ON DIMENSION–DEPENDENT CONCENTRATION FOR CONVEX LIPSCHITZ FUNCTIONS IN PRODUCT SPACES

HAN HUANG AND KONSTANTIN TIKHOMIROV

Abstract. Let $n \geq 1$, $K > 0$, and let $X = (X_1, X_2, \dots, X_n)$ be a random vector in $\mathbb{R}^n$ with independent K–subgaussian components. We show that for every 1–Lipschitz convex function f in $\mathbb{R}^n$ (the Lipschitzness with respect to the Euclidean metric),

$$\max\big(\mathbb{P}\big\{f(X) - \mathrm{Med}\, f(X) \geq t\big\}, \mathbb{P}\big\{f(X) - \mathrm{Med}\, f(X) \leq -t\big\}\big) \leq \exp\bigg(-\frac{c\, t^2}{K^2 \log\big(2 + \frac{n}{t^2/K^2}\big)}\bigg), \quad t > 0,$$

where $c > 0$ is a universal constant. The estimates are optimal in the sense that for every $n \geq \tilde{C}$ and $t > 0$ there exist a product probability distribution X in $\mathbb{R}^n$ with K–subgaussian components, and a 1–Lipschitz convex function f, with

$$\mathbb{P}\big\{\big|f(X) - \mathrm{Med}\, f(X)\big| \geq t\big\} \geq \tilde{c}\, \exp\bigg(-\frac{\tilde{C}\, t^2}{K^2 \log\big(2 + \frac{n}{t^2/K^2}\big)}\bigg).$$

The obtained deviation estimates for subgaussian variables are in sharp contrast with the case of variables with bounded $\|X_i\|_{\psi_p}$–norms for $p \in [1, 2)$.

1. Introduction

Concentration in product probability spaces is an active research direction with numerous available results (see, in particular, monographs [18, 7]). Among classical examples of such results are Bernstein–type inequalities [7, Chapter 2] for linear combinations of independent random variables, and the isoperimetric inequality in the Gauss space which implies subgaussian dimension–free concentration [25, 6] (see also [9, 3] as well as [21, Theorem V.1]).

Let $(\Omega_i, \Sigma_i, \mu_i)$, $i \geq 1$, be probability spaces, and for a given $n \geq 1$, let $\mathcal{F}_n$ be a subset of real valued measurable functions f on the product space $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \Sigma_i, \mu_1 \times \dots \times \mu_n)$. A question is to estimate for every $t > 0$ the quantity

$$\sup_{f \in \mathcal{F}_n} \max\big((\mu_1 \times \dots \times \mu_n)\big\{f - \mathrm{Med}\, f \geq t\big\}, (\mu_1 \times \dots \times \mu_n)\big\{f - \mathrm{Med}\, f \leq -t\big\}\big) \tag{1}$$

(we focus on deviation from the median, for concreteness).

First, let $\mathcal{F}_n$ be the class of 1–Lipschitz functions in $\mathbb{R}^n$ (here and further in this note, the Lipschitzness is with respect to the standard Euclidean metric in $\mathbb{R}^n$), and $\mu_1, \dots, \mu_n$ be Borel probability measures in $\mathbb{R}$. In particular, it is known that whenever measures μ_i satisfy a *Poincaré inequality* with a non-trivial constant $\lambda > 0$, i.e

$$\lambda \mathrm{Var}_{\mu_i} h \leq \mathbb{E}_{\mu_i} |h'|^2, \quad 1 \leq i \leq n, \quad \text{for every smooth function } h : \mathbb{R} \to \mathbb{R},$$

then the product measure $\mu_1 \times \dots \times \mu_n$ satisfies the Poincaré inequality in $\mathbb{R}^n$ with the same constant, which in turn implies subexponential dimension–free upper bound $\exp(-ct)$ for (1), where $c > 0$ depends only on the Poincaré constant [14] (see also, for example, [29, Chapter 2]). Conversely, if $\mu = \mu_1 = \mu_2 = \dots$ is a probability measure on $\mathbb{R}$, and for some $t > 0$, (1) is uniformly (over n) upper bounded by a quantity strictly less than $1/2$ then necessarily μ satisfies a Poincaré inequality with a non-trivial constant [11].

A connection between concentration and measure transport inequalities was first highlighted in [19, 20]. In particular, it has been established in the literature (see [23, Section 7], [10, Section 5], [4, Corollary 5.1]) that exponential dimension–free concentration for $\mu^{\times n}$, $n \geq 1$, is equivalent to the inequality

$$\inf_{X \sim \mu,\, Y \sim \nu} \mathbb{E} \min\big(|X - Y|, |X - Y|^2\big) \leq C \int_{\mathbb{R}} \frac{d\nu}{d\mu} \log\big(\frac{d\nu}{d\mu}\big)\, d\mu \quad \text{for every probab. measure } \nu \text{ absolutely cts w.r.t } \mu,$$

where the infimum is taken over all pairs of random variables X, Y, with $X \sim \mu$ and $Y \sim \nu$.

A complete characterization of product measures which enjoy dimension–free *subgaussian* concentration was obtained in [10] (see also earlier work [28]). It was shown in [10] that given a measure μ on $\mathbb{R}$, the quantity in (1) is

Date: June 14, 2021.
K.T. is partially supported by the Sloan Research Fellowship.

upper bounded by $C\exp(-ct^2)$ for some $C,c>0$ (independent of n) if and only if there is a constant $D>0$ such that μ satisfies the following measure transportation inequality (the T_2*–inequality*):

$$\inf_{X\sim\mu,\,Y\sim\nu}\mathbb{E}\,|X-Y|^2\le 2D\int_{\mathbb{R}}\frac{d\nu}{d\mu}\log\big(\frac{d\nu}{d\mu}\big)\,d\mu\quad\text{for every probability measure }\nu\text{ absolutely continuous w.r.t }\mu,$$

where the infimum is over all pairs of random variables X,Y on $\mathbb{R}$ with $X\sim\mu$ and $Y\sim\nu$. We refer to [10] for a more general statement.

We would like to mention the *logarithmic Sobolev inequality* as a well known sufficient condition for subgaussian concentration [8], [18, Chapter 5], as well as inequalities interpolating between log–Sobolev and Poincaré [17] as sufficient conditions for dimension–free concentration estimates of the form $\exp(-ct^p)$ for the quantities in (1).

Following works of Talagrand [26, 27], it has been shown in various settings that by restricting the class of Lipschitz functions to convex (or concave) functions, the worst–case concentration estimates can be significantly improved. As an illustration, it is well known that for every $n\ge 1$, there exists a (non-convex) 1–Lipschitz function f_n in $\mathbb{R}^n$ such that for the random vector $X^{(n)}$ uniformly distributed on vertices of the cube $\{-1,1\}^n$, one has $\mathrm{Var}\,f_n(X^{(n)})=\theta(\sqrt{n})$ (see, for example, [29, Problem 4.9]). On the other hand, a classical result of Talagrand [26, 27] asserts that there is a universal constant $c>0$ such that, with $\mathcal{F}_n:=\{\text{Convex 1–Lipschitz functions in }\mathbb{R}^n\}$, and with $\mu_1=\mu_2=\dots=\mu_n$ being the uniform measure on $\{-1,1\}$, the quantity in (1) is upper bounded by $2\exp(-c\,t^2)$, for a universal constant $c>0$. An extension of Talargand's argument shows that (1) can be upper bounded by $2\exp(-c\,t^2)$ for the class of convex 1–Lipschitz functions whenever $\mu_1,\dots,\mu_n$ are measures with bounded supports (then the constant $c>0$ depends on the largest support diameter) [18, Chapter 4]. A complete characterization of probability measures μ on $\mathbb{R}$ such that (1) admits dimension–free subgaussian concentration for convex 1–Lipschitz functions with $\mu=\mu_1=\mu_2=\dots$, was obtained in [12, 13] (see also [1] for an earlier result in this direction). Both necessary and sufficient condition in that setting is $\mu((t+s,\infty))\le 2\exp(-cs^2)\mu((t,\infty))$ and $\mu((-\infty,-t-s))\le 2\exp(-cs^2)\mu((-\infty,-t))$ for all $s,t>0$ for some constant $c>0$, which can be interpreted as the condition that the distribution μ has "no gaps". The convex subgaussian concentration, in turn, is implied by the *convex* log–Sobolev inequality (see [24]). For results dealing with dimension–free subexponential–type concentration for convex Lipschitz functions, we refer to [5, 13, 2].

Whereas necessary and sufficient conditions for dimension–free concentration are well understood, those conditions are rather strong. For example, it is easy to construct a sugaussian distribution which does not satisfy the condition for dimension–free subgaussian concentration mentioned above. As another illustration, take n i.i.d Bernoulli(q) random variables $b_1,b_2,\dots,b_n$, where $q>0$ is a small parameter. Talagrand's convex distance concentration inequality then implies that for every 1–Lipschitz convex function f in $\mathbb{R}^n$, $\mathbb{P}\big\{|f(b_1,\dots,b_n)-\mathrm{Med}\,f(b_1,\dots,b_n)|\ge t\big\}\le 2\exp(-ct^2)$, $t>0$, for a universal constant $c>0$. However, when $q\to 0$ as $n\to\infty$ and t is sufficiently large, it can be checked that the bound is suboptimal.

The main purpose of this note is to give *optimal dimension–dependent* concentration bound in the class of *subgaussian product measures* for convex 1–Lipschitz functions. However, we would like to start with a discussion of $\|\cdot\|_{\psi_p}$–bounded variables for $p\in[1,2)$, to emphasize the difference in tail behaviour. We recall the definition of the $\|\cdot\|_{\psi_p}$–norm. Given a real valued random variable Y, we set

$$\|Y\|_{\psi_p}:=\inf\big\{\lambda>0:\ \mathbb{E}\exp(|Y|^p/\lambda^p)\le 2\big\},\quad p\ge 1.$$

Theorem 1.1. *For every $p\in[1,2)$ there is a $c_p>0$ depending only on p with the following property. Let $K>0$, $n\ge 2$, and let $X=(X_1,X_2,\dots,X_n)$ be a vector of independent random variables with $\|X_i\|_{\psi_p}\le K$, $1\le i\le n$. Then for every 1–Lipschitz convex function f in $\mathbb{R}^n$, we have*

$$\mathbb{P}\big\{|f(X)-\mathrm{Med}\,f(X)|\ge t\big\}\le 2\exp\big(-c_p\,t^p/K^p\big)+2\exp\big(-c_p\,t^2/\big(K^2(\log n)^{2/p}\big)\big),\quad t>0.$$

We were not able to locate the above theorem in the literature, and provide its proof for completeness. Theorem 1.1 is obtained by a simple reduction to Talagrand's inequality for bounded variables. We note here that the two–level tail behavior for functions of independent variables is a common phenomenon within high–dimensional probability, starting with the classical Bernstein's inequality. It can be informally justified by saying that while deviation of individual variables from the above theorem are controlled by $\exp(-\Theta(t^p))$, linear combinations of variables of the form $\sum_{i=1}^n a_iX_i$ (with $\|a\|_\infty\ll\|a\|_2$) exhibit subgaussian behaviour in a certain range. Notice that, in the above statement, $2\exp\big(-c_p\,t^2/\big(K^2(\log n)^{2/p}\big)$ is the dominating term on the right hand side when $\frac{t}{K}=O\big((\log n)^{\frac{2}{p(2-p)}}\big)$. Further, there is no concentration phenomenon when $\frac{t}{K}=O((\log n)^{1/p})$. For $t\gg K(\log n)^{\frac{2}{p(2-p)}}$, the tail is estimated by $O\big(\exp\big(-c_p\,t^p/K^p\big)\big)$.

