]

TRANSPORTATION COST SPACES AND THEIR EMBEDDINGS INTO L_1, A SURVEY

TH. SCHLUMPRECHT

ABSTRACT. These notes present a basic survey on Transportation cost spaces (aka Lipschitzfree spaces, Wasserstein spaces) and their bi-Lipschitz and linear embeddings into L_1 spaces. To make these notes as self-contained as possible, we added the proofs of several relevant results from computational graph theory in the appendix.

CONTENTS

2010 *Mathematics Subject Classification.* 46B85, 68R12, 46B20.
Key words and phrases. Transportation Cost Spaces, Lipschitzfree Spacs, L_1-distortion.
The author was supported by the National Science Foundation under Grant Number DMS-2054443.

0. Introduction

Transportation cost spaces are of high theoretical interest, and they also are fundamental in applications in many areas of applied mathematics, engineering, physics, computer science, finance, social sciences, and more. For this reason, they appear under several different other names in the literature: *Lipschitz Free Spaces*, *Wasserstein Spaces* or more precisely *Wasserstein 1-Spaces*, *Arens Eals spaces*, *Earthmover spaces*.

Depending on the context, we will use the names *Transportation cost space* and *Lipschitz free space*. If σ and τ are two probabilities on a metric space (M,d), both having finite support, then the difference $\sigma-\tau$ can be seen as an element of the Lipschitz free space $\mathcal{F}(M)$ of (M,d). Thus $d_{\mathrm{Wa}}(\sigma,\tau)=\|\sigma-\tau\|$ is a metric on the space of probabilities on M with finite support. We will call that metric space *Wasserstein space of* M.

We investigate the relationship between the Transportation cost space $\mathcal{F}(M)$ over a metric space (M,d) and L_1 spaces; in particular, we concentrate on the question of how well $\mathcal{F}(M)$ is embeddable into an L_1. We will mostly consider finite metric spaces.

For better readability, we tried to make these notes as self-contained as possible and added the proof of some results from computational graph theory in the appendix.

1. Basic facts about Lipschitz free spaces

In this section, we will introduce the *Lipschitz free space* $\mathcal{F}(M)$ over a metric space (M,d) and only concentrate on the very basic facts, which we will need later. For a more comprehensive account of the properties of $\mathcal{F}(M)$, we refer the reader to the papers [9,13]. For two metric spaces (M,d_M) and (N,d_N), and a map ϕ we denote the Lipschitz constant by $\mathrm{Lip}(\phi)$, *i.e.,*

$$\mathrm{Lip}(\phi)=\sup_{m,m'\in M, m\neq m'}\frac{d_N(\phi(m),\phi(m'))}{d_M(m,m')}.$$

Let (M,d) be a metric space with a special point 0. Let $\mathrm{Lip}_0(M)$ be the Banach space of Lipschitz functions f on M with value in $\mathbb{R}$, for which $f(0)=0$. In that case $\mathrm{Lip}(\cdot)$ becomes a norm on $\mathrm{Lip}_0(M)$ and we write for $f\in\mathrm{Lip}_0(M)$

$$\|f\|_{\mathrm{Lip}}=\mathrm{Lip}(f)=\sup_{x\neq y, x,y\in M}\frac{|f(y)-f(x)|}{d(x,y)}.$$

Proposition 1.1. *Assume that M is a metric space and X a Banach space. Then $\|\cdot\|_L$ is a norm on $\mathrm{Lip}_0(M,X)$ which turns $\mathrm{Lip}_0(M,X)$ into a Banach space.*

Let $\mathcal{M}(M)$ be the space of linear combinations of Dirac measures μ on M. The elements of $\mathcal{M}(M)$ can be seen as elements of the dual of $\mathrm{Lip}_0(M)$ and we define $\mathcal{F}(M)$ to be closure of $\mathcal{M}(M)$ in $\mathrm{Lip}_0^*(M)$. Thus $\mathcal{F}(M)$ is the completion of $\mathcal{M}_0(M)$ with respect to the norm

$$\|\mu\|_{\mathcal{F}} = \sup_{\substack{f\in\mathrm{Lip}_0(M)\\ \|f\|_{\mathrm{Lip}}\le 1}} \int f(x)\,d\mu(x) = \sup_{\substack{f\in\mathrm{Lip}_0(M)\\ \|f\|_{\mathrm{Lip}}\le 1}} \sum_{j=1}^n a_j f(x_j) \text{ for } \mu = \sum_{j=1}^n a_j \delta_{x_j} \in \mathcal{M}(M).$$

$\mathcal{F}(M)$ can be seen as a "linearization of M" as the following observation suggests.

