HEAT FLOW AND QUANTITATIVE DIFFERENTIATION

TUOMAS HYTÖNEN AND ASSAF NAOR

Abstract. For every Banach space $(Y, \|\cdot\|_Y)$ that admits an equivalent uniformly convex norm we prove that there exists $c = c(Y) \in (0,\infty)$ with the following property. Suppose that $n \in \mathbb{N}$ and that X is an n-dimensional normed space with unit ball B_X. Then for every 1-Lipschitz function $f : B_X \to Y$ and for every $\varepsilon \in (0, 1/2]$ there exists a radius $r \geqslant \exp(-1/\varepsilon^{cn})$, a point $x \in B_X$ with $x + rB_X \subseteq B_X$, and an affine mapping $\Lambda : X \to Y$ such that $\|f(y) - \Lambda(y)\|_Y \leqslant \varepsilon r$ for every $y \in x+rB_X$. This is an improved bound for a fundamental quantitative differentiation problem that was formulated by Bates, Johnson, Lindenstrauss, Preiss and Schechtman (1999), and consequently it yields a new proof of Bourgain's discretization theorem (1987) for uniformly convex targets. The strategy of our proof is inspired by Bourgain's original approach to the discretization problem, which takes the affine mapping Λ to be the first order Taylor polynomial of a time-t Poisson evolute of an extension of f to all of X and argues that, under appropriate assumptions on f, there must exist a time $t \in (0,\infty)$ at which Λ is (quantitatively) invertible. However, in the present context we desire a more stringent conclusion, namely that Λ well-approximates f on a macroscopically large ball, in which case we show that for our argument to work one cannot use the Poisson semigroup. Nevertheless, our strategy does succeed with the Poisson semigroup replaced by the heat semigroup. As a crucial step of our proof, we establish a new uniformly convex-valued Littlewood–Paley–Stein $\mathcal{G}$-function inequality for the heat semigroup; influential work of Martínez, Torrea and Xu (2006) obtained such an inequality for subordinated Poisson semigroups but left the important case of the heat semigroup open. As a byproduct, our proof also yields a new and simple approach to the classical Dorronsoro theorem (1985) even for real-valued functions.

Contents

Date: May 12, 2017.
Key words and phrases. Quantitative differentiation, uniform convexity, Littlewood–Paley theory, heat semigroup.
T.H. was supported by the ERC Starting Grant "AnProb" and the Academy of Finland, CoE in Analysis and Dynamics Research. A. N. was supported by BSF grant 2010021, the Packard Foundation and the Simons Foundation.

1. Introduction

Denote the unit ball of a Banach space $(Y, \|\cdot\|_Y)$ by $B_Y \stackrel{\mathrm{def}}{=} \{y \in Y : \|y\|_Y \leqslant 1\}$. Recall that the norm $\|\cdot\|_Y$ is said to be uniformly convex if for every $\varepsilon \in (0,2]$ there exists $\delta \in (0,1]$ such that

$$\forall\, x, y \in B_Y, \qquad \|x-y\|_Y \geqslant \varepsilon \implies \|x+y\|_Y \leqslant 2(1-\delta). \tag{1}$$

Following Bates, Johnson, Lindenstrauss, Preiss and Schechtman [6], given a pair of Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, the space $\mathrm{Lip}(X,Y)$ of Lipschitz functions from X to Y is said to have the *uniform approximation by affine property* if for every $\varepsilon \in (0,\infty)$ there exists $r \in (0,1)$ such that for every 1-Lipschitz function $f : B_X \to Y$ there exists a radius $\rho \geqslant r$, a point $x \in X$ with $x + \rho B_X \subseteq B_X$, and an affine mapping $\Lambda : X \to Y$ such that $\|f(y) - \Lambda(y)\|_Y \leqslant \varepsilon\rho$ for every $y \in x + \rho B_X$. Denote the supremum over those r by $r^{X\to Y}(\varepsilon)$. The following theorem is due to [6].

Theorem 1 (Bates–Johnson–Lindenstrauss–Preiss–Schechtman)**.** *Suppose that* $(X, \|\cdot\|_X)$ *and* $(Y, \|\cdot\|_Y)$ *are Banach spaces with* $\dim(X) < \infty$*. Then* $\mathrm{Lip}(X,Y)$ *has the uniform approximation by affine property* **if and only if** Y *admits an equivalent uniformly convex norm.*

A function is differentiable if it admits arbitrarily good (after appropriate rescaling) affine approximations on infinitesimal balls. The uniform approximation by affine property was introduced as a way to make this phenomenon quantitative by requiring that the affine approximation occurs on a macroscopically large ball of a definite size that is independent of the specific 1-Lipschitz function. In addition to being a natural question in its own right, obtaining such quantitative information has geometric applications. In particular, Bates, Johnson, Lindenstrauss, Preiss and Schechtman introduced it in order to study [6] nonlinear quotient mappings between Banach spaces. Here we obtain the following improved lower bound on the modulus $r^{X\to Y}(\varepsilon)$ in the context of Theorem 1.

Theorem 2. *Suppose that* $(Y, \|\cdot\|_Y)$ *is a Banach space that admits an equivalent uniformly convex norm. Then there exists* $c = c(Y) \in (0,\infty)$ *such that for every* $n \in \mathbb{N}$*, every* n*-dimensional normed space* $(X, \|\cdot\|_X)$*, and every* $\varepsilon \in (0, 1/2]$ *we have* $r^{X\to Y}(\varepsilon) \geqslant \exp(-1/\varepsilon^{cn})$*.*

Theorem 2 answers Question 8 of [40] positively. We defer the description of (and comparison to) previous related works to later in the Introduction, after additional notation is introduced so as to facilitate such a discussion; see Section 1.F below. It suffices to summarize at the present juncture that the proof of Bates, Johnson, Lindenstrauss, Preiss and Schechtman in [6] did not yield any quantitative information on $r^{X\to Y}(\varepsilon)$. The first bound on this quantity, which is weaker as $n \to \infty$ than that of Theorem 2, was obtained by Li and and the second named author in [55]. A bound that is similar to that of Theorem 2 (though weaker in terms of the implicit dependence on the geometry of Y) was obtained by Li and both authors in [40] under an analytic assumption on Y that is strictly more stringent than the requirement that it admits an equivalent uniformly convex norm (which, by Theorem 1, is the correct setting for quantitative differentiation). The best known [40] upper bound on $r^{X\to Y}(\varepsilon)$ asserts that for every $p \in [2,\infty)$ there exists a uniformly convex Banach space Y such that if X is an n-dimensional Hilbert space then $r^{X\to Y}(\varepsilon) \leqslant \exp(-n(K/\varepsilon)^p)$ for every $\varepsilon \in (0, \varepsilon_0]$, with $K, \varepsilon_0 > 0$ being universal constants. The above estimates (both upper and lower) on $r^{X\to Y}(\varepsilon)$ are the best known even when X and Y are both Hilbert spaces and, say, $\varepsilon = 1/4$.

Our proof of Theorem 2 has conceptual significance that goes beyond the mere fact that it yields an asymptotically improved bound in the maximal possible generality. Firstly, Theorem 2 expresses the best quantitative differentiation result that is obtainable by available approaches, relying on a definitive local approximation estimate of independent interest (see Section 1.C below) that we derive here as a crucial step towards Theorem 2. Briefly, it seems that the L_p methods (and the corresponding Littlewood–Paley theory) that were used thus far have now reached their limit with Theorem 2, and in order to obtain a better lower bound on $r^{X\to Y}(\varepsilon)$ (if at all possible) one would need to work directly with L_∞ estimates, which would likely require a markedly different strategy.

Secondly, our proof of Theorem 2 contains contributions to Littlewood–Paley theory that are of significance in their own right. We rely on a novel semigroup argument (yielding as a side-product a new approach to classical results in harmonic analysis even for scalar-valued functions), but it turns out that our strategy is sensitive to the choice of semigroup, despite the semigroup's purely auxiliary role towards the geometric statement of Theorem 2. Specifically, our argument fails for the Poisson semigroup (even when Y is a Hilbert space) but does work for the heat semigroup. As a key step, we desire a Littlewood–Paley–Stein estimate for the corresponding $\mathcal{G}$-function for mappings that take values in uniformly convex Banach spaces. Such a theory has been developed for the Poisson semigroup initially by Xu [89], and in a definitive form in important work of Martínez, Torrea and Xu [58]. The availability of [58] has already played a decisive role in purely geometric questions [50], and it is therefore tempting to also try to use it in our context, but it turns out that obtaining the vector-valued Littlewood–Paley–Stein inequality for the heat semigroup was left open in [58]. We remedy this by proving new Littlewood–Paley–Stein $\mathcal{G}$-function estimates for the heat semigroup with values in uniformly convex targets, and using them to prove Theorem 2. The rest of the Introduction is devoted to a formal explanation of the above overview.

1.A. **Bourgain's strategy for the discretization problem.** Prior to stating the analytic results that we obtain here as steps towards the proof of Theorem 2, it would be beneficial to first present a geometric question due to Bourgain [13], known today as *Bourgain's discretization problem*, since it served both as inspiration for our subsequent proofs, as well as one of the reasons for our desire to obtain a lower bound on the modulus $r^{X\to Y}(\varepsilon)$. The formal link between the uniform approximation by affine property and Bourgain's discretization problem was clarified in [55], but the idea to use semigroup methods in the present context is new, motivated by an approach that Bourgain took within the proof of his discretization theorem in [13]. As an interesting "twist," we shall show that a "vanilla" adaptation of Bourgain's approach to our setting does not work, and in the process of overcoming this difficulty we shall obtain new results in vector-valued Littlewood–Paley theory.

The (bi-Lipschitz) distortion of a metric space (M, d_M) in a metric space (Z, d_Z) is denoted (as usual) by $c_Z(M) \in [1,\infty]$. Thus, the quantity $c_Z(M)$ is the infimum over those $D \in [1,\infty]$ for which there exists an embedding $\phi : M \to Z$ and (a scaling factor) $s \in (0,\infty)$ such that $sd_M(x,y) \leqslant d_Z(\phi(x),\phi(y)) \leqslant Dsd_M(x,y)$ for every $x,y \in M$. When Z is $\ell_2(M)$ (or any sufficiently large Hilbert space), $c_Z(M)$ is called the Euclidean distortion of M and is denoted $c_2(M)$.

Fix $n \in \mathbb{N}$. Let $(X, \|\cdot\|_X)$ be an n-dimensional normed space and let $(Y, \|\cdot\|_Y)$ be an arbitrary infinite dimensional Banach space. Bourgain's discretization problem asks for a lower estimate on the largest possible $\delta \in (0,1)$ such that for any δ-net $\mathcal{N}_\delta \subseteq B_X$ of B_X we have $c_Y(X) \leqslant 2c_Y(\mathcal{N}_\delta)$. Thus, the question at hand is to find the coarsest possible discrete approximation of B_X with the property that if it embeds into Y with a certain distortion then the entire space X also embeds into Y with at most twice that distortion (the factor 2 is an arbitrary choice; see [39] for a generalization). Bourgain's discretization theorem [13, 39] (see also Chapter 9 of the monograph [70]) asserts that

$$\delta \geqslant \exp\left(-c_Y(X)^{Kn}\right) \geqslant \exp\left(-n^{Kn}\right), \tag{2}$$

where $K \in [1,\infty)$ is a universal constant. The second inequality in (2) holds true because we always have $c_Y(X) \leqslant \sqrt{n}$ by John's theorem [42] and Dvoretzky's theorem [28].

The above discretization problem was introduced in [13] as an alternative (quantitative) approach to an important rigidity theorem of Ribe [82]. Additional applications to embedding theory appear in [68, 39, 69]. To date, the bound (2) remains the best known, even under the additional restriction that Y is uniformly convex. When Y is uniformly convex, a different proof that $\delta \geqslant \exp(-n^{Kn})$ for some $K = K(Y) \in [1,\infty)$ was obtained in [55] using the uniform approximation by affine property, and our Theorem 2 yields the stronger estimate $\delta \geqslant \exp(-c_Y(X)^{Kn})$ by [55, Remark 1.1].

The proof of (2) in [13] starts with a bi-Lipschitz embedding $\phi : \mathcal{N}_\delta \to Y$ and proceeds to construct an auxiliary mapping $f : X \to Y$. This is achieved through *Bourgain's almost extension* *theorem* [13], which is a nontrivial step but for the present purposes we do not need to recall the precise properties of f other than to state that f is Lipschitz, compactly supported, and that it well-approximates ϕ on the net $\mathcal{N}_\delta$. Having obtained a mapping f that is defined on all of X, [13] proceeds to examine the evolutes $\{P_tf\}_{t\in(0,\infty)}$ of f under the Poisson semigroup $\{P_t\}_{t\in(0,\infty)}$, i.e.,

$$\forall\, x\in\mathbb{R}^n,\qquad P_tf(x)\stackrel{\mathrm{def}}{=}p_t*f(x)=\int_{\mathbb{R}^n}p_t(y)f(x-y)\,\mathrm{d}y,$$

where the Poisson kernel $p_t:\mathbb{R}^n\to[0,\infty)$ is given by

$$\forall (x,t)\in\mathbb{R}^n\times(0,\infty),\qquad p_t(x)\stackrel{\mathrm{def}}{=}\frac{\Gamma\left(\frac{n+1}{2}\right)t}{(\pi t^2+\pi|x|^2)^{\frac{n+1}{2}}}.$$

Note that here we implicitly identified X with $\mathbb{R}^n$, with $|\cdot|$ being the standard Euclidean norm on $\mathbb{R}^n$; this issue will become important later, as discussed in Sections 1.B.2 and 1.C below.

A clever argument (by contradiction) in [13] now shows that since f is close to ϕ on the net $\mathcal{N}_\delta$ and ϕ itself is bi-Lipschitz, provided that the granularity δ of the net $\mathcal{N}_\delta$ is small enough there must exist a time $t\in(0,\infty)$ and a location $x\in X$ such that the derivative of the Poisson evolute P_tf at x is a (linear) bi-Lipschitz embedding of X into Y, with distortion at most a constant multiple of the distortion of ϕ. Here, since P_tf is obtained from f by averaging and f is Lipschitz, the fact that its derivative is Lipschitz is automatic. The difficulty is therefore to show that this derivative is invertible with good control on the operator norm of its inverse.

If an affine mapping is invertible on a sufficiently fine net of B_X then it is also invertible globally on X. So, if one could show that the first order Taylor polynomial of P_tf at x is sufficiently close to f on a sub-ball of B_X (and hence also close to ϕ on the intersection of that sub-ball with $\mathcal{N}_\delta$) whose radius is at least a sufficiently large constant multiple of δ, then this would imply the desired (quantitative) invertibility of the derivative of P_tf at x. Here, due to scale-invariance, "sufficiently close" means closeness after normalization by the radius of the sub-ball. This is the reason why a good lower bound on the modulus $r^{X\to Y}(\varepsilon)$ is helpful for Bourgain's discretization problem. Of course, one cannot hope to prove the bound (2) in this way in full generality, since (2) holds for any Banach space Y while by Theorem 1 we know that for $r^{X\to Y}(\varepsilon)$ to be positive we need Y to admit an equivalent uniformly convex norm (thus this approach is doomed to fail when, e.g., $Y=\ell_1$).

Nevertheless, when Y is uniformly convex one could take the fact that Bourgain's strategy does succeed as a hint to try to use the first order Taylor polynomial of P_tf as the affine mapping that is hopefully close to f on some macroscopically large sub-ball, thus obtaining a lower bound on $r^{X\to Y}(\varepsilon)$. This motivates the approach of the present article, eventually leading to Theorem 2.

An important issue here is that in any such argument one must find a way to use the fact that Y admits an equivalent uniformly convex norm, which by the work of Martínez, Torrea and Xu [58] is equivalent to the validity of a certain Y-valued Littlewood–Paley–Stein inequality for the Poisson semigroup; see Section 1.E below for a precise formulation. So, since the only underlying assumption on Y is equivalent to a certain L_q estimate, it is natural to use it to bound the L_q distance of $f(x)$ to the first order Taylor polynomial of P_tf at x, for an appropriate measure on the pairs $(x,t)\in X\times(0,\infty)$ of locations and times. By scale-invariance considerations, one arrives at a natural candidate L_q inequality that asserts that an appropriately normalized distance from $f(x)$ to the first order Taylor polynomial of P_tf at x is a Carleson measure; see Section 1.C below.

However, in Section 7 below we show that the desired L_q inequality does not hold true even when Y is a Hilbert space. The computations of Section 7 do suggest that for our purposes it would be better to use the heat semigroup in place of the Poisson semigroup. Unfortunately, the possible validity of the vector-valued Littlewood–Paley–Stein inequality for the heat semigroup for uniformly convex targets was previously unknown, being left open in [58] as part of a more general question that remains open in its full generality. So, as a key tool of independent interest, in the present article we also establish the desired vector-valued Littlewood–Paley–Stein inequality for the heat semigroup (which, by a standard subordination argument, is formally stronger than the corresponding inequality of [58] for the Poisson semigroup); see Section 1.E below. With this tool at hand, we proceed to prove Theorem 2 using the heat semigroup via the strategy outlined above.

1.B. **Geometric invariants.** Theorem 2 is a consequence of the analytic statement that is contained in Theorem 5 below. To formulate it, we need to first introduce notation related to (well-studied) geometric parameters that govern the ensuing arguments. We also recall the following standard conventions for asymptotic notation. Given $a, b \in (0,\infty)$, the notations $a \lesssim b$ and $b \gtrsim a$ mean that $a \leqslant cb$ for some universal constant $c \in (0,\infty)$. The notation $a \asymp b$ stands for $(a \lesssim b) \wedge (b \lesssim a)$. If we need to allow for dependence on parameters, we indicate this by subscripts. For example, in the presence of an auxiliary parameter q, the notation $a \lesssim_q b$ means that $a \leqslant c_q b$, where $c_q \in (0,\infty)$ is allowed to depend only on q, and similarly for the notations $a \gtrsim_q b$ and $a \asymp_q b$.

1.B.1. *The geometry of Y.* Despite the fact that in the definition (1) of uniform convexity the parameter $\delta \in (0,\infty)$ is allowed to have an arbitrary dependence on $\varepsilon \in (0,2]$, the following deep theorem of Pisier [76] asserts that by passing to an equivalent norm one can always ensure that δ is at least a constant multiple of a fixed power of ε.

Theorem 3 (Pisier's renorming theorem). *Suppose that $(Y, \|\cdot\|_Y)$ is a uniformly convex Banach space. Then there exists a norm $\|\cdot\|$ on Y that is equivalent to $\|\cdot\|_Y$ (thus there are $a, b \in (0,\infty)$ such that $a\|y\|_Y \leqslant \|y\| \leqslant b\|y\|_Y$ for all $y \in Y$) and constants $C, q \in [2,\infty)$ such that for every $x, y \in Y$ with $\|x\|, \|y\| \leqslant 1$ we have $\|x+y\| \leqslant 2 - \frac{1}{C^q}\|x-y\|^q$.*

In the literature, a Banach space that admits an equivalent uniformly convex norm is often called a superreflexive Banach space. Also, the conclusion of Theorem 3 is commonly referred to as the assertion that Y admits an equivalent norm with modulus of uniform convexity of power type q.

For norms that satisfy the conclusion of Theorem 3, Pisier proved [76] the following important martingale inequality. To state it, recall that a sequence of Y-valued random variables $\{M_k\}_{k=1}^\infty$ on a probability space $(\mathcal{S}, \mathcal{F}, \mu)$ is said to be a martingale if there exists an increasing sequence of sub-σ-algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}$ such that $\mathbb{E}[M_{k+1}|\mathcal{F}_k] = M_k$ for every $k \in \mathbb{N}$. Here $\mathbb{E}[\cdot|\mathcal{F}_k]$ stands for the conditional expectation relative to the σ-algebra $\mathcal{F}_k$ and we are assuming that $M_k \in L_1(\mu; Y)$ for every $k \in \mathbb{N}$, where for $q \in [1,\infty]$ the corresponding vector-valued Lebesgue space $L_q(\mu; Y)$ consists (as usual) of all the $\mathcal{F}$-measurable mappings $f : \mathcal{S} \to Y$ for which $\|f\|_{L_q(\mu;Y)}^q = \int_{\mathcal{S}} \|f\|_Y^q \,\mathrm{d}\mu < \infty$.

Theorem 4 (Pisier's martingale inequality). *Fix $C \in (0,\infty)$ and $q \in [2,\infty)$. Suppose that $(Y, \|\cdot\|_Y)$ is a Banach space such that $\|x+y\|_Y \leqslant 2 - \frac{1}{C^q}\|x-y\|_Y^q$ for every $x, y \in Y$ with $\|x\|_Y, \|x\|_Y \leqslant 1$. Then every martingale $\{M_k\}_{k=1}^\infty \subseteq L_q(\mu; Y)$ satisfies*

$$\bigg(\sum_{k=1}^{\infty} \|M_{k+1} - M_k\|_{L_q(\mu;Y)}^q\bigg)^{\frac{1}{q}} \lesssim C \sup_{k \in \mathbb{N}} \|M_k\|_{L_q(\mu;Y)}. \tag{3}$$

For the proof of (3) as stated above (i.e., with the constant factor that appears in the right hand side of (3) being proportional to the constant C of the assumption on Y), see [63, Section 6.3] combined with the proof of [5, Proposition 7] (the case $q = 2$ of this argument is due to K. Ball [4]).

Inspired by Theorem 4, Pisier introduced the following terminology in [80]. Given a Banach space $(Y, \|\cdot\|_Y)$ and $q \geqslant 2$, the martingale cotype q constant of Y, denoted $\mathfrak{m}_q(Y, \|\cdot\|_Y) \in [1,\infty]$ or simply $\mathfrak{m}_q(Y)$ if the norm is clear from the context, is the supremum of $(\sum_{k=1}^\infty \int_{\mathcal{S}} \|M_{k+1} - M_k\|_Y^q \,\mathrm{d}\mu)^{1/q}$ over all martingales $\{M_k\}_{k=1}^\infty \subseteq L_q(\mu; Y)$ with $\sup_{k \in \mathbb{N}} \int_{\mathcal{S}} \|M_k\|_Y^q \,\mathrm{d}\mu = 1$ (and over all probability spaces $(\mathcal{S}, \mathcal{F}, \mu)$). If $\mathfrak{m}_q(Y) < \infty$ then we say that Y has martingale cotype q. Pisier's work [80] yields the remarkably satisfactory characterization that Y admits an equivalent norm whose modulus of uniform convexity has power type q if and only if Y has martingale cotype q (with the relevant constants being within universal constant factors of each other).

The UMD constant of a Banach space $(Y, \|\cdot\|_Y)$, commonly denoted $\beta(Y, \|\cdot\|_Y) \in (0, \infty]$ or simply $\beta(Y)$ if the norm is clear from the context, is the infimum over those $\beta \in (0, \infty]$ such that for every martingale $\{M_k\}_{k=1}^{\infty} \subseteq L_2(\mu; Z)$, every $n \in \mathbb{N}$ and every $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$ we have

$$\Big\|M_1 + \sum_{k=1}^{n} \varepsilon_k (M_{k+1} - M_k)\Big\|_{L_2(\mu;Y)} \leqslant \beta \|M_{k+1}\|_{L_2(\mu;Y)}.$$

If $\beta(Y) < \infty$ then Y is said to be a UMD space. There exist [77] uniformly convex Banach spaces that are not UMD, and there even exist such Banach lattices [11, 81]. If Y is UMD then it admits an equivalent uniformly convex norm [60]. As a quantitative form of this assertion (that will be used below), it follows from [40, Section 4.4] that there exists $2 \leqslant q \lesssim \beta(Y)$ such that $\mathfrak{m}_q(Y) \lesssim \beta(Y)^2$.

1.B.2. *The geometry of X.* Recalling that $(X, \|\cdot\|_X)$ is an n-dimensional (real) normed space, once we fix a Hilbertian norm $|\cdot|$ on X we can identify it (as a real vector space) with $\mathbb{R}^n$. The specific choice of Euclidean structure will be very important later, but at this juncture we shall think of $|\cdot|$ as an arbitrary Hilbertian norm on X and derive an inequality that holds in such (full) generality.