It can be verified that the statement of Theorem 1.1 is optimal in the following sense:

Proposition 1.2. *For every $p \in [1,2)$ there is a $C_p > 0$ depending only on p with the following property. Let $n \geq C_p$, $t > 0$, and $K > 0$. Then there exist a random vector $X = (X_1, X_2, \dots, X_n)$ of independent random variables with $\|X_i\|_{\psi_p} \leq K$, $1 \leq i \leq n$, and a convex 1–Lipschitz function f such that*

$$\mathbb{P}\big\{f(X) - \mathrm{Med}\, f(X) \geq t\big\} \geq \tilde{c} \max\big(\exp\big(-\tilde{C}\, t^2/\big(K^2 (\log n)^{2/p}\big), \exp\big(-\tilde{C}\, t^p/K^p\big)\big)$$

and

$$\mathbb{P}\big\{f(X) - \mathrm{Med}\, f(X) \leq -t\big\} \geq \tilde{c} \max\big(\exp\big(-\tilde{C}\, t^2/\big(K^2 (\log n)^{2/p}\big), \exp\big(-\tilde{C}\, t^p/K^p\big)\big).$$

Here, $\tilde{c}, \tilde{C} > 0$ are universal constants.

Now, let $X = (X_1, \dots, X_n)$ be a vector of independent K–subgaussian random variables, that is, $\|X_i\|_{\psi_2} \leq K$, $1 \leq i \leq n$. It is elementary to see that $\|X\|_\infty := \max\limits_{i \leq n} |X_i| = O(\sqrt{\log n})$ with probability, say, $1 - n^{-10}$, where the implicit constant in $O(\cdot)$ depends on K. By considering the vector of truncations $\big(X_i\, \mathbf{1}_{\{|X_i| \leq C\sqrt{\log n}\}}\big)_{i=1}^n$ (for an appropriate choice of C) and applying the Talagrand convex distance inequality, it is elementary to deduce that for every 1–Lipschitz convex function f in $\mathbb{R}^n$,

$$\mathrm{Var}\, f(X_1, \dots, X_n) = O(\log n),$$

where the implicit constant depends on K only. However, getting optimal upper estimates for $\mathbb{P}\{|f(X_1, \dots, X_n) - \mathrm{Med}\, f(X_1, \dots, X_n)| \geq t\}$ in the entire range $t \in (0, \infty)$ appears to require additional arguments rather than the straightforward reduction to the case of bounded variables.

The main statement of this note is

Theorem 1.3. *There is a universal constant $c > 0$ with the following property. Let $K > 0$, $n \geq 2$, and let $X = (X_1, X_2, \dots, X_n)$ be a vector of independent K–subgaussian random variables. Then for every 1–Lipschitz convex function f in $\mathbb{R}^n$, we have*

$$\max\big(\mathbb{P}\big\{f(X) - \mathrm{Med}\, f(X) \geq t\big\}, \mathbb{P}\big\{f(X) - \mathrm{Med}\, f(X) \leq -t\big\}\big) \leq \exp\bigg(-\frac{c\, t^2}{K^2 \log\big(2 + \frac{K^2 n}{t^2}\big)}\bigg), \quad t > 0.$$

The estimate provided by the theorem is optimal in the following sense:

Proposition 1.4. *Let $K > 0$, $n \geq \tilde{C}$, and $t > 0$. Then there exist a vector $X = (X_1, X_2, \dots, X_n)$ of independent K–subgaussian random variables, and a convex 1–Lipschitz function f such that*

$$\mathbb{P}\big\{f(X) - \mathrm{Med}\, f(X) \geq t\big\} \geq \tilde{c} \exp\bigg(-\frac{\tilde{C}\, t^2}{K^2 \log\big(2 + \frac{K^2 n}{t^2}\big)}\bigg),$$

and

$$\mathbb{P}\big\{f(X) - \mathrm{Med}\, f(X) \leq -t\big\} \geq \tilde{c} \exp\bigg(-\frac{\tilde{C}\, t^2}{K^2 \log\big(2 + \frac{K^2 n}{t^2}\big)}\bigg).$$

Here, $\tilde{c}, \tilde{C} > 0$ are universal constants.

The structure of the note is as follows. In Section 2, we provide a proof of Theorem 1.1. Section 3 is devoted to proving Propositions 1.2 and 1.4. Finally, in Section 4 we consider the main result of the note, Theorem 1.3.

2. Proof of Theorem 1.1

Fix $p \in [1,2)$, $K > 0$, a natural number $n \geq 2$, and a 1–Lipschitz convex function f in $\mathbb{R}^n$. To prove the theorem, it is sufficient to verify a deviation inequality for the parameter $t \geq CK(\log n)^{1/p}$, where $C > 0$ is a large constant depending on p. Let $X = (X_1, \dots, X_n)$ be a vector of independent variables with $\|X_i\|_{\psi_p} \leq K$, $1 \leq i \leq n$.

For each number $k \geq 1$, denote

$$Y_i^{(k)} := X_i\, \mathbf{1}_{\{|X_i| \leq 2 \cdot 2^k\, K (\log n)^{1/p}\}}.$$

Further, let $m \geq 1$ be the largest integer such that

$$\frac{t^2}{2^{2m+6}\, K^2 (\log n)^{2/p}} \geq 1,$$

and define

$$u_k := \tilde{c}\, 2^{-(2-p)|m-k|/4}, \quad k \geq 1,$$

where the constant $\tilde{c} = \tilde{C}(p) > 0$ is defined via the relation

$$\tilde{c}\sum_{k=1}^{\infty} 2^{-(2-p)|m-k|/4} = \frac{1}{2}.$$

We start by writing

$$\begin{aligned}&\mathbb{P}\big\{|f(X) - \mathrm{Med}\, f(X)| \geq t\big\}\\ &\quad\leq \mathbb{P}\big\{|f(Y_1^{(1)},\dots,Y_n^{(1)}) - \mathrm{Med}\, f(X)| \geq t/2\big\} + \sum_{k=1}^{\infty}\mathbb{P}\big\{|f(Y_1^{(k+1)},\dots,Y_n^{(k+1)}) - f(Y_1^{(k)},\dots,Y_n^{(k)})| \geq u_k\, t\big\}.\end{aligned}$$

To estimate the probability $\mathbb{P}\big\{|f(Y_1^{(1)},\dots,Y_n^{(1)}) - \mathrm{Med}\, f(X)| \geq t/2\big\}$, we observe that

$$\begin{aligned}&\min\big(\mathbb{P}\big\{f(Y_1^{(1)},\dots,Y_n^{(1)}) \geq \mathrm{Med}\, f(X)\big\}, \mathbb{P}\big\{f(Y_1^{(1)},\dots,Y_n^{(1)}) \leq \mathrm{Med}\, f(X)\big\}\big)\\ &\geq \frac{1}{2} - n\max_{i\leq n}\mathbb{P}\big\{|X_i| \geq 4K(\log n)^{1/p}\big\} \geq \frac{1}{2} - \frac{2}{n^3} \geq \frac{1}{4}.\end{aligned}$$

Hence, applying Talagrand's convex distance inequality for bounded variables, we get

$$\mathbb{P}\big\{|f(Y_1^{(1)},\dots,Y_n^{(1)}) - \mathrm{Med}\, f(X)| \geq t/2\big\} \leq 2\exp\bigg(-\frac{ct^2}{K^2(\log n)^{2/p}}\bigg),$$

for a universal constant $c > 0$.

Further, for every $k \geq 1$ we have

$$\begin{aligned}&\mathbb{P}\big\{|f(Y_1^{(k+1)},\dots,Y_n^{(k+1)}) - f(Y_1^{(k)},\dots,Y_n^{(k)})| \geq u_k\, t\big\}\\ &\quad\leq \mathbb{P}\big\{\big\|\big(Y_i^{(k+1)} - Y_i^{(k)}\big)_{i=1}^n\big\|_2 \geq u_k\, t\big\}\\ &\quad\leq \mathbb{P}\bigg\{\sum_{i=1}^n \mathbf{1}_{\{Y_i^{(k+1)} - Y_i^{(k)} \neq 0\}} \geq \max\Big(1, \frac{u_k^2\, t^2}{2^{2k+6}\, K^2(\log n)^{2/p}}\Big)\bigg\},\end{aligned}$$

where $\mathbf{1}_{\{Y_i^{(k+1)} - Y_i^{(k)} \neq 0\}}$, $1 \leq i \leq n$, are independent Bernoulli random variables with

$$\mathbb{P}\big\{Y_i^{(k+1)} - Y_i^{(k)} \neq 0\big\} \leq \mathbb{P}\big\{|X_i| \geq 2\cdot 2^k\, K(\log n)^{1/p}\big\} \leq \frac{2}{\exp(2^p\cdot 2^{kp}\log n)}, \quad 1 \leq i \leq n.$$

Applying Chernoff's inequality, we get

$$\begin{aligned}&\mathbb{P}\big\{|f(Y_1^{(k+1)},\dots,Y_n^{(k+1)}) - f(Y_1^{(k)},\dots,Y_n^{(k)})| \geq u_k\, t\big\}\\ &\leq \bigg(\frac{2en}{\exp(2^p\cdot 2^{kp}\log n)\max(1, \frac{u_k^2\, t^2}{2^{2k+6}\, K^2(\log n)^{2/p}})}\bigg)^{\max\big(1, \frac{u_k^2\, t^2}{2^{2k+6}\, K^2(\log n)^{2/p}}\big)}\\ &\leq \exp\bigg(-c\, 2^p\cdot 2^{kp}\log n\ \ \max\Big(1, \frac{u_k^2\, t^2}{2^{2k+6}\, K^2(\log n)^{2/p}}\Big)\bigg),\end{aligned}$$

for some universal constant $c > 0$.

For $k \leq m$, we write

$$\begin{aligned}\exp\bigg(-c\, 2^p\cdot 2^{kp}\log n\ \ \max\Big(1, \frac{u_k^2\, t^2}{2^{2k+6}\, K^2(\log n)^{2/p}}\Big)\bigg) &\leq \exp\bigg(-c\,\tilde{c}^2\, 2^p\, 2^{mp}\cdot(\log n)\, \frac{2^{kp-mp-(2-p)(m-k)/2}\, t^2}{2^{2(k-m)}2^{2m+6}\, K^2(\log n)^{2/p}}\bigg)\\ &= \exp\bigg(-c\,\tilde{c}^2\, 2^p\, 2^{mp}\cdot(\log n)\ \ \frac{2^{(2-p)(m-k)/2}\, t^2}{2^{2m+6}\, K^2(\log n)^{2/p}}\bigg)\\ &\leq \exp\big(-c\,\tilde{c}^2\, 2^p\, 2^{mp}\cdot(\log n)\ \ 2^{(2-p)(m-k)/2}\big).\end{aligned}$$

Using the definition of m and assuming the constant C in the assumption for t is sufficiently large, we get

$$\sum_{k\leq m}\mathbb{P}\big\{|f(Y_1^{(k+1)},\dots,Y_n^{(k+1)}) - f(Y_1^{(k)},\dots,Y_n^{(k)})| \geq u_k\, t\big\} \leq \exp\bigg(-\frac{\hat{c}\, t^p}{K^p}\bigg)$$

for some $\hat{c} > 0$ depending only on p.

For $k > m$, we simply write

$$\mathbb{P}\big\{|f(Y_1^{(k+1)},\dots,Y_n^{(k+1)}) - f(Y_1^{(k)},\dots,Y_n^{(k)})| \geq u_k\, t\big\} \leq \exp\big(-c\, 2^p\cdot 2^{(k-m)p}2^{mp}\log n\big),$$

and essentially repeating the above computations, get

$$\sum_{k>m}\mathbb{P}\big\{|f(Y_1^{(k+1)},\dots,Y_n^{(k+1)})-f(Y_1^{(k)},\dots,Y_n^{(k)})|\geq u_k\,t\big\}\leq \exp\Big(-\frac{c''\,t^p}{K^p}\Big)$$

for some $c''>0$ depending only on p.

The result follows.

