Majorizing Measures for the Optimizer

Sander Borst
Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
sander.borst@cwi.nl

Daniel Dadush
Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
d.n.dadush@cwi.nl

Neil Olver
London School of Economics and Political Science, UK
n.olver@lse.ac.uk

Makrand Sinha
Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
makrand.sinha@cwi.nl

Abstract

The theory of majorizing measures, extensively developed by Fernique, Talagrand and many others, provides one of the most general frameworks for controlling the behavior of stochastic processes. In particular, it can be applied to derive quantitative bounds on the expected suprema and the degree of continuity of sample paths for many processes.

One of the crowning achievements of the theory is Talagrand's tight alternative characterization of the suprema of Gaussian processes in terms of majorizing measures. The proof of this theorem was difficult, and thus considerable effort was put into the task of developing both shorter and easier to understand proofs. A major reason for this difficulty was considered to be theory of majorizing measures itself, which had the reputation of being opaque and mysterious. As a consequence, most recent treatments of the theory (including by Talagrand himself) have eschewed the use of majorizing measures in favor of a purely combinatorial approach (the *generic chaining*) where objects based on sequences of partitions provide roughly matching upper and lower bounds on the desired expected supremum.

In this paper, we return to majorizing measures as a primary object of study, and give a viewpoint that we think is natural and clarifying from an optimization perspective. As our main contribution, we give an algorithmic proof of the majorizing measures theorem based on two parts:

- We make the simple (but apparently new) observation that finding the best majorizing measure can be cast as a convex program. This also allows for efficiently computing the measure using off-the-shelf methods from convex optimization.
- We obtain tree-based upper and lower bound certificates by *rounding*, in a series of steps, the primal and dual solutions to this convex program.

While duality has conceptually been part of the theory since its beginnings, as far as we are aware no explicit link to convex optimization has been previously made.

2012 ACM Subject Classification Mathematics of computing → Stochastic processes; Mathematics of computing → Convex optimization; Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases Majorizing measures, Generic chaining, Gaussian processes, Convex optimization, Dimensionality Reduction

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.73

Related Version `https://arxiv.org/abs/2012.13306`

Funding *Sander Borst*: Supported by the ERC Starting grant QIP–805241.
Daniel Dadush: Supported by the ERC Starting grant QIP–805241.
Neil Olver: Supported by NWO VIDI grant 016.Vidi.189.087.
Makrand Sinha: Supported by the NWO VICI grant 639.023.812.

12th Innovations in Theoretical Computer Science Conference (ITCS 2021).
Editor: James R. Lee; Article No. 73; pp. 73:1–73:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Let $(Z_x)_{x \in X}$ denote a family of centered (mean zero) jointly Gaussian random variables, indexed by points of a set X. A fundamental statistic of such a process is the expected supremum $\mathbb{E}[\sup_{x \in X} Z_x]$, which provides an important measure of the size of the process. This statistic has applications in a wide variety of areas. We list some relevant examples. In convex geometry, one can associate a process to any symmetric convex body K, whose supremum gives lower bounds on the size of the largest nearly spherical sections of K [13]. In the context of dimensionality reduction, one can associate a Gaussian process to any point set S in $\mathbb{R}^d$ whose squared expected supremum upper bounds the projection dimension needed to approximately preserve distances between points in S [7, 14]. In the study of Markov Chains, the square of the expected supremum of the Gaussian free field of a graph G was shown to characterize the cover time of the simple random walk on G [3].

The above list of applications, which is by no means exhaustive, help motivate the interest in many areas of Mathematics for obtaining a fine grained understanding of such suprema. We now retrace some of the key developments in the theory of Gaussian processes leading up to Talagrand's celebrated majorizing measure theorem [17], which gives an alternate characterization of Gaussian suprema in terms of an optimization problem over measures on X. The goal of this paper is to give a novel optimization based perspective on this theory, as well as a new *constructive* proof of Talagrand's theorem. For this purpose, some of the earlier concepts, in particular, majorizing measures, will be central to the exposition. We will also cover some generalizations of the theory to the non-Gaussian setting, as our results will be applicable there as well. Throughout our exposition, we rely on the terminology introduced by van Handel [22] for the various combinatorial objects within the theory (i.e., labelled nets, admissible nets and packing trees).

1.1 Bounding the Supremum of Stochastic Processes

In what follows we use the notation $A \lesssim B$ ($A \gtrsim B$) if there exists an absolute constant $c > 0$ such that $A \leq cB$ ($cA \geq B$). We use $A \asymp B$ to denote $A \lesssim B$ and $A \gtrsim B$.

A first basic question one may ask is what information about the Gaussian process $(Z_x)_{x \in X}$ is sufficient to exactly characterize the expected supremum? An answer to this problem was given by Sudakov [16], strengthening a result of Slepian [15]. Sudakov showed that it is uniquely identified by the natural (pseudo) distance metric

$$d(u, v) := \mathbb{E}[(Z_u - Z_v)^2]^{1/2}, \qquad \forall u, v \in X. \tag{1.1}$$

In fact, Sudakov proved the following stronger comparison theorem: if $(Y_x)_{x \in X}$ and $(Z_x)_{x \in X}$ are Gaussian processes on the same index set X and for every $u, v \in X$, it holds that $\mathbb{E}[(Y_u - Y_v)^2] \leq \mathbb{E}[(Z_u - Z_v)^2]$, then $\mathbb{E}[\sup_x Y_x] \leq \mathbb{E}[\sup_x Z_x]$.

Given the above, it is natural to wonder what properties of the metric space X allow us to obtain upper and lower bounds on $\mathbb{E}[\sup_{x \in X} Z_x]$? A first intuitively relevant quantity is the diameter of X defined by $\mathbf{D}(X) := \sup_{u,v \in X} d(u, v)$. For any $u, v \in X$, we have the following simple lower bound:

$$\begin{aligned}\mathbb{E}[\sup_x Z_x] \geq \mathbb{E}[\max\{Z_u, Z_v\}] &= \mathbb{E}[\max\{Z_u - Z_v, 0\}] + \mathbb{E}[Z_v] \\ &= \tfrac{1}{2}\mathbb{E}[|Z_u - Z_v|] = \frac{d(u, v)}{\sqrt{2\pi}}.\end{aligned} \tag{1.2}$$

Here, we use that $\mathbb{E}[Z_v] = 0$, and that $Z_u - Z_v$ is Gaussian with variance $d(u, v)^2$. Thus $\mathbb{E}[\sup_x Z_x] \geq \mathbf{D}(X)/\sqrt{2\pi}$.

Instead of looking at two maximally separated points, one might expect to get stronger lower bounds using a large set of well-separated points in X. Such an inequality was given by Sudakov [16], who showed that

$$\max_{r>0} r\sqrt{\log N_X(r)} \lesssim \mathbb{E}[\sup_x Z_x],$$

where $N_X(r) := \min\{|S| \mid S \subseteq X, \forall x \in X, \min_{s\in S} d(x,s) \leq r\}$ is the minimum size of an r-net of X. This is in fact a direct consequence of Sudakov's comparison theorem. Precisely, the restriction of the process $\{Z_x\}_{x\in X}$ to a suitable r-net S, chosen greedily so that every two points in S are at distance at least r, majorizes the maximum of $|S| \geq N_X(r)$ independent Gaussians with standard deviation $r/\sqrt{2}$, where a standard computation then yields the left-hand side.

On the upper bound side, Dudley [4] proved that the covering numbers can in fact be *chained* together to upper bound the supremum:

$$\mathbb{E}[\sup_x Z_x] \lesssim \int_0^\infty \sqrt{\log N_X(\mathsf{r})} d\mathsf{r}. \tag{1.3}$$

Note that the integral can be restricted to the range $\mathsf{r} \in (0, \mathbf{D}(X)]$, since $\log N_X(\mathbf{D}(X)) = \log 1 = 0$. Dudley's proof of this inequality was extremely influential and showed the power of combining simple tail bounds on pairs of variables $Z_u - Z_v$ to get a global bound on the supremum. In particular, the main inequality used in Dudley's proof is the standard Gaussian tail bound: for $u, v \in X$, and for any $s > 0$

$$\mathbb{P}[|Z_u - Z_v| \geq d(u,v) \cdot s] \leq 2e^{-s^2/2}. \tag{1.4}$$

The strategy of combining the above inequalities to control the maximum of a process is what is now called chaining.

Basics of Chaining. The concept of chaining is central to this paper, so we explain the basic mechanics here. As it will be more convenient for the exposition, we will more directly work with symmetric version of the supremum

$$\sup_{x_1,x_2\in X} Z_{x_1} - Z_{x_2}$$

which is always non-negative. Note that since $(Z_x)_{x\in X}$ and $(-Z_x)_{x\in X}$ are identically distributed,

$$\mathbb{E}\left[\sup_{x_1,x_2\in X} Z_{x_1} - Z_{x_2}\right] = \mathbb{E}\left[\sup_{x\in X} Z_x\right] + \mathbb{E}\left[\sup_{x\in X} -Z_x\right] = 2\mathbb{E}\left[\sup_{x\in X} Z_x\right],$$

and thus the expected supremum is the same after dividing by 2.

