Id\| \leq Ck^2 \varepsilon = CN^2 d^2 \varepsilon.$$

Thus (19) holds and the proof is complete. \square

4 Orthogonal decompositions

This section is concerned with probability measures on \mathbb{R}^n that may be decomposed as a mixture, whose components are mostly ensembles of approximately-orthogonal vectors. Later on, we will apply a random projection, and use Lemma 3.3 in order to show that the projection of most elements in the mixture is typically ε -radial for small $\varepsilon > 0$.

Definition 4.1 *Let ℓ, n be positive integers and let $\varepsilon > 0$. Let $v_1, \dots, v_\ell \in \mathbb{R}^n$ be non-zero vectors, and consider $v = (v_1, \dots, v_\ell)$, an ℓ -tuple of vectors. We say that v is “ ε -orthogonal” if there exist orthonormal vectors $w_1, \dots, w_\ell \in \mathbb{R}^n$ and real numbers $(a_{ij})_{1 \leq j \leq i \leq \ell}$ such that*

$$v_i = \sum_{j=1}^i a_{ij} w_j \quad \text{for } i = 1, \dots, \ell$$

and $|a_{ij}| \leq \varepsilon |a_{ii}|$ for $j < i$.

Note that (v_1, \dots, v_ℓ) is ε -orthogonal if and only if $(\mathcal{R}(v_1), \dots, \mathcal{R}(v_\ell))$ is ε -orthogonal. We write $\mathcal{O}_{\ell, \varepsilon} \subset (\mathbb{R}^n)^\ell$ for the collection of all ε -orthogonal ℓ -tuples $v = (v_1, \dots, v_\ell) \in (\mathbb{R}^n)^\ell$. For a subspace $E \subset \mathbb{R}^n$ denote by $Proj_E$ the orthogonal projection operator onto E in \mathbb{R}^n .

Lemma 4.2 *Let ℓ, n be positive integers. Suppose that μ is a Borel probability measure on the unit sphere S^{n-1} such that for any unit vector $\theta \in S^{n-1}$,*

$$\int_{S^{n-1}} (x \cdot \theta)^2 d\mu(x) \leq \frac{1}{\ell^{50}}.$$

Let X_1, \dots, X_ℓ be independent random vectors in S^{n-1} , all distributed according to μ . Then, with positive probability, (X_1, \dots, X_ℓ) is $(\ell^{-20}/2)$ -orthogonal.

Proof: We may assume that $\ell \geq 2$, otherwise the lemma is vacuously true. Write $W_1, \dots, W_\ell \in \mathbb{R}^n$ for the vectors obtained from X_1, \dots, X_ℓ through the Gram-Schmidt orthogonalization process. (If the X_i 's are not linearly independent, then some of the W_i 's might be zero). For $i \geq 2$ denote by E_i the linear span of X_1, \dots, X_{i-1} . Then, for $i \geq 2$,

$$\mathbb{E}|Proj_{E_i}(X_i)|^2 = \mathbb{E} \sum_{j=1}^{i-1} (X_i \cdot W_j)^2 = \mathbb{E} \sum_{j=1}^{i-1} \int_{S^{n-1}} (x \cdot W_j)^2 d\mu(x) \leq \frac{i-1}{\ell^{50}} \leq \ell^{-49},$$

as X_i is independent of W_1, \dots, W_{i-1} . By Chebyshev's inequality,

$$\mathbb{P} \{ \exists 2 \leq i \leq \ell; |Proj_{E_i}(X_i)| \geq \ell^{-21}/2 \} \leq (\ell-1) \frac{\ell^{-49}}{\ell^{-42}/4} \leq 4\ell^{-6} < 1.$$

Therefore, with positive probability, $|Proj_{E_i}(X_i)| < \ell^{-21}/2$ for all $i \geq 2$. In this event, the vectors X_1, \dots, X_ℓ are linearly independent, and W_1, \dots, W_ℓ are orthonormal vectors. Furthermore, in this event $a_{ii} := \sqrt{1 - |Proj_{E_i}(X_i)|^2} \geq 1 - \ell^{-21}/2$ while $a_{ij} := X_i \cdot W_j$ satisfies

$$|a_{ij}| \leq |Proj_{E_i}(X_i)| \leq \ell^{-21}/2 \quad \text{for } j < i.$$

Thus, with positive probability, $X_i = \sum_{j < i} a_{ij} W_j$ for all i , with $|a_{ij}| \leq (\ell^{-20}/2)|a_{ii}|$ for $j < i$ and with W_1, \dots, W_ℓ being orthonormal vectors. This completes the proof. Note that the ‘‘positive probability’’ is in fact greater than $1 - \ell^{-5}$. \square

The next lemma is essentially a measure-theoretic variant of a lemma going back to Bourgain, Lindenstrauss and Milman [6, Lemma 12], with the main difference being that the logarithmic dependence is improved upon to a power law. For two Borel measures μ and ν on a compact K we say that $\mu \leq \nu$ if

$$\int_K \varphi d\mu \leq \int_K \varphi d\nu \quad \text{for any continuous } \varphi : K \rightarrow [0, \infty). \quad (20)$$

Recall that a point is *not* in the support of a measure if and only if it has an open neighborhood of measure zero. We abbreviate $\mathcal{O}_\ell = \mathcal{O}_{\ell, \ell^{-20}}$. For $v = (v_1, \dots, v_\ell) \in \mathcal{O}_\ell$ denote

$$\mu_v = \frac{1}{\ell} \sum_{i=1}^{\ell} \delta_{v_i},$$

a Borel probability measure on \mathbb{R}^n (in the notation of the previous section, $\mu_v = \mathcal{E}(v)$). When $K \subseteq \mathbb{R}^n$, we write $\mathcal{O}_\ell(K) \subseteq \mathcal{O}_\ell$ for the collection of all $(v_1, \dots, v_\ell) \in \mathcal{O}_\ell$ with $v_i \in K$ for all i . Then $\mathcal{O}_\ell(K) = \mathcal{O}_\ell \cap K^\ell \subseteq (\mathbb{R}^n)^\ell$, and it is straightforward to verify that $\mathcal{O}_\ell(K)$ is compact whenever $K \subset \mathbb{R}^n$ is a compact that does not contain the origin.

Lemma 4.3 *Let ℓ, n be positive integers and let $0 < \varepsilon < 1/2$. Let μ be a Borel probability measure on \mathbb{R}^n with $\mu(\{0\}) = 0$. Assume that*

$$\sup_{\theta \in S^{n-1}} \int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_* \mu(x) < \frac{\varepsilon}{\ell^{50}}. \quad (21)$$

Then, there exists a Borel probability measure λ on \mathcal{O}_ℓ such that

$$d_{TV} \left(\mu, \int_{\mathcal{O}_\ell} \mu_v d\lambda(v) \right) < \varepsilon. \quad (22)$$

Proof: Since $\mu(\{0\}) = 0$ then for any $\delta > 0$ we may find a large punctured ball $K = \{x \in \mathbb{R}^n; r \leq |x| \leq R\}$ with $0 < r < R$ such that $\mu(K) \geq 1 - \delta$. We may assume that μ is supported on a compact set that does not contain the origin: Otherwise, replace μ with $\mu|_K$ for a large punctured ball $K \subset \mathbb{R}^n$ with $\mu(K) \geq 1 - \delta$ as above, and observe that $d_{TV}(\mu, \mu|_K) \leq \delta$, so the effect of the replacement on the inequalities (21) and (22) is bounded by δ , which can be made arbitrarily small. Write $K \subset \mathbb{R}^n$ for the support of μ , a compact which does not contain the origin. Denote by \mathcal{F} the collection of all Borel measures λ supported on $\mathcal{O}_\ell(K)$ for which

$$\int_{\mathcal{O}_\ell} \mu_v d\lambda(v) \leq \mu. \quad (23)$$

The condition (23) defining \mathcal{F} is closed in the weak* topology. Furthermore, $\lambda(\mathcal{O}_\ell) \leq 1$ for all $\lambda \in \mathcal{F}$ (use (23)), and take $\varphi \equiv 1$ in the definition (20)). Hence \mathcal{F} is a weak* closed subset of the unit ball of the Banach space of signed finite Borel measures on the compact $\mathcal{O}_\ell(K)$. From the Banach-Alaoglu theorem, \mathcal{F} is compact in the weak* topology. Therefore the continuous functional $\lambda \mapsto \lambda(\mathcal{O}_\ell)$ attains its maximum on \mathcal{F} at some $\lambda_0 \in \mathcal{F}$. Clearly $\lambda_0(\mathcal{O}_\ell) \leq 1$. To prove the lemma, it suffices to show that

$$\lambda_0(\mathcal{O}_\ell) > 1 - \varepsilon. \quad (24)$$

Indeed, if (24) holds, then we may define a probability measure $\lambda_1 = \lambda_0 + \tilde{\lambda}$, where $\tilde{\lambda}$ is any Borel measure on \mathcal{O}_ℓ of total mass $1 - \lambda_0(\mathcal{O}_\ell)$. Then clearly

$$d_{TV} \left(\mu, \int_{\mathcal{O}_\ell} \mu_v d\lambda_1(v) \right) \leq \tilde{\lambda}(\mathcal{O}_\ell) < \varepsilon,$$

and the lemma follows. We thus focus on the proof of (24). Assume by contradiction that (24) fails. Denote $\nu = \mu - \int_{\mathcal{O}_\ell} \mu_v d\lambda_0(v)$. Then ν is a non-negative Borel measure on $K \subset \mathbb{R}^n$, according to (23), and also $\nu \leq \mu$. Moreover, $\nu(K) \geq \varepsilon$, since we assume that

