

RADEMACHER TYPE AND ENFLO TYPE COINCIDE

PAATA IVANISVILI, RAMON VAN HANDEL, AND ALEXANDER VOLBERG

ABSTRACT. A nonlinear analogue of the Rademacher type of a Banach space was introduced in classical work of Enflo. The key feature of Enflo type is that its definition uses only the metric structure of the Banach space, while the definition of Rademacher type relies on its linear structure. We prove that Rademacher type and Enflo type coincide, settling a long-standing open problem in Banach space theory. The proof is based on a novel dimension-free analogue of Pisier's inequality on the discrete cube.

1. INTRODUCTION AND MAIN RESULTS

Let $(X, \|\cdot\|)$ be a Banach space. We say that X has *Rademacher type* $p \in [1, 2]$ if there exists $C \in (0, \infty)$ so that for all $n \geq 1$ and $x_1, \dots, x_n \in X$

$$\mathbf{E} \left\| \sum_{j=1}^n \varepsilon_j x_j \right\|^p \leq C^p \sum_{j=1}^n \|x_j\|^p.$$

We denote by $T_p^R(X)$ the smallest possible constant C in this inequality.

A nonlinear notion of type was introduced by Enflo [4]: a Banach space has *Enflo type* p if there exists $C \in (0, \infty)$ so that for all $n \geq 1$ and $f : \{-1, 1\}^n \rightarrow X$

$$\mathbf{E} \left\| \frac{f(\varepsilon) - f(-\varepsilon)}{2} \right\|^p \leq C^p \sum_{j=1}^n \mathbf{E} \|D_j f(\varepsilon)\|^p,$$

and we denote by $T_p^E(X)$ the smallest possible constant C in this inequality. Here we define the discrete partial derivatives on the cube $\{-1, 1\}^n$ as

$$D_j f(\varepsilon) := \frac{f(\varepsilon_1, \dots, \varepsilon_j, \dots, \varepsilon_n) - f(\varepsilon_1, \dots, -\varepsilon_j, \dots, \varepsilon_n)}{2}.$$

The key feature of Enflo type is that its definition depends only on the metric structure of X , that is, it involves only distances between two points. This notion therefore extends naturally to the setting of general metric spaces. In contrast, the definition of Rademacher type relies on the linear structure of X .

The study of metric properties of Banach spaces, known as the “Ribe program”, has been of central importance in Banach space theory in recent decades [12]. Understanding the relationship between Rademacher type and Enflo type is a fundamental question in this program. That Enflo type p implies Rademacher type p follows immediately by choosing the linear function $f(\varepsilon) = \sum_{j=1}^n \varepsilon_j x_j$ in the definition of Enflo type. Whether the converse is also true, that is, that Rademacher type p implies Enflo type p , is a long-standing problem that dates back to Enflo's

2010 *Mathematics Subject Classification.* 46B09; 46B07; 60E15.

Key words and phrases. Rademacher type; Enflo type; Pisier's inequality; Banach spaces.

original paper [4] from 1978. Despite a number of partial results in this direction [2, 3, 1, 14, 13, 11, 7, 6], the question has remained open.

Here we settle Enflo's question in the affirmative: Rademacher type p is equivalent to Enflo type p . In other words, Enflo type provides a characterization of Rademacher type using only the metric structure of X .

Theorem 1.1. *We have*

$$T_p^R(X) \leq T_p^E(X) \leq \frac{\pi}{\sqrt{2}} T_p^R(X)$$

for every $p \in [1, 2]$ and Banach space X .

The key new ingredient in the proof of Theorem 1.1 is a novel dimension-free analogue of a classical inequality of Pisier.

1.1. Pisier's inequality. Let $p \geq 1$, let $f : \{-1, 1\}^n \rightarrow X$ and let ε, δ be independent random vectors that are uniformly distributed on the discrete cube $\{-1, 1\}^n$. As part of his investigation of metric type, Pisier discovered the following class of Sobolev-type inequalities for vector-valued functions on the discrete cube:

$$\mathbf{E} \|f(\varepsilon) - \mathbf{E} f(\varepsilon)\|^p \leq C^p \mathbf{E} \left\| \sum_{j=1}^n \delta_j D_j f(\varepsilon) \right\|^p. \quad (1.1)$$

If such an inequality were to hold with a constant C that is independent of dimension n , then Enflo's problem would be solved: if X has Rademacher type p , then applying this property to the right-hand side of (1.1) conditionally on ε would yield immediately the definition of Enflo type p . Unfortunately, Pisier was able to prove (1.1) only with a dimension-dependent constant $C \sim \log n$ [14, Lemma 7.3], and it was subsequently shown by Talagrand [15, section 6] that this order of growth is optimal: that is, there exist Banach spaces X for which the optimal constant in Pisier's inequality must grow logarithmically with dimension.

In order to resolve Enflo's problem, however, it is not necessary to establish Pisier's inequality for an *arbitrary* Banach space: it suffices to show that (1.1) holds with a dimension-free constant under the additional assumption that X has nontrivial type. For this reason, subsequent work has focused on identifying conditions on the Banach space X under which (1.1) holds with a constant that depends only on the geometry of X (but not on n). Notably, Naor and Schechtman [13] proved that (1.1) holds with a dimension-free constant under the stronger assumption that X is an UMD Banach space (see also [7, 5]). Very recently, Eskenazis and Naor [6] proved that for superreflexive Banach spaces X , the constant in Pisier's inequality can be improved to $\log^\alpha n$ for some $\alpha < 1$.

