.

SPECTRAL CALCULUS AND LIPSCHITZ EXTENSION FOR BARYCENTRIC METRIC SPACES

MANOR MENDEL AND ASSAF NAOR

Dedicated to Nigel Kalton

ABSTRACT. The metric Markov cotype of barycentric metric spaces is computed, yielding the first class of metric spaces that are not Banach spaces for which this bi-Lipschitz invariant is understood. It is shown that this leads to new nonlinear spectral calculus inequalities, as well as a unified framework for Lipschitz extension, including new Lipschitz extension results for $CAT(0)$ targets. An example that elucidates the relation between metric Markov cotype and Rademacher cotype is analyzed, showing that a classical Lipschitz extension theorem of Johnson, Lindenstrauss and Benyamini is asymptotically sharp.

1. INTRODUCTION

Our main purpose here is to compute a bi-Lipschitz invariant, called *metric Markov cotype*, for *barycentric metric spaces*; an important class of metric spaces that contains all uniformly convex Banach spaces as well as complete simply connected metric spaces that are nonpositively curved in the sense of Aleksandrov.

The notion of metric Markov cotype arises from the deep work [3] of K. Ball on the Lipschitz extension problem. Based mainly on Ball's ideas in [3], combined with some additional geometric ingredients, we establish a fully nonlinear version of Ball's extension theorem that allows for targets that are not necessarily Banach spaces. Due to our computation of metric Markov cotype for barycentric spaces, this yields a versatile Lipschitz extension theorem that contains as special cases many Lipschitz extension theorems that appeared in the literature, as well as Lipschitz extension results that were previously unknown.

Another use of metric Markov cotype is due to [43], where it is shown to yield spectral calculus inequalities for nonlinear spectral gaps. Consequently, our computation of metric Markov cotype for barycentric metric spaces implies new nonlinear spectral calculus inequalities which, in the special case of $CAT(0)$ spaces, lay the groundwork for our forthcoming construction [44] of expanders with respect to certain Hadamard spaces and random graphs.

Finally, we show that a beautiful construction of Kalton [29] yields a closed linear subspace X of L_1 (thus in particular X has Rademacher cotype 2) that fails to have finite metric Markov cotype. By obtaining a quantitative version of Kalton's result, we show that a classical Lipschitz extension theorem of Johnson, Lindenstrauss and Benyamini [23] is asymptotically sharp.

In order to give precise formulations of the above results one needs to recall some background. This will be done in the subsequent sections that contain a detailed description of the contents of this paper.

1.1. **Markov type and metric Markov cotype.** Given $n \in \mathbb{N}$ and $\pi \in \Delta^{n-1} \stackrel{\mathrm{def}}{=} \{x \in [0,1]^n : \sum_{i=1}^n x_i = 1\}$, recall that a stochastic matrix $A = (a_{ij}) \in M_n(\mathbb{R})$ (here and in what follows, $M_n(\mathbb{R})$ denotes as usual the n by n matrices with real entries) is said to be reversible relative to the probability vector π if $\pi_i a_{ij} = \pi_j a_{ji}$ for all $i, j \in \{1, \ldots, n\}$. The following important definition is due to K. Ball [3].

Definition 1.1 (Markov type p)**.** A metric space (X, d_X) is a said to have Markov type $p \in (0, \infty)$ with constant $M \in (0, \infty)$ if for every $n, t \in \mathbb{N}$ and every $\pi \in \Delta^{n-1}$, if $A = (a_{ij}) \in M_n(\mathbb{R})$ is a stochastic matrix that is reversible relative to π then every $x_1, \ldots, x_n \in X$ satisfy

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i (A^t)_{ij} d_X(x_i, x_j)^p \leqslant M^p t \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i a_{ij} d_X(x_i, x_j)^p. \quad (1)$$

The infimum over those $M \in (0, \infty)$ satisfying (1) is denoted $M_p(X)$.

The triangle inequality implies that $M_1(X) = 1$ for every metric space (X, d_X), and Ball proved in [3] that $M_p(\ell_p) = 1$ for $p \in [1, 2]$. In [50] it is shown that $M_2(\ell_p) \lesssim \sqrt{p}$ for $p \in [2, \infty)$ (here, and in what follows, $A \lesssim B$ and $B \gtrsim A$ denotes the estimate $A \leqslant CB$ for some absolute constant $C \in (0, \infty)$). Additional examples of computations of Markov type will be discussed in Section 1.5.

Markov type is a bi-Lipschitz invariant that has proved itself useful to a variety problems in metric geometry, one of which will be recalled below. We refer to [3] for the natural probabilistic interpretation of (1) that explains the above terminology (this interpretation is not needed in the present paper, but it is important elsewhere).

Definition 1.2 (metric Markov cotype p)**.** A metric space (X, d_X) is said to have metric Markov cotype $p \in (0, \infty)$ with constant $N \in (0, \infty)$ if for every $n, t \in \mathbb{N}$ and every $\pi \in \Delta^{n-1}$, if $A = (a_{ij}) \in M_n(\mathbb{R})$ is a stochastic matrix that is reversible relative to π then for every

$x_1, \ldots, x_n \in X$ there exist $y_1, \ldots, y_n \in X$ satisfying

$$\sum_{i=1}^{n} \pi_i d_X(x_i, y_i)^p + t \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i a_{ij} d_X(y_i, y_j)^p \leqslant N^p \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i \left(\frac{1}{t} \sum_{s=1}^{t} A^s \right)_{ij} d_X(x_i, x_j)^p. \quad (2)$$

The infimum over those $N \in (0, \infty)$ satisfying (1) is denoted $N_p(X)$.

Definition 1.2 is taken from [43]. In [3] Ball suggested a seemingly different notion of Markov cotype, but it is in fact equivalent to Definition 1.2, as explained in Section 7. Due to applications of (2) that will be described later, we believe that it is beneficial to work with the above definition of metric Markov cotype rather than Ball's original formulation. See Section 7 for a description of Ball's approach.

Condition (2) originates from an attempt to introduce an invariant that is "dual" to Markov type by reversing the inequality in (1). However, no non-singleton metric space can satisfy (1) with the direction of the inequality reversed (this follows formally from observations in [51] and [42], and can be also easily verified directly). (2) achieves a similar reversal of (1) by allowing one to pass from the initial points $x_1, \ldots, x_n \in X$ to new points $y_1, \ldots, y_n \in X$. The first summand in the left hand side of (2) ensures that on average (with respect to π) y_i is close to x_i. The remaining terms in (2) correspond to the reversal of (1), with $\{x_i\}_{i=1}^n$ replaced by $\{y_i\}_{i=1}^n$ in the left hand side, and the power A^t replaced by the Cesàro average $\frac{1}{t} \sum_{s=1}^{t} A^s$.

Due to [3, 43], Banach spaces that admit an equivalent norm whose modulus of convexity has power type p have metric Markov cotype p, in particular $N_p(\ell_p) \lesssim 1$ for $p \in [2, \infty)$ and $N_2(\ell_p) \lesssim 1/\sqrt{p-1}$ for $p \in (1, 2]$. Prior to the present work this was the only nontrivial class of metric spaces whose metric Markov cotype was known. Here we enrich the repertoire of metric spaces for which one can prove a metric Markov cotype inequality such as (2), treating also spaces that are not necessarily Banach spaces.

1.2. **Barycentric metric spaces.** In order to avoid measurability considerations that are irrelevant to the discussion at hand, we will tacitly assume throughout this article that all measures are finitely supported and all σ-algebras are finite.

The set of probability measures on a set X is denoted $\mathscr{P}_X$. Denoting the point mass at $x \in X$ by $\delta_x \in \mathscr{P}_X$, every $\mu \in \mathscr{P}_X$ can be written uniquely as $\mu = \sum_{i=1}^{n} \lambda_i \delta_{x_i}$ for some $n \in \mathbb{N}$, distinct $x_1, \ldots, x_n \in X$

and $(\lambda_1,\ldots,\lambda_n)\in\Delta^{n-1}\cap(0,1]^n$. A coupling of $\mu,\nu\in\mathscr{P}_X$ is a measure $\pi\in\mathscr{P}_{X\times X}$ such that $\sum_{z\in X}\pi(x,z)=\mu(x)$ and $\sum_{z\in X}\pi(z,y)=\nu(y)$ for every $x,y\in X$ (both of these sums are finite). The set of all the couplings of μ and ν is denoted $\Pi(\mu,\nu)\subseteq\mathscr{P}_{X\times X}$. If (X,d_X) is a metric space and $p\in[1,\infty)$ then the corresponding Wasserstein p metric on $\mathscr{P}_X$ is defined as usual by

$$\forall\,\mu,\nu\in\mathscr{P}_X,\qquad W_p(\mu,\nu)\stackrel{\mathrm{def}}{=}\inf_{\pi\in\Pi(\mu,\nu)}\left(\int_{X\times X}d_X(x,y)^p d\pi(x,y)\right)^{1/p}.$$

Definition 1.3 (W_p barycentric metric space)**.** Fix $p,\Gamma\in[1,\infty)$. A metric space (X,d_X) is said to be W_p barycentric with constant Γ if there exists a mapping $\mathfrak{B}:\mathscr{P}_X\to X$ satisfying $\mathfrak{B}(\delta_x)=x$ for every $x\in X$, and

$$\forall\,\mu,\nu\in\mathscr{P}_X,\qquad d_X(\mathfrak{B}(\mu),\mathfrak{B}(\nu))\leqslant\Gamma W_p(\mu,\nu).\tag{3}$$

(X,d_X) is said to be W_p barycentric if it is W_p barycentric with constant Γ for some $\Gamma\in[1,\infty)$.

In what follows, a mapping $\mathfrak{B}:\mathscr{P}_X\to X$ satisfying $\mathfrak{B}(\delta_x)=x$ for every $x\in X$ will be called a *barycenter map*. The notion of W_p barycentric metric spaces was studied by several authors: see e.g. [33, 17, 60, 20, 54, 53]. Note that if (X,d_X) is W_p barycentric with constant Γ then it is also W_q barycentric with constant Γ for every $q\geqslant p$. Normed spaces are W_1 barycentric with constant $\Gamma=1$, as exhibited by the barycenter map $\mathfrak{B}(\mu)=\int_X x d\mu(x)$. Metric spaces that are nonpositively curved in the sense of Busemann (see [9]) are also W_1 barycentric with constant $\Gamma=1$, as shown in [17, 53].

Definition 1.4 (p-barycentric metric space)**.** Fix $p,K\in[1,\infty)$. A metric space (X,d_X) is said to be p-barycentric with constant K if there exists a mapping $\mathfrak{B}:\mathscr{P}_X\to X$ such that for every $x\in X$ and $\mu\in\mathscr{P}_X$ we have

$$d_X\left(x,\mathfrak{B}(\mu)\right)^p+\frac{1}{K^p}\int_X d_X(\mathfrak{B}(\mu),y)^p d\mu(y)\leqslant\int_X d_X(x,y)^p d\mu(y).\tag{4}$$

(X,d_X) is said to be p-barycentric if it is p-barycentric with constant K for some $K\in[1,\infty)$.

The appearance of the constant K in the left hand side of (4) is natural from the point of view of Banach space theory (see [3, Lem 3.1] and [43, Lem. 6.5]), but note that it means that (unless $K=1$) $\mathfrak{B}(\mu)$ need not be a point $x\in X$ which minimizes the right hand side of (4). In many cases of interest one defines barycenters as minimizers of the right hand side of (4), but as will become clear from the ensuing considerations, Definition 1.4 suffices for many purposes.

In what follows, whenever we say that a metric space is W_p barycentric with constant Γ and also q-barycentric with constant K we mean that Definition 1.3 and Definition 1.4 are satisfied with respect to *the same* barycenter map $\mathfrak{B} : \mathscr{P}_X \to X$. Examples of such spaces include all complete $CAT(0)$ metric spaces (Hadamard spaces), which are W_1 barycentric with constant 1 (see [33, Lem. 4.2] or [60, Thm. 6.3]) and also 2-barycentric with constant 1 (see [33, Lem. 4.1] or [60, Lem. 4.4]). Banach spaces whose modulus of uniform convexity have power type $p \in [2, \infty)$ are W_1 barycentric with constant 1 and also p-barycentric (see [3, Lem 3.1] for $p = 2$ and [43, Lem. 6.5] for $p \in [2, \infty)$).

We refer to the books [6, 26, 9] for an extensive discussion of the important class of $CAT(0)$ metric spaces, which includes e.g. complete simply connected Riemannian manifolds with nonpositive sectional curvature and Euclidean Tits buildings. For the sake of readers who are not familiar with this notion we state that the definition of the class of $CAT(0)$ metric spaces can be taken to be those metric spaces (X, d_X) for which there exists a mapping $\mathfrak{B} : \mathscr{P}_X \to X$ that satisfies (4) with $p = 2$ and $K = 1$ for probability measures μ that are supported on at most two points [60, Thm. 4.9]. Readers who are not familiar with the theory of uniformly convex Banach spaces are referred to [18, 5].

1.3. **Metric Markov cotype for barycentric metric spaces.**

In Section 3 we prove the following result.

Theorem 1.5. *Fix $p, K, \Gamma \in [1, \infty)$. Suppose that (X, d_X) is a metric space that is W_p barycentric with constant Γ and also p-barycentric with constant K. Then (X, d_X) has metric Markov cotype p with*

$$N_p(X) \lesssim \Gamma K. \tag{5}$$

The special case of Theorem 1.5 when X is a Banach space whose modulus of uniform convexity has power type p was proved in [3, 43]. Our proof of Theorem 1.5 is based on an extension of the method of [43] to the present nonlinear setting. In particular we prove for this purpose a nonlinear analogue of Pisier's martingale cotype inequality [57]; see Section 2 below.

Remark 1.6. The property of having metric Markov cotype p is clearly a bi-Lipschitz invariant. Similarly, the property of being W_p barycentric is a bi-Lipschitz invariant, but this is not the case for the property of being p-barycentic. Thus Theorem 1.5 leaves something to be desired, since its assumption is not invariant under bi-Lipschitz deformations while its conclusion is. By examining the proof of Theorem 1.5 one can extract a somewhat tedious bi-Lipschitz invariant condition that implies the same conclusion (5). It would be interesting to obtain a clean intrinsic characterization of those metric spaces (X, d_X) that are bi-Lipschitz equivalent to a p-barycentric metric space. For Banach spaces this was done in [45], the desired metric invariant being the notion of *Markov p-convexity* (see [45] for the definition). The method of [45] relies on the Banach space structure, so it remains open to characterize intrinsically those W_p barycentric metric spaces that are bi-Lipschitz equivalent to a p-barycentric metric space. It would also be interesting to characterize those Finsler manifolds that are p-barycentric.

1.4. **Calculus for nonlinear spectral gaps.** Let $A = (a_{ij}) \in M_n(\mathbb{R})$ be a symmetric stochastic matrix. Denote the decreasing rearangment of the eigenvalues of A by $1 = \lambda_1(A) \geqslant \lambda_2(A) \geqslant \lambda_n(A) \geqslant -1$, and write $\lambda(A) = \max_{i\in\{2,\ldots,n\}} |\lambda_i(A)|$.

Following [43], given a metric space (X, d_X) and $p \in (0, \infty)$ let $\gamma_+(A, d_X^p)$ denote the infimum over those $\gamma_+ \in (0, \infty]$ for which every $x_1, \ldots, x_n, y_1, \ldots, y_n \in X$ satisfy

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i, y_j)^p \leqslant \frac{\gamma_+}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} d_X(x_i, y_j)^p. \tag{6}$$

Letting $d_{\mathbb{R}}$ denote the standard metric on $\mathbb{R}$, i.e., $d_{\mathbb{R}}(x, y) = |x-y|$, by simple linear algebra we see that $\gamma_+(A, d_{\mathbb{R}}^2) = 1/(1-\lambda(A))$. One should therefore think of the quantity $\gamma_+(A, d_X^p)$ as measuring the magnitude of the *nonlinear absolute spectral gap* of the matrix A with respect to the geometry of X. We refer to [43] for a detailed discussion of nonlinear spectral gaps and their applications.

Despite the fact that we call inequalities such as (6) "spectral inequalities", there is no actual spectrum present here, and therefore tools that are straightforward in the linear setting due to the link to linear algebra fail to hold true in general. This is especially important in the context of *nonlinear spectral calculus*, where one aims to relate $\gamma_+(A^t, d_X^p)$ to $\gamma_+(A, d_X^p)$. We refer to [43] for an explanation of the importance of this problem, where the following theorem is proved.

Theorem 1.7. *There exists a universal constant $\kappa \in (0, \infty)$ with the following property. Suppose that $p \in [1, \infty)$ and that (X, d_X) is a metric space that has metric Markov cotype p. Then for every $n, t \in \mathbb{N}$, every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ satisfies*

$$\gamma_+\left(\frac{1}{t}\sum_{s=1}^{t} A^s, d_X^p\right) \leqslant (\kappa N_p(X))^p \max\left\{1, \frac{\gamma_+(A, d_X^p)}{t}\right\}.$$

By combining Theorem 1.5 and Theorem 1.7 we conclude that the following result holds true.

Theorem 1.8. *There exists a universal constant $c \in (1,\infty)$ such that for every $p, K, \Gamma \in [1,\infty)$, if (X,d_X) is a metric space that is W_p barycentric with constant Γ and p-barycentric with constant K then for every $n,t \in \mathbb{N}$, every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ satisfies*

$$\gamma_+\left(\frac{1}{t}\sum_{s=1}^{t} A^s, d_X^p\right) \leqslant (c\Gamma K)^p \max\left\{1, \frac{\gamma_+(A,d_X^p)}{t}\right\}. \tag{7}$$

For future applications it is worthwhile to single out the following special case of Theorem 1.8.

Corollary 1.9. *Suppose that (X,d_X) is a $CAT(0)$ space. Then for every $n,t \in \mathbb{N}$, every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ satisfies*

$$\gamma_+\left(\frac{1}{t}\sum_{s=1}^{t} A^s, d_X^2\right) \lesssim \max\left\{1, \frac{\gamma_+(A,d_X^2)}{t}\right\}.$$

Corollary 1.9 was the main motivation for the investigations that led to the present paper, since it plays a key role in our forthcoming work [44] that establishes for the first time the existence of expanders with respect to certain Hadamard spaces and random graphs.

For the purpose of the applications in [44], the fact that the spectral calculus inequality (7) involves Cesàro averages of A rather than powers of A is immaterial, but it is natural to ask if it is possible to relate $\gamma_+(A^t, d_X^p)$ to $\gamma_+(A, d_X^p)$. In the setting of general barycentric metric spaces this question remains open, but for $CAT(0)$ spaces, or more generally under the requirement $K=1$ in (4), it is indeed possible to do so, albeit via an upper bound on $\gamma_+(A^t, d_X^p)$ in terms of $\gamma_+(A, d_X^p)$ that is weaker than the right hand side of (7).

Theorem 1.10. *There is a universal constant $C \in (0,\infty)$ with the following property. Fix $p, \Gamma \in [1,\infty)$ and suppose that (X,d_X) is a metric space that is W_p barycentric with constant Γ and p-barycentric with constant $K=1$. Then for every $n,t \in \mathbb{N}$, every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ satisfies*

$$\gamma_+\left(A^t, d_X^p\right) \leqslant (C\Gamma)^p \left(\max\left\{1, p\cdot\frac{\gamma_+(A,d_X^p)}{t}\right\}\right)^p.$$

Our proof of Theorem 1.10 relies on ideas from [43, Sec. 6], where a similar treatment is given to uniformly convex Banach spaces (in this special context the conclusion of Theorem 1.10 holds true even without the restriction $K=1$). In the present nonlinear setting several modifications of the argument of [43] are required; see Section 4 below.