(24) fails. Denote $\tilde{\nu} = \nu/\nu(K)$, a probability measure on $K \subset \mathbb{R}^n$. Then $\tilde{\nu} \leq \nu/\varepsilon$ and hence $\mathcal{R}_*(\tilde{\nu}) \leq \mathcal{R}_*(\nu)/\varepsilon \leq \mathcal{R}_*(\mu)/\varepsilon$. For any unit vector $\theta \in S^{n-1}$,

$$\int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_* \tilde{\nu}(x) \leq \varepsilon^{-1} \int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_* \nu(x) \leq \varepsilon^{-1} \int_{\mathbb{R}^n} (x \cdot \theta)^2 d\mathcal{R}_* \mu(x) \leq \frac{1}{\ell^{50}},$$

from our assumption (21). Lemma 4.2 thus asserts the existence of $\tilde{x}_1, \dots, \tilde{x}_\ell \in S^{n-1}$ in the support of $\mathcal{R}_*(\tilde{\nu})$ such that $(\tilde{x}_1, \dots, \tilde{x}_\ell)$ is $(\ell^{-20}/2)$ -orthogonal. Consequently, there exist non-zero vectors $x_1, \dots, x_\ell \in \mathbb{R}^n$ in the support of ν such that (x_1, \dots, x_ℓ) is $(\ell^{-20}/2)$ -orthogonal. Let $U_1, \dots, U_\ell \subset \mathbb{R}^n$ be small open neighborhoods of x_1, \dots, x_ℓ , respectively, such that

$$(y_1, \dots, y_\ell) \in \mathcal{O}_{\ell, \ell^{-20}} = \mathcal{O}_\ell \quad \text{for all } y_1 \in U_1, \dots, y_\ell \in U_\ell.$$

The U_i 's are necessarily disjoint and $U_1 \times \dots \times U_\ell \subseteq \mathcal{O}_\ell$. Denote $\eta = \min_{i=1, \dots, \ell} \nu(U_i)$. Then $\eta > 0$, since U_i is an open neighborhood of the point x_i , and the point x_i belongs to the support of ν . We set $\nu_i = \nu|_{U_i}$, the conditioning of ν to U_i . Then ν_i is a probability measure supported on $K \subset \mathbb{R}^n$, and

$$\eta \nu_i \leq \nu(U_i) \nu_i \leq \nu = \mu - \int_{\mathcal{O}_\ell} \mu_v d\lambda_0(v) \quad \text{for } i = 1, \dots, \ell.$$

Therefore, also

$$\eta \int_{U_1 \times \dots \times U_\ell} \left(\frac{1}{\ell} \sum_{i=1}^{\ell} \delta_{y_i} \right) d\nu_1(y_1) \dots d\nu_\ell(y_\ell) = \eta \frac{\sum_{i=1}^{\ell} \nu_i}{\ell} \leq \mu - \int_{\mathcal{O}_\ell} \mu_v d\lambda_0(v).$$

Consequently, the non-negative measure $\lambda = \lambda_0 + \eta(\nu_1 \times \dots \times \nu_\ell)$ on $\mathcal{O}_\ell(K)$ satisfies (23). Hence $\lambda \in \mathcal{F}$, but $\lambda(\mathcal{O}_\ell) = \lambda_0(\mathcal{O}_\ell) + \eta > \lambda_0(\mathcal{O}_\ell)$, in contradiction to the maximality of λ_0 . We thus conclude that (24) must be true, and the lemma is proven. \square

A $d \times n$ matrix Γ will be called a ‘‘standard Gaussian random matrix’’ if the entries of Γ are independent standard Gaussian random variables (of mean zero and variance one). Suppose that w_1, \dots, w_ℓ are orthonormal vectors in \mathbb{R}^n and that Γ is a $d \times n$ standard Gaussian random matrix. Observe that in this case, $\Gamma(w_1), \dots, \Gamma(w_\ell)$ are independent standard Gaussian random vectors in \mathbb{R}^d .

Lemma 4.4 *Let $d \leq \ell \leq n$ be positive integers, let $0 < \varepsilon < 1$ and assume that*

$$\ell \geq (C/\varepsilon)^{Cd}. \tag{25}$$

Suppose that λ is a Borel probability measure on \mathcal{O}_ℓ , and denote $\mu = \int_{\mathcal{O}_\ell} \mu_v d\lambda(v)$. Let Γ be a $d \times n$ standard Gaussian random matrix.

Then, with positive probability of selecting the random matrix Γ , the measure $\Gamma_ \mu$ on \mathbb{R}^d is ε -radial. Here, $C > 1$ is a universal constant. (In fact, this probability is at least $1 - \ell^{-8}$.)*

Proof: Fix $v = (v_1, \dots, v_\ell) \in \mathcal{O}_\ell$. Consider the measure $\tilde{\mu}_v := \Gamma_*(\mu_v)$ on \mathbb{R}^d . Then $\tilde{\mu}_v = \Gamma_*\left(\frac{1}{\ell} \sum_{i=1}^{\ell} \delta_{v_i}\right) = \frac{1}{\ell} \sum_{i=1}^{\ell} \delta_{\Gamma(v_i)}$. Let $E(v)$ be the following event:

- The measure $\tilde{\mu}_v$ is $C\ell^{-c/d}$ -radial, where $C > 1$ and $0 < c < 1$ are the universal constants from Lemma 3.3.

Let us emphasize that for any $v \in \mathcal{O}_\ell$, the event $E(v)$ might either hold or not, depending on the Gaussian random matrix Γ . We are going to apply Lemma 3.3. Since $v \in \mathcal{O}_\ell = \mathcal{O}_{\ell, \ell^{-20}}$, then there exist orthonormal vectors $w_1, \dots, w_\ell \in \mathbb{R}^n$ and numbers a_{ij} such that $v_i = \sum_{j \leq i} a_{ij} w_j$ and $|a_{ij}| \leq \ell^{-20} |a_{ii}|$ for $j < i$, with $a_{ii} \neq 0$ for all i . Denote $X_i = \Gamma(w_i)$ and $Y_i = \sum_{j \leq i} a_{ij} X_j$. Then X_1, \dots, X_ℓ are independent standard Gaussian random vectors in \mathbb{R}^d , and $\tilde{\mu}_v = \ell^{-1} \sum_{i=1}^\ell \delta_{Y_i}$. We may thus apply Lemma 3.3 (with $N = \ell$ and $\varepsilon = \ell^{-20}$) and conclude that for any $v \in \mathcal{O}_\ell$,

$$\mathbb{P}(E(v)) \geq 1 - C \exp(-c\ell^{1/4}) - C\ell^2 d^2 \ell^{-20} \geq 1 - C'\ell^{-16}. \quad (26)$$

Let $\mathcal{F} \subseteq \mathcal{O}_\ell$ be the collection of all $v \in \mathcal{O}_\ell$ for which the event $E(v)$ holds. Then \mathcal{F} is a random subset of \mathcal{O}_ℓ (depending on the random matrix Γ). According to (26),

$$\mathbb{E}\lambda(\mathcal{F}) = \mathbb{E} \int_{\mathcal{O}_\ell} 1_{\mathcal{F}}(v) d\lambda(v) = \int_{\mathcal{O}_\ell} \mathbb{E} 1_{\mathcal{F}}(v) d\lambda(v) = \int_{\mathcal{O}_\ell} \mathbb{P}(E(v)) d\lambda(v) \geq 1 - C'\ell^{-16}$$

where $1_{\mathcal{F}}$ is the characteristic function of \mathcal{F} . Therefore, by Chebyshev's inequality,

$$\mathbb{P}(\lambda(\mathcal{F}) \leq 1 - 2C'\ell^{-8}) \leq \frac{\mathbb{E}(1 - \lambda(\mathcal{F}))}{2C'\ell^{-8}} \leq \ell^{-8}/2 < 1.$$

We may assume that $C'\ell^{-8} \leq 1/2$, thanks to (25). We showed that with positive probability $\lambda(\mathcal{F}) \geq 1 - 2C'\ell^{-8}$. Recall that $\tilde{\mu}_v = \Gamma_*(\mu_v)$ is $C\ell^{-c/d}$ -radial for any $v \in \mathcal{F}$. Hence, according to Lemma 2.3(c), with positive probability of selecting the Gaussian random matrix Γ , the measure

$$\int_{\mathcal{O}_\ell} \Gamma_*(\mu_v) d\lambda(v) = \Gamma_* \left(\int_{\mathcal{O}_\ell} \mu_v d\lambda(v) \right) = \Gamma_*(\mu)$$

is $C'\ell^{-c'/d}$ -radial on \mathbb{R}^d . □

The Grassmannian $G_{n,k}$ of all k -dimensional subspaces in \mathbb{R}^n carries a unique rotationally invariant probability measure, which will be referred to as the uniform probability measure on $G_{n,k}$. When Γ is a $d \times n$ standard Gaussian random matrix, the kernel of Γ is a random $(n-d)$ -dimensional subspace, that is distributed uniformly in the Grassmannian $G_{n,n-d}$. For a subspace $E \subseteq \mathbb{R}^n$ we write $E^\perp = \{x \in \mathbb{R}^n; \forall y \in E, x \cdot y = 0\}$ for its orthogonal complement.