Beside the inequality (1.1), Pisier also proved [14, Theorem 2.2] a more general counterpart of his inequality in Gauss space: if $f : \mathbb{R}^n \rightarrow X$ is locally Lipschitz, G, G' are independent standard Gaussian vectors in $\mathbb{R}^n$, and $\Phi : X \rightarrow \mathbb{R}$ is convex and satisfies a mild regularity assumption, then

$$\mathbf{E} [\Phi(f(G) - \mathbf{E} f(G))] \leq \mathbf{E} \left[\Phi \left(\frac{\pi}{2} \sum_{j=1}^n G'_j \frac{\partial f}{\partial x_j}(G) \right) \right]. \quad (1.2)$$

One obtains an inequality analogous to (1.1) by choosing $\Phi(x) = \|x\|^p$. Remarkably, the Gaussian inequality is dimension-free for an arbitrary Banach space X , in sharp contrast to the inequality on the cube. Unfortunately, its proof is very special to the

Gaussian case: one defines $G(\theta) := G \sin \theta + G' \cos \theta$, and notes that $(G(\theta), \frac{d}{d\theta}G(\theta))$ has the same distribution as (G, G') for each θ by rotation-invariance of the Gaussian measure. Then (1.2) follows by expressing $f(G) - f(G') = \int_0^{\pi/2} \frac{d}{d\theta} f(G(\theta)) d\theta$ and applying Jensen's inequality. If one attempts to repeat this idea on the discrete cube, the absence of rotational symmetry makes the argument inherently inefficient, and one cannot do better than (1.1) with constant $C \sim \log n$.

Despite the apparent obstructions, we will prove in this paper a completely general dimension-free analogue of (1.2) on the discrete cube. The existence of such an inequality appears at first sight to be quite unexpected. It will turn out, however, that the dimension-dependence of (1.1) is not an intrinsic feature of the discrete cube, but is simply a reflection of the fact that (1.1) is not the “correct” analogue of the corresponding Gaussian inequality. To obtain a dimension-free inequality, we will replace δ by a vector of *biased* Rademacher variables $\delta(t)$ which arises naturally in our proof by differentiating the discrete heat kernel.

1.2. A dimension-free Pisier inequality. The following random variables will appear frequently in the sequel, so we fix them once and for all. Let ε be a random vector that is uniformly distributed on the cube $\{-1, 1\}^n$. Given $t > 0$, we let $\xi(t)$ be a random vector in the cube, independent of ε , whose coordinates $\xi_i(t)$ are independent and identically distributed with

$$\mathbf{P}\{\xi_i(t) = 1\} = \frac{1 + e^{-t}}{2}, \quad \mathbf{P}\{\xi_i(t) = -1\} = \frac{1 - e^{-t}}{2}.$$

We also define the standardized vector $\delta(t)$ by

$$\delta_i(t) := \frac{\xi_i(t) - \mathbf{E}\xi_i(t)}{\sqrt{\text{Var } \xi_i(t)}} = \frac{\xi_i(t) - e^{-t}}{\sqrt{1 - e^{-2t}}}.$$

The following analogue of (1.2) lies at the heart of this paper.

Theorem 1.2. *For any linear space X , function $f : \{-1, 1\}^n \rightarrow X$, and convex function $\Phi : X \rightarrow \mathbb{R}$, we have*

$$\mathbf{E}[\Phi(f(\varepsilon) - \mathbf{E}f(\varepsilon))] \leq \int \mathbf{E} \left[\Phi \left(\frac{\pi}{2} \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right) \right] \mu(dt), \quad (1.3)$$

where μ is the probability measure on $\mathbb{R}_+$ with density $\mu(dt) := \frac{2}{\pi} \frac{1}{\sqrt{e^{2t} - 1}} dt$.

Even though (1.3) is formulated in terms of the biased variables $\delta_j(t)$ as opposed to the Rademacher variables δ_j that appear in (1.1), the proof of Theorem 1.1 will follow readily by a routine symmetrization argument. For this purpose the precise distribution of the random variables $\delta_i(t)$ is in fact immaterial: it suffices that they are independent, centered, and have bounded variance. However, other applications (such as Theorem 1.5 below) do require more precise information on the distribution of $\delta_i(t)$, which can be read off from its definition.