1.5. **Lipschitz extension.** If (X, d_X) and (Y, d_Y) are metric spaces then for every $S \subseteq X$ denote by $e(X, S, Y)$ the infimum over those $L \in (0, \infty)$ such that for *every* Lipschitz function $f : S \to Y$ there exists $F : X \to Y$ with $F(x) = f(x)$ for every $x \in S$ such that $\|F\|_{\mathrm{Lip}} \leqslant L\|f\|_{\mathrm{Lip}}$, where $\|f\|_{\mathrm{Lip}}$ denotes the Lipschitz constant of f. If no such L exists then set $e(X, S, Y) = \infty$.

Defining

$$e(X, Y) \stackrel{\mathrm{def}}{=} \sup \{e(X, S, Y) : \ S \subseteq X\},$$

the goal of the *Lipschitz extension problem* is to understand which pairs of metric spaces $(X, d_X), (Y, d_Y)$ satisfy $e(X, Y) < \infty$, and, when that happens, to obtain good bounds on $e(X, Y)$. Due to its intrinsic importance as well as many applications in analysis and geometry, the Lipschitz extension problem has been extensively investigated over the past century. We shall not attempt to indicate the vast literature on this topic, referring instead to the book [11] and the references therein.

K. Ball introduced [3] the notions of Markov type and cotype in order to prove an important Lipschitz extension theorem known today as *Ball's extension theorem.* Based on Ball's ideas in [3], the following result is proved in Section 5.

Theorem 1.11 (generalized Ball extension theorem)**.** *Fix $p, \Gamma \in [1, \infty)$. Let (X, d_X) be a metric space of Markov type p and let (Y, d_Y) be a metric space of metric Markov cotype p that is W_p barycentric with constant Γ. Suppose that $Z \subseteq X$ and $f : Z \to Y$ is Lipschitz. Then for every finite subset $S \subseteq X$ there exists $F : S \to Y$ with $F|_{S\cap Z} = f|_{S\cap Z}$ and*

$$\|F\|_{\mathrm{Lip}} \lesssim \Gamma M_p(X) N_p(Y) \|f\|_{\mathrm{Lip}}. \tag{8}$$

By combining Theorem 1.11 with Theorem 1.5 we deduce the following Lipschitz extension result.

Corollary 1.12. *Fix $p, K, \Gamma \in [1, \infty)$. Suppose that (X, d_X) is a metric space of Markov type p and that (Y, d_Y) is a metric space that is W_p barycentric with constant Γ and also p-barycentric with constant K. Suppose that $Z \subseteq X$ and $f : Z \to Y$ is Lipschitz. Then for every finite subset $S \subseteq X$ there exists $F : S \to Y$ with $F|_{S\cap Z} = f|_{S\cap Z}$ and*

$$\|F\|_{\mathrm{Lip}} \lesssim \Gamma^2 K M_p(X) \|f\|_{\mathrm{Lip}}.$$

In [3] Ball obtained the conclusion of Theorem 1.11 when Y is a Banach space, under the assumption that it satisfies a certain *linear* invariant that he called Markov cotype 2. He also proved that Banach spaces that admit an equivalent norm whose modulus of uniform convexity has power type 2 satisfy this assumption.

In [3, Sec. 6], Ball proposed a way to define Markov cotype 2 for metric spaces: he first defined a bi-Lipschitz invariant of metric spaces that he called "approximate convexity", and for approximately convex metric spaces he defined a notion of metric Markov cotype which is the same as (2), except that in the right hand side of (2) the Cesàro average of A is replaced by a certain Green's matrix corresponding to A. The precise formulation of these concepts is recalled in Section 7, where we show that Ball's notion of metric Markov cotype coincides with the notion of metric Markov cotype as in Definition 2. Our contribution here is to show that Ball's strategy yields the desired Lipschitz extension result, with the following differences: the W_p barycentric condition is used in a key duality step (Lemma 5.2 below), and Lemma 5.1 below removes the need to use the notion of approximate convexity. Other than these changes and some expository simplifications, Section 5 is nothing more than a realization of Ball's plan as he originally envisaged it.

Theorem 1.11 yields an extension of f to finitely many additional points, with a bound on the Lipschitz constant that is independent of the number of the additional points. This result is the main geometric content of the Lipschitz extension phenomenon studied here, but using standard arguments one can formally deduce from Theorem 1.11 bona fide solutions of the Lipschitz extension problem.

Specifically, let I denote the set of all finite subsets of X and let $\mathscr{U}$ be a free ultrafilter on I. Denoting by $Y_{\mathscr{U}}$ the associated ultrapower of Y (see [30] for background on ultrapowers of metric spaces), Y is canonically embedded in $Y_{\mathscr{U}}$ and it follows formally from Theorem 1.11 that there exists a mapping $\Phi : X \to Y_{\mathscr{U}}$ that extends f and satisfies (8). If for some $\lambda \in [1,\infty)$ there were a λ-Lipschitz retraction from $Y_{\mathscr{U}}$ onto Y, then by composing Φ with this retraction we would deduce that

$$e(X,Y) \lesssim \lambda \Gamma M_p(X) N_p(Y). \tag{9}$$

If Y is Banach space then $Y_{\mathscr{U}}$ is also a Banach space, and, as proved in [22], it follows from the principle of local reflexivity [38, 25] that there is a linear isometry $T : Y^{**} \to Y_{\mathscr{U}}$ such that $T(Y^{**})$ contains the canonical image of Y in $Y_{\mathscr{U}}$, and there is a norm 1 projection of $Y_{\mathscr{U}}$ onto $T(Y^{**})$. It therefore follows from Theorem 1.11 that

$$e(X,Y^{**}) \lesssim M_p(X) N_p(Y).$$

If in addition there is a λ-Lipschitz retraction from Y^{**} onto Y then it would follow that (9) holds true (with $\Gamma = 1$).

It is a long standing open problem whether for every separable Banach space Y there is a Lipschitz retraction from Y^{**} onto Y, but in the nonseparable setting it has been recently proved by Kalton [28] that this need not hold true. A dual Banach space is always canonically norm 1 complemented in its bi-dual, and in [29, Sec. 5] Kalton proved that if Y either has an unconditional finite dimensional decomposition (UFDD) or is a separable order continuous Banach lattice then there is a Lipschitz retraction from Y^{**} onto Y.

If Y is a complete $CAT(0)$ metric space then so is $Y_{\mathscr{U}}$, and moreover Y is a closed convex subset of $Y_{\mathscr{U}}$. In this case there is a 1-Lipschitz retraction from $Y_{\mathscr{U}}$ onto Y (the nearest point map); see [9, Ch. II.2].

The above discussion yields a variety of target spaces Y for which the assumptions of Theorem 1.11 implies that $e(X,Y) < \infty$. We single out in particular the following statement.

Corollary 1.13. *Under the assumptions of Theorem 1.11, if in addition Y is a dual Banach space then $e(X,Y) \lesssim M_p(X)N_p(Y)$. If $p=2$ and Y is a complete $CAT(0)$ metric space then $e(X,Y) \lesssim M_2(X)$.*

The Markov type of several important classes of metric spaces has been computed in the literature, and when one takes (X,d_X) to be one of those spaces Corollary 1.13 becomes a versatile Lipschitz extension theorem that encompasses a wide range of seemingly disparate Lipschitz extension results, that have been previously proved mostly via completely different methods.

Specifically, in [50] it was proved that Banach spaces that admit an equivalent norm whose modulus of uniform smoothness has power type p have Markov type p. It was also proved in [50] that trees, hyperbolic groups, complete simply connected Riemannian manifolds of pinched sectional curvature and Laakso graphs all have Markov type 2, and that spaces that admit a padded random partition (see [36]), in particular doubling metric spaces and planar graphs, have Markov type p for all $p \in (0,2)$. In [10] it was shown that series parallel graphs have Markov type 2, and finally in the recent work [13] it was shown that spaces that admit a padded random partition have Markov type 2. Thus, in particular, doubling spaces and planar graphs have Markov type 2. In [52] it was shown that spaces with finite Nagata dimension admit a padded random partition, and so by [13] they too have Markov type 2. In [55] it was shown that Aleksandrov spaces of nonnegative curvature have Markov type 2, and in [1] the Markov type of certain Wasserstein spaces was computed.

In light of these results, taking as an example the case when (Y,d_Y) is a Hadamard space in Corollary 1.13, we see that if (X,d_X) is a doubling space, planar graph, or a space with finite Nagata dimension, then $e(X,Y)$ is finite. These results were previously proved in [36] via the method of random partitions (Lipschitz extension for spaces of bounded

Nagata dimension was previously treated in [34] and only later it was shown in [52] that they admit a padded random partition and therefore the corresponding extension results are a special case of [36]). It also follows that if (X, d_X) has nonnegative curvature in the sense of Aleksandrov and (Y, d_Y) is a Hadamard space then $e(X, Y) \lesssim 1$, a result that has been previously proved in [35], as a special case of an elegant generalization of the classical Kirszbraun extension theorem [31].

Given a metric space (X, d_X) and $\alpha \in (0, 1]$ let X^α denote the metric space (X, d_X^α). By the triangle inequality X^α has Markov type p with constant 1 for every $p \in (1, 1/\alpha]$. It therefore follows from the above discussion that $e(X^\alpha, Y) < \infty$ for every metric space X, provided that Y has metric Markov cotype $p \in (1, 1/\alpha]$ and there is a Lipschitz retraction from $Y_{\mathscr{U}}$ onto Y. In particular, every 1/2-Hölder mapping from a subset of a metric space X into a Hadamard space Y can be extended to a Y-valued 1/2-Hölder mapping defined on all of X; this statement was previously known when Y is a Hilbert space due to the work of Minty [47]. One can state several additional examples of this type, but we single out only one more special case of Corollary 1.13 that does not seem to follow from previously known theorems: if X is a Banach space whose modulus of smoothness has power type 2 (thus by [50] X has Markov type 2), e.g. X can be an $L_p(\mu)$ space or the Schatten trace class S_p for $p \in [2, \infty)$, and Y is a Hadamard space, then $e(X, Y) < \infty$.

1.6. **On a construction of Kalton.** Kalton recently used his "method of sections" to obtain several striking results on the nonlinear geometry of Banach spaces. Using Kalton's beautiful work in [29], we prove the following result in Section 6.

Theorem 1.14. *There exists a closed linear subspace of ℓ_1 that fails to have metric Markov cotype p for every $p \in (0, \infty)$.*

Much of the impetus for research on bi-Lipschitz invariants stems from the search for nonlinear formulations of key concepts in Banach space theory; see the surveys [4, 49] and the references therein for more on this program. In particular, the use of the term "cotype" in Definition 1.2 arises from an analogy with the Banach space notion of *Rademacher cotype* (see e.g. [41]). ℓ_1, and hence all of its linear subspaces, has Rademacher cotype 2, so Theorem 1.14 shows that for Banach spaces metric Markov cotype and Rademacher cotype are different notions. Nevertheless, it would be very interesting to understand the metric Markov cotype of ℓ_1 itself rather than its closed subspaces (note that, due to the existential quantifier in Definition 1.2, metric Markov cotype is not trivially inherited by subspaces).

Question 1.15. Does ℓ_1 have metric Markov cotype 2? Less ambitiously, does ℓ_1 have metric Markov cotype p for some $p \in [2,\infty)$?

If ℓ_1 had metric Markov cotype 2 then it would follow from Corollary 1.13 that $e(\ell_2,\ell_1) < \infty$. Whether or not $e(\ell_2,\ell_1)$ is finite is a long-standing open question that was asked by Ball in [3]; see [39] for algorithmic ramifications of this important question.

The proof of Theorem 1.14 yields the following quantitative statement. For every $n \in \mathbb{N}$ there exists an n-dimensional subspace Z_n of ℓ_1 such that

$$N_2(Z_n) \gtrsim \sqrt[4]{\log n}. \tag{10}$$

Any n-dimensional subspace X of ℓ_1 satisfies $N_2(X) \lesssim \sqrt{\log n}$. Indeed, by [61] we know that X is 2-isomorphic to a subspace of ℓ_1^k, with $k \lesssim n \log n$ (for our purpose we can also use the weaker bound on k of [58]). By Hölder's inequality ℓ_1^k is $O(1)$-isomorphic to a subspace of ℓ_p with $p = 1 + 1/\log k$, so the desired upper bound on $N_2(X)$ follows from [3, 43]. We ask whether (10) can be sharpened.

Question 1.16. Is it true that for arbitrarily large $n \in \mathbb{N}$ there exists an n-dimensional subspace X of ℓ_1 with $N_2(X) \gtrsim \sqrt{\log n}$?

An interesting byproduct of our quantitative analysis of Kalton's construction is that it shows for the first time that an old Lipschitz extension result of Benyamini, Johnson and Lindenstrauss [23] cannot be improved. Given $\varepsilon \in (0,1)$ and spaces $(X,d_X),(Y,d_Y)$, denote

$$e_\varepsilon(X,Y) \stackrel{\mathrm{def}}{=} \sup\left\{ e(X,S,Y):\ S \subseteq X,\ \text{and} \inf_{\substack{x,y\in S\\ x\neq y}} d_X(x,y) \geqslant \varepsilon\, \mathrm{diam}(S) \right\},$$

where $\mathrm{diam}(S) = \sup_{x,y\in S} d_X(x,y)$ is the diameter of S. In other words, we are interested in the extension of Y-valued Lipschitz functions from ε-separated subsets of X, where ε-separated means that all positive distances in the subset are at least an ε-fraction of its diameter.

In [23] it was shown that for every $\varepsilon \in (0,1)$, every metric space (X,d_X) and every Banach space (Y,d_Y) we have

$$e_\varepsilon(X,Y) \lesssim \frac{1}{\varepsilon}. \tag{11}$$

Specifically, a first proof of (11) was given by Johnson and Lindenstrauss in [23] when Y is a Hilbert space, and in the appendix of the same paper Johnson and Lindenstrauss include a different argument that was subsequently found by Benyamini establishing (11) when Y is a general Banach space. A very short proof of (11) was later found by Johnson, Lindenstrauss and Schechtman [24].

Johnson and Lindenstrauss proved [23] that $e_\varepsilon(\ell_1, \ell_2) \gtrsim 1/\sqrt[4]{\varepsilon}$. Constructions of Johnson, Lindenstrauss and Schechtman [24] and Lang [32] yield the estimate $e_\varepsilon(\ell_\infty, \ell_2) \gtrsim 1/\sqrt[3]{\varepsilon}$. Here we show that (11) is sharp up to absolute constant factors, even when X is Hilbert space and Y is an appropriately chosen closed subspace of ℓ_1.

Theorem 1.17. *There exists a closed subspace Y of ℓ_1 that satisfies $e_\varepsilon(\ell_2, Y) \gtrsim 1/\varepsilon$ for every $\varepsilon \in (0,1)$. Specifically, for every $n \in \mathbb{N}$ there exists a 5^n-dimensional subspace Y_n of ℓ_1 and a $1/\sqrt[4]{n}$ net $\mathcal{N}$ of the unit ball of ℓ_2^n such that $e(\ell_2^n, \mathcal{N}, Y_n) \gtrsim \sqrt[4]{n}$.*

It would be very interesting to understand those pairs of Banach spaces X, Y for which $e_\varepsilon(X, Y) = o(1/\varepsilon)$ as $\varepsilon \to 0$. We do not even know if there exist Banach spaces X, Y such that $e_\varepsilon(X, Y) = o(1/\varepsilon)$ yet $e(X, Y) = \infty$. Our interest in this natural question is partially motivated by the forthcoming work [2], where it is asked whether $e_\varepsilon(\ell_1, \ell_1) = o(1/\varepsilon)$, and it is shown that a positive answer to this question would have applications to dimension reduction in ℓ_1 (e.g., it is shown in [2] that if $e_\varepsilon(\ell_1, \ell_1) = o(1/\varepsilon)$ then any n-point subset of ℓ_1 embeds with distortion $O(1)$ into some Banach space of dimension $(\log n)^{O(1)}$). Due to Theorem 1.17, one is tempted to believe that in fact $e_\varepsilon(\ell_1, \ell_1) \gtrsim 1/\varepsilon$, but the present approach does not seem to shed light on this question.

2. Pisier's martingale inequality in barycentric spaces

Martingales in metric spaces have been studied for several decades; see e.g. [14, 16, 17, 59, 12]. Here we will use a natural notion of martingale in barycentric metric spaces, the main goal being to extend an important martingale inequality of Pisier [57] from the setting of uniformly convex Banach spaces to the setting of barycentric metric spaces. This inequality will be used crucially in the proof of Theorem 1.5.

Let Ω be a finite set and $\mu \in \mathscr{P}_\Omega$ be a probability measure such that $\mu(\omega) > 0$ for every $\omega \in \Omega$. Suppose that (X, d_X) is a metric space and fix a barycenter map $\mathfrak{B} : \mathscr{P}_X \to X$. Let $\mathcal{F} \subseteq 2^\Omega$ be a σ-algebra. For every $\omega \in \Omega$ let $\mathcal{F}(\omega) \subseteq \Omega$ be the unique atom of $\mathcal{F}$ to which ω belongs. Given an X-valued random variable $Z : \Omega \to X$, its conditional barycenter $\mathfrak{B}(Z|\mathcal{F}):\Omega\to X$ is defined as

$$\forall\,\omega\in\Omega,\quad \mathfrak{B}(Z|\mathcal{F})(\omega)\stackrel{\mathrm{def}}{=}\mathfrak{B}\left(\frac{1}{\mu(\mathcal{F}(\omega))}\sum_{a\in\mathcal{F}(\omega)}\mu(a)\delta_{Z(a)}\right). \tag{12}$$

If $m\in\mathbb{N}$ and $\{\Omega,\emptyset\}=\mathcal{F}_0\subseteq\mathcal{F}_1\subseteq\ldots\subseteq\mathcal{F}_m\subseteq 2^\Omega$ are increasing σ-algebras (a filtration) then a sequence of X-valued random variables $Z_0,\ldots,Z_n:\Omega\to X$ is said to be a martingale if for every $i\in\{1,\ldots,m\}$ we have $\mathfrak{B}(Z_i|\mathcal{F}_{i-1})=Z_{i-1}$. We warn that in contrast to the usual setting of martingales in Banach spaces, this definition does not necessarily imply that $\mathfrak{B}(Z_i|\mathcal{F}_j)=Z_j$ for every $j\in\{0,\ldots,i-2\}$. Nevertheless, the above notion of martingale suffices to prove the following inequality.