From here, instead of bounding the expectation, we focus on upper bounding the median of $\sup_{x_1,x_2\in X} Z_{x_1} - Z_{x_2}$, which is known to be within a constant factor of the expectation. Precisely, we seek to compute a number $M > 0$ such that $\mathbb{P}[\sup_{x_1,x_2\in X} Z_{x_1} - Z_{x_2} \geq M] \leq 1/2$. To arrive at such bounds, we define the notion of a *chaining tree.*

▶ **Definition 1.1** (Chaining Tree). *A (Gaussian) chaining tree $\mathcal{C}$ for a finite metric space (X,d) is a rooted spanning tree on X, with root node $w \in X$, together with probability labels $\mathsf{p}_e \in (0,1/2)$, for each edge $e \in E[\mathcal{C}]$. The edge probabilities are required to satisfy*

$\sum_{e\in E[\mathcal{C}]} p_e \le 1/2$. For each edge $\{u,v\} = e \in E[\mathcal{C}]$, we define the induced edge length $l_e := l_e(p_e, e)$ to satisfy

$$\mathbb{P}_{Z\in\mathcal{N}(0,d(u,v)^2)}[|Z| \ge l_e] = p_e. \tag{1.5}$$

For each $x \in X$, let $\mathcal{P}_x$ denote the unique path from x to the root w in $\mathcal{C}$. We define the value of $\mathcal{C}$ to be

$$\mathsf{val}(\mathcal{C}) = \max_{x\in X} \sum_{e\in\mathcal{P}_x} l_e. \tag{1.6}$$

For a Gaussian process $(Z_x)_{x\in X}$, where d is the induced metric as in (1.1), for any chaining tree $\mathcal{C}$ on X, we now show that

$$\mathbb{P}\left[\sup_{x_1,x_2} Z_{x_1} - Z_{x_2} \ge 2\cdot\mathsf{val}(\mathcal{C})\right] \le 1/2. \tag{1.7}$$

By construction, for any edge $\{u,v\} \in E[\mathcal{C}]$ we first note that

$$\mathbb{P}[|Z_u - Z_v| \ge l_e] = p_e,$$

recalling that $Z_u - Z_v$ is distributed as $\mathcal{N}(0, d(u,v)^2)$. Since $\sum_{e\in\mathcal{C}} p_e \le 1/2$, by the union bound the event $\mathcal{E}$ defined as "$|Z_u - Z_v| \le l_{u,v}$, $\forall\{u,v\} \in E[\mathcal{C}]$", holds with probability at least 1/2. For $x \in X$, let us now define $\mathcal{P}_x$ to be the unique path from the root w to x in $\mathcal{C}$.

Conditioning on the event $\mathcal{E}$, by the triangle inequality

$$|Z_x - Z_w| \le \sum_{\{u,v\}\in\mathcal{P}_x} |Z_u - Z_v| \le \sum_{e\in\mathcal{P}_x} l_e. \tag{1.8}$$

Applying the triangle inequality again, we have that

$$\sup_{x_1,x_2} Z_{x_1} - Z_{x_2} \le 2\sup_{x\in X} |Z_x - Z_w| \le 2\cdot\mathsf{val}(\mathcal{C}).$$

The bound (1.7) now follows since the above occurs with probability at least 1/2.

To work with such chaining trees, it is important to have easy approximations of the edge lengths used above. For $e = \{u,v\} \in E[\mathcal{C}]$ and $p_e \in (0, 1/2]$, it is well known that

$$l_e \asymp d(u,v)\sqrt{\log(1/p_e)}. \tag{1.9}$$

Note that by the standard Gaussian tail bounds (1.4), for any $p \in (0,1)$, we have the upper bound, $l_e \le d(u,v)\sqrt{2\log(2/p_e)}$.

To relate to earlier lower bounds, it is instructive to see that $\mathsf{val}(\mathcal{C}) \gtrsim \mathbf{D}(X)$, for any chaining tree. Firstly, since each $p_{u,v} \in (0, 1/2]$, by (1.9), it follows that $l_{u,v} \gtrsim d(u,v)$. From here, for any pair of points $u, v \in X$, by the triangle inequality $2\cdot\mathsf{val}(\mathcal{C})$ pays for the cost of going from u to the root w and from w to v, yielding the desired upper bound on the diameter.

Chaining beyond Gaussians. Importantly, in the above framework, the only element specific to Gaussian processes is the edge length function (1.5). As our results will apply to this more general setting, we explain how chaining can straightforwardly be adapted to work with processes satisfying appropriate tail bounds.

Let us examine a jointly distributed sequence of random variables $(Z_x)_{x\in X}$ indexed by a metric space (X,d). To constrain the process we will make the following assumptions on the tails. Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous and non-increasing probability density function on the non-negative reals and let $F(s) = \int_s^\infty f(t)dt$ denote the complementary cumulative distribution function. Then, for all $x_1, x_2 \in X$ and $s \ge 0$, we assume that

$$\mathbb{P}[|Z_{x_1} - Z_{x_2}| \ge d(x_1,x_2)\cdot s] \le F(s). \tag{1.10}$$

▶ **Definition 1.2** (Chaining Functional). *We define the* chaining functional *induced by f to be $h(p) := h_f(p) = F^{-1}(p)$, for $p \in (0,1]$, which is well-defined since f is non-increasing. Note that $h(1) = 0$ and that $h(p)$ is strictly decreasing on $(0,1]$. We say that h is of* log-concave type *if the density f is log-concave.*

A property that we make crucial use of is that $h(p)$ is, in fact, a convex function of $p \in (0,1]$. To see this, for $p \in (0,1)$, a direct computation yields that $h'(p) = 1/F'(h(p)) = -1/f(h(p))$, where the derivative is well-defined since f is continuous and non-decreasing. Since $f(s)$ is non-increasing and $h(p)$ is strictly decreasing, $h'(p)$ is non-decreasing and hence, h is convex. Throughout the rest of the paper, we will mainly be interested in chaining functionals of log-concave type.

To apply the chaining framework to the process $(Z_x)_{x\in X}$, we use a chaining tree $\mathcal{C}$ exactly as in Definition 1.1 except that we now compute the edge lengths according to the chaining functional h. Specifically, for $e = \{u,v\} \in E[\mathcal{C}]$ and probability $p_e \in (0,1)$, we define

$$l_e := l_e(e, p_e) := d(u,v) \cdot h(p_e). \tag{1.11}$$

We now define $\mathsf{val}_h(\mathcal{C})$ exactly as in (1.6), using h to compute the edge lengths.

With this setup, with an identical proof to the previous section, we have the inequality

$$\mathbb{P}\left[\sup_{x_1,x_2\in X} Z_{x_1} - Z_{x_2} \geq 2 \cdot \mathsf{val}_h(\mathcal{C})\right] \leq \tfrac{1}{2}.$$

As in the Gaussian setup, it is useful to keep in mind what the "trivial" diameter lower bound on $\mathsf{val}_h(\mathcal{C})$ should be. Since the edge probability $p_e \in (0,1/2]$, for $e = \{u,v\} \in E[\mathcal{C}]$, we have that $l_e \geq d(u,v) \cdot h(1/2)$. Therefore, for any $u,v \in X$, by the triangle inequality, the cost of the paths from u or v to the root w is at least $d(u,v) \cdot h(1/2)$ for any chaining tree $\mathcal{T}$. In particular, for any chaining tree $\mathcal{C}$, we derive the lower bound

$$2 \cdot \mathsf{val}_h(\mathcal{C}) \geq \mathbf{D}(X) \cdot h(1/2). \tag{1.12}$$

It is important to note that the Gaussian chaining setup is indeed a special case of the above. Precisely, in that setup the edge lengths are $l_{u,v} := d(u,v) \cdot h_f(p_{u,v})$, where $f(s) = \sqrt{\frac{2}{\pi}} e^{-s^2/2}$ is the density of the absolute value of the standard Gaussian.

Dudley's Construction. To gain intuition about how to apply the chaining framework, we now explain how to build and analyze the chaining tree used in Dudley's inequality. For simplicity of notation, let us assume that the diameter $\mathbf{D}(X) = 1$. For each $k \geq 0$, let $\mathcal{N}_k$ denote a 2^{-k}-net of X of minimum size, i.e., satisfying $|\mathcal{N}_k| = N_X(2^{-k})$. By our diameter assumption, $\mathcal{N}_0 = \{w\}$ is clearly a single point, which gives the root of the tree $\mathcal{C}$. From here, we construct the tree by induction on $k \geq 1$. At iteration $k \geq 1$, we attach each element of $\mathcal{N}_k$ not already in $\mathcal{C}$ to a closest point in $\mathcal{C}$. From here, we set the edge probability $p_{u,v} = p_k := 2^{-(k+1)}/|\mathcal{N}_k|$ and let $l_{u,v} > 0$ be minimal subject to $\mathbb{P}[|Z_u - Z_v| \geq l_{u,v}] \leq p_k$. This completes the construction.

To analyze the tree $\mathcal{C}$, we make the following observations. Firstly, the number of edges we add to the tree at iteration k is at most $|\mathcal{N}_k|$. Therefore, the total probability sum is at most $\sum_{k=1}^{\infty} |\mathcal{N}_k| \cdot p_k = 1/2$, and hence $\mathcal{C}$ is a valid chaining tree. Second, any edge $\{u,v\}$ added during iteration k satisfies $d(u,v) \leq 2^{-k+1}$. Consequently, by the Gaussian tail bound (1.4),

$$l_{u,v} \lesssim d(u,v)\sqrt{\log(1/p_k)} \lesssim 2^{-k+1}\left(\sqrt{\log N_X(2^{-k})} + \sqrt{k+1}\right).$$

In particular, the value of $\mathcal{C}$ satisfies

$$\mathsf{val}(\mathcal{C}) \lesssim \sum_{k=1}^{\infty} 2^{-k+1} \left(\sqrt{\log N_X(2^{-k})} + \sqrt{k+1} \right) \lesssim 1 + \sum_{k=1}^{\infty} 2^{-k+1} \sqrt{\log(N_X(2^{-k}))}.$$

One can now easily show that the above expression is upper bounded by (1.3) by discretizing the range of the integral along powers of 2 (recalling that $\mathbf{D}(X) = 1$).