Remark 1.3. It is interesting to note that (1.3) is not just an analogue of (1.2) on the cube: it is in fact a strictly stronger result, as the Gaussian inequality can be derived from Theorem 1.2 by the central limit theorem. To see why, assume $f : \mathbb{R}^n \rightarrow X$ is a sufficiently smooth function with compact support and let $\Phi : X \rightarrow \mathbb{R}$

be a sufficiently regular convex function. Define $f_N : \{-1, 1\}^{n \times N} \rightarrow X$ by

$$f_N(\varepsilon) := f\left(\frac{\sum_{j=1}^N \varepsilon_{1j}}{\sqrt{N}}, \dots, \frac{\sum_{j=1}^N \varepsilon_{nj}}{\sqrt{N}}\right),$$

and note that for $1 \leq i \leq n$ and $1 \leq j \leq N$

$$D_{ij}f_N(\varepsilon) = \frac{\varepsilon_{ij}}{\sqrt{N}} \frac{\partial f}{\partial x_i} \left(\frac{\sum_{j=1}^N \varepsilon_{1j}}{\sqrt{N}}, \dots, \frac{\sum_{j=1}^N \varepsilon_{nj}}{\sqrt{N}} \right) + o\left(\frac{1}{\sqrt{N}}\right)$$

as $N \rightarrow \infty$. Thus by Theorem 1.2

$$\begin{aligned} & \mathbf{E}[\Phi(f_N(\varepsilon) - \mathbf{E}f_N(\varepsilon))] \\ & \leq \int \mathbf{E} \left[\Phi \left(\frac{\pi}{2} \sum_{i=1}^n \frac{\sum_{j=1}^N \delta_{ij}(t) \varepsilon_{ij}}{\sqrt{N}} \frac{\partial f}{\partial x_i} \left(\frac{\sum_{j=1}^N \varepsilon_{1j}}{\sqrt{N}}, \dots, \frac{\sum_{j=1}^N \varepsilon_{nj}}{\sqrt{N}} \right) \right) \right] \mu(dt) + o(1). \end{aligned}$$

Letting $N \rightarrow \infty$ now yields (1.2) by the multivariate central limit theorem, as $\{(\varepsilon_{1j}, \delta_{1j}(t)\varepsilon_{1j}, \dots, \varepsilon_{nj}, \delta_{nj}(t)\varepsilon_{nj})\}_{j \leq N}$ are i.i.d. random vectors with unit covariance matrix. The requisite regularity assumptions on f and Φ can subsequently be removed by routine approximation arguments.

The above discussion also shows that the constant in Theorem 1.2 is optimal. Indeed, as (1.3) implies (1.2), it suffices to show that (1.2) is sharp. But this is already known to be the case when $X = \mathbb{R}$ and $\Phi(x) = |x|$ [10, Chapter 8].

When $\Phi(x) = \|x\|^p$, the conclusion of Theorem 1.2 may be slightly improved. As the improvement will be needed in the sequel, we spell out this variant separately.

Theorem 1.4. *Let μ be as in Theorem 1.2. Then for any Banach space $(X, \|\cdot\|)$, function $f : \{-1, 1\}^n \rightarrow X$, and $1 \leq p < \infty$, we have*

$$(\mathbf{E}\|f(\varepsilon) - \mathbf{E}f(\varepsilon)\|^p)^{1/p} \leq \frac{\pi}{2} \int \left(\mathbf{E} \left\| \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right\|^p \right)^{1/p} \mu(dt). \quad (1.4)$$

In this setting, the difference between (1.3) and (1.4) is that in the former the exponent $1/p$ appears outside the $\mu(dt)$ integral on the right-hand side.

We now briefly describe the idea behind the proofs of Theorems 1.2 and 1.4, which was inspired by Gaussian semigroup methods of Ledoux [10, Chapter 8]. Instead of using rotational invariance as in the proof of (1.2) to interpolate between $f(G)$ and $f(G')$, we use the heat semigroup on the discrete cube to interpolate between f and $\mathbf{E}f$. The resulting expressions involve derivatives of the form $D_j e^{t\Delta} f$. We now observe that rather than applying the derivative to f , we may differentiate the heat kernel instead. A short computation (Lemma 2.1) shows that the gradient of the heat kernel on the cube yields the biased Rademacher vector $\delta(t)$. This elementary observation, analogous to the classical smoothing property of diffusion semigroups, leads us to discover (1.3) in a completely natural manner.

1.3. Pisier's inequality and cotype. Theorems 1.2 and 1.4 provide dimension-free analogues of Pisier's inequality on the cube for an *arbitrary* Banach space X . With these results in hand, however, we can now revisit the question of what additional assumption must be imposed on X in order that Pisier's original inequality (1.1) holds with a dimension-independent constant.

Recall that a Banach space $(X, \|\cdot\|)$ has (Rademacher) *cotype* $q \in [2, \infty)$ if there exists $C \in (0, \infty)$ so that for all $n \geq 1$ and $x_1, \dots, x_n \in X$

$$\sum_{j=1}^n \|x_j\|^q \leq C^q \mathbf{E} \left\| \sum_{j=1}^n \varepsilon_j x_j \right\|^q.$$

We denote by $C_q(X)$ the smallest possible constant C in this inequality. The significance of cotype in the present context is twofold:

- If X has finite cotype, one can estimate biased Rademacher averages by regular Rademacher averages [14, Proposition 3.2], so that Theorem 1.4 yields (1.1).
- If X does not have finite cotype, it contains ℓ_∞^n uniformly [14, Theorem 3.3], which enables us to embed Talagrand's example [15, section 6] for every n .

Both theorems applied here are classical results of Maurey and Pisier. These observations give rise to the following characterization.