Lemma 2.1 (Pisier's inequality for barycentric spaces)**.** *Fix $m\in\mathbb{N}$, $p,K\in[1,\infty)$ and a metric space (X,d_X) that is p-barycentric with constant K. Let Ω be a finite set and fix $\mu\in\mathscr{P}_\Omega$ with $\mu(\omega)>0$ for every $\omega\in\Omega$. Suppose that $\{\Omega,\emptyset\}=\mathcal{F}_0\subseteq\mathcal{F}_1\subseteq\ldots\subseteq\mathcal{F}_m\subseteq 2^\Omega$ is a filtration with respect to which $Z_0,\ldots,Z_n:\Omega\to X$ is an X-valued martingale. Then for every $z\in X$ we have*

$$K^pd_X(Z_0,z)^p+\sum_{t=0}^{m-1}\int_\Omega d_X(Z_{t+1},Z_t)^pd\mu\leqslant K^p\int_\Omega d_X(Z_m,z)^pd\mu. \tag{13}$$

Proof. Fix $t\in\{0,\ldots,m-1\}$ and $\omega\in\Omega$. Recalling (12), an application of (4) to the probability measure $\frac{1}{\mu(\mathcal{F}_t(\omega))}\sum_{a\in\mathcal{F}_t(\omega)}\mu(a)\delta_{Z_{t+1}(a)}$ yields the estimate

$$\begin{aligned}d_X(Z_t(\omega),z)^p+\frac{1}{K^p}\sum_{a\in\mathcal{F}_t(\omega)}\frac{\mu(a)}{\mu(\mathcal{F}_t(\omega))}d_X(Z_t(\omega),Z_{t+1}(a))^p&\\ \leqslant\sum_{a\in\mathcal{F}_t(\omega)}\frac{\mu(a)}{\mu(\mathcal{F}_t(\omega))}d_X(z,Z_{t+1}(a))^p,&\end{aligned} \tag{14}$$

where we used the martingale assumption $Z_t=\mathfrak{B}(Z_{t+1}|\mathcal{F}_t)$. Let $A_1,\ldots,A_k\in\mathcal{F}_t$ be the distinct atoms of $\mathcal{F}_t$ and fix $\omega_i\in A_i$ for every $i\in\{1,\ldots,k\}$ (thus $A_i=\mathcal{F}_t(\omega_i)$). It follows from (14) that for every $i\in\{1,\ldots,k\}$,

$$\begin{aligned}&\sum_{a\in A_i}\mu(a)d_X(Z_t(\omega_i),Z_{t+1}(a))^p\\ &\qquad\leqslant K^p\left(\sum_{a\in A_i}\mu(a)d_X(z,Z_{t+1}(a))^p-\mu(A_i)d_X(Z_t(\omega_i),z)^p\right).\end{aligned} \tag{15}$$

Since Z_t is constant on each of the sets $A_1,\ldots,A_k$, by summing (15) over $i\in\{1,\ldots,k\}$ we obtain the estimate

$$\frac{1}{K^p}\int_\Omega d_X(Z_{t+1},Z_t)^p d\mu \leqslant \int_\Omega d_X(Z_{t+1},z)^p d\mu - \int_\Omega d_X(Z_t,z)^p d\mu. \quad (16)$$

The desired inequality (13) now follows since the right hand side of (16) telescopes upon summation over $t\in\{0,\ldots,m-1\}$. □

3. Proof of Theorem 1.5

Throughout the remainder of this paper it will be convenient to use the following notation for Cesáro averages.

$$\mathscr{A}_t(A) \stackrel{\mathrm{def}}{=} \frac{1}{t}\sum_{s=1}^{t} A^s. \quad (17)$$

The following simple lemma will be used in the proof of Theorem 1.5, as well as in Section 7.

Lemma 3.1. *Fix $p\in[1,\infty)$ and $n,t\in\mathbb{N}$. Suppose that $A\in M_n(\mathbb{R})$ is a stochastic matrix that is reversible relative to $\pi\in\Delta^{n-1}$. Then for every metric space (X,d_X) and every $x_1,\ldots,x_n\in X$,*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i(A^t)_{ij}d_X(x_i,x_j)^p \leqslant 2^p\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\mathscr{A}_t(A)_{ij}d_X(x_i,x_j)^p.$$

Proof. By the triangle inequality, for every $i,j,k\in\{1,\ldots,n\}$ we have

$$d_X(x_i,x_j)^p \leqslant 2^{p-1}\big(d_X(x_i,x_k)^p + d_X(x_k,x_j)^p\big). \quad (18)$$

Consequently,

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i(A^t)_{ij}d_X(x_i,x_j)^p\\
&= \frac{1}{t}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\left(\sum_{k=1}^{n}(A^s)_{ik}(A^{t-s})_{kj}\right)d_X(x_i,x_j)^p\\
&\stackrel{(18)}{\leqslant} \frac{2^{p-1}}{t}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\pi_i(A^s)_{ik}(A^{t-s})_{kj}\big(d_X(x_i,x_k)^p + d_X(x_k,x_j)^p\big)\\
&\stackrel{(17)}{=} 2^{p-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\left(\mathscr{A}_t(A)_{ij} + \frac{t-1}{t}\mathscr{A}_{t-1}(A)_{ij}\right)d_X(x_i,x_j)^p\\
&\leqslant 2^p\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\mathscr{A}_t(A)_{ij}d_X(x_i,x_j)^p.
\end{aligned}$$

□

Proof of Theorem 1.5. Fix $p,K,\Gamma\in[1,\infty)$, a metric space (X,d_X) and a barycenter map $\mathfrak{B}:\mathscr{P}_X\to X$ with respect to which (X,d_X) is both W_p barycentric with constant Γ and p-barycentric with constant K. We also fix $n,t\in\mathbb{N}$, a probability vector $\pi\in\Delta^{n-1}$ and a stochastic matrix $A=(a_{ij})$ that is reversible relative to π. Given $x_1,\ldots,x_n\in X$ our goal is to prove that there exist $y_1,\ldots,y_n\in X$ such that (2) is satisfied with $N\lesssim \Gamma K$.

In the proof of Theorem 1.5 we may assume that the right hand side of (2) is nonzero. By restricting to the support of π we may also assume that $\pi\in(0,1)^n$. Letting $\Pi\in M_n(\mathbb{R})$ be given by $\Pi_{ij}=\pi_j$, choose $\varepsilon\in(0,1/2)$ small enough so that for $B=(1-\varepsilon)A+\varepsilon\Pi$ we have

$$\sum_{i=1}^n\sum_{j=1}^n \pi_i\mathscr{A}_t(B)_{ij}d_X(x_i,x_j)^p\lesssim \sum_{i=1}^n\sum_{j=1}^n \pi_i\mathscr{A}_t(A)_{ij}d_X(x_i,x_j)^p.$$

Since $B_{ij}\geqslant(1-\varepsilon)a_{ij}\geqslant\frac12 a_{ij}$, for every $y_1,\ldots,y_n\in X$ we have

$$\begin{aligned}\sum_{i=1}^n \pi_i d_X(x_i,y_i)^p&+t\sum_{i=1}^n\sum_{j=1}^n \pi_iB_{ij}d_X(y_i,y_j)^p\\ &\gtrsim \sum_{i=1}^n \pi_i d_X(x_i,y_i)^p+t\sum_{i=1}^n\sum_{j=1}^n \pi_ia_{ij}d_X(y_i,y_j)^p.\end{aligned}$$

Since the matrix B is stochastic and reversible relative to π and none of its entries vanish, this shows that it suffices to prove Theorem 1.5 under the assumption that $\pi_i,a_{ij}>0$ for every $i,j\in\{1,\ldots,n\}$.

Denote $\Omega=\{1,\ldots,n\}^t$. Set $\mathcal{F}_0=\{\emptyset,\Omega\}$ and for every $s\in\{1,\ldots,t\}$ let $\mathcal{F}_s\subseteq 2^\Omega$ be the σ-algebra generated by the first s coordinates, i.e., the atoms of $\mathcal{F}_s$ are $\{E_\tau\}_{\tau\in\{1,\ldots,n\}^s}$, where we denote for every $(i_1,\ldots,i_s)\in\{1,\ldots,n\}^s$,

$$E_{(i_1,\ldots,i_s)}\stackrel{\mathrm{def}}{=}\big\{(j_1,\ldots,j_t)\in\Omega:\ (j_1,\ldots,j_s)=(i_1,\ldots,i_s)\big\}.$$

Fix $\ell\in\{1,\ldots,n\}$ and define $\mu_\ell\in\mathscr{P}_\Omega$ on Ω by

$$\forall\,(i_1,\ldots,i_t)\in\Omega,\qquad \mu_\ell(i_1,\ldots,i_t)\stackrel{\mathrm{def}}{=}a_{\ell,i_1}\prod_{s=1}^{t-1}a_{i_s,i_{s+1}}.$$

Thus (Ω,μ_ℓ) is the probability space of trajectories of length t of the Markov chain on $\{1,\ldots,n\}$ that starts at ℓ and whose transition matrix is A. By definition, $\mu_\ell(\omega)>0$ for every $\omega\in\Omega$.

We next define inductively mappings $M_0^{(\ell,t)},\ldots,M_t^{(\ell,t)}:\Omega\to X$ as follows. For every $(i_1,\ldots,i_t)\in\Omega$,

$$M_t^{(\ell,t)}(i_1,\ldots,i_t)\stackrel{\mathrm{def}}{=}x_{i_t},$$

and for every $s\in\{0,\ldots,t-1\}$,

$$M_s^{(\ell,t)}(i_1,\ldots,i_t)\stackrel{\mathrm{def}}{=}\mathfrak{B}\left(\sum_{j=1}^n a_{i_s,j}\cdot\delta_{M_{s+1}^{(\ell,t)}(i_1,\ldots,i_s,j,1\ldots,1)}\right), \qquad (19)$$

where $i_0=\ell$. Thus $M_s^{(\ell,t)}(i_1,\ldots,i_t)$ depends only on $(i_1,\ldots,i_s)$. We may therefore think of $M_s^{(\ell,t)}$ as an X-valued function defined on $\{1,\ldots,n\}^s$, and $M_0^{(\ell,t)}$ as a point in X.

Definition (19) implies that for $s\in\{0,\ldots,t\}$ and $(i_1,\ldots,i_t)\in\Omega$,

$$M_s^{(\ell,t)}(i_1,\ldots,i_t)=M_0^{(i_s,t-s)}. \qquad (20)$$

Moreover, recalling (12) it follows from (19) that for $s\in\{0,\ldots,t-1\}$,

$$M_s^{(\ell,t)}=\mathfrak{B}\left(M_{s+1}^{(\ell,t)}\Big|\mathcal{F}_s\right).$$

Thus $M_0^{(\ell,t)},\ldots,M_t^{(\ell,t)}$ is an X-valued martingale with respect to the measure μ_ℓ and the filtration $\mathcal{F}_0\subseteq\ldots,\subseteq\mathcal{F}_t=2^\Omega$. An application of Lemma 2.1 (with $z=x_\ell$) therefore yields the following estimate.

$$\sum_{s=1}^t\sum_{i=1}^n(A^{s-1})_{\ell,i}\sum_{j=1}^n a_{ij}d_X\left(M_0^{(i,t-s+1)},M_0^{(j,t-s)}\right)^p$$
$$\leqslant K^p\left(\sum_{j=1}^n(A^t)_{\ell j}d_X\left(x_j,x_\ell\right)^p-d_X\left(x_\ell,M_0^{(\ell,t)}\right)^p\right). \qquad (21)$$

Multiplying (21) by π_ℓ and summing over $\ell\in\{1,\ldots,n\}$ while using the fact that A^{s-1} is stochastic and reversible relative to π shows that

$$\sum_{s=1}^t\sum_{i=1}^n\sum_{j=1}^n\pi_i a_{ij}d_X\left(M_0^{(i,s)},M_0^{(j,s-1)}\right)^p$$
$$\leqslant K^p\left(\sum_{\ell=1}^n\sum_{j=1}^n\pi_\ell(A^t)_{\ell j}d_X(x_\ell,x_j)^p-\sum_{\ell=1}^n\pi_\ell d_X\left(x_\ell,M_0^{(\ell,t)}\right)^p\right). \qquad (22)$$

In order to bound the left hand side of (22) from below, observe that for every $i,j\in\{1,\ldots,n\}$ condition (3) of our assumption that (X,d_X) is W_p barycentric with constant Γ implies that

$$\sum_{s=1}^t d_X\left(M_0^{(i,s)},M_0^{(j,s-1)}\right)^p$$
$$\geqslant\frac{t}{\Gamma^p}d_X\left(\mathfrak{B}\left(\frac1t\sum_{s=1}^t\delta_{M_0^{(i,s)}}\right),\mathfrak{B}\left(\frac1t\sum_{s=1}^t\delta_{M_0^{(j,s-1)}}\right)\right)^p. \qquad (23)$$

Moreover, by the triangle inequality and convexity of $u \mapsto u^p$ on $[0,\infty)$,

$$\begin{aligned} & d_X\left(\mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(i,s)}}\right), \mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(j,s)}}\right)\right)^p \\ & \leqslant 2^{p-1} d_X\left(\mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(i,s)}}\right), \mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(j,s-1)}}\right)\right)^p \\ & \quad + 2^{p-1} d_X\left(\mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(j,s)}}\right), \mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(j,s-1)}}\right)\right)^p. \end{aligned} \tag{24}$$

Another application of (3) shows that

$$\begin{aligned} & d_X\left(\mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(j,s)}}\right), \mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(j,s-1)}}\right)\right) \\ & \qquad \leqslant \frac{\Gamma}{t^{1/p}} d_X\left(M_0^{(j,t)}, M_0^{(j,0)}\right) = \frac{\Gamma}{t^{1/p}} d_X\left(M_0^{(j,t)}, x_j\right). \end{aligned} \tag{25}$$

Consequently, if we define

$$y_i \stackrel{\mathrm{def}}{=} \mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^t \delta_{M_0^{(i,s)}}\right), \tag{26}$$

then it follows from (23), (24) and (25) that

$$\begin{aligned} & \sum_{s=1}^t \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_X\left(M_0^{(i,s)}, M_0^{(j,s-1)}\right)^p \\ & \qquad \geqslant \frac{2t}{(2\Gamma)^p} \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_X(y_i,y_j)^p - \sum_{j=1}^n \pi_j d_X\left(M_0^{(j,t)}, x_j\right)^p. \end{aligned} \tag{27}$$

A substitution of (27) into (22) now yields the following estimate.

$$\begin{aligned} & t\sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_X(y_i,y_j)^p \\ & \leqslant \frac{(2\Gamma K)^p}{2} \sum_{i=1}^n \sum_{j=1}^n \pi_i (A^t)_{ij} d_X(x_i,x_j)^p \\ & \quad - 2^{p-1}\Gamma^p\left(K^p-1\right) \sum_{j=1}^n \pi_j d_X\left(M_0^{(j,t)}, x_j\right)^p \\ & \leqslant \frac{(2\Gamma K)^p}{2} \sum_{i=1}^n \sum_{j=1}^n \pi_i (A^t)_{ij} d_X(x_i,x_j)^p. \end{aligned} \tag{28}$$

By combining (28) and Lemma 3.1 we deduce that

$$t\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i a_{ij} d_X(y_i,y_j)^p \leqslant (4\Gamma K)^p \sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i \mathscr{A}_t(A)_{ij} d_X(x_i,x_j)^p. \quad (29)$$

Next, we need to bound the quantity $\sum_{i=1}^{n}\pi_i d_X(x_i,y_i)^p$. We first claim that for every $i\in\{1,\ldots,n\}$, every $s\in\{0,\ldots,t\}$ and every $z\in X$ we have

$$d_X\left(z, M_0^{(i,t)}\right)^p \leqslant \sum_{j=1}^{n}(A^s)_{ij} d_X\left(z, M_0^{(j,t-s)}\right)^p. \quad (30)$$

The proof of (30) is by induction on s. For $s=0$ the desired inequality (30) holds as equality. Assuming the validity of (30) for some $s\in\{0,\ldots,t-1\}$, and recalling (19) and (20), observe that for every $j\in\{1,\ldots,n\}$ we have

$$M_0^{(j,t-s)} = \mathfrak{B}\left(\sum_{k=1}^{n} a_{jk}\delta_{M_0^{(k,t-s+1)}}\right).$$

Consequently, it follows from (4) that

$$d_X\left(z, M_0^{(j,t-s)}\right)^p \leqslant \sum_{k=1}^{n} a_{jk} d_X\left(z, M_0^{(k,t-s+1)}\right)^p. \quad (31)$$

So,

$$\begin{aligned} d_X\left(z, M_0^{(i,t)}\right)^p &\overset{(30)\wedge(31)}{\leqslant} \sum_{j=1}^{n}(A^s)_{ij}\sum_{k=1}^{n} a_{jk} d_X\left(z, M_0^{(k,t-s+1)}\right)^p \\ &= \sum_{k=1}^{n}(A^{s+1})_{ik} d_X\left(z, M_0^{(k,t-s+1)}\right)^p, \end{aligned}$$

thus completing the inductive verification of (30).

When $s=t$ inequality (30) becomes

$$d_X\left(z, M_0^{(i,t)}\right)^p \leqslant \sum_{j=1}^{n}(A^t)_{ij} d_X\left(z, x_j\right)^p. \quad (32)$$

Hence, for every $i\in\{1,\ldots,n\}$,

$$\begin{aligned} d_X(x_i,y_i)^p &\overset{(26)}{=} d_X\left(x_i, \mathfrak{B}\left(\frac{1}{t}\sum_{s=1}^{t}\delta_{M_0^{(i,s)}}\right)\right)^p \overset{(4)}{\leqslant} \frac{1}{t}\sum_{s=1}^{t} d_X\left(x_i, M_0^{(i,s)}\right)^p \\ &\overset{(32)}{\leqslant} \frac{1}{t}\sum_{s=1}^{t}\sum_{j=1}^{n}(A^s)_{ij} d_X\left(x_i,x_j\right)^p \overset{(17)}{=} \sum_{j=1}^{n}\mathscr{A}_t(A)_{ij} d_X\left(x_i,x_j\right)^p. \end{aligned}$$

Consequently,

$$\sum_{i=1}^{n} \pi_i d_X(x_i, y_i)^p \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i \mathscr{A}_t(A)_{ij} d_X\left(x_i, x_j\right)^p . \qquad (33)$$

By summing (29) and (33) we conclude that the desired inequality (2) holds true with $y_1, \ldots, y_n$ chosen as in (26) and $N^p = (4\Gamma K)^p + 1$. This completes the proof of Theorem 1.5. $\square$

4. Proof of Theorem 1.10

Theorem 1.10 is a consequence of Lemma 4.2 and Lemma 4.3 below. These lemmas are meaningful without the restriction $K = 1$ of Theorem 1.10: this more stringent assumption will only be used later, when we combine Lemma 4.2 and Lemma 4.3 to deduce Theorem 1.10.

In order to state our results we need to first introduce a small amount of notation. Given a metric space (X, d_X) and $p \in [1, \infty)$, for every $n \in \mathbb{N}$ we denote by $L_p^n(X)$ the space of all function $f : \{1, \ldots, n\} \to X$, equipped with the metric

$$\forall\, f, g \in L_p^n(X), \qquad d_{L_p^n(X)}(f, g) \stackrel{\mathrm{def}}{=} \left(\frac{1}{n}\sum_{i=1}^{n} d_X(f(i), g(i))^p\right)^{1/p}.$$

Suppose that $\mathfrak{B} : \mathscr{P}_X \to X$ is a barycenter map. In what follows it will be convenient to use the following slight abuse of notation: for every $f : \{1, \ldots, n\} \to X$ write

$$\mathfrak{B}(f) \stackrel{\mathrm{def}}{=} \mathfrak{B}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{f(i)}\right).$$

For a symmetric stochastic matrix $A = (a_{ij}) \in M_n(\mathbb{R})$ define a mapping $A \otimes I_X^n : L_p^n(X) \to L_p^n(X)$ by setting for every $i \in \{1, \ldots n\}$ and $f : \{1, \ldots, n\} \to X$,

$$(A \otimes I_X^n)(f)(i) \stackrel{\mathrm{def}}{=} \mathfrak{B}\left(\sum_{j=1}^{n} a_{ij}\delta_{f(j)}\right).$$

We warn that, unlike in the setting of Banach space valued mappings, given two symmetric stochastic matrices $A, B \in M_n(\mathbb{R})$ the composition $(A \otimes I_X^n) \circ (B \otimes I_X^n)$ need not be of the form $C \otimes I_X^n$ for some symmetric stochastic matrix $C \in M_n(\mathbb{R})$, and in particular the identity $(A \otimes I_X^n) \circ (B \otimes I_X^n) = (AB) \otimes I_X$ need not hold true.