3. Proof of Propositions 1.2 and 1.4

First, consider the following basic example. Let $p\in[1,2]$, $\tilde K>0$, and let μ be the probability measure on $\mathbb{R}$ defined via the relation

$$\mu([t,\infty))=\mu((-\infty,-t])=\frac{1}{2}\exp\big(-(t/\tilde K)^p\big),\quad t\geq 0.$$

It is easy to see that, with the random vector X in $\mathbb{R}^n$ distributed according to $\mu^{\times n}$, the components of X have $\|\cdot\|_{\psi_p}$–norms bounded by $O(\tilde K)$ (with the absolute implicit constant). On the other hand, with the function $f:\mathbb{R}^n\to\mathbb{R}$ given by

$$f(x_1,x_2,\dots,x_n):=x_1,\quad (x_1,x_2,\dots,x_n)\in\mathbb{R}^n,$$

we have

$$\mathbb{P}\{f(X)\leq -t\}=\mathbb{P}\{f(X)\geq t\}=\frac{1}{2}\exp\big(-(t/\tilde K)^p\big),\ t>0,$$

which gives the required estimates for $t\geq \tilde K(\log n)^{\frac{2}{p(2-p)}}$ in the statement of Proposition 1.2, and for $t\geq\tilde K\sqrt{n}$ in Proposition 1.4.

The main statement of this section is the following:

Proposition 3.1. *There exists a universal constant $C>1$ so that the following holds: Let $n\geq C$, $p\in[1,2]$, $K>0$. Further, let $0\leq t\leq\frac{K}{C}\sqrt{n}$. Then there exists a random vector $X=(X_1,\dots,X_n)$ with i.i.d components whose $\|\cdot\|_{\psi_p}$–norm is bounded above by K such that*

$$\mathbb{P}\big\{\|X\|_2-\operatorname{Med}\|X\|_2\geq t\big\}\geq\frac{1}{C}\exp\Bigg(-C\cdot\frac{t^2}{K^2\big(\log\big(2+\frac{K^2n}{t^2}\big)\big)^{2/p}}\Bigg),\qquad \textit{and}$$

$$\mathbb{P}\big\{\|X\|_2-\operatorname{Med}\|X\|_2\leq -t\big\}\geq\frac{1}{C}\exp\Bigg(-C\cdot\frac{t^2}{K^2\big(\log\big(2+\frac{K^2n}{t^2}\big)\big)^{2/p}}\Bigg).$$

Together with the above example, Proposition 3.1 implies Propositions 1.2 and 1.4. The "test" distribution we use to prove Proposition 3.1 is the n–fold product of a 2–point probability measure defined by $\mu(\{0\})=1-\theta$ and $\mu(\{K\log(1/\theta)^{1/p}\})=\theta$ where $\theta=\theta(t)$ is an appropriately chosen parameter.

The proof of the proposition relies on a precise lower bound for the tail probability of a Binomial random variable. We need the following result:

Lemma 3.2. *There exists universal constants $c_b>0$ and $C_b>1$ so that the following holds. Let n be a sufficiently large integer. For $\theta\in\Big[\frac{1}{c_b\,n},c_b\Big]$, let $Y_1,\dots,Y_n$ be i.i.d Bernoulli random variables with a parameter $\theta>0$. Then, for any $0\leq r\leq n-\theta n$, we have*

$$\mathbb{P}\Big\{\sum_{i=1}^n Y_i\geq\theta n+r\Big\}\geq\frac{1}{C_b}\exp\left(-C_b\log\left(2+\frac{\theta n+r}{\theta n}\right)\frac{r^2}{\theta n+r}\right).\tag{2}$$

Remark 3.3. The term $\frac{r^2}{\theta n+r}$ corresponds to the usual Bernstein–type tail estimate, and $\log\big(2+\frac{\theta n+r}{\theta n}\big)$ is the "extra" factor emerging when $\theta n=o(r)$.

Although the above statement is based on completely standard calculations, we provide its proof for completeness.

Proof of Lemma 3.2. We will assume that $\sqrt{\theta n}$ (and θn) is greater than a sufficiently large universal constant and at the same time θ is smaller than another small universal constant. Those conditions on θ can be imposed by adjusting the constant c_b in the statement of the lemma. For every $k\leq n$, let $P_k:=\mathbb{P}\left\{\sum_{i=1}^n Y_i=k\right\}$ and $P_{\geq k}:=\mathbb{P}\left\{\sum_{i=1}^n Y_i\geq k\right\}$.

We claim that in order to prove the lemma it is sufficient to establish the following inequalities:

$$\forall\, 0\leq r\leq n-\theta n \text{ with } \theta n+r\in\mathbb{N},\quad P_{\geq\theta n+r}\geq\begin{cases}\frac{1}{\tilde C}\exp\Big(-\tilde C\log\Big(2+\frac{r}{\theta n}\Big)r\Big) & \text{if } r\geq\frac{1}{10}\theta n,\\ \frac{1}{\tilde C}\exp\Big(-\tilde C\frac{r^2}{\theta n+r}\Big) & \text{if } 0\leq r<\frac{1}{10}\theta n,\end{cases}\tag{3}$$

for some universal constants $\tilde{C}, \bar{C} > 1$.

To verify the claim, fix any θ (satisfying assumptions from the beginning of the proof) and any r with $0 < r \leq n - \theta n$. We have $P_{\geq \theta n + r} = P_{\geq \lceil \theta n + r \rceil}$.

First, consider the case $\lceil \theta n + r \rceil - \theta n \geq \frac{1}{10}\theta n$. Since θn is greater than a large universal constant, we have $\lceil \theta n + r \rceil \leq \theta n + 2r$, whence, applying (3) with parameters θ and $\lceil \theta n + r \rceil - \theta n$,

$$P_{\geq \lceil \theta n + r \rceil} \geq \frac{1}{\bar{C}} \exp\Big(- \bar{C} \log\Big(2 + \frac{2r}{\theta n}\Big) \cdot 2r\Big) \geq \frac{1}{\bar{C}} \exp\Big(- 4\bar{C} \log\Big(2 + \frac{r}{\theta n}\Big) r\Big),$$

where the last inequality holds since $\log(2 + 2x) \leq \log\big((2 + x)^2\big) = 2\log(2 + x)$ for $x \geq 0$. Further, under the condition $\lceil \theta n + r \rceil - \theta n \geq \frac{\theta n}{10}$ and assuming that θn is larger than a big universal constant, we have $\frac{12r}{\theta n + r} \geq 1$. Therefore,

$$P_{\geq \lceil \theta n + r \rceil} \geq \frac{1}{\bar{C}} \exp\Big(- 4\bar{C} \log\Big(2 + \frac{r}{\theta n}\Big) r\Big) \geq \frac{1}{\bar{C}} \exp\Big(- 4 \cdot 12\bar{C} \log\Big(2 + \frac{\theta n + r}{\theta n}\Big) \frac{r^2}{\theta n + r}\Big).$$

Next, consider the case $0 < r$, $\lceil \theta n + r \rceil - \theta n < \frac{\theta n}{10}$. Clearly, $\lceil \theta n + r \rceil - \theta n \leq r + 1$, and hence

$$P_{\geq \lceil \theta n + r \rceil} \geq \frac{1}{\tilde{C}} \exp\Big(- \tilde{C} \frac{(r+1)^2}{\theta n + r}\Big) \geq \frac{1}{\tilde{C}} \exp\Big(- \tilde{C} \frac{r^2}{\theta n + r} - \tilde{C}\Big),$$

where the last inequality holds since $r \leq \frac{\theta n}{10}$ and θn is sufficiently large. As $\log\Big(2 + \frac{r}{\theta n}\Big) \geq \log(2)$, we obtain

$$P_{\geq \lceil \theta n + r \rceil} \geq \frac{1}{\tilde{C}} \exp(-\tilde{C}) \exp\Big(- \frac{\tilde{C}}{\log(2)} \log\Big(2 + \frac{r}{\theta n}\Big) \frac{r^2}{\theta n + r}\Big),$$

and derivation of (2) from (3) is complete.

From now on, we assume $r \geq$ and $\theta n + r \in \mathbb{N}$. Obviously,

$$P_{\theta n + r} = \binom{n}{\theta n + r} \theta^{\theta n + r} (1 - \theta)^{n - \theta n - r}. \tag{4}$$

Case 1: $\frac{\theta n}{10} \leq r \leq n - \theta n$.

By the standard estimate, $\binom{n}{\theta n + r} \geq \big(\frac{n}{\theta n + r}\big)^{\theta n + r}$, and so

$$P_{\theta n + r} \geq \left(\frac{\theta n}{\theta n + r}\right)^{\theta n + r} (1 - \theta)^{n - \theta n - r} = \exp\left(- \log\left(\frac{\theta n + r}{\theta n}\right)(\theta n + r)\right)(1 - \theta)^{n - \theta n - r}.$$

Since $(1 - \theta) \geq \exp(-2\theta)$ whenever $\theta > 0$ is small enough, we get

$$(1 - \theta)^{n - \theta n - r} \geq (1 - \theta)^n \geq \exp(-2\theta n),$$

and therefore

$$P_{\geq \theta n + r} \geq P_{\theta n + r} \geq \exp\left(- \log\left(\frac{\theta n + r}{\theta n}\right)(\theta n + r) - 2\theta n\right) \geq \exp\left(- C \log\left(\frac{\theta n + r}{\theta n}\right) r\right)$$

for a universal constant $C > 1$. This completes the proof of (3) in the regime $r \geq \frac{\theta n}{10}$.

Case 2: $0 \leq r < \frac{\theta n}{10}$.

In view of Stirling's formula,

$$\begin{aligned} P_{\theta n + r} \geq\, & c \sqrt{\frac{n}{(\theta n + r)\,(n - \theta n - r)}} \left(\frac{n}{\theta n + r}\right)^{\theta n + r} \left(\frac{n}{n - \theta n - r}\right)^{n - \theta n - r} \theta^{\theta n + r} (1 - \theta)^{n - \theta n - r} \\ \geq\, & \frac{c}{\sqrt{\theta n + r}} \left(\frac{\theta n}{\theta n + r}\right)^{\theta n + r} \left(\frac{n - \theta n}{n - \theta n - r}\right)^{n - \theta n - r}, \end{aligned}$$

where $c > 0$ is a universal constant. Since $\log(1 + x) \geq x - x^2$ for $x > 0$, we get

$$\left(\frac{n - \theta n}{n - \theta n - r}\right)^{n - \theta n - r} = \left(1 + \frac{r}{n - \theta n - r}\right)^{n - \theta n - r} \geq \exp\left(r - \frac{r^2}{n - n\theta - r}\right).$$

Similarly, since $\log(1 - x) \geq -x - 2x^2$ for $x \in [0, \frac{1}{2}]$ and $\frac{r}{\theta n + r} \in [0, \frac{1}{2}]$ for $0 \leq r \leq \frac{1}{10}\theta n$,

$$\left(\frac{\theta n}{\theta n + r}\right)^{\theta n + r} = \left(1 - \frac{r}{\theta n + r}\right)^{\theta n + r} \geq \exp\left(-r - \frac{2r^2}{\theta n + r}\right).$$

Hence, together using that $\frac{1}{n-\theta n-r} \le \frac{1}{n-\frac{11}{10}\theta n} \le \frac{1}{\frac{11}{10}\theta n} \le \frac{1}{\theta n+r}$ when $0<\theta<\frac{1}{3}$, we get

$$P_{\theta n+r} \ge \frac{c}{\sqrt{\theta n+r}} \exp\left(-\frac{3r^2}{\theta n+r}\right). \tag{5}$$

The bound $P_{\ge \theta n+r} \ge P_{\theta n+r}$ is insufficient to get (3) when r is small. We will bound $P_{\ge \theta n+r}$ by comparing it with the sum of a geometric sequence starting with $P_{\theta n+r}$.