Throughout what follows, the scalar product of two vectors $x, y \in \mathbb{R}^n$ is denoted $x \cdot y \in \mathbb{R}$, the volume of a Lebesgue measurable subset $\Omega \subseteq \mathbb{R}^n$ is denoted $|\Omega|$, and integration with respect to the Lebesgue measure on $\mathbb{R}^n$ is indicated by $\mathrm{d}x$. The Euclidean unit ball in $\mathbb{R}^n$ is denoted $B^n = \{x \in \mathbb{R}^n : \ |x| \leqslant 1\}$. Thus $|B^n| = \pi^{n/2}/\Gamma(1 + n/2)$. The Euclidean unit sphere is denoted (as usual) $S^{n-1} = \partial B^n = \{x \in \mathbb{R}^n : \ |x| = 1\}$, integration with respect to the surface area measure on S^{n-1} is indicated by $\mathrm{d}\sigma$ and, while slightly abusing notation, we denote the surface area of a Lebesgue measurable subset $A \subseteq S^{n-1}$ by $|A|$. Thus $|S^{n-1}| = n|B^n| = 2\pi^{n/2}/\Gamma(n/2)$.

If $\Omega \subseteq \mathbb{R}^n$ is Lebesgue measurable and has positive finite volume then it will be convenient to use the following notation for the average over Ω of an integrable function $f : \Omega \to \mathbb{R}$.

$$\fint_{\Omega} f(x)\, \mathrm{d}x \stackrel{\mathrm{def}}{=} \frac{1}{|\Omega|} \int_{\Omega} f(x)\, \mathrm{d}x. \tag{4}$$

Analogously, write $\fint_A \phi \, \mathrm{d}\sigma \stackrel{\mathrm{def}}{=} \frac{1}{|A|} \int_A \phi(\sigma)\, \mathrm{d}\sigma$ for measurable $A \subseteq S^{n-1}$ and integrable $\phi : A \to \mathbb{R}$.

Using standard notation of asymptotic convex geometry (as in e.g. [14]), denote[1]

$$M(X) \stackrel{\mathrm{def}}{=} \fint_{S^{n-1}} \|\sigma\|_X \, \mathrm{d}\sigma \qquad \text{and} \qquad \forall\, q \in (0, \infty], \qquad I_q(X) \stackrel{\mathrm{def}}{=} \bigg(\fint_{B_X} |x|^q \, \mathrm{d}x \bigg)^{\frac{1}{q}}. \tag{5}$$

In what follows, the quantity $I_q(X)M(X)$ has an important role. We shall present a nontrivial upper bound on it (for an appropriate choice of Euclidean norm $|\cdot|$ on X) as a quick consequence of powerful results from asymptotic convex geometry, and we shall formulate conjectures about the possible availability of better bounds; some of these conjectures may be quite difficult, however, because they relate to longstanding open problems in convex geometry. We postpone these discussions for the moment since it will be more natural to treat them after we present Theorem 5.

Fixing from now on a target Banach space $(Y, \|\cdot\|_Y)$, for $p \in [1, \infty]$ the corresponding Y-valued Lebesgue–Bochner space on a domain $\Omega \subseteq \mathbb{R}^n$ will be denoted $L_p(\Omega; Y)$ (the underlying measure on Ω will always be understood to be the Lebesgue measure). When $Y = \mathbb{R}$ we shall use the usual simpler notation $L_p(\Omega; \mathbb{R}) = L_p(\Omega)$ for the corresponding scalar-valued function space.

---

[1]In the literature it is common to suppress the Euclidean norm $|\cdot|$ in this notation, but these quantities do depend on it. A possible more precise notation would have been to use $M(\|\cdot\|_X, |\cdot|)$ and $I_q(\|\cdot\|_X, |\cdot|)$. However, this more cumbersome notation isn't necessary here because the ambient Euclidean norm will always be clear from the context.

The Y-valued heat semigroup on $\mathbb{R}^n$ will be denoted by $\{H_t\}_{t\in(0,\infty)}$. Thus, for every $t\in(0,\infty)$ and $f\in L_1(\mathbb{R}^n;Y)$ the function $H_tf:\mathbb{R}^n\to Y$ is defined by

$$\forall\, x\in\mathbb{R}^n,\qquad H_tf(x)\stackrel{\mathrm{def}}{=} h_t*f(x)=\int_{\mathbb{R}^n} h_t(z)f(x-z)\,\mathrm{d}z,$$

where the corresponding heat kernel $h_t:\mathbb{R}^n\to[0,\infty)$ is given by

$$\forall (t,x)\in(0,\infty)\times\mathbb{R}^n,\qquad h_t(x)\stackrel{\mathrm{def}}{=}\frac{1}{(4\pi t)^{\frac{n}{2}}}e^{-\frac{|x|^2}{4t}}=\frac{1}{t^{\frac{n}{2}}}h_1\Big(\frac{x}{\sqrt{t}}\Big).$$

The first order Taylor polynomial at a point $x\in\mathbb{R}^n$ of a differentiable function $f:\mathbb{R}^n\to Y$ will be denoted below by $\mathrm{Taylor}^1_x(f):\mathbb{R}^n\to Y$. Thus, $\mathrm{Taylor}^1_x(f)$ is the affine function given by

$$\forall\, x,y\in\mathbb{R}^n,\qquad \mathrm{Taylor}^1_x(f)(y)\stackrel{\mathrm{def}}{=} f(x)+(y-x)\cdot\nabla f(x),$$

where for every $x,z\in\mathbb{R}^n$ we set (as usual) $z\cdot\nabla f(x)=\sum_{j=1}^n z_j\partial_j f(x)=\lim_{\varepsilon\to0}(f(x+\varepsilon z)-f(x))/\varepsilon$ to be the corresponding Y-valued directional derivative of f.

Despite the fact that in our setting $\mathbb{R}^n$ is endowed with two metrics, namely those that are induced by $\|\cdot\|_X$ and $|\cdot|$, when discussing Lipschitz constants of mappings from subsets of $\mathbb{R}^n$ to Y we will adhere to the convention that they are exclusively with respect to the metric that is induced by the norm $\|\cdot\|_X$. In particular, we shall use the following notation for a mapping $f:\mathbb{R}^n\to Y$.

$$\|f\|_{\mathrm{Lip}(X,Y)}\stackrel{\mathrm{def}}{=}\sup_{\substack{x,y\in\mathbb{R}^n\\ x\neq y}}\frac{\|f(x)-f(y)\|_Y}{\|x-y\|_X}\qquad\text{and}\qquad \|f\|_{\mathrm{Lip}(B_X,Y)}\stackrel{\mathrm{def}}{=}\sup_{\substack{x,y\in B_X\\ x\neq y}}\frac{\|f(x)-f(y)\|_Y}{\|x-y\|_X}.$$

Hence, if f is differentiable then $\|f\|_{\mathrm{Lip}(X,Y)}=\sup_{z\in\partial B_X}\|z\cdot\nabla f\|_{L_\infty(\mathbb{R}^n;Y)}$.

1.C. **Dorronsoro estimates.** Our proof of Theorem 2 uses Theorem 5 below, which shows that at most scales and locations the first order Taylor polynomial of a heat evolute of a 1-Lipschitz function $f:\mathbb{R}^n\to Y$ must be close to f itself. Using standard terminology, our arguments imply that $(t^{-q}\fint_{x+tB_X}\|f(y)-\mathrm{Taylor}^1_x(H_{\gamma t^2}f)(y)\|_Y^q\,\mathrm{d}y)\frac{\mathrm{d}x\,\mathrm{d}t}{t}$ is a Carleson measure for a certain $\gamma>0$.

Theorem 5. *There exists a universal constant $\kappa\in[2,\infty)$ with the following property. Suppose that $q\in[2,\infty)$ and $n\in\mathbb{N}$, and that $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ are Banach spaces that satisfy $\dim(X)=n$ and $\mathfrak{m}_q(Y)<\infty$. Let $|\cdot|$ be any Hilbertian norm on X, thus identifying X with $\mathbb{R}^n$. Define $\gamma,K\in(0,\infty)$ by*

$$\gamma=\gamma(q,X)\stackrel{\mathrm{def}}{=}\frac{I_q(X)}{\sqrt{n}M(X)}\qquad\text{and}\qquad K=K(q,n,X,Y)\stackrel{\mathrm{def}}{=}\kappa\sqrt[4]{n}\cdot\mathfrak{m}_q(Y)\sqrt{I_q(X)M(X)}.\tag{6}$$

Then every compactly supported Lipschitz function $f:\mathbb{R}^n\to Y$ satisfies the following estimate.

$$\bigg(\int_{\mathbb{R}^n}\int_0^\infty\fint_{x+tB_X}\frac{\|f(y)-\mathrm{Taylor}^1_x(H_{\gamma t^2}f)(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\bigg)^{\frac1q}\leqslant K|\operatorname{supp}(f)|^{\frac1q}\|f\|_{\mathrm{Lip}(X,Y)}.\tag{7}$$

Remark 6. As we discussed in Section 1.A, the analogue of Theorem 5 for the Poisson semigroup is not true. Specifically, in Section 7 we show that if $(\mathscr{H},\|\cdot\|_{\mathscr{H}})$ is a Hilbert space then for every nonconstant Lipschitz function $f:\ell_2^n\to\mathscr{H}$ and every $\gamma\in(0,\infty)$ the following integral diverges.

$$\bigg(\int_{\mathbb{R}^n}\int_0^\infty\fint_{x+tB^n}\frac{\|f(y)-\mathrm{Taylor}^1_x(P_{\gamma t}f)(y)\|_{\mathscr{H}}^2}{t^3}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\bigg)^{\frac12}.\tag{8}$$

Note that (7) considers the heat evolute of f at time γt^2 while (8) considers the Poisson evolute of f at time γt because these time choices are determined by the requirement that when the argument of f is rescaled the relevant quantities (namely, the left hand side of (7) when $q=2$ and the quantity appearing in (8)) have the same order of homogeneity as the right hand side of (7) (when $q=2$).

Inequality (7) is not determined solely by intrinsic geometric properties of X due to the auxiliary choice of the Hilbertian norm $|\cdot|$ on X, which influences the quantities γ and K that appear in (6), as well as the meaning of the heat semigroup $\{H_t\}_{t\in[0,\infty)}$. To deduce an intrinsic statement from Theorem 5, namely a statement that refers only to geometric characteristics of X and Y without any additional (a priori arbitrary) choices, let $\mathcal{A}(X,Y)$ denote the space of affine mappings from X to Y. Then, continuing with the notations of Theorem 5, for every $x\in X$ and $t\in(0,\infty)$ we have

$$\fint_{x+tB_X}\frac{\|f(y)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y \geqslant \inf_{\substack{\Lambda\in\mathcal{A}(X,Y)\\ \|\Lambda\|_{\mathrm{Lip}(X,Y)}\leqslant\|f\|_{\mathrm{Lip}(X,Y)}}}\fint_{x+tB_X}\frac{\|f(y)-\Lambda(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y,$$

where we used the fact that in the above integrand, since $H_{\gamma t^2}f$ is obtained from f by convolution with a probability measure, we have $\|H_{\gamma t^2}f\|_{\mathrm{Lip}(X,Y)}\leqslant\|f\|_{\mathrm{Lip}(X,Y)}$, and consequently also the affine mapping $\mathrm{Taylor}_x^1(H_{\gamma t^2}f)$ has Lipschitz constant at most $\|f\|_{\mathrm{Lip}(X,Y)}$. Therefore (7) implies that

$$\bigg(\int_X\int_0^\infty \inf_{\substack{\Lambda\in\mathcal{A}(X,Y)\\ \|\Lambda\|_{\mathrm{Lip}(X,Y)}\leqslant\|f\|_{\mathrm{Lip}(X,Y)}}}\fint_{x+tB_X}\frac{\|f(y)-\Lambda(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\bigg)^{\frac{1}{q}}$$
$$\lesssim \sqrt[4]{n}\cdot\mathfrak{m}_q(Y)\sqrt{I_q(X)M(X)}\cdot|\operatorname{supp}(f)|^{\frac{1}{q}}\|f\|_{\mathrm{Lip}(X,Y)}. \qquad (9)$$

The inequality (9) depends on the auxiliary Hilbertian norm $|\cdot|$ on X only through the quantity $I_q(X)M(X)$ that appears on the right hand side of (9), and it is clearly in our interest to choose the Hilbertian structure on X so as to make this quantity as small as possible. Since the definitions of $M(X)$ and $I_q(X)$ in (5) involve averagings, if for some $D\in[1,\infty)$ we have $\|x\|_X\leqslant|x|\leqslant D\|x\|_X$ for every $x\in X$ then $I_q(X)M(X)\leqslant D$. By John's theorem [42], if B^n is the ellipsoid of maximum volume contained in B_X then $D\leqslant\sqrt{n}$, so it is always the case that $I_q(X)M(X)\leqslant\sqrt{n}$ for some choice of Hilbertian structure on X. Of course, it would be better to choose here the Hilbertian norm $|\cdot|$ so as to minimize D, in which case (using a standard differentiation argument [8]) D becomes the Euclidean distortion $c_2(X)$. So, we always have $I_q(X)M(X)\leqslant c_2(X)\leqslant\sqrt{n}$ for some Euclidean norm $|\cdot|$ on X, but it turns out that this estimate is very crude. Firstly, one can improve it (up to constant factors) to the assertion that there exists a Euclidean norm $|\cdot|$ on X for which $I_q(X)M(X)\lesssim_q T_2(X)$, where $T_2(X)\leqslant c_2(X)$ is the Rademacher type 2 constant of X; see Remark 28 below, where the definition of Rademacher type is recalled and this estimate is justified. In terms of the dependence on the dimension n, we have for example $c_2(\ell_\infty^n)=c_2(\ell_1^n)=\sqrt{n}$ while by a direct computation one sees that $I_q(\ell_\infty^n)M(\ell_\infty^n)\asymp_q\sqrt{\log n}$ and $I_q(\ell_1^n)M(\ell_1^n)\asymp_q 1$ (more generally, for $p\in[1,\infty)$ one computes that $I_q(\ell_p^n)M(\ell_p^n)\asymp_{p,q}1$).[2] Also, if X has a C-unconditional basis for some $C\in[1,\infty)$ then $I_q(X)M(X)\lesssim_q C^2\sqrt{\log n}$, as explained in Remark 29 below.

Conjecture 7. *For every $n\in\mathbb{N}$ and $q\in[1,\infty)$, every n-dimensional normed space $(X,\|\cdot\|_X)$ admits a Hilbertian norm $|\cdot|$ with respect to which we have $I_q(X)M(X)\lesssim_q\sqrt{\log n}$.*

The currently best known upper bound on $I_q(X)M(X)$ in terms of $n=\dim(X)$ occurs when the Hilbertian structure is chosen so as to make X isotropic, where we recall that X is said to be isotropic if the Hilbertian norm $|\cdot|$ satisfies $|B_X|=1$ and there is $L_X\in(0,\infty)$ such that

$$\forall\, y\in\mathbb{R}^n,\qquad \bigg(\int_{B_X}(x\cdot y)^2\,\mathrm{d}x\bigg)^{\frac{1}{2}}=L_X|y|. \qquad (10)$$

[2]Here, and throughout the rest of this article, given an integer $n\in\mathbb{N}$ we shall use the nonconventional interpretation of the quantity $\log n$ as being equal to the usual natural logarithm when $n\geqslant 2$, but equal to 1 when $n=1$. This is done only for the purpose of ensuring that all the ensuing statements are correct also in the one-dimensional setting without the need to write more cumbersome expressions. Alternatively, one can assume throughout that $n\geqslant 2$.

Every finite dimensional normed space X admits a unique Hilbertian norm with respect to which it is isotropic. The quantity L_X in (10) is called the isotropic constant of X; see the monograph [14] for more about isotropicity. In Section 5 below we explain how a direct combination of (major) results in convex geometry shows that if X is an n-dimensional isotropic normed space and $q \in [1,\infty)$ then

$$n \geqslant q^2 \implies I_q(X)M(X) \lesssim (n\log n)^{\frac{2}{5}}. \tag{11}$$

The restriction $n \geqslant q^2$ in (11) corresponds to the most interesting range of parameters, but in Section 5 we also present the currently best known bound when $n \leqslant q^2$; see inequality (68) below. We make no claim that (11) is best possible, the main point being that (11) is asymptotically better as $n \to \infty$ than the bound of $\sqrt{n}$ that follows from John's theorem. It is tempting to speculate that the upper bound on $I_q(X)M(X)$ of Conjecture (7) holds true already when X is isotropic. This refined version of Conjecture (7) seems challenging because, as we explain in Remark 32 below, we always have $I_q(X)M(X) \gtrsim L_X$, so a positive answer would yield the estimate $L_X \lesssim \sqrt{\log n}$, which would be much stronger than the currently best known [47] bound $L_X \lesssim \sqrt[4]{n}$ (the longstanding Slicing Problem [12, 3, 65] asks whether L_X could be bounded from above by a universal constant).

By substituting (11) into (9) we obtain Theorem 8 below, which is an intrinsic version of Theorem 2. Of course, any future improvement over (11) (for any Hilbertian structure on X) would immediately imply an improved dependence on n in Theorem 8.

Theorem 8 (Intrinsic vector-valued Dorronsoro estimate)**.** *Suppose that $q \in [2,\infty)$ and $n \in \mathbb{N}$ satisfy $n \geqslant q^2$. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces that satisfy $\dim(X) = n$ and $\mathfrak{m}_q(Y) < \infty$. Then every compactly supported 1-Lipschitz function $f : X \to Y$ satisfies*

$$\bigg(\int_X \int_0^\infty \inf_{\substack{\Lambda \in \mathcal{A}(X,Y)\\ \|\Lambda\|_{\mathrm{Lip}(X,Y)} \leqslant 1}} \fint_{x+tB_X} \frac{\|f(y)-\Lambda(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\bigg)^{\frac{1}{q}} \lesssim n^{\frac{9}{20}} \sqrt[5]{\log n}\cdot \mathfrak{m}_q(Y)|\operatorname{supp}(f)|^{\frac{1}{q}}. \tag{12}$$

The above nomenclature arises from important classical work of Dorronsoro [27], who obtained Theorem (8) when $Y = \mathbb{R}$ (in which case $\mathfrak{m}_q(\mathbb{R}) \asymp 1$ for every $q \geqslant 2$) and $X = \ell_2^n$, but with much weaker (implicit) dependence on the dimension n. As we shall see in Section 7 below, in the special case $X = \ell_2^n$, $Y = \ell_2$ and $q = 2$, a more careful analysis yields the validity of (12) with the right hand side being dimension independent (in forthcoming work of Danailov and Fefferman [25], an even sharper result is obtained in this Hilbertian setting, yielding the precise value of the implicit universal constant). We do not know if it is possible to obtain a dimension independent version of Theorem 8 in its full generality, but even if that were possible then it would not influence the statement of Theorem 2 (only the value of the universal constant c will be affected).

In their classical (scalar-valued) form, Dorronsoro estimates are very influential in several areas, including singular integrals (e.g. [44]), geometric measure theory (e.g. [26]), local approximation spaces (e.g. [87]), PDE (e.g. [49]), calculus of variations (e.g. [48]). Our semigroup proof of Theorem 8 is via a strategy that differs from Dorronsoro's original approach [27] as well as the subsequent approaches of Fefferman [33], Jones [44], Seeger [85], Triebel [87], Kristensen–Mingione [48] and Azzam–Schul [1]. Importantly, this semigroup strategy is what makes it possible for us to obtain for the first time the validity of Theorem 8 when Y is superreflexive (i.e., Y admits an equivalent uniformly convex norm), thus leading to Theorem 2. The best previously known result [40] was that a variant of Theorem 8 holds true in the more restrictive setting when Y is a UMD Banach space; this was achieved in [40] via a (quite subtle) adaptation of Dorronsoro's original interpolation-based method [27], and we do not see how to make such an approach apply to superreflexive targets. Theorem 9 below shows that the validity of (12) for any fixed n, q, X and with any constant multiplying $|\operatorname{supp}(f)|^{1/q}$ in the right hand side of (12) implies that Y is superreflexive, hence Theorem 8 as stated above yields a vector-valued Dorronsoro estimate in the maximal possible generality.

Theorem 9 (Characterization of superreflexivity in terms of a Dorronsoro estimate)**.** *The following conditions are equivalent for a Banach space $(Y,\|\cdot\|_Y)$.*

(1) *Y admits an equivalent uniformly convex norm (Y is superreflexive).*
(2) *There exists $q\in[2,\infty)$ and for every $n\in\mathbb{N}$ there exists $C=C(n,Y)\in(0,\infty)$ such that for every n-dimensional normed space $(X,\|\cdot\|_X)$ and every 1-Lipschitz compactly supported function $f:X\to Y$ we have*

$$\int_X\int_0^\infty \inf_{\Lambda\in\mathcal{A}(X,Y)} \fint_{x+tB_X} \frac{\|f(y)-\Lambda(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\leqslant C|\operatorname{supp}(f)|. \tag{13}$$

(3) *There exist $q,C\in[2,\infty)$, $n_0\in\mathbb{N}$ and an n_0-dimensional normed space $(X_0,\|\cdot\|_{X_0})$ such that (13) holds true for every 1-Lipschitz compactly supported function $f:X_0\to Y$.*

Note that, in contrast to (12), the infimum in (13) is over all $\Lambda\in\mathcal{A}(X,Y)$ without any restriction on the Lipschitz constant of Λ. It is natural to ask whether or not one could refine Theorem 9 so as to yield a characterization of those Banach spaces $(Y,\|\cdot\|_Y)$ that admit an equivalent norm whose modulus of uniform convexity is of power type q, or equivalently that Y has martingale cotype q.

Question 10. Does the validity of (13) imply that $\mathfrak{m}_q(Y)<\infty$?

While we did not dedicate much effort to try to answer Question 10, some partial results are obtained in Section 6 below, including the assertion that if Y is a Banach lattice that satisfies (13) then $\mathfrak{m}_{q+\varepsilon}(Y)<\infty$ for every $\varepsilon\in(0,\infty)$.

Remark 11. The literature also contains [27, 85, 87] Dorronsoro estimates corresponding to local approximation by higher degree polynomials rather than by degree 1 polynomials as in (12). We made no attempt to study such extensions in our setting, since the goal of the present article is the geometric application of Theorem 2. Nevertheless, an inspection of our proofs reveals that they do yield mutatis mutandis vector-valued Dorronsoro estimates for local approximation by polynomials of any degree of sufficiently smooth functions with values in uniformly convex targets.

1.D. **A local Dorronsoro estimate and L_q affine approximation.** To explain the link between Theorem 2 and Theorem 5, note first that Theorem 2 deals with functions that are defined on the unit ball B_X, while Theorem 5 deals with functions that are defined globally on all of $\mathbb{R}^n$. So, in order to prove Theorem 2 we shall first establish Theorem 12 below, which is a localized version of Theorem 5. To state it, it will be notationally convenient (and harmless) to slightly abuse (but only when $n=1$) the notation for averages that was introduced in (4) as follows. Given $a,A\in(0,\infty)$ with $a<A$ and $\psi:\mathbb{R}\to\mathbb{R}$ such that the mapping $\rho\mapsto\psi(\rho)/\rho$ is in $L_1([a,A])$, denote the average of ψ with respect to the measure $\frac{\mathrm{d}\rho}{\rho}$ over the interval $[a,A]$ by $\fint_a^A\psi(\rho)\frac{\mathrm{d}\rho}{\rho}=\frac{1}{\log(A/a)}\int_a^A\psi(\rho)\frac{\mathrm{d}\rho}{\rho}$.

Theorem 12 (Local vector-valued Dorronsoro estimate)**.** *There is a universal constant $c\in(0,1/4)$ with the following properties. Suppose that $q\in[2,\infty)$, $n\in\mathbb{N}$, and that $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ are Banach spaces that satisfy $\dim(X)=n$ and $\mathfrak{m}_q(Y)<\infty$. Let $|\cdot|$ be any Hilbertian norm on X, thus identifying X with $\mathbb{R}^n$. Let $K\in(0,\infty)$ be defined as in (6), and define also*

$$T\stackrel{\mathrm{def}}{=}\frac{c}{n^{\frac54}\sqrt{I_q(X)M(X)\log n}}. \tag{14}$$

Then $T\leqslant 1/(2n)$. Moreover, for every 1-Lipschitz function $f:B_X\to Y$ and every $r\in(0,T^2]$,

$$\fint_r^T\left(\fint_{(1-\frac{1}{2n})B_X}\inf_{\substack{\Lambda\in\mathcal{A}(X,Y)\\ \|\Lambda\|_{\mathrm{Lip}(X,Y)}\leqslant 2}}\frac{\fint_{x+\rho B_X}\|f(y)-\Lambda(y)\|_Y^q\,\mathrm{d}y}{\rho^q}\,\mathrm{d}x\right)\frac{\mathrm{d}\rho}{\rho}\leqslant\frac{(9Kn)^q}{|\log r|}. \tag{15}$$

The assertion of Theorem 12 that $T \leqslant 1/(2n)$ is needed only in order to make the integrals that appear in the left hand side of (15) well-defined, ensuring that for every $x \in (1-1/(2n))B_X$ and $\rho \in [r,T]$, every point y in the ball $x+\rho B_X$ is also in B_X, i.e., y is in the domain of f and the integrand makes sense. This upper bound on T follows automatically from the restriction $c \leqslant 1/4$, since it is always the case that $I_q(X)M(X) \geqslant 1/2$, as explained in Corollary 31 below. The heart of the matter is therefore to obtain the estimate (15). See Section 3 below for the deduction of Theorem 12 from Theorem 5, where the additional information in Theorem 5 that the approximating affine function Λ is actually a heat evolute is used to show that Λ is 2-Lipschitz, a property that we shall soon use crucially to deduce Theorem 2.