Definition 4.1. Fix $p \in [1, \infty)$ and a metric space (X, d_X) equipped with a barycenter map $\mathfrak{B} : \mathscr{P}_X \to X$. Given $T : L_p^n(X) \to L_p^n(X)$, define $\lambda_p(T) \in [0,\infty]$ to be the infimum over those $\lambda \in (0,\infty]$ for which every $f \in L_p^n(X)$ satisfies

$$d_{L_p^n(X)}\left(T(f), \mathfrak{B}\left(T(f)\right)\right) \leqslant \lambda \cdot d_{L_p^n(X)}\left(f, \mathfrak{B}(f)\right).$$

Lemma 4.2 below relates the quantities $\gamma_+(A^t, d_X^p)$ and $\lambda_p\left(A \otimes I_X^n\right)$. Note that it assumes that (X, d_X) is p-barycentric with constant K, but K does not appear in the conclusion (34). The reason for this is that the proof of Lemma 4.2 uses a weaker version of (4) in which the rightmost term on the left hand side of (4) is dropped, i.e., the assumption that (X, d_X) is p-barycentric with constant K is used in Lemma 4.2 only through the requirement that every $\mu \in \mathscr{P}_X$ satisfies

$$\forall\, x \in X, \qquad d_X(x, \mathfrak{B}(\mu))^p \leqslant \int_X d_X(x,y)^p d\mu(y).$$

Lemma 4.2. *Fix $p, K, \Gamma \in [1,\infty)$ and $n, t \in \mathbb{N}$. Suppose that (X, d_X) is a metric space that is both W_p barycentric with constant Γ and p-barycentric with constant K. Let $A = (a_{ij}) \in M_n(\mathbb{R})$ be a symmetric stochastic matrix such that $\lambda_p\left(A \otimes I_X^n\right) < 1$. Then*

$$\gamma_+(A^t, d_X^p) \leqslant \left(\Gamma + \frac{4(\Gamma+1)}{1 - \lambda_p\left(A \otimes I_X^n\right)^{2t}}\right)^p. \tag{34}$$

In what follows, for $p, K \in [1,\infty)$ we denote by $\beta_p(K)$ the unique $\beta \in [1,\infty)$ satisfying

$$\beta^p + K^p(\beta-1)^p = K^p. \tag{35}$$

Thus in particular,

$$\beta_2(K) = \frac{2K^2}{K^2+1}.$$

Observe that $\beta_p(K) \in [1, \min\{K, 2\}]$ and $\beta_p(K) = 1$ if and only if $K = 1$. We also have

$$\beta_p(K) = \max_{\beta \in [1,2]} \min\left\{\beta, K\left(1-(\beta-1)^p\right)^{1/p}\right\}. \tag{36}$$

To verify (36) note that the function $\beta \mapsto K\left(1-(\beta-1)^p\right)^{1/p}$ decreases from K to 0 on $[1,2]$, so the maximum that appears in (36) is attained when $\beta = K\left(1-(\beta-1)^p\right)^{1/p}$, or, due to (35), when $\beta = \beta_p(K)$. An equivalent way to state (36) is that for every $a \in [0,\infty)$ we have

$$0 \leqslant b \leqslant a \implies \min\left\{a+b, K\left(a^p - b^p\right)^{1/p}\right\} \leqslant \beta_p(K)a. \tag{37}$$

To deduce (37) from (36) simply write $b = (\beta-1)a$ for some $\beta \in [1,2]$.

Lemma 4.3. *Fix $p, K \in [1,\infty)$ and $n \in \mathbb{N}$. Suppose that (X, d_X) is a p-barycentric metric space with constant K. Then every symmetric stochastic matrix $A = (a_{ij}) \in M_n(\mathbb{R})$ satisfies*

$$\lambda_p\left(A \otimes I_X^n\right) \leqslant \beta_p(K)\left(\frac{K^{2p}\gamma_+(A, d_X^p) - 1}{K^{2p}\gamma_+(A, d_X^p) + K^p}\right)^{1/p}.$$

Assuming the validity of Lemma 4.2 and Lemma 4.3 for the moment, we now show how they imply Theorem 1.10.

Proof of Theorem 1.10. We are now assuming that (X, d_X) is both W_p barycentric with constant Γ and p-barycentric with constant $K = 1$. Under the latter assumption the conclusion of Lemma 4.3 becomes

$$\lambda_p\left(A \otimes I_X^n\right) \leqslant \left(\frac{\gamma_+(A, d_X^p) - 1}{\gamma_+(A, d_X^p) + 1}\right)^{1/p}. \tag{38}$$

Thus in particular $\lambda_p\left(A \otimes I_X^n\right) < 1$, so we may use Lemma 4.2 in conjunction with (38) to obtain the estimate

$$\gamma_+(A^t, d_X^p)^{1/p} \lesssim \frac{\Gamma}{1 - \left(\frac{\gamma_+(A, d_X^p) - 1}{\gamma_+(A, d_X^p) + 1}\right)^{2t/p}}.$$

It now remains to note the elementary inequality

$$\forall\, p, \gamma, t \in [1,\infty), \qquad 1 - \left(\frac{\gamma - 1}{\gamma + 1}\right)^{2t/p} \gtrsim \min\left\{1, \frac{t}{p\gamma}\right\},$$

which follows by considering the cases $t \geqslant p\gamma$ and $t \leqslant p\gamma$ separately. $\square$

We now proceed to prove Lemma 4.2 and Lemma 4.3. To this end, given $n, t \in \mathbb{N}$ and $T : L_p^n(X) \to L_p^n(X)$, we denote the t-fold iterate of T by $T^{[t]}$, i.e.,

$$T^{[t]} \stackrel{\mathrm{def}}{=} \underbrace{T \circ \ldots \circ T}_{t \text{ times}}.$$

We also use the convention that $T^{[0]}$ is the identity mapping. If X is a Banach space then $(A \otimes I_X^n)^{[t]} = A^t \otimes I_X^n$, but this need not hold true when X is not a Banach space. Observe that a direct iterative application of Definition 4.1 implies that

$$\lambda_p\left(T^{[t]}\right) \leqslant \lambda_p(T)^t. \tag{39}$$

Lemma 4.4. *Fix $p, K \in [1,\infty)$. Suppose that (X, d_X) is a p-barycentric metric space with constant K. Then for every $n, t \in \mathbb{N}$, every symmetric stochastic matrix $A = (a_{ij}) \in M_n(\mathbb{R})$ and every $f \in L_p^n(X)$,*

$$d_{L_p^n(X)}\left(f, (A \otimes I_X^n)^{[t]}(f)\right)^p \leqslant \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n (A^t)_{ij} d_X(f(i), f(j))^p. \tag{40}$$

Proof. We will prove by induction on t that if $B = (b_{ij}) \in M_n(\mathbb{R})$ has nonnegative entries then

$$\sum_{i=1}^n\sum_{j=1}^n b_{ij} d_X\left(f(i), (A \otimes I_X^n)^{[t]}(f)(j)\right)^p \leqslant \sum_{i=1}^n\sum_{j=1}^n (BA^t)_{ij} d_X(f(i), f(j))^p. \tag{41}$$

The desired inequality (40) is then the special case of (41) when B is the identity matrix.

(41) holds as equality when $t = 0$, so assume inductively that (41) holds true for some $t \in \mathbb{N} \cup \{0\}$. Fix $i, j \in \{1, \ldots, n\}$ and consider the probability measure $\mu_j \in \mathscr{P}_X$ given by

$$\mu_j \stackrel{\mathrm{def}}{=} \sum_{k=1}^n a_{jk} \delta_{(A\otimes I_X^n)^{[t]}(f)(k)}.$$

Then $\mathfrak{B}(\mu_j) = (A \otimes I_X^n)^{[t+1]}(f)(j)$. By (4) with $\mu = \mu_j$ and $x = f(i)$,

$$d_X\left(f(i), (A \otimes I_X^n)^{[t+1]}(f)(j)\right)^p \leqslant \sum_{k=1}^n a_{jk} d_X\left(f(i), (A \otimes I_X^n)^{[t]}(f)(k)\right)^p. \tag{42}$$

Hence,

$$\begin{aligned}
&\sum_{i=1}^n\sum_{j=1}^n b_{ij} d_X\left(f(i), (A \otimes I_X^n)^{[t+1]}(f)(j)\right)^p \\
&\leqslant \sum_{i=1}^n\sum_{j=1}^n b_{ij}\sum_{k=1}^n a_{jk} d_X\left(f(i), (A \otimes I_X^n)^{[t]}(f)(k)\right)^p \qquad (43)\\
&= \sum_{i=1}^n\sum_{k=1}^n (BA)_{ik} d_X\left(f(i), (A \otimes I_X^n)^{[t]}(f)(k)\right)^p \\
&\leqslant \sum_{i=1}^n\sum_{k=1}^n (BA^{t+1})_{ik} d_X(f(i), f(k))^p, \qquad (44)
\end{aligned}$$

where in (43) uses (42) and (44) uses the inductive hypothesis. □

Lemma 4.5. *Fix $p, \Gamma \in [1,\infty)$. Suppose that (X, d_X) is a W_p barycentric metric space with constant Γ. Then for every $f, g \in L_p^n(X)$,*

$$\left| d_{L_p^n(X)}(f, \mathfrak{B}(f)) - d_{L_p^n(X)}(g, \mathfrak{B}(g)) \right| \leqslant (\Gamma+1) d_{L_p^n(X)}(f,g). \tag{45}$$

Proof. By (3) we have

$$d_X(\mathfrak{B}(f), \mathfrak{B}(g)) \leqslant \Gamma \cdot d_{L_p^n(X)}(f,g). \tag{46}$$

By the triangle inequality in $L_p^n(\mathbb{R})$ we have

$$\begin{aligned} &\left| d_{L_p^n(X)}(f, \mathfrak{B}(f)) - d_{L_p^n(X)}(g, \mathfrak{B}(g)) \right| \\ &\qquad \leqslant \left(\frac{1}{n} \sum_{i=1}^{n} \left| d_X(f(i), \mathfrak{B}(f)) - d_X(g(i), \mathfrak{B}(g)) \right|^p \right)^{1/p}. \end{aligned} \tag{47}$$

For every $i \in \{1, \ldots, n\}$ the triangle inequality in (X, d_X) implies that

$$|d_X(f(i), \mathfrak{B}(f)) - d_X(g(i), \mathfrak{B}(g))| \leqslant d_X(f(i), g(i)) + d_X(\mathfrak{B}(f), \mathfrak{B}(g)).$$

In combination with (47) and another application of the triangle inequality in $L_p^n(\mathbb{R})$ we deduce that

$$\left| d_{L_p^n(X)}(f, \mathfrak{B}(f)) - d_{L_p^n(X)}(g, \mathfrak{B}(g)) \right| \leqslant d_{L_p^n(X)}(f,g) + d_X(\mathfrak{B}(f), \mathfrak{B}(g)).$$

Due to (46), this implies the desired inequality (45). □

Lemma 4.6. *Fix $p, K, \Gamma \in [1,\infty)$ and $n, t \in \mathbb{N}$. Suppose that (X, d_X) is a metric space that is both W_p barycentric with constant Γ and p-barycentric with constant K. Let $A \in M_n(\mathbb{R})$ be a symmetric stochastic matrix such that $\lambda_p\left((A \otimes I_X^n)^{[2t]}\right) < 1$. Then*

$$\gamma_+(A^t, d_X^p) \leqslant \left(\Gamma + \frac{4(\Gamma+1)}{1 - \lambda_p\left((A \otimes I_X^n)^{[2t]}\right)} \right)^p. \tag{48}$$

Proof. For every $f, g : \{1, \ldots, n\} \to X$ we have

$$\begin{aligned} &\left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_X(f(i), g(j))^p \right)^{1/p} \\ &\qquad \leqslant d_X(\mathfrak{B}(f), \mathfrak{B}(g)) + d_{L_p^n(X)}(f, \mathfrak{B}(f)) + d_{L_p^n(X)}(g, \mathfrak{B}(g)). \end{aligned} \tag{49}$$

We proceed to bound each of the terms on the right hand side of (49) separately.

First, define $\mu_f, \mu_g \in \mathscr{P}_X$ by

$$\mu_f \stackrel{\mathrm{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \delta_{f(i)} \qquad \text{and} \qquad \mu_g \stackrel{\mathrm{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \delta_{g(i)}.$$

Then $\mathfrak{B}(f) = \mathfrak{B}(\mu_f)$ and $\mathfrak{B}(g) = \mathfrak{B}(\mu_g)$. If we define $\pi \in \mathscr{P}_{X\times X}$ by

$$\pi = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(A^t)_{ij}\delta_{(f(i),g(j))},$$

then, since A^t is a symmetric stochastic matrix, $\pi \in \Pi(\mu_f, \mu_g)$, i.e., π is a coupling of μ_f and μ_g. It therefore follows from (3) that

$$d_X(\mathfrak{B}(f), \mathfrak{B}(g)) \leqslant \Gamma \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(A^t)_{ij} d_X(f(i), g(j))^p\right)^{1/p}. \tag{50}$$

Next, since

$$\begin{aligned} d_{L_p^n(X)} &\left((A\otimes I_X^n)^{[2t]}(f), \mathfrak{B}\left((A\otimes I_X^n)^{[2t]}(f)\right)\right) \\ &\leqslant \lambda_p\left((A\otimes I_X^n)^{[2t]}\right) d_{L_p^n(X)}(f, \mathfrak{B}(f)), \end{aligned}$$

and $\lambda_p\left((A\otimes I_X^n)^{[2t]}\right) < 1$, we have

$$\frac{d_{L_p^n(X)}(f, \mathfrak{B}(f)) - d_{L_p^n(X)}\left((A\otimes I_X^n)^{[2t]}(f), \mathfrak{B}\left((A\otimes I_X^n)^{[2t]}(f)\right)\right)}{1-\lambda_p\left((A\otimes I_X^n)^{[2t]}\right)} \geqslant d_{L_p^n(X)}(f, \mathfrak{B}(f)). \tag{51}$$

Now,

$$\begin{aligned} &d_{L_p^n(X)}(f, \mathfrak{B}(f)) - d_{L_p^n(X)}\left((A\otimes I_X^n)^{[2t]}(f), \mathfrak{B}\left((A\otimes I_X^n)^{[2t]}(f)\right)\right) \\ &\leqslant (\Gamma+1) d_{L_p^n(X)}\left(f, (A\otimes I_X^n)^{[2t]}(f)\right) \qquad (52) \\ &\leqslant (\Gamma+1)\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(A^{2t})_{ij} d_X(f(i), f(j))^p\right)^{1/p}. \qquad (53) \end{aligned}$$

where (52) uses Lemma 4.5 and (53) uses Lemma 4.4.

Observe that

$$
\begin{aligned}
&\sum_{i=1}^n\sum_{j=1}^n (A^{2t})_{ij} d_X(f(i),f(j))^p\\
&= \sum_{i=1}^n\sum_{j=1}^n \left(\sum_{k=1}^n (A^t)_{ik}(A^t)_{kj}\right) d_X(f(i),f(j))^p\\
&\leqslant 2^{p-1}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n (A^t)_{ik}(A^t)_{kj}\big(d_X(f(i),g(k))^p + d_X(g(k),f(j))^p\big)\\
&= 2^p \sum_{i=1}^n\sum_{j=1}^n (A^t)_{ij} d_X(f(i),g(j))^p. \qquad (54)
\end{aligned}
$$

By combining (51) with (53) and (54) we have,

$$
\begin{aligned}
&d_{L_p^n(X)}(f,\mathfrak{B}(f))\\
&\quad\leqslant \frac{2(\Gamma+1)}{1-\lambda_p\left((A\otimes I_X^n)^{[2t]}\right)}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n (A^t)_{ij} d_X(f(i),g(j))^p\right)^{1/p}. \qquad (55)
\end{aligned}
$$

By the same reasoning,

$$
\begin{aligned}
&d_{L_p^n(X)}(g,\mathfrak{B}(g))\\
&\quad\leqslant \frac{2(\Gamma+1)}{1-\lambda_p\left((A\otimes I_X^n)^{[2t]}\right)}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n (A^t)_{ij} d_X(f(i),g(j))^p\right)^{1/p}. \qquad (56)
\end{aligned}
$$

Finally, a substitution of (50), (55) and (56) into (49) shows that

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(f(i),g(j))^p\\
&\leqslant \left(\Gamma + \frac{4(\Gamma+1)}{1-\lambda_p\left((A\otimes I_X^n)^{[2t]}\right)}\right)^p\cdot\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n (A^t)_{ij} d_X(f(i),g(j))^p. \qquad (57)
\end{aligned}
$$

Since (57) holds true for every $f,g:\{1,\ldots,n\}\to X$, the desired estimate (48) now follows by recalling the definition of $\gamma_+(A,d_X^p)$. $\square$

Proof of Lemma 4.2. The desired estimate (34) is a consequence of (39) and Lemma 4.6. $\square$

We now proceed to prove Lemma 4.3. Recalling the definition of $\beta_p(K)$ in (35), we first establish the following estimate.

