Enseign. Math. (2) 69 (2023), 315–333
DOI 10.4171/LEM/1052

Wasserstein distance and metric trees

Maxime Mathey-Prevot and Alain Valette

Abstract. We study the Wasserstein (or earthmover) metric on the space $P(X)$ of probability measures on a metric space X. We show that, if a finite metric space X embeds stochastically with distortion D in a family of finite metric trees, then $P(X)$ embeds bi-Lipschitz into ℓ^1 with distortion D. Next, we re-visit the closed formula for the Wasserstein metric on finite metric trees due to Evans–Matsen (2012). We advocate that the right framework for this formula is real trees, and we give two proofs of extensions of this formula: one making the link with Lipschitz-free spaces from Banach space theory, the other one algorithmic (after reduction to finite metric trees).

Mathematics Subject Classification (2020). Primary 05C05; Secondary 05C12, 46B85.

Keywords. Metric trees, Wasserstein metric, bi-Lipschitz embedding, Banach spaces.

A la mémoire de Vaughan, ami et bon vivant

1. Introduction

Embeddings of metric spaces, especially discrete metric spaces like graphs, into the Banach spaces $\ell^1 = \ell^1(\mathbf{N})$ or $L^1 = L^1([0,1], dx)$, form a well-established part of metric geometry, with applications ranging from computer science to topology: we refer to [19], [7, Part I], or [22, Chapter 1]. In this paper we will be concerned with embeddings of Wasserstein spaces, that we now recall.

Let (X, d) be a metric space and let $P_1(X)$ be the space of probability measures μ on X with finite first moment, i.e.

$$\int_X d(x_0, x)\, d\mu(x) < +\infty$$

for some (hence any) base-point $x_0 \in X$. For compact X, the space $P_1(X)$ coincides with the space $P(X)$ of all probability measures on X.

The *Wasserstein metric* is a distance function on $P_1(X)$. Intuitively, given $\mu, \nu \in P_1(X)$, the distance $\mathrm{Wa}(\mu, \nu)$ represents the amount of work necessary to transform μ

into ν. More precisely, a probability measure $\pi \in P(X \times X)$ is a *coupling* between μ and ν if its marginals are μ and ν, i.e. $\mu(A) = \pi(A \times X)$ and $\nu(A) = \pi(X \times A)$ for any Borel subset $A \subset X$. And the Wasserstein distance $\mathrm{Wa}(\mu, \nu)$ is defined as

$$\mathrm{Wa}(\mu,\nu) = \inf\Big\{\int_{X\times X} d(x,y)\, d\pi(x,y) : \pi \text{ coupling between } \mu \text{ and } \nu\Big\}.$$

Note that X embeds isometrically in $P_1(X)$ by $x \mapsto \delta_x$ (the Dirac mass at x). See [23, Chapter 5] or [25, Chapter 7] for more on the Wasserstein distance, also called *Kantorovich–Rubinstein distance* or *earthmover distance* (EMD) in computer science papers. We denote by $\mathrm{Wa}(X)$ the space $P_1(X)$ endowed with the Wasserstein distance, and call it *the Wasserstein space* of X. For a coupling π, the *cost* of π is the quantity $\int_{X\times X} d(x,y)\, d\pi(x,y)$.

If (X, d) is a metric space and B is a Banach space, we say that X *embeds bi-Lipschitz in B with distortion at most $D \geq 1$* if there exists $f: X \to B$ such that, for all $x, x' \in X$:

$$(x,x') \leq \|f(x) - f(x')\|_B \leq D \cdot d(x,x').$$

Let $\mathcal{Y} = (Y_i, d_i)_{i\in I}$ be a finite family of metric spaces. We say that a metric space (X, d) *embeds stochastically in $\mathcal{Y}$ with distortion $D \geq 1$* if there exists non-negative numbers $(p_i)_{i\in I}$ summing up to 1, and maps $f_i: X \to Y_i$ (for each $i \in I$) such that:

- Each f_i is non-contracting, i.e. for every $x, y \in X$, we have
$$d_i(f_i(x), f_i(y)) \geq d(x,y).$$
- For every $x, y \in X$, we have
$$\sum_{i\in I} p_i d_i(f_i(x), f_i(y)) \leq D \cdot d(x,y).$$

The above definition has a natural probabilistic interpretation: a metric space X embeds stochastically in $\mathcal{Y}$ with distortion D, if there is a randomly chosen metric space $Y \in \mathcal{Y}$ and a randomly chosen non-contracting map $f: X \to Y$ such that for all $x, y \in X$:

$$\mathbb{E}(d_Y(f(x), f(y))) \leq D \cdot d(x,y).$$

The first aim of this paper is to prove the following result:

Theorem 1.1. *Assume that the finite metric space (X, d) embeds stochastically with distortion D into a finite family of finite metric trees. Then* $\mathrm{Wa}(X)$ *embeds bi-Lipschitz into ℓ^1 with distortion at most D.*

Here, by a metric tree, we mean a tree $T = (V, E)$ endowed with a positive weight function $w: E \to \mathbf{R}_{>0}: e \mapsto w_e$. For $x, y \in V$ we denote by $[x, y]$ the set of edges on the unique path from x to y and we endow V with the distance $d_T(x, y) = \sum_{e \in [x,y]} w_e$.

We learned Theorem 1.1 from the paper [14] by P. Indyk and N. Thaper, who get a less precise $O(D)$ for the distortion of the embedding into ℓ^1, and provide a rather frustrating comment that prompted our desire to provide a direct proof of Theorem 1.1.[1]

It was shown by J. Fakcharoenphol, S. Rao and K. Talwar (see [11, Theorem 2]) that any finite metric space on n points embeds stochastically with distortion $O(\log n)$ into a family of finite metric trees (and this bound is optimal). Using this it was shown by F. Baudier, P. Motakis, G. Schlumprecht and A. Zsák ([2, Corollary 8]) that, for X a finite metric space on n points, the lamplighter metric space $La(X)$ embeds into ℓ^1 with distortion $O(\log n) = O(\log \log |La(X)|)$. Using the same result from [11], our Theorem 1.1 immediately implies:

Corollary 1.2. *For any finite metric space X on n points, the Wasserstein space* $\mathrm{Wa}(X)$ *embeds bi-Lipschitz into ℓ^1 with distortion $O(\log n)$.*

Combining with the isometric embedding $X \to \mathrm{Wa}(X): x \mapsto \delta_x$, we get as corollary a celebrated result by J. Bourgain [4].[2]

Corollary 1.3. *Any finite metric space on n points, embeds bi-Lipschitz into ℓ^1 with distortion $O(\log n)$.*

It turns out that on finite metric trees there is a remarkable closed formula for the Wasserstein distance. It originated in papers in computer science in 2002 and probably earlier: see Charikar [5], for measures supported on the leaves of the tree.[3] For general probability measures on a finite metric tree, the formula appears in a paper in biomathematics (see S. N. Evans and F. A. Matsen [9, Section 2]). We believe it deserves to be better known in mathematical circles. To understand it, let $T = (V, E)$ be a metric tree, fix a base-vertex $x_0 \in V$ (so that T appears as a rooted tree). Any edge $e \in E$ separates T into two half-trees, and we denote by T_e the set of vertices of the half-tree NOT containing x_0: if we view the tree as hanging from the root, T_e is the subtree hanging below the edge e.

