ON THE SUBGAUSSIAN COMPARISON THEOREM

RAMON VAN HANDEL

Abstract. The aim of this expository note is to prove that any 1-subgaussian random vector is dominated in the convex ordering by a universal constant times a standard Gaussian vector. This strengthens Talagrand's celebrated subgaussian comparison theorem. The proof combines a tensorization argument due to J. Liu with ideas that date back to the work of Fernique.

1. Introduction

A random vector X in $\mathbb{R}^n$ is said to be 1-subgaussian if $\mathbf{E}[X] = 0$ and

$$\mathbf{P}\big[|\langle v, X\rangle| > x\big] \le 2e^{-x^2/2}$$

for all $x \ge 0$ and $v \in S^{n-1}$, that is, if it is centered and the tail probabilities of its linear projections are dominated by those of a standard Gaussian random variable. The main result of this note is that this weak form of domination implies a much stronger form of domination for the distribution of X.

Theorem 1.1. *Let X be any 1-subgaussian random vector in $\mathbb{R}^n$ and $G \sim N(0, I_n)$ be a standard Gaussian vector in $\mathbb{R}^n$. Then*

$$\mathbf{E}[f(X)] \le \mathbf{E}[f(cG)]$$

for every convex function $f : \mathbb{R}^n \to \mathbb{R}$, where c is a universal constant.[1]

As we will recall below, the conclusion of Theorem 1.1 for 1-*homogeneous* convex functions is a direct consequence of the celebrated majorizing measure theorem of Talagrand [8, §3]. That such a comparison principle holds for *arbitrary* convex functions however appears to have been overlooked. This stronger form of domination is fundamentally more powerful and leads to a better structural understanding of subgaussian vectors. For example, the following corollary provides an equivalent formulation of Theorem 1.1 by a classical result of Strassen [7].

Corollary 1.2. *There is a universal constant c such that for every 1-subgaussian vector X in $\mathbb{R}^n$, we can construct X and a standard Gaussian vector $G \sim N(0, I_n)$ on a common probability space such that $X = c\mathbf{E}[G|X]$.*

Theorem 1.1 will follow almost immediately by observing its connection with some old and recent ideas in the study of suprema of random processes. Beside the formulation of Theorem 1.1 and the more general Theorem 1.3 below, the expository aim of this note is to draw attention to these developments.

2020 *Mathematics Subject Classification.* 60E15; 60G15.

[1] As every convex function is lower bounded by an affine function, the expectations $\mathbf{E}[f(X)]$ and $\mathbf{E}[f(cG)]$ are well defined for every convex function f and take values in $(-\infty, +\infty]$.

1.1. **Random processes.** We begin by formulating a more general form of the subgaussian comparison principle in terms of random processes.

To avoid irrelevant technicalities, we will consider only random processes defined on a finite index set T; the extension of the result below to more general index sets is routine. Let $(G_t)_{t\in T}$ be any centered Gaussian process, and denote by

$$d(t,s) = \|G_t - G_s\|_2$$

the associated natural metric on T. Let $(X_t)_{t\in T}$ be any centered random process that is subgaussian and dominated by $(G_t)_{t\in T}$ in the sense that

$$\mathbf{P}\big[|X_t - X_s| > x\big] \le 2e^{-x^2/2d(t,s)^2}$$

for all $t,s\in T$. Finally, let $(m_t)_{t\in T}$ be any family of real numbers $m_t\in\mathbb{R}$ defined on the same index set. We will prove the following.

Theorem 1.3. *For any centered Gaussian process $(G_t)_{t\in T}$, centered random process $(X_t)_{t\in T}$, and $(m_t)_{t\in T}$ satisfying the above assumptions, we have*

$$\mathbf{E}\bigg[\sup_{t\in T}\big\{X_t + m_t\big\}\bigg] \le \mathbf{E}\bigg[\sup_{t\in T}\big\{cG_t + m_t\big\}\bigg],$$

where c is a universal constant.

Theorem 1.1 follows readily from Theorem 1.3. Indeed, applying Theorem 1.3 with $T\subset\mathbb{R}^n$, $X_t = \langle t, X\rangle$, $G_t = \langle t, G\rangle$ yields the conclusion of Theorem 1.1 for any function $f(x) = \sup_{t\in T}\{\langle t,x\rangle + m_t\}$ that is a finite maximum of affine functions. As any convex function $f:\mathbb{R}^n\to\mathbb{R}$ is the limit of an increasing sequence of functions of this form, Theorem 1.1 follows by monotone convergence.