Theorem 1.5. *For any Banach space X and $1 \leq p < \infty$, Pisier's inequality (1.1) holds with a constant independent of dimension n if and only if X has finite cotype.*

As any Banach space with nontrivial type has finite cotype [9, Theorem 7.1.14], we obtain in particular an affirmative answer to the question posed after (1.1): Pisier's inequality holds with a dimension-free constant in any Banach space with nontrivial type. However, one may argue that this fact is no longer of great importance in view of our main results; in practice Theorems 1.2 and 1.4 may be just as easily deployed directly in applications (as we do in Theorem 1.1), and give rise to much better constants than would be obtained from Theorem 1.5.

A quantitative formulation of Theorem 1.5 will be given in section 4.

1.4. Organization of this paper. The rest of this paper is organized as follows. Section 2 is devoted to the proofs of Theorems 1.2 and 1.4. We subsequently deduce Theorem 1.1 in section 3. Finally, Theorem 1.5 is proved in section 4.

2. PROOF OF THEOREMS 1.2 AND 1.4

The Laplacian on the discrete cube is defined by

$$\Delta f := - \sum_{j=1}^n D_j f.$$

We denote by P_t the standard heat semigroup on the cube, that is,

$$P_t := e^{t\Delta}.$$

Recall that Δ is self-adjoint on $L^2(\{-1, 1\}^n)$ with quadratic form

$$-\mathbf{E}[f(\varepsilon) \Delta g(\varepsilon)] = \sum_{j=1}^n \mathbf{E}[D_j f(\varepsilon) D_j g(\varepsilon)].$$

The basis for the proof of Theorem 1.2 is the following probabilistic representation of the heat semigroup and its discrete partial derivatives.

Lemma 2.1. *We have*

$$P_t f(x) = \mathbf{E}[f(x_1 \xi_1(t), \dots, x_n \xi_n(t))] \quad \text{for } t \geq 0,$$

and

$$D_j P_t f(x) = \frac{1}{\sqrt{e^{2t}-1}} \mathbf{E}[\delta_j(t) f(x_1 \xi_1(t), \dots, x_n \xi_n(t))] \quad \text{for } t > 0.$$

Proof. Let $Q_t f(x) := \mathbf{E}[f(x_1 \xi_1(t), \dots, x_n \xi_n(t))]$. By the definition of $\xi_j(t)$, we have

$$Q_t f(x) = \sum_{\xi \in \{-1,1\}^n} \left[\prod_{i=1}^n \frac{1+e^{-t}\xi_i}{2} \right] f(x_1 \xi_1, \dots, x_n \xi_n).$$

Note also that

$$Q_t f(x_1, \dots, -x_j, \dots, x_n) = \sum_{\xi \in \{-1,1\}^n} \left[\prod_{i=1}^n \frac{1+e^{-t}\xi_i}{2} \right] \frac{1-e^{-t}\xi_j}{1+e^{-t}\xi_j} f(x_1 \xi_1, \dots, x_n \xi_n).$$

We now observe that

$$\frac{1}{2} \left(1 - \frac{1-e^{-t}\xi_j}{1+e^{-t}\xi_j} \right) = \frac{e^{-t}\xi_j}{1+e^{-t}\xi_j} = \frac{e^{-t}}{1-e^{-2t}} (\xi_j - e^{-t}).$$

We have therefore shown that

$$D_j Q_t f(x) = \frac{1}{\sqrt{e^{2t}-1}} \mathbf{E}[\delta_j(t) f(x_1 \xi_1(t), \dots, x_n \xi_n(t))].$$

It remains to show that $Q_t f = P_t f$. To this end, note that $Q_0 f = f$ and

$$\begin{aligned} \frac{d}{dt} Q_t f(x) &= - \sum_{j=1}^n \sum_{\xi \in \{-1,1\}^n} \left[\prod_{i=1}^n \frac{1+e^{-t}\xi_i}{2} \right] \frac{e^{-t}\xi_j}{1+e^{-t}\xi_j} f(x_1 \xi_1, \dots, x_n \xi_n) \\ &= - \sum_{j=1}^n D_j Q_t f(x) = \Delta Q_t f(x). \end{aligned}$$

Thus Q_t satisfies the Kolmogorov equation for the semigroup P_t . $\square$

We are now ready to prove Theorem 1.2.

Proof of Theorem 1.2. We may assume without loss of generality that X is finite-dimensional (as $f(\{-1,1\}^n)$ spans a space of dimension at most 2^n). Write

$$\Phi(x) = \sup_{z \in X^*} \{ \langle z, x \rangle - \Phi^*(z) \}$$

where $\Phi^* : X^* \rightarrow (-\infty, \infty]$ is the convex conjugate of Φ . Then

$$\mathbf{E}[\Phi(f(\varepsilon) - \mathbf{E}f(\varepsilon))] = \sup_{g: \{-1,1\}^n \rightarrow X^*} \{ \mathbf{E}[\langle g(\varepsilon), f(\varepsilon) - \mathbf{E}f(\varepsilon) \rangle] - \mathbf{E}[\Phi^*(g(\varepsilon))] \}.$$