Lemma 4.5 *Let $0 \leq k \leq n-1$ be integers, and let μ be a Borel probability measure on \mathbb{R}^n with $\mu(\{0\}) = 0$. Suppose that E is a random k -dimensional subspace, distributed uniformly in $G_{n,k}$. Then $\mu(E) = 0$ with probability one of selecting E .*

Proof: By induction on k . The case $k = 0$ holds trivially. Suppose now that $k \geq 1$, let n be such that $k \leq n-1$, and let μ be a Borel probability measure on \mathbb{R}^n with $\mu(\{0\}) = 0$. Since $\mu(\{0\}) = 0$, there are at most countably many one-dimensional subspaces $\ell \subset \mathbb{R}^n$ with $\mu(\ell) > 0$. Let ℓ be a random one-dimensional subspace, distributed

uniformly in $G_{n,1}$. Then with probability one, $\mu(\ell) = 0$. Denote $\nu = (Proj_{\ell^\perp})_*\mu$, a measure supported on an $(n-1)$ -dimensional subspace, with $\nu(\{0\}) = 0$. Let F be a random $(k-1)$ -dimensional subspace in ℓ^\perp , distributed uniformly. By the induction hypothesis, $\nu(F) = 0$ with probability one. Denoting $E = Proj_{\ell^\perp}^{-1}(F)$, we see that $\mu(E) = \nu(F) = 0$ with probability one. From the uniqueness of the Haar measure, E is distributed uniformly in $G_{n,k}$, and the lemma follows. \square

Corollary 4.6 *Let $1 \leq d \leq n$ be integers and let $0 < \varepsilon < 1/2$. Let μ be a Borel probability measure on \mathbb{R}^n with $\mu(\{0\}) = 0$. Assume that*

$$\sup_{\theta \in S^{n-1}} \int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_*\mu(x) \leq (\tilde{c}\varepsilon)^{\tilde{C}d}. \quad (27)$$

Let Γ be a $d \times n$ standard Gaussian random matrix. Then, with positive probability of selecting Γ , the measure Γ_μ on \mathbb{R}^d is ε -radial proper. Here, $0 < \tilde{c} < 1$ and $\tilde{C} > 1$ are universal constants. (In fact, we have a lower bound of $1 - (\tilde{c}\varepsilon)^{\tilde{C}d/10}$ for the aforementioned probability.)*

Proof: Throughout this proof, we write C for the universal constant from Lemma 4.4. We define $\tilde{c} = (10C)^{-1}$ and $\tilde{C} = 100C$. It is elementary to verify that with this choice of universal constants, there exists an integer ℓ such that

$$\ell \geq (5C/\varepsilon)^{Cd} \quad \text{and} \quad (\tilde{c}\varepsilon)^{\tilde{C}d} \leq \frac{\varepsilon^2}{50\ell^{50}}.$$

Note that the left-hand side of (27) is at least $1/n$. Indeed,

$$\begin{aligned} & \sup_{\theta \in S^{n-1}} \int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_*\mu(x) \\ & \geq \int_{S^{n-1}} \int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_*\mu(x) d\sigma_{d-1}(\theta) = \int_{S^{n-1}} \frac{|x|^2}{n} d\mathcal{R}_*(\mu)(x) = \frac{1}{n}. \end{aligned}$$

We conclude that $d < \ell \leq n$, and

$$\sup_{\theta \in S^{n-1}} \int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_*\mu(x) \leq \frac{\varepsilon^2}{50\ell^{50}}.$$

According to Lemma 4.3, there exists a Borel probability measure λ on \mathcal{O}_ℓ such that

$$d_{TV} \left(\mu, \int_{\mathcal{O}_\ell} \mu_v d\lambda(v) \right) \leq \varepsilon^2/25. \quad (28)$$

Denote $\nu = \int_{\mathcal{O}_\ell} \mu_v d\lambda(v)$. From Lemma 4.4 the measure $\Gamma_*(\nu)$ is $(\varepsilon/5)$ -radial, with probability at least $1 - \ell^{-8}$ of selecting Γ , because $\ell \geq (5C/\varepsilon)^{Cd}$. Additionally, $d_{TV}(\Gamma_*\mu, \Gamma_*\nu) \leq \varepsilon^2/25$, by (28). From Lemma 2.3(a) we thus learn that $\Gamma_*(\mu)$ is ε -radial, with positive probability of selecting Γ . Moreover, $\Gamma_*(\mu)(\{0\}) = \mu(\Gamma^{-1}(0)) = 0$ with probability one, according to Lemma 4.5. Hence, with positive probability, $\Gamma_*(\mu)$ is ε -radial proper. \square

5 Selecting a position

Our goal in this section is to find an appropriate invertible linear transformation T on \mathbb{R}^n such that $T_*\mu$ satisfies the requirements of Corollary 4.6. Our analysis is very much related to the results of Barthe [2], Carlen and Cordero-Erausquin [7] and Carlen, Lieb and Loss [8]. For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ we write $x \otimes x$ for the $n \times n$ matrix whose entries are $(x_i x_j)_{i,j=1, \dots, n}$. For a probability measure μ on the unit sphere S^{n-1} define

$$M(\mu) = \int_{S^{n-1}} (x \otimes x) d\mu(x).$$

Then $M(\mu)$ is a positive semi-definite matrix of trace one. Clearly, for any $\theta \in \mathbb{R}^n$, we have $M(\mu)\theta \cdot \theta = \int_{S^{n-1}} (x \cdot \theta)^2 d\mu(x)$. More generally, for any subspace $E \subseteq \mathbb{R}^n$,

$$\int_{S^{n-1}} |Proj_E(x)|^2 d\mu(x) = Tr(Proj_E M(\mu)) = Tr(M(\mu) Proj_E),$$

the trace of the matrix $M(\mu) Proj_E$. A Borel probability measure μ on S^{n-1} is called isotropic if $M(\mu) = Id/n$, where Id is the identity matrix. Observe that when μ is isotropic, for any subspace $E \subseteq \mathbb{R}^n$,

$$\int_{S^{n-1}} |Proj_E x|^2 d\mu(x) = \frac{\dim(E)}{n}. \quad (29)$$

In particular, $\mu(E) \leq \dim(E)/n$ and hence an isotropic probability measure is necessarily decent in the sense of Definition 1.1. A Borel probability measure μ on \mathbb{R}^n with $\mu(\{0\}) = 0$ is called “potentially isotropic” if there exists an invertible linear map T on \mathbb{R}^n such that $(\mathcal{R} \circ T)_*\mu$ is isotropic.

Lemma 5.1 *Let μ be a Borel probability measure on S^{n-1} such that*

$$\mu(H) = 0 \quad (30)$$

for any hyperplane $H \subset \mathbb{R}^n$ through the origin. Then μ is potentially isotropic.

Proof: Given an invertible linear map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ we abbreviate $M_\mu(T) = M((\mathcal{R} \circ T)_*\mu)$. Then $M_\mu(T)$ is a positive semi-definite matrix of trace one, and by the arithmetic-geometric means inequality, $\det M_\mu(T) \leq n^{-n}$. Note that $M_\mu(T) = M_\mu(\lambda T)$ for any $\lambda > 0$. Consider the supremum of the continuous functional

$$T \mapsto \det M_\mu(T) \quad (31)$$

over the space of all invertible linear operators $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ of Hilbert-Schmidt norm one.

We claim that the supremum is attained. Indeed, let T_1, T_2, \dots be a maximizing sequence of matrices. By passing to a subsequence if necessary, we may assume that $T_i \rightarrow T$, for a certain matrix T of Hilbert-Schmidt norm one. We need to show that T is

invertible. Denote by E the image of T , a subspace of \mathbb{R}^n . We need to show that $E = \mathbb{R}^n$. For any $x \in S^{n-1}$ which is not in the kernel of T , we have $T_i x \rightarrow Tx \in E \setminus \{0\}$, hence

$$|Proj_{E^\perp}(\mathcal{R} \circ T_i)(x)| \xrightarrow{i \rightarrow \infty} 0. \quad (32)$$

The kernel of T is at most $(n-1)$ -dimensional, since the Hilbert-Schmidt norm of T is one. According to (30), the convergence in (32) occurs μ -almost-everywhere in x . Therefore,

$$Tr(M_\mu(T_i)Proj_{E^\perp}) = \int_{S^{n-1}} |Proj_{E^\perp}(\mathcal{R} \circ T_i)(x)|^2 d\mu(x) \xrightarrow{i \rightarrow \infty} 0.$$

We conclude that if $E \neq \mathbb{R}^n$, then $\det M_\mu(T_i) \rightarrow 0$, in contradiction to the maximizing property of the sequence $(T_i)_{i \geq 1}$. Hence $E = \mathbb{R}^n$ and T is invertible. Thus the supremum of the functional (31) is attained for some invertible matrix T_0 of Hilbert-Schmidt norm one. We will show that $(\mathcal{R} \circ T_0)_* \mu$ is isotropic. Without loss of generality we assume that $T_0 = Id$ (otherwise, replace μ with $(\mathcal{R} \circ T_0)_* \mu$ and note that this replacement does not affect the validity of the assumptions and the conclusions of the lemma).

The matrix $M(\mu) = M_\mu(Id)$ is a positive semi-definite matrix of trace one. It is non-singular, thanks to (30), and therefore $M(\mu)$ is in fact positive definite. Moreover, for any function $u : S^{n-1} \rightarrow \mathbb{R}$ which is positive μ -almost-everywhere and for any $\theta \in S^{n-1}$,

$$\int_{S^{n-1}} u(x)(x \cdot \theta)^2 d\mu(x) > 0. \quad (33)$$

Assume by contradiction that $M(\mu)$ is not a scalar matrix. Denote by λ_1 the largest eigenvalue of $M(\mu) = M_\mu(Id)$, and let $E \subset \mathbb{R}^n$ be the eigenspace corresponding to the eigenvalue λ_1 . Then $1 \leq \dim(E) \leq n-1$. For $0 \leq \delta < 1$ consider the linear operator

$$L_\delta(x) = x - \delta Proj_E(x) \quad (x \in \mathbb{R}^n).$$

Then $Proj_E L_\delta = (1-\delta)Proj_E$ while $Proj_{E^\perp} L_\delta = Proj_{E^\perp}$. This means that $\mathcal{R} \circ L_\delta$ strengthens the E^\perp -component of a given point in \mathbb{R}^n , at the expense of its E -component. More precisely, for any $x \in S^{n-1}$ and $0 \leq \delta < 1$ there exists $\varepsilon_x^\delta \geq 0$ such that

$$Proj_{E^\perp}(\mathcal{R}(L_\delta x)) = (1 + \varepsilon_x^\delta) Proj_{E^\perp}(x).$$

Moreover, when $x \notin E \cup E^\perp$ we have the inequality $\varepsilon_x^\delta \geq \varepsilon(x)\delta$ for some $\varepsilon(x) > 0$ depending only on x . Consequently, for any $0 \leq \delta < 1$ and a non-zero vector $\theta \in E^\perp$,

$$\begin{aligned} \int_{S^{n-1}} (\mathcal{R}(L_\delta x) \cdot \theta)^2 d\mu(x) &= \int_{S^{n-1}} (1 + \varepsilon_x^\delta)^2 (x \cdot \theta)^2 d\mu(x) \\ &\geq \int_{S^{n-1}} (x \cdot \theta)^2 d\mu(x) + 2\delta \int_{S^{n-1}} \varepsilon(x)(x \cdot \theta)^2 d\mu(x). \end{aligned} \quad (34)$$

The symmetric matrix $M_\mu(L_\delta)$ is of trace one, and it depends smoothly on δ . Denote $D = dM_\mu(L_\delta)/d\delta|_{\delta=0}$, a traceless symmetric matrix. According to our assumption (30), the condition $x \notin E \cup E^\perp$ holds μ -almost-everywhere as $1 \leq \dim(E) \leq n-1$.