The Method of Majorizing Measures. Given the above, it is natural to wonder how one might construct an *optimal* chaining tree for a given process $(Z_x)_{x \in X}$. A principal goal of this paper will be to give efficient constructions for such trees. At first sight, this may seem like a daunting task, as one must somehow simultaneously optimize over all spanning trees and edge probabilities. Nevertheless, a major step towards this goal was achieved for Gaussian processes by Fernique [5], who proved the following remarkable theorem:

$$\mathbb{E}\left[\sup_{x \in X} Z_x\right] \lesssim \boldsymbol{\gamma}_2(X) := \inf_{\mu} \sup_{x \in X} \int_0^{\infty} g(\mu(B(x, \mathsf{r})))d\mathsf{r}. \tag{1.13}$$

Some definitions are in order. Firstly, the infimum over μ is taken over all probability measures on X. Secondly, $g(p) := \sqrt{\log(1/p)}$, for $p \in [0,1]$ corresponds to (an approximation of) the Gaussian edge length function in (1.5). Lastly, $B(x, \mathsf{r}) = \{y \in X : d(x,y) \leq r\}$ is the metric ball of radius r around x, where d is the canonical metric induced by the Gaussian process.

Importantly, the natural analogue of $\boldsymbol{\gamma}_2$ for the general setup in (1.10) also yields upper bounds on the expected supremum, provided the tails of f decay sufficiently quickly. In particular, for $(Z_x)_{x \in X}$ satisfying (1.10), for any "nice enough" f, we have that

$$\mathbb{E}\left[\sup_{x_1, x_2 \in X} Z_{x_1} - Z_{x_2}\right] \lesssim \boldsymbol{\gamma}_h(X) := \inf_{\mu} \sup_{x \in X} \int_0^{\infty} h(\mu(B(x, \mathsf{r})))d\mathsf{r}, \tag{1.14}$$

where h is as in (1.11). Note that since $h(1) = 0$, one can truncate the range of the integral to $r \in [0, \mathbf{D}(X)]$. Very general results of the above type can be found in [18, 1]. We note that the requirements of the process in these works are parametrized is a somewhat different way in terms of Orlicz norms. In this work, we will focus on the setting where the chaining functional h is of log-concave type (where the tail density f is log-concave), where these different parametrizations are equivalent. Prototypical examples in this class are the tail densities of *exponential type*, which are proportional to e^{-x^q}, $x \geq 0$, for $q \geq 1$, and where $h(p) \asymp \ln^{1/q}(1/p)$ for $p \in (0, 1/2)$.

Given that any probability measure μ can be used to upper bound the expected supremum, Fernique dubbed the above technique the method of *majorizing measures*. It is worthwhile to note that Fernique did not prove inequality (1.13) via chaining. He relied instead on a more general technique, which first proves a generic concentration inequality for real valued functions on the metric space, and recovers the desired inequality by averaging over the ensemble of functions induced by the process. The fact that one can recover the same bound via chaining for Gaussian processes would only be proved later, at first implicitly in Talagrand [17], and explicitly in [21], where the latter work also covered processes of exponential type mentioned above.

As noted above, the quantity $\boldsymbol{\gamma}_2(X)$ and more generally $\boldsymbol{\gamma}_h(X)$ (for h of log-concave type), rather miraculously models the value of the best chaining tree as a continuous optimization problem. As majorizing measures may seem like rather opaque objects at first sight, we believe it is instructive to note that from a chaining tree $\mathcal{C}$, one can construct a measure μ whose value in (1.13) is at most $3 \cdot \mathsf{val}_h(\mathcal{C})$. The construction is simple: set $\mu_w = 1/2$ on the root w, and for each $e = \{u, v\} \in E[\mathcal{C}]$ (with v closer to the root than u), set $\mu_u = p_e$. The details of the comparison can be found in the appendix of the full version of the paper.

From the above discussion, we see that the majorizing measures are indeed powerful tools for upper bounding suprema. Given this, together with the many tools for *lower bounding* Gaussian suprema (which are not available in general), Fernique [5] conjectured that majorizing measures should fully characterize the expected supremum of Gaussian processes. This conjecture was verified in the ground-breaking work of Talagrand [17], which is now called the majorizing measures theorem:

▶ **Theorem 1.3** (Fernique-Talagrand [5, 17]). *For any centered Gaussian process $(Z_x)_{x \in X}$ over the metric space $X = (X, d)$, where d is the canonical metric induced by the process, we have*

$$\mathbb{E}\Big[\sup_{x \in X} Z_x\Big] \asymp \gamma_2(X).$$

The original proof of the majorizing measure theorem [17] was considered notoriously difficult. Due to its importance in the theory of stochastic processes, many simpler as well as different proofs were found [19, 20, 21, 12, 2, 22], often by Talagrand himself.

As stated at the beginning of the introduction, the goal of this paper is to give an alternative constructive proof of this theorem using a convex optimization approach. In particular, our starting point is the simple observation that $\gamma_h(X)$ is in fact a convex program, which follows directly from the convexity of h[1]. While simple (and most certainly known to experts), we have not seen this observation leveraged in earlier proofs. In our context, convexity will allow for near-optimal solutions to $\gamma_h(X)$ to be efficiently computed using off-the-shelf methods. Furthermore, convex duality will allow us to inspect the structure of solutions to natural dual program(s) for $\gamma_h(X)$, enabling us to reason about lower bounds. Our proof will operate entirely at the level of the metric space, and will produce a natural combinatorial variant of an optimal primal-dual solution pair for $\gamma_h(X)$, namely a chaining tree and packing tree (defined shortly). These solutions will in fact be obtained by “rounding” solutions to the corresponding continuous programs. Specializing to the Gaussian case, we recover the majorizing measure theorem by an easy comparison between the value of the Gaussian supremum and the value of the combinatorial solutions (which are tailor made for this purpose). This strategy has the benefit of clearly separating the role of the metric space and the role of the Gaussian process, which are often intertwined in difficult to disentangle ways in many proofs.

We now review some of the key ideas in known proofs, which will be important for our approach as well. In particular, we will require appropriate dual analogue to chaining trees.

Primal Proof Strategies. Given what we have seen so far, a main missing ingredient is a stronger form of lower bound for the value of the Gaussian supremum (noting that chaining already provides the upper bound). For this purpose, we examine the natural functional induced by the process on subsets of X, defining

$$G(S) := \mathbb{E}\left[\sup_{x \in S} Z_x\right], \quad \forall S \subseteq X. \tag{1.15}$$

[1] The formulation $\gamma_2(X)$ is “essentially convex”. This is because $g(p)$ is only convex on the interval $[0, 1/\sqrt{e}]$, which is easily remedied. We note this non-convexity is principally due to $g(p)$ being a poor approximation of (1.5) for $p \in [1/\sqrt{e}, 1]$.

The following functional inequality, named the "super-Sudakov" inequality in [22], was proven in [19]: there exists $\gamma \in (0,1)$, such that given an r-separated (non-empty) subsets $A_1, \dots, A_N \subseteq S$, i.e., satisfying $d(A_i, A_j) \geq r$, $\forall i \neq j$, and $\mathbf{D}(A_i) \leq \gamma r$, $\forall i \in [N]$, then

$$G(S) \geq \gamma \cdot r \cdot g(1/N) + \min_{i \in [N]} G(A_i), \tag{1.16}$$

where $g(x) = \sqrt{\log 1/x}$ for $x \in [0,1]$, as before.

In [19], Talagrand gave a construction which takes a functional G on X satisfying (1.16), and produces (a variant of) a chaining tree $\mathcal{C}$ satisfying $\mathsf{val}(\mathcal{C}) \lesssim G(X)$. Talagrand's construction is based on a recursive partitioning scheme, where the partitions roughly correspond to subtrees, which greedily chooses metric balls of large G value to construct the partition. We note that this construction comes in different flavors, each yielding more structured versions of chaining trees (i.e., labelled nets [19] and admissible nets [20]). By instantiating G to be the functional given by (1.15) immediately yields Theorem 1.3. While Talagrand's construction was certainly algorithmic, the Gaussian functional G is not easy to compute (at least deterministically). As mentioned previously, in [21], Talagrand also gave another procedure that directly converts measures to chaining trees. Note that this yields a good chaining tree from a good measure, but by itself does not yield Theorem 1.3.

Dual Proof Strategies. One reason the "difficult" Gaussian functional G was required to prove Theorem 1.3 is that there was no simple dual object to compare to that certifies a lower bound. From the convex optimization perspective, this should morally correspond to a solution to the dual of $\boldsymbol{\gamma}_2(X)$ (or $\boldsymbol{\gamma}_h(X)$). Such an object, called a *packing tree* in the terminology of [22], was in fact developed in Talagrand's original proof [17] for the Gaussian case, and extended to general chaining functionals in [10].