Lemma 4.7. *Fix $p, K \in [1,\infty)$. Suppose that (X, d_X) is a p-barycentric metric space with constant K. Then for every $n \in \mathbb{N}$, every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ and every $f \in L_p^n(X)$ we have*

$$d_{L_p^n(X)}\left((A\otimes I_X^n)(f), \mathfrak{B}\left((A\otimes I_X^n)(f)\right)\right) \leqslant \beta_p(K)\cdot d_{L_p^n(X)}\left((A\otimes I_X^n)(f), \mathfrak{B}(f)\right). \quad (58)$$

Proof. Write

$$a \stackrel{\mathrm{def}}{=} d_{L_p^n(X)}\left((A\otimes I_X^n)(f), \mathfrak{B}(f)\right), \quad (59)$$

and

$$b \stackrel{\mathrm{def}}{=} d_{L_p^n(X)}\left(\mathfrak{B}(f), \mathfrak{B}\left((A\otimes I_X^n)(f)\right)\right). \quad (60)$$

Then by the triangle inequality in $L_p^n(X)$,

$$d_{L_p^n(X)}\left((A\otimes I_X^n)(f), \mathfrak{B}\left((A\otimes I_X^n)(f)\right)\right) \leqslant a+b. \quad (61)$$

Next, define $\nu \in \mathscr{P}_X$ by

$$\nu \stackrel{\mathrm{def}}{=} \frac{1}{n}\sum_{i=1}^n \delta_{(A\otimes I_X^n)(f)(i)}. \quad (62)$$

Then

$$\mathfrak{B}(\nu) = \mathfrak{B}\left((A\otimes I_X^n)(f)\right), \quad (63)$$

and therefore

$$d_{L_p^n(X)}\left((A\otimes I_X^n)(f), \mathfrak{B}\left((A\otimes I_X^n)(f)\right)\right) = \left(\int_X d_X(y,\mathfrak{B}(\nu))^p d\nu(y)\right)^{1/p}. \quad (64)$$

It also follows from (59) and (62) that

$$a = \left(\int_X d_X(y,\mathfrak{B}(f))^p d\nu(y)\right)^{1/p}. \quad (65)$$

By combining (60), (63), (64) and (65), an application of (4) to the measure $\mu=\nu$ with $x=\mathfrak{B}(f)$ yields the estimate

$$b^p + \frac{1}{K^p}\cdot d_{L_p^n(X)}\left((A\otimes I_X^n)(f), \mathfrak{B}\left((A\otimes I_X^n)(f)\right)\right)^p \leqslant a^p. \quad (66)$$

The desired estimate (58) now follows by combining (61) and (66) with (37). $\square$

Lemma 4.8. *Fix $p, K \in [1,\infty)$. Suppose that (X, d_X) is a p-barycentric metric space with constant K. Then for every $n \in \mathbb{N}$, every symmetric stochastic matirx $A = (a_{ij}) \in M_n(\mathbb{R})$ and every $f \in L_p^n(X)$ we have*

$$d_{L_p^n(X)}\left((A \otimes I_X^n)(f), \mathfrak{B}(f)\right) \leqslant \left(\frac{K^{2p}\gamma_+(A, d_X^p) - 1}{K^{2p}\gamma_+(A, d_X^p) + K^p}\right)^{1/p} d_{L_p^n(X)}\left(f, \mathfrak{B}(f)\right). \quad (67)$$

Proof. For every $i \in \{1,\ldots,n\}$ define $\nu_i \in \mathscr{P}_X$ by

$$\nu_i \stackrel{\mathrm{def}}{=} \sum_{i=1}^n a_{ij}\delta_{f(j)}. \quad (68)$$

Thus

$$(A \otimes I_X^n)(f)(i) = \mathfrak{B}(\nu_i),$$

$$\int_X d_X(y, \mathfrak{B}(\nu_i))^p d\nu_i(y) = \sum_{j=1}^n a_{ij} d_X\left(f(j), (A \otimes I_X^n)(f)(i)\right)^p,$$

and

$$\int_X d_X(y, \mathfrak{B}(f))^p d\nu_i(y) = \sum_{j=1}^n a_{ij} d_X(f(j), \mathfrak{B}(f))^p.$$

An application of (4) with $\mu = \nu_i$ and $x = \mathfrak{B}(f)$ therefore implies that

$$d_X\left((A \otimes I_X^n)(f)(i), \mathfrak{B}(f)\right)^p + \frac{1}{K^p}\sum_{j=1}^n a_{ij} d_X\left(f(j), (A \otimes I_X^n)(f)(i)\right)^p \leqslant \sum_{j=1}^n a_{ij} d_X(f(j), \mathfrak{B}(f))^p. \quad (69)$$

By averaging (69) over $i \in \{1,\ldots,n\}$ we conclude that

$$d_{L_p^n(X)}\left((A \otimes I_X^n)(f), \mathfrak{B}(f)\right)^p + \frac{1}{nK^p}\sum_{i=1}^n\sum_{j=1}^n a_{ij} d_X\left(f(j), (A \otimes I_X^n)(f)(i)\right)^p \leqslant \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij} d_X(f(j), \mathfrak{B}(f))^p = d_{L_p^n(X)}\left(f, \mathfrak{B}(f)\right)^p. \quad (70)$$

Next, the definition of $\gamma_+(A, d_X^p)$ implies that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} d_X\left(f(j), (A\otimes I_X^n)(f)(i)\right)^p \\ \geqslant \frac{1}{\gamma_+(A, d_X^p)}\cdot \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X\left(f(j), (A\otimes I_X^n)(f)(i)\right)^p. \quad (71)$$

For every $i\in\{1,\ldots,n\}$, an application of (4) with $\mu = \frac{1}{n}\sum_{j=1}^n \delta_{f(j)}$ and $x = (A\otimes I_X^n)(f)(i)$ implies the estimate

$$d_X\left((A\otimes I_X^n)(f)(i), \mathfrak{B}(f)\right)^p + \frac{1}{K^p}\cdot d_{L_p^n(X)}(f, \mathfrak{B}(f))^p \\ \leqslant \frac{1}{n}\sum_{j=1}^{n} d_X\left(f(j), (A\otimes I_X^n)(f)(i)\right)^p. \quad (72)$$

By averaging (72) over $i\in\{1,\ldots,n\}$ we see that

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X\left(f(j), (A\otimes I_X^n)(f)(i)\right)^p \\ \geqslant d_{L_p^n(X)}\left((A\otimes I_X^n)(f), \mathfrak{B}(f)\right)^p + \frac{1}{K^p}\cdot d_{L_p^n(X)}(f, \mathfrak{B}(f))^p. \quad (73)$$

By substituting (73) into (71), and plugging the resulting estimate into (70), we conclude that

$$\left(1+\frac{1}{K^p\gamma_+(A, d_X^p)}\right) d_{L_p^n(X)}\left((A\otimes I_X^n)(f), \mathfrak{B}(f)\right)^p \\ \leqslant \left(1-\frac{1}{K^{2p}\gamma_+(A, d_X^p)}\right) d_{L_p^n(X)}(f, \mathfrak{B}(f))^p,$$

which simplifies to give the desired inequality (67). □

Proof of Lemma 4.3. Simply combine Lemma 4.7 and Lemma 4.8. □

5. Proof of Theorem 1.11

Lemma 5.1 below plays an important role in our proof of Theorem 1.11. It was proved by the second named author in collaboration with M. Csörnyei (2001); we thank her for letting us include it here.

Lemma 5.1. *Fix $m, n\in\mathbb{N}$ and $p\in[1,\infty)$. Let $B=(b_{ij})\in M_{n\times m}(\mathbb{R})$ and $C=(c_{ij})\in M_n(\mathbb{R})$ be stochastic matrices (of dimensions n by m and n by n, respectively). Fix $\pi\in\Delta^{n-1}$ and suppose that C is* *reversible relative to π. Then for every metric space (X,d_X) and every $z_1,\ldots,z_m\in X$ there exist $w_1,\ldots,w_n\in X$ such that*

$$\max\left\{\sum_{i=1}^n\sum_{r=1}^m \pi_i b_{ir} d_X(w_i,z_r)^p,\sum_{i=1}^n\sum_{j=1}^n \pi_i c_{ij} d_X(w_i,w_j)^p\right\}\leqslant 3^p\sum_{r=1}^m\sum_{s=1}^m (B^*D_\pi CB)_{rs} d_X(z_r,z_s)^p, \quad (74)$$

where

$$D_\pi=\begin{pmatrix}\pi_1 & 0 & \ldots & 0\\ 0 & \pi_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \ldots & 0 & \pi_n\end{pmatrix}\in M_n(\mathbb{R}). \quad (75)$$

Proof. Let $f:\{z_1,\ldots,z_m\}\to \ell_\infty^m$ be an isometric embedding (e.g. one can take $f(z)=\sum_{r=1}^m d_X(z,z_r)e_r$, where $\{e_r\}_{r=1}^m$ is the standard basis of $\mathbb{R}^m$). Define $y_1,\ldots,y_n\in\ell_\infty^m$ by

$$\forall\, i\in\{1,\ldots,n\},\quad y_i\stackrel{\mathrm{def}}{=}\sum_{r=1}^m b_{ir}f(z_r). \quad (76)$$

Next, for every $i\in\{1,\ldots,n\}$ choose $w_i\in\{z_1,\ldots,z_m\}$ such that

$$\|y_i-f(w_i)\|_\infty=\min_{z\in\{z_1,\ldots,z_m\}}\|y_i-f(z)\|_\infty. \quad (77)$$

By the triangle inequality, for every $i,j\in\{1,\ldots,m\}$ we have

$$d_X(w_i,w_j)^p=\|f(w_i)-f(w_j)\|_\infty^p\leqslant 3^{p-1}\left(\|f(w_i)-y_i\|_\infty^p+\|y_i-y_j\|_\infty^p+\|y_j-f(w_j)\|_\infty^p\right).$$

Consequently, using the stochasticity of C and its reversibility relative to π,

$$\sum_{i=1}^n\sum_{j=1}^n \pi_i c_{ij} d_X(w_i,w_j)^p\leqslant 3^{p-1}\sum_{i=1}^n\sum_{j=1}^n \pi_i c_{ij}\|y_i-y_j\|_\infty^p+2\cdot 3^{p-1}\sum_{i=1}^n \pi_i\|y_i-f(w_i)\|_\infty^p. \quad (78)$$

Recalling (76), the convexity of the function $v \mapsto \|v\|_\infty^p$ implies that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i c_{ij}\|y_i-y_j\|_\infty^p \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i c_{ij}\sum_{r=1}^{m}\sum_{s=1}^{m} b_{ir}b_{js}\|f(z_r)-f(z_s)\|_\infty^p$$
$$=\sum_{r=1}^{m}\sum_{s=1}^{m}(B^*D_\pi CB)_{rs}d_X(z_r,z_s)^p. \quad (79)$$

Next, due to (77) and the fact that CB is a stochastic matrix,

$$\forall\, i\in\{1,\ldots,n\},\quad \|y_i-f(w_i)\|_\infty^p \leqslant \sum_{r=1}^{m}(CB)_{ir}\|y_i-f(z_r)\|_\infty^p.$$

Recalling (76) and using the convexity of the function $v \mapsto \|v\|_\infty^p$ once more, we deduce that

$$\sum_{i=1}^{n}\pi_i\|y_i-f(w_i)\|_\infty^p \leqslant \sum_{i=1}^{n}\pi_i\sum_{r=1}^{m}(CB)_{ir}\sum_{s=1}^{m} b_{is}\|f(z_s)-f(z_r)\|_\infty^p$$
$$=\sum_{r=1}^{m}\sum_{s=1}^{m}(B^*D_\pi CB)_{rs}d_X(z_r,z_s)^p. \quad (80)$$

A combination of (78), (79) and (80) now implies that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i c_{ij}d_X(w_i,w_j)^p \leqslant 3^p\sum_{r=1}^{m}\sum_{s=1}^{m}(B^*D_\pi CB)_{rs}d_X(z_r,z_s)^p. \quad (81)$$

Next, note that by the triangle inequality for every $i,j\in\{1,\ldots,n\}$ and $r\in\{1,\ldots,m\}$ we have

$$d_X(w_i,z_r)^p = \|f(w_i)-f(z_r)\|_\infty^p$$
$$\leqslant 3^{p-1}\left(\|f(w_i)-y_i\|_\infty^p+\|y_i-y_j\|_\infty^p+\|y_j-f(z_r)\|_\infty^p\right). \quad (82)$$

By multiplying inequality (82) by $\pi_i b_{ir}c_{ij}$, summing over $r\in\{1,\ldots,m\}$ and $i,j\in\{1,\ldots,n\}$, and using the stochasticity of B and C, we deduce that

$$\frac{1}{3^{p-1}}\sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir}d_X(w_i,z_r)^p \leqslant \sum_{i=1}^{n}\pi_i\|f(w_i)-y_i\|_\infty^p$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i c_{ij}\|y_i-y_j\|_\infty^p+\sum_{j=1}^{n}\sum_{r=1}^{m}(B^*D_\pi C)_{rj}\|y_j-f(z_r)\|_\infty^p. \quad (83)$$

Recalling (76) and using the convexity of the function $v \mapsto \|v\|_\infty^p$, we have

$$\sum_{j=1}^{n}\sum_{r=1}^{m}(B^*D_\pi C)_{rj}\|y_j - f(z_r)\|_\infty^p \leqslant \sum_{r=1}^{m}\sum_{s=1}^{m}(B^*D_\pi CB)_{rs}d_X(z_r,z_s)^p. \quad (84)$$

A combination of (83) with (79), (80), and (84) yields the estimate

$$\sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_X(w_i,z_r)^p \leqslant 3^p \sum_{r=1}^{m}\sum_{s=1}^{m}(B^*D_\pi CB)_{rs}d_X(z_r,z_s)^p,$$

which, due to (81), yields the desired estimate (74). □

The following lemma is a natural variant of [3, Lem. 1.1].

Lemma 5.2 (dual extension criterion). *Fix $p,\Lambda,\Gamma \in [1,\infty)$, an integer n, and $\varepsilon \in (0,1)$. Suppose that (X,d_X) and (Y,d_Y) are metric spaces with (Y,d_Y) being W_p barycentric with constant Γ. Fix $Z \subseteq X$, a Lipschitz function $f : Z \to Y$ and distinct $x_1,\ldots,x_n \in X$. Suppose that for every symmetric n by n matrix $H = (h_{ij})$ with nonnegative entries there exists $\Phi^H : \{x_1,\ldots,x_n\} \to Y$ with $\Phi^H|_{\{x_1,\ldots,x_n\}\cap Z} = f|_{\{x_1,\ldots,x_n\}\cap Z}$ such that*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij} d_Y\left(\Phi^H(x_i),\Phi^H(x_j)\right)^p \leqslant \Lambda^p \|f\|_{\mathrm{Lip}}^p \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij} d_X(x_i,x_j)^p.$$

Then there is $F : \{x_1,\ldots,x_n\} \to Y$ with $F|_{\{x_1,\ldots,x_n\}\cap Z} = f|_{\{x_1,\ldots,x_n\}\cap Z}$ and $\|F\|_{\mathrm{Lip}} \leqslant (1+\varepsilon)\Gamma\Lambda\|f\|_{\mathrm{Lip}}$.

Proof. We proceed via the following duality argument due to Ball [3] (which is itself inspired by the work of Maurey [40]), with a slight twist that brings in the assumption that Y is W_p barycentric.

Consider the following set of n by n symmetric matrices.

$$C \stackrel{\mathrm{def}}{=} \Big\{ (d_Y(\Phi(x_i),\Phi(x_j))^p) \in M_n(\mathbb{R}) : \Phi : \{x_1,\ldots,x_n\} \to Y \ \wedge\ \Phi|_{\{x_1,\ldots,x_n\}\cap Z} = f|_{\{x_1,\ldots,x_n\}\cap Z} \Big\}.$$

Let $D \subseteq M_n(\mathbb{R})$ be the set of all n by n symmetric matrices with nonnegative entries and define

$$E \stackrel{\mathrm{def}}{=} \overline{\mathrm{conv}\,(C+D)}.$$

For every $i,j \in \{1,\ldots,n\}$ write $t_{ij} \stackrel{\mathrm{def}}{=} \Lambda^p\|f\|_{\mathrm{Lip}}^p d_X(x_i,x_j)^p$. The assumption of Lemma 5.2 can be rephrased as

$$\sup_{H\in D}\inf_{M\in C}\sum_{i=1}^n\sum_{j=1}^n h_{ij}\left(m_{ij}-t_{ij}\right)\leqslant 0. \tag{85}$$

It follows that the matrix $T=(t_{ij})$ belongs to E, since otherwise by the separation theorem (Hahn-Banach) there would exist a symmetric matrix $H=(h_{ij})$ such that

$$\inf_{M\in E}\sum_{i=1}^n\sum_{j=1}^n h_{ij}m_{ij} > \sum_{i=1}^n\sum_{j=1}^n h_{ij}t_{ij} \tag{86}$$

Since $E\supseteq C+D$ it follows from (86) that the entries of H are non-negative, i.e., $H\in D$. Now (86) contradicts (85) since $E\supseteq C$.

Having shown that $T\in E$, we deduce that there exists $m\in\mathbb{N}$ and $\lambda_1,\ldots,\lambda_m\in[0,1]$ with $\sum_{k=1}^m\lambda_k=1$, and in addition there are $\Phi^1,\ldots,\Phi^m:\{x_1,\ldots,x_n\}\to Y$ with $\Phi^k|_{\{x_1,\ldots,x_n\}\cap Z}=f|_{\{x_1,\ldots,x_n\}\cap Z}$ for all $k\in\{1,\ldots,m\}$, such that for every $i,j\in\{1,\ldots,m\}$,

$$(1+\varepsilon)^p\Lambda^p\|f\|_{\mathrm{Lip}}^p d_X(x_i,x_j)^p-\sum_{k=1}^m\lambda_k d_Y\left(\Phi^k(x_i),\Phi^k(x_j)\right)^p\geqslant 0. \tag{87}$$

For every $i\in\{1,\ldots,n\}$ consider the finitely supported probability measure μ_i on Y given by

$$\mu_i\stackrel{\mathrm{def}}{=}\sum_{k=1}^m\lambda_k\delta_{\Phi^k(x_i)},$$

and set

$$F(x_i)\stackrel{\mathrm{def}}{=}\mathfrak{B}\left(\mu_i\right).$$

If $x_i\in Z$ then $\mu_i=\delta_{f(x_i)}$ and therefore $F(x_i)=f(x_i)$. In other words, $F|_{\{x_1,\ldots,x_n\}\cap Z}=f|_{\{x_1,\ldots,x_n\}\cap Z}$. Also, for every $i,j\in\{1,\ldots,n\}$ we have

$$\begin{aligned}
d_Y(F(x_i),F(x_j)) &\stackrel{(3)}{\leqslant} \Gamma W_p(\mu_i,\mu_j)\\
&\leqslant \Gamma\left(\sum_{k=1}^m\lambda_k d_Y(\Phi^k(x_i),\Phi^k(x_j))^p\right)^{1/p}\\
&\stackrel{(87)}{\leqslant} (1+\varepsilon)\Gamma\Lambda\|f\|_{\mathrm{Lip}}d_Y(x_i,x_j).
\end{aligned}$$

Thus $\|F\|_{\mathrm{Lip}}\leqslant(1+\varepsilon)\Gamma\Lambda\|f\|_{\mathrm{Lip}}$, as required. □

Proof of Theorem 1.11. Fix $m,n\in\mathbb{N}$. Take $x_1,\ldots,x_n\in X\setminus Z$ and $z_1,\ldots,z_m\in Z$. If $H=(h_{ij})\in M_{n+m}(\mathbb{R})$ is symmetric then write

$$H=\begin{pmatrix} U(H) & W(H)^* \\ W(H) & V(H)\end{pmatrix},$$

where $U(H)\in M_m(\mathbb{R})$, $V(H)\in M_n(\mathbb{R})$ and $W(H)\in M_{n\times m}(\mathbb{R})$. With this notation, define

$$\begin{aligned} R_H \stackrel{\mathrm{def}}{=} &\sum_{r=1}^{m}\sum_{s=1}^{m} U(H)_{rs} d_X(z_r,z_s)^p \\ &+2\sum_{i=1}^{n}\sum_{r=1}^{m} W(H)_{ir} d_X(x_i,z_r)^p+\sum_{i=1}^{n}\sum_{j=1}^{n} V(H)_{ij} d_X(x_i,x_j)^p,\end{aligned}$$

and for every $y_1,\ldots,y_n\in Y$,

$$\begin{aligned} L_H(y_1,\ldots,y_n) \stackrel{\mathrm{def}}{=} &\sum_{r=1}^{m}\sum_{s=1}^{m} U(H)_{rs} d_Y(f(z_r),f(z_s))^p \\ &+2\sum_{i=1}^{n}\sum_{r=1}^{m} W(H)_{ir} d_Y(y_i,f(z_r))^p+\sum_{i=1}^{n}\sum_{j=1}^{n} V(H)_{ij} d_Y(y_i,y_j)^p.\end{aligned}$$

Fix from now on $M\in(M_p(X),2M_p(X)]$ and $N\in(N_p(Y),2N_p(Y)]$. Due to Lemma 5.2 it suffices to show that for every symmetric matrix $H=(h_{ij})\in M_{n+m}(\mathbb{R})$ with nonnegative entries and for every $\delta\in(0,1)$ one can find $y_1,\ldots,y_n\in Y$ such that

$$L_H(y_1,\ldots,y_n)\leqslant \Lambda(R_H+\delta),$$

where

$$\Lambda\stackrel{\mathrm{def}}{=}\frac{18^p}{3}(N^p+1)M^p\|f\|_{\mathrm{Lip}}^p. \tag{88}$$

Since the Lipschitz condition of f on $\{z_1,\ldots,z_m\}$ implies that

$$\sum_{r=1}^{m}\sum_{s=1}^{m} U(H)_{rs} d_Y(f(z_r),f(z_s))^p\leqslant \|f\|_{\mathrm{Lip}}^p\sum_{r=1}^{m}\sum_{s=1}^{m} U(H)_{rs} d_X(z_r,z_s)^p,$$

it suffices to establish the existence of $y_1,\ldots,y_n\in Y$ that satisfy the inequality

$$\begin{aligned} &\frac{1}{\Lambda}\left(2\sum_{i=1}^{n}\sum_{r=1}^{m} W(H)_{ir} d_Y(y_i,f(z_r))^p+\sum_{i=1}^{n}\sum_{j=1}^{n} V(H)_{ij} d_Y(y_i,y_j)^p\right) \\ &\leqslant 2\sum_{i=1}^{n}\sum_{r=1}^{m} W(H)_{ir} d_X(x_i,z_r)^p+\sum_{i=1}^{n}\sum_{j=1}^{n} V(H)_{ij} d_X(x_i,x_j)^p+\delta. \end{aligned} \tag{89}$$

Fix $t \in [2,\infty)$ and $\varepsilon \in (0,1)$ that will be determined later. Note that the diagonal entries $\{V(H)_{ii}\}_{i=1}^n$ are irrelevant for the validity of (89), so we may assume from now on that $V(H)_{ii} = 0$ for all $i \in \{1,\ldots,n\}$.