In the same manner, the special case of Theorem 1.3 with $m_t\equiv 0$ yields the conclusion of Theorem 1.1 for functions of the form $f(x)=\sup_{t\in T}\langle t,x\rangle$, that is, for 1-homogeneous convex functions. This special case follows from the celebrated majorizing measure theorem of Talagrand [8], which states that

$$\frac{1}{c}\mathbf{E}\bigg[\sup_{t\in T} X_t\bigg] \le \gamma_2(T,d) \le c\mathbf{E}\bigg[\sup_{t\in T} G_t\bigg]$$

where $\gamma_2(T,d)$ is an explicit functional that defined in terms of the geometry of the metric space (T,d). As $(G_t)_{t\in T}$ is itself a subgaussian process, replacing $(X_t)_{t\in T}$ by $(G_t)_{t\in T}$ on the left-hand side of this inequality shows that $\gamma_2(T,d)$ characterizes the expected supremum of any *centered* Gaussian process up to a universal constant. Since $\gamma_2(T,d)$ is difficult to compute in concrete situations, however, the application of the majorizing measure theorem as a subgaussian comparison principle has proved to be one of its most useful features in practice.

Theorem 1.3 naturally leads us to seek a generalization of the majorizing measure theorem to *non-centered* Gaussian processes. Despite that the suprema of non-centered processes arise in many applications, the problem of achieving sharp bounds for such processes does not appear to have been discussed in the literature. A common method for handling non-centered processes, the "peeling device", is to split T into slices on which the value of m_t is roughly constant and to estimate the supremum on each slice separately; see, e.g., [10]. While effective in various applications, such a procedure need not lead to sharp bounds.

We presently aim to explain that a form of the majorizing measure theorem for non-centered Gaussian processes is nonetheless already implicitly contained in another, largely forgotten, part of Talagrand's paper [8, §4].

1.2. **Fernique's functional.** We will need the following notion.

Definition 1.4. Let T be a finite set, and let P_X and μ be probability measures on $\mathbb{R}^T$ and T, respectively, with $\int \|x\|\,P_X(dx) < \infty$. We define

$$\mathscr{F}(P_X,\mu) = \sup_{\substack{X\sim P_X\\ Z\sim\mu}} \mathbf{E}[X_Z],$$

where the supremum is over all couplings of P_X and μ. Given a random process $X = (X_t)_{t\in T}$, we will also write $\mathscr{F}(X,\mu) = \mathscr{F}(P_X,\mu)$ where P_X is the law of X.

In other words, the quantity $\mathscr{F}(X,\mu)$ is the largest expected value of the random process X evaluated at a random index with distribution μ. This functional was first introduced by Fernique [3, 4] as a tool for understanding the expected suprema of random processes; some additional comments on the original motivation behind this quantity can be found in section 2 below.

Returning to the setting of Theorem 1.3, we now make a

Trivial observation. *We can write*

$$\mathbf{E}\left[\sup_{t\in T}\{X_t+m_t\}\right] = \sup_{\mu}\left\{\mathscr{F}(X,\mu) + \int m_t\,\mu(dt)\right\},$$

where the supremum is taken over all probability measures μ on T.

Indeed, this follows immediately from $\mathbf{E}[\sup_{t\in T}\{X_t+m_t\}] = \sup_Z \mathbf{E}[X_Z+m_Z]$ where the supremum is over all random variables Z in T.

In the final section of his paper [8, §4], by an elaboration of the methods used to prove the majorizing measure theorem, Talagrand also provides a characterization of Fernique's functional $\mathscr{F}(G,\mu)$ for any centered Gaussian process G and measure μ, up to a universal constant, in terms of a certain geometric functional $\mathrm{I}_\mu(T,d)$ (for example, one may take the quantity Q_3 in [8, Theorem 30] as its definition). When combined with the above trivial observation, this provides the following majorizing measure theorem for non-centered Gaussian processes:

$$\mathbf{E}\left[\sup_{t\in T}\left\{\frac{1}{c}G_t+m_t\right\}\right] \le \sup_{\mu}\left\{\mathrm{I}_\mu(T,d) + \int m_t\,\mu(dt)\right\} \le \mathbf{E}\left[\sup_{t\in T}\{cG_t+m_t\}\right].$$

To complete the proof of Theorem 1.3, it only remains to show that the first inequality remains valid if the Gaussian process $(G_t)_{t\in T}$ is replaced by the subgaussian process $(X_t)_{t\in T}$ on the left-hand side. It seems likely that the methods of Fernique and Talagrand can be used to show that this is the case, but this is not immediately obvious from the proof that is presented in [8].