---

[1] We provide the comment for completeness: "The embedding can be seen as resulting from a combination of the following two results:

1. The result of [5], who (implicitly) showed that the techniques of [16] imply the following: if a metric M can be probabilistically embedded into trees with distortion c, then the EMD over M can be embedded into ℓ^1 with distortion $O(c)$.
2. The result of [6] who showed that the Euclidean metric over $\{1, \ldots, \Delta\}^d$ can be probabilistically embedded into trees with distortion $O(d \log \Delta)$. Again that result is implicit in that paper."

[2] Of course all the difficulty becomes hidden in [11]!

[3] A proof for this special case appears in [17, Lemma 3.1].

Theorem 1.4. *Let $T = (V, E)$ be a finite, rooted metric tree. Then for $\mu, \nu \in P(V)$:*

$$\mathrm{Wa}(\mu, \nu) = \sum_{e \in E} w_e |\mu(T_e) - \nu(T_e)|. \tag{1}$$

This formula has numerous implications, we state two of them: first, the right-hand side is independent of the choice of the root; second, it shows that the Wasserstein metric on $P_1(T)$ is a L^1-metric (see Lemma 2.4 below).

Our second aim in this paper is to give two new proofs of Theorem 1.4. The first one advocates that the right framework for Theorem 1.4 is real trees: by exploiting a connection with the theory of Lipschitz-free spaces from Banach space theory, we will extend the result to metric trees with countably many vertices. The second proof is by double inequality: the inequality

$$\mathrm{Wa}(\mu, \nu) \geq \sum_{e \in E} w_e |\mu(T_e) - \nu(T_e)|$$

follows by considering the canonical embedding of the tree into ℓ^1 and its barycentric extension to $P_1(V)$. The converse inequality is proved by first reducing to finite metric trees and, for those, given $\mu, \nu \in P(V)$, by providing an algorithmic construction of a coupling π with

$$\int_{V \times V} d(x, y)\, d\pi(x, y) = \sum_{e \in E} w_e |\mu(T_e) - \nu(T_e)|.$$

We emphasize that it is a rather unique situation to enjoy a closed formula like (1) for the Wasserstein space of a finite graph. Even for simple graphs like cycles, such a formula is not known (see, however, [18] where an approximation is obtained for a cycle by deleting one edge and reducing to a linear tree). For the n-dimensional Hamming cube H_n, it was shown by S. Khot and A. Naor (see [15, Corollary 2]) that $\mathrm{Wa}(H_n)$ embeds bi-Lipschitz into L^1 with distortion $\Omega(n)$ (it was known previously that $\mathrm{Wa}(H_n)$ embeds into L^1 with distortion $O(n)$, see [5, 14]). This illustrates the curious phenomenon of a family of spaces embedding isometrically in L^1, but with Wasserstein spaces embedding poorly.

The paper is organized as follows. In Section 2 we prove Theorem 1.1, taking Theorem 1.4 for granted. Sections 3 and 4 present our two proofs of Theorem 1.4, suitably generalized to metric trees with countably many vertices (see Theorem 3.3). Finally, the appendix provides a comparison between various σ-algebras of sets on a real tree, that appeared in the literature.

2. Stochastic embeddings

We will prove Theorem 1.1 by means of a series of lemmas. The first one is [2, Lemma 3], to which we also refer for the proof.

Lemma 2.1. *Assume that the metric space (X,d) embeds stochastically in $\mathcal{Y} = (Y_i, d_i)_{i\in I}$ with distortion D, and that for every $i \in I$ the space Y_i embeds bi-Lipschitz into ℓ^1 with distortion C_i. Then X embeds bi-Lipschitz into ℓ^1 with distortion at most CD, where $C = \max_{i\in I} C_i$.*

The second lemma was suggested to us by F. Baudier.

Lemma 2.2. *If the finite metric space (X,d) embeds stochastically into $\mathcal{Y} = (Y_i, d_i)_{i\in I}$ with distortion D, then $\mathrm{Wa}(X)$ embeds stochastically into $(\mathrm{Wa}(Y_i))_{i\in I}$ with distortion D.*

Proof. For $i \in I$, let $p_i \geq 0$ and $f_i\colon X \to Y_i$ be realizing the stochastic embedding with distortion D of X into $\mathcal{Y}$. Consider then $(f_i)_*\colon P(X) \to P(Y_i)\colon \mu \mapsto (f_i)_*(\mu)$, where $(f_i)_*(\mu)$ denotes the push-forward of the measure μ. We claim that the stochastic embedding with distortion D of $\mathrm{Wa}(X)$ into the family $(\mathrm{Wa}(Y_i))_{i\in I}$ is realized by the p_i's and the $(f_i)_*$'s; to see this, we check the two points in the definition of a stochastic embedding. Fix $\mu, \nu \in \mathrm{Wa}(X)$.

- Fix $i \in I$. Let π_i be a coupling between $(f_i)_*(\mu)$ and $(f_i)_*(\nu)$ such that

$$\mathrm{Wa}((f_i)_*(\mu), (f_i)_*(\nu)) = \sum_{y,y'\in Y_i} d_{Y_i}(y,y')\pi_i(y,y').$$

For $y \in Y_i \setminus f_i(X)$, we have

$$\sum_{y'\in Y_i} \pi_i(y,y') = (f_i)_*(\mu)(y) = \mu(f_i^{-1}(y)) = 0,$$

hence $\pi_i(y,y') = 0$ for every $y' \in Y_i$. So π_i vanishes outside of $f_i(X) \times f_i(X)$. Hence, we may define $\sigma_i \in P(X \times X)$ by $\sigma_i(x,x') = \pi_i(f_i(x), f_i(x'))$ and σ_i is a coupling between μ and ν. Then

$$\begin{aligned}\mathrm{Wa}((f_i)_*(\mu), (f_i)_*(\nu)) &= \sum_{y,y'\in Y_i} d_{Y_i}(y,y')\pi_i(y,y') \\ &= \sum_{x,x'\in X} d_{Y_i}(f_i(x), f_i(x'))\pi_i(f_i(x), f_i(x')) \\ &= \sum_{x,x'\in X} d_{Y_i}(f_i(x), f_i(x'))\sigma_i(x,x') \geq \sum_{x,x'\in X} d(x,x')\sigma_i(x,x') \geq \mathrm{Wa}(\mu,\nu),\end{aligned}$$

where the first inequality follows from the fact that f_i is non-contracting. So $(f_i)_*$ is non-contracting as well.