Define $\pi \in \mathbb{R}^n$ by

$$\forall\, i \in \{1,\ldots,n\}, \quad \pi_i \stackrel{\mathrm{def}}{=} \frac{m\varepsilon + \sum_{r=1}^m W(H)_{ir}}{mn\varepsilon + \sum_{j=1}^n \sum_{r=1}^m W(H)_{jr}}. \tag{90}$$

Thus $\pi \in \Delta^{n-1}$. Next, define $B = (b_{ir}) \in M_{n\times m}(\mathbb{R})$ by setting for every $i \in \{1,\ldots,n\}$ and $r \in \{1,\ldots,m\}$,

$$b_{ir} \stackrel{\mathrm{def}}{=} \frac{\varepsilon + W(H)_{ir}}{m\varepsilon + \sum_{s=1}^m W(H)_{is}}. \tag{91}$$

Thus B is a stochastic matrix. Finally, define $A = (a_{ij}) \in M_n(\mathbb{R})$ by setting for $i \in \{1,\ldots,n\}$,

$$a_{ii} \stackrel{\mathrm{def}}{=} 1 - \frac{1}{t}\cdot \frac{\sum_{j=1}^n V(H)_{ij}}{m\varepsilon + \sum_{r=1}^m W(H)_{ir}}, \tag{92}$$

and for distinct $i,j \in \{1,\ldots,n\}$,

$$a_{ij} \stackrel{\mathrm{def}}{=} \frac{1}{t}\cdot \frac{V(H)_{ij}}{m\varepsilon + \sum_{r=1}^m W(H)_{ir}}. \tag{93}$$

The role of ε is only to ensure that the denominators that appear in (90), (91), (92) and (93) do not vanish. Fixing ε, we will assume from now on that t is sufficiently large so as to ensure that $a_{11},\ldots,a_{nn}$ are all nonnegative. Thus A is a stochastic matrix. Note also that since $V(H)$ is symmetric, an inspection of (90) and (93) reveals that A is reversible relative to π.

Set

$$\tau \stackrel{\mathrm{def}}{=} \left\lceil \frac{t}{2^p} \right\rceil. \tag{94}$$

By Lemma 5.1 applied with $C = \mathscr{A}_\tau(A)$ there exist $w_1,\ldots,w_n \in Y$ such that

$$\max\left\{ \sum_{i=1}^n \sum_{r=1}^m \pi_i b_{ir} d_Y(w_i, f(z_r))^p, \sum_{i=1}^n \sum_{j=1}^n \pi_i \mathscr{A}_\tau(A)_{ij} d_Y(w_i,w_j)^p \right\} \\ \leqslant 3^p \sum_{r=1}^m \sum_{s=1}^m \left(B^* D_\pi \mathscr{A}_\tau(A) B\right)_{rs} d_Y\left(f(z_r), f(z_s)\right)^p. \tag{95}$$

Since $N > N_p(Y)$ there exist $y_1, \ldots, y_n \in Y$ such that

$$\sum_{i=1}^n \pi_i d_Y(w_i, y_i)^p + \frac{t}{2^p} \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_Y(y_i, y_j)^p \leqslant N^p \sum_{i=1}^n \sum_{j=1}^n \pi_i \mathscr{A}_\tau(A)_{ij} d_Y(w_i, w_j)^p. \quad (96)$$

We will now show that the points $y_1, \ldots, y_n$ thus found satisfy the desired inequality (89).

To estimate the left hand side of (89) from above, denote

$$\theta \stackrel{\mathrm{def}}{=} mn\varepsilon + \sum_{i=1}^n \sum_{r=1}^m W(H)_{ir}, \quad (97)$$

and observe that due to (90), (91) and (97) we have

$$\forall (i, r) \in \{1, \ldots, n\} \times \{1, \ldots, m\}, \qquad W(H)_{ir} + \varepsilon = \theta \pi_i b_{ir}. \quad (98)$$

Similarly, due to (90), (92) and (93) we have

$$(i, j \in \{1, \ldots, n\} \ \wedge \ i \neq j) \implies V(H)_{ij} = \theta t \pi_i a_{ij}. \quad (99)$$

Hence,

$$\frac{1}{\theta}\left(2 \sum_{i=1}^n \sum_{r=1}^m W(H)_{ir} d_Y(y_i, f(z_r))^p + \sum_{i=1}^n \sum_{j=1}^n V(H)_{ij} d_Y(y_i, y_j)^p\right) \leqslant 2 \sum_{i=1}^n \sum_{r=1}^m \pi_i b_{ir} d_Y(y_i, f(z_r))^p + t \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_Y(y_i, y_j)^p. \quad (100)$$

By the triangle inequality, for every $i \in \{1, \ldots, n\}$ and $r \in \{1, \ldots, m\}$ we have

$$d_Y(y_i, f(z_r))^p \leqslant 2^{p-1} d_Y(y_i, w_i)^p + 2^{p-1} d_Y(w_i, f(z_r))^p.$$

Consequently,

$$\sum_{i=1}^n \sum_{r=1}^m \pi_i b_{ir} d_Y(y_i, f(z_r))^p \leqslant 2^{p-1} \sum_{i=1}^n \pi_i d_Y(y_i, w_i)^p + 2^{p-1} \sum_{i=1}^n \sum_{r=1}^m \pi_i b_{ir} d_Y(w_i, f(z_r))^p. \quad (101)$$

We can therefore bound the right hand side of (100) as follows.

$$
\begin{aligned}
&2\sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(y_i,f(z_r))^p + t\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i a_{ij} d_Y(y_i,y_j)^p\\
&\overset{(101)}{\leqslant} 2^p\sum_{i=1}^{n}\pi_i d_Y(y_i,w_i)^p + t\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i a_{ij} d_Y(y_i,y_j)^p\\
&\quad + 2^p\sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(w_i,f(z_r))^p\\
&\overset{(96)}{\leqslant} (2N)^p\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i \mathscr{A}_\tau(A)_{ij} d_Y(w_i,w_j)^p + 2^p\sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(w_i,f(z_r))^p\\
&\overset{(95)}{\leqslant} 6^p\left(N^p+1\right)\sum_{r=1}^{m}\sum_{s=1}^{m}\left(B^*D_\pi\mathscr{A}_\tau(A)B\right)_{rs} d_Y(f(z_r),f(z_s))^p\\
&\leqslant 6^p\left(N^p+1\right)\|f\|_{\mathrm{Lip}}^p\sum_{r=1}^{m}\sum_{s=1}^{m}\left(B^*D_\pi\mathscr{A}_\tau(A)B\right)_{rs} d_X(z_r,z_s)^p. \qquad (102)
\end{aligned}
$$

For every $i,j\in\{1,\ldots,n\}$ and $r,s\in\{1,\ldots,m\}$ we have

$$
d_X(z_r,z_s)^p \leqslant 3^{p-1}d_X(z_r,x_i)^p + 3^{p-1}d_X(x_i,x_j)^p + 3^{p-1}d_X(x_j,z_s)^p.
$$

Consequently,

$$
\begin{aligned}
&\sum_{r=1}^{m}\sum_{s=1}^{m}\left(B^*D_\pi\mathscr{A}_\tau(A)B\right)_{rs} d_X(z_r,z_s)^p\\
&=\sum_{r=1}^{m}\sum_{s=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ir}b_{js}\pi_i\mathscr{A}_\tau(A)_{ij} d_X(z_r,z_s)^p\\
&\leqslant 3^{p-1}(S_1+S_2+S_3), \qquad (103)
\end{aligned}
$$

where, using the stochasticity of A and B,

$$
\begin{aligned}
S_1 &\stackrel{\mathrm{def}}{=} \sum_{r=1}^{m}\sum_{s=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ir}b_{js}\pi_i\mathscr{A}_\tau(A)_{ij} d_X(z_r,x_i)^p\\
&= \sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_X(x_i,z_r)^p, \qquad (104)
\end{aligned}
$$

using the stochasticity of A and B and the fact that X has Markov type p with $M > M_p(X)$,

$$\begin{aligned}
S_2 &\stackrel{\mathrm{def}}{=} \sum_{r=1}^{m}\sum_{s=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ir}b_{js}\pi_i\mathscr{A}_\tau(A)_{ij}d_X(x_i,x_j)^p \\
&= \frac{1}{\tau}\sum_{k=1}^{\tau}\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i(A^k)_{ij}d_X(x_i,x_j)^p \\
&\stackrel{(1)}{\leqslant} \frac{M^p}{\tau}\sum_{k=1}^{\tau}k\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i a_{ij}d_X(x_i,x_j)^p \\
&= \frac{M^p(\tau+1)}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i a_{ij}d_X(x_i,x_j)^p \\
&\stackrel{(94)}{\leqslant} \frac{(2^{p+1}+t)M^p}{2^{p+1}}\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i a_{ij}d_X(x_i,x_j)^p, \qquad (105)
\end{aligned}$$

and, using the stochasticity of A and B, and the reversibility of $\mathscr{A}_\tau(A)$ relative to π,

$$\begin{aligned}
S_3 &\stackrel{\mathrm{def}}{=} \sum_{r=1}^{m}\sum_{s=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ir}b_{js}\pi_i\mathscr{A}_\tau(A)_{ij}d_X(x_j,z_s)^p \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{m} b_{js}\pi_j\mathscr{A}_\tau(A)_{ji}d_X(x_j,z_s)^p \\
&= \sum_{j=1}^{n}\sum_{s=1}^{m}\pi_j b_{js}d_X(x_j,z_s)^p = S_1. \qquad (106)
\end{aligned}$$

By substituting (104), (105) and (106) into (103), and combining the resulting estimate with (102) and (100) while recalling the definition of Λ in (88), we arrive at

$$\begin{aligned}
&\frac{1}{\Lambda}\left(2\sum_{i=1}^{n}\sum_{r=1}^{m}W(H)_{ir}d_Y(y_i,f(z_r))^p+\sum_{i=1}^{n}\sum_{j=1}^{n}V(H)_{ij}d_Y(y_i,y_j)^p\right) \\
&\leqslant \frac{2}{M^p}\sum_{i=1}^{n}\sum_{r=1}^{m}\theta\pi_i b_{ir}d_X(x_i,z_r)^p+\frac{2^{p+1}+t}{2^{p+1}t}\sum_{i=1}^{n}\sum_{j=1}^{n}t\theta\pi_i a_{ij}d_X(x_i,x_j)^p \\
&\leqslant 2\sum_{i=1}^{n}\sum_{r=1}^{m}\left(W(H)_{ir}+\varepsilon\right)d_X(x_i,z_r)^p+\sum_{i=1}^{n}\sum_{j=1}^{n}V(H)_{ij}d_X(x_i,x_j)^p,
\end{aligned}$$

where we used $M > M_p(X) \geqslant 1$ and $2^{p+1}t \geqslant 2^{p+1}+t$ (recall that $t \geqslant 2$), combined with the identities (98) and (99). Choosing $\varepsilon \in (0,1)$ so that $2\varepsilon \sum_{i=1}^n \sum_{r=1}^m d_X(x_i, z_r)^p \leqslant \delta$ yields the desired estimate (89). □

6. Proof of Theorem 1.14 and Theorem 1.17

Both Theorem 1.14 and Theorem 1.17 rely on a quantitative variant of a beautiful construction of Kalton [27, 29]. Before passing to the construction itself, we record some basic facts on metric Markov cotype.

Lemma 6.1. *Fix $p \in (0,\infty)$ and let (X, d_X) be a metric space with metric Markov cotype p. Suppose that $S \subseteq X$ is a Lipschitz retract of X, i.e., there exists a Lipschitz mapping $\rho : X \to S$ such that $\rho(s) = s$ for every $s \in S$. Then S also has metric Markov cotype p, and in fact*

$$N_p(S) \leqslant \|\rho\|_{\mathrm{Lip}} N_p(X).$$

Proof. Fix $N > N_p(X)$ and $n, t \in \mathbb{N}$. Suppose that $A = (a_{ij}) \in M_n(\mathbb{R})$ is a stochastic matrix that is reversible relative to $\pi \in \Delta^{n-1}$. For every $x_1, \dots, x_n \in S$ there exist $y_1, \dots, y_n \in X$ such that

$$\sum_{i=1}^n \pi_i d_X(x_i, y_i)^p + t \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_X(y_i, y_j)^p \leqslant N^p \sum_{i=1}^n \sum_{j=1}^n \pi_i \mathscr{A}_t(A)_{ij} d_X(x_i, x_j)^p.$$

Then, since $\rho(x_i) = x_i$ for every $i \in \{1, \dots, n\}$,

$$\begin{aligned}
&\sum_{i=1}^n \pi_i d_X(x_i, \rho(y_i))^p + t \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_X(\rho(y_i), \rho(y_j))^p \\
&\leqslant \|\rho\|_{\mathrm{Lip}}^p \left(\sum_{i=1}^n \pi_i d_X(x_i, y_i)^p + t \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_X(y_i, y_j)^p \right) \\
&\leqslant \|\rho\|_{\mathrm{Lip}}^p N^p \sum_{i=1}^n \sum_{j=1}^n \pi_i \mathscr{A}_t(A)_{ij} d_X(x_i, x_j)^p. \qquad \square
\end{aligned}$$

We next show that the real line $\mathbb{R}$ (equipped with the standard metric) fails to have metric Markov cotype p for any $p \in (0, 2)$. Lemma 6.2 below contains a simple explicit example that exhibits this fact, but when $p \in [1, 2)$ there is a also a roundabout way to see that $\mathbb{R}$ fails to have metric Markov cotype p via the link to Lipschitz extension. Indeed, it follows from the definition of metric Markov cotype that $N_p(\ell_p) = N_p(\mathbb{R})$. Ball proved [3] that the Markov type p constant of

ℓ_p satisfies $M_p(\ell_p) = 1$ for $p \in [1, 2)$. Corollary 1.13 therefore implies that $e(\ell_p, \ell_p) \lesssim N_p(\mathbb{R})$. But it is known that $e(\ell_p, \ell_p) = \infty$ for every $p \in [1, 2)$: for $p \in (1, 2)$ see [48], and for $p = 1$ it is observed in [39] that this follows from [8] or [19] ([39] also provides an interesting third proof of the fact that $e(\ell_1, \ell_1) = \infty$).

Lemma 6.2. *For every $p \in (0, 2)$ the real line $\mathbb{R}$ (equipped with the standard metric) fails to have metric Markov cotype p.*

Proof. Fix $p \in (0, 2)$ and suppose for the sake of obtaining a contradiction that $N_p(\mathbb{R}) < \infty$. Fixing $n \in \mathbb{N}$, define $A = (a_{ij}) \in M_n(\mathbb{R})$ by $a_{11} = a_{nn} = 1/2$ and $a_{i,i+1} = a_{i+1,i} = 1/2$ for every $i \in \{1, \ldots, n-1\}$, and the remaining entries of A vanish. Thus A is a symmetric stochastic matrix, corresponding to the standard random walk on $\{1, \ldots, n\}$ in which if the walker is at either 1 or n then with probability 1/2 it does nothing in the next step, and with probability 1/2 it moves in the next step to its unique neighbor in $\{2, n-1\}$. Let $\{W_0, W_1, W_2, \ldots\}$ denote this walk, i.e., W_0 is uniformly distributed on $\{1, \ldots, n\}$ and conditioned on $W_t = i$ the probability that $W_{t+1} = j$ equals a_{ij}. Thus, for every $t \in \mathbb{N}$ we have

$$\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(A^t)_{ij}|i-j|^p\right)^{1/p} = \left(\mathbb{E}\left[|W_t - W_0|^p\right]\right)^{1/p} \leqslant \left(\mathbb{E}\left[|W_t - W_0|^2\right]\right)^{1/2} \leqslant \sqrt{t}. \quad (107)$$

To justify the final inequality in (107), proceed by induction on t as follows. Since $|W_{t+1} - W_t| \leqslant 1$ point-wise,

$$\mathbb{E}\left[|W_{t+1} - W_0|^2\right] \leqslant \mathbb{E}\left[|W_t - W_0|^2\right] + 2\,\mathbb{E}\left[(W_t - W_0)(W_{t+1} - W_t)\right] + 1,$$

so for the induction step it suffices to show that for every $t \in \mathbb{N}$ we have $\mathbb{E}\left[(W_t - W_0)(W_{t+1} - W_t)\right] \leqslant 0$. By conditioning on W_0, W_t, it suffices to check the point-wise inequality

$$(W_t - W_0)\,\mathbb{E}\left[W_{t+1} - W_t \middle| W_0, W_t\right] \leqslant 0. \quad (108)$$

(108) is easy to verify: if $W_t \in \{2, \ldots, n-1\}$ then $W_{t+1} - W_t$ is uniformly distributed on $\{-1, 1\}$ and therefore $\mathbb{E}\left[W_{t+1} - W_t \middle| W_0, W_t\right] = 0$, if $W_t = 1$ then $W_t - W_0 \leqslant 0$ and $W_{t+1} - W_t \geqslant 0$, and if $W_t = n$ then $W_t - W_0 \geqslant 0$ and $W_{t+1} - W_t \leqslant 0$.