Proposition 1.2. *Let M be a metric space. The map $\delta_M : M \to \mathcal{F}(M)$, $m \mapsto \delta_m$, is an isometry. We will, from now on, identify elements of M with their image in $\mathcal{F}(M)$ under δ_M.*

Proof. For $m, m' \in M$

$$\|\delta_m - \delta_{m'}\|_{\mathrm{Lip}^*} = \sup_{f\in\mathrm{Lip}_0(M),\|f\|_L\le 1} \|f(m) - f(m')\| \le d(m,m').$$

On the other hand, define for $m \in M$

$$f_m(m') = d(m,m') - d(m,0), \text{ for } m' \in M.$$

then $f \in \mathrm{Lip}_0(M)$ with $\|f\|_L = 1$, and

$$\langle \delta_{m'} - \delta_m, f\rangle = f(m') - f(m) = d(m',m),$$

and thus $\|\delta_m - \delta_{m'}\|_{L^*} \ge d(m,m')$. □

Remark 1.3. The measure δ_0, as elment of $\mathrm{Lip}_0^*(M)$ is the 0-functional. Therefore the family $(\delta_m : m \in M)$ is not linear independent in $\mathrm{Lip}_0^*(M)$. But it is easy to see that the family $(\delta_m : m \in M \setminus \{0\})$ is linear independent, and the linear span of it is dense in $\mathcal{F}(M)$.

Let $\mathcal{M}_0(M)$ be the elements $\mu \in \mathcal{M}(M)$ for which $\mu(M) = 0$. After adding an appropriate multiple of δ_0 to $\mu \in \mathcal{M}$, which will not change how μ acts on elements of $\mathrm{Lip}_0(M)$, we can assume that $\mu \in \mathcal{M}_0(M)$ and thus we also can see $\mathcal{F}(M)$ as the closure of $\mathcal{M}_0(M)$ in $\mathrm{Lip}_0^*(M)$.

We will show (cf. [19, Section 2]) that $\mathrm{Lip}_0(M)$ is, in a natural way, isometrically equivalent to the dual of $\mathcal{F}(M)$. The following observation is easy to see and crucial.

Proposition 1.4. *Assume that $(f_i)_{i\in I}$ is a net in $\mathrm{Lip}_0(M)$, with $\|f\|_{\mathrm{Lip}} \le 1$, for $i \in I$ and $f(m) = \lim_{i\in I} f_i(m)$ exists for every $m \in M$, then $f \in \mathrm{Lip}_0(M)$, and $\|f\|_{\mathrm{Lip}} \le 1$.*

Theorem 1.5. *Let M be a metric space and Z a Banach space then the canonical map*

$$\mathrm{Lip}_0(M) \to \mathcal{F}^*(M), \quad f \mapsto \chi_f, \text{ with } \chi_f(\mu) = \langle \mu, f\rangle, \text{ for } \mu \in \mathcal{F}(M)$$

is an isometry from $\mathrm{Lip}_0(M)$ onto $\mathcal{F}^*(M)$.

We will need the following well known result by Dixmier:

Theorem 1.6. [5, 18] *Assume that U is a Banach space and V is a closed subspace of U^* so that B_U, the unit ball in U, is compact in the topology $\sigma(U,V)$, the topology on U generated by V.*

Then V^* is isometrically isomorphic to U and the map

$$T : U \to V^*, \quad u \mapsto F_u, \text{ with } F_u(v) = \langle u, v\rangle, \text{ for } u \in U \text{ and } v \in V,$$

is an isometrical isomorphism onto V^*.