As $P_0 f = f$ and $\lim_{t \rightarrow \infty} P_t f = \mathbf{E}f(\varepsilon)$ (this follows, e.g., from Lemma 2.1), we can write by the fundamental theorem of calculus

$$\begin{aligned} \mathbf{E}[\langle g(\varepsilon), f(\varepsilon) - \mathbf{E}f(\varepsilon) \rangle] &= - \int_0^\infty \mathbf{E} \left[\left\langle g(\varepsilon), \frac{d}{dt} P_t f(\varepsilon) \right\rangle \right] dt \\ &= - \int_0^\infty \mathbf{E}[\langle g(\varepsilon), \Delta P_t f(\varepsilon) \rangle] dt \\ &= \int_0^\infty \sum_{j=1}^n \mathbf{E}[\langle D_j P_t g(\varepsilon), D_j f(\varepsilon) \rangle] dt, \end{aligned}$$

where we used in the last line that Δ is self-adjoint and commutes with P_t . To proceed, we note that by Lemma 2.1

$$\sum_{j=1}^n \mathbf{E}[\langle D_j P_t g(\varepsilon), D_j f(\varepsilon) \rangle] = \frac{1}{\sqrt{e^{2t} - 1}} \mathbf{E} \left[\left\langle g(\varepsilon \xi(t)), \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right\rangle \right],$$

where $\varepsilon \xi(t) := (\varepsilon_1 \xi_1(t), \dots, \varepsilon_n \xi_n(t))$. Moreover, $\mathbf{E}[\Phi^*(g(\varepsilon))] = \mathbf{E}[\Phi^*(g(\varepsilon \xi(t)))]$, as the random vectors $\varepsilon \xi(t)$ and ε have the same distribution. Thus

$$\begin{aligned} & \mathbf{E}[\langle g(\varepsilon), f(\varepsilon) - \mathbf{E}f(\varepsilon) \rangle] - \mathbf{E}[\Phi^*(g(\varepsilon))] \\ &= \int \mathbf{E} \left[\left\langle g(\varepsilon \xi(t)), \frac{\pi}{2} \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right\rangle - \Phi^*(g(\varepsilon \xi(t))) \right] \mu(dt) \\ &\leq \int \mathbf{E} \left[\Phi \left(\frac{\pi}{2} \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right) \right] \mu(dt), \end{aligned}$$

and the conclusion follows. $\square$

The proof of Theorem 1.4 is almost identical.

Proof of Theorem 1.4. In this case we use [8, Proposition 1.3.1]

$$(\mathbf{E}\|f(\varepsilon) - \mathbf{E}f(\varepsilon)\|^p)^{1/p} = \sup_{\mathbf{E}\|g(\varepsilon)\|^q \leq 1} \mathbf{E}[\langle g(\varepsilon), f(\varepsilon) - \mathbf{E}f(\varepsilon) \rangle]$$

with $\frac{1}{p} + \frac{1}{q} = 1$. Proceeding exactly as in the proof of Theorem 1.2, we obtain

$$\begin{aligned} \mathbf{E}[\langle g(\varepsilon), f(\varepsilon) - \mathbf{E}f(\varepsilon) \rangle] &= \int \mathbf{E} \left[\left\langle g(\varepsilon \xi(t)), \frac{\pi}{2} \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right\rangle \right] \mu(dt) \\ &\leq \int (\mathbf{E}\|g(\varepsilon \xi(t))\|^q)^{1/q} \left(\mathbf{E} \left\| \frac{\pi}{2} \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right\|^p \right)^{1/p} \mu(dt) \end{aligned} \quad (2.1)$$

using Hölder's inequality. Recalling that $\mathbf{E}\|g(\varepsilon \xi(t))\|^q = \mathbf{E}\|g(\varepsilon)\|^q$ as the random vectors $\varepsilon \xi(t)$ and ε have the same distribution, the conclusion follows readily. $\square$

Remark 2.2 (Alternative approach to the proofs of Theorems 1.2 and 1.4). Using that ε and $\varepsilon \xi(t)$ have the same distribution and that (2.1) holds for all g , it is readily seen that (2.1) implies the *pointwise* identity

$$f(x) - \mathbf{E}f(\varepsilon) = \int \mathbf{E} \left[\frac{\pi}{2} \sum_{j=1}^n \delta_j(t) D_j f(x \xi(t)) \right] \mu(dt) \quad (2.2)$$

for $x \in \{-1, 1\}^n$. By using this identity one can organize the proofs in a manner that is closer to the proof of (1.2). For example, to prove Theorem 1.2 we can upper bound $\Phi(f(x) - \mathbf{E}f(\varepsilon))$ pointwise by applying Jensen's inequality to the right-hand side of (2.2), and then (1.3) follows by taking the expectation of the resulting expression and using that ε and $\varepsilon \xi(t)$ have the same distribution.