For $r'>0$ with $\theta n+r' \in \mathbb{N}$ and $n-\theta n-r'>0$, by (4) we have

$$\frac{P_{\theta n+r'+1}}{P_{\theta n+r'}} = \frac{n-\theta n-r'}{\theta n+r'+1}\frac{\theta}{1-\theta} = \frac{1-\frac{r'}{(1-\theta)n}}{1+\frac{1+r'}{\theta n}}.$$

Since $\frac{1}{1+x} \ge 1-x$ for all $x \ge 0$,

$$\frac{P_{\theta n+r'+1}}{P_{\theta n+r'}} \ge \left(1-\frac{r'}{(1-\theta)n}\right)\left(1-\frac{1+r'}{\theta n}\right) \ge 1-\frac{r'}{(1-\theta)n}-\frac{1+r'}{\theta n}.$$

Next, with $\frac{\theta}{1-\theta} \le \frac{c_b}{1-c_b} \le \frac{1}{3}$ when $c_b>0$ is small enough,

$$\frac{P_{\theta n+r'+1}}{P_{\theta n+r'}} \ge 1-\frac{1+\frac{4}{3}r'}{\theta n}.$$

Notice that for $0 \le i \le \max(\lceil r\rceil, \lceil\sqrt{\theta n}\rceil) := u$, we have $1-\frac{1+\frac{4}{3}(r+i)}{\theta n} \ge 1-\frac{4u}{\theta n}$ where we used that $\sqrt{\theta n}$ is greater than a large absolute constant. Hence, for $1 \le i \le u$,

$$P_{\theta n+r+i} \ge P_{\theta n+r}\Big(1-\frac{4u}{\theta n}\Big)^i.$$

Then,

$$P_{\ge\theta n+r} \ge \sum_{i=0}^{u} P_{\theta n+r+i} \ge P_{\theta n+r}\cdot\left(\sum_{i=0}^{u}\Big(1-\frac{4u}{\theta n}\Big)^i\right) = P_{\theta n+r}\cdot\frac{1-\big(1-\frac{4u}{\theta n}\big)^{u+1}}{\frac{4u}{\theta n}} \ge P_{\theta n+r}\cdot\frac{\theta n}{8u}$$

where the last inequality holds since $\big(1-\frac{4u}{\theta n}\big)^{u+1} \le \exp\big(-\frac{4u^2}{\theta n}\big) \le \exp(-4) \le \frac{1}{2}$ since $u \ge \sqrt{\theta n}$. Together with (5), we obtain

$$P_{\ge\theta n+r} \ge \frac{\theta n}{8u}\frac{c}{\sqrt{\theta n+r}}\exp\left(-\frac{3r^2}{\theta n+r}\right).$$

With $\theta n \ge \frac{\theta n+r}{2}$ (since $r \le \frac{\theta n}{10}$) and $u \le 2\max(r, \sqrt{\theta n})$ (if θn is large enough), $\frac{\theta n}{8u}\frac{c}{\sqrt{\theta n+r}} \ge \frac{c}{32}\frac{\sqrt{\theta n+r}}{\max(r,\sqrt{\theta n})}$. Finally, it is easy to check that

$$\frac{\sqrt{\theta n+r}}{\max(r,\sqrt{\theta n})} \ge \exp\Big(-\frac{r^2}{\theta n+r}\Big).$$

Now we conclude that

$$P_{\ge\theta n+r} \ge \frac{c}{32}\exp\left(-\frac{4r^2}{\theta n+r}\right),$$

and the proof of (3) is finished. □

Lemma 3.4. *There exist constants $c_b>0$ and $\tilde{C}_b>1$ so that the following holds. Let n be a sufficiently large integer and let $\alpha>0$. For $\theta \in \Big[\frac{1}{c_b n}, c_b\Big]$, let $Y_1,\dots,Y_n$ be i.i.d Bernoulli random variables with parameter θ. Set $X=(X_1,X_2,\dots,X_n)$, with $X_i=\alpha Y_i$, $i \le n$. Then, for all $t \in \Big[0, \frac{\alpha\sqrt{n}}{4}\Big]$,*

$$\mathbb{P}\big\{\|X\|_2 \ge \mathrm{Med}\|X\|_2+t\big\} \ge \frac{1}{\tilde{C}_b}\exp\left(-\tilde{C}_b\log\left(2+\frac{t^2}{\theta n\alpha^2}\right)\frac{t^2}{\alpha^2}\right). \tag{6}$$

Proof. Clearly,

$$\|X\|_2 = \alpha\sqrt{\sum_{i=1}^n Y_i}.$$

Since the mapping $y \mapsto \alpha\sqrt{y}$ is monotone increasing for $y \ge 0$, the median estimate for Binomial random variable

$$\lfloor\theta n\rfloor \le \mathrm{Med}\Big(\sum_{i=1}^n Y_i\Big) \le \lceil\theta n\rceil$$

(see [16]) implies

$$\alpha\sqrt{\lfloor\theta n\rfloor} \leq \mathrm{Med}\|X\|_2 \leq \alpha\sqrt{\lceil\theta n\rceil}. \tag{7}$$

Thus,

$$|\mathrm{Med}\|X\|_2 - \alpha\sqrt{\theta n}| \leq \alpha\sqrt{\lceil\theta n\rceil} - \alpha\sqrt{\lfloor\theta n\rfloor} \leq \alpha,$$

where the last inequality holds when $\theta n \geq 1$.

We claim that in order to verify the lemma, it is sufficient to establish the following bound:

$$\forall t \in \left[0, \alpha\frac{\sqrt{n}}{2}\right],\ \mathbb{P}\big\{\|X\|_2 \geq \alpha\sqrt{\theta n} + t\big\} \geq \frac{1}{C}\exp\left(-C\log\left(2+\frac{t^2}{\theta n\alpha^2}\right)\frac{t^2}{\alpha^2}\right) \tag{8}$$

for a universal constant $C > 1$. Indeed, suppose (8) holds. For $t \in [0, \frac{\alpha\sqrt{n}}{4}]$,

$$\mathbb{P}\big\{\|X\|_2 \geq \mathrm{Med}\|X\|_2 + t\big\} \geq \mathbb{P}\big\{\|X\|_2 \geq \alpha\sqrt{\theta n} + \alpha + t\big\} \geq \frac{1}{C}\exp\left(-C\log\left(2+\frac{1}{\theta n}\left(\frac{t}{\alpha}+1\right)^2\right)\left(\frac{t}{\alpha}+1\right)^2\right),$$

where the last inequality follows from (8) since $\alpha + t \in \left[0, \alpha\frac{\sqrt{n}}{2}\right]$, under the assumption $n \geq 16$. Since $(\frac{t}{\alpha}+1)^2 \leq 2(\frac{t}{\alpha})^2 + 2$, we get

$$\log\left(2+\frac{1}{\theta n}\left(\frac{t}{\alpha}+1\right)^2\right) \leq \log\left(2+\frac{2}{\theta n}+\frac{2}{\theta n}\left(\frac{t}{\alpha}\right)^2\right) \leq \log\left(2\cdot\left(2+\frac{t^2}{\theta n\alpha^2}\right)\right) \leq 2\log\left(2+\frac{t^2}{\theta n\alpha^2}\right),$$

where we used $\frac{1}{\theta n} \leq 1$ in the second inequality. Then, applying the bounds $(\frac{t}{\alpha}+1)^2 \leq 2(\frac{t}{\alpha})^2+2$ and $\frac{1}{\theta n} \leq 1$ again, we obtain

$$\log\left(2+\frac{1}{\theta n}\left(\frac{t}{\alpha}+1\right)^2\right)\left(\frac{t}{\alpha}+1\right)^2 \leq 4\log\left(2+\frac{t^2}{\theta n\alpha^2}\right)\left(\left(\frac{t}{\alpha}\right)^2+1\right) \leq 4\log(3) + 8\log\left(2+\frac{t^2}{\theta n\alpha^2}\right)\left(\frac{t}{\alpha}\right)^2,$$

where we applied the inequality $\log\left(2+\frac{t^2}{\theta n\alpha^2}\right) \leq \max\left(\log(3), \log\left(2+\frac{t^2}{\theta n\alpha^2}\right)\left(\frac{t}{\alpha}\right)^2\right)$. Therefore,

$$\mathbb{P}\big\{\|X\|_2 \geq \mathrm{Med}\|X\|_2 + t\big\} \geq \frac{1}{C}\exp(-4\log(3)C)\exp\left(-8C\log\left(2+\frac{t^2}{\theta n\alpha^2}\right)\left(\frac{t}{\alpha}\right)^2\right),$$

and (6) follows from (8) with $C_b = \max(C\exp(4\log(3)C), 8C)$. The claim is established.

Now we prove (8). First, since $\|X\|_2 = \alpha\sqrt{\sum_{i=1}^n Y_i}$,

$$\mathbb{P}\big\{\|X\|_2 \geq \alpha\sqrt{\theta n} + t\big\} = \mathbb{P}\left\{\sum_{i=1}^n Y_i - \theta n \geq \underbrace{2\sqrt{\theta n}\frac{t}{\alpha} + \frac{t^2}{\alpha^2}}_{r}\right\}.$$

For $0 \leq \frac{t}{\alpha} \leq \sqrt{\theta n}$, we have $0 \leq r \leq 3\theta n$. We apply Lemma 3.2 and use that $\log(2+\frac{\theta n + r}{\theta n}) \leq \log(6)$, to conclude

$$\begin{aligned}\mathbb{P}\big\{\|X\|_2 \geq \sqrt{\alpha\theta n} + t\big\} &\geq \frac{1}{C_b}\exp\left(-C_b\log(6)\cdot\frac{r^2}{\theta n}\right)\\ &\geq \frac{1}{C_b}\exp\left(-C_b\log(6)\cdot 9\frac{t^2}{\alpha^2}\right) \geq \frac{1}{C_b}\exp\left(-C_b\frac{9\log(6)}{\log(2)}\log\left(2+\frac{t^2}{\theta n\alpha^2}\right)\frac{t^2}{\alpha^2}\right).\end{aligned}$$

For $\sqrt{\theta n} \leq \frac{t}{\alpha} \leq \frac{1}{2}\sqrt{n}$, we have $\theta n \leq r \leq \frac{3t^2}{\alpha^2} \leq \frac{3}{4}n \leq n - \theta n$ where the last inequality holds when $c_b > 0$ is chosen small enough. Applying Lemma 3.2 again, we obtain

$$\mathbb{P}\big\{\|X\|_2 \geq \alpha\sqrt{\theta n} + t\big\} \geq \frac{1}{C_b}\exp\left(-C_b\log\left(2+\frac{6t^2}{\theta n\alpha^2}\right)\cdot\frac{3t^2}{\alpha^2}\right).$$

We have $\log\left(2+\frac{6t^2}{\theta n\alpha^2}\right) \leq 3\log\left(2+\frac{t^2}{\theta n\alpha^2}\right)$, and hence

$$\mathbb{P}\big\{\|X\|_2 \geq \alpha\sqrt{\theta n} + t\big\} \geq \frac{1}{C_b}\exp\left(-9C_b\log\left(2+\frac{t^2}{\theta n\alpha^2}\right)\frac{t^2}{\alpha^2}\right).$$

Now (8) follows by choosing $C := \max\left(\frac{9\log(6)}{\log(2)}, 9\right)C_b$. $\square$

Proof of Proposition 3.1. Let $X(\theta)=(X_1(\theta),\dots,X_n(\theta))$ be the random vector defined in Lemma 3.4 with parameters $\theta\in[\frac{1}{c_b n},c_b]$ and $\alpha:=K(\log(1/\theta))^{1/p}$ (the actual choice of θ will be made later in the proof). Then, $\{X_i(\theta)\}_{i=1}^n$ are i.i.d random variables with the $\|\cdot\|_{\psi_p}$-norm bounded above by K. We want to emphasize that the distribution of X depends on the parameter θ, and that our future choice of θ will also depend on t.