Therefore $\varepsilon(x) > 0$ for μ -almost every $x \in S^{n-1}$. From (33) and (34) we learn that for any $0 \neq \theta \in E^\perp$,

$$D\theta \cdot \theta = \frac{d}{d\delta} \left(\int_{S^{n-1}} (\mathcal{R}(L_\delta x) \cdot \theta)^2 d\mu(x) \right) \Big|_{\delta=0} > 0. \quad (35)$$

Recall that E is the eigenspace corresponding to the maximal eigenvalue λ_1 of $M(\mu)$. Denote by λ_2 the second-largest eigenvalue, which is still positive but is strictly smaller than λ_1 . Then $Proj_{E^\perp} M(\mu)^{-1} \geq \lambda_2^{-1} Proj_{E^\perp}$ in the sense of symmetric matrices. Using elementary linear algebra, we deduce from (35) that

$$Tr(Proj_{E^\perp} M(\mu)^{-1} D) \geq \lambda_2^{-1} Tr(Proj_{E^\perp} D) > 0.$$

Since $Tr(D) = 0$ then $Tr(Proj_E D) = -Tr(Proj_{E^\perp} D)$ and

$$\begin{aligned} \frac{d \log \det M_\mu(L_\delta)}{d\delta} \Big|_{\delta=0} &= Tr(M(\mu)^{-1} D) \\ &= Tr(Proj_E M(\mu)^{-1} D) + Tr(Proj_{E^\perp} M(\mu)^{-1} D) \\ &= \frac{Tr(Proj_E D)}{\lambda_1} + Tr(Proj_{E^\perp} M(\mu)^{-1} D) \geq \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right) Tr(Proj_{E^\perp} D) > 0, \end{aligned}$$

in contradiction to the maximality of $\det M(\mu)$. Hence our assumption that $M(\mu)$ is not a scalar matrix was absurd. Since $Tr(M(\mu)) = 1$ then $M(\mu) = Id/n$ and μ is isotropic. \square

For a subspace $E \subset \mathbb{R}^n$ and $\delta > 0$ we write

$$\mathcal{N}_\delta(E) = \{rx; |x| = 1, r \geq 0, d(x, E \cap S^{n-1}) \leq \delta\} \quad (36)$$

where $d(x, A) = \inf_{y \in A} |x - y|$. Then $\mathcal{N}_\delta(E)$ is the projective δ -neighborhood of E . We will need the following auxiliary continuity lemma. It is the only time in this text where the non-degeneracy conditions of Definition 1.1 are used.

Lemma 5.2 *Let $n \geq 1$ be an integer and let μ be a probability measure on \mathbb{R}^n with $\mu(\{0\}) = 0$ such that*

$$\mu(E) < \dim(E)/n$$

for any subspace $E \subset \mathbb{R}^n$ other than \mathbb{R}^n and $\{0\}$. Suppose there exists a sequence of potentially isotropic probability measures on S^{n-1} that converges to $\mathcal{R}_ \mu$ in the weak* topology. Then μ is potentially isotropic.*

Proof: From the assumptions of the lemma, there exist Borel probability measures μ_1, μ_2, \dots on S^{n-1} and invertible linear maps T_1, T_2, \dots for which the following holds:

- $\mu_i \xrightarrow{i \rightarrow \infty} \mathcal{R}_* \mu$ in the weak* topology; and
- $(\mathcal{R} \circ T_i)_* \mu_i$ is isotropic for all i .

Without loss of generality we may assume that the T_i 's are positive definite operators of trace one: If not, we will replace the operator T_i by rU_iT_i , where U_i is an orthogonal transformation such that U_iT_i is positive definite and r^{-1} is the trace of U_iT_i . Such a replacement does not affect the isotropicity of $(\mathcal{R} \circ T_i)_*\mu_i$. Furthermore, replacing T_i ($i = 1, 2, \dots$) with a subsequence, we may assume that $T_i \rightarrow T$, where T is a positive semi-definite matrix of trace one.

We claim that T is invertible. Assume by contradiction that T is singular. Denote by $E \subset \mathbb{R}^n$ the kernel of T , and set $k = \dim(E)$. Then $1 \leq k \leq n - 1$ and hence $\mu(E) < k/n$. Since $E = \bigcap_{\delta > 0} \mathcal{N}_\delta(E)$, then there exists $\delta > 0$ such that

$$\mu(\mathcal{N}_\delta(E)) < k/n.$$

The set $\mathcal{N}_\delta(E) \cap S^{n-1}$ is closed in S^{n-1} . Since $\mu_i \rightarrow \mathcal{R}_*\mu$ in the weak* topology then

$$\limsup_{i \rightarrow \infty} \mu_i(\mathcal{N}_\delta(E)) \leq \mathcal{R}_*\mu(\mathcal{N}_\delta(E)) = \mu(\mathcal{N}_\delta(E)) < \frac{k}{n}. \quad (37)$$

Recall that $T_i \rightarrow T$, that the T_i 's are self-adjoint, and that E is the kernel of T , hence E^\perp is the image of T . This entails, roughly speaking, that for any $x \notin E$, the sequence $T_i x$ is ‘‘approaching E^\perp ’’. In more precise terms, we conclude that for any $x \notin \mathcal{N}_\delta(E)$,

$$\left| Proj_{E^\perp} \left(\frac{T_i x}{|T_i x|} \right) \right| \xrightarrow{i \rightarrow \infty} 1. \quad (38)$$

Moreover, the convergence in (38) is uniform over $x \in \mathbb{R}^n \setminus \mathcal{N}_\delta(E)$. Consequently, from (37) and (38),

$$\liminf_{i \rightarrow \infty} \int_{S^{n-1} \setminus \mathcal{N}_\delta(E)} \left| Proj_{E^\perp} \left(\frac{T_i x}{|T_i x|} \right) \right|^2 d\mu_i(x) = \liminf_{i \rightarrow \infty} \mu_i(S^{n-1} \setminus \mathcal{N}_\delta(E)) > 1 - \frac{k}{n}. \quad (39)$$

Recall that $(\mathcal{R} \circ T_i)_*\mu_i$ is isotropic. According to (29),

$$\int_{S^{n-1}} \left| Proj_{E^\perp} \left(\frac{T_i x}{|T_i x|} \right) \right|^2 d\mu_i(x) = \frac{\dim(E^\perp)}{n} = 1 - \frac{k}{n},$$

in contradiction to (39). Thus our assumption that T is singular was absurd, and T is necessarily invertible.

Since $T_i \rightarrow T$ with T being invertible, we know that for any $x \in S^{n-1}$,

$$\frac{T_i x}{|T_i x|} \xrightarrow{i \rightarrow \infty} \frac{T x}{|T x|}$$

and the convergence is uniform in S^{n-1} . Therefore, for any $\theta \in S^{n-1}$,

$$\int_{S^{n-1}} \left(\frac{T_i x}{|T_i x|} \cdot \theta \right)^2 d\mu_i(x) \xrightarrow{i \rightarrow \infty} \int_{S^{n-1}} \left(\frac{T x}{|T x|} \cdot \theta \right)^2 d\mathcal{R}_*\mu(x). \quad (40)$$

However, the left-hand side of (40) is always $1/n$. We see that $(\mathcal{R} \circ T \circ \mathcal{R})_*\mu = (\mathcal{R} \circ T)_*\mu$ is isotropic, and therefore μ is potentially isotropic. \square

Corollary 5.3 *Let n be a positive integer and let μ be a Borel probability measure on \mathbb{R}^n with $\mu(\{0\}) = 0$ such that*

$$\mu(E) < \dim(E)/n \quad (41)$$

for any subspace $E \subset \mathbb{R}^n$ other than \mathbb{R}^n and $\{0\}$. Then μ is potentially isotropic.

Proof: Consider a sequence μ_1, μ_2, \dots of Borel probability measures on S^{n-1} , absolutely continuous with respect to the Lebesgue measure on S^{n-1} , that converges to $\mathcal{R}_*\mu$ in the weak* topology. The μ_i 's are potentially isotropic by Lemma 5.1. Therefore μ is potentially isotropic according to Lemma 5.2. \square

A clever proof of Corollary 5.3 for the case where the measure μ is discrete and has finite support appears in the works of Barthe [2], Carlen and Cordero-Erausquin [7, Lemma 3.5] and Carlen, Lieb and Loss [8]. We were not able to generalize their argument to the case of a general measure satisfying (41). The proof presented above is unfortunately longer, but perhaps it has the advantage of being geometrically straightforward.