Due to (107), for every $t \in \mathbb{N}$ we have

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathscr{A}_t(A)_{ij}|i-j|^p \leqslant \frac{1}{t}\sum_{s=1}^{t} s^{p/2} \leqslant t^{p/2}.$$

The definition of metric Markov cotype p therefore implies that there exist $y_1,\ldots,y_n\in\mathbb{R}$ such that

$$\frac{1}{n}\sum_{i=1}^{n}|i-y_i|^p+\frac{t}{n}\sum_{i=1}^{n-1}|y_{i+1}-y_i|^p\leqslant N_p(\mathbb{R})^p t^{p/2}. \tag{109}$$

Suppose first that $p\in[1,2)$. In this case we choose

$$t=\left\lceil (4N_p(\mathbb{R}))^{\frac{2p}{2-p}}\right\rceil. \tag{110}$$

Using Hölder's inequality and (109) we deduce that

$$\sum_{i=1}^{n-1}|y_{i+1}-y_i|\leqslant (n-1)^{1-\frac{1}{p}}\left(\sum_{i=1}^{n-1}|y_{i+1}-y_i|^p\right)^{1/p}\leqslant \frac{N_p(\mathbb{R})n}{t^{\frac{1}{p}-\frac{1}{2}}}.$$

Consequently, for every $i\in\{2,\ldots,n\}$ we have

$$|y_i-y_1|\leqslant\sum_{j=1}^{i-1}|y_{j+1}-y_j|\leqslant\frac{N_p(\mathbb{R})n}{t^{\frac{1}{p}-\frac{1}{2}}}\overset{(110)}{\leqslant}\frac{n}{4}. \tag{111}$$

(111) implies that if $y_1\leqslant n/2$ then $y_i\leqslant 3n/4$ for every $i\in\{1,\ldots,n\}$ and if $y_1\geqslant n/2$ then $y_i\geqslant n/4$ for every $i\in\{1,\ldots,n\}$. Hence,

$$\frac{1}{n}\sum_{i=1}^{n}|i-y_i|^p\geqslant\frac{1}{n}\sum_{i=1}^{\lfloor n/8\rfloor}|i-y_i|^p+\frac{1}{n}\sum_{i=\lceil 7n/8\rceil}^{n}|i-y_i|^p\gtrsim n^p. \tag{112}$$

By substituting (112) into (109) and recalling (110) we conclude that

$$n\leqslant 4^{\frac{p}{2-p}}N_p(\mathbb{R})^{\frac{2}{2-p}},$$

which is a contradiction for large enough n.

It remains to deal with the case $p\in(0,1]$. Now our choice of t is

$$t=\left\lceil (4N_p(\mathbb{R}))^{\frac{2p}{2-p}}\,n^{\frac{2(1-p)}{2-p}}\right\rceil. \tag{113}$$

Observe that since $p\in(0,1]$, for every $i\in\{2,\ldots,n\}$,

$$\begin{aligned}|y_i-y_1|&\leqslant\sum_{j=1}^{i-1}|y_{j+1}-y_j|\\&\leqslant\left(\sum_{j=1}^{i-1}|y_{j+1}-y_j|^p\right)^{1/p}\overset{(109)}{\leqslant}\frac{N_p(\mathbb{R})n^{\frac{1}{p}}}{t^{\frac{1}{p}-\frac{1}{2}}}\overset{(113)}{\leqslant}\frac{n}{4}.\end{aligned}$$

We thus arrived at the same conclusion as (111), and therefore (112) holds true. In combination with (109) and our current choice of t in (113), we see that

$$n \lesssim N_p(\mathbb{R})^{\frac{2}{2-p}} n^{\frac{1-p}{2-p}},$$

which simplifies to $n \lesssim N_p(\mathbb{R})^2$, a contradiction for large enough n. $\square$

Corollary 6.3. *Every Banach space $(X, \|\cdot\|_X)$ fails to have metric Markov cotype p for all $p \in (0,2)$.*

Proof. X contains an isometric copy of $\mathbb{R}$. Since $\mathbb{R}$ is a 1-absolute Lipschitz retract (see [7, Ch. 1]), it follows from Lemma 6.1 that $N_p(X) \geqslant N_p(\mathbb{R})$. For $p \in (0,2)$ we have $N_p(\mathbb{R}) = \infty$ by Lemma 6.2. $\square$

6.1. **Kalton's construction.** For $p \in [1,\infty)$ let ℓ_p^n denote (as usual) the space $\mathbb{R}^n$ equipped with the norm $\|x\|_p = (|x_1|^p + \ldots + |x_n|^p)^{1/p}$. The unit ball of ℓ_p^n is denoted below $B_p^n = \{x \in \mathbb{R}^n : \|x\|_p \leqslant 1\}$. For every $\varepsilon \in (0,1)$ there exists a subset $\mathcal{N}$ of B_2^n such that $|\mathcal{N}| \leqslant (1+2/\varepsilon)^n$ and $\min_{y\in\mathcal{N}} \|x-y\|_2 \leqslant \varepsilon$ for every $x \in B_2^n$ (see e.g. [46]). In particular, there exists a linear operator $Q_n : \ell_1^{5^n} \to \ell_2^n$ such that

$$B_2^n \supseteq Q_n\left(B_1^{5^n}\right) \supseteq \frac{1}{2} B_2^n. \tag{114}$$

Indeed, choose $\{x_1, \ldots, x_{5^n}\} \subseteq B_2^n$ with $\min_{i\in\{1,\ldots,5^n\}} \|x - x_i\|_2 \leqslant 1/2$ for every $x \in B_2^n$. This implies that the convex hull of $\{\pm x_1, \ldots \pm x_{5^n}\}$ contains $\frac{1}{2}B_2^n$. Hence, if we set $Q_n(e_i) = x_i$, where $e_1, \ldots, e_{5^n}$ is the standard basis of $\ell_1^{5^n}$, then the linear extension of Q_n satisfies (114).

In what follows we fix a linear mapping $Q_n : \ell_1^{5^n} \to \ell_2^n$ for which (114) holds true, and we also fix a mapping $\phi_n : B_2^n \to 2B_1^{5^n}$ such that $\phi_n(-x) = -\phi_n(x)$ for every $x \in B_2^n$ and $Q_n \circ \phi_n$ is the identity mapping on B_2^n. The fact that such a ϕ_n exists is an immediate consequence of (114): simply choose a section $\varphi : B_2^n \to 2B_1^{5^n}$ of Q_n and define $\phi_n(x) = (\varphi(x) - \varphi(-x))/2$.

For every $\theta \in (0,1]$ consider the following linear subspace of $\ell_1^{5^n} \oplus \ell_2^n$.

$$Y_\theta^n \stackrel{\mathrm{def}}{=} \left\{\left(\frac{x}{n^{\theta/4}}, Q_n(x)\right) :\ x \in \ell_1^{5^n}\right\} \subseteq \ell_1^{5^n} \oplus \ell_2^n. \tag{115}$$

Below it will always be understood that Y_n is equipped with the norm inherited from $(\ell_1^{5^n} \oplus \ell_2^n)_1$, i.e., $\|(x,y)\|_{(\ell_1^{5^n}\oplus\ell_2^n)_1} = \|x\|_1 + \|y\|_2$ for every $(x,y) \in \ell_1^{5^n} \oplus \ell_2^n$.

Let $A_n \subseteq S^{n-1} = \partial B_2^n$ be a maximal (with respect to inclusion) symmetric set (i.e., $x \in A_n \iff -x \in A_n$) such that $\|a-b\|_2 > 1/\sqrt[4]{n}$ for every distinct $a, b \in A_n$. Thus $\min_{a\in A_n} \|x-a\|_2 \leqslant 1/\sqrt[4]{n}$ for every

$x \in S^{n-1}$. Define a mapping $f_\theta^n : A_n \to Y_n$ by

$$\forall\, a \in A_n, \qquad f_\theta^n(a) \stackrel{\mathrm{def}}{=} \left(\frac{\phi_n(a)}{n^{\theta/4}}, a\right). \tag{116}$$

Observe that $f_\theta^n(-a) = -f_\theta^n(a)$ for every $a \in A_n$.

Lemma 6.4. *For every $n \in \mathbb{N}$, $\theta \in (0,1]$ and $\tau \in [\theta, 1]$ we have*

$$\forall\, a, b \in A_n, \qquad \|f_\theta^n(a) - f_\theta^n(b)\|_{Y_\theta^n} \lesssim n^{(\tau-\theta)/4}\|a-b\|_2^\tau.$$

Proof. We may assume that $a \neq b$, in which case $1/\sqrt[4]{n} \leqslant \|a-b\|_2 \leqslant 2$. Consequently,

$$\begin{aligned}\|f_\theta^n(a) - f_\theta^n(b)\|_{Y_\theta^n} &= \frac{\|\phi_n(a)-\phi_n(b)\|_1}{n^{\theta/4}} + \|a-b\|_2 \\ &\leqslant \frac{\|\phi_n(a)\|_1 + \|\phi_n(b)\|_1}{n^{\theta/4}} + 2^{1-\tau}\|a-b\|_2^\tau \\ &\leqslant \left(4n^{(\tau-\theta)/4} + 2^{1-\tau}\right)\|a-b\|_2^\tau. \qquad \square\end{aligned}$$

Lemma 6.5. *Fix $n \in \mathbb{N}$, $L \in (0,\infty)$, $\theta \in (0,1]$ and $\tau \in [\theta,1]$. Suppose that $F : S^{n-1} \to Y_\theta^n$ satisfies*

$$\forall\, x, y \in S^{n-1}, \qquad \|F(x) - F(y)\|_{Y_\theta^n} \leqslant Ln^{(\tau-\theta)/4}\|x-y\|_2^\tau, \tag{117}$$

and that $F(a) = f_\theta^n(a)$ for every $a \in A_n$. Then $L \gtrsim n^{\theta/4}$.

Proof. By replacing $F(x)$ with $(F(x)-F(-x))/2$ we may assume without loss of generality that $F(-x) = -F(x)$ for every $x \in S^{n-1}$. Since F takes values in Y_θ^n, it follows from (115) that there exists a mapping $\psi : S^{n-1} \to \ell_1^{5^n}$ such that

$$\forall\, x \in S^{n-1}, \qquad F(x) = \left(\frac{\psi(x)}{n^{\theta/4}}, Q_n(\psi(x))\right). \tag{118}$$

Let σ_{n-1} denote the normalized Haar measure on S^{n-1}. We claim that for every $y \in B_\infty^{5^n} = [-1,1]^{5^n}$ we have

$$\int_{S^{n-1}} |\langle \psi(x), y\rangle| d\sigma_{n-1}(x) \lesssim \frac{L}{n^{\tau/4}}. \tag{119}$$

The proof of (119) is a standard application of the concentration of measure phenomenon on S^{n-1}. Indeed, consider the set

$$U_y \stackrel{\mathrm{def}}{=} \{x \in S^{n-1} :\ \langle \psi(x), y\rangle \leqslant 0\} \subseteq S^{n-1}.$$

Since $\psi(-x) = -\psi(x)$ for every $x \in S^{n-1}$ we have $\sigma_{n-1}(U_y) \geqslant \frac{1}{2}$. For $x \in S^{n-1}$ and $t \in (0,\infty)$ note that

$$\langle \psi(x), y\rangle \geqslant t \implies \inf_{u \in U_y} \|x-u\|_2 \geqslant \left(\frac{t}{Ln^{\tau/4}}\right)^{1/\tau}. \tag{120}$$

Indeed, if $\langle \psi(x), y\rangle \geqslant t$ then for every $u \in U_y$, since $\langle \psi(u), y\rangle \leqslant 0$,

$$\begin{aligned} t \leqslant \langle \psi(x)-\psi(u), y\rangle &\leqslant \|\psi(x)-\psi(u)\|_1 \\ &\overset{(118)}{\leqslant} n^{\theta/4}\|F(x)-F(u)\|_{Y_\theta^n} \overset{(117)}{\leqslant} Ln^{\tau/4}\cdot \|x-u\|_2^\tau, \end{aligned}$$

implying (120). By the isoperimetric inequality on S^{n-1} (see e.g. [46]) it follows from (120) that

$$\sigma_{n-1}\left(\left\{x\in S^{n-1}:\ \langle \psi(x), y\rangle \geqslant t\right\}\right) \lesssim \exp\left(-cn\left(\frac{t}{Ln^{\tau/4}}\right)^{2/\tau}\right),$$

where $c\in (0,\infty)$ is a universal constant. By symmetry, the same estimate holds true for $\sigma_{n-1}\left(\{x\in S^{n-1}:\ \langle \psi(x), y\rangle \leqslant -t\}\right)$, and therefore

$$\sigma_{n-1}\left(\left\{x\in S^{n-1}:\ |\langle \psi(x), y\rangle| \geqslant t\right\}\right) \lesssim \exp\left(-c\left(\frac{tn^{\tau/4}}{L}\right)^{2/\tau}\right).$$

Consequently,

$$\begin{aligned} \int_{S^{n-1}} |\langle \psi(x), y\rangle| d\sigma_{n-1}(x) &\lesssim \int_0^\infty \exp\left(-c\left(\frac{tn^{\tau/4}}{L}\right)^{2/\tau}\right) dt \\ &= \frac{L\tau}{2c^{\tau/2}n^{\tau/4}}\int_0^\infty s^{\frac{\tau}{2}-1}e^{-s}ds = \frac{L\Gamma(1+\tau/2)}{c^{\tau/2}n^{\tau/4}} \lesssim \frac{L}{n^{\tau/4}}, \end{aligned}$$

completing the proof of (119).

The Pietsch Domination Theorem [56] (see also [37, Prop. 3.1]) implies that there exists a Borel probability measure μ on $B_\infty^{5^n}$ such that

$$\forall\, x\in \ell_1^{5^n}, \qquad \|Q_n(x)\|_2 \leqslant \pi_1(Q_n)\int_{B_\infty^{5^n}} |\langle x, y\rangle| d\mu(y), \tag{121}$$

where $\pi_1(Q_n)$ is the 1-summing norm of Q_n, i.e.,

$$\pi_1(Q_n) = \sup_{k\in\mathbb{N}}\sup\left\{\sum_{i=1}^k \|Q_n(x_i)\|_2 :\ \sup_{y\in B_\infty^{5^n}}\sum_{i=1}^k |\langle x_i, y\rangle| \leqslant 1\right\}.$$

A theorem of Grothendieck [21, Cor. 1] (see also [37, Thm. 4.1]) implies that $\pi_1(Q_n)\leqslant K_G\|Q_n\|_{\ell_1^{5^n}\to\ell_2^n}$, where $K_G\in[1,2]$ is the Grothendieck constant. Recalling (114), we have $\|Q_n\|_{\ell_1^{5^n}\to \ell_2^n}\leqslant 1$. Hence,

$$\begin{aligned} &\int_{S^{n-1}} \|Q_n(\psi(x))\|_2 d\sigma_{n-1}(x) \\ &\qquad \overset{(121)}{\lesssim} \int_{S^{n-1}}\int_{B_\infty^{5^n}} |\langle \psi(x), y\rangle| d\mu(y) d\sigma_{n-1}(x) \overset{(119)}{\lesssim} \frac{L}{n^{\tau/4}}. \end{aligned} \tag{122}$$

For $x \in S^{n-1}$ choose $a \in A_n$ such that $\|x-a\|_2 \leqslant 1/\sqrt[4]{n}$. Recalling (116) and (118), since $F(a) = f_\theta^n(a)$ we have $Q_n(\psi(a)) = a$. Hence,

$$\begin{aligned}\|Q_n(\psi(x))\|_2 &\geqslant 1 - \|Q_n(\psi(x)) - Q_n(\psi(a))\|_2 \\ &\overset{(118)}{\geqslant} 1 - \|F(x) - F(a)\|_{Y_\theta^n} \overset{(117)}{\geqslant} 1 - \frac{Ln^{(\tau-\theta)/4}}{n^{\tau/4}} = 1 - \frac{L}{n^{\theta/4}}. \quad (123)\end{aligned}$$

By combining (122) and (123), and recalling that $\tau \geqslant \theta$, the proof of Lemma 6.5 is complete. □

Proof of Theorem 1.14. The ensuing deduction of Theorem 1.14 from Lemma 6.5 follows an idea of [48]. Fix $n \in \mathbb{N}$ and $p \in [2, n]$. Consider the metric on ℓ_2 given by

$$\forall\, x, y \in \ell_2, \qquad d_p(x,y) \stackrel{\mathrm{def}}{=} \|x-y\|_2^{2/p}.$$

Since by [3] the Markov type 2 constant of ℓ_2 satisfies $M_2(\ell_2) = 1$, the Markov type p constant of (ℓ_2, d_p) satisfies

$$M_p(\ell_2, d_p) = M_2(\ell_2) = 1. \quad (124)$$

Lemma 6.4 with $\theta = 2/n$ and $\tau = 2/p$ asserts that the function

$$f_{2/n}^{2^{n^2}} : A_{2^{n^2}} \to Y_{2/n}^{2^{n^2}}$$

is K-Lipschitz in the metric d_p, where

$$K \lesssim 2^{\frac{n^2}{4}\left(\frac{2}{p} - \frac{2}{n}\right)}.$$

In light of (124), since the $Y_{2/n}^{2^{n^2}}$ is finite dimensional it follows from Corollary 1.13 that there exists a function

$$F : \ell_2^{2^{n^2}} \to Y_{2/n}^{2^{n^2}}$$

that extends $f_{2/n}^{2^{n^2}}$ and satisfies (117) with $\theta = 2/n$, $\tau = 2/p$ and

$$L \lesssim N_p\left(Y_{2/n}^{2^{n^2}}\right).$$

We therefore deduce from Lemma 6.5 that

$$\forall\, p \in [2,n], \qquad N_p\left(Y_{2/n}^{2^{n^2}}\right) \gtrsim 2^{n/2}. \quad (125)$$

Consider the ℓ_1 direct sum

$$Y \stackrel{\mathrm{def}}{=} \left(\bigoplus_{n=1}^{\infty} Y_{2/n}^{2^{n^2}}\right)_1.$$

For every $n \in \mathbb{N}$ the restriction to the nth coordinate is a 1-Lipschitz retraction from Y onto $Y_{2/n}^{2^{n^2}}$, so by Lemma 6.1 we have

$$\forall\, p \in [2,\infty), \qquad N_p(Y) \geqslant \sup_{n\in\mathbb{N}} N_p\left(Y_{2/n}^{2^{n^2}}\right) \stackrel{(125)}{=} \infty.$$

By Corollary 6.3 we also have $N_p(Y)=\infty$ for $p\in(0,2)$, so in order to complete the proof of Theorem 1.14 it remains to show that Y is isomorphic to a subspace of ℓ_1. Since Y is the ℓ_1 direct sum of the spaces

$$Y_{2/n}^{2^{n^2}} \subseteq \ell_1^{5^{2^{n^2}}} \oplus \ell_2^{2^{n^2}},$$

it remains to recall that ℓ_2^k is $(1+\varepsilon)$-isomorphic to a subspace of ℓ_1 for every $\varepsilon\in(0,1)$ and $k\in\mathbb{N}$ (e.g. by Dvoretzky's theorem [15]). □

Proof of Theorem 1.17. By Lemma 6.4 and Lemma 6.5 with $\theta=\tau=1$,

$$e\left(S^{n-1}, A_n, Y_1^n\right) \gtrsim \sqrt[4]{n}. \tag{126}$$

Since the diameter of A_n is at most 2 and the minimal nonzero distance in A_n is at least $1/\sqrt[4]{n}$, the proof of Theorem 1.17 is complete. □

Observe in passing that since Y_1^n is 5^n-dimensional, (126) also implies (10) with $Z_{5^n}=Y_1^n$. For general $m\in\mathbb{N}$, choose $n\in\mathbb{N}$ such that $5^{n-1}\leqslant m<5^n$ and set $Z_m=\left(Y_1^{n-1}\oplus \ell_1^{m-5^{n-1}}\right)_1$.