Proof of Theorem 1.5. We will verify that the assumptions of Theorem 1.6 hold for $U = \mathrm{Lip}_0(M)$ and $V = \mathcal{F}(M) \subset \mathrm{Lip}_0^*(M)$.

Let $(f_i)_{i\in I} \subset B_{\mathrm{Lip}_0(M)}$ be a net in $\subset B_{\mathrm{Lip}_0(M)}$. We have to show that there is a subnet $(g_j)_{j\in J}$ of $(f_i)_{i\in I}$ which converges to some element $f \in B_{\mathrm{Lip}_0(M)}$ with respect to $\sigma\big(\mathrm{Lip}_0(M), \mathcal{F}(M)\big)$. Let

$$K = \prod_{m\in M} [0, d(0, m)]$$

which is compact in the product topology. It follows that some subnet $(g_j)_{j\in J}$ of $(f_i)_{i\in I}$ is pointwise converging to $g \in K$, and from Proposition 1.4 it follows that $g \in \mathrm{Lip}_0(M)$, and $\|g\|_{\mathrm{Lip}_0(M)} \le 1$. Thus $\langle \mu, g_j\rangle$ converges to $\langle \mu, g\rangle$ for each μ in the linear span of $(\delta_m : m \in M)$, and since $(f_i)_{i\in I}$ is bounded it follows that $\langle \mu, g_j\rangle$ converges to $\langle \mu, g\rangle$ for all $\mu \in \mathcal{F}(M)$. We deduce therefore our claim from Theorem 1.6. □

1.1. **Example.**

Example 1.7. Let $M = \mathbb{R}$. We want to find a concrete representation of the space $\mathcal{F}(\mathbb{R})$. Every Lipschitz function on $\mathbb{R}$ is absolutely continuous and thus, by the Fundamental Theorem of Calculus [7, Theorem 3.35] almost everywhere differentiable. Moreover the derivative

$$D : \mathrm{Lip}_0(\mathbb{R}) \to L_\infty(\mathbb{R}), \quad f \mapsto f'$$

is an isometry whose inverse is

$$I : L_\infty(\mathbb{R}) \to \mathrm{Lip}_0(\mathbb{R}), \quad f \mapsto F, \text{ with } F(x) = \int_0^x f(t)\,dt \text{ for } x \in \mathbb{R}.$$

We claim that the map $\delta_x \mapsto 1_{[0,x]}$, for $x \ge 0$ and $\delta_x \mapsto 1_{[x,0]}$, if $x < 0$, extends to an isometric isomorphism between $\mathcal{F}(\mathbb{R})$ and $L_1(\mathbb{R})$.

Indeed, we represent the span of the δ_x, $x \in \mathbb{R}$ by

$$\tilde{\mathcal{F}} = \left\{ \sum_{j=-m}^{n-1} \xi_j(\delta_{x_{j+1}} - \delta_{x_j}) : \begin{matrix} m, n \in \mathbb{N},\ x_{-m} < x_{-m+1} < \ldots < x_{-1} < x_0 = 0 < x_1 < \ldots x_n \\ \text{and } (\xi_j)_{j=-m}^n \subset \mathbb{R} \end{matrix} \right\}.$$

Then, we consider the map

$$T : \tilde{\mathcal{F}} \to L_1(\mathbb{R}), \quad \sum_{j=-m}^{n-1} \xi_j(\delta_{x_{j+1}} - \delta_{x_j}) \mapsto \sum_{j=-m}^{n-1} \xi_j 1_{[x_j, x_{j+1})}.$$

Since T has a dense image, we only need to show that T is an isometry.

For $\mu = \sum_{j=-m}^{n-1} \xi_j(\delta_{x_{j+1}} - \delta_{x_j}) \in \tilde{\mathcal{F}}$ it follows that

$$\begin{aligned}\|T(\mu)\| &= \sup_{f\in L_\infty(\mathbb{R}),\|f\|_\infty\le 1} \left| \sum_{j=-m}^{n-1} \xi_j \int_{x_j}^{x_{j+1}} f(t)\,dt \right| \\ &= \sup_{f\in L_\infty(\mathbb{R}),\|f\|_\infty\le 1} \sum_{j=-m}^{n-1} \xi_j\big(I(f)(x_{j+1}) - I(f)(x_j)\big) \\ &= \sup_{F\in \mathrm{Lip}_0(\mathbb{R}),\|f\|_\infty\le 1} \sum_{j=-m}^{n-1} \xi_j\big(F(x_{j+1}) - F(x_j)\big) = \|\mu\|_{\mathrm{Lip}^*},\end{aligned}$$

which verifies our claim.