Lemma 5.4 *Let n be a positive integer and $\alpha > 0$. Suppose that μ is an α -decent probability measure on \mathbb{R}^n . Then for any $0 < \varepsilon < 1$, there exists a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $\nu = T_*(\mu)$ satisfies*

$$\int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_*\nu(x) \leq \alpha + \varepsilon \quad \text{for all } \theta \in S^{n-1}.$$

Proof: By induction on the dimension n . The case $n = 1$ is obvious. Suppose that $n \geq 2$. We may assume that $\mu(H) < 1$ for any hyperplane $H \subset \mathbb{R}^n$ that passes through the origin (otherwise, invoke the induction hypothesis). We may also assume that $\alpha = \sup_{E \subset \mathbb{R}^n} \mu(E)/\dim(E)$ where the supremum runs over all subspaces $\{0\} \neq E \subseteq \mathbb{R}^n$. Corollary 5.3 takes care of the case where

$$\mu(E) < \dim(E)/n$$

for any subspace $E \subset \mathbb{R}^n$ with $1 \leq \dim(E) \leq n - 1$. We may thus focus on the case where there exists a proper subspace $E \subset \mathbb{R}^n$ with $\mu(E) \geq \dim(E)/n$. Clearly $\alpha \geq 1/n$. Consequently, there is a subspace $E \subset \mathbb{R}^n$, with $1 \leq \dim(E) \leq n - 1$, such that $\alpha - \varepsilon/(3n) \leq \mu(E) < 1$. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the map defined by

$$T(x) = \begin{cases} Proj_{E^\perp} x & x \notin E \\ x & x \in E \end{cases}$$

The map T may be viewed as a ‘‘stratified linear map’’ as in Furstenberg [9]. Set $\lambda = \mu(E) > 0$. The probability measure $T_*\mu$ on \mathbb{R}^n is supported on $E \cup E^\perp$, and it may be decomposed as

$$T_*\mu = \lambda\mu_E + (1 - \lambda)\mu_{E^\perp}$$

where $\mu_E = \mu|_E$ is the conditioning of μ to E , and μ_{E^\perp} is a certain probability measure supported on E^\perp . Clearly, $\mu_E = \mu|_E$ is (α/λ) -decent. Regarding μ_{E^\perp} , let us select a subspace $F \subseteq E^\perp$. Then, $\mu_{E^\perp}(F) = 0$ if $F = \{0\}$ and otherwise

$$\begin{aligned} (1 - \lambda)\mu_{E^\perp}(F) &= \mu(T^{-1}(F \setminus \{0\})) = \mu((F \oplus E) \setminus E) = \mu(F \oplus E) - \mu(E) \\ &\leq \alpha(\dim(E) + \dim(F)) - (\alpha - \varepsilon/(3n))\dim(E) \leq (\alpha + \varepsilon/3)\dim(F) \end{aligned}$$

where $E \oplus F = \{x + y; x \in E, y \in F\}$ is the subspace spanned by E and F . Consequently, μ_{E^\perp} is an $((\alpha + \varepsilon/3)/(1 - \lambda))$ -decent measure on E^\perp . We may apply the induction hypothesis for μ_E and μ_{E^\perp} . We conclude that there exists a linear transformation $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$, with $S(E) \subseteq E$ and $S(E^\perp) \subseteq E^\perp$, such that

$$\int_{S^{n-1}} (x \cdot \theta)^2 d(\mathcal{R} \circ S \circ T)_* \mu(x) \leq \alpha + \varepsilon/2 \quad \text{for any } \theta \in S^{n-1}. \quad (42)$$

The problem is that $S \circ T$ is not a linear map. However, it is easy to approximate it by a linear map: For $0 < \delta < 1$ denote $T_\delta x = x - \delta \text{Proj}_E x$. Then $(\mathcal{R} \circ T)(x) = \lim_{\delta \rightarrow 1^-} (\mathcal{R} \circ T_\delta)(x)$ for any $0 \neq x \in \mathbb{R}^n$. Consequently,

$$(\mathcal{R} \circ S \circ T_\delta)_* \mu = (\mathcal{R} \circ S \circ \mathcal{R} \circ T_\delta)_* \mu \xrightarrow{\delta \rightarrow 1^-} (\mathcal{R} \circ S \circ \mathcal{R} \circ T)_* \mu = (\mathcal{R} \circ S \circ T)_* \mu$$

in the weak* topology. We conclude that the matrices $M((\mathcal{R} \circ S \circ T_\delta)_* \mu)$ tend to $M((\mathcal{R} \circ S \circ T)_* \mu)$ as $\delta \rightarrow 1^-$. Hence, by (42), for some $\delta_0 < 1$,

$$\int_{S^{n-1}} (x \cdot \theta)^2 d(\mathcal{R} \circ S \circ T_{\delta_0})_* \mu(x) \leq \alpha + \varepsilon \quad \text{for any } \theta \in S^{n-1}.$$

The map $S \circ T_{\delta_0}$ is the desired linear transformation. This completes the proof. \square

6 Proof of the main theorem and some remarks

Proof of Theorem 1.3: Suppose that μ is an η -decent probability measure on \mathbb{R}^n . According to Lemma 5.4, there exists a linear map $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $\nu = S_* \mu$ satisfies

$$\int_{S^{n-1}} (x \cdot \theta)^2 d\mathcal{R}_* \nu(x) \leq 2\eta \quad \text{for all } \theta \in S^{n-1}.$$

We invoke Corollary 4.6 for the measure ν . We see that if the positive integer d and $0 < \varepsilon < 1/2$ are such that

$$2\eta \leq (\tilde{c}\varepsilon)^{\tilde{C}d},$$

then there exists a $d \times n$ matrix Γ for which the measure $\Gamma_* \nu$ on \mathbb{R}^d is ε -radial proper. Setting $T = \Gamma S$, a $d \times n$ matrix, we conclude that $T_* \mu = \Gamma_* \nu$ is a measure on \mathbb{R}^d which is ε -radial proper. \square

Proof of Corollary 1.4: We may assume that n exceeds a given large universal constant. Denote $d = \bar{c} \lceil \sqrt{\log n} \rceil$ and $\delta = e^{-d}/500$, for a small universal constant $0 < \bar{c} < 1$ such that $n \geq (C/\delta)^{C^d}$ where C is the universal constant from Theorem 1.3. According to Theorem 1.3, we may pass to a d -dimensional marginal and assume that our measure μ is a proper δ -radial measure on \mathbb{R}^d . For $t \in \mathbb{R}$ and $L > 0$ let $\chi_{t,L}$ be the L -Lipschitz function on the real line which equals zero on $(-\infty, t]$ and one on $[t + 1/L, \infty)$. Recall the Kantorovich-Rubinstein duality as in (5) above. Then, for any probability measure ν on the unit sphere S^{d-1} and $0 < t < c \leq 1/2$,

$$\nu(\{x; x_1 \geq t\}) \geq \int_{S^{d-1}} \chi_{t,d}(x_1) d\nu(x) \geq \int_{S^{d-1}} \chi_{t,d}(x_1) d\sigma_{d-1}(x) - dW_1(\nu, \sigma_{d-1}),$$

where $x = (x_1, \dots, x_d)$ are the coordinates of $x \in S^{d-1}$. The integral with respect to σ_{d-1} may be estimated directly, and it is bounded from below by ce^{-Ct^2d} (note that the marginal of σ_{d-1} on the first coordinate has a density that is proportional to $(1-t^2)_+^{(d-3)/2}$ on $[-1, 1]$). We conclude that for any $0 < t < c$ and an interval $J = [a, b] \subset (0, \infty)$ with $\mu(S(J)) \geq \delta$,

$$\begin{aligned} \mu|_{S(J)}(\{x; x_1 \geq at\}) &\geq \mathcal{R}_*(\mu|_{S(J)})(\{x; x_1 \geq t\}) \\ &\geq ce^{-Ct^2d} - d \cdot W_1(\mathcal{R}_*(\mu|_{S(J)}), \sigma_{d-1}) \geq ce^{-Ct^2d} - d\delta \geq c'e^{-C't^2d}. \end{aligned} \quad (43)$$

Similarly, for any interval $J = [a, b] \subset (0, \infty)$ with $\mu(S(J)) \geq \delta$,

$$\begin{aligned} \mu|_{S(J)}(\{x; |x_1| \geq 20b/\sqrt{d}\}) &\leq \mathcal{R}_*(\mu|_{S(J)})(\{x; |x_1| \geq 20/\sqrt{d}\}) \\ &\leq \int_{S^{d-1}} \chi_{20/\sqrt{d}-d^{-1}, d}(|x_1|) d\sigma_{d-1}(x) + d \cdot W_1(\mathcal{R}_*(\mu|_{S(J)}), \sigma_{d-1}) \leq 1/5, \end{aligned} \quad (44)$$

where the integral with respect to σ_{d-1} is estimated in a straightforward manner. Let $\tilde{M} > 0$ be a quantile with

$$\mu(\{x; |x| \leq \tilde{M}\}) \geq 3/4 \quad \text{and} \quad \mu(\{x; |x| \geq \tilde{M}\}) \geq 1/4.$$

Let $a > 0$ be such that the interval $J = [a, \tilde{M}]$ satisfies $\mu(S(J)) \geq 2/3$. We apply (44) for the interval $J = [a, \tilde{M}]$ to deduce that

$$\mu(\{x; |x_1| \geq 20\tilde{M}/\sqrt{d}\}) \leq \frac{1}{3} + \frac{2}{3} \cdot \mu|_{S(J)}(\{x; |x_1| \geq 20\tilde{M}/\sqrt{d}\}) \leq \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{5} < \frac{1}{2}. \quad (45)$$

Suppose that $M > 0$ satisfies (3) with the linear functional $\varphi(x) = x_1$. We learn from (45) that necessarily $M \leq 20\tilde{M}/\sqrt{d}$. Let $b > 0$ be such that the interval $J = [\tilde{M}, b]$ satisfies $\mu(S(J)) \geq 1/5$. We apply (43) for the interval $J = [\tilde{M}, b]$ and conclude that, for any $0 \leq t \leq c\sqrt{d}/20$,

$$\mu(\{x; x_1 \geq tM\}) \geq \frac{1}{5} \cdot \mu|_{S(J)}(\{x; x_1 \geq 20t\tilde{M}/\sqrt{d}\}) \geq \frac{c'}{5} \exp(-400C't^2).$$

Since $c\sqrt{d}/20 \geq \tilde{c} \log^{1/4} n$, the proof of the lower bound for $\mu(\{x; x_1 \geq tM\})$ is complete. The proof of the lower bound for $\mu(\{x; x_1 \leq -tM\})$ is almost entirely identical. The corollary is thus proven. \square

Remarks.