- Let $\pi \in P(X \times X)$ be a coupling between μ and ν such that

$$\mathrm{Wa}(\mu, \nu) = \sum_{x,x' \in X} d(x, y)\pi(x, x').$$

Set $\tau_i = (f_i \times f_i)_*(\pi) \in P(Y_i \times Y_i)$. Then τ_i is a coupling between $(f_i)_*(\mu)$ and $(f_i)_*(\nu)$ and

$$\begin{aligned}
\sum_{i \in I} p_i \,\mathrm{Wa}((f_i)_*(\mu), (f_i)_*(\nu)) &\le \sum_{i \in I} p_i \sum_{y,y' \in Y_i} d_{Y_i}(y, y')\tau_i(y, y') \\
&= \sum_{i \in I} p_i \sum_{x,x' \in X} d_{Y_i}(f_i(x), f_i(x'))\tau_i(f_i(x), f_i(x')) \\
&= \sum_{i \in I} p_i \sum_{x,x' \in X} d_{Y_i}(f_i(x), f_i(x'))\pi(x, x') \\
&= \sum_{x,x' \in X} \pi(x, x') \sum_{i \in I} p_i d_{Y_i}(f_i(x), f_i(x')) \\
&\le D \cdot \sum_{x,x' \in X} \pi(x, x')d(x, x') = D \cdot \mathrm{Wa}(\mu, \nu),
\end{aligned}$$

where the second inequality follows from the fact that the f_i's provide a stochastic embedding. This concludes the proof. ∎

Combining Lemmas 2.1 and 2.2 we immediately get:

Corollary 2.3. *If the finite metric space (X, d) embeds stochastically into $\mathcal{Y} = (Y_i, d_i)_{i \in I}$ with distortion D, and each $\mathrm{Wa}(Y_i)$ embeds bi-Lipschitz in ℓ^1 with distortion C_i, then $\mathrm{Wa}(X)$ embeds bi-Lipschitz into ℓ^1 with distortion at most CD, where $C = \max_{i \in I} C_i$.*

To prove Theorem 1.1, in view of Corollary 2.3, it is therefore enough to observe:

Lemma 2.4. *If $T = (V, E)$ is any finite metric tree, then $\mathrm{Wa}(T)$ embeds isometrically into ℓ^1.*

Proof. Fix a root $x_0 \in V$ and, for any edge $e \in E$, let $T_e \subset V$ be defined as in the Introduction. The map

$$\mathrm{Wa}(T) \to \ell^1(E) \colon \mu \mapsto (e \mapsto w_e \mu(T_e))$$

is an isometric embedding of $\mathrm{Wa}(T)$, by Theorem 1.4. ∎

This concludes the proof of Theorem 1.1 (taking Theorem 1.4 for granted).

3. First proof of Theorem 1.4

3.1. Lipschitz-free spaces. For a metric space (X, d) with a base-point $x_0 \in X$, we denote by $\mathrm{Lip}_0(X)$ the Banach space of Lipschitz functions on X vanishing at x_0, endowed with the Lipschitz norm. The space $\mathrm{Lip}_0(X)$ has a canonical pre-dual, called the *Lipschitz-free space* of X (see e.g. [26, Chapter 2], [22, Chapter 10]) and denoted by $\mathcal{F}(X)$: it is the closed linear subspace of the dual space $\mathrm{Lip}_0(X)^*$ generated by the point evaluations δ_x ($x \in X \setminus \{x_0\}$).

For $\mu \in P_1(X)$, the linear form $f \mapsto \int_X f(x)\, d\mu(x)$ defines an element of the dual $\mathrm{Lip}_0(X)^*$: this way we get an embedding of $\mathrm{Wa}(X)$ into $\mathrm{Lip}_0(X)^*$. When X is a complete separable metric space, it can be shown that this is actually an isometric embedding of $\mathrm{Wa}(X)$ into $\mathcal{F}(X)$ (see [21, Theorem 1.13] or [20, Section 2]).

3.2. Real trees. Let x, y be points in a metric space (T, d). We say that an *arc* from x to y is the image of an injective continuous map $\sigma\colon [0,1] \to T$ such that $\sigma(0) = x$ and $\sigma(1) = y$. A *geodesic segment* from x to y is the image of an isometric embedding $\sigma\colon [0, d(x,y)] \to T$ such that $\sigma(0) = x$ and $\sigma(d(x,y)) = y$. We say that (T, d) is a *real tree* if, for every $x, y \in T$, there is a unique arc from x to y, which moreover is a geodesic segment. In that case, there is a unique isometry $\sigma_{xy}\colon [0, d(x,y)]$ with $\sigma_{xy}(0) = x$ and $\sigma_{xy}(d(x,y)) = y$. Equivalently, (T, d) is a real tree if and only if T is a geodesic metric space which is 0-hyperbolic in the sense of Gromov (see [3, Lemma 2.13] for the equivalence).

For $x, y \in T$, we set $[x, y] =: \mathrm{Im}(\sigma_{xy})$ and call it the *segment* between x and y. A point $x \in T$ is a *branching point* if $T \setminus \{x\}$ has at least 3 connected components; we denote by $\mathrm{Branch}(T)$ the set of branching points of T. Fix a base-point $x_0 \in T$. For $x \in T$, we set

$$T_x = \{y \in T : x \in [x_0, y]\} = \{y \in T : d(x_0, y) = d(x_0, x) + d(x, y)\};$$

so letting T hang from the root x_0, the set T_x is the part of T lying below x.

Following A. Godard [12], we say that a subset $A \subset T$ is *measurable* if, for every $x, y \in T$, the set $\sigma_{xy}^{-1}(A)$ is Lebesgue-measurable in $[0, d(x,y)]$. On the σ-algebra $\mathcal{G}$ of measurable subsets, there is a unique measure λ such that $\lambda([x,y]) = d(x,y)$: we call λ the *length measure*[4]. It is defined as follows: for $[x, y]$ a segment in T, let λ_{xy} denote Lebesgue measure on $[0, d(x,y)]$. Then, for $A \in \mathcal{G}$, if $R \subset T$ is a finite disjoint union of segments, say $R = \bigcup_{i=1}^k [x_i, y_i]$, we set

$$\lambda_R(A) = \sum_{i=1}^k \lambda_{x_i y_i}(\sigma_{x_i y_i}^{-1}(A)).$$

[4]See the appendix below for a comparison of various σ-algebras associated with real trees.