Example 1.8. Similarly the following can be shown. If $M = \mathbb{Z}$ with its usual metric, then

$$T : \mathrm{Lip}_0(\mathbb{Z}) \to \ell_\infty(\mathbb{Z}),\ f \mapsto x_f,\ \text{with } x_f(n) = f(n) - f(n-1), \quad \text{for } n\in\mathbb{Z},$$

is an isometric isomorphism onto $\ell_\infty(\mathbb{Z})$, and it is the adjoint of the operator

$$S : \ell_1(\mathbb{Z}) \to \mathcal{F}(\mathbb{Z}), \quad e_n \mapsto \delta_n - \delta_{n-1},$$

which is also an isometric isomorphism.

1.2. **The Universality property of $\mathcal{F}(M)$.** An important property is the following extension property of $\mathcal{F}(M)$.

Proposition 1.9. *Let M and N be metric spaces, with special 0-points 0_M and 0_N. Then every Lipschitz map $\varphi : M \to N$, with $\varphi(0_M) = 0_N$, can be extended to a linear bounded map $\hat{\varphi} : \mathcal{F}(M) \to \mathcal{F}(N)$ (meaning that $\hat{\varphi}(\delta_m) = \delta_{\varphi(m)}$, for $m \in M$) so that for the operator norm of $\hat{\varphi}$ we have*

$$\|\hat{\varphi}\|_{\mathcal{F}(M)\to\mathcal{F}(N)} \le \mathrm{Lip}(\varphi).$$

Proof. Let $\lambda = \mathrm{Lip}(\varphi)$. We define

$$\varphi^\# : \mathrm{Lip}_0(N) \to \mathrm{Lip}_0(M), \qquad f \mapsto f \circ \varphi.$$

Thus, $\varphi^\#$ is a linear bounded operator, with $\|\varphi^\#\| \le \lambda$. We claim that $\varphi^\#$ is also w^* continuous, more precisely $\varphi^\#$ is $\sigma(\mathrm{Lip}_0(N), \mathcal{F}(N))$- $\sigma(\mathrm{Lip}_0(M), \mathcal{F}(M))$-continuous.

Let $(f_i)_{i\in I}$ be a net in $\mathrm{Lip}_0(N)$ which w^*-converges to 0, and let $\mu \in \mathcal{F}(M)$, say $\mu = \lim_{k\to\infty} \mu_k$ in $\mathcal{F}(M)$, and thus in $\mathrm{Lip}_0^*(M)$, where μ_k is of the form

$$\mu_k = \sum_{j=1}^{l_k} a_{(k,j)}\delta_{m(k,j)},\ \text{with } (a_{(k,j)})_{j=1}^{l_k} \subset \mathbb{R}, \text{ and } (m_{(k,j)})_{j=1}^{l_k} \subset M.$$

Then

$$\tilde{\mu}_k = \sum_{j=1}^{l_k} a_{(k,j)}\delta_{\varphi(m(k,j))}$$

converges in $\mathcal{F}(N)$ to some element $\tilde{\mu} \in \mathcal{F}(N)$ with the property that $\langle \tilde{\mu}, f \rangle = \langle \mu, f \circ \varphi \rangle$, for $f \in \mathrm{Lip}_0(N)$. Indeed, note that for $k, k' \in \mathbb{N}$

$$\begin{aligned}\|\tilde{\mu}_k - \tilde{\mu}_{k'}\|_{\mathrm{Lip}^*} &= \sup_{f \in \mathrm{Lip}_0(N), \|f\|_L \le 1} \langle f,\ \tilde{\mu}_k - \tilde{\mu}_{k'} \rangle \\ &= \sup_{f \in \mathrm{Lip}_0(N), \|f\|_L \le 1} \langle f \circ \varphi,\ \mu_k - \mu_{k'} \rangle \\ &\le \lambda \sup_{g \in \mathrm{Lip}_0(M), \|g\|_L \le 1} \langle g, \mu_k - \mu_{k'} \rangle = \|\mu_k - \mu_{k'}\|_{\mathrm{Lip}^*}.\end{aligned}$$

It follows therefore that

$$\lim_{i \in I} \langle f_i \circ \varphi, \mu \rangle = \lim_{i \in I} \langle f_i, \tilde{\mu} \rangle = 0$$

and thus we verified that $\varphi^{\#}$ is w^*-continuous.