Instead of pursuing this route, we aim to draw attention to a striking new approach to the majorizing measure theorem that was recently discovered by J. Liu [6] which, as a byproduct, readily yields the comparison principle for $\mathscr{F}(X,\mu)$ (cf.

[6, Corollary 2]) that is needed to complete the proof of Theorem 1.3. The rest of this note is devoted to a short exposition of the proof of this result. A feature that is emphasized in our presentation is that it is now possible to prove comparison theorems such as Theorem 1.3 in an elementary manner that circumvents the need to achieve a complete geometric characterization of the quantities in question.

1.3. **Organization of this note.** The remainder of this note is organized as follows. In section 2, we recall Fernique's classical work on the suprema of Gaussian processes and include some historical comments. Section 3 presents a simple tensorization principle that forms the basis of the work of J. Liu. Finally, section 4 combines these ingredients to complete the proof of Theorem 1.3.

2. On the work of Fernique

The systematic study of the suprema of general Gaussian processes dates back to the work of Dudley and Sudakov in the 1960s. The program of characterizing such suprema in geometric terms was subsequently taken up by Fernique. A major breakthrough, presented in Fernique's 1974 Saint Flour lectures [2], was the complete solution of this problem for *stationary* Gaussian processes.

Given a metric space (T, d), the *covering number* $N(T, d, \varepsilon)$ is the smallest number of ε-balls with respect to the metric d that cover T. A random process $(G_t)_{t\in T}$ will be called *stationary* if there is a group Γ that acts transitively on T such that $(G_{\gamma(t)})_{t\in T}$ has the same distribution as $(G_t)_{t\in T}$ for every $\gamma \in \Gamma$.

Theorem 2.1 (Dudley; Fernique)**.** *Let $(X_t)_{t\in T}$ and $(G_t)_{t\in T}$ be as defined in section 1.1, and suppose that $(G_t)_{t\in T}$ is stationary. Then*

$$\frac{1}{c}\mathbf{E}\left[\sup_{t\in T} X_t\right] \le \int_0^\infty \sqrt{\log N(T, d, \varepsilon)}\, d\varepsilon \le c\mathbf{E}\left[\sup_{t\in T} G_t\right].$$

The first inequality is due to Dudley and the second is due to Fernique. The proofs of both inequalities are based on elementary chaining arguments that are essentially straightforward by modern standards. A simple direct proof of this theorem is sketched at the end of [5, Chapter 6].

The stationarity assumption plays a key role in Theorem 2.1: it ensures that the geometry of (T, d) is self-similar. Major difficulties arise when this assumption is dropped, since the process can then behave in a completely nonhomogeneous manner; indeed, it was known already in the 1960s that the suprema of non-stationary Gaussian processes cannot be characterized in terms of covering numbers. To capture the nonhomogeneity, Fernique introduced a system of weights in his chaining arguments which led him to an improvement of the first inequality of Theorem 2.1 in terms of the notion of a *majorizing measure*. Fernique conjectured that this new upper bound is sharp for all centered Gaussian processes.

A conceptual obstacle to a proof of this conjecture was the lack of a clear probabilistic interpretation of the majorizing measure, which arises in a purely nonprobabilistic manner in the upper bound. Fernique's intuition was that the majorizing measure should be closely connected to the distribution of the maximizer of the Gaussian process. The functional $\mathscr{F}(X, \mu)$ was introduced in [3, 4] in order to elucidate the relation between these notions; see [4, §3.3] and [8, p. 105]. However, this approach does not appear to have led to significant progress. Not the least remarkable aspect of Talagrand's celebrated resolution of Fernique's conjecture [8] is that his proof was entirely geometric in nature, avoiding the need to understand the majorizing measure itself; indeed, majorizing measures play only an incidental role in the definitive contemporary treatment of this subject [9].

Very recently, however, a remarkable idea of J. Liu [6] has led to an unexpected new proof of the majorizing measure theorem that completely bypasses the methods introduced in Talagrand's work. At the core of Liu's approach is the observation that, by a simple tensorization argument that is explained in the following section, the computation of $\mathscr{F}(X,\mu)$ can be reduced to computing the expected supremum of an auxiliary *stationary* random process. The analysis of $\mathscr{F}(X,\mu)$ therefore reduces to the much more elementary setting of Theorem 2.1. This provides a new probabilistic approach to majorizing measures that appears to be much closer in spirit to the program that was originally envisioned by Fernique.