▶ **Definition 1.4** (Packing Tree). *Let $\alpha \in (0, \frac{1}{10}]$. An α-packing tree $\mathcal{T}$ on a finite metric space (X, d) is a rooted tree on subsets of X, with root node $W \subseteq X$, together with a labelling $\chi : \mathcal{T} \to \mathbb{Z}_{\geq 0}$. We enforce that every leaf node $V \in \mathcal{T}$ is a singleton, i.e., $V = \{x\}$ for some $x \in X$. We denote* $\mathrm{leaf}(\mathcal{T}) \subseteq X$ *to be the union of all leaf nodes of $\mathcal{T}$. Every node $V \in \mathcal{T}$ has a (possibly empty) set of children $C_1, \dots, C_k \subseteq V$ which are pairwise disjoint. We let* $\deg_+(V) := k$ *denote the* number of children *of V. We enforce the follow metric properties on $\mathcal{T}$:*

1. *For any child C of $V \in \mathcal{T}$, we have that $\mathbf{D}(C) \leq \alpha^{\chi(V)+1} \cdot \mathbf{D}(X)$.*
2. *For $V \in \mathcal{T}$ and distinct children C_1, C_2 of V, we have $d(C_1, C_2) \geq \frac{1}{10}\alpha^{\chi(V)} \cdot \mathbf{D}(X)$.*

The value of an α-packing tree $\mathcal{T}$ with respect to a chaining functional h is defined as

$$\mathsf{val}_h(\mathcal{T}) := \inf_{x \in \mathrm{leaf}(\mathcal{T})} \sum_{V \in \mathcal{P}_x \setminus \{x\}} \alpha^{\chi(V)} \cdot \mathbf{D}(X) \cdot h(1/\deg_+(V)), \tag{1.17}$$

where $\mathcal{P}_x$ is the unique path from the root W to the leaf $\{x\}$. We use the shorthand $\mathsf{val}_2(\mathcal{T})$ *to denote the value with respect to the Gaussian functional g.*

We remark that we do not count the edge going to the parent in $\deg_+(V)$ mostly for notational convenience – in this case, nodes with a sole child do not contribute to the value of the packing tree. Also, there is quite a bit of flexibility in the parameters of the packing tree, which are chosen above for convenience. Packing trees are objects that allow us to chain lower bounds together in analogy to upper bounds via chaining trees. The combinatorial structure of a packing tree is more constrained than that of a chaining tree however, and their construction (at least more from the perspective of the analysis) is more delicate.

In the Gaussian setting, an α-packing tree $\mathcal{T}$ is perfectly tailored for combining the "super-Sudakov" inequalities given by (1.16). In particular, for $\alpha = 1/(2\gamma)$, a direct proof by induction starting from the leaves of the tree certifies that $G(X) \gtrsim \mathsf{val}_2(\mathcal{T})$ (see Theorem 6.36 in [22]). This was in fact first established in [17] using Slepian's lemma instead of (1.16). Independently of any process however, they also directly serve as combinatorial lower bounds for $\gamma_h(X)$.

▶ **Lemma 1.5.** *Let $\alpha \in (0, \frac{1}{10}]$. For a finite metric space (X, d), an α-packing tree $\mathcal{T}$ on X, and any chaining functional h, we have*

$$\gamma_h(X) \geq \frac{1}{2}(1-\alpha) \cdot \mathsf{val}_h(\mathcal{T}).$$

While known to experts, it is not so easy to find combinatorial proofs of the above inequality, i.e. not related to a process, in the literature (see for example Exercise 6.12 in [22] or Lemma 3.7 in [3]). We include a proof in the appendix of the full version of the paper.

Talagrand's original proof of the majorizing measures theorem worked almost entirely on the dual side. As generalized in [10], the main work in the proof was in fact to construct an α-packing tree $\mathcal{T}$ satisfying $\mathsf{val}_h(\mathcal{T}) \gtrsim \gamma_h(\mathcal{T})$ (for h of log-concave type). As for the primal side, the construction is based on similar greedy ball (sub-)partitioning using an appropriate functional H on X satisfying a so-called "super-chaining" inequality in the terminology of [22]. Specifically, for any set $S \subseteq X$, and a partition $S = \sqcup_{i=1}^{N} P_i$, H satisfies

$$H(S) \leq \max_{i \in [N]} \beta \cdot \mathbf{D}(S) h(1/(i+1)) + H(P_i), \tag{1.18}$$

for some absolute constant $\beta > 0$. Interestingly, the functional H used in [17, 10] was a variant of $\gamma_h(X)$, which is deterministically computable, and not the Gaussian functional in the case $h = g$ (though this works as well [22]). This construction was in fact leveraged in [3] to give a deterministic polynomial time dynamic programming algorithm for computing a nearly optimal packing tree.

1.2 Our Results

1.2.1 A Constructive Min-Max Theorem

The main result of this paper is the following constructive variant of the combinatorial core of the majorizing measure theorem.

▶ **Theorem 1.6.** *Let (X, d) be an n point metric space, h be a chaining functional of log-concave type. Then there is a deterministic algorithm which computes a chaining tree $\mathcal{C}^*$ and an 1/10-packing tree $\mathcal{T}^*$ satisfying*

$$\mathsf{val}_h(\mathcal{C}^*) \asymp \mathsf{val}_h(\mathcal{T}^*),$$

using $\tilde{O}(n^{\omega+1})$ arithmetic operations and evaluations of h and h', where $\omega \leq 2.373$ is the matrix multiplication constant.

We note that the packing tree parameter 1/10 can be made smaller at the cost increasing the hidden constant in the $\asymp$ notation. Recall that for any pair of trees $\mathcal{C}$ and $\mathcal{T}$ as above, we have already seen that

$$\mathsf{val}_h(\mathcal{C}) \gtrsim \gamma_h(X) \gtrsim \mathsf{val}_h(\mathcal{T}), \tag{1.19}$$

so the pair in Theorem 1.6 form a nearly-optimal primal-dual pair. Furthermore, in the Gaussian setting where $h = g$, replacing $\boldsymbol{\gamma}_2(X)$ above by $\mathbb{E}[\sup_x Z_x]$ corresponds to the "easy direction" of the majorizing measures theorem. Plugging in the solutions from Theorem 1.6 immediately yield the hard direction of the theorem. This allows us to view the metric space part of the majorizing measure theorem as an instance of a combinatorial min-max theorem. We remark that such a combinatorial min-max characterization of the Majorizing Measures theorem was already observed by Guédon and Zvavitch [8]. They showed that the value of the optimal packing tree defines a functional that satisfies the super-Sudakov inequality; when combined with Talagrand's framework, this implies the combinatorial min-max theorem described above. This does not directly yield deterministic constructions of nearly-optimal chaining or packing trees, however.

Ding, Lee and Peres [3] essentially used this observation of [8] along with a suitable dynamic program to give an efficient deterministic algorithm to compute nearly optimal packing trees, as mentioned previously. For nearly optimal chaining trees, we make the simple observation (which seems to have gone unnoticed) that these can extracted from Talagrand's [21] "rounding" algorithm applied to a nearly optimal solution for the efficiently solvable convex program $\boldsymbol{\gamma}_h(X)$. By themselves however, these algorithms do not directly say much about how the values of these different trees relate to each other.

In Theorem 1.6, we build further on the convex programming approach. At a high level, we build the primal and dual solutions at the same time and rely on convex programming duality to ensure they have (nearly) the same value. In essence, we replace the "magic functionals" satisfying super-Sudakov or super-chaining inequalities that appear in Talagrand's constructions with convex duality. As we will see in the next section, the dual objects will also correspond to probability measures. In contrast to the primal however, where the rounding to a suitable chaining tree can be done in one shot, the dual measures will require multiple levels of rounding.

The primal and dual solutions we require correspond to nearly optimal primal and dual measures associated with a saddle-point formulation of $\boldsymbol{\gamma}_h(X)$ (see (1.20) in the next subsection). There are in fact many existing solvers that are able to compute nearly optimal solutions to such saddle point problems, where we will rely on a recent fast solver of [9]. This computation in fact forms the bulk of the running time of the algorithm in Theorem 1.6. The details of this part of the algorithm can be found in the full version of this paper. An interesting open problem is whether one can reduce the running time of Theorem 1.6 to $\widetilde{O}(n^2)$, which would be nearly-linear in the input size (recall that an n point metric consists of n^2 distances). The main bottleneck is the use of an all purpose blackbox solver [9] to approximately solve (1.20), and it seems likely that an appropriately tailored first-order method could bring the running time down to $\tilde{O}(n^2)$.

While our main contribution is conceptual, we expect and hope that novel and interesting applications of an "algorithmic" theory of chaining will be found. As a contribution on this front, we give an application of Theorem 1.6 in the context of derandomization: we give a deterministic algorithm for computing Johnson-Lindenstrauss projections achieving the guarantees of Gordon's theorem [7], where we rely on the chaining based proof from [14]. As far as we are aware, no prior deterministic construction was known.

1.2.2 Simplifying the Dual of $\gamma_h(X)$

For simplicity of notation, throughout this section (and most of the paper), we will assume that (X, d) is a fixed n-point metric space and that h is a chaining functional of log-concave type satisfying $|h'(1)| = 1$ (interpreted as the left directional derivative). Under this normalization on h, the trivial diameter lower bound on $\gamma_h(X)$ will be at least $\mathbf{D}(X)/4$, which we will use to convert additive errors to multiplicative ones. This normalization is without loss of generality, and can be achieved by appropriately scaling the h and the metric d so that $\gamma_h(X)$ remains unchanged (see Section 2.1 for a full explanation).