Applying Lemma 3.4 with $0\le t\le \frac{K\sqrt{n}}{4}\le\frac{\alpha\sqrt{n}}{4}$ and any $\theta\in[\frac{1}{c_b n},c_b]$, we get

$$\mathbb{P}\big\{\|X(\theta)\|_2\ge \mathrm{Med}\|X(\theta)\|_2+t\big\}\ge\frac{1}{\tilde C_b}\exp\bigg(-\tilde C_b\log\Big(2+\frac{t^2}{K^2\theta n(\log(1/\theta))^{2/p}}\Big)\frac{t^2}{K^2(\log(1/\theta))^{2/p}}\bigg). \tag{9}$$

Case 1: $t\in\Big[\sqrt{\frac{K^2(\log n)^{2/p}}{3c_b}},\sqrt{\frac{c_bK^2n}{3}}\Big]$. In this case, we define

$$\theta:=\theta(t)=\bigg(\frac{K^2n}{3t^2}\Big(\log\Big(\frac{K^2n}{3t^2}\Big)\Big)^{2/p}\bigg)^{-1}.$$

Since $t\mapsto\theta(t)$ is a monotone increasing function for $t\le K\sqrt{n/3}$, our choice of θ satisfies

$$\frac{1}{c_bn}\le\underbrace{\frac{(\log n)^{2/p}}{c_bn\Big(\log\Big(\frac{c_bn}{(\log n)^{2/p}}\Big)\Big)^{2/p}}}_{\text{when } t=\sqrt{\frac{K^2(\log n)^{2/p}}{3c_b}}}\le\theta\le\underbrace{\frac{c_b}{\Big(\log\Big(\frac{1}{c_b}\Big)\Big)^{2/p}}}_{\text{when } t=\sqrt{\frac{c_bK^2n}{3}}}\le c_b,$$

which conforms to the conditions in Lemma 3.4, and therefore the estimate (9) is valid. Our choice of θ implies $\log(1/\theta)\ge\log\big(\frac{K^2n}{3t^2}\big)$ and thus

$$\begin{aligned}\log\Big(2+\frac{3t^2}{K^2\theta n(\log(1/\theta))^{2/p}}\Big)\frac{3t^2}{K^2(\log(1/\theta))^{2/p}}&=\log\bigg(2+\frac{\big(\log\big(\frac{K^2n}{3t^2}\big)\big)^{2/p}}{(\log(1/\theta))^{2/p}}\bigg)\frac{3t^2}{K^2(\log(1/\theta))^{2/p}}\\&\le\frac{3\log(3)\,t^2}{K^2\big(\log\big(\frac{K^2n}{3t^2}\big)\big)^{2/p}}.\end{aligned}$$

Further, the assumption that $t\le\sqrt{\frac{c_bK^2n}{3}}$ and $c_b>0$ is sufficiently small implies that $\frac{K^2n}{t^2}\ge9$ and therefore

$$\log\Big(\frac{K^2n}{3t^2}\Big)\ge\frac12\log\Big(\frac{K^2n}{t^2}\Big)=\frac14\log\Big(\Big(\frac{K^2n}{t^2}\Big)^2\Big)\ge\frac14\log\Big(2+\frac{K^2n}{t^2}\Big). \tag{10}$$

We conclude that

$$\begin{aligned}\mathbb{P}\big\{\|X(\theta(t))\|_2\ge\mathrm{Med}\|X(\theta(t))\|_2+t\big\}&\ge\frac{1}{\tilde C_b}\exp\bigg(-\tilde C_b\frac{3\log(3)\,t^2}{K^2\big(\log\big(\frac{K^2n}{3t^2}\big)\big)^{2/p}}\bigg)\\&\ge\frac{1}{\tilde C_b}\exp\bigg(-3\cdot4^{2/p}\tilde C_b\log(3)\,\frac{t^2}{K^2\big(\log\big(2+\frac{K^2n}{t^2}\big)\big)^{2/p}}\bigg).\end{aligned}$$

Next, we will handle the lower tail estimate. We can assume that $\lfloor\theta n\rfloor\ge\theta n/3$ since $\theta n\ge\frac{1}{c_b}$ and $c_b>0$ is sufficiently small. Then, by (7) we have

$$\begin{aligned}\mathrm{Med}\|X(\theta(t))\|_2&\ge K(\log(1/\theta))^{1/p}\sqrt{\lfloor\theta n\rfloor}\ge K(\log(1/\theta))^{1/p}\sqrt{\theta n/3}\\&=\sqrt{\bigg(\log\Big(\frac{K^2n}{3t^2}\Big(\log\Big(\frac{K^2n}{3t^2}\Big)\Big)^{2/p}\Big)\bigg)^{2/p}\frac{t^2}{(\log(K^2n/3t^2))^{2/p}}}\ge t.\end{aligned}$$

As a consequence,

$$\begin{aligned}\mathbb{P}\big\{\|X(\theta(t))\|_2\le\mathrm{Med}\|X(\theta(t))\|_2-t\big\}&\ge\mathbb{P}\left\{\|X(\theta(t))\|_2=0\right\}=(1-\theta)^n\ge\exp\big(-2\theta n\big)\\&=\exp\bigg(-\frac{6t^2}{K^2\big(\log(K^2n/3t^2)\big)^{2/p}}\bigg).\end{aligned}$$

Finally, by (10),

$$\mathbb{P}\big\{\|X(\theta(t))\|_2\le\mathrm{Med}\|X(\theta(t))\|_2-t\big\}\ge\exp\bigg(-6\cdot4^{2/p}\frac{t^2}{K^2\big(\log\big(2+\frac{K^2n}{t^2}\big)\big)^{2/p}}\bigg).$$

We have shown that for $\sqrt{\frac{K^2(\log n)^{2/p}}{3c_b}} \le t \le \sqrt{\frac{c_b K^2 n}{3}}$, the proposition holds with $C = \max(48\tilde{C}_b \log(3), 6\cdot 16)$, since $p \ge 1$.

Case 2: $0 \le t \le \sqrt{\frac{K^2(\log n)^{2/p}}{3c_b}}$. Set $t_0 := \sqrt{\frac{K^2(\log n)^{2/p}}{3c_b}}$, and let $\tilde{X} := X(\theta(t_0))$. We have, by the above,

$$\mathbb{P}\big\{\|\tilde{X}\|_2 - \mathrm{Med}\,\|\tilde{X}\|_2 \ge t_0\big\} \ge \frac{1}{C}\exp\bigg(-C\cdot\frac{t_0^2}{K^2\big(\log\big(2+\frac{K^2 n}{t_0^2}\big)\big)^{2/p}}\bigg).$$

When n is greater than a sufficiently large constant,

$$\frac{t_0^2}{K^2\big(\log\big(2+\frac{K^2 n}{t_0^2}\big)\big)^{2/p}} = \frac{(\log n)^{2/p}}{3c_b\Big(\log\Big(2+\frac{3c_b n}{(\log n)^{2/p}}\Big)\Big)^{2/p}} \le \frac{(\log n)^{2/p}}{3c_b\big(\log(\sqrt{n})\big)^{2/p}} \le \frac{2}{3c_b},$$

where we used that $p \ge 1$. We conclude that for $t \in [0, t_0]$,

$$\mathbb{P}\big\{\|\tilde{X}\|_2 - \mathrm{Med}\,\|\tilde{X}\|_2 \ge t\big\} \ge \mathbb{P}\big\{\|\tilde{X}\|_2 - \mathrm{Med}\,\|\tilde{X}\|_2 \ge t_0\big\} \ge \frac{1}{C}\exp\Big(-\frac{2C}{3c_b}\Big).$$

The lower tail is treated the same way. By adjusting the constant C, it implies the proposition for $t \in [0, t_0]$, and completes the proof. □

4. Proof of Theorem 1.3

Our proof of Theorem 1.3 is based on a modification of *the induction method* of Talagrand. In fact, the first part of the proof which deals with setting up a recursive relation for a modified convex distance, essentially repeats, up to minor changes, the standard account of the method (see, for example, [18]).

Definition 4.1. *Given a point $x \in \mathbb{R}^n$ and a non-empty subset A of $\mathbb{R}^n$, we define the* **modified convex distance** *between x and A as*

$$\mathrm{dist}^c(x, A) := \max_{a:\,\|a\|_2=1}\min_{y\in A}\sum_{i=1}^n a_i\,|x_i - y_i|.$$

Remark 4.2. In the standard notion of the convex distance, the indicators $\mathbf{1}_{\{x_i\neq y_i\}}$ are considered instead of the differences $|x_i - y_i|$. Since we work with measures with (possibly) unbounded supports, it is crucial for us to track the "quantitative" distance between x_i and y_i, $i \le n$.

Given a non-empty $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we denote by $U(x, A)$ the set of all vectors in $\mathbb{R}^n_+$ of the form

$$U(x, A) := \bigg\{\Big(|x_i - y_i|\Big)_{i=1}^n \,:\, y \in A\bigg\},$$

and let $V(x, A) \subset \mathbb{R}^n$ be the convex hall of $U(x, A)$.

Lemma 4.3. *We have*

$$\mathrm{dist}^c(x, A) = \mathrm{dist}(0, V(x, A)),$$

where the distance on the right hand side is the usual Euclidean distance in $\mathbb{R}^n$. Furthermore, when A is convex,

$$\mathrm{dist}^c(x, A) = \mathrm{dist}(x, A).$$

Proof. We will provide the proof for the second assertion of the lemma for Reader's convenience. Let A be a non-empty convex set. Without loss of generality, A is closed. There exists a vector $y \in x - A$ such that for all $z \in x - A$, we have $z\cdot y \ge y\cdot y$, so that $\mathrm{dist}(x, A) = \mathrm{dist}(0, x - A) = \|y\|_2$.

Now, for any $z \in \mathbb{R}^n$, let $\tilde{z}$ be the vector obtained from z by replacing each component of z by its absolute value. For each point $z' \in U(x, A)$, there exists $z \in x - A$ such that $z' = \tilde{z}$. Since $\tilde{z}\cdot\tilde{y} \ge z\cdot y \ge \|y\|_2^2$, the set $U(x, A)$ is contained in the half-space $\{w \in \mathbb{R}^n \,:\, w\cdot\tilde{y} \ge \|y\|_2^2\}$, and the same is true for its convex hall $V(x, A)$. Therefore, we conclude that $\mathrm{dist}(0, V(x, A)) = \|\tilde{y}\|_2 = \|y\|_2$ since $\tilde{y} \in V(x, A)$. □

The main technical result in this section is the following:

Proposition 4.4. *Let $K > 0$, and let $\mu_1, \mu_2, \dots, \mu_n$ be K–subgaussian probability measures in $\mathbb{R}$. Let $X = (X_1, X_2, \dots, X_n)$ be distributed in $\mathbb{R}^n$ according to $\mu_1\times\mu_2\times\cdots\times\mu_n$, and let $A \subset \mathbb{R}^n$ be a non-empty Borel subset. Then, for any $\delta \in (0, \frac{1}{2}]$,*

$$\mathbb{E}\exp\bigg(\frac{\tilde{c}\,(\mathrm{dist}^c(X, A))^2}{K^2\log\big(2+\frac{n}{\log(2+1/\delta)}\big)}\bigg) \le \frac{4}{\mathbb{P}\{X \in A\}\delta},$$

where $\tilde{c} > 0$ is a universal constant.

Before we consider the proof, let us show how to derive Theorem 1.3 from the above proposition.