3. Liu's tensorization principle

3.1. **Statement of the principle.** Let T be a finite set, and let

$$\mathcal{P}_K = \left\{ \tfrac{1}{K}\textstyle\sum_{i=1}^K \delta_{t_i} : t_1,\ldots,t_K \in T \right\}$$

be the set of probability measures on T whose atom probabilities are integer multiples of $\frac{1}{K}$. Given any $K,N\in\mathbb{N}$ and $\mu\in\mathcal{P}_K$, we let

$$\mathcal{T}_N(\mu) = \left\{ \mathbf{t}\in T^{NK} : \tfrac{1}{NK}\textstyle\sum_{i=1}^{NK} \delta_{\mathbf{t}_i} = \mu \right\}$$

be the set of sequences in which each $t\in T$ appears exactly $NK\mu(\{t\})$ times.

Proposition 3.1 (Liu's tensorization principle)**.** *Let T be a finite set, $X=(X_t)_{t\in T}$ be a random process with $\max_t \|X_t\|_1<\infty$, and $\mu\in\mathcal{P}_K$. Define*

$$X_{\mathbf{t}} = \tfrac{1}{M}\textstyle\sum_{i=1}^M X^{(i)}_{\mathbf{t}_i}$$

for every $M\in\mathbb{N}$ and $\mathbf{t}\in T^M$, where $X^{(1)},X^{(2)},\ldots$ are i.i.d. copies of X. Then

$$\mathscr{F}(X,\mu) = \lim_{N\to\infty} \mathbf{E}\Bigg[\sup_{\mathbf{t}\in\mathcal{T}_N(\mu)} X_{\mathbf{t}}\Bigg].$$

The point here is that the random process $(X_{\mathbf{t}})_{\mathbf{t}\in\mathcal{T}_N(\mu)}$ is stationary. Indeed, let the symmetric group S_{NK} act on $\mathcal{T}_N(\mu)$ by defining $\sigma(\mathbf{t}) = (\mathbf{t}_{\sigma(1)},\ldots,\mathbf{t}_{\sigma(NK)})$ for every $\mathbf{t}\in\mathcal{T}_N(\mu)$ and $\sigma\in\mathrm{S}_{NK}$. This action is clearly transitive. Moreover, as

$$X_{\sigma(\mathbf{t})} = \tfrac{1}{NK}\textstyle\sum_{i=1}^{NK} X^{(\sigma^{-1}(i))}_{\mathbf{t}_i}$$

and $X^{(1)},X^{(2)},\ldots$ are exchangeable, the processes $(X_{\sigma(\mathbf{t})})_{\mathbf{t}\in\mathcal{T}_N(\mu)}$ and $(X_{\mathbf{t}})_{\mathbf{t}\in\mathcal{T}_N(\mu)}$ have the same distribution for every $\sigma\in\mathrm{S}_{NK}$. Thus Proposition 3.1 reduces the computation of Fernique's functional for an arbitrary random process to the computation of the expected supremum of a stationary process.

Proposition 3.1 is a variant of [6, Lemma 5]. For completeness, we include a short proof of this result in the remainder of this section. We emphasize that this requires no new idea as compared to the arguments in [6].

3.2. **Two simple lemmas.** Recall that the Wasserstein distance between probability measures P_X, P'_X on $\mathbb{R}^T$ is defined as

$$W_1(P_X, P'_X) = \inf_{\substack{X\sim P_X\\ X'\sim P'_X}} \mathbf{E}\|X - X'\|,$$

where the infimum is taken over all couplings of P_X and P'_X. The following straightforward continuity property will be used below.

Lemma 3.2. *Let T be a finite set, μ be a probability measure on T, and P_X, P'_X be probability measures on $\mathbb{R}^T$ with $\int \|x\|\, P_X(dx) < \infty$, $\int \|x\|\, P'_X(dx) < \infty$. Then*

$$|\mathscr{F}(P_X, \mu) - \mathscr{F}(P'_X, \mu)| \le W_1(P_X, P'_X).$$

Proof. Given a pair of random processes (X, X') distributed according to any coupling of $X \sim P_X$ and $X' \sim P'_X$, we can readily estimate

$$\mathscr{F}(P_X, \mu) - \mathscr{F}(P'_X, \mu) \le \sup_{Z\sim\mu} \mathbf{E}[X_Z - X'_Z] \le \mathbf{E}\|X - X'\|,$$

where the supremum is over all couplings of μ with the law of (X, X'). By exchanging the role of P_X and P'_X, the inequality remains valid if the take the absolute value of the left-hand side. It remains to take the infimum over all couplings (X, X'). $\square$

We also recall the following routine consequence of the law of large numbers and the metric properties of the Wasserstein distance.