Finally, we set

$$\lambda(A) = \sup_{R \in \mathcal{R}} \lambda_R(A), \tag{2}$$

where $\mathcal{R}$ is the set of subsets of T that can be expressed as finite disjoint unions of segments.

For A a closed subset of T containing x_0, still following [12], we define a function $L_A : A \to \mathbf{R}^+$ by

$$L_A(a) = \inf\{d(a,x) : x \in A \cap [x_0, a[\}.$$

So $L_A(a) > 0$ if and only if a is isolated in $A \cap [x_0, a]$. We then define a measure μ_A on A by

$$\mu_A = \lambda|_A + \sum_{a \in A} L_A(a)\delta_a.$$

In [12, Theorem 3.2], it is proved that, if A is a closed subset of T containing $\mathrm{Branch}(T)$, then $\mathcal{F}(A)$ is isometrically isometric to $L^1(A, \mu_A)$.

Assume from now on that the real tree T is complete and separable. Then by the previous subsection, for A a closed subset of T containing $\mathrm{Branch}(T)$, the space $\mathrm{Wa}(A)$ isometrically embeds into $L^1(A, \mu_A)$. This embedding is not written explicitly in [12]; by making it explicit we get a closed formula for the Wasserstein distance on closed subsets of real trees.

Proposition 3.1. *Let (T, d) be a complete, separable real tree and let A be a closed subset of T containing* $\mathrm{Branch}(T)$. *For $\mu, \nu \in \mathrm{Wa}(A)$, we have*

$$\mathrm{Wa}(\mu, \nu) = \int_A |\mu(T_x \cap A) - \nu(T_x \cap A)| \, d\mu_A(x). \tag{3}$$

Proof. By the proof of [12, Theorem 3.2], the map

$$\Phi : L^\infty(A, \mu_A) \to \mathrm{Lip}_0(A) : g \mapsto \Big(a \mapsto \int_{[x_0,a] \cap A} g(x) \, d\mu_A(x)\Big)$$

is an isometric isomorphism which is weak*-weak* continuous, so its transpose Φ^* realizes the desired isometric isomorphism $\mathcal{F}(A) \to L^1(A, \mu_A)$. Denoting by $\chi_{[x,y]}$ the characteristic function of the interval $[x, y]$, the previous formula may be re-written:

$$(\Phi(g))(a) = \int_A \chi_{[x_0,a]}(x) g(x) \, d\mu_A(x).$$

For $\nu \in \mathrm{Wa}(A)$, we compute $\Phi^*(\nu)$. For $g \in L^\infty(A, \mu_A)$, we have

$$\begin{aligned}(\Phi^*(\nu), g) &= (\nu, \Phi(g)) \\ &= \int_A (\Phi(g))(a) \, d\nu(a) = \int_A \Big(\int_A \chi_{[x_0,a]}(x) g(x) \, d\mu_A(x)\Big) \, d\nu(a).\end{aligned}$$

As the real tree T is separable, the measure μ_A is σ-finite, and since ν is a probability measure we may appeal to the Fubini theorem to exchange integrals[5]:

$$\begin{aligned}(\Phi^*(\nu), g) &= \int_A g(x)\Big(\int_A \chi_{[x_0,a]}(x)\, d\nu(a)\Big)\, d\mu_A(x)\\ &= \int_A g(x)\nu(T_x \cap A)\, d\mu_A(x).\end{aligned}$$

Since this holds for every $g \in L^\infty(A, \mu_A)$ we deduce that, for almost every $x \in A$:

$$(\Phi^*(\nu))(x) = \nu(T_x \cap A).$$

Equation (3) follows. ∎

Remark 3.2. When $A = T$, Proposition 3.1 becomes, for T a complete separable real tree and $\mu, \nu \in \mathrm{Wa}(T)$:

$$\mathrm{Wa}(\mu, \nu) = \int_T |\mu(T_x) - \nu(T_x)|\, d\lambda(x). \tag{4}$$

When T is the geometric realization of a finite metric tree, equation (4) appears as [9, equation (5)]; the proof is different.

Theorem 3.3. *Let $T = (V, E)$ be a rooted metric tree with countably many vertices. Then for $\mu, \nu \in P_1(V)$:*

$$\mathrm{Wa}(\mu, \nu) = \sum_{e \in E} w_e |\mu(T_e) - \nu(T_e)|.$$

Proof. Fix $\mu, \nu \in P_1(V)$. For an edge e, let e^+, e^- be the vertices of e, chosen so that $d(x_0, e^+) < d(x_0, e^-)$. Then on the arc $[e^+, e^-]$ the function $x \mapsto |\mu(T_x) - \nu(T_x)|$ is constant, equal to $|\mu(T_e) - \nu(T_e)|$. So by formula (4):

$$\begin{aligned}\mathrm{Wa}(\mu, \nu) &= \sum_{e \in E} \int_{[e^+, e^-]} |\mu(T_x) - \nu(T_x)|\, d\lambda(x)\\ &= \sum_{e \in E} w_e |\mu(T_e) - \nu(T_e)|.\end{aligned}$$ ∎

4. Second proof of Theorem 1.4

We now proceed with our second proof.

[5]There are well-known counterexamples to Fubini when one of the measures is not σ-finite, we refer e.g. to the Wikipedia entry on the Fubini theorem.

4.1. Lipschitz maps to Banach spaces.

Proposition 4.1. *Let (X, d) be a metric space, and let E be a Banach space. Any C-Lipschitz map $\beta\colon X \to E$ extends canonically to a C-Lipschitz map $\widetilde{\beta}\colon \mathrm{Wa}(X) \to E\colon \mu \mapsto \widetilde{\beta}(\mu)$ defined as the barycenter of $\beta(X)$ with respect to μ, i.e.*

$$\widetilde{\beta}(\mu) = \int_X \beta(x)\, d\mu(x).$$

Proof. Let x_0 be a base-point in X. Composing β with a translation in E, we may assume that $\beta(x_0) = 0$. Then, as $\|\beta(x) - \beta(y)\| \leq C \cdot d(x, y)$ for any $x, y \in X$, we get

$$\|\beta(x)\| \leq C \cdot d(x_0, x),$$

hence

$$\|\widetilde{\beta}(\mu)\| \leq \int_X \|\beta(x)\|\, d\mu(x) \leq C \int_X d(x_0, x)\, d\mu(x) < +\infty.$$

So $\widetilde{\beta}$ is well defined.