It follows therefore that $\varphi^{\#}$ is the adjoint of an operator $\hat{\varphi} : \mathcal{F}(M) \to \mathcal{F}(N)$.

Since for $f \in \mathrm{Lip}_0(N)$ and $m \in M$

$$\langle \hat{\varphi}(\delta_m), f \rangle = \langle \delta_m, \varphi^{\#}(f) \rangle = f(\varphi(m)) = \langle \delta_{\varphi(m)}, f \rangle,$$

it follows that $\hat{\varphi}(\delta_m) = \delta_{\varphi(n)}$, which finishes our proof. □

The following extension result of $\mathcal{F}(M)$ is the reason why Godefroy and Kalton coined the name *Lipschitz free space over M* for $\mathcal{F}(M)$.

Proposition 1.10. [19, Theorem 2.2.4] *Let M be a metric space and X a Banach space, and let $L : M \to X$ be a Lipschitz function. Then a unique linear and bounded* extension $\hat{L} : \mathcal{F}(M) \to X$ *of L exists. This means that $\hat{L}(\delta_m) = L(m)$, for all $m \in \mathbb{N}$.*

Moreover we have in this case that $\|\hat{L}\|_{\mathcal{F}(M) \to X} = \|L\|_L$.

Proof. Since the δ_m, $m \in M \setminus \{0\}$ are linearly independent as elements of $\mathcal{F}(M)$ we can at least extend L linearly to $\mathrm{span}(\delta_m : m \in M \setminus \{0\})$. We denote this extension by $\tilde{L}$, and we have to show that $\tilde{L}$ extends to a bounded linear operator on $\mathcal{F}(M)$, whose operator norm coincides with $\|L\|_{\mathrm{Lip}}$. Let $\mu = \sum_{j=1}^n a_j \delta_{m_j} \in \mathrm{span}(\delta_m : m \in M)$. Then there is some $x^* \in B_{X^*}$, so that $\langle x^*, \tilde{L}(\mu) \rangle = \|\tilde{L}(\mu)\|$. It follows that $x^* \circ L$ is in $\mathrm{Lip}_0(M)$ and $\|x^* \circ L\|_{\mathrm{Lip}} \le \|L\|_{\mathrm{Lip}}$, and thus,

$$\|\tilde{L}(\mu)\|_X = \langle x^*, \tilde{L}(\mu) \rangle \le \|x^* \circ L\|_{\mathrm{Lip}} \cdot \|\mu\|_{\mathrm{Lip}^*} \le \mathrm{Lip}(L) \cdot \|\mu\|_{\mathrm{Lip}^*}. \tag{1}$$

Thus, $\tilde{L}$ can be extended to a linear bounded operator $\hat{L}$ on all of $\mathcal{F}(M)$. $\hat{L}$ is of course also Lipschitz with $\mathrm{Lip}(\hat{L}) = \|\hat{L}\|_{\mathcal{F}(M) \to X}$ and since M isometrically embeds into $\mathcal{F}$, it follows that

$$\|\hat{L}\|_{\mathcal{F}(M) \to X} = \mathrm{Lip}(\hat{L}) \ge \sup_{m, m' \in M, m \ne m'} \frac{\|L(m) - L(m')\|}{d(m, m')} = \mathrm{Lip}(L),$$

and thus, together with (1), we deduce that $\|\hat{L}\|_{\mathcal{F}(M) \to X} = \mathrm{Lip}(L)$. □

Proposition 1.11. *If M is a metric space and $N \subset M$, then $\mathcal{F}(N)$ is (in the natural way) a subspace of $\mathcal{F}(M)$.*

The proof of Proposition 1.11 follows from the following extension result for Lipschitz functions.