Lemma 3.3. *Let T be a finite set, and let P_X be a probability measure on $\mathbb{R}^T$ with $\int \|x\|\, P_X(dx) < \infty$. Let $X^{(1)}, X^{(2)}, \ldots$ be i.i.d. copies of $X \sim P_X$. Then*

$$\lim_{N\to\infty} \mathbf{E}\big[W_1(P_X, \hat{P}^N)\big] = 0,$$

where $\hat{P}^N = \frac{1}{N}\sum_{i=1}^N \delta_{X^{(i)}}$ denotes the empirical distribution.

Proof. That $W_1(P_X, \hat{P}^N) \to 0$ a.s. follows directly from the law of large numbers and [12, Theorem 7.12]. Moreover, as $W_1(P_X, \hat{P}^N) \le \int \|x\|\, P_X(dx) + \int \|x\|\, \hat{P}^N(dx)$ and the right-hand side converges in L^1 by the law of large numbers, the sequence $(W_1(P_X, \hat{P}^N))_{N\ge 1}$ is uniformly integrable and thus also $\mathbf{E}[W_1(P_X, \hat{P}^N)] \to 0$. $\square$

3.3. **Proof of the tensorization principle.**

Proof of Proposition 3.1. Let P_X be the law of X, and define $\hat{P}^N$ as in Lemma 3.3. Fix an arbitrary $\mathbf{t}' \in \mathcal{T}_N(\mu)$. Then any coupling of $\hat{P}^{NK}$ and μ can be realized by selecting each pair $(X^{(i)}, \mathbf{t}'_j)$ with probability $\frac{1}{NK}\Pi_{ij}$, where Π is an $NK \times NK$ bistochastic matrix. Denoting the set of such matrices as B_{NK}, we have

$$\mathscr{F}(\hat{P}^{NK}, \mu) = \sup_{\Pi\in \mathrm{B}_{NK}} \tfrac{1}{NK}\textstyle\sum_{i,j=1}^{NK} \Pi_{ij} X^{(i)}_{\mathbf{t}'_j} = \sup_{\sigma\in \mathrm{S}_{NK}} \tfrac{1}{NK}\textstyle\sum_{i=1}^{NK} X^{(i)}_{\sigma(\mathbf{t}')_i} = \sup_{\mathbf{t}\in\mathcal{T}_N(\mu)} X_{\mathbf{t}},$$

where we used that the set of bistochastic matrices is the convex hull of the set of permutation matrices by Birkhoff's theorem [12, p. 5]. Taking the expectation and applying Lemmas 3.2 and 3.3 concludes the proof. $\square$

4. Proof of Theorem 1.3

With the above ingredients in hand, the proof of Theorem 1.3 only requires some minor technicalities. We need the following result that is similar to Lemma 3.2.

Lemma 4.1. *Let T be a finite set, and let P_X be a probability measure on $\mathbb{R}^T$ with $\int \|x\| \, P_X(dx) < \infty$. Then $\mu \mapsto \mathscr{F}(P_X, \mu)$ is continuous in total variation.*

Proof. Let μ, μ' be probability measures on T and let (Z, Z') be any coupling of $Z \sim \mu$ and $Z' \sim \mu'$. Arguing as in the proof of Lemma 3.2, we obtain

$$|\mathscr{F}(P_X, \mu) - \mathscr{F}(P_X, \mu')| \le \sup_{X \sim P_X} \mathbf{E}[X_Z - X_{Z'}] \le 2 \sup_{X \sim P_X} \mathbf{E}[1_{Z \neq Z'} \|X\|]$$

where the supremum is over all couplings of P_X with the law of (Z, Z'). Estimating

$$\mathbf{E}[1_{Z \neq Z'} \|X\|] \le r \mathbf{P}[Z \neq Z'] + \mathbf{E}[\|X\| 1_{\|X\| > r}]$$

and taking the infimum over all couplings (Z, Z'), we can estimate for any $r \ge 0$

$$|\mathscr{F}(P_X, \mu) - \mathscr{F}(P_X, \mu')| \le r \|\mu - \mu'\|_{\mathrm{TV}} + 2 \, \mathbf{E}[\|X\| 1_{\|X\| > r}]$$

using the coupling characterization of the total variation metric [12, p. 7]. □

We can now conclude the proof.