We now describe the dual formulation of $\gamma_h(X)$ and describe the process of simplifying it. For this purpose, we start with the basic saddle-point formulation of $\gamma_h(X)$:

$$\gamma_h(X) = \min_\mu \max_{x \in X} \int_0^\infty h(\mu(B(x,\mathsf{r})))d\mathsf{r} = \min_\mu \max_\nu \int_X \int_0^\infty h(\mu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x) \tag{1.20}$$

where ν also ranges over all probability measures on X (the optimal ν above puts mass 1 on any maximizer of $\int_0^\infty h(\mu(B(x,\mathsf{r}))d\mathsf{r})$.

To obtain the dual program to $\gamma_h(X)$, we interchange μ and ν:

$$\gamma_h(X) \geq \max_\nu \min_\mu \int_X \int_0^\infty h(\mu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x) := \gamma_h^*(X). \tag{1.21}$$

In particular, for any fixed dual measure ν, we have

$$\gamma_h(X) \geq \min_\mu \int_X \int_0^\infty h(\mu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x). \tag{1.22}$$

Since the objective $\int_X \int_0^\infty h(\mu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x)$ is convex in μ and linear in ν, and the probability simplex is compact and convex, by Sion's theorem the value of both convex programs is equal. That is, $\gamma_h(X) = \gamma_h^*(X)$. The measures required within the construction in Theorem 1.6 will be nearly optimal primal and dual measures μ^* and ν^* to $\gamma_h(X)$ and $\gamma_h^*(X)$ respectively.

Unfortunately, it is not clear at this point that the dual is terribly useful. In particular, even evaluating the objective in (1.22) for a given dual measure ν requires solving a non-trivial convex optimization problem (note that the corresponding objective of $\gamma_h(X)$ can be computed by simply evaluating n integrals). Rather surprisingly, it turns out that for h of log-concave type, one can in fact "guess" a near-optimal μ in (1.22), namely, we can set $\mu = \nu$.

▶ **Lemma 1.7.** *For any probability measure ν on X, we have that*

$$\int_X \int_0^\infty h(\nu(B(x,\mathsf{r}))d\mathsf{r}d\nu(x) \leq 2 \min_\mu \int_X \int_0^\infty h(\mu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x) + \mathbf{D}(X)/e,$$

where the minimum is taken over all probability measures μ.

The proof of the above proceeds on a "per scale" basis. More precisely, for a given $\mathsf{r} > 0$, we show that $\int_X h(\nu(B(x,2\mathsf{r}))d\nu(x) \leq \int_X h(\mu(B(x,\mathsf{r})))d\nu(x) + 1/e$. This statement is easily restated in graph-theoretic terms, by defining a graph $G = (X_1 \cup X_2, E)$, with X_1, X_2 both being copies of X and where $x_1 \in X_1$ is adjacent to $x_2 \in X_2$ if $d(x_1, x_2) \leq \mathsf{r}$. Then $\mu(B(x,\mathsf{r}))$ corresponds to the mass under μ within the neighborhood of x, and $\nu(B(x,2\mathsf{r}))$ to the mass under ν within the two-hop neighborhood of x. In this setting, we use a tool from combinatorial optimization, namely, a generalization of Hall's theorem. We note the two properties needed from h above are that h be decreasing and $\max_{a\in(0,1]} |ah'(a)| \leq 1$. The latter property in fact follows from h being of log-concave type and the normalization $|h'(1)| = 1$.

Motivated by the above, we consider the following simplification of $\gamma_h^*(X)$, which we call the *entropic dual*:

$$\delta_h^{\mathbf{Ent}}(X) := \max_\nu \int_X \int_0^\infty h(\nu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x). \tag{1.23}$$

This corresponds to the value ν using the nearly optimal guess for μ in (1.22), which is now readily computable. The following direct corollary relates the value of the entropic dual to the actual dual.

▶ **Corollary 1.8.**

$$\gamma_h^*(X) \leq \boldsymbol{\delta}_h^{\mathbf{Ent}}(X) \leq 2\boldsymbol{\gamma}_h^*(X) + \mathbf{D}(X)/e.$$

In terms of the additive error above, as mentioned at the beginning of the section, $\mathbf{D}(X)/4$ will be the trivial lower bound on $\boldsymbol{\gamma}_h^*(X) = \boldsymbol{\gamma}_h(X)$. Therefore, the right hand in Theorem 1.8 is at most $(2 + 4/e)\boldsymbol{\gamma}_h^*(X)$ in the worst-case. In most interesting cases however, one would expect $\boldsymbol{\gamma}_h(X)$ to be far from the trivial lower bound, in which case one can think of the right hand side as $(2 + o(1))\boldsymbol{\gamma}_h^*(X)$.

We are now ready to give the final simplified form of the dual whose value will most directly relate to the value of packing trees: we define the *simplified dual* by

$$\boldsymbol{\delta}_h(X) := \max_{\nu} \min_{x \in X, \nu(x)>0} \int_0^\infty h(\nu(B(x, \mathsf{r})))d\mathsf{r}. \tag{1.24}$$

Note that we restrict to the minimum of the points supported by ν. These will correspond to the potential leaf nodes in the packing tree. Furthermore, the minimum in (1.24) is in direct analogy to the minimum cost of a path down a packing tree.

Trivially, since we replaced the average by a minimum, we have that $\boldsymbol{\delta}_h^{\mathbf{Ent}}(X) \geq \boldsymbol{\delta}_h(X)$. We show that the reverse direction also holds up to additive error. For a probability measure ν on X and for any subset $S \subseteq X$, satisfying $\nu(S) > 0$, define ν_S by

$$\nu_S(A) := \nu_S(A \cap S)/\nu(S), \forall A \subseteq X, \tag{1.25}$$

i.e., ν_S is the conditional probability measure induced by ν on S. The following lemma shows that one can easily convert a measure ν with large $\boldsymbol{\delta}_h^{\mathbf{Ent}}$ value to one with large $\boldsymbol{\delta}_h$ value via conditioning.

▶ **Lemma 1.9.** *For any probability measure ν on X, there exists $S \subseteq \{x \in X : \nu(x) > 0\}$ such that*

$$\int_X \int_0^\infty h(\nu(B(x, \mathsf{r})))dr \leq \min_{x \in S} \int_0^\infty h(\nu_S(B(x, \mathsf{r})))d\mathsf{r} \ + \ \mathbf{D}(X).$$

Furthermore, S can be computed using at most $O(n^3)$ arithmetic operations and evaluations of h.

The algorithm achieving the above is in fact very simple: we start with $S = \{x \in X : \nu(x) > 0\}$, and iteratively kick out the element $x \in S$ with lowest value as long as the above inequality is not met.

Combining Corollary 1.8 and Lemma 1.9, we obtain the following relations between the dual program $\boldsymbol{\gamma}_h^*(X)$ and the simplified dual $\boldsymbol{\delta}_h(X)$.

▶ **Theorem 1.10.**

$$\gamma_h^*(X) - \mathbf{D}(X) \leq \boldsymbol{\delta}_h(X) \leq 2\boldsymbol{\gamma}_h^*(X) + \mathbf{D}(X)/e.$$

Furthermore, given any probability measure ν on X, one can compute $S \subseteq \{x \in X : \nu(x) > 0\}$ satisfying

$$\int_X \int_0^\infty h(\nu(B(x, \mathsf{r})))d\mathsf{r}d\nu(x) \leq \min_{x \in S} \int_0^\infty h(\nu_S(B(x, \mathsf{r})))d\mathsf{r} \ + \ \mathbf{D}(X),$$

using at most $O(n^3)$ arithmetic operations and evaluations of g.

We are in fact not the first to examine the $\boldsymbol{\delta}_h^{\mathbf{Ent}}(X)$ and $\boldsymbol{\delta}_h(X)$ programs. The analysis of solutions to $\boldsymbol{\delta}_2^{\mathbf{Ent}}(X)$ (i.e., $h = g$) in fact already goes back all the way to Fernique [5]. Starting from a Gaussian process $(Z_x)_{x\in X}$, Fernique examined the measure ν on X satisfying $\nu(u) = \mathbb{P}[Z_u = \max_{x\in X} Z_x]$, $\forall u \in X$, where we assume X is finite and that the maximum is uniquely attained with probability 1. For this "argmax" measure ν, it was shown that

$$\mathbb{E}\left[\sup_{x\in X} Z_x\right] \asymp \int_X \int_0^\infty g(\nu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x),$$

where the inequalities $\lesssim$ and $\gtrsim$ where proven by Fernique [5] and Talagrand [17] respectively.

The relationship between $\boldsymbol{\gamma}_h(X)$ and $\boldsymbol{\delta}_h(X)$ was also already studied by Naor and Mendel [11] as well as Bednorz [2]. In particular, for any continuous h satisfying $\lim_{x\to 0^+} h(x) = \infty$ (i.e., not necessarily of log-concave type), [11, 2] showed that $\boldsymbol{\gamma}_h(X) \le \boldsymbol{\delta}_h(X)$. This was proved using Brouwer's fixed-point theorem, which was used to find a measure μ where the quantities $\int_0^\infty h(\mu(B(x,\mathsf{r})))d\mathsf{r}$ are equal for all $x \in X$. Recalling that $\boldsymbol{\gamma}_h(X) = \boldsymbol{\gamma}_h^*(X)$, this bound is stronger than $\boldsymbol{\gamma}_h^*(X) - \mathbf{D}(X) \le \boldsymbol{\delta}_h(X)$ in Theorem 1.10. However, we do not require $\lim_{x\to 0^+} h(x) = \infty$ (e.g., $h(x) = 1 - x$ is valid for us), which is crucially used to prove the existence of the above Brouwer measure. Mendel and Naor [11] further prove that $\boldsymbol{\delta}_h(X) \lesssim \boldsymbol{\gamma}_h(X)$, for any decreasing h satisfying $h(x^2) \lesssim h(x)$. This is achieved by rounding any dual measure ν to what they call an *ultrametric skeleton*, which one can interpret as a very sophisticated analogue of a packing tree which is agnostic to h. These skeletons are also used to derive optimal bounds for the largest subset of a metric space embeddable into ℓ_2 with small distortion (known as a non-linear Dvoretzky theorem). [11] asked whether one can improve their bound to $\boldsymbol{\delta}_h(X) \le (2 + o(1))\boldsymbol{\gamma}_h(X)$, where the factor of 2 is tight up to $o(1)$ factors for an n-point star-metric with $h = g$. Up to the additive constant (which is often $o(1)$ compared to $\boldsymbol{\gamma}_h(X)$) and the restriction that h be of log-concave type, Theorem 1.10 resolves their question in the affirmative.