The deduction of Theorem 2 from Theorem 12 is quick. Suppose that $f: B_X \to Y$ is 1-Lipschitz and fix $\delta \in (0,1/2)$. Let $C \in [9,\infty)$ be a (large enough) universal constant ensuring that if we define $r = \exp(-(CKn/\delta)^q)$ then $r \leqslant T^2$, where K is defined in (6) and T is defined in (14); the existence of such a universal constant follows immediately from the fact that $I_q(X)M(X) \geqslant 1/2$ and the definitions of K and T. Now, by (15) there exists a radius $\rho \geqslant r$, a point $x \in B_X$ with $x+\rho B_X \subseteq B_X$, and an affine mapping $\Lambda : X \to Y$ with $\|\Lambda\|_{\mathrm{Lip}(X,Y)} \leqslant 2$ such that

$$\left(\fint_{x+\rho B_X} \|f(y)-\Lambda(y)\|_Y^q \,\mathrm{d}y\right)^{\frac{1}{q}} \leqslant \delta\rho. \tag{16}$$

Next, the (simple) argument of [40, Section 2.1] shows that given $\varepsilon \in (0,1/2)$, if the L_q-closeness of f to Λ that appears in (16) is exponentially small in n then necessarily $\|f(y)-\Lambda(y)\|_Y \leqslant \varepsilon\rho$ for *every* $y \in x+\rho B_X$. Specifically, this holds true if $\delta = (\eta\varepsilon)^{1+n/q}$ for a sufficiently small universal constant $\eta \in (0,1)$. Briefly, the reason for this fact is that since $\|f\|_{\mathrm{Lip}(B_X,Y)} \leqslant 1$ and $\|\Lambda\|_{\mathrm{Lip}(X,Y)} \leqslant 2$ we have $\|f-\Lambda\|_{\mathrm{Lip}(B_X,Y)} \leqslant 3$, and consequently if $\|f(y_0)-\Lambda(y_0)\|_Y$ were larger than $\varepsilon\rho$ for some $y_0 \in x+\rho B_X$ then it would follow that $\|f(y)-\Lambda(y)\|_Y$ is larger than a constant multiple of $\varepsilon\rho$ on a sub-ball of $x+\rho B_X$ of radius that is at least a constant multiple of $\varepsilon\rho$, thus making the left hand side of (16) be greater than $(\eta\varepsilon)^{1+n/q}\rho$ for some universal constant $\eta \in (0,1)$. So, by choosing $\delta = (\eta\varepsilon)^{1+n/q}$, recalling that we defined $r = \exp(-(CKn/\delta)^q)$ and recalling also the definition of K in (6), if we set $\mathfrak{a} = \eta/(C\kappa)$ (with κ being the universal constant of Theorem 5) we obtain the following refined version of Theorem 2.

Corollary 13. *There exists a universal constant $\mathfrak{a} \in (0,1)$ such that for every $q \in [2,\infty)$, $n \in \mathbb{N}$, and every two Banach spaces $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ that satisfy $\dim(X) = n$ and $\mathfrak{m}_q(Y) < \infty$, if $|\cdot|$ is any Hilbertian norm on X and $\varepsilon \in (0,1/2)$ then*

$$r^{X\to Y}(\varepsilon) \geqslant \exp\left(-\frac{\left(n^{\frac{5}{4}}\sqrt{I_q(X)M(X)}\cdot\mathfrak{m}_q(Y)\right)^q}{(\mathfrak{a}\varepsilon)^{n+q}}\right).$$

Consequently, since an appropriate choice of the Hilbertian norm $|\cdot|$ ensures that $I_q(X)M(X) \leqslant \sqrt{n}$, for every $\varepsilon \in (0,1/2)$ we have

$$r^{X\to Y}(\varepsilon) \geqslant \exp\left(-\frac{n^{\frac{3q}{2}}\mathfrak{m}_q(Y)^q}{(\mathfrak{a}\varepsilon)^{n+q}}\right). \tag{17}$$

Remark 14. Observe that the fact that $\|\Lambda\|_{\mathrm{Lip}(X,Y)} \leqslant 2$ in (16) was used crucially in the above deduction of Corollary 13 from Theorem 12. Any universal constant in place of 2 would work just as well for the purpose of this deduction, but an upper bound on $\|\Lambda\|_{\mathrm{Lip}(X,Y)}$ that grows to ∞ with n would result in an asymptotically weaker lower bound in Theorem 2.

Remark 15. In the setting of Theorem 12, notation for an L_q affine approximation modulus was introduced as follows in [40, Definition 1]. Given $\delta \in (0,1)$ let $r_q^{X\to Y}(\delta)$ be the supremum over those $r \in [0,1]$ such that for every 1-Lipschitz function $f : B_X \to Y$ there exists a radius $\rho \geqslant r$, a point $x \in B_X$ with $x + \rho B_X \subseteq B_X$, and an affine mapping $\Lambda : X \to Y$ with $\|\Lambda\|_{\mathrm{Lip}(X,Y)} \leqslant 3$ such that the estimate (16) is satisfied. The constant 3 of the definition of this modulus was chosen in [40] essentially arbitrarily (any universal constant that is at least $1 + 2\delta$ would work equally well for the purposes of [40]). The above argument shows that in the setting of Theorem 12 we have

$$\forall\, \delta \in (0,1), \qquad r_q^{X\to Y}(\delta) \geqslant \exp\left(-\frac{(CKn)^q}{\delta^q}\right).$$

As we mentioned earlier in the Introduction, in Section 7 below we show that if $X = \ell_2^n$, $Y = \ell_2$ and $q = 2$ then (7) holds true with K replaced by a universal constant. This translates to the validity of (15) with K replaced by a universal constant. By reasoning as above, we therefore deduce that there exists a universal constant $\mathfrak{c} \in (1,\infty)$ for which the following lower bound holds true.

$$\forall\, \delta \in (0,1), \qquad r_2^{\ell_2^n \to \ell_2}(\delta) \geqslant \exp\left(-\frac{\mathfrak{c} n^2}{\delta^2}\right). \tag{18}$$

This improves over the bound $r_2^{\ell_2^n\to\ell_2}(\delta) \geqslant \exp(-\mathfrak{c}(n\log n)^2/\delta^2)$ of [40]. The modulus $r_2^{\ell_2^n\to\ell_2}(\cdot)$ is currently not known to have geometric applications, but it is a natural Hilbertian quantity and it would be of interest to determine its asymptotic behavior; see also Question 9 in [40].

Remark 16. It is instructive to examine the reason for the doubly exponential dependence on n in Theorem 2. Continuing with the notation and hypotheses of Theorem 12, the above reasoning proceeded in two steps. The first step deduced the L_q approximation (16) on a sub-ball of radius at least $\exp(-(CKn/\delta)^q)$. The second step argued that if $\delta = (\eta\varepsilon)^{1+n/q}$ (with $\eta \in (0,1)$ being a universal constant) then the average $(\delta\rho)$-closeness to Λ that is exhibited in (16) automatically "upgrades" to $(\varepsilon\rho)$-closeness in $L_\infty(x + \rho B_X; Y)$. This second step trivially requires δ to decay like $\varepsilon^{c_q n}$: Take for example $\Lambda = 0$ and f to be a real-valued 1-Lipschitz function that vanishes everywhere except for a ball of radius $\varepsilon\rho$ on which its maximal value equals $2\varepsilon\rho$. A substitution of an exponentially decaying δ into $r_q^{X\to Y}(\delta)$ would at best result in a doubly exponential lower bound on $r^{X\to Y}(\varepsilon)$ if $r_q^{X\to Y}(\delta)$ decays at least exponentially fast in $-1/\delta^{c_q}$. This must indeed be the case in general, since there exist examples of spaces X, Y that satisfy the assumptions of Theorem 12 yet $r^{X\to Y}(\delta) \leqslant \exp(-(c/\delta)^q)$, where $c \in (0,1)$ is a universal constant. Specifically, by [40, Lemma 16] this holds for $X = \ell_\infty^n$ and $Y = \ell_q$. So, in order to obtain a better bound in Theorem 2 it seems that one should somehow argue about L_∞ bounds directly, despite the fact that in our setting the only assumption on Y is that it has martingale cotype q, which is by [58] equivalent to an L_q Littlewood–Paley inequality. This could be viewed as an indication that perhaps Theorem 2 cannot be improved in the stated full generality, though we leave this very interesting question open.

1.E. **Littlewood–Paley–Stein theory.** An elegant and useful theorem of Martínez, Torrea and Xu [58] asserts that a Banach space $(Y, \|\cdot\|_Y)$ has martingale cotype $q \geqslant 2$ if and only if we have

$$\forall\, n \in \mathbb{N},\ \forall\, f \in L_q(\mathbb{R}^n; Y), \qquad \|\mathcal{G}_q f\|_{L_q(\mathbb{R}^n;Y)} \lesssim_{n,Y} \|f\|_{L_q(\mathbb{R}^n;Y)}, \tag{19}$$

where $\mathcal{G}_q f : \mathbb{R}^n \to Y$ is the (generalized) Littlewood–Paley–Stein $\mathcal{G}$-function which is defined by

$$\forall\, x \in \mathbb{R}^n, \qquad \mathcal{G}_q f(x) \stackrel{\mathrm{def}}{=} \left(\int_0^\infty \|t\partial_t P_t f(x)\|_Y^q \,\frac{\mathrm{d}t}{t}\right)^{\frac{1}{q}},$$

with $\{P_t\}_{t\in(0,\infty)}$ being the Poisson semigroup. More generally, the Littlewood–Paley–Stein theory of [58] applies to abstract semigroups provided that they are so-called *subordinated diffusion semigroups*, of which the heat semigroup (in contrast to Poisson) is *not* an example.

Problem 2 of [58] asks whether or not (19) holds true when Y has martingale cotype q for any diffusion semigroup in the sense of Stein [86] (with the implicit constant in (19) being allowed to also depend on the semigroup), in which case (19) would apply to the heat semigroup as well. This question remains open in full generality, with the best known partial result being due to [58], where it is shown that the answer is positive if Y is a Banach lattice of martingale cotype q. Here we obtain the following positive answer in the important special case of the heat semigroup.

Theorem 17 (Temporal Littlewood–Paley–Stein inequality for the heat semigroup)**.** *Fix $q \in [2,\infty)$ and $n \in \mathbb{N}$. Suppose that $(Y, \|\cdot\|_Y)$ is a Banach space that admits an equivalent norm with modulus of uniform convexity of power type q. Then for every $f \in L_q(\mathbb{R}^n; Y)$ we have*

$$\left(\int_0^\infty \|t\partial_t H_t f\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t}\right)^{\frac{1}{q}} \lesssim \sqrt{n} \cdot \mathfrak{m}_q(Y) \|f\|_{L_q(\mathbb{R}^n;Y)}. \tag{20}$$

By considering the direct sum of all the heat semigroups on $\mathbb{R}^n$ as n ranges over $\mathbb{N}$, a positive answer to the abstract question [58, Problem 2] would imply a dimension-independent bound in (20). However, at present it remains open whether or not the (mild) dimension dependence that appears in (20) can be removed altogether. While this is an interesting open question, it isn't relevant to the investigations of the present article because in order to prove Theorem 5 we actually need the following new result about the spatial derivatives of the heat semigroup.

Theorem 18 (Spatial Littlewood–Paley–Stein inequalities for the heat semigroup)**.** *Fix $q \in [2,\infty)$ and $n \in \mathbb{N}$. Suppose that $(Y, \|\cdot\|_Y)$ is a Banach space that admits an equivalent norm with modulus of uniform convexity of power type q. Then for every $\vec{f} \in \ell_q^n(L_q(\mathbb{R}^n; Y))$ we have*

$$\left(\int_0^\infty \left\|\sqrt{t}\,\mathrm{div}\, H_t \vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t}\right)^{\frac{1}{q}} \lesssim \sqrt{n} \cdot \mathfrak{m}_q(Y) \fint_{S^{n-1}} \left\|\sigma \cdot \vec{f}\right\|_{L_q(\mathbb{R}^n;Y)} \mathrm{d}\sigma, \tag{21}$$

where (21) uses the following (standard) notation, in which $\vec{f} = (f_1, \ldots, f_n)$.

$$H_t \vec{f} \stackrel{\mathrm{def}}{=} (H_t f_1, \ldots, H_t f_n) \qquad \text{and} \qquad \mathrm{div}\, H_t \vec{f} \stackrel{\mathrm{def}}{=} \sum_{j=1}^n \partial_j (H_t f_j). \tag{22}$$

Moreover, for every $f \in L_q(\mathbb{R}^n; Y)$ and $z \in \mathbb{R}^n$ we have

$$\left(\int_0^\infty \left\|\sqrt{t}(z \cdot \nabla) H_t f\right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t}\right)^{\frac{1}{q}} \lesssim |z| \mathfrak{m}_q(Y) \|f\|_{L_q(\mathbb{R}^n;Y)}. \tag{23}$$

We stated Theorem 17 despite the fact that only Theorem 18 is needed in our proof of Theorem 5 because it is directly related to [58, Problem 2], and moreover we shall establish Theorem 17 without much additional effort. It is well known that Littlewood–Paley–Stein inequalities as above for the time derivatives and spatial derivatives often come hand-in-hand, and indeed a version of Theorem 18 for the spatial derivatives of the Poisson semigroup was deduced in [58]. However, for abstract diffusion semigroups as in [58, Problem 2] there is no intrinsic notion of spatial derivatives.

The temporal Littlewood–Paley–Stein inequality (19) for the Poisson semigroup was previously used for geometric purposes in [50]. In that setting, the Poisson semigroup sufficed due to parabolic scaling that was afforded by the geometry of the Heisenberg group. Parabolic scaling also makes our proof of Theorem 5 go through, but this time it occurs due to our use of the heat semigroup in place of the Poisson semigroup. The only step in our proof that uses the heat semigroup and fails for the Poisson semigroup occurs in equation (42) below, which was inspired by the proof of [50, Lemma 2.5]. An inspection of that step reveals that we could have also worked with the fractional semigroup $t \mapsto \exp(-t(-\Delta)^\alpha)$ for any $\alpha \in (1/2, 1)$ in place of the heat semigroup. In this fractional (hence subordinated) setting one can prove the required version of Theorem 18 by adapting (in the spatial setting) the argument of [58], though for our purposes one needs to also take care to derive polynomial dependence on dimension, which we checked is possible but this leads to a significantly more involved and less natural argument than the one that we obtain below for the heat semigroup (Theorem 18 formally implies the corresponding statement for these fractional semigroups as well as the Poisson semigroup because these semigroups are subordinated to the heat semigroup).

1.F. **Comparison to previous work.** The previously best known bound [55] in the setting of Theorem 2 was that there exists a universal constant $C \in (0,\infty)$ such that for every integer $n \geqslant 2$ if $(X, \|\cdot\|_X)$ in an n-dimensional normed space and $(Y, \|\cdot\|_Y)$ is a Banach space whose modulus of uniform convexity is of power type q for some $q \in [2,\infty)$ then for every $\varepsilon \in (0,1/2]$ we have

$$r^{X\to Y}(\varepsilon) \geqslant \exp\left(-\frac{(Cn)^{20(n+q)}\mathfrak{m}_q(Y)^q \log\left(\frac{1}{\varepsilon}\right)}{\varepsilon^{2n+2q-2}}\right). \tag{24}$$

So, our new bound (17) is stronger than (24) both as $\varepsilon \to 0$ and as $n \to \infty$.

Under the more stringent assumption that $(Y, \|\cdot\|_Y)$ is a UMD Banach space with UMD constant $\beta = \beta(Y)$, in [40] it was shown that for every $\varepsilon \in (0,1/2]$ we have

$$r^{X\to Y}(\varepsilon) \geqslant \exp\left(-\frac{(\beta n)^{c\beta}}{\varepsilon^{c(n+\beta)}}\right), \tag{25}$$

where $c \in (0,\infty)$ is a universal constant. Since, as we recalled in the end of Section 1.B.1, there exists $2 \leqslant q \lesssim \beta$ for which $\mathfrak{m}_q(Y) \lesssim \beta^2$, the estimate (25) is weaker than our new bound (17).

No vector-valued Dorronsoro estimate was previously known for uniformly convex targets. So, Theorem 8 and Theorem 12 that are obtained here are qualitatively new statements that answer Question 8 in [40] and are definitive due to Theorem 9 and the results of Section 6 below. Previously, vector-valued Dorronsoro estimates were known only when the target Banach space Y is UMD. Specifically, by [40, Lemma 10] and [40, Theorem 19] there exists a universal constant $\kappa \in (0,\infty)$ with the following properties. Suppose that $(X, \|\cdot\|_X)$ is an n-dimensional normed space, equipped with the Hilbertian norm $|\cdot|$ that is induced by the ellipsoid of maximal volume that is contained in B_X (John position), so as to identify X with $\mathbb{R}^n$. Suppose also that $(Y, \|\cdot\|_Y)$ is a UMD Banach space with $\beta = \beta(Y)$ and $f : \mathbb{R}^n \to Y$ is 1-Lipschitz and compactly supported. Then

$$\left(\int_{\mathbb{R}^n}\int_0^\infty \inf_{\substack{\Lambda\in\mathcal{A}(X,Y)\\ \|\Lambda\|_{\mathrm{Lip}(X,Y)}\leqslant 1}} \fint_{x+tB^n} \frac{\|f(y)-\Lambda(y)\|_Y^{\kappa\beta}}{t^{\kappa\beta+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\right)^{\frac{1}{\kappa\beta}} \lesssim n^{\frac{5}{2}}\beta^{15}|\operatorname{supp}(f)|^{\frac{1}{\kappa\beta}}. \tag{26}$$

The relatively large power of the UMD constant β that occurs in the right hand side of (26) reflects the fact that the proof of (26) in [40] is quite involved, in particular using the UMD property of Y fifteen times (also through equivalent formulations of the UMD property, like the boundedness of the Y-valued Hilbert transform). In contrast, the proof of the Dorronsoro estimate (12) that we obtain here does not only address the correct generality (of all uniformly convex targets), but it also achieves this by a new and simpler argument that seems like the correct approach to the problem at hand. This proof is different from, but not any more complicated than, the existing ones for $Y = \mathbb{R}$, and is of interest even as a new route to the original result of Dorronsoro [27].

Among the approaches to Dorronsoro-type estimates that appeared in the literature, some seem to be inherently Hilbertian, such as Fefferman's identity (77) in [33] or the argument of Kristensen and Mingione in [48]. Nevertheless, this point of view yields more precise estimates (actually, identities) in the Hilbertian setting [25], where the problem of understanding quantitative differentiation remains open, with the best known bound on, say, $r^{\ell_2^n\to\ell_2}(1/4)$ currently having the same asymptotic form as the general bound of Theorem 2. Our proof of Theorem 5 is closer in spirit to the approaches of Jones [44], Seeger [85] and Triebel [87] (which are related to each other in terms of the underlying principles), though we do not see how to use these approaches to obtain a proof of Theorem 5. As we stated earlier, despite significant effort it seems that a vector-valued adaptation of Dorronsoro's original strategy [27] requires the UMD property, as in [40]. Finally, a possible direction for future research would be to investigate whether the approaches of Schul [84], Azzam–Schul [1, 2] and Li–Naor [50] could be adapted so as to yield a Dorronsoro-type estimate for uniformly convex targets, though (if at all possible) it seems that this route could at best yield a version of (12) with weaker dependence on n that is insufficient for proving Theorem 2 (a version of (12) that leads to a bound similar to (24) might be within reach through such an approach).

Our estimate (12) is also a quantitative improvement of (26) because, as we recalled in the end of Section 1.B.1, there exists a universal constant $\kappa \in (0,\infty)$ such that when Y is a UMD Banach space with $\beta = \beta(Y)$, if we set $q = \kappa\beta$ then we have $\mathfrak{m}_q(Y) \lesssim \beta^2$. Consequently, it follows formally from (12) that a variant of (26) holds true (using intrinsic averaging; see Section 1.F.1 below) with the quantity $n^{5/2}\beta(Y)^{15}$ in the right hand side replaced by the smaller quantity $n^{9/20}\sqrt[5]{\log n}\cdot\beta^2$.

1.F.1. *Intrinsic averages.* A convenient advantage of (12) over (26) is that the averaging in the left hand side of (12) occurs over the intrinsic balls $x + tB_X$ while the averaging in (26) is over the auxiliary Euclidean balls $x + tB^n$. This difference reflects a conceptual rather than technical geometric difficulty that arose in [40] and is circumvented here altogether due to the use of the heat semigroup rather than a more complicated (but natural) operator that was used in [40]. As we have seen in Section 1.D, the fact that we can work here with averages over intrinsic balls leads to a quick and direct deduction of our bound on $r^{X\to Y}(\varepsilon)$, while the fact that in (26) the averages are over Euclidean balls requires an additional argument that is carried out in [40, Section 3.1] in order to relate (26) to the uniform approximation by affine property with the stated asymptotic dependence on n. In [40] (as well as in [27, 33, 25]) the affine function Λ that is used to approximate f on the ball $x + tB^n$ is $(\mathsf{Proj}\otimes \mathrm{Id}_Y)f$, where Id_Y is the identity on Y and Proj is the orthogonal projection from $L_2(x + tB^n)$ onto its linear subspace $\mathcal{A}(X,\mathbb{R})\cap L_2(x + tB^n)$ that consists of all the restrictions to $x + uB^n$ of affine functions from $\mathbb{R}^n$ to $\mathbb{R}$. Lemma 10 of [40] asserts that $\|(\mathsf{Proj}\otimes \mathrm{Id}_Y)f\|_{\mathrm{Lip}(x+tB^n,Y)} \leqslant \|f\|_{\mathrm{Lip}(x+tB^n,Y)}$, but its proof uses the rotational symmetry of the Euclidean ball $x+tB^n$ and the analogous statement is unknown with $x+tB^n$ replaced by $x+tB_X$. The need to address such issues does not arise in our approach because the Lipschitz constant of a heat evolute can be easily bounded without any need for additional symmetries of B_X. Nevertheless, it would be independently interesting to understand Question 19 below since the operator Proj is a natural object whose Lipschitz properties are equivalent to the potential availability of certain approximate distributional symmetries of high dimensional centrally symmetric convex bodies.

Question 19. Let $(X,\|\cdot\|_X)$ be an n-dimensional normed space, equipped with a Hilbertian norm $|\cdot|$ with respect to which it is isotropic with isotropic constant L_X, i.e., $|B_X| = 1$ and (10) holds true. Consider the orthogonal projection Proj from $L_2(B_X)$ onto the subspace of affine mappings $\mathcal{A}(X,\mathbb{R})\cap L_2(B_X)$. Thus, for every $f\in L_2(B_X)$ and $x\in X$ we have

$$\mathsf{Proj}f(x) = \int_{B_X} f(z)\,\mathrm{d}z + \frac{1}{L_X^2}\sum_{j=1}^n x_j\int_{B_X} z_j f(z)\,\mathrm{d}z. \tag{27}$$

The formula (27) also makes sense when $f\in L_1(B_X;Y)$ for any Banach space $(Y,\|\cdot\|_Y)$, and we shall use the notation $\mathsf{Proj}f$ in this case as well (i.e., slightly abusing notation by identifying $\mathsf{Proj}\otimes\mathrm{Id}_Y$ with Proj). Is it true that for every Banach space Y and every Lipschitz function $f : B_X \to Y$ we have $\|\mathsf{Proj}f\|_{\mathrm{Lip}(B_X,Y)} \lesssim \|f\|_{\mathrm{Lip}(B_X,Y)}$?

In Section 5.A below we show that it suffices to treat the above question when $Y = \mathbb{R}$, i.e., the following operator norm identity holds true for every Banach space $(Y,\|\cdot\|_Y)$.

$$\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,Y)\to\mathrm{Lip}(B_X,Y)} = \|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})}. \tag{28}$$

Moreover, the quantity $\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})}$ has the following geometric interpretation.