Proof of Theorem 1.3. By the trivial observation in section 1.2, it suffices to prove

$$\mathscr{F}(X, \mu) \le c \mathscr{F}(G, \mu)$$

for every probability measure μ on T.

Let us first fix $K, N \in \mathbb{N}$ and $\mu \in \mathcal{P}_K$, and consider the two random processes $(X_{\mathbf{t}})_{\mathbf{t} \in \mathcal{T}_N(\mu)}$ and $(G_{\mathbf{t}})_{\mathbf{t} \in \mathcal{T}_N(\mu)}$ as defined by Proposition 3.1. Then

$$d_N(\mathbf{t}, \mathbf{s}) := \|G_{\mathbf{t}} - G_{\mathbf{s}}\|_2 = \tfrac{1}{NK} \sqrt{\textstyle\sum_{i=1}^{NK} d(\mathbf{t}_i, \mathbf{s}_i)^2},$$

and it is elementary (see, e.g., [11, Theorem 2.6.2]) that

$$\mathbf{P}\big[|X_{\mathbf{t}} - X_{\mathbf{s}}| > Cx\big] \le 2e^{-x^2/2d_N(\mathbf{t}, \mathbf{s})^2}$$

for a universal constant C. Since $(G_{\mathbf{t}})_{\mathbf{t} \in \mathcal{T}_N(\mu)}$ is stationary, we obtain

$$\mathscr{F}(X, \mu) = \lim_{n \to \infty} \mathbf{E}\Bigg[\sup_{\mathbf{t} \in \mathcal{T}_N(\mu)} X_{\mathbf{t}}\Bigg] \lesssim \lim_{n \to \infty} \mathbf{E}\Bigg[\sup_{\mathbf{t} \in \mathcal{T}_N(\mu)} G_{\mathbf{t}}\Bigg] = \mathscr{F}(G, \mu)$$

by Theorem 2.1 and Proposition 3.1. This proves the desired inequality for every $K \in \mathbb{N}$ and $\mu \in \mathcal{P}_K$. It remains to note that the conclusion extends to an arbitrary probability measure μ on T by continuity using Lemma 4.1. □

For the purpose of proving a subgaussian comparison theorem, the approach that we have followed here completely avoids the need to obtain a geometric characterization of $\mathscr{F}(G, \mu)$. The latter can also be achieved, however: using Theorem 2.1 and Proposition 3.1, this problem reduces to understanding the asymptotics of the covering numbers $N(\mathcal{T}_N(\mu), d_N, \varepsilon)$ as $N \to \infty$, which is a classical problem of coding theory [1]. Such an analysis is developed in detail in the work of J. Liu [6], leading to a new formulation and proof of the majorizing measure theorem.

Acknowledgments. The author thanks Antoine Song for an interesting conversation which motivated the observations that are recorded in this note, and Jingbo Liu for many discussions on his approach to the majorizing measure theorem. This work was supported in part by NSF grant DMS-2347954.

References

[1] I. Csiszár and J. Körner. *Information theory*. Cambridge University Press, Cambridge, second edition, 2011. Coding theorems for discrete memoryless systems.

[2] X. Fernique. Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École d'Été de Probabilités de Saint-Flour, IV-1974*, volume 480 of *Lecture Notes in Math.*, pages 1–96. Springer, 1975.

[3] X. Fernique. Évaluations de processus gaussiens composés. In *Probability in Banach spaces (Proc. First Internat. Conf., Oberwolfach, 1975)*, volume 526 of *Lecture Notes in Math.*, pages 67–83. Springer, 1976.

[4] X. Fernique. Évaluations de certaines fonctionnelles associées à des fonctions aléatoires gaussiennes. *Probab. Math. Statist.*, 2(1):1–29, 1981.

[5] M. Ledoux. Isoperimetry and Gaussian analysis. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, volume 1648 of *Lecture Notes in Math.*, pages 165–294. Springer, 1996.

[6] J. Liu. Simple and sharp generalization bounds via lifting, 2025. arxiv:2508.18682v3.

[7] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.

[8] M. Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987.

[9] M. Talagrand. *Upper and lower bounds for stochastic processes—decomposition theorems*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer, 2nd edition, 2021.

[10] S. A. van de Geer. *Applications of empirical process theory*, volume 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2000.

[11] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018.

[12] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

Department of Mathematics, Princeton University, Princeton, NJ 08544, USA
Email address: `rvan@math.princeton.edu`