The pointwise identity (2.2) can also be proved directly, which leads to proofs of Theorems 1.2 and 1.4 that avoid the use of duality. The following argument was

communicated to us by Jingbo Liu. First, note two basic properties of the discrete cube: $D_j^2 = D_j$ and $D_j P_t = P_t D_j$ for every j . Thus we can write

$$\begin{aligned} f(x) - \mathbf{E}f(\varepsilon) &= - \int_0^\infty \frac{d}{dt} P_t f(x) dt = - \int_0^\infty \Delta P_t f(x) dt \\ &= \int_0^\infty \sum_{j=1}^n D_j^2 P_t f(x) dt = \int_0^\infty \sum_{j=1}^n D_j P_t D_j f(x) dt. \\ &= \int_0^\infty \frac{1}{\sqrt{e^{2t} - 1}} \mathbf{E} \left[\sum_{j=1}^n \delta_j(t) D_j f(x\xi(t)) \right] dt, \end{aligned}$$

using Lemma 2.1 in the last step. While conceptually appealing, the disadvantage of this argument is that it relies on special properties of calculus on the discrete cube. In contrast, the proofs that are based on duality use nothing else than the quadratic form $-\langle f, \Delta g \rangle_{\ell^2(\{-1,1\})} = \langle Df, Df \rangle_{\ell^2(\{-1,1\})}$ and the gradient formula of Lemma 2.1, providing a more direct route to extensions beyond the discrete cube.

3. PROOF OF THEOREM 1.1

Theorem 1.1 follows from Theorem 1.2 by a routine symmetrization argument.

Proof of Theorem 1.1. The first inequality $T_p^R(X) \leq T_p^E(X)$ follows readily by choosing $f(\varepsilon) = \sum_{j=1}^n \varepsilon_j x_j$ in the definition of Enflo type.

In the converse direction, note first that as ε and $-\varepsilon$ have the same distribution, and as $x \mapsto \|x\|^p$ is convex, we can estimate

$$\mathbf{E} \left\| \frac{f(\varepsilon) - f(-\varepsilon)}{2} \right\|^p = \mathbf{E} \left\| \frac{f(\varepsilon) - \mathbf{E}f(\varepsilon) - f(-\varepsilon) + \mathbf{E}f(-\varepsilon)}{2} \right\|^p \leq \mathbf{E} \|f(\varepsilon) - \mathbf{E}f(\varepsilon)\|^p.$$

Applying Theorem 1.2 with $\Phi(x) = \|x\|^p$ yields

$$\mathbf{E} \left\| \frac{f(\varepsilon) - f(-\varepsilon)}{2} \right\|^p \leq \int \mathbf{E} \left\| \frac{\pi}{2} \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right\|^p \mu(dt).$$

To estimate the right-hand side we use a standard symmetrization argument. Let $\xi'(t)$ be an independent copy of $\xi(t)$ and ε' be an independent copy of ε . Then

$$\begin{aligned} \mathbf{E} \left\| \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right\|^p &\leq \mathbf{E} \left\| \sum_{j=1}^n \frac{\xi_j(t) - \xi'_j(t)}{\sqrt{\text{Var } \xi_j(t)}} D_j f(\varepsilon) \right\|^p \\ &= \mathbf{E} \left\| \sum_{j=1}^n \varepsilon'_j \frac{\xi_j(t) - \xi'_j(t)}{\sqrt{\text{Var } \xi_i(t)}} D_j f(\varepsilon) \right\|^p \\ &\leq T_p^R(X)^p \sum_{j=1}^n \mathbf{E} \left| \frac{\xi_j(t) - \xi'_j(t)}{\sqrt{\text{Var } \xi_i(t)}} \right|^p \mathbf{E} \|D_j f(\varepsilon)\|^p, \end{aligned}$$

where we used Jensen's inequality in the first line; that $\xi(t) - \xi'(t)$ has the same distribution as $\varepsilon'(\xi(t) - \xi'(t))$ (by symmetry and independence) in the second line; and the definition of Rademacher type conditionally on $\xi(t), \xi'(t), \varepsilon$ and that $\xi(t), \xi'(t), \varepsilon, \varepsilon'$ are independent in the third line. But as $p \leq 2$, we obtain

$$\mathbf{E} \left| \frac{\xi_j(t) - \xi'_j(t)}{\sqrt{\text{Var } \xi_i(t)}} \right|^p \leq \left(\frac{\mathbf{E}[(\xi_j(t) - \xi'_j(t))^2]}{\text{Var } \xi_i(t)} \right)^{p/2} = 2^{p/2}$$

by Jensen's inequality. Thus we have shown

$$\mathbf{E} \left\| \frac{f(\varepsilon) - f(-\varepsilon)}{2} \right\|^p \leq \left(\frac{\pi}{\sqrt{2}} T_p^R(X) \right)^p \sum_{j=1}^n \mathbf{E} \|D_j f(\varepsilon)\|^p,$$

which implies $T_p^E(X) \leq \frac{\pi}{\sqrt{2}} T_p^R(X)$. $\square$

4. PROOF OF THEOREM 1.5

The following contraction principle is a classical result of Maurey and Pisier (see, e.g., [14, Proposition 3.2]). We spell out a version with explicit constants.