7. Comparison with Ball's approach

Fix $n\in\mathbb{N}$, $t\in[1,\infty)$ and let $A\in M_n(\mathbb{R})$ be a stochastic matrix that is reversible with respect to $\pi\in\Delta^{n-1}$. Since A has norm 1 when viewed as an operator on $L_2(\pi)$, we can consider the following matrix.

$$\mathscr{B}_t(A) \stackrel{\mathrm{def}}{=} \frac{1}{t}\sum_{s=1}^{\infty}\left(1-\frac{1}{t}\right)^s A^s \in M_n(\mathbb{R}). \tag{127}$$

We also denote the corresponding Green's matrix by

$$\mathcal{G}_t(A) \stackrel{\mathrm{def}}{=} \frac{1}{t}I_n + \mathscr{B}_t(A) = \frac{1}{t}\left(I_n-\left(1-\frac{1}{t}\right)A\right)^{-1},$$

where $I_n\in M_n(\mathbb{R})$ denotes the identity matrix.

In [3] Ball worked with following *linear* invariant of Banach spaces. For $p\in(0,\infty)$ say that a Banach space $(X,\|\cdot\|_X)$ has Markov cotype p with constant $N\in(0,\infty)$ if for every $n\in\mathbb{N}$ and $t\in[1,\infty)$, if

$A = (a_{ij}) \in M_n(\mathbb{R})$ is a symmetric stochastic matrix that is reversible relative to $\pi \in \Delta^{n-1}$ and $x_1, \ldots, x_n \in X$ then

$$(t-1)\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i a_{ij}\left\|\sum_{k=1}^{n}\mathcal{G}_t(A)_{ik}x_k - \sum_{\ell=1}^{n}\mathcal{G}_t(A)_{j\ell}x_\ell\right\|_X^p + \sum_{i=1}^{n}\pi_i\left\|x_i - \sum_{j=1}^{n}\mathcal{G}_t(A)_{ij}x_j\right\|_X^p \leqslant N^p\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\mathscr{B}_t(A)_{ij}\|x_i - x_j\|_X^p.$$

As we shall see shortly, in Banach spaces Markov cotype p implies metric Markov cotype p as in Definition 2, but their equivalence remains open. Note that Ball proved in [3] that ℓ_1 fails to have Markov cotype 2, so this question has relevance to Question 1.15.

In the closing remarks of his paper [3], Ball proposed the following two step definition of *metric* Markov cotype for metric spaces. First, given $p \in [1, \infty)$ say that a metric space (X, d_X) is p-approximately convex if there exists $K \in (0, \infty)$ with the following property. Fix $m, n \in \mathbb{N}$ and let $B = (b_{ij}) \in M_{n\times m}(\mathbb{R})$ and $C = (c_{ij}) \in M_n(\mathbb{R})$ be stochastic matrices, such that C is reversible relative to $\pi \in \Delta^{n-1}$. Then for every $z_1, \ldots, z_m \in X$ there exist $w_1, \ldots, w_n \in X$ such that

$$\sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_X(w_i, z_r)^p + \sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i c_{ij} d_X(w_i, w_j)^p \leqslant K^p\sum_{r=1}^{m}\sum_{s=1}^{m}(B^*D_\pi CB)_{rs} d_X(z_r, z_s)^p, \quad (128)$$

where $D_\pi \in M_n(\mathbb{R})$ is given as in (75), i.e., it is the diagonal matrix whose diagonal equals π. Assuming that (X, d_X) is approximately convex, Ball defined it to have metric Markov cotype p if there exists $N \in (0, \infty)$ such that for every $n, t \in \mathbb{N}$, if $A = (a_{ij}) \in M_n(\mathbb{R})$ is stochastic and reversible relative to $\pi \in \Delta^{n-1}$ then for every $x_1, \ldots, x_n \in X$ there exist $y_1, \ldots, y_n \in X$ such that

$$\sum_{i=1}^{n}\pi_i d_X(x_i, y_i)^p + (t-1)\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\pi_i d_X(y_i, y_j)^p \leqslant N^p\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\mathscr{B}_t(A)_{ij} d_X(x_i, x_j)^p. \quad (129)$$

Denote by $N_p^B(X)$ the infimum over those $N \in (0, \infty)$ for which (129) holds true.

By Lemma 5.1 every metric space is a p-approximately convex (with K in (128) at most 6). So, the first step of Ball's definition is not needed. Observe also that for Banach spaces Markov cotype p trivially implies (129). Moreover, there is an immediate link between (129) and (2): due to (127) we have $\mathscr{A}_t(A)_{ij} \lesssim \mathscr{B}_t(A)_{ij}$ for every integer $t \geqslant 2$ and $i,j \in \{1,\ldots,n\}$. Therefore every metric space (X,d_X) satisfies $N_p^B(X) \lesssim N_p(X)$. Despite the fact that one cannot bound from above (entry-wise) the matrix $\mathscr{B}_t(A)$ by a constant multiple of the matrix $\mathscr{A}_t(A)$, the following lemma implies that $N_p(X) \lesssim N_p^B(X)$.

Lemma 7.1. *Let (X,d_X) be a metric space. Suppose that $n,t \in \mathbb{N}$ and $A \in M_n(\mathbb{R})$ is a stochastic matrix that is reversible relative to $\pi \in \Delta^{n-1}$. Then for every $p \in [1,\infty)$ and $x_1,\ldots,x_n \in X$ we have*

$$\sum_{i=1}^n\sum_{j=1}^n \pi_i \mathscr{B}_t(A)_{ij} d_X(x_i,x_j)^p \lesssim p2^p \sum_{i=1}^n\sum_{j=1}^n \pi_i \mathscr{A}_{\lceil pt\rceil}(A)_{ij} d_X(x_i,x_j)^p. \tag{130}$$

Proof. Write $u \stackrel{\mathrm{def}}{=} \lceil pt\rceil$ and note that for every $m \in \mathbb{N}$ we have

$$\begin{aligned}
&\sum_{i=1}^n\sum_{j=1}^n \pi_i (A^{mu})_{ij} d_X(x_i,x_j)^p\\
&= \sum_{i\in\{1,\ldots,n\}^{m+1}} \pi_{i_1}\left(\prod_{a=1}^m (A^u)_{i_a,i_{a+1}}\right) d_X(x_{i_1},x_{i_{m+1}})^p\\
&\leqslant \sum_{i\in\{1,\ldots,n\}^{m+1}} \pi_{i_1}\left(\prod_{a=1}^m (A^u)_{i_a,i_{a+1}}\right) m^{p-1}\sum_{a=1}^m d_X(x_{i_a},x_{i_{a+1}})^p \qquad (131)\\
&= m^p \sum_{i=1}^n\sum_{j=1}^n \pi_i (A^u)_{ij} d_X(x_i,x_j)^p\\
&\leqslant (2m)^p \sum_{i=1}^n\sum_{j=1}^n \pi_i \mathscr{A}_u(A)_{ij} d_X(x_i,x_j)^p, \qquad (132)
\end{aligned}$$

where in (131) we used the triangle inequality and Hölder's inequality, and in (132) we used Lemma 3.1.

For every $m \in \mathbb{N} \cup \{0\}$ and $r \in \{0, \ldots, u-1\}$ we have

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i (A^{mu+r})_{ij} d_X(x_i,x_j)^p \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \pi_i (A^{mu})_{ik}(A^r)_{kj} d_X(x_i,x_j)^p \\
&\leqslant 2^{p-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \pi_i (A^{mu})_{ik}(A^r)_{kj}\big(d_X(x_i,x_k)^p + d_X(x_k,x_j)^p\big) \\
&= 2^{p-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i (A^{mu})_{ij} d_X(x_i,x_j)^p + 2^{p-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i (A^{r})_{ij} d_X(x_i,x_j)^p \\
&\leqslant \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i \big((4m)^p \mathscr{A}_u(A)_{ij} + 2^p (A^r)_{ij}\big) d_X(x_i,x_j)^p, \qquad (133)
\end{aligned}$$

where in (133) we used (132).

Recalling the definition of $\mathscr{B}_t(A)$ in (127), by writing every $s \in \mathbb{N}$ as $s = mu + r$ for unique $m \in \mathbb{N} \cup \{0\}$ and $r \in \{0, \ldots, u-1\}$, it follows from (133) that

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i \mathscr{B}_t(A)_{ij} d_X(x_i,x_j)^p \\
&= \frac{1}{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=0}^{\infty}\sum_{r=0}^{u-1}\left(1-\frac{1}{t}\right)^{mu+r} \pi_i (A^{mu+r})_{ij} d_X(x_i,x_j)^p \\
&\leqslant \frac{4^p}{t}\left(\sum_{m=0}^{\infty}\sum_{r=0}^{u-1} m^p \left(1-\frac{1}{t}\right)^{mu+r}\right)\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i \mathscr{A}_u(A)_{ij} d_X(x_i,x_j)^p \\
&\quad + \frac{2^p}{t}\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i \left(\sum_{m=0}^{\infty}\sum_{r=1}^{u}\left(1-\frac{1}{t}\right)^{mu+r} (A^r)_{ij}\right) d_X(x_i,x_j)^p. \qquad (134)
\end{aligned}$$

Recalling that $u = \lceil pt \rceil$, we have

$$\begin{aligned}
\frac{1}{t}\sum_{m=0}^{\infty}\sum_{r=0}^{u-1} m^p \left(1-\frac{1}{t}\right)^{mu+r} &\lesssim p\sum_{m=1}^{\infty}\frac{m^p}{e^{pm}} \\
&\leqslant \frac{p}{e^p} + p\int_0^{\infty}\frac{x^p}{e^{px}}dx \qquad (135) \\
&= \frac{p}{e^p} + \frac{\Gamma(p+1)}{p^p} \lesssim \frac{p}{e^p}, \qquad (136)
\end{aligned}$$

where (135) uses the fact that $x \mapsto x^p e^{-px}$ achieves its global maximum on $[0,\infty)$ at $x = 1$, and (136) uses Stirling's formula.

Next, for every $i, j \in \{1, \ldots, n\}$ we have

$$\frac{1}{t}\sum_{m=0}^{\infty}\sum_{r=1}^{u}\left(1-\frac{1}{t}\right)^{mu+r}(A^r)_{ij} \leqslant \frac{u}{t}\left(\sum_{m=0}^{\infty} e^{-mu/t}\right)\mathscr{A}_u(A)_{ij}$$
$$\lesssim p\mathscr{A}_u(A)_{ij}. \tag{137}$$

(130) now follows by substituting (136) and (137) into (134). □

Acknowledgements. We are grateful to the anonymous referees for their helpful suggestions. M. M. was supported by ISF grants 221/07 and 93/11, BSF grant 2010021, and NSF grant CCF-0832797. Part of this work was completed while M. M. was a member of the Institute for Advanced Study at Princeton, NJ. A. N. was supported by NSF grant CCF-0832795, BSF grant 2010021, the Packard Foundation and the Simons Foundation. Part of this work was completed while A. N. was visiting Université Pierre et Marie Curie, Paris, France.

References

[1] A. Andoni, A. Naor, and O. Neiman. Snowflake universality of Wasserstein spaces. Preprint, 2010.

[2] A. Andoni, A. Naor, and O. Neiman. On isomorphic dimension reduction in ℓ_1. Preprint, 2011.

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

[4] K. Ball. The Ribe programme. Séminaire Bourbaki, exposé 1047, 2012.

[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] W. Ballmann. *Lectures on spaces of nonpositive curvature*, volume 25 of *DMV Seminar*. Birkhäuser Verlag, Basel, 1995. With an appendix by Misha Brin.

[7] 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.

[8] J. Bourgain. A counterexample to a complementation problem. *Compositio Math.*, 43(1):133–144, 1981.

[9] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[10] B. Brinkman, A. Karagiozova, and J. R. Lee. Vertex cuts, random walks, and dimension reduction in series-parallel graphs. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 621–630. ACM, New York, 2007.

[11] A. Brudnyi and Y. Brudnyi. *Methods of geometric analysis in extension and trace problems. Volume 2*, volume 103 of *Monographs in Mathematics*. Birkhäuser/Springer Basel AG, Basel, 2012.

[12] T. Christiansen and K. T. Sturm. Expectations and martingales in metric spaces. *Stochastics*, 80(1):1–17, 2008.
[13] J. Ding, J. R. Lee, and Y. Peres. Markov type and threshold embeddings. Preprint available at `http://arxiv.org/abs/1208.6088`, 2012.
[14] S. Doss. Moyennes conditionnelles et martingales dans un espace métrique. *C. R. Acad. Sci. Paris*, 254:3630–3632, 1962.
[15] 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, 1961.
[16] M. Émery. *Stochastic calculus in manifolds.* Universitext. Springer-Verlag, Berlin, 1989. With an appendix by P.-A. Meyer.
[17] A. Es-Sahib and H. Heinich. Barycentre canonique pour un espace métrique à courbure négative. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 355–370. Springer, Berlin, 1999.
[18] T. Figiel. On the moduli of convexity and smoothness. *Studia Math.*, 56:121–155, 1976.
[19] T. Figiel, W. B. Johnson, and G. Schechtman. Factorizations of natural embeddings of l_p^n into L_r. I. *Studia Math.*, 89(1):79–103, 1988.
[20] M. Gromov. Random walk in random groups. *Geom. Funct. Anal.*, 13(1):73–146, 2003.
[21] A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.
[22] S. Heinrich. Ultraproducts in Banach space theory. *J. Reine Angew. Math.*, 313:72–104, 1980.
[23] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.
[24] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. Extensions of Lipschitz maps into Banach spaces. *Israel J. Math.*, 54(2):129–138, 1986.
[25] W. B. Johnson, H. P. Rosenthal, and M. Zippin. On bases, finite dimensional decompositions and weaker structures in Banach spaces. *Israel J. Math.*, 9:488–506, 1971.
[26] J. Jost. *Nonpositive curvature: geometric and analytic aspects.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1997.
[27] N. J. Kalton. Spaces of Lipschitz and Hölder functions and their applications. *Collect. Math.*, 55(2):171–217, 2004.
[28] N. J. Kalton. Lipschitz and uniform embeddings into ℓ_∞. *Fund. Math.*, 212(1):53–69, 2011.
[29] N. J. Kalton. The uniform structure of Banach spaces. *Math. Ann.*, 354(4):1247–1288, 2012.
[30] M. Kapovich and B. Leeb. On asymptotic cones and quasi-isometry classes of fundamental groups of 3-manifolds. *Geom. Funct. Anal.*, 5(3):582–603, 1995.
[31] M. D. Kirszbraun. Über die zusammenziehenden und Lipschitzchen Transformationen. *Fundam. Math.*, 22:77–108, 1934.
[32] U. Lang. Extendability of large-scale Lipschitz maps. *Trans. Amer. Math. Soc.*, 351(10):3975–3988, 1999.

[33] U. Lang, B. Pavlović, and V. Schroeder. Extensions of Lipschitz maps into Hadamard spaces. *Geom. Funct. Anal.*, 10(6):1527–1553, 2000.

[34] U. Lang and T. Schlichenmaier. Nagata dimension, quasisymmetric embeddings, and Lipschitz extensions. *Int. Math. Res. Not.*, (58):3625–3655, 2005.

[35] U. Lang and V. Schroeder. Kirszbraun's theorem and metric spaces of bounded curvature. *Geom. Funct. Anal.*, 7(3):535–560, 1997.

[36] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[37] J. Lindenstrauss and A. Pełczyński. Absolutely summing operators in L_p-spaces and their applications. *Studia Math.*, 29:275–326, 1968.

[38] J. Lindenstrauss and H. P. Rosenthal. The $\mathscr{L}_p$ spaces. *Israel J. Math.*, 7:325–349, 1969.

[39] K. Makarychev and Y. Makarychev. Metric extension operators, vertex sparsifiers and Lipschitz extendability. In *51th Annual IEEE Symposium on Foundations of Computer Science*, pages 255–264, 2010.

[40] B. Maurey. *Théorèmes de factorisation pour les opérateurs linéaires à valeurs dans les espaces L^p*. Société Mathématique de France, Paris, 1974. With an English summary, Astérisque, No. 11.

[41] B. Maurey. Type, cotype and K-convexity. In *Handbook of the geometry of Banach spaces, Vol. 2*, pages 1299–1332. North-Holland, Amsterdam, 2003.

[42] M. Mendel and A. Naor. Metric cotype. *Ann. of Math. (2)*, 168(1):247–298, 2008.

[43] M. Mendel and A. Naor. Nonlinear spectral calculus and super-expanders. To appear in Inst. Hautes Études Sci. Publ. Math., available at `http://arxiv.org/abs/1207.4705`, 2012.

[44] M. Mendel and A. Naor. Expanders with respect to Hadamard spaces and random graphs. Preprint, 2013.

[45] M. Mendel and A. Naor. Markov convexity and local rigidity of distorted metrics. *J. Eur. Math. Soc. (JEMS)*, 15(1):287–337, 2013.

[46] 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.

[47] G. J. Minty. On the extension of Lipschitz, Lipschitz-Hölder continuous, and monotone functions. *Bull. Amer. Math. Soc.*, 76:334–339, 1970.

[48] A. Naor. A phase transition phenomenon between the isometric and isomorphic extension problems for Hölder functions between L_p spaces. *Mathematika*, 48(1-2):253–271 (2003), 2001.

[49] A. Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.

[50] A. Naor, Y. Peres, O. Schramm, and S. Sheffield. Markov chains in smooth Banach spaces and Gromov-hyperbolic metric spaces. *Duke Math. J.*, 134(1):165–197, 2006.

[51] A. Naor and G. Schechtman. Remarks on non linear type and Pisier's inequality. *J. Reine Angew. Math.*, 552:213–236, 2002.

[52] A. Naor and L. Silberman. Poincaré inequalities, embeddings, and wild groups. *Compos. Math.*, 147(5):1546–1572, 2011.

[53] A. Navas. An L^1 ergodic theorem with values in a non-positively curved space via a canonical barycenter map. *Ergodic Theory Dynam. Systems*, FirstView:1–15.

[54] S.-i. Ohta. Extending Lipschitz and Hölder maps between metric spaces. *Positivity*, 13(2):407–425, 2009.

[55] S.-i. Ohta. Markov type of Alexandrov spaces of non-negative curvature. *Mathematika*, 55(1-2):177–189, 2009.

[56] A. Pietsch. Absolut p-summierende Abbildungen in normierten Räumen. *Studia Math.*, 28:333–353, 1966/1967.

[57] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.

[58] G. Schechtman. More on embedding subspaces of L_p in l_r^n. *Compositio Math.*, 61(2):159–169, 1987.

[59] K.-T. Sturm. Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature. *Ann. Probab.*, 30(3):1195–1222, 2002.

[60] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002)*, volume 338 of *Contemp. Math.*, pages 357–390. Amer. Math. Soc., Providence, RI, 2003.

[61] M. Talagrand. Embedding subspaces of L_1 into l_1^N. *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.

Mathematics and Computer Science Department, The Open University of Israel, 1 University Road, P.O. Box 808, Raanana 43107,Israel
E-mail address: manorme@openu.ac.il

Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York NY 10012, USA
E-mail address: naor@cims.nyu.edu