1. It is conceivable that a more delicate analysis yields a better bound for R_n in Corollary 1.4. However, note that $R_n \leq C\sqrt{\log n}$ as is shown by the example where μ is distributed uniformly on n linearly independent vectors in \mathbb{R}^n . Compare the ‘‘super-gaussian’’ tail behavior of Corollary 1.4 with the almost sub-gaussian bounds in the convex case in [13] and in Giannopoulos, Paouris and Pajor [10].
2. The central limit theorem for convex bodies [14, 15] states that any uniform probability measure on a high-dimensional convex set has some low-dimensional marginals

that are approximately Gaussian. It is clear that there are perfectly regular probability measures in high dimension (e.g., a mixture of two Gaussians) without any approximately Gaussian marginals. Therefore, a geometric condition such as convexity is indeed relevant when we look for approximately Gaussian marginals. For arbitrary high-dimensional measures without convexity properties, we may still state the more modest conclusion that some of the marginals are approximately spherically-symmetric, according to Theorem 1.3. There is no hope for approximate Gaussians.

Theorem 1.3 bears a strong relation to the proof of the central limit theorem for convex bodies presented in [14, 15] (see [16] for another proof, which at present works only for a subclass of convex bodies). That proof begins by showing that marginals of the uniform measure on a convex body are approximately spherically-symmetric. The approximation in [14, 15] is rather strong compared to Theorem 1.3, but nevertheless, a simple compactness argument enables us to leverage Theorem 1.3 in order to obtain the desired type of approximation. In principle, this approach yields a slightly different proof of the central limit theorem for convex sets, albeit with weaker estimates.

The Euclidean structure with respect to which a random projection “works” with high probability seems *a priori* different in Theorem 1.3 and in the central limit theorem for convex bodies. In Theorem 1.3 we use the Euclidean structure with respect to which the covariance matrix of $\mathcal{R}_*\mu$ is scalar, while in the central limit theorem for convex bodies, the most natural position is to require the covariance matrix of μ itself to be a scalar matrix (compare also with [18], [21]). For convex bodies, these Euclidean structures are close to each other, since most of the mass of a normalized convex body is located very close to a sphere (see [17]).

3. The linear map T in Theorem 1.3 may be assumed to be an orthogonal projection. This follows from the following simple observation we learned from G. Schechtman: Any n -dimensional ellipsoid has an $\lceil n/2 \rceil$ -dimensional projection which is precisely a Euclidean ball. Therefore, in order to show that T may be chosen to be an orthogonal projection, one essentially has to verify that a $\lceil d/2 \rceil$ -dimensional marginal of an ε -radial measure on \mathbb{R}^d is $100\varepsilon^{1/8}$ -radial. We omit the details.
4. The isoperimetric inequality on the high-dimensional sphere, which is the cornerstone of the concentration of measure phenomenon (see Milman and Schechtman [24]), is not used in the proof of Theorem 1.3. We do apply Levy’s lemma, which embeds the isoperimetric inequality, in the proof of Lemma 2.4, but only in d dimensions. The dimension d here is typically not very large.
5. For a positive integer d and $\varepsilon > 0$ denote by $N_0(\varepsilon, d)$ the minimal dimension with the following property: Whenever $N \geq N_0(\varepsilon, d)$, any N -dimensional Banach space has a d -dimensional subspace which is ε -close to a Hilbert space. The classical Dvoretzky’s theorem states that $N_0(\varepsilon, d) \leq \exp(Cd/\varepsilon^2)$, where $C > 0$ is a universal constant (see Milman [23] and references therein). The power of $1/\varepsilon$ in the exponent in the bound for $N_0(\varepsilon, d)$ can be made arbitrarily close to one at the

expense of increasing the universal constant C (see Schechtman [27]). It is conceivable, however, that these bounds are still far from optimal; perhaps $N_0(\varepsilon, d)$ can be made as small as $(C/\varepsilon)^{C^d}$? See Milman [22] for a discussion of this conjecture. An affirmative answer for the case $d = 2$ was given by Gromov [22], using a topological argument which does not seem to generalize to higher dimensions.

The analogy with the present article suggests to try and use Theorem 1.3, or ideas from its proof, in order to improve the bounds in Dvoretzky's theorem. Furthermore, the operation of marginal is dual, via the Fourier transform, to the operation of restriction to a subspace. So, for instance, suppose a norm $\|\cdot\|$ in \mathbb{R}^n may be represented as

$$\|x\| = \int_{\mathbb{R}^n} |x \cdot \theta| d\mu(\theta) \quad (46)$$

for a compactly-supported probability measure μ on \mathbb{R}^n . In this case, we may consider subspaces $E \subset \mathbb{R}^n$ for which $(Proj_E)_*\mu$ is ε -radial, and expect that the restriction of $\|\cdot\|$ to these subspaces is close, in a certain sense, to the Euclidean norm. See Koldobsky [19, Chapter 6] for a comprehensive discussion of norms admitting representations in the spirit of (46).

While this approach may possibly yield some meaningful estimates for some classes of normed spaces, it has limitations. Theorem 1.3 is proven by considering a *random* marginal with respect to an appropriate Euclidean structure, i.e., a projection of the given measure to a subspace which is distributed uniformly over the Grassmannian of all d -dimensional subspaces in \mathbb{R}^n . However, for Banach spaces such as ℓ_∞^N , a random subspace is not sufficiently close to a Hilbert space (see Schechtman [28]), and there are better choices than the random one. (Indeed, the ℓ_∞^N norm cannot be represented as in (46) or in a similar way, see Theorem 6.13 in Koldobsky [19], going back to Misiewicz). A direct application of Theorem 1.3 is thus quite unlikely to provide new information regarding approximately Hilbertian subspaces for *all* finite-dimensional normed spaces.

6. In principle, the measures $T_*(\mu)$ in Theorem 1.3 are not only approximately radial, but are also approximately a composition of isotropic Gaussians. Indeed, it is well-known that any d -dimensional marginal of the measure σ_{k-1} , for $d \ll k$, is approximately an isotropic d -dimensional Gaussian measure. Thus, we may project an approximately-radial measure on \mathbb{R}^k to any d -dimensional subspace, and obtain a measure which is approximately, in some sense, a composition of isotropic Gaussians. We did not rigorously investigate this approximation property on a precise, quantitative level.

7 Infinite-dimensional spaces

This section contains a corollary to Theorem 1.3, pertaining to probability measures supported on infinite-dimensional spaces. We begin with a lemma regarding distributions on finite-dimensional spaces. Let $n \geq 1$ be an integer, suppose that μ is a Borel probability measure on \mathbb{R}^n and let $0 < a \leq 1$. A subspace $E \subseteq \mathbb{R}^n$ is “ a -basic for μ ” if

- (i) $\mu(E) \geq a$
- (ii) $\mu(F) < a$ for any proper subspace $F \subsetneq E$.

Note that any subspace $E \subseteq \mathbb{R}^n$ with $\mu(E) \geq a$ contains an a -basic subspace. Also, suppose $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear map, and let $E \subseteq \mathbb{R}^n$ be an a -basic subspace for μ containing the kernel of T . Then $T(E)$ is a -basic for $T_*(\mu)$.

Lemma 7.1 *Let $n \geq 1$ be an integer, $0 < a \leq 1$, and let μ be a Borel probability measure on \mathbb{R}^n . Then there are only finitely many subspaces $E \subseteq \mathbb{R}^n$ that are a -basic for μ .*

Proof: Let $k \geq 0$ be an integer and $0 < a \leq 1$. We will prove by induction on k the following statement: For any integer $n \geq k$ and for any Borel probability measure μ on \mathbb{R}^n , there are at most finitely many subspaces $E \subseteq \mathbb{R}^n$ whose dimension is at most k , that are a -basic for the measure μ . The statement clearly implies the lemma. The case $k = 0$ is easy, as there is only one 0-dimensional subspace in \mathbb{R}^n .

Let $k \geq 1$. Suppose that $n \geq k$ is an integer, $0 < a \leq 1$ and let μ be a Borel probability measure on \mathbb{R}^n . Denote by \mathcal{G} the family of all subspaces $E \subseteq \mathbb{R}^n$ whose dimension is at most k that are a -basic for the measure μ . We need to show that

$$\#(\mathcal{G}) < \infty. \quad (47)$$

First, note that it is sufficient to prove (47) under the additional assumption that $\mu(\{0\}) = 0$. Indeed, denote $\varepsilon = \mu(\{0\})$. If $a \leq \varepsilon$ then there is only one a -basic subspace in \mathbb{R}^n , which is the subspace $\{0\}$, and (47) clearly holds. In the non-trivial case where $a > \varepsilon$, we may replace μ by $(\mu - \varepsilon\delta_0)/(1 - \varepsilon)$ and a by $(a - \varepsilon)/(1 - \varepsilon)$. The family of basic subspaces remains exactly the same. From now on, we will thus assume that $\mu(\{0\}) = 0$.