Proof of Theorem 1.3. First, note that it is sufficient to prove the statement for $t \geq C'K\sqrt{\log n}$ for a large constant $C' > 1$. For the upper tail, we let $A := \{x \in \mathbb{R}^n : f(x) \leq \mathrm{Med}\, f(X)\}$. By Proposition 4.4, for any $\delta \in (0, \frac{1}{2}]$,

$$\mathbb{E} \exp\bigg(\frac{\tilde{c}\,(\mathrm{dist}^c(X,A))^2}{K^2 \log\big(2 + \frac{n}{\log(2+1/\delta)}\big)}\bigg) \leq \frac{8}{\delta}.$$

Let $A_t := \{x \in \mathbb{R}^n : \mathrm{dist}^c(x,A) < t\}$. Observe that, since f is convex, so is the set A, and therefore $A_t = \{x \in \mathbb{R}^n : \mathrm{dist}(x,A) < t\}$, in view of Lemma 4.3. Applying Markov's inequality, we get

$$\begin{aligned}\mathbb{P}\{f(X) \geq \mathrm{Med}\, f(X) + t\} &\leq \mathbb{P}\{X \notin A_t\}\\ &\leq \frac{8}{\delta} \exp\bigg(-\frac{\tilde{c}\,t^2}{K^2 \log\big(2 + \frac{n}{\log(2+1/\delta)}\big)}\bigg).\end{aligned}$$

We choose $\delta := \exp\big(-\frac{\tilde{c}\,t^2/4}{K^2\log(2+\frac{K^2 n}{\tilde{c}\,t^2/4})}\big)$ (we can assume that $\delta \leq 1/2$ if C' is sufficiently large). Observe that $\log(2+\frac{1}{\delta}) \geq \log(1/\delta) = \frac{\tilde{c}\,t^2/4}{K^2\log(2+\frac{K^2 n}{\tilde{c}\,t^2/4})}$, and hence

$$\log\Big(2 + \frac{n}{\log(2+1/\delta)}\Big) \leq \log\bigg(2 + \frac{K^2 n}{\tilde{c}\,t^2/4}\log\Big(2+\frac{K^2 n}{\tilde{c}\,t^2/4}\Big)\bigg) \leq 2\log\Big(2+\frac{K^2 n}{\tilde{c}\,t^2/4}\Big).$$

Therefore,

$$\mathbb{P}\{f(X) \geq \mathrm{Med}\, f(X) + t\} \leq 8\exp\bigg(\frac{\tilde{c}\,t^2/4}{K^2\log(2+\frac{K^2 n}{\tilde{c}\,t^2/4})} - \frac{\tilde{c}\,t^2}{2K^2\log\big(2+\frac{K^2 n}{\tilde{c}\,t^2/4}\big)}\bigg) = 8\exp\bigg(-\frac{ct^2}{K^2\log(2+\frac{K^2 n}{ct^2})}\bigg),$$

where $c := \frac{1}{4}\tilde{c}$. By assuming $C' > 1$ to be sufficiently large, we get

$$8\exp\bigg(-\frac{ct^2}{K^2\log(2+\frac{K^2 n}{ct^2})}\bigg) \leq \exp\bigg(-\frac{ct^2/2}{K^2\log(2+\frac{K^2 n}{t^2})}\bigg),$$

which completes treatment of the upper tail.

For the lower tail, we take $A := \{x \in \mathbb{R}^n : f(x) \leq \mathrm{Med}\, f(X) - t\}$ and define $A_t := \{x \in \mathbb{R}^n : \mathrm{dist}^c(x,A) < t\} = \{x \in \mathbb{R}^n : \mathrm{dist}(x,A) < t\}$ (with the last equality due to convexity of A). Then $\{x \in \mathbb{R}^n : f(x) \geq \mathrm{Med}\, f(X)\} \subset A_t^c$ and therefore $\mathbb{P}\{X \in A_t^c\} \geq \frac{1}{2}$. For $\delta \in (0, \frac{1}{2}]$, we have, in view of Proposition 4.4 and Markov's inequality,

$$\frac{1}{2} \leq \mathbb{P}\{X \in A_t^c\} \leq \frac{4}{\mathbb{P}\{X \in A\}\delta} \exp\bigg(-\frac{\tilde{c}\,t^2}{\log\big(2+\frac{n}{\log(2+1/\delta)}\big)}\bigg),$$

which implies

$$\begin{aligned}\mathbb{P}\{f(X) \leq \mathrm{Med}\, f(X) - t\} &= \mathbb{P}\{X \in A\}\\ &\leq \frac{8}{\delta}\exp\bigg(-\frac{\tilde{c}\,t^2}{\log\big(2+\frac{n}{\log(2+1/\delta)}\big)}\bigg).\end{aligned}$$

Now, the same choice of δ leads to the desired bound. $\square$

As we have mentioned above, the proof of Proposition 4.4 is based on the inductive argument (with the dimension as a parameter of the induction). The next proposition sets up the argument:

Proposition 4.5. *Let $n \geq 1$, and let $\mu_1, \mu_2, \dots, \mu_{n+1}$ be probability measures in $\mathbb{R}$. Let $A \subset \mathbb{R}^{n+1}$ be a non-empty subset, and for each $\alpha \in \mathbb{R}$, denote*

$$A(\alpha) := \big\{v \in \mathbb{R}^n : (v, \alpha) \in A\big\}.$$

Let $X = (X_1, X_2, \dots, X_{n+1})$ be distributed in $\mathbb{R}^{n+1}$ according to $\mu_1 \times \mu_2 \times \dots \times \mu_{n+1}$, and X' be the vector of first n components of X. Then for every $\kappa > 0$,

$$\begin{aligned}&\mathbb{E}\exp\big(\kappa \cdot (\mathrm{dist}^c(X,A))^2\big)\\ &\leq \mathbb{E}\inf_{\nu}\bigg[\exp\bigg(\kappa\cdot\Big(\int_{\mathbb{R}} |X_{n+1}-\alpha|\, d\nu(\alpha)\Big)^2\bigg)\prod_{\alpha\in\mathbb{R}}\big(\mathbb{E}\exp\big(\kappa\cdot(\mathrm{dist}^c(X', A(\alpha)))^2\big)\big)^{d\nu(\alpha)}\bigg],\end{aligned}$$

where the infimum is taken over all discrete probability measures ν in $\mathbb{R}$ with a finite support.

Proof. Take arbitrary element $(x,s)\in\mathbb{R}^n\times\mathbb{R}=\mathbb{R}^{n+1}$. Observe that

$$U((x,s),A)=\bigcup_{\alpha\in\mathbb{R}:\,A(\alpha)\neq\emptyset}\big(U(x,A(\alpha))\oplus(|s-\alpha|)\big),$$

where the notation "$\oplus$" should be understood as vector-wise concatenation producing vectors in $\mathbb{R}^{n+1}$. Therefore, every vector of the form

$$\int_{\mathbb{R}}\big(v(\alpha)\oplus(|s-\alpha|)\big)\,d\nu(\alpha)=\Big(\int_{\mathbb{R}}v(\alpha)\,d\nu(\alpha),\int_{\mathbb{R}}|s-\alpha|\,d\nu(\alpha)\Big)\in\mathbb{R}^{n+1},$$

where $v(\alpha)\in V(x,A(\alpha))$, $\alpha\in\mathbb{R}$, and ν is a discrete probability measure on $\mathbb{R}$ with a finite support, belongs to the convex hull $V((x,s),A)$ of $U((x,s),A)$.

Further, we have for every Borel probability measure ν on $\mathbb{R}$ and every choice of $v(\alpha)\in V(x,A(\alpha))$:

$$\Big\|\int_{\mathbb{R}}v(\alpha)\,d\nu(\alpha)\Big\|_2^2\le\Big(\int_{\mathbb{R}}\|v(\alpha)\|_2\,d\nu(\alpha)\Big)^2\le\int_{\mathbb{R}}\|v(\alpha)\|_2^2\,d\nu(\alpha),$$

by Jensen's inequality. Hence,

$$\Big\|\Big(\int_{\mathbb{R}}v(\alpha)\,d\nu(\alpha),\int_{\mathbb{R}}|s-\alpha|\,d\nu(\alpha)\Big)\Big\|_2^2\le\int_{\mathbb{R}}\|v(\alpha)\|_2^2\,d\nu(\alpha)+\Big(\int_{\mathbb{R}}|s-\alpha|\,d\nu(\alpha)\Big)^2.$$

Recall that $\mathrm{dist}^c((x,s),A)=\mathrm{dist}(0,V((x,s),A))$ and $\mathrm{dist}^c(x,A(\alpha))=\mathrm{dist}(0,V(x,A(\alpha)))$ (see Lemma 4.3). Thus, taking $v(\alpha)\in V(x,A(\alpha))$ so that $\|v(\alpha)\|_2=\mathrm{dist}^c(x,A(\alpha))$ for all α, we obtain that

$$\big(\mathrm{dist}^c((x,s),A)\big)^2\le\inf_{\nu}\bigg(\int_{\mathbb{R}}\big(\mathrm{dist}^c(x,A(\alpha))\big)^2\,d\nu(\alpha)+\Big(\int_{\mathbb{R}}|s-\alpha|\,d\nu(\alpha)\Big)^2\bigg),$$

where the infimum is taken over all discrete probability measures ν on $\mathbb{R}$ with a finite support. Clearly, $\mathbb{E}\exp\big(\kappa\cdot(\mathrm{dist}^c(X,A))^2\big)=\mathbb{E}_{X_{n+1}}\mathbb{E}_{X'}\exp\big(\kappa\cdot(\mathrm{dist}^c((X',X_{n+1}),A))^2\big)$. Further, applying the above observation, we get

$$\begin{aligned}
&\mathbb{E}_{X'}\exp\big(\kappa\cdot(\mathrm{dist}^c((X',X_{n+1}),A))^2\big)\\
&\quad\le\mathbb{E}_{X'}\inf_{\nu}\exp\bigg(\kappa\cdot\int_{\mathbb{R}}\big(\mathrm{dist}^c(X',A(\alpha))\big)^2\,d\nu(\alpha)+\kappa\cdot\Big(\int_{\mathbb{R}}|X_{n+1}-\alpha|\,d\nu(\alpha)\Big)^2\bigg)\\
&\quad\le\inf_{\nu}\mathbb{E}_{X'}\exp\bigg(\kappa\cdot\int_{\mathbb{R}}\big(\mathrm{dist}^c(X',A(\alpha))\big)^2\,d\nu(\alpha)+\kappa\cdot\Big(\int_{\mathbb{R}}|X_{n+1}-\alpha|\,d\nu(\alpha)\Big)^2\bigg)\\
&\quad=\inf_{\nu}\bigg[\exp\bigg(\kappa\cdot\Big(\int_{\mathbb{R}}|X_{n+1}-\alpha|\,d\nu(\alpha)\Big)^2\bigg)\mathbb{E}_{X'}\exp\bigg(\kappa\cdot\int_{\mathbb{R}}\big(\mathrm{dist}^c(X',A(\alpha))\big)^2\,d\nu(\alpha)\bigg)\bigg].
\end{aligned}$$

In view of Holder's inequality, the last expression is majorized by

$$\inf_{\nu}\bigg[\exp\bigg(\kappa\cdot\Big(\int_{\mathbb{R}}|X_{n+1}-\alpha|\,d\nu(\alpha)\Big)^2\bigg)\prod_{\alpha\in\mathbb{R}}\Big(\mathbb{E}_{X'}\exp\big(\kappa\cdot\big(\mathrm{dist}^c(X',A(\alpha))\big)^2\Big)^{d\nu(\alpha)}\bigg],$$

and the result follows. ∎

Remark 4.6. The class of measures ν in the above proposition is restricted to discrete measures to avoid any discussion of measurability.