For every $x \in X \smallsetminus \{0\}$ let μ_x^+, μ_x^- be the probability measures supported on B_X whose densities are given for every $y \in X$ by

$$\mathrm{d}\mu_x^+(y) \stackrel{\mathrm{def}}{=} \frac{\max\{x\cdot y, 0\}\mathbf{1}_{B_X}(y)}{\frac12\int_{B_X}|x\cdot z|\,\mathrm{d}z}\,\mathrm{d}y \qquad \text{and} \qquad \mathrm{d}\mu_x^-(y) \stackrel{\mathrm{def}}{=} \frac{\max\{-x\cdot y, 0\}\mathbf{1}_{B_X}(y)}{\frac12\int_{B_X}|x\cdot z|\,\mathrm{d}z}\,\mathrm{d}y. \tag{29}$$

Alternatively, if we consider the half spaces $H_x^+ \stackrel{\mathrm{def}}{=} \{y\in\mathbb{R}^n:\ x\cdot y \geqslant 0\}$, $H_x^- \stackrel{\mathrm{def}}{=} \{y\in\mathbb{R}^n:\ x\cdot y \leqslant 0\}$ then μ_x^+, μ_x^- are supported on the half balls $H_x^+\cap B_X, H_x^-\cap B_X$, respectively, and each of them has density $y\mapsto |x\cdot y|$ on the corresponding half ball. We are interested in the extent to which the convex body B_X is approximately symmetric about the hyperplane $x^\perp \subseteq \mathbb{R}^n$ in the sense that the probability measures μ_x^+, μ_x^- are distributionally close to each other. In Section 5.A we show that

$$\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})} \asymp \sup_{x\in X\smallsetminus\{0\}} \frac{|x|}{L_X\|x\|_X}\mathsf{W}_1^{\|\cdot\|_X}(\mu_x^+,\mu_x^-), \tag{30}$$

where $\mathsf{W}_1^{\|\cdot\|_X}(\cdot,\cdot)$ denotes the Wasserstein-1 (transportation cost) distance induced by $\|\cdot\|_X$. Consequently, $\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})} = O(1)$ if and only if the Wasserstein-1 distance between μ_x^+ and μ_x^- is at most a constant multiple of $L_X\|x\|_X/|x|$ for every $x\in X\smallsetminus\{0\}$. The latter statement implies that B_X has the following "Wasserstein symmetry." By [65] we have $|x|/(nL_X) \lesssim \|x\|_X \lesssim |x|/L_X$ for every $x\in X$. Hence (30) implies that if $\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})} = O(1)$ then for every $x \in X\smallsetminus\{0\}$ such that $\|x\|_X$ is not within $O(1)$ factors of the Euclidean norm $|x|/L_X$, the Wasserstein-1 distance between μ_x^+ and μ_x^- must necessarily be $o(1)$.

1.F.2. *Beyond Banach spaces.* Quantitative differentiation is a widely studied topic of importance to several mathematical disciplines, often (but not only) as a tool towards proofs of rigidity theorems. Given an appropriate (case-specific) replacement for the notation of "affine mapping," one can formulate notions of "differentiation" in many settings that do not necessarily involve linear spaces; examples of such "qualitative" metric differentiation results include [71, 46, 16, 73, 45, 51, 18, 19, 20]. Corresponding results about quantitative differentiation, which lead to refined (often quite subtle and important) rigidity results can be found in [7, 43, 67, 59, 21, 52, 74, 75, 22, 53, 17, 29, 62, 30, 31, 24, 23, 2, 54, 32]. Due to the prominence of this topic and the fact that many of the quoted results are probably not sharp, it would be of interest to develop new methods to prove sharper quantitative differentiation results. While the argument of [40] yielded the best-known bound for UMD Banach spaces, the methods of [40] relied extensively on the underlying linear structure. The present article uses the linear structure as well, but it suggests that heat flow methods may be useful for obtaining quantitative differentiation results in situations where heat flow makes sense but the underlying metric space is not a Banach space. It seems to be worthwhile to investigate whether "affine approximations" (appropriately defined) of an appropriate evolute (which is a regularized object) must be close to the initial mapping on some macroscopically large ball. We did not attempt to investigate this approach when the underlying spaces are not Banach spaces, but we believe that this is an open-ended yet worthwhile direction for future research.

2. The Lipschitz constant of heat evolutes

In this short section we shall establish an estimate on the Lipschitz constant of heat evolutes. This control will be needed later in order to deduce the localized Dorronsoro estimate of Theorem 12 from the Carleson measure estimate for the heat semigroup of Theorem 5. Throughout, we are given an n-dimensional normed space $(X,\|\cdot\|_X)$ that is also equipped with a Hilbertian norm $|\cdot|$ through which X is identified as a real vector space with $\mathbb{R}^n$. In this setting, for $p\in(0,\infty]$ let

$$M_p(X) \stackrel{\mathrm{def}}{=} \bigg(\fint_{S^{n-1}} \|\sigma\|_X^p\,\mathrm{d}\sigma\bigg)^{\frac1p}. \tag{31}$$

So, $M_\infty(X)=\sup_{\sigma\in S^{n-1}}\|\sigma\|_X$, but we shall use below the more common notation $b(X)\stackrel{\mathrm{def}}{=}M_\infty(X)$. Also, recalling the notation (5), we have $M(X)=M_1(X)$.

Lemma 20. *There exists a universal constant $C\in(0,\infty)$ with the following property. Fix $n\in\mathbb{N}$, an n-dimensional normed space $(X,\|\cdot\|_X)$, and $L\in[1,\infty)$. Let $|\cdot|$ be a Hilbertian norm on X, thus identifying X with $\mathbb{R}^n$. Suppose that $f:X\to Y$ satisfies $\|f\|_{\mathrm{Lip}(B_X,Y)}\leqslant 1$ and $\|f\|_{\mathrm{Lip}(X,Y)}\leqslant L$. Then for every $x\in B_X$ we have*

$$0<t\leqslant \frac{(1-\|x\|_X)^2}{C\left(M_1(X)\sqrt{n}+b(X)\sqrt{\log L}\right)^2}\implies \left\|\mathrm{Taylor}_x^1(H_tf)\right\|_{\mathrm{Lip}(X,Y)}\leqslant 2. \tag{32}$$

Proof. By convolving f with a smooth bump function with arbitrarily small support we may assume without loss of generality that f is smooth, in which case $\|(z\cdot\nabla)f(w)\|_Y\leqslant \mathbf{1}_{B_X}(w)+L\mathbf{1}_{\mathbb{R}^n\smallsetminus B_X}(w)$ for every $w\in\mathbb{R}^n$ and $z\in\partial B_X$. Consequently, if we let G denote a standard Gaussian vector in $\mathbb{R}^n$, i.e., the density of G is proportional to $e^{-|x|^2/2}$, then for every $z\in\partial B_X$ we have

$$\begin{aligned}\|(z\cdot\nabla)H_tf(x)\|_Y&=\left\|(z\cdot\nabla)\mathbb{E}\big[f\big(x-\sqrt{2t}G\big)\big]\right\|_Y\leqslant\mathbb{E}\big[\big\|(z\cdot\nabla)f\big(x-\sqrt{2t}G\big)\big\|_Y\big]\\&\leqslant\mathsf{Prob}\big[\big\|x-\sqrt{2t}G\big\|_X\leqslant 1\big]+L\mathsf{Prob}\big[\big\|x-\sqrt{2t}G\big\|_X\geqslant 1\big]\leqslant 1+L\mathsf{Prob}\left[\|G\|_X\geqslant\frac{1-\|x\|_X}{\sqrt{2t}}\right].\end{aligned}$$

Hence, using Markov's inequality we see that

$$\left\|\mathrm{Taylor}_x^1(H_tf)\right\|_{\mathrm{Lip}(X,Y)}=\sup_{z\in\partial B_X}\|(z\cdot\nabla)H_tf(x)\|_Y\leqslant 1+L\inf_{p\in(0,\infty)}\left(\frac{\sqrt{2t}}{1-\|x\|_X}\right)^p\mathbb{E}\big[\|G\|_X^p\big]. \tag{33}$$

For every $p\in(0,\infty)$, by integrating in polar coordinates we have

$$\mathbb{E}\big[\|G\|_X^p\big]=\frac{|B^n|}{(2\pi)^{\frac{n}{2}}}\left(\int_0^\infty nr^{n+p-1}e^{-\frac{r^2}{2}}\,\mathrm{d}r\right)\fint_{S^{n-1}}\|\sigma\|_X^p\,\mathrm{d}\sigma=\frac{2^{\frac{p}{2}}\Gamma\left(\frac{n+p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}M_p(X)^p. \tag{34}$$

Also, a theorem of Litvak, Milman and Schechtman [57] (see also [14, Theorem 5.2.4]) asserts that

$$\forall\, p\in[1,\infty],\qquad M_p(X)\asymp M(X)+\frac{\sqrt{p}}{\sqrt{n+p}}b(X). \tag{35}$$

By combining (34) with (35) and Stirling's formula we therefore see that

$$\forall\, p\in[1,\infty],\qquad \big(\mathbb{E}\big[\|G\|_X^p\big]\big)^{\frac1p}\asymp\sqrt{n+p}\left(M(X)+\frac{\sqrt{p}}{\sqrt{n+p}}b(X)\right)\asymp M(X)\sqrt{n}+b(X)\sqrt{p}. \tag{36}$$

Suppose that t satisfies the assumption that appears in (32), with $C\in(0,\infty)$ being a large enough universal constant that will be specified presently. A substitution of (36) into (33) shows that there exists a universal constant $\kappa\in(0,\infty)$ such that

$$\begin{aligned}\left\|\mathrm{Taylor}_x^1(H_tf)\right\|_{\mathrm{Lip}(X,Y)}&\leqslant 1+L\inf_{p\in(0,\infty)}\left(\frac{\kappa\sqrt{2t}\left(M(X)\sqrt{n}+b(X)\sqrt{p}\right)}{1-\|x\|_X}\right)^p\\&\leqslant 1+L\inf_{p\in(0,\infty)}\left(\frac{\kappa\sqrt{2}}{\sqrt{C}}\cdot\frac{M(X)\sqrt{n}+b(X)\sqrt{p}}{M(X)\sqrt{n}+b(X)\sqrt{\log L}}\right)^p,\end{aligned} \tag{37}$$

where in the last step of (37) we used the upper bound on t that appears in (32). Our choice of p in (37) is $p=C\left(M(X)\sqrt{n}+b(X)\sqrt{\log L}\right)^2/(8e^2\kappa^2b(X)^2)$. Since $b(X)\lesssim M(X)\sqrt{n}$ (e.g., this follows from the case $p=1$ of (35)), provided C is a large enough universal constant we have $p\geqslant 1$ and $p\geqslant nM(X)^2/b(X)^2$. This implies that $M(X)\sqrt{n}+b(X)\sqrt{p}\leqslant 2b(X)\sqrt{p}$, and therefore,

$$\left(\frac{\kappa\sqrt{2}}{\sqrt{C}}\cdot\frac{M(X)\sqrt{n}+b(X)\sqrt{p}}{M(X)\sqrt{n}+b(X)\sqrt{\log L}}\right)^p\leqslant\left(\frac{\kappa\sqrt{2}}{\sqrt{C}}\cdot\frac{2b(X)\sqrt{p}}{M(X)\sqrt{n}+b(X)\sqrt{\log L}}\right)^p=\frac{1}{e^p}\leqslant\frac{1}{L^{\frac{C}{8e^2\kappa^2}}}\leqslant\frac{1}{L},$$

where in the penultimate step we used the fact that $p\geqslant(C/8\kappa^2e^2\kappa^2)\log L$ and the final step holds true provided $C\geqslant 8e^2\kappa^2$. By (37), this concludes the proof of the desired implication (32). $\square$

3. Deduction of Theorem 5 and Theorem 12 from Theorem 18

Theorem 18, i.e., the Littlwood–Paley–Stein inequalities for the heat semigroup, will be proven in Section 4 below. In this section we will assume the validity of Theorem 18 for the moment and show how Theorem 5 and Theorem 12 follow from it. This implies our main result on quantitative differentiation, namely Theorem 2, as we explained in Section 1.D. The main step is Theorem 21 below, which asserts a statement in the spirit of Theorem 5 but with the assumption that f is Lipschitz replaced by the weaker assumption that certain Sobolev $W^{1,q}$ norms of f are finite. Similar refinements already appear in Dorronsoro's original work [27] for scalar-valued functions.

Theorem 21. *Fix $q\in[2,\infty)$ and $n\in\mathbb{N}$. Suppose that $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ are Banach spaces that satisfy $\dim(X)=n$ and $\mathfrak{m}_q(Y)<\infty$. Let $|\cdot|$ be a Hilbertian norm on X, thus identifying X with $\mathbb{R}^n$. Then for every $\gamma\in(0,\infty)$ and every smooth function $f:\mathbb{R}^n\to Y$ we have*

$$\begin{aligned}\bigg(\int_{\mathbb{R}^n}\int_0^\infty\fint_{x+tB_X}&\frac{\|f(y)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\bigg)^{\frac1q}\\&\lesssim\frac{\mathfrak{m}_q(Y)}{\sqrt{\gamma}}\bigg(\fint_{B_X}|x|^q\,\|x\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}^q\,\mathrm{d}x\bigg)^{\frac1q}+\mathfrak{m}_q(Y)\sqrt{\gamma n}\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma.\end{aligned}\tag{38}$$

Consequently, by choosing $\gamma\in(0,\infty)$ so as to minimize the right hand side of (38), if we define

$$\gamma(f)\stackrel{\mathrm{def}}{=}\frac{1}{\sqrt{n}}\cdot\frac{\Big(\fint_{B_X}|x|^q\,\|x\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}^q\,\mathrm{d}x\Big)^{\frac1q}}{\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma},$$

then

$$\begin{aligned}\bigg(\int_{\mathbb{R}^n}\int_0^\infty\fint_{x+tB_X}&\frac{\|f(y)-\mathrm{Taylor}_x^1\big(H_{\gamma(f)t^2}f\big)(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\bigg)^{\frac1q}\\&\lesssim\mathfrak{m}_q(Y)\sqrt[4]{n}\bigg(\fint_{B_X}|x|^q\,\|x\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}^q\,\mathrm{d}x\bigg)^{\frac{1}{2q}}\bigg(\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma\bigg)^{\frac12}.\end{aligned}\tag{39}$$

Proof. The validity of (39) follows from substituting (the optimal choice) $\gamma=\gamma(f)$ into (38). So, it remains to prove (38). To do so, we shall prove the following two estimates.

$$\begin{aligned}J_1&\stackrel{\mathrm{def}}{=}\bigg(\int_{\mathbb{R}^n}\int_0^\infty\fint_{B_X}\frac{\|H_{\gamma t^2}f(x+tz)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(x+tz)\|_Y^q}{t^{q+1}}\,\mathrm{d}z\,\mathrm{d}t\,\mathrm{d}x\bigg)^{\frac1q}\\&\lesssim\frac{\mathfrak{m}_q(Y)}{\sqrt{\gamma}}\bigg(\fint_{B_X}|x|^q\,\|x\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}^q\,\mathrm{d}x\bigg)^{\frac1q},\end{aligned}\tag{40}$$

and

$$J_2 \stackrel{\mathrm{def}}{=} \left(\int_0^\infty \int_{\mathbb{R}^n} \frac{\|f(y)-H_{\gamma t^2}f(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\right)^{\frac{1}{q}} \lesssim \mathfrak{m}_q(Y)\sqrt{\gamma n}\fint_{S^{n-1}} \|\sigma\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma. \tag{41}$$

Once proven, the validity of (41) and (40) would imply Theorem 21 because, by adding and subtracting $H_{\gamma t^2}f(y)$ and applying the triangle inequality in $L_q(\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}^n;Y)$, we have

$$\left(\int_{\mathbb{R}^n}\int_0^\infty \fint_{x+tB_X} \frac{\|f(y)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\right)^{\frac{1}{q}} \leqslant J_1+J_2.$$

To prove (41), observe that $\dot{H}_tf \stackrel{\mathrm{def}}{=} \partial_tH_tf=\Delta H_tf=\operatorname{div} H_t\nabla f$, and therefore

$$\begin{aligned}
J_2 &= \left(\int_0^\infty \left\|\int_0^1 \gamma t^2\dot{H}_{u\gamma t^2}f\,\mathrm{d}u\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}t}{t^{q+1}}\right)^{\frac{1}{q}}\\
&\leqslant \gamma\int_0^1\left(\int_0^\infty t^{q-1}\left\|\dot{H}_{u\gamma t^2}f\right\|_{L_q(\mathbb{R}^n;Y)}^q\,\mathrm{d}t\right)^{\frac{1}{q}}\mathrm{d}u\\
&= \sqrt{\gamma}\left(\int_0^1\frac{\mathrm{d}u}{\sqrt{u}}\right)\left(\frac12\int_0^\infty\left\|\sqrt{s}\dot{H}_sf\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}s}{s}\right)^{\frac{1}{q}}\\
&= 2^{1-\frac{1}{q}}\sqrt{\gamma}\left(\int_0^\infty\left\|\sqrt{s}\operatorname{div} H_t\nabla f\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}s}{s}\right)^{\frac{1}{q}}.
\end{aligned}\tag{42}$$

The desired estimate (41) on J_2 now follows from an application of (21) with $\vec{f}=\nabla f$.

To prove (40), observe first that by the integral representation for the error in Taylor's formula, for every $x,z\in\mathbb{R}^n$ and $t\in\mathbb{R}$ we have

$$H_{\gamma t^2}f(x+tz)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(x+tz)=\int_0^1(tz\cdot\nabla)^2H_{\gamma t^2}f(x+stz)(1-s)\,\mathrm{d}s.$$

Consequently, using Jensen's inequality we see that

$$\begin{aligned}
J_1 &\leqslant \int_0^1\left(\int_{\mathbb{R}^n}\int_0^\infty\fint_{B_X} t^{q-1}\left\|(z\cdot\nabla)^2H_{\gamma t^2}f(x+stz)\right\|_Y^q\,\mathrm{d}z\,\mathrm{d}t\,\mathrm{d}x\right)^{\frac{1}{q}}(1-s)\,\mathrm{d}s\\
&= \frac12\left(\int_0^\infty\fint_{B_X} t^{q-1}\left\|(z\cdot\nabla)^2H_{\gamma t^2}f\right\|_{L_q(\mathbb{R}^n;Y)}^q\,\mathrm{d}z\,\mathrm{d}t\right)^{\frac{1}{q}}.
\end{aligned}\tag{43}$$

Since the operators $\{z\cdot\nabla\}_{z\in\mathbb{R}^n}$ and $\{H_s\}_{s\in[0,\infty)}$ commute, for every $z\in B_X$ we have

$$\begin{aligned}
\left(\int_0^\infty t^{q-1}\left\|(z\cdot\nabla)^2H_{\gamma t^2}f\right\|_{L_q(\mathbb{R}^n;Y)}^q\,\mathrm{d}t\right)^{\frac{1}{q}} &= \frac{1}{\sqrt{\gamma}}\left(\frac12\int_0^\infty\left\|\sqrt{s}(z\cdot\nabla)H_s(z\cdot\nabla)f\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}s}{s}\right)^{\frac{1}{q}}\\
&\lesssim \frac{|z|\mathfrak{m}_q(Y)}{\sqrt{\gamma}}\left\|z\cdot\nabla f\right\|_{L_q(\mathbb{R}^n;Y)},
\end{aligned}\tag{44}$$

where in the final step of (44) we used (23) with f replaced by $(z\cdot\nabla)f$. The desired estimate (40) now follows from a substitution of (44) into (43). □

Proof of Theorem 5. By rescaling we may assume that $\|f\|_{\mathrm{Lip}(X,Y)}=1$, and by convolving f with a smooth bump function of arbitrarily small support we may assume that f is itself smooth. Then $\|z\cdot\nabla f(x)\|_Y\leqslant\|z\|_X\mathbf{1}_{\mathrm{supp}(f)}(x)$ for every $x,z\in\mathbb{R}^n$. So, $\|z\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}^q\leqslant\|z\|_X^q|\operatorname{supp}(f)|$. Hence,

$$\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma\leqslant|\operatorname{supp}(f)|^{\frac{1}{q}}\fint_{S^{n-1}}\|\sigma\|_X\,\mathrm{d}\sigma=M(X)|\operatorname{supp}(f)|^{\frac{1}{q}},$$

and

$$\left(\fint_{B_X} |x|^q \,\|x\cdot\nabla f\|_{L_q(\mathbb{R}^n;Y)}^q \,\mathrm{d}x\right)^{\frac{1}{q}} \leqslant \left(\fint_{B_X} |x|^q \,\mathrm{d}x\right)^{\frac{1}{q}} |\operatorname{supp}(f)|^{\frac{1}{q}} = I_q(X)|\operatorname{supp}(f)|^{\frac{1}{q}}.$$

A substitution of these estimates into (38) shows that for every $\gamma \in (0,\infty)$ we have

$$\left(\int_{\mathbb{R}^n}\int_0^\infty \fint_{x+tB_X} \frac{\|f(y)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\right)^{\frac{1}{q}} \lesssim \mathfrak{m}_q(Y)\left(\sqrt{\gamma n}M(X)+\frac{I_q(X)}{\sqrt{\gamma}}\right)|\operatorname{supp}(f)|^{\frac{1}{q}}. \tag{45}$$

The value of γ in (6) minimizes the right hand side of (45), thus yielding the desired estimate (7). □

Proof of Theorem 12. Suppose that $f : B_X \to Y$ satisfies $\|f\|_{\mathrm{Lip}(B_X,Y)} \leqslant 1$. Since Theorem 12 is translation-invariant, we may assume without loss of generality that $f(0) = 0$. Define $F : X \to Y$ by setting $F(x) = f(x)$ for $x \in B_X$ and $F(x) = \max\{0, n+1-n\|x\|_X\}\cdot f(x/\|x\|_X)$ for $x \in \mathbb{R}^n \smallsetminus B_X$. Then $\operatorname{supp}(F) \subseteq (1 + 1/n)B_X$ and it is straightforward to check that $\|F\|_{\mathrm{Lip}(X,Y)} \leqslant n + 2$.

Fix $x \in \mathbb{R}^n$ with $\|x\|_X \leqslant 1-1/(2n)$ and $t \in (0,\infty)$. Recalling the definition of γ in (6) and letting $C \in (0,\infty)$ be the constant of Lemma 20, we know that $\left\|\mathrm{Taylor}_x^1(H_{\gamma t^2}F)\right\|_{\mathrm{Lip}(X,Y)} \leqslant 2$ provided

$$\gamma t^2 = \frac{I_q(X)t^2}{\sqrt{n}M(X)} \leqslant \frac{1}{4Cn^2\left(M(X)\sqrt{n}+b(X)\sqrt{\log(n+2)}\right)^2}. \tag{46}$$

Since by (35) (with $p = 1$) we have $b(X) \lesssim M(X)\sqrt{n}$, there exists a universal constant $c \in (0, 1/4)$ such that the condition (46) is satisfied for every $t \in (0,T]$, where T is defined in (14). Hence, for $t \in (0,T]$ the mapping $\mathrm{Taylor}_x^1(H_{\gamma t^2}F) \in \mathcal{A}(X,Y)$ is 2-Lipschitz. Consequently,

$$\fint_r^T \left(\fint_{(1-\frac{1}{2n})B_X} \inf_{\substack{\Lambda\in\mathcal{A}(X,Y)\\ \|\Lambda\|_{\mathrm{Lip}(X,Y)}\leqslant 2}} \frac{\fint_{x+\rho B_X}\|f(y)-\Lambda(y)\|_Y^q\,\mathrm{d}y}{\rho^q}\,\mathrm{d}x\right)\frac{\mathrm{d}\rho}{\rho}$$
$$\leqslant \frac{1}{\left(1-\frac{1}{2n}\right)^n|B_X|\log\left(\frac{T}{r}\right)}\int_{(1-\frac{1}{2n})B_X}\int_r^T\fint_{x+tB_X}\frac{\|F(y)-\mathrm{Taylor}_x^1(H_{\gamma t^2}F)(y)\|_Y^q}{t^{q+1}}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x.$$

Hence, by Theorem 5 (which we have already proved assuming the validity of Theorem 18) applied to F, the left hand side of the desired inequality (15) is bounded from above by

$$\frac{K^q|\operatorname{supp}(F)|(n+2)^q}{\left(1-\frac{1}{2n}\right)^n|B_X|\log\left(\frac{T}{r}\right)} \leqslant \frac{2(3Kn)^q\left(1+\frac{1}{n}\right)^n}{\left(1-\frac{1}{2n}\right)^n|\log r|} \leqslant \frac{(9Kn)^q}{|\log r|},$$

where we used the fact that $\|F\|_{\mathrm{Lip}(X,Y)} \leqslant n+2$, that the support of F is contained in $(1+1/n)B_X$ and therefore $|\operatorname{supp}(F)| \leqslant (1+1/n)^n|B_X|$, that $r \leqslant T^2 \leqslant 1$ and therefore $\log(T/r) \geqslant |\log r|/2$, that the sequence $\{(1+1/n)^n/(1-1/(2n))^n = ((1+3/(2n-1))^{(2n-1)/3+1})^{3n/(2n+2)}\}_{n=1}^\infty$ is decreasing and therefore bounded from above by 4, and that $q \geqslant 2$. This completes the proof of Theorem 12. □

4. $\mathcal{G}$-function estimates

Here we shall prove Theorem 17 and Theorem 18. The argument naturally splits into a part that holds true for general symmetric diffusion semigroups in Section 4.A below, followed by steps that use more special properties of the heat semigroups in Section 4.B and Section 4.C below.