To check that $\widetilde{\beta}$ is C-Lipschitz, observe that for $\mu, \nu \in P_1(X)$ and π a coupling between μ and ν, we have

$$\begin{aligned}\|\widetilde{\beta}(\mu) - \widetilde{\beta}(\nu)\| &= \left\| \int_X \beta(x)\, d\mu(x) - \int_X \beta(y)\, d\nu(y) \right\| \\ &= \left\| \int_{X\times X} \beta(x)\, d\pi(x, y) - \int_{X\times X} \beta(y)\, d\pi(x, y) \right\| \\ &\leq \int_{X\times X} \|\beta(x) - \beta(y)\|\, d\pi(x, y) \leq C \int_{X\times X} d(x, y)\, d\pi(x, y).\end{aligned}$$

The result follows by taking the infimum over all couplings π. ∎

Remark 4.2. Observe that, if β in Proposition 4.1 is bi-Lipschitz, in general its extension $\widetilde{\beta}$ is not. Indeed take $E = \mathbf{R}$, and let $X \subset \mathbf{R}$ be any subset with at least 3 elements, the inclusion $\beta\colon X \to \mathbf{R}$ is isometric, but $\widetilde{\beta}$ is not even injective.

Let $T = (V, E)$ be a metric tree; we denote by $\chi_{[x,y]}$ the characteristic function of the set of edges in $[x, y]$. There is a well-known isometric embedding $\beta\colon V \to \ell^1(E, w)\colon x \mapsto \chi_{[x_0,x]}$ (it is hard to locate the first appearance of this embedding in the literature: we learned it from [13]). By Proposition 4.1, we extend it to a 1-Lipschitz map $\widetilde{\beta}\colon P_1(V) \to \ell^1(E, w)$. Ultimately we will see that $\widetilde{\beta}$ is isometric. For the moment we prove:

Proposition 4.3. *Let $T = (V, E)$ be a metric tree. For $\mu, \nu \in P_1(V)$:*

$$\|\widetilde{\beta}(\mu) - \widetilde{\beta}(\nu)\|_1 = \sum_{e\in E} w_e |\mu(T_e) - \nu(T_e)| \leq \mathrm{Wa}(\mu, \nu).$$

Proof. The inequality follows from Proposition 4.1, we focus on the equality. But,

$$\|\widetilde{\beta}(\mu) - \widetilde{\beta}(\nu)\|_1 = \sum_{e\in E} w_e |\widetilde{\beta}(\mu)(e) - \widetilde{\beta}(\nu)(e)|.$$

So it is enough to prove that $\widetilde{\beta}(\mu)(e) = \mu(T_e)$. So we compute:

$$\widetilde{\beta}(\mu)(e) = \sum_{x\in V} \beta(x)(e)\mu(x) = \sum_{x\in V} \chi_{[x_0,x]}(e)\mu(x) = \sum_{x\in T_e} \mu(x) = \mu(T_e)$$

as $\chi_{[x_0,x]}(e) = 1$ if and only if $x \in T_e$. ∎

Our aim now is to prove that the inequality in Proposition 4.3 is actually an equality, i.e. for metric trees $T = (V, E)$ with countably many vertices we wish to prove the reverse inequality

$$\sum_{e\in E} w_e |\mu(T_e) - \nu(T_e)| \geq \mathrm{Wa}(\mu, \nu). \tag{5}$$

[21, Theorem 1.13] implies that the set of finitely supported probability measures is dense in $(P_1(V), Wa)$. Of course $\mathrm{Wa}(\cdot,\cdot)\colon P_1(V) \times P_1(V) \to \mathbf{R}$ is continuous, and

$$P_1(V) \times P_1(V) \to \mathbf{R}\colon (\mu,\nu) \mapsto \sum_{e\in E} w_e |\mu(T_e) - \nu(T_e)|$$

is continuous too, as an immediate consequence of Proposition 4.3. So to show (5) we may restrict for finitely supported measures, i.e. we may restrict to finite metric trees.

4.2. An algorithm for finite metric trees. Let $T = (V, E)$ be a finite metric tree. Recall from the proof of Theorem 3.3 that if $e \in E$ is an edge, we write e^+ and e^- its two extremities chosen so $d(x_0, e^+) < d(x_0, e^-)$, moreover if $v, w \in V$, we say that w is a descendant of v if $v \in [x_0, w]$ (notice that a vertex is its own descendant) and we say that w is a child of v - and that v is the parent of w - if w is a descendant of v and $[v, w] = \{w, v\}$. If $v \in V$ we write T_v for the half tree with set of vertices the set of all descendants of v, hence $T_{x_0} = T$ and if $e \in E$, then $T_e = T_{e^-}$.

To show that

$$\mathrm{Wa}(\mu, \nu) \leq \sum_{e\in E} w_e |\mu(T_e) - \nu(T_e)|$$

we provide an algorithm which transforms a probability measure μ', initially set to μ into ν. In parallel, this algorithm keeps track of a variable (here a matrix)

$$\pi' = (\pi'(x,y))_{x,y\in V} := (\pi'_{x,y})_{x,y\in V}$$

that, all the way through the running of the algorithm, provides a coupling between μ and μ'. When the algorithm stops we will have $\mu' = \nu$ and the cost of the coupling π'

will be $\sum_{e\in E} w_e|\mu(T_e)-\nu(T_e)|$. This algorithm runs in two *phases*; intuitively speaking the first phase brings up (towards the root) the excess of mass from those subtrees T_e with $\mu(T_e) > \nu(T_e)$, and the second phase let that mass fall (towards the leaves) in the subtrees T_e with $\mu(T_e) < \nu(T_e)$. Still intuitively, for every vertex x, the quantity $\pi'_{x,y}$ is the mass attributed by μ' to x coming from y; the coupling remembers where the mass comes from. We consider that the vertices of T are numbered with $1, 2, \ldots, n := |V|$, in such a way that given two vertices that are at distinct depths in the tree, the deeper one is associated with a lower number than the other. The algorithm is such that it moves first the mass coming from vertices with a low number.

Algorithm.