Theorem 4.1. *Let $(X, \|\cdot\|)$ be a Banach space of cotype $q < \infty$, let $\eta_1, \dots, \eta_n$ be i.i.d. symmetric random variables, and let ε be uniformly distributed on $\{-1, 1\}^n$. Then for any $n \geq 1$, $x_1, \dots, x_n \in X$, and $1 \leq p < \infty$, we have*

$$\left(\mathbf{E} \left\| \sum_{j=1}^n \eta_j x_j \right\|^p \right)^{1/p} \leq L_{q,p} \int_0^\infty \mathbf{P}\{|\eta_1| > t\}^{\frac{1}{\max(q,p)}} dt \left(\mathbf{E} \left\| \sum_{j=1}^n \varepsilon_j x_j \right\|^p \right)^{1/p}$$

with $L_{q,p} = L C_q(X) \max(1, (q/p)^{1/2})$, where L is a universal constant.

Proof. As η_i are symmetric random variables, they have the same distribution as $\varepsilon_i \eta_i$. The conclusion for the special case $p = q$ follows from [9, Theorem 7.2.6]. For the general case, we consider two distinct cases.

For the case $p > q$, recall that a Banach space with cotype q also has cotype p for all $r > q$, with $C_r(X) \leq C_q(X)$ [9, p. 55]. Thus the conclusion follows readily from [9, Theorem 7.2.6] by choosing $q = p$.

For the case $p < q$, we bound the L^p -norm on the left-hand side by the L^q -norm, and then apply the inequality for the case $p = q$. This yields

$$\left(\mathbf{E} \left\| \sum_{j=1}^n \eta_j x_j \right\|^p \right)^{1/p} \leq L C_q(X) \int_0^\infty \mathbf{P}\{|\eta_1| > t\}^{\frac{1}{q}} dt \left(\mathbf{E} \left\| \sum_{j=1}^n \varepsilon_j x_j \right\|^q \right)^{1/q}.$$

We conclude by using the Kahane-Khintchine inequality [9, Theorem 6.2.4] to bound the L^q -norm on the right-hand side by the L^p -norm, which incurs the additional factor $\lesssim (q/p)^{1/2}$. This completes the proof. $\square$

We are now ready to prove one direction of Theorem 1.5: if X has finite cotype, then (1.1) holds with a dimension-free constant.

Proposition 4.2. *Let $(X, \|\cdot\|)$ be a Banach space of cotype q , and let ε, δ be independent uniformly distributed random vectors in $\{-1, 1\}^n$. Then for any function $f : \{-1, 1\}^n \rightarrow X$ and $1 \leq p < \infty$, we have*

$$\mathbf{E} \|f(\varepsilon) - \mathbf{E} f(\varepsilon)\|^p \leq K_{q,p}^p \mathbf{E} \left\| \sum_{j=1}^n \delta_j D_j f(\varepsilon) \right\|^p$$

with $K_{q,p} = K C_q(X) p \max(1, (q/p)^{3/2})$, where K is a universal constant.

Proof. Let $\xi'(t)$ be an independent copy of $\xi(t)$. We first note that

$$\int_0^\infty \mathbf{P}\{|\xi_j(t) - \xi'_j(t)| > s\}^{1/r} ds = 2^{1-1/r} (1 - e^{-2t})^{1/r}.$$

Thus

$$\begin{aligned} \left(\mathbf{E} \left\| \sum_{j=1}^n \delta_j(t) D_j f(\varepsilon) \right\|^p \right)^{1/p} &\leq \frac{1}{\sqrt{1-e^{-2t}}} \left(\mathbf{E} \left\| \sum_{j=1}^n (\xi_j(t) - \xi'_j(t)) D_j f(\varepsilon) \right\|^p \right)^{1/p} \\ &\leq 2L_{q,p} (1-e^{-2t})^{\frac{1}{\max(q,p)} - \frac{1}{2}} \left(\mathbf{E} \left\| \sum_{j=1}^n \delta_j D_j f(\varepsilon) \right\|^p \right)^{1/p}, \end{aligned}$$

where we used Jensen's inequality in the first line and we applied Theorem 4.1 conditionally on ε in the second line. Now note that

$$\frac{\pi}{2} \int (1-e^{-2t})^{\frac{1}{\max(q,p)} - \frac{1}{2}} \mu(dt) \leq \int_0^\infty e^{-t} (1-e^{-t})^{\frac{1}{\max(q,p)} - 1} dt = \max(q, p).$$

Thus Theorem 1.4 yields

$$(\mathbf{E} \|f(\varepsilon) - \mathbf{E} f(\varepsilon)\|^p)^{1/p} \leq 2L_{q,p} \max(q, p) \left(\mathbf{E} \left\| \sum_{j=1}^n \delta_j D_j f(\varepsilon) \right\|^p \right)^{1/p},$$

and the conclusion follows by the definition of $L_{q,p}$ in Theorem 4.1. $\square$

Remark 4.3. It should be noted that the improvement provided by Theorem 1.4 is used crucially in the proof of Theorem 1.5. Had we used Theorem 1.2 instead, we would have encountered the integral $\int_0^\infty (1-e^{-2t})^{\frac{p}{\max(q,p)} - \frac{p}{2}} \mu(dt)$ in the proof; it is readily verified that this integral diverges at some finite value of p .

It remains to show the converse direction: if X does not have finite cotype, then the constant in (1.1) is at least of order $\log n$.