4.A. **Diffusion semigroups.** Following Stein [86, page 65], a symmetric diffusion semigroup on a measure space $(\mathcal{M},\mu)$ is a one-parameter family of self-adjoint linear operators $\{T_t\}_{t\in[0,\infty)}$ that map (real-valued) measurable functions on $(\mathcal{M},\mu)$ to measurable functions on $(\mathcal{M},\mu)$, such that T_0 is the identity operator and $T_{t+s}=T_tT_s$ for every $s,t\in[0,\infty)$. Moreover, it is required that for every $t\in[0,\infty)$ and $p\in[1,\infty]$ the operator T_t maps $L_p(\mu)$ to $L_p(\mu)$ with $\|T_t\|_{L_p(\mu)\to L_p(\mu)}\leqslant 1$, for every $f\in L_2(\mu)$ we have $\lim_{t\to 0}T_tf=f$ (with convergence in $L_2(\mu)$), for every nonnegative measurable function $f:\mathcal{M}\to\mathbb{R}$ the function T_tf is also nonnegative, and that $T_t\mathbf{1}_{\mathcal{M}}=\mathbf{1}_{\mathcal{M}}$.

As explained in [58, page 433], for every Banach space Y the above semigroup $\{T_t\}_{t\in[0,\infty)}$ extends to a semigroup of contractions on $L_q(\mu;Y)$ for every $q\in[1,\infty]$. This is achieved by considering the tensor product $T_t\otimes \mathrm{Id}_Y$, but in what follows it will be convenient to slightly abuse notation by identifying $T_t\otimes\mathrm{Id}_Y$ with T_t. Note that by a standard density argument for every $q\in[1,\infty)$ and $f\in L_q(\mu,Y)$ the mapping $t\mapsto T_tf$ is continuous as a mapping from $[0,\infty)$ to $L_q(\mu;Y)$.

Proposition 22. *Fix $q\in[2,\infty)$ and a Banach space $(Y,\|\cdot\|_Y)$ of martingale cotype q. Suppose that $\{T_t\}_{t\in[0,\infty)}$ is a symmetric diffusion semigroup on a measure space $(\mathcal{M},\mu)$. Then for every $f\in L_q(\mu;Y)$, if $\{t_j\}_{j\in\mathbb{Z}}\subseteq(0,\infty)$ is an increasing sequence then*

$$\Big(\sum_{j\in\mathbb{Z}}\|T_{t_j}f-T_{t_{j+1}}f\|_{L_q(\mu;Y)}^q\Big)^{\frac1q}\leqslant \mathfrak{m}_q(Y)\|f\|_{L_q(\mu;Y)}. \tag{47}$$

Proof. It suffices to prove (47) for finite sums, i.e., that for every $0<t_0<t_1<\ldots<t_N$ we have

$$\Big(\sum_{j=0}^{N}\|T_{t_j}f-T_{t_{j+1}}f\|_{L_q(\mu;Y)}^q\Big)^{\frac1q}\leqslant \mathfrak{m}_q(Y)\|f\|_{L_q(\mu;Y)}. \tag{48}$$

Since $t\mapsto T_tf$ is a continuous mapping from $[0,\infty)$ to $L_q(\mu;Y)$, we may further assume by approximation that each t_j is an integer multiple of some $\delta\in(0,\infty)$, i.e., that $t_j=k_j\delta$ with $k_j\in\mathbb{N}$.

Denoting $Q\stackrel{\mathrm{def}}{=}T_{\delta/2}$, the desired bound (48) can be rewritten as follows.

$$\Big(\sum_{j=0}^{N}\Big\|Q^{2k_j}f-Q^{2k_{j+1}}f\Big\|_{L_q(\mu;Y)}^q\Big)^{\frac1q}\leqslant \mathfrak{m}_q(Y)\|f\|_{L_q(\mu;Y)}. \tag{49}$$

The operator Q satisfies the assumptions of Rota's representation theorem [83] in the form presented by Stein [86, page 106] (see [58, Theorem 2.5] for an explanation of the vector-valued extension that is relevant to the present context), and hence its even powers admit the following representation.

$$\forall\, k\in\mathbb{N},\qquad Q^{2k}=J^{-1}\circ E'\circ E_k\circ J, \tag{50}$$

where

- $J:L_q(\mu;Y)\to L_q(\mathcal{S},\mathcal{F}',\nu;Y)\subseteq L_q(\mathcal{S},\mathcal{F},\nu;Y)$ is an isometric isomorphism for some σ-finite σ-algebras $\mathcal{F}'\subseteq\mathcal{F}$ of a measure space $(\mathcal{S},\mathcal{F},\nu)$,
- $E_k:L_q(\mathcal{S},\mathcal{F},\nu;Y)\to L_q(\mathcal{S},\mathcal{F}_k,\nu;Y)\subseteq L_p(\mathcal{S},\mathcal{F},\nu;Y)$ is the "conditional expectation" (naturally extended from a probabilistic to a σ-finite setting), where $\mathcal{F}\supseteq\mathcal{F}_1\supseteq\mathcal{F}_2\supseteq\cdots$ is a decreasing sequence of σ-finite sub-σ-algebras, and
- $E':L_q(\mathcal{S},\mathcal{F},\nu;Y)\to L_q(\mathcal{S},\mathcal{F}',\nu;Y)\subseteq L_q(\mathcal{S},\mathcal{F},\nu;Y)$ is another such "conditional expectation" for the sub-σ-algebra $\mathcal{F}'\subseteq\mathcal{F}$.

Consequently, the desired estimate (49) is proven as follows.

$$
\begin{aligned}
\sum_{j=0}^{N}\left\|(Q^{2k_j}-Q^{2k_{j+1}})f\right\|_{L_q(\mu;Y)}^q &= \sum_{j=0}^{N}\left\|J^{-1}E'(E_{k_j}-E_{k_{j+1}})Jf\right\|_{L_q(\mu;Y)}^q \\
&\leqslant \sum_{j=0}^{N}\left\|(E_{k_j}-E_{k_{j+1}})Jf\right\|_{L_q(\mathcal{S},\mathcal{F},\nu;Y)}^q \leqslant \mathfrak{m}_q(Y)^q\|Jf\|_{L_q(\mathcal{S},\mathcal{F}',\nu;Y)}^q = \mathfrak{m}_q(Y)^q\|f\|_{L_q(\mu;Y)}^q, \qquad (51)
\end{aligned}
$$

where the first step of (51) uses (50), the second step of (51) uses the fact that J^{-1} is an isometry and that E' is a contraction, the third step of (51) uses the definition of $\mathfrak{m}_q(Y)$ applied to the (reverse) martingale $\{E_{k_j}Jf)\}_{j=0}^N$, and the final step of (51) uses the fact that J is an isometry. $\square$

Remark 23. The above argument used the definition of martingale cotype q when the martingales are with respect to σ-finite measures rather than probability measures. While the traditional way to define martingale cotype q uses probability measures, as we have done in Section 1.B.1, this is equivalent (with the same constant) to the case of σ-finite measures by a general approximation result [41, Theorem 3.95]. Alternatively, one can check that the available proofs of Pisier's inequality (Theorem 4) extend effortlessly to the setting of martingales with respect to σ-finite measures.

Lemma 24. *Fix $q\in[2,\infty)$, $\alpha\in(1,\infty)$ and a Banach space $(Y,\|\cdot\|_Y)$ of martingale cotype q. Suppose that $\{T_t\}_{t\in[0,\infty)}$ is a symmetric diffusion semigroup on a measure space $(\mathcal{M},\mu)$. Then*

$$
\forall\, f\in L_q(\mu;Y),\qquad \left(\int_0^\infty \|(T_t-T_{\alpha t})f\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}t}{t}\right)^{\frac1q}\leqslant(\log\alpha)^{\frac1q}\mathfrak{m}_q(Y)\|f\|_{L_q(\mu;Y)}.
$$

Proof. The desired estimate is proven by the following computation.

$$
\begin{aligned}
\int_0^\infty \|(T_t-T_{\alpha t})f\|_{L_q(\mu;Y)}^q\frac{\mathrm{d}t}{t} &= \sum_{j\in\mathbb{Z}}\int_{\alpha^j}^{\alpha^{j+1}}\|(T_t-T_{\alpha t})f\|_{L_q(\mu;Y)}^q\frac{\mathrm{d}t}{t}\\
&=\int_1^\alpha\sum_{j\in\mathbb{Z}}\|(T_{\alpha^j t}-T_{\alpha^{j+1}t})f\|_{L_q(\mu;Y)}^q\frac{\mathrm{d}t}{t}\leqslant \mathfrak{m}_q(Y)^q\|f\|_{L_q(\mu;Y)}^q\int_1^\alpha\frac{\mathrm{d}t}{t}, \qquad (52)
\end{aligned}
$$

where the last step of (52) is an application of Proposition 22 with $t_j=\alpha^j t$. $\square$

4.B. **The spatial derivatives of the heat semigroup.** Here we shall prove Theorem 18.

Lemma 25. *Fix $q\in[1,\infty]$ and a Banach space $(Y,\|\cdot\|_Y)$. For every $\vec{f}\in\ell_q^n(L_q(\mathbb{R}^n;Y))$ we have*

$$
\left\|\sqrt{t}\operatorname{div}H_t\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\leqslant\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac n2\right)}\fint_{S^{n-1}}\left\|\sigma\cdot\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\mathrm{d}\sigma\lesssim\sqrt{n}\fint_{S^{n-1}}\left\|\sigma\cdot\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\mathrm{d}\sigma. \qquad (53)
$$

Proof. Observe that $\partial_j h_t(y)=-y_jh_t(y)/(2t)$ for every $j\in\{1,\ldots,n\}$, $y=(y_1,\ldots,y_n)\in\mathbb{R}^n$ and $t\in[0,\infty)$. Hence, for every $x\in\mathbb{R}^n$ and $t\in[0,\infty)$ we have

$$
\operatorname{div}H_t\vec{f}(x)=\sum_{j=1}^n\partial_jh_t*f_j(x)=-\frac{1}{2t}\int_{\mathbb{R}^n}h_t(y)y\cdot\vec{f}(x-y)\,\mathrm{d}y=-\frac{1}{2t}\int_{\mathbb{R}^n}h_t(y)y\cdot\vec{f}_y(x)\,\mathrm{d}y, \qquad (54)
$$

where $\vec{f}_y:\mathbb{R}^n\to\ell_q^n(L_q(\mathbb{R}^n;Y))$ is defined by setting $\vec{f}_y(x)=\vec{f}(x-y)$ for every $x,y\in\mathbb{R}^n$.

By translation invariance $\left\|y \cdot \vec{f}_y\right\|_{L_q(\mathbb{R}^n;Y)} = \left\|y \cdot \vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}$ for all $y \in \mathbb{R}^n$, so (54) implies that

$$\begin{aligned}
\left\|\sqrt{t}\,\mathrm{div}\, H_t\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)} &\leqslant \frac{1}{2\sqrt{t}}\int_{\mathbb{R}^n} h_t(y)\left\|y\cdot\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}y = \frac{1}{2}\int_{\mathbb{R}^n} h_1(z)\left\|z\cdot\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}z\\
&= \left(\frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty r^n\frac{e^{-\frac{r^2}{4}}}{(4\pi)^{\frac{n}{2}}}\,\mathrm{d}r\right)\fint_{S^{n-1}}\left\|\sigma\cdot\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\fint_{S^{n-1}}\left\|\sigma\cdot\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma,
\end{aligned}$$

where the penultimate step follows from passing to polar coordinates. This completes the proof of the first inequality in (53). The second inequality in (53) follows from Stirling's formula. □

Lemma 26. *Suppose that $q \in [1,\infty]$. For every $f \in L_q(\mathbb{R}^n;Y)$, $t \in [0,\infty)$ and $z \in \mathbb{R}^n$ we have*

$$\left\|\sqrt{t}(z\cdot\nabla)H_t f\right\|_{L_q(\mathbb{R}^n;Y)} \leqslant \frac{|z|}{\sqrt{\pi}}\|f\|_{L_q(\mathbb{R}^n;Y)}.$$

Proof. Since for every $y \in \mathbb{R}^n$ we have $(z\cdot\nabla)h_t(y) = -(z\cdot y)h_t(y)/(2t)$, every $x \in \mathbb{R}^n$ satisfies

$$(z\cdot\nabla)H_t f(x) = (z\cdot\nabla)h_t * f = -\frac{1}{2t}\int_{\mathbb{R}^n}(z\cdot y)h_t(y)f(x-y)\,\mathrm{d}y = -\frac{1}{2t}\int_{\mathbb{R}^n}(z\cdot y)h_t(y)f_y(x)\,\mathrm{d}y.$$

Consequently,

$$\begin{aligned}
\left\|\sqrt{t}(z\cdot\nabla)H_t f\right\|_{L_q(\mathbb{R}^n;Y)} &\leqslant \frac{1}{2\sqrt{t}}\int_{\mathbb{R}^n}|z\cdot y|h_t(y)\|f_y\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}y\\
&= \frac{1}{2}\|f\|_{L_q(\mathbb{R}^n;Y)}\int_{\mathbb{R}^n}|z\cdot y|h_1(y)\,\mathrm{d}y = \frac{|z|}{2}\|f\|_{L_q(\mathbb{R}^n;Y)}\int_{\mathbb{R}^n}|y_1|h_1(y)\,\mathrm{d}y = \frac{|z|}{\sqrt{\pi}}\|f\|_{L_q(\mathbb{R}^n;Y)}.
\end{aligned}$$ □

Proof of Theorem 18. Since $q \in (1,\infty)$, we have $\lim_{t\to\infty}\left\|H_t\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)^n} = 0$. Consequently,

$$H_t\vec{f} = \sum_{k=-1}^{\infty}\big(H_{2^{k+1}t} - H_{2^{k+2}t}\big)\vec{f} = \sum_{k=-1}^{\infty} H_{2^k t}\big(H_{2^k t} - H_{3\cdot 2^k t}\big)\vec{f}. \tag{55}$$

By the triangle inequality in $L_q((0,\infty),\mathrm{d}t/t;L_q(\mathbb{R}^n;Y))$, it follows from (55) that

$$\begin{aligned}
\left(\int_0^\infty\left\|\sqrt{t}\,\mathrm{div}\,H_t\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}t}{t}\right)^{\frac{1}{q}} &\leqslant \sum_{k=-1}^{\infty}\left(\int_0^\infty\left\|\sqrt{t}\,\mathrm{div}\,H_{2^k t}\big(H_{2^k t} - H_{3\cdot 2^k t}\big)\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}t}{t}\right)^{\frac{1}{q}}\\
&= \sum_{k=-1}^{\infty}\frac{1}{2^{\frac{k}{2}}}\left(\int_0^\infty\left\|\sqrt{s}\,\mathrm{div}\,H_s(H_s - H_{3s})\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}s}{s}\right)^{\frac{1}{q}} \qquad (56)\\
&\asymp \left(\int_0^\infty\left\|\sqrt{s}\,\mathrm{div}\,H_s(H_s - H_{3s})\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}s}{s}\right)^{\frac{1}{q}}, \qquad (57)
\end{aligned}$$

where in (56) we made the change of variable $s = 2^k t$ in each of the summands.

For every $s \in [0,\infty)$, an application of Lemma 25 with $\vec{f}$ replaced by $(H_s - H_{3s})\vec{f}$ shows that

$$\begin{aligned}
\left\|\sqrt{s}\,\mathrm{div}\,H_s(H_s - H_{3s})\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)} &\lesssim \sqrt{n}\fint_{S^{n-1}}\left\|\sigma\cdot(H_s - H_{3s})\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma\\
&= \sqrt{n}\fint_{S^{n-1}}\left\|(H_s - H_{3s})\left(\sigma\cdot\vec{f}\right)\right\|_{L_q(\mathbb{R}^n;Y)}\,\mathrm{d}\sigma. \qquad (58)
\end{aligned}$$

By combining (57) and (58) we therefore conclude that

$$\begin{aligned}\left(\int_0^\infty \left\|\sqrt{t}\,\mathrm{div}\, H_t\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t}\right)^{\frac1q} &\lesssim \sqrt{n}\left(\int_0^\infty \left(\fint_{S^{n-1}} \left\|(H_s-H_{3s})\left(\sigma\cdot\vec{f}\right)\right\|_{L_q(\mathbb{R}^n;Y)} \mathrm{d}\sigma\right)^q \frac{\mathrm{d}s}{s}\right)^{\frac1q}\\ &\leqslant \sqrt{n}\fint_{S^{n-1}} \left(\int_0^\infty \left\|(H_s-H_{3s})\left(\sigma\cdot\vec{f}\right)\right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}s}{s}\right)^{\frac1q}\mathrm{d}\sigma\\ &\lesssim \sqrt{n}\cdot\mathfrak{m}_q(Y)\fint_{S^{n-1}}\left\|\sigma\cdot\vec{f}\right\|_{L_q(\mathbb{R}^n;Y)}\mathrm{d}\sigma, \end{aligned}\tag{59}$$

where the penultimate step of (59) uses the triangle inequality in $L_q((0,\infty),\mathrm{d}s/s)$ and the final step of (59) uses Lemma 24. This completes the proof of (21).

To prove (23), by the triangle inequality applied to the identity (55) with $\vec{f}$ replaced by f, and making the same changes of variable as in (56), we see that

$$\begin{aligned}\left(\int_0^\infty \left\|\sqrt{t}(z\cdot\nabla)H_tf\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}t}{t}\right)^{\frac1q} &\lesssim \left(\int_0^\infty \left\|\sqrt{s}(z\cdot\nabla)H_s(H_s-H_{3s})f\right\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}s}{s}\right)^{\frac1q}\\ &\lesssim |z|\left(\int_0^\infty \|(H_s-H_{3s})f\|_{L_q(\mathbb{R}^n;Y)}^q\frac{\mathrm{d}s}{s}\right)^{\frac1q} \lesssim |z|\mathfrak{m}_q(Y)\|f\|_{L_q(\mathbb{R}^n;Y)},\end{aligned}\tag{60}$$

where the penultimate step of (60) uses Lemma 26 and the final step of (60) uses Lemma 24. □

4.C. **The time derivative of the heat semigroup.** Here we shall prove Theorem 17.

Lemma 27. *Suppose that $q\in[1,\infty]$. Then for every $f\in L_q(\mathbb{R}^n;Y)$ and $t\in[1,\infty)$ we have*

$$\left\|t\dot{H}_tf\right\|_{L_q(\mathbb{R}^n;Y)} \lesssim \sqrt{n}\cdot\|f\|_{L_q(\mathbb{R}^n;Y)}.$$

Proof. We claim that for every $t\in(0,\infty)$ we have

$$\left\|t\dot{h}_t\right\|_{L_1(\mathbb{R}^n)} = \frac{2}{\Gamma\left(\frac n2\right)}\left(\frac{n}{2e}\right)^{\frac n2} \asymp \sqrt{n}.\tag{61}$$

This would imply the desired estimate because

$$\left\|t\dot{H}_tf\right\|_{L_q(\mathbb{R}^n;Y)} = \left\|(t\dot{h}_t)*f\right\|_{L_q(\mathbb{R}^n;Y)} \leqslant \left\|t\dot{h}_t\right\|_{L_1(\mathbb{R}^n)}\cdot\|f\|_{L_q(\mathbb{R}^n;Y)} \asymp \sqrt{n}\cdot\|f\|_{L_q(\mathbb{R}^n;Y)}.$$

Verifying the validity of the identity (61) amounts to the following simple computation. By direct differentiation we see that every $x\in\mathbb{R}^n$ and $t\in(0,\infty)$ satisfy

$$\dot{h}_t(x) = -\frac{1}{2t}\left(n-\frac{|x|^2}{2t}\right)h_t(x).$$

Hence, by passing to polar coordinates we have

$$\forall\, t\in(0,\infty),\qquad \left\|t\dot{h}_t\right\|_{L_1(\mathbb{R}^n)} = \frac{\pi^{\frac n2}}{\Gamma\left(\frac n2\right)}\int_0^\infty \left|n-\frac{r^2}{2t}\right| h_t(r)r^{n-1}\,\mathrm{d}r.\tag{62}$$

It therefore remains to evaluate the integral in the right hand side of (62) as follows.

$$\begin{aligned}\int_0^\infty \left|n-\frac{r^2}{2t}\right| h_t(r)r^{n-1}\,\mathrm{d}r &= \int_0^{\sqrt{2tn}} \frac{n-\frac{r^2}{2t}}{(4\pi t)^{\frac n2}}e^{-\frac{r^2}{4t}}r^{n-1}\,\mathrm{d}r + \int_{\sqrt{2tn}}^\infty \frac{\frac{r^2}{2t}-n}{(4\pi t)^{\frac n2}}e^{-\frac{r^2}{4t}}r^{n-1}\,\mathrm{d}r\\ &= \frac{1}{(4\pi t)^{\frac n2}}\int_0^{\sqrt{2tn}}\frac{\partial}{\partial r}\left(r^ne^{-\frac{r^2}{4t}}\right)\mathrm{d}r - \frac{1}{(4\pi t)^{\frac n2}}\int_{\sqrt{2tn}}^\infty\frac{\partial}{\partial r}\left(r^ne^{-\frac{r^2}{4t}}\right)\mathrm{d}r = 2\left(\frac{n}{2\pi e}\right)^{\frac n2}.\end{aligned}$$

□

Proof of Theorem 17. The semigroup identity $H_{t+s} = H_t H_s$ implies that $\dot{H}_{t+s} = \dot{H}_t H_s$ for every $s, t \in (0,\infty)$. Hence, recalling (55), we have

$$\forall\, t \in (0,\infty), \qquad \dot{H}_t f = \sum_{k=-1}^{\infty} \left(\dot{H}_{2^{k+1}t} - \dot{H}_{2^{k+2}t} \right) f = \sum_{k=-1}^{\infty} \dot{H}_{2^k t} \left(H_{2^k t} - H_{3\cdot 2^k t} \right) f.$$

Consequently, by arguing as in (57), we conclude that

$$\left(\int_0^\infty \left\| t\dot{H}_t f \right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t} \right)^{\frac{1}{q}} \leqslant \sum_{k=-1}^{\infty} \left(\int_0^\infty \left\| t\dot{H}_{2^k t}(H_{2^k t} - H_{3\cdot 2^k t}) f \right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t} \right)^{\frac{1}{q}}$$
$$= \sum_{k=-1}^{\infty} \frac{1}{2^k} \left(\int_0^\infty \left\| s\dot{H}_s(H_s - H_{3s}) f \right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}s}{s} \right)^{\frac{1}{q}} \lesssim \left(\int_0^\infty \left\| t\dot{H}_t(H_t - H_{3t}) f \right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t} \right)^{\frac{1}{q}}.$$

It remains to note that, by an application of Lemma 27 (with f replaced by $(H_t - H_{3t})f$) followed by integration with respect to t and an application of Lemma 24, we have

$$\left(\int_0^\infty \left\| t\dot{H}_t(H_t - H_{3t}) f \right\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t} \right)^{\frac{1}{q}} \lesssim \sqrt{n} \left(\int_0^\infty \|(H_t - H_{3t}) f\|_{L_q(\mathbb{R}^n;Y)}^q \frac{\mathrm{d}t}{t} \right)^{\frac{1}{q}}$$
$$\lesssim \sqrt{n} \cdot \mathfrak{m}_q(Y) \|f\|_{L_q(\mathbb{R}^n;Y)}. \qquad \square$$

5. Auxiliary geometric estimates when X is isotropic

In this section we shall present several geometric estimates on $(X, \|\cdot\|_X)$ that include justifications of statements that were already presented in the Introduction, such as the bound (11) on $I_q(X)M(X)$ when X is isotropic. Recall that the quantities $I_q(X)$ and $M(X)$ were defined in (5). Also, recall the notation $M_p(X)$ in (31) and that we use the more common notation $b(X) = M_\infty(X)$. Some of the ensuing inequalities are elementary, while others are quite deep, since they are deduced below (in a straightforward manner) from a combination of major results in convex geometry.