Remark 4.7. By considering two-point probability measures ν of the form $\lambda\delta_{X_{n+1}} + (1-\lambda)\delta_y$, we get from the last proposition

$$\begin{aligned}
&\mathbb{E}\exp\big(\kappa\cdot(\mathrm{dist}^c(X,A))^2\big)\\
&\leq \mathbb{E}\inf_{\nu=\lambda\delta_{X_{n+1}}+(1-\lambda)\delta_y,\,\lambda\in[0,1],\,y\in\mathbb{R}}\Bigg[\exp\bigg(\kappa\cdot\Big(\int_{\mathbb{R}}|X_{n+1}-\alpha|\,d\nu(\alpha)\Big)^2\bigg)\\
&\qquad\qquad\cdot\prod_{\alpha\in\mathbb{R}}\Big(\mathbb{E}\,\exp\big(\kappa\cdot(\mathrm{dist}^c(X',A(\alpha)))^2\big)\Big)^{d\nu(\alpha)}\Bigg]\\
&=\mathbb{E}\inf_{\nu=\lambda\delta_{X_{n+1}}+(1-\lambda)\delta_y,\,\theta\in[0,1],\,y\in\mathbb{R}}\Bigg[\exp\bigg(-\lambda\log\frac{1}{\mathbb{E}\,\exp\big(\kappa\cdot(\mathrm{dist}^c(X',A(X_{n+1})))^2\big)}\\
&\qquad\qquad-(1-\lambda)\log\frac{1}{\mathbb{E}\,\exp\big(\kappa\cdot(\mathrm{dist}^c(X',A(y)))^2\big)}+\kappa\cdot(X_{n+1}-y)^2\,(1-\lambda)^2\bigg)\Bigg].
\end{aligned}$$

Next, we record an elementary fact:

Lemma 4.8. *Let $-\infty \leq b \leq a < +\infty$, and let $c_0 > 0$, $R > 0$. Then*

$$\min_{\lambda\in[0,1]}\Big(-\lambda b-(1-\lambda)a+c_0R^2(1-\lambda)^2\Big)=\begin{cases}-a+c_0R^2, & \text{if } (a-b)\geq 2c_0R^2\\ -b-\frac{(a-b)^2}{4c_0R^2}, & \text{if } (a-b)\leq 2c_0R^2.\end{cases}\tag{11}$$

Proof. We have

$$\begin{aligned}
\min_{\lambda\in[0,1]}\Big(-\lambda b-(1-\lambda)a+c_0R^2(1-\lambda)^2\Big)&=-a+\min_{\lambda\in[0,1]}\Big(c_0R^2(1-\lambda)^2+\lambda(a-b)\Big)\\
&=-a+c_0R^2\min_{\lambda\in[0,1]}\Big(1+\Big(\frac{a-b}{c_0R^2}-2\Big)\lambda+\lambda^2\Big).
\end{aligned}$$

The expression $\big(1+\big(\frac{a-b}{c_0R^2}-2\big)\lambda+\lambda^2\big)$, $\lambda\in[0,1]$, is minimized at $\lambda=\max(0,1-\frac{a-b}{2c_0R^2})$. And (11) follows since

$$1+\Big(\frac{a-b}{c_0R^2}-2\Big)\Big(1-\frac{a-b}{2c_0R^2}\Big)+\Big(1-\frac{a-b}{2c_0R^2}\Big)^2=1-\Big(1-\frac{a-b}{2c_0R^2}\Big)^2=\frac{a-b}{c_0R^2}-\frac{(a-b)^2}{4c_0^2R^4}.$$

□

As an immediate consequence of Remark 4.7 and Lemma 4.8, by considering two-point measures we get

Proposition 4.9. *Let $n\geq 1$, and let $\mu_1,\mu_2,\ldots,\mu_{n+1}$ be probability measures in $\mathbb{R}$. Let $A\subset\mathbb{R}^{n+1}$ be a non-empty subset, and for each $\alpha\in\mathbb{R}$, denote*

$$A(\alpha):=\big\{v\in\mathbb{R}^n:\ (v,\alpha)\in A\big\}.$$

Let $X=(X_1,X_2,\ldots,X_{n+1})$ be distributed in $\mathbb{R}^{n+1}$ according to $\mu_1\times\mu_2\times\cdots\times\mu_{n+1}$, and X' be the vector of first n components of X. Then for every $\kappa>0$,

$$\mathbb{E}\exp\big(\kappa\cdot(\mathrm{dist}^c(X,A))^2\big)\leq\mathbb{E}\inf_{y\in\mathbb{R}}\exp\big(H(X_{n+1},y)\big),$$

where

$$H(t,y):=\min_{\lambda\in[0,1]}\big(-\lambda h(t)-(1-\lambda)h(y)+\kappa(1-\lambda)^2\,(y-t)^2\big),$$

and $h:\mathbb{R}\to\mathbb{R}$ is any function satisfying

$$h(x)\leq\log\frac{1}{\mathbb{E}\,\exp\big(\kappa\cdot(\mathrm{dist}^c(X',A(x)))^2\big)},\quad x\in\mathbb{R}.$$

Moreover, the function $H(t,y)$ can be represented as

$$H(t,y):=\begin{cases}-h(y)+\kappa\,(y-t)^2, & \text{if } h(y)-h(t)\geq 2\kappa\,(y-t)^2,\\ -h(t)-\frac{(h(y)-h(t))^2}{4\kappa(y-t)^2}, & \text{if } 0\leq h(y)-h(t)\leq 2\kappa\,(y-t)^2,\\ -h(t), & \text{if } h(y)-h(t)\leq 0.\end{cases}$$

Remark 4.10. Repeating the optimization argument from [18, p. 74], we get for every pair numbers t, x with $h(y) \geq h(t)$, and for every number $Q \geq 4\kappa\,(y-t)^2$:

$$\begin{aligned} H\,(t,y) &= -h(y) + \min_{\lambda\in[0,1]} \bigg(4\kappa\,(y-t)^2\bigg((1-\lambda)^2/4 - \lambda\,\frac{h(t)-h(y)}{4\kappa\,(y-t)^2}\bigg)\bigg) \\ &\leq -h(y) + Q\log\bigg(2-\exp\bigg(\frac{h(t)-h(y)}{Q}\bigg)\bigg). \end{aligned}$$

The next lemma encapsulates the initial step of the induction:

Lemma 4.11. *Let μ be a K–subgaussian probability measure on $\mathbb{R}$, and X be distributed according to μ. Then for any choice of the parameter $L \geq \sqrt{2}\,K$ and any non-empty Borel subset $A \subset \mathbb{R}$,*

$$\mathbb{E}\exp\bigg(\frac{(\mathrm{dist}^c(X,A))^2}{L^2}\bigg) \leq \frac{4}{\mu(A)}.$$

Proof. WLOG, the set A is closed. Let $x \in A$ be a point with $\mathrm{dist}^c(0,A) = \mathrm{dist}(0,A) = \|x\|_2$. Then

$$\mathbb{E}\exp\bigg(\frac{(\mathrm{dist}^c(X,A))^2}{L^2}\bigg) \leq \mathbb{E}\exp\bigg(\frac{2X^2+2x^2}{L^2}\bigg) \leq 2\exp\bigg(\frac{2x^2}{L^2}\bigg).$$

It remains to note that

$$\mu(A) \leq \mathbb{P}\{|X| \geq x\} = \mathbb{P}\{\exp(2X^2/L^2) \geq \exp(2x^2/L^2)\} \leq \exp(-2x^2/L^2)\,\mathbb{E}\,\exp(2X^2/L^2).$$

The result follows. □

In the next lemma, we deal with "the main part" of the induction argument. The basic idea is to split the argument into two cases, according to how much of the "total mass" of a set A is located far from the origin.

Lemma 4.12. *Let $m \geq 2$, and let A be a non-empty Borel subset of $\mathbb{R}^m$. For each $x \in \mathbb{R}$, let*

$$A(x) := \big\{y \in \mathbb{R}^{m-1} :\ (y,x) \in A\big\}.$$

Further, let $\mu_1, \mu_2, \dots, \mu_m$ be K–subgaussian measures on $\mathbb{R}$, let $X = (X_1, \dots, X_m)$ be distributed according to $\mu_1 \times \mu_2 \times \dots \times \mu_m$, and let X' be the vector of first $m-1$ components of X. Assume that for some $R \geq 1$, $L \geq 16K$ and every $x \in \mathbb{R}$,

$$\mathbb{E}\,\exp\bigg(\frac{(\mathrm{dist}^c(X', A(x)))^2}{L^2}\bigg) \leq \frac{R}{\mathbb{P}\{X' \in A(x)\}}.$$

Then

$$\mathbb{E}\exp\bigg(\frac{(\mathrm{dist}^c(X,A))^2}{L^2}\bigg) \leq \frac{R(1-\exp(-L^2/(64K^2)))^{-2}}{\mathbb{P}\{X \in A\}}.$$

Proof. Without loss of generality, A is closed. Moreover, by an approximation argument, we can (and will) assume that the measures $\mu_1, \dots, \mu_{m-1}$ are supported on finitely many points. Define parameters $\tilde{L} := L/4$ and $M := L/(8K)$.

We consider two cases. First, assume that $\mathbb{P}\{X \in A \text{ and } X_m \in [-\tilde{L}, \tilde{L}])\} \geq (1-\exp(-M^2))\mathbb{P}\{X \in A\}$. In this case, we essentially repeat the standard "induction method" argument employed in the proof of dimension–free subgaussian concentration on the cube. Let x_b be a point in $[-\tilde{L}, \tilde{L}]$ such that $\mathbb{P}\{X' \in A(x_b)\} \geq \mathbb{P}\{X' \in A(x)\}$ for all $x \in [-\tilde{L}, \tilde{L}]$. In view of Proposition 4.9,

$$\mathbb{E}\exp\big((\mathrm{dist}^c(X,A))^2/L^2\big) \leq \mathbb{E}\,\exp\big(H(X_m, x_b)\big),$$

where

$$\begin{aligned} H\,(t, x_b) &: = \min_{\lambda\in[0,1]} \big(-\lambda h(t) - (1-\lambda)h(x_b) + (1-\lambda)^2\,(x_b - t)^2/L^2\big) \\ &= \begin{cases} -h(x_b) + (x_b-t)^2/L^2, & \text{if } h\,(x_b) - h\,(t) \geq 2\,(x_b-t)^2/L^2, \\ -h\,(t) - \frac{L^2(h(x_b)-h(t))^2}{4(x_b-t)^2}, & \text{if } 0 \leq h\,(x_b) - h\,(t) \leq 2\,(x_b-t)^2/L^2, \\ -h\,(t), & \text{if } h\,(x_b) - h\,(t) \leq 0, \end{cases} \end{aligned}$$

and

$$h\,(u) := \log\bigg(\frac{\mathbb{P}\{X' \in A(u)\}}{R}\bigg) \leq \log\frac{1}{\mathbb{E}\,\exp\big((\mathrm{dist}^c(X', A(u)))^2/L^2\big)}, \quad u \in \mathbb{R}. \tag{12}$$

Using the definition of x_b, the equation $\frac{16\tilde{L}^2}{L^2}=1$, and Remark 4.10 with parameter $Q:=1$, we get

$$H(X_m,x_b)\le -h(x_b)+\log\big(2-\exp\big(h(X_m)-h(x_b)\big)\big),\quad \text{whenever } X_m\in[-\tilde{L},\tilde{L}].$$

On the other hand, for all realizations of $X_m\notin[-\tilde{L},\tilde{L}]$ we can crudely bound the function as

$$H(X_m,x_b)\le -h(x_b)+(x_b-X_m)^2/L^2.$$

Combining the relations, we get

$$\begin{aligned}
&\mathbb{E}\,\exp\big((\mathrm{dist}^c(X,A))^2/L^2\big)\\
&\le \mathbb{E}\Big[\exp(-h(x_b))\big(2-\exp\big(h(X_m)-h(x_b)\big)\big)\,\mathbf{1}_{\{X_m\in[-\tilde{L},\tilde{L}]\}}\\
&\qquad+\exp\big(-h(x_b)+(x_b-X_m)^2/L^2\big)\,\mathbf{1}_{\{X_m\notin[-\tilde{L},\tilde{L}]\}}\Big]\\
&\le \frac{R}{\mathbb{P}\{X'\in A(x_b)\}}\mathbb{E}\bigg(\bigg(2-\frac{\mathbb{P}_{X'}\{X'\in A(X_m)\}}{\mathbb{P}\{X'\in A(x_b)\}}\bigg)\mathbf{1}_{\{X_m\in[-\tilde{L},\tilde{L}]\}}+\exp\big(4X_m^2/L^2\big)\,\mathbf{1}_{\{X_m\notin[-\tilde{L},\tilde{L}]\}}\bigg)\\
&=\frac{R}{\mathbb{P}\{X'\in A(x_b)\}}\bigg(2\mathbb{P}\{|X_m|\le\tilde{L}\}-\frac{\mathbb{P}\{X\in A\text{ and }X_m\in[-\tilde{L},\tilde{L}]\}}{\mathbb{P}\{X'\in A(x_b)\}}+\mathbb{E}\big[\exp\big(4X_m^2/L^2\big)\,\mathbf{1}_{\{|X_m|>\tilde{L}\}}\big]\bigg).
\end{aligned}$$