1.2.3 Rounding Measures to Trees

Towards proving our main theorem (Theorem 1.6), we show how to round a measure ρ to chaining and packing trees with values approximately $\boldsymbol{\gamma}_h(\rho, X)$ and $\boldsymbol{\delta}_h(\rho, X)$ respectively. As remarked before, the primal rounding strategy was already introduced by Talagrand [21] himself. While the algorithm is simple, it is based on a not terribly intuitive variant of ball partitioning, which is perhaps due to the structure of the object he converts to (an admissible sequence). In the present work, we show that Talagrand's basic greedy ball partitioning scheme with the functional replaced by a measure, very transparently yields a construction of good chaining trees. The greedy ball partition algorithm iteratively selects centers that maximize the measure of balls of a smaller radius and removes a ball of a larger radius centered at the previously chosen points. One does this until the removed pieces form a partition and then proceeds recursively on those pieces. Here we show that this can be implemented in near linear time in the input size which is $O(n^2)$ for an n-point metric space.

▶ **Theorem 1.11.** *There exists a deterministic algorithm that runs in $O(n^2 \log n)$ time and given a probability measure ρ on an n-point metric space X, finds a chaining tree $\mathcal{C}$ such that* $\mathsf{val}_h(\mathcal{C}) \lesssim \boldsymbol{\gamma}_h(\rho, X)$.

On the dual side, as far as we are aware there were no rounding strategies to compute packing trees starting from a measure ρ. The previous approaches for rounding [17, 10, 3] were based on defining a functional that satisfies the previously mentioned "super-chaining" inequality and used a greedy partitioning procedure based on the value of this functional.

This analysis was rather delicate and somewhat mysterious. Moreover, the corresponding functionals in these instances were themselves solutions to optimization problem on the metric space, so implementing this strategy deterministically was rather slow. In fact, Ding, Lee and Peres [3] showed that using a carefully constructed dynamic program, one can implement the above strategy and compute a packing tree in polynomial time in the input size when the metric space is given as the input. Although, they don't specify a precise bound on the running time, one can directly infer a $O(n^4)$ deterministic running time for building a packing tree; with an additional observation, this can in fact be improved to a $O(n^3 \log n)$ bound.[2]

In this work, we revisit the above approach and show that in fact, one can round a probability measure ρ on the metric space to a packing tree with approximately the same value as $\boldsymbol{\delta}_h(\rho, X)$. This has certain advantages over a rounding strategy using functionals as it can be implemented in near linear time in the input size. Moreover, this rounding algorithm is quite similar to the primal rounding algorithm and in our opinion clarifies why such a construction works. In particular, the basic strategy of choosing centers that maximize the measure of a smaller ball remains exactly the same – one just selects smaller balls to add as children and recurses on them instead, followed by some post-processing.

▶ **Theorem 1.12.** *There exists a deterministic algorithm that runs in $O(n^2 \log n)$ time and given a probability measure ρ on an n-point metric space X, finds a $(1/10)$-packing tree $\mathcal{T}$ such that $\boldsymbol{\delta}_h(p, X) \lesssim \mathsf{val}_h(\mathcal{T})$.*

1.2.4 Proof of the Combinatorial Min-Max Theorem (Theorem 1.6)

Of course, without a way to compute measures which have almost optimal values for $\boldsymbol{\gamma}_h(\rho, X)$ and $\boldsymbol{\delta}_h(\rho, X)$, the above rounding algorithms would not have been very useful. Fortunately, we can obtain almost optimal primal and dual measures μ and ν by solving the saddle point formulation (1.20) of $\boldsymbol{\gamma}_h(X)$. Plugging the measure μ in Theorem 1.11 gives us a chaining tree $\mathcal{C}^*$ such that $\mathsf{val}_h(\mathcal{C}^*) \lesssim \boldsymbol{\gamma}(\mu, X) \asymp \boldsymbol{\gamma}_h(X)$.

On the dual side, the measure ν is not enough as it might not be a good solution to the approximate dual $\boldsymbol{\delta}_h(X)$. However, using Theorem 1.10, one can find a set $S \subset X$ such that the probability measure ν_S obtained by conditioning ν on the set S satisfies $\boldsymbol{\delta}_h(\nu_S, X) \asymp \boldsymbol{\gamma}_h(X)$. Plugging the measure ν_S in Theorem 1.12, gives us a packing tree $\mathcal{T}^*$ satisfying $\boldsymbol{\delta}_h(\nu_S, X) \lesssim \mathsf{val}_h(\mathcal{T}^*)$. Combining the two yields that

$$\mathsf{val}_h(\mathcal{C}^*) \lesssim \boldsymbol{\gamma}_h(X) \asymp \boldsymbol{\delta}_h(\nu_S, X) \lesssim \mathsf{val}_h(\mathcal{T}^*). \tag{1.26}$$

Recall that the weak duality relation (1.19) between chaining and packing trees implies the reverse inequality for $\mathsf{val}_h(\mathcal{T}) \lesssim \mathsf{val}_h(\mathcal{C})$ for any chaining tree $\mathcal{C}$ and any packing tree $\mathcal{T}$. Thus, (1.26) gives us the combinatorial min-max statement given by Theorem 1.6. Moreover, this can be made algorithmic by solving the saddle point formulation and using the algorithm to find the set S given by Theorem 1.10. The details for solving the saddle point formulation are given in the appendix of the full version of this paper and the algorithm to find the set S is presented in the proof of Theorem 1.10.

[2] Their running time is $O(n^2)$ times the number of "distance scales" in the metric, meaning the number of dyadic intervals $[2^k, 2^{k+1})$ containing at least one distance $d(u,v)$ in the metric. To bound the number of distance scales by $O(n \log n)$, compute a minimum spanning tree T in the complete graph describing the metric. Consider any edge $e = \{v, w\}$ not in the tree, and let $m(v,w)$ denote an edge on the path between v and w in T of maximum length. Then e has length at least the length of $m(v,w)$ since T is an MST, but not more than n times larger. That is, the distance scale of e is in a range of size $O(\log n)$ of the distance scale of $m(v,w)$. Now consider assigning each non-tree edge $\{v,w\}$ to $m(v,w)$; there are $O(\log n)$ distance scales assigned to each tree edge, so $O(n \log n)$ in total.

1.3 Organization

In Section 2, we list some basic notation as well as the main useful properties of chaining functionals of log-concave type. In Section 3, we simplify the dual program $\boldsymbol{\gamma}_h^*(X)$, proving the main inequalities relating it to the entropic and simplified duals $\boldsymbol{\delta}_h^{\mathbf{Ent}}(X)$ and $\boldsymbol{\delta}_h(X)$ respectively. Other details including near-linear time rounding algorithms from measures to chaining trees and packing trees, deterministic construction of Johnson-Lindenstrauss projections achieving Gordon's bound, using a black-box solver to compute nearly-optimal primal and dual measures for the saddle-point formulation of $\boldsymbol{\gamma}_h(X)$, and proofs of other technical statements can be found in the full version of this paper.

2 Preliminaries

Notation. Throughout this paper, log denotes the natural logarithm unless the base is explicitly mentioned. We use $[k]$ to denote the set $\{1, 2, \ldots, k\}$. For a vector $z \in \mathbb{R}^n$, we will use z_i or $z(i)$ interchangeably to denote the i-th coordinate of z. Given a probability measure μ over a set X, we use $\mathbb{E}_{x\sim\mu}[f(x)]$ to denote the expectation of $f(x)$ where x is sampled from μ.

2.1 Properties of Chaining Functionals

Recall that we work with a chaining functional $h : (0,1] \to \mathbb{R}_+$ defined by $h(p) = F^{-1}(p)$, where $F(s) = \int_s^\infty f(x)dx$ and where f is non-decreasing and continuous probability density on $\mathbb{R}_+$. We assume throughout the rest of the paper that h is of log-concave type, i.e., that f is log-concave.

The fundamental property of such functionals, that we make extensive use of, is the following:

▶ **Proposition 2.1.** *For a chaining functional h of log-concave type,*

$$h(ab) \leq h(a) + h(b) \quad \textit{for} \quad a, b \in (0,1].$$

Before giving a proof, we note that the above property appears in [10] as the base assumption for the chaining functionals they consider. The above shows that this condition is very natural and applies to a wide variety of functionals.