Proof of (11). Recall that in the setting of (11) we are given an n-dimensional normed space $(X, \|\cdot\|_X)$ and a Hilbertian norm $|\cdot|$ on X (thus identifying X with $\mathbb{R}^n$) such that the isotropicity requirement (10) holds true. By applying (10) with the vector $y \in \mathbb{R}^n$ ranging over an orthonormal basis and summing the (squares of) the resulting identities, we see that $I_2(X) = \sqrt{n} L_X$.

Note that $I_\infty(X) = \max_{x\in B_X} |x|$ is the circumradius of B_X and $b(X) = \max_{x\in S^{n-1}} \|x\|_X$ is the reciprocal of the inradius of B_X, and therefore by [65] (see also [14, Section 3.2.1]) we have

$$I_\infty(X) \lesssim \sqrt{n} I_2(X) \qquad \text{and} \qquad b(X) \lesssim \frac{\sqrt{n}}{I_2(X)} \tag{63}$$

A theorem of Giannopoulos and E. Milman [37] asserts that

$$M(X) \lesssim \frac{(n\log n)^{\frac{2}{5}}}{I_2(X)}. \tag{64}$$

A substitution of (64) and the second inequality in (63) into the result (35) of Litvak, Milman and Schechtman that we already used earlier in the proof of Lemma 20 shows that

$$\forall\, p \in [1,\infty), \qquad M_p(X) \lesssim \left((n\log n)^{\frac{2}{5}} + \frac{\sqrt{pn}}{\sqrt{n+p}} \right) \frac{1}{I_2(X)}. \tag{65}$$

By a theorem of Paouris [72] we have

$$\forall\, q \in [2,\infty), \qquad I_q(X) \lesssim \left(1 + \frac{q\sqrt{n}}{n+q} \right) I_2(X). \tag{66}$$

Note that (66) is stated in [72, Theorem 1.2] only for the range $q \in [2,n]$, but when $q \geqslant n$ the estimate (66) becomes $I_q(X) \lesssim \sqrt{n} I_2(X)$, which follows from the first inequality in (63) since $I_q(X) \leqslant I_\infty(X)$. In conclusion, it follows from (65) and (66) that for all $p \geqslant 1$ and $q \geqslant 2$ we have

$$I_q(X)M_p(X) \lesssim \left(1+\frac{q\sqrt{n}}{n+q}\right)\left((n\log n)^{\frac{2}{5}}+\frac{\sqrt{pn}}{\sqrt{n+p}}\right). \tag{67}$$

The estimate (67) is the currently best known bound on $I_q(X)M_p(X)$ when X is isotropic, the case $p=1$ of which becomes

$$I_q(X)M(X) \lesssim \begin{cases} (n\log n)^{\frac{2}{5}} & \text{if } n \geqslant q^2, \\ \frac{q(\log q)^{\frac{2}{5}}}{\sqrt[10]{n}} & \text{if } q \leqslant n \leqslant q^2, \\ n^{\frac{9}{10}}(\log n)^{\frac{2}{5}} & \text{if } n \leqslant q. \end{cases} \tag{68}$$

The first range in the right hand side of (68) is precisely the desired estimate (11). $\square$

Remark 28. For $p \in [1,2]$, the Rademacher type p constant of a normed space $(X, \|\cdot\|_X)$, denoted $T_p(X)$, is the smallest $T \in [1,\infty)$ such that for every $k \in \mathbb{N}$ and every $x_1,\ldots,x_k \in X$ we have

$$\frac{1}{2^k}\sum_{\varepsilon\in\{-1,1\}^k}\Big\|\sum_{j=1}^k \varepsilon_j x_k\Big\|_X \leqslant T\bigg(\sum_{j=1}^k \|x_j\|_X^p\bigg)^{\frac{1}{p}}.$$

Observe that $T_2(\ell_2) \leqslant 1$ by the parallelogram identity, and therefore $T_2(X) \leqslant c_2(X) \leqslant \sqrt{\dim(X)}$.

For an isotropic n-dimensional normed space $(X, \|\cdot\|_X)$ with good control on $T_2(X)$, the estimate (11) can be improved by incorporating the work of E. Milman [64] into the above argument. Indeed, it is shown in [64] (see also [14, Theorem 9.3.3]) that $M(X) \lesssim T_2(X)/I_2(X)$. This bound, in combination with the second inequality in (63) and the estimates (35) and (66) shows that

$$\forall\, q \in [2,\infty), \qquad n \geqslant q^2 \implies I_q(X)M(X) \lesssim T_2(X).$$

More generally, for every $p,q \in [1,\infty)$ we have

$$I_q(X)M_p(X) \lesssim \left(1+\frac{q\sqrt{n}}{n+q}\right)\left(T_2(X)+\frac{\sqrt{pn}}{\sqrt{n+p}}\right).$$

In particular, $I_q(X)M(X) \lesssim \sqrt{q}\cdot T_2(X)$ for every $q \in [1,\infty)$.

Remark 29. Fix $C \geqslant 1$ and suppose that $(X, \|\cdot\|_X)$ is C-unconditional with respect to a Hilbertian norm $|\cdot|$ on X, i.e., after identification with $\mathbb{R}^n$ we have $\|(\varepsilon_1 x_1,\ldots,\varepsilon_n x_n)\|_X \leqslant C\|x\|_X$ for all $x \in \mathbb{R}^n$ and $\varepsilon_1,\ldots,\varepsilon_n \in \{-1,1\}$. We stated in the Introduction that $I_q(X)M(X) \lesssim_q C^2\sqrt{\log n}$ for every $q \in [2,\infty)$. Indeed, by a result of Milman and Pajor [65, Proposition b] we have $I_2(X) \lesssim C\sqrt{n}$. Hence, using (66) we see that $I_q(X) \lesssim C(\sqrt{n}+\min\{n,q\})$. Also, by [9, Proposition 2.5] we have $\|x\|_X \lesssim C\|x\|_{\ell_\infty^n}$ for every $x \in \mathbb{R}^n$, and therefore $M_p(X) \lesssim CM_p(\ell_\infty^n)$ for every $p \in [1,\infty)$. A standard computation (see e.g. [66, Section 5.7]) shows that $M(\ell_\infty^n) \asymp \sqrt{\log n}/\sqrt{n}$. Since $b(\ell_\infty^n)=1$, it therefore follows from (35) that $M_p(\ell_\infty^n) \asymp \sqrt{p+\log n}/\sqrt{p+n}$. The above bounds yield

$$I_q(X)M_p(X) \lesssim C^2\left(\sqrt{n}+\min\{n,q\}\right)\sqrt{\frac{p+\log n}{p+n}}.$$

In particular, when $p=1$ we see that for every $q \in [2,\infty)$ we have

$$I_q(X)M(X) \lesssim \begin{cases} C^2\sqrt{\log n} & \text{if } n \geqslant q^2, \\ \frac{C^2 q\sqrt{\log q}}{\sqrt{n}} & \text{if } q \leqslant n \leqslant q^2, \\ C^2\sqrt{n\log n} & \text{if } n \leqslant q. \end{cases}$$

This implies the desired estimate, and in fact it gives that $I_q(X)M(X) \lesssim C^2\sqrt{q\log n}$.

We next record some elementary volumetric estimates that yield lower bounds on the quantity $I_q(X)M(X)$ that were already used in the Introduction. The following lemma, whose simple proof is included for the sake of completeness, is known; see e.g. [65, Section 2] for a different derivation.

Lemma 30. *Fix $n \in \mathbb{N}$ and $q \in (0,\infty)$. Suppose that $\|\cdot\|_U$ and $\|\cdot\|_V$ are two norms on $\mathbb{R}^n$, with unit balls B_U and B_V, respectively. Then*

$$\left(\fint_{B_U} \|u\|_V^q \,\mathrm{d}u\right)^{\frac{1}{q}} \geqslant \left(\frac{n}{n+q}\right)^{\frac{1}{q}} \left(\frac{|B_U|}{|B_V|}\right)^{\frac{1}{n}}. \tag{69}$$

Proof. Fixing a Euclidean norm $|\cdot|$ on $\mathbb{R}^n$, by integrating in polar coordinates (twice) we see that

$$\begin{aligned}\int_{B_U} \left(\frac{\|u\|_U}{\|u\|_V}\right)^n \mathrm{d}u &= \int_{\mathbb{R}^n} \left(\frac{\|u\|_U}{\|u\|_V}\right)^n \mathbf{1}_{\{\|u\|_U \leqslant 1\}} \,\mathrm{d}u = |B^n| \fint_{S^{n-1}} \left(\frac{\|\sigma\|_U}{\|\sigma\|_V}\right)^n \int_0^{\frac{1}{\|\sigma\|_U}} nr^{n-1}\,\mathrm{d}r\,\mathrm{d}\sigma \\ &= |B^n| \fint_{S^{n-1}} \frac{\mathrm{d}\sigma}{\|\sigma\|_V^n} = |B^n| \fint_{S^{n-1}} \int_0^{\frac{1}{\|\sigma\|_V}} nr^{n-1}\,\mathrm{d}r\,\mathrm{d}\sigma = \int_{\mathbb{R}^n} \mathbf{1}_{\{\|v\|_V \leqslant 1\}} \,\mathrm{d}v = |B_V|. \end{aligned} \tag{70}$$

Similarly,

$$\begin{aligned}\int_{B_U} \left(\frac{\|u\|_V}{\|u\|_U}\right)^q \mathrm{d}u &= |B^n| \fint_{S^{n-1}} \left(\frac{\|\sigma\|_V}{\|\sigma\|_U}\right)^q \int_0^{\frac{1}{\|\sigma\|_U}} nr^{n-1}\,\mathrm{d}r\,\mathrm{d}\sigma = |B^n| \fint_{S^{n-1}} \frac{\|\sigma\|_V^q}{\|\sigma\|_U^{q+n}}\,\mathrm{d}\sigma \\ &= |B^n| \fint_{S^{n-1}} \|\sigma\|_V^q \int_0^{\frac{1}{\|\sigma\|_U}} (n+q) r^{n+q-1}\,\mathrm{d}r\,\mathrm{d}\sigma = \frac{n+q}{n} \int_{B_U} \|u\|_V^q \,\mathrm{d}u. \end{aligned} \tag{71}$$

Hence, by Hölder's inequality with exponents $(n+q)/n$ and $(n+q)/q$, we see that

$$\begin{aligned}|B_U| &= \int_{B_U} \left(\frac{\|u\|_V}{\|u\|_U}\right)^{\frac{nq}{n+q}} \left(\frac{\|u\|_U}{\|u\|_V}\right)^{\frac{nq}{n+q}} \mathrm{d}u \leqslant \left(\int_{B_U} \left(\frac{\|u\|_V}{\|u\|_U}\right)^q \mathrm{d}u\right)^{\frac{n}{n+q}} \left(\int_{B_U} \left(\frac{\|u\|_U}{\|u\|_V}\right)^n \mathrm{d}u\right)^{\frac{q}{n+q}} \\ &\overset{(70)\wedge(71)}{=} \left(\frac{n+q}{n} \int_{B_U} \|u\|_V^q \,\mathrm{d}u\right)^{\frac{n}{n+q}} |B_V|^{\frac{q}{n+q}}. \end{aligned} \tag{72}$$

Now, the inequality (72) simplifies to give the desired estimate (69). □

Corollary 31. *Fix $p, q \in (0,\infty)$ and $n \in \mathbb{N}$. Suppose that $(X, \|\cdot\|_X)$ is an n-dimensional normed space and that $|\cdot|$ is a Hilbertian norm on X, thus identifying X with $\mathbb{R}^n$. Then*

$$I_q(X)M_p(X) \geqslant \left(\frac{n}{n+q}\right)^{\frac{1}{q}}. \tag{73}$$

Proof. By an application of Lemma 30 with $\|\cdot\|_U = |\cdot|$ and $\|\cdot\|_V = \|\cdot\|_X$, combined with integration in polar coordinates, we see that

$$\begin{aligned}\left(\frac{n}{n+p}\right)^{\frac{1}{p}} \left(\frac{|B^n|}{|B_X|}\right)^{\frac{1}{n}} &\leqslant \left(\fint_{B^n} \|x\|_X^p \,\mathrm{d}x\right)^{\frac{1}{p}} \\ &= \left(\fint_{S^{n-1}} \int_0^1 nr^{n+p-1} \|\sigma\|_X^p \,\mathrm{d}r\,\mathrm{d}\sigma\right)^{\frac{1}{p}} = \left(\frac{n}{n+p}\right)^{\frac{1}{p}} M_p(X).\end{aligned}$$

Hence,

$$M_p(X) \geqslant \left(\frac{|B^n|}{|B_X|}\right)^{\frac{1}{n}}. \tag{74}$$

Also, another application of Lemma 30, this time with $\|\cdot\|_U = \|\cdot\|_X$ and $\|\cdot\|_V = |\cdot|$ shows that

$$I_q(X) \geqslant \left(\frac{n}{n+q}\right)^{\frac{1}{q}} \left(\frac{|B_X|}{|B^n|}\right)^{\frac{1}{n}}. \tag{75}$$

The desired lower bound (73) now follows by taking the product of (74) and (75). □

Remark 32. Fix $p, q \in [2, \infty)$. Since $\sqrt[n]{|B^n|} \asymp 1/\sqrt{n}$ and when X is isotropic we have $|B_X| = 1$, it follows from (74) that $M_p(X) \gtrsim 1/\sqrt{n}$. Also, $I_q(X) \geqslant I_2(X) = L_X\sqrt{n}$, so $I_q(X)M_p(X) \gtrsim L_X$.

5.A. **Wasserstein symmetries.** Here we shall provide justifications for statements that we made in Question 19. These issues relate to geometric questions that originated from investigations into quantitative differentiation, but are of interest in their own right. As such, the contents of this section are not needed for the purpose of proving the new results that we stated in the Introduction.

Recall that in Question 19 we are assuming that X is isotropic. For every Banach space $(Y, \|\cdot\|_Y)$ and $f \in L_1(B_X; Y)$, the definition of $\mathsf{Proj}f$ in (27) says that for every $x \in X$ we have

$$\mathsf{Proj}f(x) = \int_{B_X} f(z)\,\mathrm{d}z + \frac{1}{L_X^2}\sum_{j=1}^n x_j \int_{B_X} z_j f(z)\,\mathrm{d}z = \int_{B_X} f(z)\,\mathrm{d}z + \frac{1}{L_X^2}\int_{B_X} (x\cdot z) f(z)\,\mathrm{d}z. \tag{76}$$

Recall that if μ, ν are nonnegative Borel measures on B_X with $\mu(B_X) = \nu(B_X) < \infty$ then a coupling of μ and ν is a Borel measure π on $B_X \times B_X$ such that $\pi(A \times B_X) = \mu(A)$ and $\pi(B_X \times A) = \nu(A)$ for all Borel $A \subseteq B_X$. The set of all coupling of μ and ν is denoted $\Pi(\mu,\nu)$. Note that $\Pi(\mu,\nu) \neq \varnothing$ because μ and ν have the same total mass (specifically, $(\mu\times\nu)/\mu(B_X) \in \Pi(\mu,\nu)$). The Wasserstein-1 distance between μ and ν associated to the metric that is induced by $\|\cdot\|_X$ is

$$\mathsf{W}_1^{\|\cdot\|_X}(\mu,\nu) \stackrel{\mathrm{def}}{=} \inf_{\pi\in\Pi(\mu,\nu)} \iint_{B_X\times B_X} \|x-y\|_X\,\mathrm{d}\pi(x,y).$$

If τ is a Borel measure on X with $|\tau|(B_X) < \infty$ and $\tau(B_X) = 0$ then write

$$\|\tau\|_{\mathsf{W}_1(B_X,\|\cdot\|_X)} \stackrel{\mathrm{def}}{=} \mathsf{W}_1^{\|\cdot\|_X}(\tau^+,\tau^-), \tag{77}$$

where $\tau = \tau^+ - \tau^-$ and τ^+, τ^- are nonnegative Borel measures on B_X, which have the same total mass since we are assuming that $\tau(B_X) = 0$, so that the definition (77) makes sense. This definition turns the space of all Borel measures τ on B_X with $|\tau|(B_X) < \infty$ and $\tau(B_X) = 0$ into a Banach space, which we denote below by $\mathsf{W}_1(B_X, \|\cdot\|_X)$.

Let $\mathrm{Lip}_0(B_X, \mathbb{R})$ denote the space of all functions $f : B_X \to \mathbb{R}$ with $f(0) = 0$, equipped with the norm $\|\cdot\|_{\mathrm{Lip}(B_X,\mathbb{R})}$. By the Kantorovich–Rubinstein duality theorem (see [88, Theorem 1.14]), we have $\mathrm{Lip}_0(B_X,\mathbb{R})^* = \mathsf{W}_1(B_X, \|\cdot\|_X)$, with the identification being that a measure $\mu \in \mathsf{W}_1(B_X, \|\cdot\|_X)$ acts on a function $f \in \mathrm{Lip}_0(B_X, \mathbb{R})$ through integration, i.e., $\mu(f) = \int_{B_X} f(y)\,\mathrm{d}\mu(y)$.

Since $\mathsf{Proj}f = f$ for every constant function f, we have

$$\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})} = \|\mathsf{Proj}\|_{\mathrm{Lip}_0(B_X,\mathbb{R})\to\mathrm{Lip}_0(B_X,\mathbb{R})}.$$

Moreover, if we define an operator $T : \mathrm{Lip}_0(B_X, \mathbb{R}) \to X^*$ by setting

$$\forall\, x \in X, \qquad Tf(x) \stackrel{\mathrm{def}}{=} \frac{1}{L_X^2}\sum_{j=1}^n x_j \int_{B_X} z_j f(z)\,\mathrm{d}z,$$

then by (76) for every $f \in \mathrm{Lip}_0(B_X, \mathbb{R})$ the linear part of the affine mapping $\mathsf{Proj}f$ is precisely $\frac{1}{L_X^2}Tf$. Hence $L_X^2\|\mathsf{Proj}f\|_{\mathrm{Lip}(B_X,\mathbb{R})} = \|Tf\|_{X^*}$, and therefore

$$L_X^2\|\mathsf{Proj}\|_{\mathrm{Lip}_0(B_X,\mathbb{R})\to\mathrm{Lip}_0(B_X,\mathbb{R})} = \|T\|_{\mathrm{Lip}_0(B_X,\mathbb{R})\to X^*} = \|T^*\|_{X\to\mathsf{W}_1(B_X,\|\cdot\|_X)}.$$

One computes directly that the adjoint operator $T^*: X \to \mathrm{Lip}_0(B_X,\mathbb{R})^* = \mathsf{W}_1(B_X,\|\cdot\|_X)$ is such that T^*x is the measure whose density is $y \mapsto (x\cdot y)\mathbf{1}_{B_X}(y)$. Altogether, these observations give

$$\sup_{x\in\partial B_X} \mathsf{W}_1^{\|\cdot\|_X}\big(y\mapsto (x\cdot y)^+\mathbf{1}_{B_X}(y), y\mapsto (x\cdot y)^-\mathbf{1}_{B_X}(y)\big) = L_X^2\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to \mathrm{Lip}(B_X,\mathbb{R})}. \qquad (78)$$

Proof of (28). Since trivially $\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})} \leqslant \|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,Y)\to\mathrm{Lip}(B_X,Y)}$ for every Banach space Y with $\dim(Y)\neq 0$, the goal here is to establish the reverse inequality. Fix $\varepsilon\in(0,1)$ and $x\in X$. Let ν_x^+,ν_x^- be the measures supported on B_X whose densities are $w\mapsto (x\cdot w)^+$ and $w\mapsto (x\cdot w)^-$, respectively. By (78) there exists a coupling $\pi_x^\varepsilon\in\Pi(\nu_x^+,\nu_x^-)$ with

$$\iint_{B_X\times B_X}\|w-z\|_X\,\mathrm{d}\pi_x^\varepsilon(w,z)\leqslant L_X^2\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})}(\|x\|_X+\varepsilon). \qquad (79)$$

Suppose that $f\in \mathrm{Lip}(B_X,Y)$. Then for every $x,y\in B_X$ we have

$$\begin{aligned}\|\mathsf{Proj}f(x)-\mathsf{Proj}f(y)\|_Y &= \frac{1}{L_X^2}\bigg\|\int_{B_X} f(w)\,\mathrm{d}\nu_{x-y}^+(w)-\int_{B_X} f(w)\,\mathrm{d}\nu_{x-y}^-(w)\bigg\|_Y\\ &= \frac{1}{L_X^2}\bigg\|\iint_{B_X\times B_X}\big(f(w)-f(z)\big)\,\mathrm{d}\pi_{x-y}^\varepsilon(w,z)\bigg\|_Y \leqslant \frac{\|f\|_{\mathrm{Lip}(B_X,Y)}}{L_X^2}\iint_{B_X\times B_X}\|w-z\|_X\,\mathrm{d}\pi^\varepsilon_{x-y}(w,z).\end{aligned}$$

Hence, $\|\mathsf{Proj}f\|_{\mathrm{Lip}(B_X,Y)}\leqslant \|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})}\|f\|_{\mathrm{Lip}(B_X,Y)}$, by the above estimate combined with (79) (and letting $\varepsilon\to 0$). So, $\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,Y)\to\mathrm{Lip}(B_X,Y)}\leqslant\|\mathsf{Proj}\|_{\mathrm{Lip}(B_X,\mathbb{R})\to\mathrm{Lip}(B_X,\mathbb{R})}$. □

Proof of (30). By Borell's lemma [10] (see [14, Theorem 2.4.6]) for every $x\in X$ we have

$$\int_{B_X}|x\cdot z|\,\mathrm{d}z\asymp\bigg(\int_{B_X}(x\cdot y)^2\,\mathrm{d}y\bigg)^{\frac12}\overset{(10)}{=}L_X|x|.$$

Hence, recalling the definition of μ_x^+,μ_x^- in (29), for every $x\in X\smallsetminus\{0\}$ we have

$$\mathsf{W}_1^{\|\cdot\|_X}(\mu_x^+,\mu_x^-)\asymp\frac{\mathsf{W}_1^{\|\cdot\|_X}(y\mapsto(x\cdot y)^+\mathbf{1}_{B_X}(y),y\mapsto(x\cdot y)^-\mathbf{1}_{B_X}(y))}{L_X|x|}.$$

Therefore (30) follows from (78). □

6. From L_q affine approximation to Rademacher cotype

The notion of Rademacher type p of a Banach space was already recalled in Remark 28. Specifically, a Banach space $(Y,\|\cdot\|_Y)$ is said to have Rademacher type $p\in[1,2]$ if $T_p(Y)<\infty$, with $T_p(Y)$ as in Remark 28. A Banach space $(Y,\|\cdot\|_Y)$ is said to have Rademacher cotype $q\in[2,\infty]$ if there exists $C\in(0,\infty)$ such that for every $k\in\mathbb{N}$ and every $x_1,\ldots,x_k\in Y$ we have

$$\bigg(\sum_{j=1}^k\|x_j\|_Y^q\bigg)^{\frac1q}\leqslant\frac{C}{2^k}\sum_{\varepsilon\in\{-1,1\}^k}\bigg\|\sum_{j=1}^k\varepsilon_jx_k\bigg\|_Y.$$

The supremum over those $p\in[1,2]$ for which $(Y,\|\cdot\|_Y)$ has Rademacher type p is denoted p_Y. The infimum over those $q\in[2,\infty]$ for which $(Y,\|\cdot\|_Y)$ has Rademacher cotype q is denoted q_Y.