% <u>*Initialization*</u>:
$\mu' \leftarrow \mu$.
for all v
  $\pi'_v \leftarrow \vec{0}$
  $\pi'_{v,v} \leftarrow \mu(v)$
end for
$M \leftarrow 0$ % *This variable is used just for the proof*
% <u>*Phase* (1)</u>:
for N **depth level, from the deeper up to** 1:
  for all T_e **subtree whose root** e^- **is at depth** N: % *Loop* $(*)$
    if $\mu'(T_e) > \nu(T_e)$ **then**
      % *"we bring* $(\mu'(T_e)-\nu(T_e))$ *up one level":*
      $x \leftarrow (\mu'(T_e)-\nu(T_e))$
      $\mu'(e^-) \leftarrow \mu'(e^-) - x$ % *and simultaneously*
      $\mu'(e^+) \leftarrow \mu'(e^+) + x$.
      $j \leftarrow \min\{k : \sum_{i=1}^{k} \pi'_{e^-,i} \geq x\}$
      for $i < j$
        $\pi'_{e^+,i} \leftarrow \pi'_{e^+,i} + \pi'_{e^-,i}$
      end for
      $\pi'_{e^+,j} \leftarrow \pi'_{e^+,j} + (x - \sum_{n=1}^{j-1} \pi_{e^-,n})$
      $\pi'_{e^-,j} \leftarrow \pi'_{e^+,j} - (x - \sum_{n=1}^{j-1} \pi_{e^-,n})$
      for all $i < j$
        $\pi_{e^-,i} \leftarrow 0$
      end for

end if

$M \leftarrow M + 1$

end for

end for

% Phase (2):

for N **depth level from** 0 **to the deepest level in the tree**-1:

for all T **subtree whose root** r **is at depth** N:

let $s_1, \dots, s_n$ **be the sons of** r

if $\mu'(r) > \nu(r)$:

for $i = 1, \dots, n$ % *Loop* $(**)$:

if $\nu(T_{s_i}) > \mu'(T_{s_i})$:

% *"We let* $(\nu(T_{s_i}) - \mu'(T_{s_i}))$ *fall one level":*

$x \leftarrow (\nu(T_{s_i}) - \mu'(T_{s_i}))$

$\mu'(s_i) \leftarrow \mu'(s_i) + x$ % *and simultaneously*

$\mu'(r) \leftarrow \mu'(r) - x.$

$j \leftarrow \min\{k : \sum_{n=1}^{k} \pi_{r,n} \geq x\}$

for $k < j$

$\pi_{s_i,k} \leftarrow \pi_{s_i,k} + \pi_{r,k}$

end for

$\pi_{s_i,j} \leftarrow \pi_{s_i,j} + (x - \sum_{n=1}^{j-1} \pi_{r,n})$

$\pi_{r,j} \leftarrow \pi_{s_i,j} - (x - \sum_{n=1}^{j-1} \pi_{r,n})$

for $k < j$

$\pi_{r,k} \leftarrow 0$

end for

end if

$M \leftarrow M + 1$

end for

end if

end for

end for

We must now prove that the algorithm works as intended, that is:

- that π' is always a coupling between μ and μ';
- second, that when the algorithm terminates we have $\mu' = \nu$;

- finally, that the cost of π' is $\sum_{e\in E} w_e \mid \mu(T_e) - \nu(T_e) \mid$.

Proof. A probability measure on a tree T is determined by the measure attributed to all subtrees T_e. To see that $\mu' = \nu$ when the algorithm terminates, we thus show that $\mu'(T_e) = \nu(T_e)$ for all subtree T_e:

- If $\mu(T_e) = \nu(T_e)$ then neither phase (1) nor (2) modifies $\mu'(T_e) = \mu(T_e) = \nu(T_e)$ (even though the distribution may vary).

- If $\mu(T_e) > \nu(T_e)$ then phase (1) removes the adequate quantity of mass from $\mu'(T_e)$ so that once phase (1) is over we have $\mu'(T_e) = \nu(T_e)$. Then phase (2) does not change the quantity $\mu'(T_e) = \nu(T_e)$ (even though it could change the distribution on that subtree).

- If $\mu(T_e) < \nu(T_e)$ phase (1) does not change the quantity $\mu'(T_e) = \mu(T_e)$ (even though it could change the distribution on that subtree). We write μ^N for the measure μ' after all subtrees whose root is at depth N have been treated by phase (2) (N going from 0 to the deepest level in the tree -1). Then we proceed by induction on N, assuming e^+ is at depth N. The initial step consists in seeing that $\mu^{N=0}$, the measure μ' just after phase (1), is a probability measure on T; ν being one too it follows $\mu^{N=0}(T) = \nu(T) = 1$. For the induction step, we write $e^- = v_1, \dots, v_m$ the children of e^+ and assume that (induction hypothesis) for all $i = 1, \dots, m$:

$$\mu^N(T_{e^+}) = \nu(T_{e^+}) = \mu^N(e^+) + \sum_{i=1}^{m} \mu^N(T_{v_i}) = \nu(e^+) + \sum_{i=1}^{m} \nu(T_{v_i}).$$

Since phase (1) is over $\mu^N(T_{v_i}) \le \nu(T_{v_i})$, hence $\mu^N(e^+) \ge \nu(e^+)$. Then, phase (2) of the algorithm modifies:

$$\mu^{N+1}(v_i) = \mu^N(v_i) + (\nu(T_{v_i}) - \mu^N(T_{v_i})).$$

And:

$$\begin{aligned}\mu^{N+1}(T_{v_i}) &= \mu^N(T_{v_i}) - \mu^N(v_i) + \mu^{N+1}(v_i) \\ &= \mu^N(T_{v_i}) - \mu^N(v_i) + \mu^N(v_i) + (\nu(T_{v_i}) - \mu^N(T_{v_i})) \\ &= \nu(T_{v_i}).\end{aligned}$$

Then we have the desired fact for $i = 1$.

Eventually when the algorithm stops $\mu' = \nu$.

The measures μ' and π' are modified only during loops $(*)$ and $(**)$, we write μ^M and π^M for the values of μ' and π' after M rounds through loops $(*)$ or $(**)$. Then $\pi^M = (\pi^M(x, y))_{x,y\in V} := (\pi^M_{x,y})_{x,y\in V}$ is a coupling between μ and μ^M: just after initialization, it is clear that $\pi' = \pi^0$ is a coupling between μ and $\mu^0 = \mu$, it follows by induction that π^M is a coupling between μ and μ^M (treating separately the case where moving from M to $M+1$ is done during phase (1) and the case where this move is done during phase (2)). About the cost of the coupling, if moving from M to $M+1$ is done during phase (1), in loop $(*)$ we have:

$$\begin{aligned}
\sum_{x,i} d(x,i)\pi_{x,i}^{M+1} &= \sum_x \sum_i d(x,i)\pi_{x,i}^{M+1} \\
&= \sum_i d(s,i)\pi_{s,i}^{M+1} + \sum_i d(r,i)\pi_{r,i}^{M+1} + \sum_{x\neq r,s} \sum_i d(x,i)\pi_{x,i}^M \\
&= \sum_{i<j} d(s,i)(\pi_{s,i}^M + \pi_{r,i}^M) + d(s,j)\left(\pi_{s,j}^M + \left(x - \sum_{n=1}^{j-1} \pi_{r,n}^M\right)\right) \\
&\quad + \sum_{i>j} d(s,i)\pi_{s,i}^M + \sum_{i<j} d(r,i)\cdot 0 + d(r,j)\left(\pi_{r,j}^M - \left(x - \sum_{n=1}^{j-1} \pi_{r,n}^M\right)\right) \\
&\quad + \sum_{i>j} d(r,i)\pi_{s,i}^M + \sum_{x\neq r,s} \sum_i d(x,i)\pi_{x,i}^M \\
&= \sum_{x,i} d(x,i)\pi_{x,i}^M + (d(s,j) - d(r,j))\left(x - \sum_{n=1}^{j-1} \pi_{r,n}^M\right) \\
&\quad + \sum_{i<j} (d(s,i) - d(r,i))\pi_{r,i}^M \\
&= \sum_{x,i} d(x,i)\pi_{x,i}^M + x\cdot(d(s,j) - d(r,j)) \\
&= \sum_{x,i} d(x,i)\pi_{x,i}^M + x\cdot d(s,r).
\end{aligned}$$