Proof. Let us define $\varphi : [0,\infty) \to [0,\infty)$ as $\varphi(t) := h(e^{-t})$. We will show that φ is concave on its domain, which implies that h is sub-additive as

$$h(ab) = \varphi\left(\log\frac{1}{ab}\right) \leq \frac{1}{2}\varphi\left(2\log\frac{1}{a}\right) + \frac{1}{2}\varphi\left(2\log\frac{1}{b}\right) \leq \varphi\left(\log\frac{1}{a}\right) + \varphi\left(\log\frac{1}{b}\right) = h(a) + h(b),$$

where both inequalities follow from concavity and $\varphi(0) = 0$.

To see that φ is concave, we show that $\varphi'(t) = -e^{-t}h'(e^{-t}) = \frac{e^{-t}}{f(F^{-1}(e^{-t}))}$ is a decreasing function of t. Substituting $x = F^{-1}(e^{-t})$, and noting that f is decreasing, it suffices to show that $\frac{F(x)}{f(x)}$ is decreasing for x on the positive real line. Taking arbitrary $0 \leq x_1 \leq x_2$, we have that

$$\frac{F(x_1)}{f(x_1)} = \int_0^\infty \frac{f(x_1+t)}{f(x_1)}dt \geq \int_0^\infty \frac{f(x_2+t)}{f(x_2)}dt = \frac{F(x_2)}{f(x_2)},$$

where the inequality follows from the following elementary property of non-negative log-concave functions: for any four points $a \leq b \leq c \leq d$, we have that $f(b)f(c) \geq f(a)f(d)$ which in the above scenario implies that $\frac{f(x_1+t)}{f(x_1)} \geq \frac{f(x_2+t)}{f(x_2)}$. This completes the proof of the proposition. ◀

When working on a metric space (X,d) with a chaining functional h, we observe the following rescaling symmetry: for any constant $\beta > 0$, if we replace h by βh (equivalently, the density function f is replaced by $z \to \beta f(\beta z)$) and $d(u,v)$ by $d(u,v)/\beta$ for all $u,v \in X$, the values of the various quantities we consider remain unaffected. In particular, if we have an underlying process $(Z_x)_{x\in X}$ amenable to X and f as in (1.10), the condition $\mathbb{P}[|Z_{x_1} - Z_{x_2}| \geq d(x_1,x_2)s] \leq F(s)$ is identical for both scalings, and the values of the various programs and trees we consider remain unchanged.

So from now on, we make the convenient choice that $f(0) = 1$. This implies that the (left) derivative $h'(1) = -1/f(0) = -1$. We maintain this normalization for the remainder of the paper.

Given this normalization, the following useful bound is easy to show.

▶ **Proposition 2.2.** *For every $a \in (0,1]$, we have that*

$-1 \leq ah'(a) \leq 0$ *and* $h(a) \leq \log(1/a)$.

Proof. As h is decreasing, $h'(a) \leq 0$ for $a \in (0,1]$. Therefore, for any a,

$$\begin{aligned} ah'(a) = \lim_{\epsilon\to 0^+} \frac{a(h(a) - h(a(1-\epsilon)))}{a - a(1-\epsilon)} &\geq \lim_{\epsilon\to 0^+} \frac{h(a) - h(a) - h(1-\epsilon)}{1-(1-\epsilon)} \\ &= \lim_{\epsilon\to 0^+} \frac{h(1) - h(1-\epsilon)}{1-(1-\epsilon)} = h'(1) = -1, \end{aligned} \tag{2.1}$$

where the first inequality follows from sub-multiplicativity and the last equality holds since $h(1) = 0$. The first statement in the proposition follows.

To see the second statement, one can observe that as h is decreasing, (2.1) implies the following differential inequality : $h'(a) \geq -\frac{1}{a}$ for every $a \in (0,1]$. Together with the boundary condition that $h(1) = 0$, this implies that $h(a) \leq \log(1/a)$. ◀

3 Dual simplifications

In this section, we provide the main arguments that allow for rounding a solution to $\gamma_h^*(X)$ to a solution to the simplified dual, via the entropic dual, as described in Section 1.2.2.

Proof of Lemma 1.7. Our goal is to prove that taking $\mu = \nu$ in (1.22) does not cause too much error. We will prove this on a "per scale" basis:

▶ **Lemma 3.1.** *For any probability measures μ and ν on X, and any $\mathsf{r} > 0$, we have*

$$\int_X h(\nu(B(x,2\mathsf{r}))d\nu(x) \leq \int_X h(\mu(B(x,\mathsf{r})))d\nu(x) + 1/e.$$

Lemma 1.7 follows easily from this:

$$\begin{aligned} \int_X \int_0^\infty h(\nu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x) &\leq \min_\mu \int_X \int_0^{\mathbf{D}(X)} \Big(h(\mu(B(x,\mathsf{r}/2))) + 1/e\Big) d\mathsf{r}d\nu(x) \\ &\leq 2\min_\mu \int_X \int_0^\infty h(\mu(B(x,\mathsf{r})))d\mathsf{r}d\nu(x) + \mathbf{D}(X)/e. \end{aligned}$$

Proof of Lemma 3.1. We first recast the statement in graph theoretic terms. Let us define a bipartite graph on the vertex set $X_1 \cup X_2$ where each X_i for $i \in [2]$ is a copy of the index set X. We add an edge between two vertices $x_1 \in X_1, x_2 \in X_2$ if $d(x_1,x_2) \leq \mathsf{r}$ – note that every vertex has an edge incident to it. Define the weight of a vertex $x \in X_i$ as $\mu(x)$ and the weight of a vertex $x \in X_2$ as $\nu(x)$. Let E denote the set of edges in the graph, let $N(S)$ be the neighbors of a subset of the vertices $S \subset X_1 \cup X_2$ and let $N^2(S) = N(N(S))$. For brevity, we will write $N(x)$ instead of $N(\{x\})$ for singleton sets.

With this setup, the statement we want to prove is

$$\sum_{x\in X_2} \nu(x)h(\nu(N^2(x))) \le \sum_{x\in X_2} \nu(x)h(\mu(N(x)) + 1/e. \tag{3.1}$$

To prove the above, we will use the following structural result which is a consequence of the theory of principal sequences of matroids [6], applied to transversal matroids; we include a self-contained proof in the full version of the paper.

▶ **Proposition 3.2.** *There exist sequences $\emptyset = S_0 \subset S_1 \subset \cdots \subset S_k = X_1$ and $0 < \beta_1 < \beta_2 < \cdots < \beta_k$, such that*

1. *$\beta_i\mu(S_i \setminus S_{i-1}) = \nu(N(S_i) \setminus N(S_{i-1}))$ for all $i \in [k]$.*
2. *For all $i \in [k]$ and $A \subseteq X_1 \setminus S_{i-1}$, we have that $\beta_i\mu(A) \le \nu(N(A) \setminus N(S_{i-1}))$.*

The proposition above can be viewed as a kind of strengthening of Hall's theorem. For instance, if there is a fractional perfect matching between μ and ν, i.e., a way of transporting mass distributed according to μ on X_1 along edges of the graph to yield precisely the distribution ν on X_2, then the claim will be satisfied with $k = 1$, $S_1 = X_1$ and $\beta_1 = 1$, for in this case, $\mu(A) \le \nu(N(A))$ for any $A \subseteq X_1$ (essentially the easy direction of Hall's theorem). If there is no fractional perfect matching, Hall's theorem implies the existence of a set $S \subset X_1$ with $\nu(N(S)) < \mu(S)$. In the proposition, S_1 is the "least matchable" set: only a β_1 fraction of the mass in S_1 can be transported to $N(S_1)$. The full sequence is then obtained by removing S_1 and $N(S_1)$ and repeating on the remainder.

Taking the sequences guaranteed by the proposition, define $\tilde{\beta}_i := \min\{\beta_i, 1\}$.

We split the left hand side of (3.1) as follows:

$$\sum_{x\in X_2} \nu(x)h(\nu(N^2(x))) = \sum_{i=1}^{k} \sum_{x\in N(S_i)\setminus N(S_{i-1})} \nu(x)h(\nu(N^2(x))).$$

Note that for any $i \in [k]$, if $x \in X_2 \setminus N(S_{i-1})$, then $N(x) \subseteq X_1 \setminus S_{i-1}$, and hence, Proposition 3.2 implies that $\tilde{\beta}_i\mu(N(x)) \le \nu(N^2(x))$. Furthermore, using sub-multiplicativity of h and Proposition 2.2, we find that for any $i \in [k]$ and $x \in N(S_i) \setminus N(S_{i-1})$,

$$\begin{aligned} h(\nu(N^2(x))) = h\left(\mu(N(x)) \cdot \frac{\nu(N^2(x)}{\mu(N(x))}\right) &\le h(\mu(N(x))) + h\left(\min\left\{1, \frac{\nu(N^2(x))}{\mu(N(x))}\right\}\right) \\ &\le h(\mu(N(x)) + h(\tilde{\beta}_i) \le h(\mu(N(x)) + \log\left(\frac{1}{\tilde{\beta}_i}\right). \end{aligned}$$

For the first inequality above, sub-multiplicativity is used when $\frac{\nu(N^2(x))}{\mu(N(x))} > 1$; otherwise, the inequality follows because h is decreasing and $h(1) = 0$. (The first inequality does still hold if $\mu(N(x)) = 0$, taking the minimum in the second term to have value 1 in this case, again because h is decreasing.)

The second inequality again uses that h is decreasing, as well as that $\tilde{\beta}_i \le 1$ for every $i \in [k]$; and the final inequality uses Proposition 2.2.