Recalling the notation in Remark 15, the main content of Proposition 33 below is that if $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ are Banach space with $\dim(X)<\infty$ and such that for some $K,q,Q\in[2,\infty)$ we have $r_Q^{X\to Y}(\varepsilon)\geqslant\exp(-K/\varepsilon^q)$ for every $\varepsilon\in(0,1/2]$, then necessarily $q_Y\leqslant q$. The initial idea here is to use an example that was constructed in [40], which yields a sharp upper bound on the modulus of L_Q affine approximation at ε for a certain function that takes values in $\ell_{q_Y}^m$ for some $m=m(\varepsilon)$. The Maurey–Pisier theorem [61] asserts that Y contains a copy of $\ell^m_{q_Y}$, so we can certainly embed this example of [40] into Y. However, a substantial complication occurs here because the affine approximant is now allowed to take values in Y that may fall outside the given copy of $\ell_{q_Y}^m$, thus precluding our ability to apply the impossibility result of [40] as a "black box." This would not be a problem if there existed a projection from Y onto a copy of $\ell_{q_Y}^m$ with norm $O(1)$. However, obtaining such complemented copies of $\ell_{q_Y}^m$ is not possible in general, as exhibited in a remarkable example of Pisier [79]. Maurey and Pisier considered this complementation issue in [61, Remarques 2.9], obtaining a partial result along these lines when Y has Rademacher cotype q_Y, Y^* has Rademacher type p_Y^* (i.e., q_Y and p_{Y^*} are *attained*), and $1/p_{Y^*}+1/q_Y=1$. The latter condition is satisfied in our setting since Y must be superreflexive by [6], but there is no reason for the critical Rademacher type and cotype to be attained. We overcome this by adapting the proof of [61, Remarques 2.9] so as to obtain a copy of $\ell_{q_Y}^m$ in Y on which there exists a projection whose norm is bounded by a certain function of $m=m(\varepsilon)$ that grows to ∞ sufficiently slowly so as to yield the desired result (using the bound on $m(\varepsilon)$ as a function of ε that is obtained in the *proof of* [40, Lemma 16]).

Proposition 33. *Let $(Y,\|\cdot\|_Y)$ be a Banach space such that there exist $K,q,Q\in[1,\infty)$, $n\in\mathbb{N}$ and an n-dimensional Banach space $(X,\|\cdot\|_X)$ such that $r_Q^{X\to Y}(\varepsilon)\geqslant \exp(-K/\varepsilon^q)$ for every $\varepsilon\in(0,1/2]$. Then Y is superreflexive and $q_Y\leqslant q$. Hence, if in addition Y is a Banach lattice then for every $s\in(q,\infty]$ it admits an equivalent norm whose modulus of uniform convexity is of power type s.*

Proof. The conclusion that Y is superreflexive is a consequence of the work of Bates, Johnson, Lindenstrauss, Preiss and Schechtman [6], namely Theorem 1 as stated in the Introduction. Indeed, since in the present setting the dependence of the constant K on the dimension n is irrelevant, one may assume that $X=\ell_2^n$, in which case by combining [40, Lemma 10] and [40, Lemma 13] we see that the affine mapping Λ can also be taken to satisfy $\|\Lambda\|_{\mathrm{Lip}(X,Y)}\leqslant 1$. By [40, Lemma 4] it follows that the assumption of Proposition 33 implies that $r^{X\to Y}(\varepsilon)>0$ for every $\varepsilon\in(0,1/2]$, and therefore Y is superreflexive by Theorem 1. The stronger conclusion when Y is also a Banach lattice, i.e., that in this case for every $s\in(q,\infty]$ it admits an equivalent norm whose modulus of uniform convexity is of power type s, is a formal consequence of the (yet to be proven) conclusion $q_Y\leqslant q$, by well-known structural results for Banach latices [35, 34] (see also [56, Section 1.f]).

Due to these comments, the proof of Proposition 33 will be complete if we show that $q_Y\leqslant q$. To that end, by the Maurey–Pisier theorem [61] for every $M\in\mathbb{N}$ there exist $y_1,\ldots,y_M\in Y$ such that

$$\forall\, a=(a_1,\ldots,a_M)\in\mathbb{R}^M,\qquad \Big(\sum_{j=1}^M |a_j|^{q_Y}\Big)^{\frac{1}{q_Y}}\leqslant \Big\|\sum_{j=1}^M a_jy_j\Big\|_Y\leqslant 2\Big(\sum_{j=1}^M |a_j|^{q_Y}\Big)^{\frac{1}{q_Y}}. \tag{80}$$

In particular, $\|y_j\|_Y\geqslant 1$ for all $j\in\{1,\ldots,M\}$, so by Hahn–Bananach there exist $y_1^*,\ldots,y_M^*\in B_{Y^*}$ such that $y_k^*(y_j)=\delta_{kj}$ for every $k,j\in\{1,\ldots,M\}$. Consequently, for every $a\in\mathbb{R}^M$ we have

$$\Big\|\sum_{k=1}^M a_ky_k^*\Big\|_{Y^*}\geqslant \frac{\left(\sum_{k=1}^M a_ky_k^*\right)\left(\sum_{j=1}^M \mathrm{sign}(a_j)|a_j|^{\frac{1}{q_Y-1}}y_j\right)}{\left\|\sum_{j=1}^M \mathrm{sign}(a_j)|a_j|^{\frac{1}{q_Y-1}}y_j\right\|_Y}\overset{(80)}{\geqslant}\frac12\Big(\sum_{j=1}^M |a_j|^{\frac{q_Y}{q_Y-1}}\Big)^{\frac{q_Y-1}{q_Y}}. \tag{81}$$

Fix $m\in\mathbb{N}$. It follows from (81) that $\|y_j^*-y_k^*\|_{Y^*}\geqslant 2^{-1/q_Y}$, so by [61, Lemme 1.5] (which itself uses and important construction of Brunel and Sucheston [15]), provided M is large enough (as a function of m), there exist $k_1,\ldots,k_{2m}\in\{1,\ldots,M\}$ with $k_1<k_2<\ldots<k_{2m}$ such that the vectors $\{y_{k_{2j}}^*-y_{k_{2j-1}}^*\}_{j=1}^m$ are a 3-unconditional basic sequence in Y^*, i.e.,

$$\forall(b,\varepsilon)\in\mathbb{R}^m\times\{-1,1\}^m,\qquad \Big\|\sum_{j=1}^m b_j(y_{k_{2j}}^*-y_{k_{2j-1}}^*)\Big\|_{Y^*}\leqslant 3\Big\|\sum_{j=1}^m \varepsilon_jb_j(y_{k_{2j}}^*-y_{k_{2j-1}}^*)\Big\|_{Y^*}. \tag{82}$$

Since we have already shown that Y is superreflexive, by the results of [36] and [76] we have $p_Y>1$, and therefore by Pisier's K-convexity theorem [78] we have $p_{Y^*}=q_Y/(q_Y-1)$. Hence

$T_p(Y^*)<\infty$ for every $p\in[1,q_Y/(q_Y-1))$, which implies that for every $b_1,\ldots,b_m\in\mathbb{R}$ we have

$$\begin{aligned}\Big\|\sum_{j=1}^m b_j(y^*_{k_{2j}}-y^*_{k_{2j-1}})\Big\|_{Y^*}&\stackrel{(82)}{\leqslant}\frac{3}{2^m}\sum_{\varepsilon\in\{-1,1\}^m}\Big\|\sum_{j=1}^m\varepsilon_j b_j(y^*_{k_{2j}}-y^*_{k_{2j-1}})\Big\|_{Y^*}\\ &\leqslant 3T_p(Y^*)\bigg(\sum_{j=1}^m|b_j|^p\Big\|y^*_{k_{2j}}-y^*_{k_{2j-1}}\Big\|_{Y^*}^p\bigg)^{\frac1p}\leqslant 6T_p(Y^*)m^{\frac1p+\frac{1}{q_Y}-1}\bigg(\sum_{j=1}^m|b_j|^{\frac{q_Y}{q_Y-1}}\bigg)^{\frac{q_Y-1}{q_Y}}. \end{aligned}\qquad(83)$$

Consider the subspace $W=\mathrm{span}\{y_{k_2},y_{k_4},\ldots,y_{k_{2m}}\}\subseteq Y$. Let $S:\mathbb{R}^m\to W$ be defined by setting $S(a_1,\ldots,a_m)=\sum_{j=1}^m a_jy_{k_{2j}}$. Then due to (80) we know that $\|a\|_{\ell_{q_Y}^m}\leqslant\|Sa\|_Y\leqslant 2\|a\|_{\ell_{q_Y}^m}$ for every $a\in\mathbb{R}^m$. Next, consider the linear operator $P:Y\to W$ that is defined by

$$\forall\, y\in Y,\qquad Py\stackrel{\mathrm{def}}{=}\sum_{j=1}^m\Big(y^*_{k_{2j}}(y)-y^*_{k_{2j-1}}(y)\Big)\,y_{k_{2j}}.\qquad(84)$$

Then P is a projection onto W. We claim that P satisfies the following operator norm bound.

$$\|P\|_{Y\to W}\leqslant 12T_p(Y^*)m^{\frac1p+\frac{1}{q_Y}-1}\|y\|_Y.\qquad(85)$$

Indeed, fixing $y\in Y$, if we define

$$\forall\, j\in\{1,\ldots,m\},\qquad b_j(y)\stackrel{\mathrm{def}}{=}\frac{\big|y^*_{k_{2j}}(y)-y^*_{k_{2j-1}}(y)\big|^{q_Y-1}\,\mathrm{sign}\,\big(y^*_{k_{2j}}(y)-y^*_{k_{2j-1}}(y)\big)}{\Big(\sum_{j=1}^n\big|y^*_{k_{2j}}(y)-y^*_{k_{2j-1}}(y)\big|^{q_Y}\Big)^{1-\frac{1}{q_Y}}}\in\mathbb{R},\qquad(86)$$

i.e., $(b_j(y))_{j=1}^m\in B^m_{\ell_{q_Y/(q_Y-1)}}$ is the normalizing functional of $(y^*_{k_{2j}}(y)-y^*_{k_{2j-1}}(y))_{j=1}^m\in\ell_{q_Y}^m$, then

$$\begin{aligned}\|Py\|_Y&\stackrel{(80)}{\leqslant}2\bigg(\sum_{j=1}^n\Big|y^*_{k_{2j}}(y)-y^*_{k_{2j-1}}(y)\Big|^{q_Y}\bigg)^{\frac{1}{q_Y}}\stackrel{(86)}{=}2\bigg(\sum_{j=1}^m b_j(y)(y^*_{k_{2j}}-y^*_{k_{2j-1}})\bigg)(y)\\ &\leqslant 2\Big\|\sum_{j=1}^m b_j(y)(y^*_{k_{2j}}-y^*_{k_{2j-1}})\Big\|_{Y^*}\|y\|_Y\stackrel{(83)\wedge(86)}{\leqslant}12T_p(Y^*)m^{\frac1p+\frac{1}{q_Y}-1}\|y\|_Y,\end{aligned}$$

thus establishing the validity of (85).

By [40, Lemma 16] there is a universal constant $\eta\in(0,1)$ with the following property. For every $m\in\mathbb{N}$ there exists a function $\phi^m:\mathbb{R}\to\ell_{q_Y}^m$ with $\|\phi^m\|_{\mathrm{Lip}(\mathbb{R},\ell_{q_Y}^m)}\leqslant 1$ such that for every $Q\in[1,\infty]$ and every affine mapping $\Lambda:\mathbb{R}\to\ell_{q_Y}^m$, if $a,b\in[-1,1]$ satisfy $a\leqslant b$ and $b-a\geqslant 4/2^m$ then

$$\bigg(\frac{1}{b-a}\int_a^b\|\phi^m(x)-\Lambda(x)\|_{\ell_{q_Y}^m}^Q\,\mathrm{d}x\bigg)^{\frac1Q}\geqslant\frac{\eta}{m^{\frac{1}{q_Y}}}\cdot\frac{b-a}{2}.\qquad(87)$$

We note that the above assertion does not appear in the statement of Lemma 16 of [40] but it is stated explicitly in its (short) proof. In what follows it will be convenient to denote the coordinates of ϕ^m by $\phi_1^m,\ldots,\phi_m^m:\mathbb{R}\to\mathbb{R}$, thus $\phi^m(x)=(\phi_1^m(x),\ldots,\phi_m^m(x))\in\mathbb{R}^m$ for every $x\in\mathbb{R}$.

By John's theorem [42] we can identify X (as a real vector space) with $\mathbb{R}^n$ so that for every $x\in X$ we have $\|x\|_{\ell_\infty^n}\leqslant\|x\|_X\leqslant n\|x\|_{\ell_\infty^n}$. (By [38] the factor of n here can be improved to $O(n^{5/6})$, but this is not important in the present context.) Define $f^m:\mathbb{R}^n\to Y$ by

$$\forall\, x\in\mathbb{R}^n,\qquad f^m(x)\stackrel{\mathrm{def}}{=}\frac12 S\circ\phi^m(x_1)=\frac12\sum_{j=1}^m\phi_j^m(x_1)y_{k_{2j}}.\qquad(88)$$

Thus $f^m(x)$ depends only on the first coordinate of x. Since $\|\phi^m\|_{\mathrm{Lip}(\mathbb{R},\ell_{q_Y}^m)} \leqslant 1$ and $\|\cdot\|_{\ell_\infty^n} \leqslant \|\cdot\|_X$, by (80) we have $\|f^m\|_{\mathrm{Lip}(X,Y)} \leqslant 1$. So, by our underlying assumption that $r_Q^{X\to Y}(\varepsilon) \geqslant \exp(-K/\varepsilon^q)$ for every $\varepsilon \in (0,1/2]$, there exists a radius $\rho \in (0,1)$ with

$$\rho \geqslant \exp\left(-\frac{K}{\eta^q}5^{2q}T_p(Y^*)^q n^{\frac{(n+Q)q}{Q}} m^{\frac{q}{p}+\frac{2q}{q_Y}-q}\right), \tag{89}$$

a point $x \in B_X$ with $x+\rho B_X \subseteq B_X$, and an affine mapping $\Lambda : \mathbb{R}^n \to Y$ such that

$$\left(\fint_{x+\rho B_X} \|f^m(y)-\Lambda(y)\|_Y^Q \,\mathrm{d}y\right)^{\frac{1}{Q}} \leqslant \frac{\eta}{25T_p(Y^*)n^{\frac{n}{Q}+1}m^{\frac{1}{p}+\frac{2}{q_Y}-1}}\rho. \tag{90}$$

Since $\|\cdot\|_{\ell_\infty^n} \leqslant \|\cdot\|_X \leqslant n\|\cdot\|_{\ell_\infty^n}$ we have $x+\frac{\rho}{n}[-1,1]^n \subseteq x+\rho B_X \subseteq x+\rho[-1,1]^n$. So, by (90),

$$\left(\fint_{x+\frac{\rho}{n}[-1,1]^n} \|f^m(y)-\Lambda(y)\|_Y^Q \,\mathrm{d}y\right)^{\frac{1}{Q}} \leqslant \frac{\eta\rho}{25T_p(Y^*)nm^{\frac{1}{p}+\frac{2}{q_Y}-1}}. \tag{91}$$

We claim that (91) implies that necessarily $\rho/n < 1/2^{m-1}$. Once proven, this assertion may be contrasted with (89) to deduce that

$$\forall (m,q) \in \mathbb{N}\times\left[1,\frac{q_Y}{q_Y-1}\right), \qquad 2^{m-1} < n\exp\left(\frac{K}{\eta^q}5^{2q}T_p(Y^*)^q n^{\frac{(n+Q)q}{Q}} m^{\frac{q}{p}+\frac{2q}{q_Y}-q}\right). \tag{92}$$

By letting $m\to\infty$ in (92) we see that $q/p+2q/q_Y-q \geqslant 1$. By letting $p \to q_Y/(q_Y-1)$, we conclude that $q(q_Y-1)/q_Y+2q/q_Y-q \geqslant 1$, which simplifies to give the desired estimate $q_Y \leqslant q$.

It therefore remains to prove that $\rho/n < 1/2^{m-1}$. To this end, assume for the sake of obtaining a contradiction that $2\rho/n \geqslant 4/2^m$. Since $x+\rho B_X \subseteq B_X$ we have $\|x\|_{\ell_\infty^n} \leqslant \|x\|_X \leqslant (1-\rho)$. Consequently, $x_1-\rho/n, x_1+\rho/n \in [-1,1]$. Hence, by an application of (87) with $a = x_1-\rho/n$ and $b = x_1+\rho/n$ (so that our contrapositive assumption implies that indeed $b-a \geqslant 4/2^m$), for every fixed $(y_2,\ldots,y_n) \in (x_2,\ldots,x_n)+[-\rho/n,\rho/n]^{n-1}$ we may consider the affine function $(y_1\in\mathbb{R}) \mapsto S^{-1}\circ P\circ\Lambda(y_1,y_2,\ldots,y_n) \in \ell_{q_Y}^m$ to deduce that

$$\fint_{x_1-\frac{\rho}{n}}^{x_1+\frac{\rho}{n}} \left\|\phi^m(y_1)-S^{-1}\circ P\circ\Lambda(y_1,y_2,\ldots,y_n)\right\|_{\ell_{q_Y}^m}^Q \,\mathrm{d}y_1 \geqslant \left(\frac{\eta\rho}{nm^{\frac{1}{q_Y}}}\right)^Q. \tag{93}$$

By averaging (93) over $(y_2,\ldots,y_n) \in (x_2,\ldots,x_n)+[-\rho/n,\rho/n]^{n-1}$, we therefore have

$$\begin{aligned}\frac{\eta\rho}{nm^{\frac{1}{q_Y}}} &\leqslant \left(\fint_{x+\frac{\rho}{n}[-1,1]^n} \left\|\phi^m(y_1)-S^{-1}\circ P\circ\Lambda(y)\right\|_{\ell_{q_Y}^m}^Q \,\mathrm{d}y\right)^{\frac{1}{Q}}\\ &= 2\left(\fint_{x+\frac{\rho}{n}[-1,1]^n} \left\|S^{-1}\circ P\big(f^m(y)-\Lambda(y)\big)\right\|_{\ell_{q_Y}^m}^Q \,\mathrm{d}y\right)^{\frac{1}{Q}} \qquad (94)\\ &\leqslant 24T_p(Y^*)m^{\frac{1}{p}+\frac{1}{q_Y}-1}\left(\fint_{x+\frac{\rho}{n}[-1,1]^n} \|f^m(y)-\Lambda(y)\|_Y^Q \,\mathrm{d}y\right)^{\frac{1}{Q}}, \qquad (95)\end{aligned}$$

where in (94) we used the definition of f^m in (88) and the fact that $Pf^m = f^m$ (since f^m takes values in the subspace W and P is a projection onto W), and in (95) we used the norm bound (85). The desired contradiction now follows by contrasting (91) with (95). $\square$

The assumption (13) of Theorem 9 implies a local Dorronsoro inequality as in Theorem 12, which in turn implies an estimate of the form $r_q^{X\to Y}(\varepsilon) \geqslant \exp(-K/\varepsilon^q)$ as in Proposition 33. In these implications the constants deteriorate, but for the purpose of Theorem 9 and Proposition 33 constants are not important (all that matters is that they are independent of ε). So, due to Proposition 33 we have proven Theorem 9 and the statement in the paragraph that follows Question 10.

7. Explicit computations for real-valued mappings on Euclidean space

For concreteness, below we shall fix the following normalization for the Fourier transform on $\mathbb{R}^n$.

$$\forall\, f\in L_1(\mathbb{R}^n),\ \forall \xi\in\mathbb{R}^n,\qquad \widehat{f}(\xi)\stackrel{\mathrm{def}}{=}\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} f(x)e^{-ix\cdot\xi}\,\mathrm{d}x.$$

Thus, by the Plancherel theorem, every smooth compactly supported function $f:\mathbb{R}^n\to\mathbb{R}$ satisfies

$$\left(\int_{\mathbb{R}^n}|\xi|^2\cdot\left|\widehat{f}(\xi)\right|^2\mathrm{d}\xi\right)^{\frac12}=\left(\int_{\mathbb{R}^n}|\nabla f(x)|^2\,\mathrm{d}x\right)^{\frac12}=\sqrt{n}\left(\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_2(\mathbb{R}^n)}^2\,\mathrm{d}\sigma\right)^{\frac12}. \tag{96}$$

Also, $\widehat{P_tf}(\xi)=e^{-t|\xi|}\widehat{f}(\xi)$ and $\widehat{H_tf}(\xi)=e^{-t|\xi|^2}\widehat{f}(\xi)$ for every $t\in[0,\infty)$, $\xi\in\mathbb{R}^n$ and $f\in L_1(\mathbb{R}^n)$.

The heat semigroup. Fix a smooth compactly supported function $f:\mathbb{R}^n\to\mathbb{R}$ and a parameter $\gamma\in(0,\infty)$. For every $t\in(0,\infty)$ and $z\in\mathbb{R}^n$ the Fourier transform of the function

$$(x\in\mathbb{R}^n)\mapsto f(x+tz)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(x+tz)=f(x+tz)-H_{\gamma t^2}f(x)-tz\cdot\nabla H_{\gamma t^2}f(x)$$

is given by

$$(\xi\in\mathbb{R}^n)\mapsto e^{itz\cdot\xi}\widehat{f}(\xi)-\left(e^{-\gamma t^2|\xi|^2}\widehat{f}(\xi)+itz\cdot\xi e^{-\gamma t^2|\xi|^2}\widehat{f}(\xi)\right)=\left(e^{itz\cdot\xi}-(1+itz\cdot\xi)e^{-\gamma t^2|\xi|^2}\right)\widehat{f}(\xi).$$

By the Plancherel theorem we therefore have

$$\int_{\mathbb{R}^n}\left|f(x+tz)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(x+tz)\right|^2\,\mathrm{d}x=\int_{\mathbb{R}^n}\left|e^{itz\cdot\xi}-(1+itz\cdot\xi)e^{-\gamma t^2|\xi|^2}\right|^2\cdot\left|\widehat{f}(\xi)\right|^2\,\mathrm{d}\xi. \tag{97}$$

By rotation invariance the following identity holds true for every $\xi\in\mathbb{R}^n$ and $t\in(0,\infty)$.

$$\begin{aligned}\fint_{B^n}\left|e^{itz\cdot\xi}-(1+itz\cdot\xi)e^{-\gamma t^2|\xi|^2}\right|^2\,\mathrm{d}z&=\fint_{B^n}\left|e^{it|\xi|z_1}-(1+it|\xi|z_1)e^{-\gamma t^2|\xi|^2}\right|^2\,\mathrm{d}z\\&=\frac{|B^{n-1}|}{|B^n|}\int_{-1}^1\left|e^{it|\xi|u}-(1+it|\xi|u)e^{-\gamma t^2|\xi|^2}\right|^2(1-u^2)^{\frac{n-1}{2}}\,\mathrm{d}u.\end{aligned}$$

Hence, using the change of variable $s=t|\xi|$ and a substitution of the values of $|B^{n-1}|$ and $|B^n|$,

$$\begin{aligned}&\int_0^\infty\fint_{B^n}\left|e^{itz\cdot\xi}-(1+itz\cdot\xi)e^{-\gamma t^2|\xi|^2}\right|^2\,\mathrm{d}z\frac{\mathrm{d}t}{t^3}\\&\qquad=\sqrt{\pi}\cdot\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n+1}{2}\right)}|\xi|^2\int_0^\infty\int_{-1}^1\left|e^{isu}-(1+isu)e^{-\gamma s^2}\right|^2(1-u^2)^{\frac{n-1}{2}}\,\mathrm{d}u\frac{\mathrm{d}s}{s^3}.\end{aligned} \tag{98}$$

Consequently, if we introduce the notation

$$\begin{aligned}\kappa(n,\gamma)&\stackrel{\mathrm{def}}{=}n\sqrt{\pi}\cdot\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n+1}{2}\right)}\int_0^\infty\int_{-1}^1\left|e^{isu}-(1+isu)e^{-\gamma s^2}\right|^2(1-u^2)^{\frac{n-1}{2}}\,\mathrm{d}u\frac{\mathrm{d}s}{s^3}\\&\asymp n^{\frac32}\int_0^1\int_0^\infty\left(\left(\cos(su)-e^{-\gamma s^2}\right)^2+\left(\sin(su)-sue^{-\gamma s^2}\right)^2\right)\frac{(1-u^2)^{\frac{n-1}{2}}}{s^3}\,\mathrm{d}s\,\mathrm{d}u,\end{aligned} \tag{99}$$

then a combination of (96), (97) and (98) implies the validity of the following identity.

$$\left(\int_{\mathbb{R}^n}\int_0^\infty \fint_{x+tB^n} \frac{\big(f(y)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(y)\big)^2}{t^3}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x\right)^{\frac12}$$
$$=\sqrt{\frac{\hbar(n,\gamma)}{n}}\bigg(\int_{\mathbb{R}^n}|\xi|^2\cdot\Big|\widehat{f}(\xi)\Big|^2\,\mathrm{d}\xi\bigg)^{\frac12}=\sqrt{\hbar(n,\gamma)}\bigg(\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_2(\mathbb{R}^n)}^2\,\mathrm{d}\sigma\bigg)^{\frac12}. \quad (100)$$

Despite the fact that (100) is stated for real-valued mappings, the corresponding identity automatically holds true for Hilbert-space valued mappings as well by an application of (100) to each of the coordinates with respect to an orthonormal basis, i.e., if $\mathcal{H}$ is a Hilbert space then for every $n\in\mathbb{N}$, every $\gamma\in(0,\infty)$ and every smooth compactly supported $f:\mathbb{R}^n\to\mathcal{H}$ we have

$$\left(\int_{\mathbb{R}^n}\int_0^\infty \fint_{x+tB^n} \big\|f(y)-\mathrm{Taylor}_x^1(H_{\gamma t^2}f)(y)\big\|_{\mathcal{H}}^2\,\mathrm{d}y\frac{\mathrm{d}t}{t^3}\,\mathrm{d}x\right)^{\frac12}$$
$$=\sqrt{\hbar(n,\gamma)}\bigg(\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_2(\mathbb{R}^n;\mathcal{H})}^2\,\mathrm{d}\sigma\bigg)^{\frac12}. \quad (101)$$

The following lemma contains a (sharp) upper bound on the quantity $\hbar(n,\gamma)$.