Proposition 4.4. *If the Banach space X does not have finite cotype, then for every $n \geq 1$ and $1 \leq p < \infty$, there exists a function $f : \{-1, 1\}^n \rightarrow X$ so that*

$$\mathbf{E} \|f(\varepsilon) - \mathbf{E} f(\varepsilon)\|^p \geq C_{n,p}^p \mathbf{E} \left\| \sum_{j=1}^n \delta_j D_j f(\varepsilon) \right\|^p$$

with $C_{n,p} = C \log(n/9p)$, where C is a universal constant.

Proof. It was shown by Talagrand [15, section 6] that for every $n \geq 1$ and $1 \leq p < \infty$, there is a function $f : \{-1, 1\}^n \rightarrow \ell_\infty^{2^n}$ so that

$$(\mathbf{E} \|f(\varepsilon) - \mathbf{E} f(\varepsilon)\|_\infty^p)^{1/p} \geq c \log(n/9p) \left(\mathbf{E} \left\| \sum_{j=1}^n \delta_j D_j f(\varepsilon) \right\|_\infty^p \right)^{1/p}$$

for a universal constant c . But if X does not have finite cotype, then by the Maurey-Pisier theorem [9, Theorem 7.3.8] it must contain a 2-isomorphic copy of ℓ_∞^N for every $N \geq 1$. Thus we can embed Talagrand's example in X for every $n \geq 1$ and $1 \leq p < \infty$, and the proof is readily concluded. $\square$

Remark 4.5. We emphasize that our characterization of when Pisier's inequality holds with dimension-free constant assumes the Banach space X and $1 \leq p < \infty$ are fixed. When this is not the case, other phenomena can arise. For example, it follows from a result of Wagner [16] that if one chooses $p \asymp n$, then (1.1) holds with a universal constant for *any* Banach space X . This is a purely combinatorial fact that does not capture any structure of the underlying space.

Acknowledgments. The authors thank Sergey Bobkov, Alexandros Eskenazis, Dong Li, Jingbo Liu, and Gilles Pisier for helpful discussions and comments. We are particularly grateful to Assaf Naor and to the anonymous referee for suggestions that considerably improved the presentation of the paper.

P.I. was supported in part by NSF grants DMS-1856486 and CAREER-DMS-1945102. R.v.H. was supported in part by NSF grants CAREER-DMS-1148711 and DMS-1811735, by the ARO through PECASE award W911NF-14-1-0094, and by the Simons Collaboration on Algorithms & Geometry. A.V. was supported in part by NSF grants DMS-160065 and DMS-1900268.

REFERENCES

- [1] J. Bourgain, V. Milman, and H. Wolfson. On type of metric spaces. *Trans. Amer. Math. Soc.*, 294(1):295–317, 1986.
- [2] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p -spaces. *Ark. Mat.*, 8:103–105, 1969.
- [3] P. Enflo. Uniform structures and square roots in topological groups. II. *Israel J. Math.*, 8:230–252; *ibid.* 8 (1970), 253–272, 1970.
- [4] P. Enflo. On infinite-dimensional topological groups. In *Séminaire sur la Géométrie des Espaces de Banach (1977–1978)*, pages Exp. No. 10–11, 11. École Polytech., Palaiseau, 1978.
- [5] A. Eskenazis. On Pisier’s inequality for UMD targets, 2020. Preprint arXiv:2002.10396.
- [6] A. Eskenazis and A. Naor. Discrete Littlewood-Paley-Stein theory and Pisier’s inequality for superreflexive targets, 2019. Preprint.
- [7] T. Hytönen and A. Naor. Pisier’s inequality revisited. *Studia Math.*, 215(3):221–235, 2013.
- [8] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach spaces. Vol. I. Martingales and Littlewood-Paley theory*, volume 63. Springer, Cham, 2016.
- [9] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach spaces. Vol. II*, volume 67. Springer, Cham, 2017. Probabilistic methods and operator theory.
- [10] 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, Berlin, 1996.
- [11] M. Mendel and A. Naor. Scaled Enflo type is equivalent to Rademacher type. *Bull. Lond. Math. Soc.*, 39(3):493–498, 2007.
- [12] A. Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.
- [13] A. Naor and G. Schechtman. Remarks on non linear type and Pisier’s inequality. *J. Reine Angew. Math.*, 552:213–236, 2002.
- [14] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.
- [15] M. Talagrand. Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis’ graph connectivity theorem. *Geom. Funct. Anal.*, 3(3):295–314, 1993.
- [16] R. Wagner. Notes on an inequality by Pisier for functions on the discrete cube. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 263–268. Springer, Berlin, 2000.

(P.I.) DEPARTMENT OF MATHEMATICS, NORTH CAROLINA STATE UNIVERSITY, RALEIGH, NC 27695, USA; AND DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, IRVINE, CA 92617, USA

E-mail address: pivanisv@uci.edu

(R.v.H.) FINE HALL 207, PRINCETON UNIVERSITY, PRINCETON, NJ 08544, USA

E-mail address: rvan@princeton.edu

(A.V.) DEPARTMENT OF MATHEMATICS, MICHIGAN STATE UNIVERSITY, EAST LANSING, MI 48823, USA

E-mail address: volberg@math.msu.edu