Denote by $\mathcal{E} \subseteq \mathcal{G}$ the collection of all subspaces $E \subseteq \mathbb{R}^n$ that are a -basic for μ , with $\dim(E) \leq k$, for which $\mu(F) < a^2/8$ for any proper subspace $F \subsetneq E$. We will prove that

$$\#(\mathcal{E}) \leq 2/a < \infty. \quad (48)$$

To that end, let $\tilde{\mathcal{E}}$ be any finite subset of \mathcal{E} , and denote $N = \#(\tilde{\mathcal{E}})$. For any two distinct subspaces $E_1, E_2 \in \tilde{\mathcal{E}}$, we have $\mu(E_1 \cap E_2) < a^2/8$ as $E_1 \cap E_2$ is a proper subspace of E_1 . According to the inclusion-exclusion principle,

$$1 \geq \mu \left(\bigcup_{E \in \tilde{\mathcal{E}}} E \right) \geq \sum_{E \in \tilde{\mathcal{E}}} \mu(E) - \sum_{\substack{E_1, E_2 \in \tilde{\mathcal{E}} \\ E_1 \neq E_2}} \mu(E_1 \cap E_2) > Na - \frac{N(N-1)}{2} a^2/8,$$

where we used the fact that $\mu(E) \geq a$ for any $E \in \tilde{\mathcal{E}}$, since E is a -basic. We conclude that

$$1 > Na - \frac{N(N-1)a^2}{16} \geq Na \left[1 - \frac{Na}{10} \right] \implies |Na - 5| > 3. \quad (49)$$

Thus, there are no finite subsets of \mathcal{E} whose cardinality is $N = \lceil 2/a \rceil$: In this case $2 \leq Na \leq 3$ which is impossible according to (49). Hence $\#(\mathcal{E}) \leq 2/a$ and (48) is proven.

Next, denote by $\tilde{\mathcal{G}}$ the family of all subspaces $E \subseteq \mathbb{R}^n$ that are a -basic, with $\dim(E) \leq k$, for which there exists a proper subspace $F \subsetneq E$ with $\mu(F) \geq a^2/8$. In view of (48), in order to deduce (47) it suffices to show that

$$\#(\tilde{\mathcal{G}}) < \infty. \quad (50)$$

Whenever a subspace $E \subseteq \mathbb{R}^n$ contains a proper subspace $F \subsetneq E$ with $\mu(F) \geq a^2/8$, it also contains an $a^2/8$ -basic proper subspace $\tilde{F} \subsetneq E$ with $\dim(\tilde{F}) \leq k-1$. By the induction hypothesis, there are only finitely many subspaces $\tilde{F} \subseteq \mathbb{R}^n$ that are $a^2/8$ -basic for μ whose dimension is at most $k-1$. Fix such an $a^2/8$ -basic subspace \tilde{F} . Let \mathcal{F} be the collection of all subspaces $E \subseteq \mathbb{R}^n$ that are a -basic, contain \tilde{F} , and satisfy $\dim(E) \leq k$. The task of proving (50) and completing the proof of the lemma is reduced to showing that

$$\#(\mathcal{F}) < \infty.$$

Note that $\dim(\tilde{F}) \geq 1$ as $\mu(\{0\}) = 0 < a^2/8$, and hence $\{0\}$ is not an $a^2/8$ -basic subspace. Denote by $P = Proj_{\tilde{F}^\perp}$ the orthogonal projection operator onto \tilde{F}^\perp in \mathbb{R}^n . Then $\nu = P_*(\mu)$ is a Borel probability measure on \tilde{F}^\perp . For any $E \in \mathcal{F}$, the subspace $P(E)$ is an a -basic subspace for the measure ν , and $\dim(P(E)) = \dim(E) - \dim(\tilde{F}) \leq k-1$. From the induction hypothesis, we see that the set $\{P(E); E \in \mathcal{F}\}$ is finite. However, $P(E_1) \neq P(E_2)$ for any distinct $E_1, E_2 \in \mathcal{F}$. Thus $\#(\mathcal{F}) < \infty$, as promised. The lemma is proven. \square

An alternative proof of Lemma 7.1 was suggested by N. Alon. His idea is to replace the first part of the proof of the induction step with the known fact that there exists a finite set $A \subset \mathbb{R}^n$ that intersects any subspace of measure at least a (see, e.g., Alon and Spencer [1, Section 13.4]).

We write \mathbb{R}^∞ for the linear space of infinite sequences $a = (a_1, a_2, \dots)$ with $a_i \in \mathbb{R}$ for all $i \geq 1$. The space \mathbb{R}^∞ is endowed with the standard product topology (also known as Tychonoff's topology) and the corresponding Borel σ -algebra. The projection map $P_n : \mathbb{R}^\infty \rightarrow \mathbb{R}^n$ is defined by

$$P_n(x) = (x_1, \dots, x_n)$$

for $x = (x_1, x_2, \dots) \in \mathbb{R}^\infty$. Then P_n is a continuous, linear map. Note that any finite-dimensional subspace $E \subset \mathbb{R}^\infty$ is a closed set. Also for any subspace $E \subseteq \mathbb{R}^\infty$ we have

$$\dim(E) = \sup_n \dim(P_n(E)). \quad (51)$$

With a slight abuse of notation, for $m \geq n \geq 1$ we also write $P_n : \mathbb{R}^m \rightarrow \mathbb{R}^n$ for the projection operator defined by $P_n(x_1, \dots, x_m) = (x_1, \dots, x_n)$. We will also use the ridiculous space $\mathbb{R}^0 = \{0\}$, and $P_0(x) = 0$ for any x . Let $\varepsilon > 0$ and let X be a measurable linear space in which all finite-dimensional subspaces are measurable. A probability measure μ on X is called ε -decent if for any finite-dimensional subspace $E \subseteq X$,

$$\mu(E) \leq \varepsilon \dim(E).$$

Lemma 7.2 *Let $\varepsilon > 0$ and let μ be a Borel probability measure on \mathbb{R}^∞ . Suppose that μ is ε -decent. Then, there exists $N \geq 1$ such that $(P_N)_*\mu$ is 2ε -decent.*

Proof: For $n \geq 0$ denote $\mu_n = (P_n)_*\mu$, a Borel probability measure on \mathbb{R}^n . We say that a subspace $E \subseteq \mathbb{R}^n$ is “thick” if $\mu_n(E) \geq 2\varepsilon \dim(E)$. A thick subspace $E \subseteq \mathbb{R}^n$ is necessarily of dimension at most $(2\varepsilon)^{-1}$. We say that E is a “primitive, thick subspace” if it is thick and additionally

$$\mu_n(F) < 2\varepsilon \dim(F)$$

for any proper subspace $F \subsetneq E$. Clearly, any thick subspace $E \subseteq \mathbb{R}^n$ contains a primitive, thick subspace. Observe also that a primitive, thick, k -dimensional subspace $E \subseteq \mathbb{R}^n$ is necessarily $2\varepsilon k$ -basic for the measure μ_n . From Lemma 7.1 we thus learn that for any n , there are only finitely many primitive, thick subspaces $E \subseteq \mathbb{R}^n$.

Denote by \mathcal{V} the collection of all pairs (E, n) such that $E \subseteq \mathbb{R}^n$ is a primitive, thick subspace. In order to prove the lemma, it suffices to show that \mathcal{V} is finite. Indeed, in this case, set $N = \max\{n + 1; \exists E \subseteq \mathbb{R}^n, (E, n) \in \mathcal{V}\}$. Then there are no primitive, thick subspaces in \mathbb{R}^N , and hence there are no thick subspaces in \mathbb{R}^N . Consequently $\mu_N = (P_N)_*(\mu)$ is (2ε) -decent, and the lemma is proven. The rest of the argument is thus concerned with the proof that \mathcal{V} is finite.

Define a directed graph structure on \mathcal{V} as follows: There is an edge going from the node $(E, n) \in \mathcal{V}$ to the node $(F, n + 1) \in \mathcal{V}$ if and only if $E \subseteq P_n(F)$. Note that for each node $(F, n + 1) \in \mathcal{V}$, the subspace $P_n(F) \subseteq \mathbb{R}^n$ is clearly thick, hence it contains a primitive, thick subspace $\tilde{E} \subseteq \mathbb{R}^n$. Therefore each node $(F, n + 1)$ is connected to a certain node $(\tilde{E}, n) \in \mathcal{V}$. We conclude that there is a path from $(\{0\}, 0) \in \mathcal{V}$ to any node in \mathcal{V} . For each $n \geq 1$ there are only finitely many nodes of the form $(E, n) \in \mathcal{V}$, since there are only finitely many primitive, thick subspaces $E \subseteq \mathbb{R}^n$. Therefore, \mathcal{V} is finite if and only if it does not contain an infinite path.

We deduce that in order to prove the lemma, it suffices to show that there is no sequences of subspaces $E_n \subseteq \mathbb{R}^n$ ($n = 0, 1, \dots$) such that for any $n \geq 0$,

$$E_n \subseteq P_n(E_{n+1}) \quad \text{and} \quad (E_n, n) \in \mathcal{V}. \quad (52)$$

Assume by contradiction that a sequence of subspaces satisfying (52) exists. Recall that a subspace of dimension larger than $(2\varepsilon)^{-1}$ cannot be thick, hence $\dim(E_n)$ is bounded by $(2\varepsilon)^{-1}$. Additionally, $\dim(E_n) \leq \dim(E_{n+1})$ for all n . Therefore, there exist $n_0 \geq 1$ and $d \leq (2\varepsilon)^{-1}$ such that

$$\dim(E_n) = d \quad \text{for all } n \geq n_0.$$

Consequently, $E_n = P_n(E_{n+1})$ for any $n \geq n_0$. Consider the direct limit

$$E = \{a \in \mathbb{R}^\infty; P_n(a) \in E_n \text{ for all } n \geq n_0\} \subseteq \mathbb{R}^\infty.$$

Then $E = \bigcap_{n \geq n_0} P_n^{-1}(E_n)$ is a subspace of \mathbb{R}^∞ with $P_n(E) = E_n$ for all $n \geq n_0$. Furthermore, $\dim(E) = d$ according to (51). Note that $P_n^{-1}(E_n) \supset P_{n+1}^{-1}(E_{n+1})$ for any $n \geq n_0$. Therefore

$$\mu(E) = \mu \left(\bigcap_{n \geq n_0} P_n^{-1}(E_n) \right) = \lim_{n \rightarrow \infty} \mu(P_n^{-1}(E_n)) = \lim_{n \rightarrow \infty} \mu_n(E_n) \geq 2\varepsilon d,$$

since $E_n \subseteq \mathbb{R}^n$ is a d -dimensional thick subspace. Hence $\mu(E) \geq 2\varepsilon d$, in contradiction to our assumption that μ is ε -decent. We conclude that there are no infinite paths in \mathcal{V} , and hence that \mathcal{V} is finite. The lemma is proven. \square

Suppose X is a topological vector space. We say that X has a countable separating family of continuous, linear functionals if there exist continuous linear functionals $f_1, f_2, \dots : X \rightarrow \mathbb{R}$ such that for any $x \in X$,

$$x = 0 \iff \forall n, f_n(x) = 0.$$

This condition is not too restrictive. For example, any separable normed space, any separable Fréchet space, and any topological vector space dual to a separable Fréchet space – admits a countable, separating family of continuous, linear functionals.