Where we have used that $d(s,j) - d(r,j) = d(s,r)$, and that every i such that $\pi_{r,i}^M$ contributes to the sum $\sum_{i<j}(d(s,j) - d(r,j))\pi_{r,i}^M$ is such that

$$d(s,i) - d(r,i) = d(s,j) - d(r,j) = d(s,r) \geq 0.$$

Each vertex i such that $\pi_{r,i}^M \neq 0$ is (non-strictly) below r in the rooted-tree (since $\pi_{r,i}^M$ is the mass μ^M in r coming from i; we let the reader check it formally). Then those vertices i such that $\pi_{r,i}^M$ contributes to the sum $\sum_{i<j}(d(s,i) - d(r,i))\pi_{r,i}^M$ are (non-strictly) below r and for those we have $d(s,i) \geq d(r,i)$, and then

$$d(s,i) - d(r,i) = d(s,r)$$

since s is the father of r. By definition of j, $\pi_{r,j}^M \neq 0$ and thus j is (non-strictly) below r, hence

$$d(s,j) - d(r,j) = d(s,r) \geq 0.$$

If moving from M to $M+1$ is done during phase (2), in loop $(**)$, we conclude similarly that the cost of the coupling is increased by $x \cdot d(s_i, r)$. During phase (1) excess measure is always brought up one level at the time, in the loop $(*)$ we thus always have

$$x = \mu'(T) - \nu(T) = \mu(T) - \nu(T).$$

And phase (1) brings up excess measure exactly from those subtrees T_e with $\mu(T_e) > \nu(T_e)$. During phase (2) measure is always brought down one level at the time, in the loop $(**)$ we thus always have

$$x = \mu'(T_i) - \nu(T_i) = \mu(T_i) - \nu(T_i).$$

And phase (2) brings down adequate quantity of measure exactly in those subtrees T_e with $\mu(T_e) < \nu(T_e)$. Since just after initialization π' has null cost, the cost of it at the end of the algorithm is thus

$$\sum_{e \in E} w_e |\mu(T_e) - \nu(T_e)|. \qquad \blacksquare$$

Remark 4.4. Let (X, d) be a Polish metric space. For $\mu, \nu \in P_1(X)$, we have from the *Kantorovich–Rubinstein duality*:

$$\mathrm{Wa}(\mu, \nu) = \sup\Big\{\int_X f(x)\, d\mu(x) - \int_X f(x)\, d\nu(x)\Big\},$$

where the supremum is taken over all 1-Lipschitz functions f (see [25, Theorem 1.3]; see also [8] for a short proof). We observe that, for a finite metric tree, our second proof of Theorem 1.4 does not appeal to Kantorovich–Rubinstein duality (in contrast e.g. with the proof in [9]).

A. σ-algebras on real trees

Let (T, d) be a real tree. Apart from Godard's construction from [12] of the σ-algebra $\mathcal{G}$ recalled above, we are aware of other constructions of σ-algebras on T and of corresponding length measures:

- The σ-algebra $\mathcal{S}$ generated by segments, see [24].
- The Borel σ-algebra $\mathcal{B}$ generated by open subsets, see [10] for compact real trees, then for locally compact real trees in [1].

All these constructions have in common that the length measure of a segment $[x, y]$ is exactly $d(x, y)$. In order to clarify the relation between $\mathcal{S}$, $\mathcal{B}$ and $\mathcal{G}$, we also introduce the σ-algebras $\mathcal{B}_0$ generated by open balls (so that $\mathcal{B}_0 \subset \mathcal{B}$) and $\overline{\mathcal{S}}$ obtained by completing $\mathcal{S}$ with respect to λ-negligible subsets.

The following proposition explains our choice to work with Godard's σ-algebra $\mathcal{G}$.

Proposition A.1. *Let T be a real tree.*

(1) *We have $\mathcal{S} \subset \mathcal{B}_0 \subset \mathcal{G}$ and $\overline{\mathcal{S}} \subset \mathcal{G}$.*

(2) *If T is separable then $\mathcal{B}_0 = \mathcal{B}$ and $\overline{\mathcal{S}} = \mathcal{G}$.*

Proof. (1) To show that $\mathcal{S} \subset \mathcal{B}_0$, fix $x, y \in T$ and let $(z_n)_{n>0}$ be a dense sequence in $[x, y]$. Then the equality

$$[x, y] = \bigcap_{m \geq 1} \Big(\bigcup_{n>0} B(z_n, 1/m) \Big)$$

shows that $[x, y] \in \mathcal{B}_0$. Now, let B be an open ball in T. For any $x, y \in T$, the intersection $B \cap [x, y]$ is convex in $[x, y]$, so it is a sub-interval in $[x, y]$. In particular, $B \cap [x, y]$ is Lebesgue-measurable in $[x, y]$, so $B \in \mathcal{G}$.

The inclusion $\overline{\mathcal{S}} \subset \mathcal{G}$ follows from the fact that $\mathcal{G}$ is complete, as can be seen from the definitions.

(2) The equality $\mathcal{B}_0 = \mathcal{B}$ holds in every separable metric space (any open set being then a countable union of open balls).

To prove the inclusion $\mathcal{G} \subset \overline{\mathcal{S}}$, we consider the subset $T^0 =: \bigcup_{x,y \in T}]x, y[$ and its complement $L = T \setminus T^0$: the latter is the set of *leaves* of T. For every segment $[x, y]$ we have $L \cap [x, y] \subset \{x, y\}$, so that $L \in \mathcal{G}$; moreover $\lambda(L) = 0$ by equation (2).