Observe further that

$$\begin{aligned}
\mathbb{E}\big[\exp\big(4X_m^2/L^2\big)\,\mathbf{1}_{\{|X_m|>\tilde{L}\}}\big]&\le 2\exp\big(-\tilde{L}^2/K^2+4\tilde{L}^2/L^2\big)+2\int_{\exp(4\tilde{L}^2/L^2)}^{\infty}s^{-\frac{L^2}{4K^2}}\,ds\\
&\le 4\exp\big(-\tilde{L}^2/K^2+4\tilde{L}^2/L^2\big),
\end{aligned}$$

since $\frac{L^2}{4K^2}\ge 2$. Moreover,

$$\frac{\mathbb{P}\{X\in A\text{ and }X_m\in[-\tilde{L},\tilde{L}]\}}{\mathbb{P}\{X'\in A(x_b)\}}\le\mathbb{P}\{|X_m|\le\tilde{L}\}.$$

Hence,

$$\begin{aligned}
&\mathbb{E}\,\exp\big((\mathrm{dist}^c(X,A))^2/L^2\big)\\
&\le\frac{R\,\mathbb{P}\{|X_m|\le\tilde{L}\}}{\mathbb{P}\{X'\in A(x_b)\}}+\frac{4R\exp\big(-\tilde{L}^2/K^2+4\tilde{L}^2/L^2\big)}{\mathbb{P}\{X'\in A(x_b)\}}\\
&\le\frac{R}{\mathbb{P}\{X\in A\text{ and }X_m\in[-\tilde{L},\tilde{L}]\}}+\frac{4R\exp\big(-\tilde{L}^2/K^2+4\tilde{L}^2/L^2\big)}{\mathbb{P}\{X'\in A(x_b)\}}\\
&\le\frac{R(1-\exp(-M^2))^{-1}}{\mathbb{P}\{X\in A\}}+\frac{4R(1-\exp(-M^2))^{-1}\exp\big(-\tilde{L}^2/K^2+4\tilde{L}^2/L^2\big)}{\mathbb{P}\{X\in A\}}\\
&\le\frac{R(1-\exp(-M^2))^{-1}}{\mathbb{P}\{X\in A\}}+\frac{R(1-\exp(-M^2))^{-1}\exp\big(-2M^2\big)}{\mathbb{P}\{X\in A\}},
\end{aligned}$$

implying the result.

Now, consider the second case: $\mathbb{P}\{X\in A\text{ and }X_m\notin[-\tilde{L},\tilde{L}]\}>\exp(-M^2)\mathbb{P}\{X\in A\}$. Observe that since

$$\mathbb{P}\{X\in A\text{ and }X_m\notin[-\tilde{L},\tilde{L}]\}=\int_{\mathbb{R}\setminus[-\tilde{L},\tilde{L}]}\mathbb{P}_{X'}\{X'\in A(s)\}\,d\mu_m(s)=\mathbb{E}_{X_m}\big(\mathbb{P}_{X'}\{X'\in A(X_m)\}\mathbf{1}_{\{|X_m|>\tilde{L}\}}\big),$$

there must exist a point $x_t\in\mathbb{R}\setminus[-\tilde{L},\tilde{L}]$ with $\mathbb{P}_{X'}\{X'\in A(x_t)\}\ge 2\mathbb{P}\{X\in A\}\exp(2x_t^2/L^2)$. Indeed, if we assume the opposite then, by the above (using that $\frac{L^2}{2K^2}\ge 2$),

$$\begin{aligned}
\exp(-M^2)\,\mathbb{P}\{X\in A\}&<\mathbb{P}\{X\in A\text{ and }X_m\notin[-\tilde{L},\tilde{L}]\}\\
&\le 2\mathbb{P}\{X\in A\}\,\mathbb{E}_{X_m}\big(\exp(2X_m^2/L^2)\mathbf{1}_{\{|X_m|>\tilde{L}\}}\big)\\
&=2\mathbb{P}\{X\in A\}\bigg(\exp(2\tilde{L}^2/L^2)\cdot\mathbb{P}\{|X_m|>\tilde{L}\}+\int_{\exp(2\tilde{L}^2/L^2)}^{\infty}\mathbb{P}\{\exp(2X_m^2/L^2)\ge s\}\,ds\bigg)\\
&\le 8\mathbb{P}\{X\in A\}\,\exp\big(-\tilde{L}^2/K^2+2\tilde{L}^2/L^2\big)\\
&\le 8\mathbb{P}\{X\in A\}\,\exp\big(-L^2/(32K^2)\big)=8\exp(-2M^2)\,\mathbb{P}\{X\in A\},
\end{aligned}$$

leading to contradiction.

Applying again Proposition 4.9, we can bound

$$\mathbb{E}\exp\big((\mathrm{dist}^c(X,A))^2/L^2\big) \le \mathbb{E}\,\exp\big(-h(x_t) + (x_t - X_m)^2/L^2\big),$$

where h is given by (12). Hence,

$$\begin{aligned}\mathbb{E}\exp\big((\mathrm{dist}^c(X,A))^2/L^2\big) &\le \frac{R}{\mathbb{P}\{X' \in A(x_t)\}}\,\exp(2x_t^2/L^2)\,\mathbb{E}\,\exp(2X_m^2/L^2)\\ &\le \frac{2R\exp(2x_t^2/L^2)}{2\mathbb{P}\{X\in A\}\exp(2x_t^2/L^2)},\end{aligned}$$

and the result follows. □

Proof of Proposition 4.4. Let $\delta \in (0, \frac{1}{2}]$ which could be an n–dependent parameter. Define a positive parameter L via the relation

$$L^2 = 512K^2\log\left(2 + \frac{n}{\log(2+1/\delta)}\right).$$

Observe that this choice of L satisfies both Lemmas 4.11 and 4.12. Hence, applying Lemma 4.11 and then Lemma 4.12 inductively $n-1$ times, we get

$$\mathbb{E}\exp\big((\mathrm{dist}^c(X,A))^2/L^2\big) \le \frac{4(1-\exp(-L^2/(64K^2)))^{-2(n-1)}}{\mathbb{P}\{X\in A\}}.$$

Note that

$$\begin{aligned}(1-\exp(-L^2/(64K^2)))^{-2(n-1)} &= \left(1 - \left(2 + \frac{n}{\log(2+1/\delta)}\right)^{-8}\right)^{-2(n-1)}\\ &\le \exp\left(4n\left(2+\frac{n}{\log(2+1/\delta)}\right)^{-8}\right)\\ &= \exp\left(\frac{4n}{2+\frac{n}{\log(2+1/\delta)}}\left(2+\frac{n}{\log(2+1/\delta)}\right)^{-7}\right)\\ &\le \big(2+1/\delta\big)^{4\left(2+\frac{n}{\log(2+1/\delta)}\right)^{-7}} < (2+1/\delta)^{\frac{1}{2}} \le 1/\delta,\end{aligned}$$

since $1/\delta \ge 2$. The result follows. □

References

[1] R. Adamczak, Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses, Bull. Pol. Acad. Sci. Math. **53** (2005), no. 2, 221–238. MR2163396

[2] R. Adamczak and M. Strzelecki, On the convex Poincaré inequality and weak transportation inequalities, Bernoulli **25** (2019), no. 1, 341–374. MR3892322

[3] S. G. Bobkov, An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space, Ann. Probab. **25** (1997), no. 1, 206–214. MR1428506

[4] S. G. Bobkov, I. Gentil and M. Ledoux, Hypercontractivity of Hamilton-Jacobi equations, J. Math. Pures Appl. (9) **80** (2001), no. 7, 669–696. MR1846020

[5] S. G. Bobkov and F. Götze, Discrete isoperimetric and Poincaré-type inequalities, Probab. Theory Related Fields **114** (1999), no. 2, 245–277. MR1701522

[6] C. Borell, The Brunn-Minkowski inequality in Gauss space, Invent. Math. **30** (1975), no. 2, 207–216. MR0399402

[7] S. Boucheron, G. Lugosi and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013. MR3185193

[8] E. B. Davies and B. Simon, Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians, J. Funct. Anal. **59** (1984), no. 2, 335–395. MR0766493

[9] A. Ehrhard, Symétrisation dans l'espace de Gauss, Math. Scand. **53** (1983), no. 2, 281–301. MR0745081

[10] N. Gozlan, A characterization of dimension free concentration in terms of transportation inequalities, Ann. Probab. **37** (2009), no. 6, 2480–2498. MR2573565

[11] N. Gozlan, C. Roberto and P.-M. Samson, From dimension free concentration to the Poincaré inequality, Calc. Var. Partial Differential Equations **52** (2015), no. 3-4, 899–925. MR3311918

[12] N. Gozlan, C. Roberto, P.-M. Samson, P. Tetali, Kantorovich duality for general transport costs and applications, J. Funct. Anal. **273** (2017), no. 11, 3327–3405. MR3706606

[13] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, P. Tetali, Characterization of a class of weak transport-entropy inequalities on the line, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 3, 1667–1693. MR3825894

[14] M. Gromov and V. D. Milman, A topological application of the isoperimetric inequality, Amer. J. Math. **105** (1983), no. 4, 843–854. MR0708367

[15] L. Gross, Logarithmic Sobolev inequalities, Amer. J. Math. **97** (1975), no. 4, 1061–1083. MR0420249

[16] R. Kaas and J.-M. Buhrman, Mean, Median, and Mode in Binomial Distribution, Statistica Neerlandica, **34** (1980), no. 1, 13-18.

[17] R. Latała and K. Oleszkiewicz, Between Sobolev and Poincaré, in *Geometric aspects of functional analysis*, 147–168, Lecture Notes in Math., 1745, Springer, Berlin. MR1796718

[18] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, 89, American Mathematical Society, Providence, RI, 2001. MR1849347

[19] K. Marton, A simple proof of the blowing-up lemma, IEEE Trans. Inform. Theory **32** (1986), no. 3, 445–446. MR0838213

[20] K. Marton, Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration, Ann. Probab. **24** (1996), no. 2, 857–866. MR1404531

[21] V. D. Milman and G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces*, Lecture Notes in Mathematics, 1200, Springer-Verlag, Berlin, 1986. MR0856576

[22] E. Nelson, The free Markoff field, J. Functional Analysis **12** (1973), 211–227. MR0343816

[23] F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. **173** (2000), no. 2, 361–400. MR1760620

[24] Y. Shu and M. Strzelecki, A characterization of a class of convex log-Sobolev inequalities on the real line, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 4, 2075–2091. MR3865667

[25] V. N. Sudakov and B. S. Tsirelson, Extremal properties of half-spaces for spherically invariant measures, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **41** (1974), 14–24, 165. MR0365680

[26] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Inst. Hautes Études Sci. Publ. Math. No. 81 (1995), 73–205. MR1361756

[27] M. Talagrand, A new look at independence, Ann. Probab. **24** (1996), no. 1, 1–34. MR1387624

[28] M. Talagrand, Transportation cost for Gaussian and other product measures, Geom. Funct. Anal. **6** (1996), no. 3, 587–600. MR1392331

[29] R. van Handel, Probability in High Dimension, Lecture notes, https://web.math.princeton.edu/∼rvan/APC550.pdf

Email address: HHUANG421@GATECH.EDU
Email address: KTIKHOMIROV6@GATECH.EDU