Corollary 7.3 *Let $\varepsilon > 0$, let $d \geq 1$ be an integer, and let X be a topological vector space with a countable separating family of continuous, linear functionals. Suppose that μ is an ε -decent Borel probability measure on X . Then, there exists a continuous linear map $T : X \rightarrow \mathbb{R}^d$ such that $T_*(\mu)$ is δ -radial proper, for $\delta = c\varepsilon^{c/d}$. Here, $c > 0$ is a universal constant.*

Proof: Let $f_1, f_2, \dots : X \rightarrow \mathbb{R}$ be the separating sequence of continuous linear functionals. Then the linear map $T : X \rightarrow \mathbb{R}^\infty$ defined by

$$T(x) = (f_1(x), f_2(x), \dots)$$

is a continuous linear embedding. Since μ is ε -decent, then also $T_*(\mu)$ is an ε -decent, Borel probability measure on \mathbb{R}^∞ . According to Lemma 7.2, there exists a finite $N \geq 1$ and a continuous linear map $P : \mathbb{R}^\infty \rightarrow \mathbb{R}^N$ such that $(P \circ T)_*(\mu)$ is a 2ε -decent measure on \mathbb{R}^N . The corollary now follows from Theorem 1.3. \square

Note that the linear map T in Corollary 7.3 is not only measurable but also continuous. In principle, we could have formulated Corollary 7.3 for a probability measure on a measurable linear space, without having to rely on an ambient topology: All we need is a linear, measurable embedding in \mathbb{R}^∞ . We refer the reader to Tsirelson [30] for a discussion of measures on infinite-dimensional linear spaces, and for an exposition of Vershik’s “de-topologization” program [31, 32]. We conclude this note with an infinite-dimensional analog of Corollary 1.4.

Corollary 7.4 *Let X be a topological vector space with a countable separating family of continuous, linear functionals. Suppose that μ is a Borel probability measure on X such that $\mu(E) = 0$ for any finite-dimensional subspace $E \subset X$. Then, for any $R > 0$, there exists a non-zero, continuous linear functional $\varphi : X \rightarrow \mathbb{R}$ such that*

$$\mu(\{x; \varphi(x) \geq tM\}) \geq c \exp(-Ct^2) \quad \text{for all } 0 \leq t \leq R$$

and

$$\mu(\{x; \varphi(x) \leq -tM\}) \geq c \exp(-Ct^2) \quad \text{for all } 0 \leq t \leq R$$

where $M > 0$ is a median, that is,

$$\mu(\{x; |\varphi(x)| \leq M\}) \geq 1/2 \quad \text{and} \quad \mu(\{x; |\varphi(x)| \geq M\}) \geq 1/2$$

and $c, C > 0$ are universal constants.

References

- [1] Alon, N., Spencer, J. H., *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York, 2000.
- [2] Barthe, F., *On a reverse form of the Brascamp-Lieb inequality*. Invent. Math., 134, no. 2, (1998), 335–361.
- [3] Beck, J., Chen, W. W. L., *Irregularities of distribution*. Cambridge Tracts in Mathematics, 89. Cambridge University Press, Cambridge, 1987.
- [4] Bobkov, S. G., *On concentration of distributions of random weighted sums*. Ann. Prob., 31, no. 1, (2003), 195–215.
- [5] Bourgain, J., Lindenstrauss, J., Milman, V. D., *Approximation of zonoids by zonotopes*. Acta Math., 162, no. 1-2, (1989), 73–141.
- [6] Bourgain, J., Lindenstrauss, J., Milman, V. D., *Minkowski sums and symmetrizations*. Geometric aspects of functional analysis, Israel seminar (1986–87), Lecture Notes in Math., 1317, Springer, Berlin, (1988), 44–66.
- [7] Carlen, E., Cordero-Erausquin, D., *Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities*. Preprint. Available under <http://arxiv.org/abs/0710.0870>
- [8] Carlen, E., Lieb, E., Loss, M., *A sharp form of Young’s inequality on S^N and related entropy inequalities*. Jour. Geom. Analysis, 14, (2004), 487 – 520.
- [9] Furstenberg, H., *A note on Borel’s density theorem*. Proc. Amer. Math. Soc., 55, no. 1, (1976), 209–212.
- [10] Giannopoulos, A., Pajor, A., Paouris, G., *A note on subgaussian estimates for linear functionals on convex bodies*. Proc. Amer. Math. Soc., 135, (2007), 2599–2606.
- [11] Gromov, M., *Dimension, nonlinear spectra and width*. Geometric aspects of functional analysis, Israel seminar (1986–87), Lecture Notes in Math., 1317, Springer, Berlin, (1988), 132–184.
- [12] Gromov, M., personal communication.
- [13] Klartag, B., *Uniform almost sub-gaussian estimates for linear functionals on convex sets*. Algebra i Analiz (St. Petersburg Math. Journal), Vol. 19, no. 1 (2007), 109–148.
- [14] Klartag, B., *A central limit theorem for convex sets*. Invent. Math., 168, (2007), 91–131.
- [15] Klartag, B., *Power-law estimates for the central limit theorem for convex sets*. J. Funct. Anal., 245, (2007), 284–310.
- [16] Klartag, B., *A Berry-Esseen type inequality for convex bodies with an unconditional basis*. Probab. Theory Related Fields, 45, no. 1, (2009), 1 – 33.
- [17] Klartag, B., *High-dimensional distributions with convexity properties*. Preprint. Available under <http://www.math.tau.ac.il/~klartagb/papers/euro.pdf>
- [18] Klartag, B., Milman, E., *On volume distribution in 2-convex bodies*. Israel J. Math., Vol. 164, (2008), 221–249.
- [19] Koldobsky, A., *Fourier analysis in convex geometry*. Mathematical Surveys and Monographs, 116, American Mathematical Society, Providence, RI, 2005.
- [20] Kolmogoroff, A., *Über das Gesetz des iterierten Logarithmus*. Math. Ann., 101, (1929), 126–135. English translation in Selected Works of A. N. Kolmogorov, vol. II, edited by A. N. Shirayev. Math. and its Applications (Soviet Series), 26, Kluwer, Dordrecht, (1992), 32–42.

- [21] Milman, E., *On Gaussian Marginals of Uniformly Convex Bodies*. J. Theoret. Probab., 22, no. 1, (2009), 256–278.
- [22] Milman, V. D., *A few observations on the connections between local theory and some other fields*. Geometric aspects of functional analysis, Israel seminar (1986–87), Lecture Notes in Math., 1317, Springer, Berlin, (1988), 283–289.
- [23] Milman, V. D., *Dvoretzky’s theorem—thirty years later*. Geom. Funct. Anal., 2, no. 4, (1992), 455–479.
- [24] Milman, V. D., Schechtman, G., *Asymptotic theory of finite-dimensional normed spaces*. Lecture Notes in Math., 1200, Springer-Verlag, Berlin, 1986.
- [25] Nagaev, S. V., *Lower bounds on large deviation probabilities for sums of independent random variables*. In Asymptotic methods in probability and statistics with applications (St. Petersburg, 1998), Stat. Ind. Technol., Birkhäuser, Boston, (2001), 277–295.
- [26] Pisier, G., *The volume of convex bodies and Banach space geometry*. Cambridge Tracts in Mathematics, 94, Cambridge University Press, Cambridge, 1989.
- [27] Schechtman, G., *Two observations regarding embedding subsets of Euclidean spaces in normed spaces*. Adv. Math., 200, no. 1, (2006), 125–135.
- [28] Schechtman, G., *The random version of Dvoretzky’s theorem in ℓ_∞^n* . Geometric aspects of functional analysis, Israel seminar (2004–05), Lecture Notes in Math., 1910, Springer, Berlin, (2007), 265–270.
- [29] Sudakov, V. N., *Typical distributions of linear functionals in finite-dimensional spaces of high-dimension*. (Russian) Dokl. Akad. Nauk. SSSR, 243, no. 6, (1978), 1402–1405. English translation in Soviet Math. Dokl., 19, (1978), 1578–1582.
- [30] Tsirelson, B., *A strange linear space with measure*. Manuscript, 1998.
- [31] Veršik, A. M., *Duality in the theory of measure in linear spaces*. Dokl. Akad. Nauk SSSR, 170, (1966), 497–500. (Russian). English translation in Soviet Math. Dokl., 7, (1966), 1210–1214.
- [32] Veršik, A. M., *The axiomatics of measure theory in linear spaces*. Dokl. Akad. Nauk SSSR, 178, (1968), 278–281. (Russian). English translation in Soviet Math. Dokl., 9, (1968), 68–72.
- [33] Villani, C., *Topics in optimal transportation*. Graduate Studies in Mathematics, 58, American Mathematical Society, 2003.

School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
e-mail address: klartagb@post.tau.ac.il