Let ℓ be maximal such that $\tilde{\beta}_\ell < 1$; so $\beta_i = \tilde{\beta}_i$ for $i \leq \ell$. Summing the last inequality over all $i \in [k]$ and $x \in N(S_i) \setminus N(S_{i-1})$, we obtain

$$\begin{aligned}\sum_{x\in X_2} \nu(x)h(\nu(N^2(x))) &\leq \sum_{x\in X_2} \nu(x)h(\mu(N(x))) + \sum_{i=1}^{k} \nu(N(S_i) \setminus N(S_{i-1})) \cdot \log\left(\frac{1}{\tilde{\beta}_i}\right) \\ &= \sum_{x\in X_2} \nu(x)h(\mu(N(x))) + \sum_{i=1}^{\ell} \tilde{\beta}_i \mu(S_i \setminus S_{i-1}) \cdot \log\left(\frac{1}{\tilde{\beta}_i}\right) \\ &\leq \sum_{x\in X_2} \nu(x)h(\mu(N(x))) + \left(\max_{\beta\in(0,1]} \beta \log\left(\frac{1}{\beta}\right)\right) \sum_{i=1}^{\ell} \mu(S_i \setminus S_{i-1}),\end{aligned}$$

where the second line follows from Proposition 3.2. From this (3.1) readily follows as $\sum_{i=1}^{\ell} \mu(S_i \setminus S_{i-1}) \leq \sum_{i=1}^{k} \mu(S_i \setminus S_{i-1}) = \mu(X_1) = 1$ and $\max_{\beta\in(0,1]} \beta \log\left(\frac{1}{\beta}\right) = 1/e$. ◀

Proof of Lemma 1.9. For convenience, define

$$H(\mu, t) := \int_0^\infty h(\mu(B(t, \mathsf{r})))d\mathsf{r} \qquad \text{and} \qquad H(\mu, \nu) := \int_X H(\mu, t)d\nu(t).$$

Start by setting $S = \{x \in X : \nu(x) > 0\}$. Consider the following greedy algorithm:

> As long as $H(\nu_S, \nu_S) > \min_{x\in S} H(\nu_S, x) + \mathbf{D}(X)$, choose $s \in S$ so that $H(\nu_S, s)$ is minimized, and remove s from S.

Note that $H(\nu_S, \nu_S) \leq \min_{x\in S} H(\nu_S, x) + \mathbf{D}(X)$ when this terminates, since it is vacuous for $S = \emptyset$.

We will now show that $H(\nu_S, \nu_S)$ can only increase during the progression of the algorithm. This suffices to prove the lemma, since then upon termination

$$H(\nu, \nu) \leq H(\nu_S, \nu_S) \leq \min_{x\in S} H(\nu_S, x) + \mathbf{D}(X).$$

So, consider a moment in the algorithm where $s \in S$ is about to be removed from S, yielding $S' := S \setminus \{s\}$. From our choice of s,

$$H(\nu_S, \nu_S) > H(\nu_S, s) + \mathbf{D}(X). \tag{3.2}$$

Let $\alpha = 1/(1 - \nu_S(s))$; so $\nu_{S'}(t) = \alpha\nu_S(t)$ for $t \neq s$, and $\nu_{S'}(s) = 0$. Thus

▷ Claim 3.3. For every $t \in X$,

$$H(\nu_{S'}, t) \geq H(\nu_S, t) - \mathbf{D}(X) \cdot (\alpha - 1).$$

Proof. Recall that h is convex and satisfies $-1 \leq ah'(a) \leq 0$ for any $a \in (0, 1]$ by Proposition 2.2. Thus for any $z \in (0, 1]$,

$$\begin{aligned}h(z) &\geq h(z/\alpha) + z(1 - 1/\alpha)h'(z/\alpha) \\ &= h(z/\alpha) - (\alpha - 1)\big|(z/\alpha)h'(z/\alpha)\big| \geq h(z/\alpha) - (\alpha - 1).\end{aligned} \tag{3.3}$$

Now notice that for any $T \subseteq S$, we have that $\nu_S(T) \geq \nu_{S'}(T)/\alpha$. Therefore, since h is decreasing,

$$\begin{aligned} H(\nu_S, t) &= \int_0^{\mathbf{D}(X)} h(\nu_S(B(t, \mathsf{r})))d\mathsf{r} \\ &\leq \int_0^{\mathbf{D}(X)} h(\nu_{S'}(B(t, \mathsf{r}))/\alpha))d\mathsf{r} \\ &\leq \int_0^{\mathbf{D}(X)} h(\nu_{S'}(B(t, \mathsf{r})))d\mathsf{r} + \mathbf{D}(X) \cdot (\alpha - 1) \;=\; H(\nu_{S'}, t) + \mathbf{D}(X) \cdot (\alpha - 1), \end{aligned}$$

where the last inequality follows from (3.3). ◁

We can now compare $H(\nu_{S'}, \nu_{S'})$ with $H(\nu_S, \nu_S)$:

$$\begin{aligned} H(\nu_{S'}, \nu_{S'}) &\geq H(\nu_S, \nu_{S'}) - \mathbf{D}(X) \cdot (\alpha - 1) && \text{(by Claim 3.3)} \\ &= \tfrac{1}{1-\nu_S(s)} \left(H(\nu_S, \nu_S) - \nu_S(s) H(\nu_S, s)\right) - \tfrac{\nu_S(s)}{1-\nu_S(s)} \cdot \mathbf{D}(X) \\ &\geq \tfrac{1}{1-\nu_S(s)} \Big(H(\nu_S, \nu_S) - \nu_S(s)(H(\nu_S, \nu_S) - \mathbf{D}(X))\Big) - \tfrac{\nu_S(s)}{1-\nu_S(s)} \cdot \mathbf{D}(X) && \text{(by (3.2))} \\ &= H(\nu_S, \nu_S, \end{aligned}$$

which finishes the proof of Lemma 1.9.

Runtime Analysis. It is easily seen that the algorithm can be implemented in $O(n^3)$ time. First, by pre-processing the input, we may assume that the pairs of points are sorted by their pairwise distances – this only adds an additional overhead of $O(n^2 \log n)$. Then, as in each iteration we remove one element, there are $O(n)$ iterations to compute the final set S'. Furthermore, in each of these iterations, we are required to compute a minimizer s of $H(\nu_S, x)$. This takes $O(n^2)$ time since for each $x \in S$, one can compute the measure of all possible balls $B(x, \mathsf{r})$ in $O(n)$ time using a straightforward dynamic program.

References

1 Witold Bednorz. A theorem on majorizing measures. *The Annals of Probability*, 34(5):1771–1781, 2006.

2 Witold Bednorz. The majorizing measure approach to sample boundedness. *Colloquium Mathematicum*, 139, November 2012.

3 Jian Ding, James R Lee, and Yuval Peres. Cover times, blanket times, and majorizing measures. *Annals of mathematics*, 175(3):1409–1471, 2012.

4 Richard M Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *Journal of Functional Analysis*, 1(3):290–330, 1967.

5 Xavier Fernique. Regularité des trajectoires des fonctions aléatoires gaussiennes. In *Ecole d'Eté de Probabilités de Saint-Flour IV–1974*, pages 1–96. Springer, 1975.

6 Satoru Fujishige. *Theory of Principal Partitions Revisited*, pages 127–162. Springer Berlin Heidelberg, 2009.

7 Yehoram Gordon. On Milman's inequality and random subspaces which escape through a mesh in Rn. In *Geometric aspects of functional analysis*, pages 84–106. Springer, 1988.

8 Olivier Guédon and Artem Zvavitch. *Supremum of a Process in Terms of Trees*, pages 136–147. Springer Berlin Heidelberg, 2003.

9 Haotian Jiang, Yin Tat Lee, Zhao Song, and Sam Chiu-wai Wong. An improved cutting plane method for convex optimization, convex-concave games, and its applications. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020*, pages 944–953. ACM, 2020.

10 Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*, volume 23. Springer Science & Business Media, 1991.

11 Manor Mendel and Assaf Naor. Ultrametric subsets with large Hausdorff dimension. *Inventiones mathematicae*, 192, June 2011.

12 Manor Mendel and Assaf Naor. Ultrametric skeletons. *Proceedings of the National Academy of Sciences*, 110(48):19256–19262, 2013.

13 Vitali D Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Prilozen*, 5:28–37, 1971.

14 Samet Oymak, Benjamin Recht, and Mahdi Soltanolkotabi. Isometric sketching of any set via the restricted isometry property. *Information and Inference: A Journal of the IMA*, 7(4):707–726, 2018.

15 David Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, 41(2):463–501, 1962.

16 Vladimir Nikolaevich Sudakov. Gaussian random processes and measures of solid angles in Hilbert space. In *Doklady Akademii Nauk*, volume 197, pages 43–45. Russian Academy of Sciences, 1971.

17 Michel Talagrand. Regularity of Gaussian processes. *Acta mathematica*, 159:99–149, 1987.

18 Michel Talagrand. Sample boundedness of stochastic processes under increment conditions. *The Annals of Probability*, pages 1–49, 1990.

19 Michel Talagrand. A simple proof of the majorizing measure theorem. *Geometric & Functional Analysis GAFA*, 2(1):118–125, 1992.

20 Michel Talagrand. Majorizing measures: the generic chaining. *Ann. Probab.*, 24(3):1049–1103, July 1996.

21 Michel Talagrand. Majorizing measures without measures. *Ann. Probab.*, 29(1):411–417, February 2001.

22 Ramon van Handel. Probability in High Dimensions. Lecture Notes. Princeton University, 2016. URL: `https://web.math.princeton.edu/~rvan/APC550.pdf`.