Lemma 34. *For every $n\in\mathbb{N}$ and $\gamma\in(0,\infty)$ we have*

$$\hbar(n,\gamma)\lesssim \gamma n+\int_0^\infty v^2e^{-v^2}\log\left(2+\frac{v^2+\gamma n}{v\sqrt{\gamma n}}\right)\mathrm{d}v. \quad (102)$$

Prior to proving Lemma 34, we record the following corollary (corresponding to a substitution of the special case $\gamma=1/n$ of Lemma 34 into (101)) that was already stated in the Introduction, where it was noted that it implies the improved estimate (18) on the modulus of L_2 affine approximation.

Corollary 35. *Suppose that $\mathcal{H}$ is a Hilbert space. Then for every $n\in\mathbb{N}$, every smooth compactly supported function $f:\mathbb{R}^n\to\mathcal{H}$ satisfies*

$$\left(\int_{\mathbb{R}^n}\int_0^\infty \fint_{x+tB^n} \Big\|f(y)-\mathrm{Taylor}_x^1\Big(H_{\frac{t^2}{n}}f\Big)(y)\Big\|_{\mathcal{H}}^2\,\mathrm{d}y\frac{\mathrm{d}t}{t^3}\,\mathrm{d}x\right)^{\frac12}\lesssim\bigg(\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_2(\mathbb{R}^n;\mathcal{H})}^2\,\mathrm{d}\sigma\bigg)^{\frac12}.$$

Proof of Lemma 34. We shall estimate the two integrals that correspond to each of the summands that appear in the right hand side of (99) separately. Firstly, consider the elementary estimate

$$\forall\, a,b\in[0,\infty),\qquad \Big|\cos(a)-e^{-b}\Big|\leqslant|\cos(a)-1|+\Big|1-e^{-b}\Big|\asymp\min\{a^2,1\}+\min\{b,1\},$$

which implies that for every $u\in(0,\infty)$ we have

$$\int_0^\infty\Big(\cos(su)-e^{-\gamma s^2}\Big)^2\frac{\mathrm{d}s}{s^3}\lesssim u^4\int_0^{\frac1u}s\,\mathrm{d}s+\int_{\frac1u}^\infty\frac{\mathrm{d}s}{s^3}+\gamma^2\int_0^{\frac{1}{\sqrt{\gamma}}}s\,\mathrm{d}s+\int_{\frac{1}{\sqrt{\gamma}}}^\infty\frac{\mathrm{d}s}{s^3}\asymp u^2+\gamma.$$

Therefore,

$$n^{\frac32}\int_0^1\int_0^\infty\Big(\cos(su)-e^{-\gamma s^2}\Big)^2\frac{(1-u^2)^{\frac{n-1}{2}}}{s^3}\,\mathrm{d}s\,\mathrm{d}u\lesssim n^{\frac32}\int_0^1(u^2+\gamma)e^{-\frac{n-1}{2}u^2}\,\mathrm{d}u\asymp 1+\gamma n. \quad (103)$$

Secondly, consider the elementary estimate

$$\forall (a,b)\in[0,1]\times[0,\infty),\qquad \Big|\sin(a)-ae^{-b}\Big|\leqslant|\sin(a)-a|+a\Big|1-e^{-b}\Big|\asymp a^2+a\min\{1,b\},$$

which implies that for every $u\in(0,1]$ we have

$$\int_0^{\frac{1}{u}}\left(\sin(su)-sue^{-\gamma s^2}\right)^2\frac{\mathrm{d}s}{s^3}\lesssim u^4\int_0^{\frac{1}{u}}s\,\mathrm{d}s+u^2\int_0^{\frac{1}{u}}\frac{\min\{1,\gamma^2s^4\}}{s}\,\mathrm{d}s$$
$$\asymp u^2+\min\left\{\frac{\gamma^2}{u^2},u^2\right\}+u^2\log\left(\max\left\{1,\frac{\sqrt{\gamma}}{u}\right\}\right)\asymp u^2\log\left(2+\frac{\sqrt{\gamma}}{u}\right).$$

Consequently, if $n\geqslant 2$ then using the elementary inequality $(1-u^2)^{(n-1)/2}\leqslant e^{-nu^2/4}$ we see that

$$n^{\frac{3}{2}}\int_0^1\int_0^{\frac{1}{u}}\left(\sin(su)-sue^{-\gamma s^2}\right)^2\frac{(1-u^2)^{\frac{n-1}{2}}}{s^3}\,\mathrm{d}s\,\mathrm{d}u\lesssim n^{\frac{3}{2}}\int_0^1u^2e^{-\frac{nu^2}{4}}\log\left(2+\frac{\sqrt{\gamma}}{u}\right)\mathrm{d}u$$
$$=8\int_0^{\frac{\sqrt{n}}{2}}v^2e^{-v^2}\log\left(2+\frac{\sqrt{\gamma n}}{2v}\right)\mathrm{d}v\lesssim\int_0^\infty v^2e^{-v^2}\log\left(2+\frac{\sqrt{\gamma n}}{2v}\right)\mathrm{d}v.\qquad(104)$$

When $n=1$ the leftmost term in (104) is bounded from above by a universal constant, and therefore it is bounded above by a constant multiple of the rightmost term in (104) in the case $n=1$ as well.

In a similar fashion, consider the elementary estimate

$$\forall(a,b)\in[1,\infty)\times[0,\infty),\qquad\left|\sin(a)-ae^{-b}\right|\leqslant|\sin(a)|+ae^{-b}\leqslant 1+ae^{-b},$$

which implies that for every $u\in(0,1]$ we have

$$\int_{\frac{1}{u}}^\infty\left(\sin(su)-sue^{-\gamma s^2}\right)^2\frac{\mathrm{d}s}{s^3}\lesssim\int_{\frac{1}{u}}^\infty\frac{\mathrm{d}s}{s^3}+u^2\int_{\frac{1}{u}}^\infty\frac{e^{-2\gamma s^2}}{s}\,\mathrm{d}s$$
$$\asymp u^2+u^2\int_{\frac{\sqrt{2\gamma}}{u}}^\infty\frac{e^{-t^2}}{t}\,\mathrm{d}t\lesssim u^2+u^2\int_{\frac{\sqrt{2\gamma}}{u}}^{\max\left\{\frac{\sqrt{2\gamma}}{u},1\right\}}\frac{\mathrm{d}t}{t}+u^2\int_1^\infty e^{-t^2}\,\mathrm{d}t\lesssim u^2\log\left(2+\frac{u}{\sqrt{\gamma}}\right).$$

By integrating this inequality with respect to u, when $n\geqslant 2$ we therefore have

$$n^{\frac{3}{2}}\int_0^1\int_{\frac{1}{u}}^\infty\left(\sin(su)-sue^{-\gamma s^2}\right)^2\frac{(1-u^2)^{\frac{n-1}{2}}}{s^3}\,\mathrm{d}s\,\mathrm{d}u\lesssim n^{\frac{3}{2}}\int_0^1u^2e^{-\frac{nu^2}{4}}\log\left(2+\frac{u}{\sqrt{\gamma}}\right)\mathrm{d}u$$
$$=8\int_0^{\frac{\sqrt{n}}{2}}v^2e^{-v^2}\log\left(2+\frac{2v}{\sqrt{\gamma n}}\right)\mathrm{d}v\lesssim\int_0^\infty v^2e^{-v^2}\log\left(2+\frac{2v}{\sqrt{\gamma n}}\right)\mathrm{d}v.\qquad(105)$$

As before, the leftmost term in (105) is bounded by a universal constant multiple of the rightmost term of (105) in the case $n=1$ as well.

By summing (104) and (105) while using the fact that $\log[(2+a)(2+1/a)]\asymp\log(2+(a^2+1)/a)$ for every $a\in(0,\infty)$, we conclude that

$$n^{\frac{3}{2}}\int_0^1\int_0^\infty\left(\sin(su)-sue^{-\gamma s^2}\right)^2\frac{(1-u^2)^{\frac{n-1}{2}}}{s^3}\,\mathrm{d}s\,\mathrm{d}u\lesssim\int_0^\infty v^2e^{-v^2}\log\left(2+\frac{v^2+\gamma n}{v\sqrt{\gamma n}}\right)\mathrm{d}v.\qquad(106)$$

Recalling (99), the desired estimate (102) now follows from (103) and (106). □

The Poisson semigroup. Fix $\gamma\in(0,\infty)$ and a nonconstant smooth compactly supported function $f:\mathbb{R}^n\to\mathbb{R}$. Arguing analogously to (97), by the Plancherel theorem we have

$$\int_{\mathbb{R}^n}\left|f(x+tz)-\mathrm{Taylor}_x^1(P_{\gamma t}f)(x+tz)\right|^2\mathrm{d}x=\int_{\mathbb{R}^n}\left|e^{itz\cdot\xi}-(1+itz\cdot\xi)e^{-\gamma t|\xi|}\right|^2\cdot\left|\widehat{f}(\xi)\right|^2\mathrm{d}\xi.$$

From here, the same reasoning that led to the identity (100) shows that

$$\int_{\mathbb{R}^n}\int_0^\infty \fint_{x+tB^n} \frac{\big(f(y)-\mathrm{Taylor}_x^1(P_{\gamma t}f)(y)\big)^2}{t^3}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x$$
$$= c_n\left(\int_0^\infty\int_{-1}^1 \left|e^{isu}-(1+isu)e^{-\gamma s}\right|^2(1-u^2)^{\frac{n-1}{2}}\,\mathrm{d}u\frac{\mathrm{d}s}{s^3}\right)\fint_{S^{n-1}}\|\sigma\cdot\nabla f\|_{L_2(\mathbb{R}^n)}^2\,\mathrm{d}\sigma, \quad (107)$$

where $c_n = n\sqrt{\pi}\Gamma(1+n/2)/\Gamma((n+1)/2)$. But, for fixed $u\in(-1,1)$ when $s\to 0$ the integrand of the first integral in the right hand side of (107) is asymptotic to $(1-u^2)^{(n-1)/2}\gamma^2/s$. So, the first integral in the right hand side of (107) diverges, implying that the left hand side of (107) is infinite.

Acknowledgements. We thank Apostolos Giannopoulos and Gilles Pisier for providing pointers to the literature. We also thank the three anonymous referees for their feedback and corrections.

References

[1] J. Azzam and R. Schul. Hard Sard: quantitative implicit function and extension theorems for Lipschitz maps. *Geom. Funct. Anal.*, 22(5):1062–1123, 2012.

[2] J. Azzam and R. Schul. A quantitative metric differentiation theorem. *Proc. Amer. Math. Soc.*, 142(4):1351–1357, 2014.

[3] K. Ball. Logarithmically concave functions and sections of convex sets in $\mathbf{R}^n$. *Studia Math.*, 88(1):69–84, 1988.

[4] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[5] K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.

[6] S. Bates, W. B. Johnson, J. Lindenstrauss, D. Preiss, and G. Schechtman. Affine approximation of Lipschitz functions and nonlinear quotients. *Geom. Funct. Anal.*, 9(6):1092–1127, 1999.

[7] Y. Benyamini. The uniform classification of Banach spaces. In *Texas functional analysis seminar 1984–1985 (Austin, Tex.)*, Longhorn Notes, pages 15–38. Univ. Texas Press, Austin, TX, 1985.

[8] Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

[9] S. G. Bobkov and F. L. Nazarov. On convex bodies and log-concave probability measures with unconditional basis. In *Geometric aspects of functional analysis*, volume 1807 of *Lecture Notes in Math.*, pages 53–69. Springer, Berlin, 2003.

[10] C. Borell. Convex measures on locally convex spaces. *Ark. Mat.*, 12:239–252, 1974.

[11] J. Bourgain. Some remarks on Banach spaces in which martingale difference sequences are unconditional. *Ark. Mat.*, 21(2):163–168, 1983.

[12] J. Bourgain. On high-dimensional maximal functions associated to convex bodies. *Amer. J. Math.*, 108(6):1467–1476, 1986.

[13] J. Bourgain. Remarks on the extension of Lipschitz maps defined on discrete sets and uniform homeomorphisms. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 157–167. Springer, Berlin, 1987.

[14] S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou. *Geometry of isotropic convex bodies*, volume 196 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2014.

[15] A. Brunel and L. Sucheston. On J-convexity and some ergodic super-properties of Banach spaces. *Trans. Amer. Math. Soc.*, 204:79–90, 1975.

[16] J. Cheeger. Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.*, 9(3):428–517, 1999.

[17] J. Cheeger. Quantitative differentiation: a general formulation. *Comm. Pure Appl. Math.*, 65(12):1641–1670, 2012.

[18] J. Cheeger and B. Kleiner. On the differentiability of Lipschitz maps from metric measure spaces to Banach spaces. In *Inspired by S. S. Chern*, volume 11 of *Nankai Tracts Math.*, pages 129–152. World Sci. Publ., Hackensack, NJ, 2006.

[19] J. Cheeger and B. Kleiner. Differentiating maps into L^1, and the geometry of BV functions. *Ann. of Math. (2)*, 171(2):1347–1385, 2010.

[20] J. Cheeger and B. Kleiner. Metric differentiation, monotonicity and maps to L^1. *Invent. Math.*, 182(2):335–370, 2010.

[21] J. Cheeger, B. Kleiner, and A. Naor. A $(\log n)^{\Omega(1)}$ integrality gap for the sparsest cut SDP. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009)*, pages 555–564. IEEE Computer Soc., Los Alamitos, CA, 2009.

[22] J. Cheeger, B. Kleiner, and A. Naor. Compression bounds for Lipschitz maps from the Heisenberg group to L_1. *Acta Math.*, 207(2):291–373, 2011.

[23] J. Cheeger and A. Naber. Lower bounds on Ricci curvature and quantitative behavior of singular sets. *Invent. Math.*, 191(2):321–339, 2013.

[24] J. Cheeger and A. Naber. Quantitative stratification and the regularity of harmonic maps and minimal currents. *Comm. Pure Appl. Math.*, 66(6):965–990, 2013.

[25] K. Danailov and C. Fefferman. A note on quantitative differentiation. Forthcoming manuscript, 2016.

[26] G. David and S. Semmes. *Analysis of and on uniformly rectifiable sets*, volume 38 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1993.

[27] J. R. Dorronsoro. A characterization of potential spaces. *Proc. Amer. Math. Soc.*, 95(1):21–31, 1985.

[28] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961.

[29] A. Eskin, D. Fisher, and K. Whyte. Coarse differentiation of quasi-isometries I: Spaces not quasi-isometric to Cayley graphs. *Ann. of Math. (2)*, 176(1):221–260, 2012.

[30] A. Eskin, D. Fisher, and K. Whyte. Coarse differentiation of quasi-isometries II: Rigidity for Sol and lamplighter groups. *Ann. of Math. (2)*, 177(3):869–910, 2013.

[31] A. Eskin, H. Masur, and K. Rafi. Large scale rank of Teichmüller space. Preprint, available at `http://arxiv.org/abs/1307.3733`, 2013.

[32] A. Eskin, H. Masur, and K. Rafi. Rigidity of of Teichmüller space. Preprint, available at `http://arxiv.org/abs/1506.04774`, 2015.

[33] C. Fefferman. The N-body problem in quantum mechanics. *Comm. Pure Appl. Math.*, 39(S, suppl.):S67–S109, 1986. Frontiers of the mathematical sciences: 1985 (New York, 1985).

[34] T. Figiel. On the moduli of convexity and smoothness. *Studia Math.*, 56(2):121–155, 1976.

[35] T. Figiel and W. B. Johnson. A uniformly convex Banach space which contains no l_p. *Compositio Math.*, 29:179–190, 1974.

[36] T. Figiel and G. Pisier. Séries aléatoires dans les espaces uniformément convexes ou uniformément lisses. *C. R. Acad. Sci. Paris Sér. A*, 279:611–614, 1974.

[37] A. Giannopoulos and E. Milman. M-estimates for isotropic convex bodies and their L_q-centroid bodies. In *Geometric aspects of functional analysis*, volume 2116 of *Lecture Notes in Math.*, pages 159–182. Springer, Cham, 2014.

[38] A. A. Giannopoulos. A note on the Banach-Mazur distance to the cube. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 67–73. Birkhäuser, Basel, 1995.

[39] O. Giladi, A. Naor, and G. Schechtman. Bourgain's discretization theorem. *Ann. Fac. Sci. Toulouse Math. (6)*, 21(4):817–837, 2012.

[40] T. Hytönen, S. Li, and A. Naor. Quantitative affine approximation for UMD targets. *Discrete Analysis*, 2016(6):1–37, 2016.

[41] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach spaces. Volume I. Martingales and Littlewood-Paley theory*, volume 63 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer, Cham, 2016.

[42] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and essays presented to R. Courant on his 60th birthday*, pages 187–204. Interscience Publishers, Inc., 1948.

[43] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. Banach spaces determined by their uniform structures. *Geom. Funct. Anal.*, 6(3):430–470, 1996.

[44] P. W. Jones. Square functions, Cauchy integrals, analytic capacity, and harmonic measure. In *Harmonic analysis and partial differential equations (El Escorial, 1987)*, volume 1384 of *Lecture Notes in Math.*, pages 24–68. Springer, Berlin, 1989.

[45] S. Keith. A differentiable structure for metric measure spaces. *Adv. Math.*, 183(2):271–315, 2004.

[46] B. Kirchheim. Rectifiable metric spaces: local structure and regularity of the Hausdorff measure. *Proc. Amer. Math. Soc.*, 121(1):113–123, 1994.

[47] B. Klartag. On convex perturbations with a bounded isotropic constant. *Geom. Funct. Anal.*, 16(6):1274–1290, 2006.

[48] J. Kristensen and G. Mingione. The singular set of Lipschitzian minima of multiple integrals. *Arch. Ration. Mech. Anal.*, 184(2):341–369, 2007.

[49] T. Kuusi, G. Mingione, and Y. Sire. A fractional Gehring lemma, with applications to nonlocal equations. *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.*, 25(4):345–358, 2014.

[50] V. Lafforgue and A. Naor. Vertical versus horizontal Poincaré inequalities on the Heisenberg group. *Israel J. Math.*, 203(1):309–339, 2014.
[51] J. R. Lee and A. Naor. L_p metrics on the Heisenberg group and the Goemans-Linial conjecture. In *Proceedings of 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 99–108, 2006.
[52] J. R. Lee and P. Raghavendra. Coarse differentiation and multi-flows in planar graphs. *Discrete Comput. Geom.*, 43(2):346–362, 2010.
[53] J. R. Lee and A. Sidiropoulos. Near-optimal distortion bounds for embedding doubling spaces into L_1 [extended abstract]. In *STOC'11—Proceedings of the 43rd ACM Symposium on Theory of Computing*, pages 765–772. ACM, New York, 2011.
[54] S. Li. Coarse differentiation and quantitative nonembeddability for Carnot groups. *J. Funct. Anal.*, 266(7):4616–4704, 2014.
[55] S. Li and A. Naor. Discretization and affine approximation in high dimensions. *Israel J. Math.*, 197(1):107–129, 2013.
[56] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin-New York, 1979. Function spaces.
[57] A. E. Litvak, V. D. Milman, and G. Schechtman. Averages of norms and quasi-norms. *Math. Ann.*, 312(1):95–124, 1998.
[58] T. Martínez, J. L. Torrea, and Q. Xu. Vector-valued Littlewood-Paley-Stein theory for semigroups. *Adv. Math.*, 203(2):430–475, 2006.
[59] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel J. Math.*, 114:221–237, 1999.
[60] B. Maurey. Système de Haar. In *Séminaire Maurey-Schwartz 1974–1975: Espaces L^p, applications radonifiantes et géométrie des espaces de Banach, Exp. Nos. I et II*, pages 26 pp. (erratum, p. 1). Centre Math., École Polytech., Paris, 1975.
[61] B. Maurey and G. Pisier. Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. *Studia Math.*, 58(1):45–90, 1976.
[62] M. Mendel and A. Naor. Markov convexity and local rigidity of distorted metrics. *J. Eur. Math. Soc. (JEMS)*, 15(1):287–337, 2013.
[63] M. Mendel and A. Naor. Nonlinear spectral calculus and super-expanders. *Publ. Math. Inst. Hautes Études Sci.*, 119:1–95, 2014.
[64] E. Milman. Dual mixed volumes and the slicing problem. *Adv. Math.*, 207(2):566–598, 2006.
[65] V. D. Milman and A. Pajor. Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 64–104. Springer, Berlin, 1989.
[66] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.
[67] A. Naor. Geometric problems in non-linear functional analysis. Master's thesis, Hebrew University, 1998.
[68] A. Naor and G. Schechtman. Planar earthmover is not in L_1. *SIAM J. Comput.*, 37(3):804–826 (electronic), 2007.
[69] A. Naor and G. Schechtman. Metric X_p inequalities. *Forum Math. Pi*, 4:e3, 81 pp., 2016.
[70] M. I. Ostrovskii. *Metric embeddings*, volume 49 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2013. Bilipschitz and coarse embeddings into Banach spaces.
[71] P. Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math. (2)*, 129(1):1–60, 1989.
[72] G. Paouris. Concentration of mass on convex bodies. *Geom. Funct. Anal.*, 16(5):1021–1049, 2006.
[73] S. D. Pauls. The large scale geometry of nilpotent Lie groups. *Comm. Anal. Geom.*, 9(5):951–982, 2001.
[74] I. Peng. Coarse differentiation and quasi-isometries of a class of solvable Lie groups I. *Geom. Topol.*, 15(4):1883–1925, 2011.
[75] I. Peng. Coarse differentiation and quasi-isometries of a class of solvable Lie groups II. *Geom. Topol.*, 15(4):1927–1981, 2011.
[76] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.
[77] G. Pisier. Un exemple concernant la super-réflexivité. In *Séminaire Maurey-Schwartz 1974–1975: Espaces L^p applications radonifiantes et géométrie des espaces de Banach, Annexe No. 2*, page 12. Centre Math. École Polytech., Paris, 1975.
[78] G. Pisier. Holomorphic semigroups and the geometry of Banach spaces. *Ann. of Math. (2)*, 115(2):375–392, 1982.
[79] G. Pisier. Counterexamples to a conjecture of Grothendieck. *Acta Math.*, 151(3-4):181–208, 1983.

[80] 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.
[81] Y. Qiu. On the UMD constants for a class of iterated $L_p(L_q)$ spaces. *J. Funct. Anal.*, 263(8):2409–2429, 2012.
[82] M. Ribe. On uniformly homeomorphic normed spaces. *Ark. Mat.*, 14(2):237–244, 1976.
[83] G.-C. Rota. An "Alternierende Verfahren" for general positive operators. *Bull. Amer. Math. Soc.*, 68:95–102, 1962.
[84] R. Schul. Bi-Lipschitz decomposition of Lipschitz functions into a metric space. *Rev. Mat. Iberoam.*, 25(2):521–531, 2009.
[85] A. Seeger. A note on Triebel-Lizorkin spaces. In *Approximation and function spaces (Warsaw, 1986)*, volume 22 of *Banach Center Publ.*, pages 391–400. PWN, Warsaw, 1989.
[86] E. M. Stein. *Topics in harmonic analysis related to the Littlewood-Paley theory.* Annals of Mathematics Studies, No. 63. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1970.
[87] H. Triebel. Local approximation spaces. *Z. Anal. Anwendungen*, 8(3):261–288, 1989.
[88] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
[89] Q. Xu. Littlewood-Paley theory for functions with values in uniformly convex spaces. *J. Reine Angew. Math.*, 504:195–226, 1998.

(T.H.) University of Helsinki, Department of Mathematics and Statistics, P.O.B. 68 (Gustaf Hällströmin katu 2b), FI-00014 Helsinki, Finland
E-mail address: tuomas.hytonen@helsinki.fi

(A.N.) Princeton University, Department of Mathematics, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA
E-mail address: naor@math.princeton.edu