Then we take $A \in \mathcal{G}$. To show $A \in \overline{\mathcal{S}}$ we use separability: let D be a countable subset of T. It is easy to see that $T^\circ = \bigcup_{x,y \in D}]x, y[$ which implies that T° is $\overline{\mathcal{S}}$-measurable, as well as its complement L. On the one hand, $A \cap L \subset L$ and $\lambda(L) = 0$, then $A \cap L$ is λ-negligible, and thus $\overline{\mathcal{S}}$-measurable. On the other hand, $A \cap T^\circ \in \overline{\mathcal{S}}$ since $A \cap]x, y[\in \overline{\mathcal{S}}$ for all $x, y \in T$ because the σ-algebra of Lebesgue-measurable subsets is the completion of the σ-algebra generated by sub-intervals, i.e. $A \cap]x, y[\in \overline{\mathcal{S}}$. This concludes the proof. ∎

Acknowledgments. We thank Florent Baudier for many insightful discussions, for providing us with Lemma 2.2 and its proof, and for useful comments on an earlier version of the paper; and Gilles Lancien for pointing out the connection with Lipschitz-free spaces.

References

[1] S. Athreya, M. Eckhoff and A. Winter, Brownian motion on $\mathbb{R}$-trees. *Trans. Amer. Math. Soc.* **365** (2013), no. 6, 3115–3150. Zbl 1278.60009 MR 3034461

[2] F. Baudier, P. Motakis, T. Schlumprecht and A. Zsák, Stochastic approximation of lamplighter metrics. *Bull. Lond. Math. Soc.* **54** (2022), no. 5, 1804–1826. MR 4512686

[3] M. Bestvina, $\mathbb{R}$-trees in topology, geometry, and group theory. In *Handbook of geometric topology*, pp. 55–91, North-Holland, Amsterdam, 2002. Zbl 0998.57003 MR 1886668

[4] J. Bourgain, On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.* **52** (1985), no. 1-2, 46–52. Zbl 0657.46013 MR 815600

[5] M. S. Charikar, Similarity estimation techniques from rounding algorithms. In *STOC'02 – Proceedings of the 34th Annual ACM Symposium on the Theory of Computing*, pp. 380–388, ACM, New York, 2002. MR 2121163

[6] M. Charikar, C. Chekuri, A. Goel, S. Guha and S. Plotkin, Approximating a finite metric by a small number of tree metrics. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Palo Alto, CA, USA, 1998)*, pp. 379–388, IEEE, 2002, DOI 10.1109/SFCS.1998.743488

[7] M. M. Deza and M. Laurent, *Geometry of cuts and metrics*. Algorithms Combin. 15, Springer, Berlin, 1997. Zbl 0885.52001 MR 1460488

[8] D. A. Edwards, A simple proof in Monge–Kantorovich duality theory. *Studia Math.* **200** (2010), no. 1, 67–77. Zbl 1218.90213 MR 2720207

[9] S. N. Evans and F. A. Matsen, The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** (2012), no. 3, 569–592. Zbl 1411.62317 MR 2925374

[10] S. N. Evans, J. Pitman and A. Winter, Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields* **134** (2006), no. 1, 81–126. Zbl 1086.60050 MR 2221786

[11] J. Fakcharoenphol, S. Rao and K. Talwar, A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.* **69** (2004), no. 3, 485–497. Zbl 1071.68082 MR 2087946

[12] A. Godard, Tree metrics and their Lipschitz-free spaces. *Proc. Amer. Math. Soc.* **138** (2010), no. 12, 4311–4320. Zbl 1222.46010 MR 2680057

[13] U. Haagerup, An example of a nonnuclear C^*-algebra, which has the metric approximation property. *Invent. Math.* **50** (1978/79), no. 3, 279–293. Zbl 0408.46046 MR 520930

[14] P. Indyk and N. Thaper, Fast image retrieval via embeddings. In *3rd International Workshop on Statistical and Computational Theories of Vision (at ICCV)*, 2003.

[15] S. Khot and A. Naor, Nonembeddability theorems via Fourier analysis. *Math. Ann.* **334** (2006), no. 4, 821–852. Zbl 1102.46051 MR 2209259

[16] J. Kleinberg and É. Tardos, Approximation algorithms for classification problems with pairwise relationships: metric labeling and Markov random fields. *J. ACM* **49** (2002), no. 5, 616–639. Zbl 1326.68336 MR 2145145

[17] B. R. Kloeckner, A geometric study of Wasserstein spaces: ultrametrics. *Mathematika* **61** (2015), no. 1, 162–178. Zbl 1312.30069 MR 3333967

[18] A. McGregor and D. Stubbs, Sketching earth-mover distance on graph metrics. In *Approximation, randomization, and combinatorial optimization*, pp. 274–286, Lecture Notes in Comput. Sci. 8096, Springer, Heidelberg, 2013. Zbl 1405.68250 MR 3126535

[19] A. Naor, Metric dimension reduction: a snapshot of the Ribe program. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. I. Plenary lectures*, pp. 759–837, World Sci. Publ., Hackensack, NJ, 2018. Zbl 1444.46019 MR 3966745

[20] A. Naor and G. Schechtman, Planar earthmover is not in L_1. *SIAM J. Comput.* **37** (2007), no. 3, 804–826. Zbl 1155.46005 MR 2341917

[21] S. Ostrovska and M. I. Ostrovskii, Generalized transportation cost spaces. *Mediterr. J. Math.* **16** (2019), no. 6, Paper No. 157. Zbl 1454.46014 MR 4029537

[22] M. I. Ostrovskii, *Metric embeddings*. De Gruyter Studies in Mathematics 49, De Gruyter, Berlin, 2013. Zbl 1279.46001 MR 3114782

[23] F. Santambrogio, *Optimal transport for applied mathematicians*. Progress in Nonlinear Differential Equations and their Applications 87, Birkhäuser/Springer, Cham, 2015. Zbl 1401.49002 MR 3409718

[24] A. Valette, Les représentations uniformément bornées associées à un arbre réel. *Bull. Soc. Math. Belg. Sér. A* **42** (1990), 747–760. Zbl 0727.43002 MR 1316222

[25] C. Villani, *Topics in optimal transportation*. Grad. Stud. Math. 58, Amer. Math. Soc., Providence, RI, 2003. Zbl 1106.90001 MR 1964483

[26] N. Weaver, *Lipschitz algebras*. World Scientific Publishing Co., River Edge, NJ, 1999. Zbl 0936.46002 MR 1832645

(Reçu le 31 octobre 2021)

Maxime Mathey-Prevot, Institut de Mathématiques, Université de Neuchâtel, Unimail, 11 rue Emile Argand, 2000 Neuchâtel, Switzerland; *e-mail:* maxmatheyp@gmail.com

Alain Valette, Institut de Mathématiques, Université de Neuchâtel, Unimail, 11 rue Emile Argand, 2000 Neuchâtel, Switzerland; *e-mail:* alain.valette@unine.ch