Lemma 1.12. *Any Lipschitz function $f : N \to \mathbb{R}$ can be extended to a Lipschitz function $F : M \to \mathbb{R}$, by defining*

$$F(m) = \inf_{n\in N}(f(n) + \|f\|_L d(n,m)), \text{ for } m \in M.$$

Moreover, it follows that $\mathrm{Lip}(F) = \mathrm{Lip}(f)$.

Proof. Assume that $m, m' \in M$, and assume without loss of generality that $F(m) \le F(m')$. Let $\varepsilon > 0$ and choose $n \in N$ so that $F(m) \ge f(n) + \mathrm{Lip}(f)d(n,m) - \varepsilon$.

Then it follows

$$\begin{aligned}0 &\le F(m') - F(m)\\ &\le f(n) + \mathrm{Lip}(f)d(n,m') - \big(f(n) + \mathrm{Lip}(f)d(n,m')\big) + \varepsilon\\ &= \mathrm{Lip}(f)(d(n,m) - d(n,m')) + \varepsilon \le \mathrm{Lip}(f)d(m',m) + \varepsilon,\end{aligned}$$

thus, we deduce the claim if we let ε tend to 0. □

2. The Transportation cost norm

2.1. **The Duality Theorem of Kantorovich.** This section presents an intrinsic definition of the space $\mathcal{F}(M)$ for a metric space (M,d), *i.e.*, a definition which only uses the metric on M. It represents $\mathcal{F}(M)$ as a *Transportation Cost Space*.

Let $\mu = \sum_{j=1}^n a_j\delta_{x_j} \in \tilde{\mathcal{F}}(M)$, $(a_j)_{j=1}^n \subset \mathbb{R}$ and $(x_j)_{j=1}^n \subset M \setminus \{0\}$. Since $\delta_0 \equiv 0$ (as a functional acting on $\mathrm{Lip}_0(M)$), we can put $a_0 = -\sum_{j=1}^n a_j$, and write μ as

$$\mu = a_0\delta_0 + \sum_{j=1}^n a_j\delta_{x_j}$$

Thus, from now on, we define

$$\tilde{\mathcal{F}}(M) = \Big\{\sum_{j=1}^n a_j\delta_{x_j} : n \in \mathbb{N}, (a_i)_{j=1}^n \subset \mathbb{R},\ (x_j)_{j=1}^n \subset M, \text{ with } \sum_{j=1}^n a_j = 0\Big\}.$$

Proposition 2.1. *Every $\mu \in \tilde{\mathcal{F}}(M)$ can be represented as*

$$\mu = \sum_{j=1}^l r_j(\delta_{x_j} - \delta_{y_j}), \text{ with } (x_j)_{j=1}^l, (y_j)_{j=1}^l \subset M, (r_j)_{j=1}^l \subset \mathbb{R}^+. \tag{2}$$

Definition 2.2. For $x, y \in M$ we call $\delta_x - \delta_y$, with $x \ne y$ a *molecule in* $\tilde{F}(M)$ and (2) a *molecular representation of* $\mu \in \tilde{\mathcal{F}}(M)$. We will always assume in that case that $x_j \ne y_j$.

Proof of Proposition 2.1. We can write $\mu \in \tilde{\mathcal{F}}(M)$ as

$$\mu = \sum_{i=1}^m a_i\delta_{x_i} - \sum_{j=1}^n b_j\delta_{y_j}$$

with $a_i, b_j > 0$, $1 \le i \le m$, $1 \le j \le n$, and $S := \sum_{i=1}^m a_i = \sum_{j=1}^n b_j$, and $(x_i)_{i=1}^m$, $(y_j)_{j=1}^n \subset M$.

$$\mu = \frac{1}{S}\Big(\sum_{j=1}^n b_j\Big)\sum_{i=1}^m a_i\delta_{x_i} - \frac{1}{S}\Big(\sum_{i=1}^m a_i\Big)\sum_{j=1}^n b_j\delta_{y_j}$$
$$= \frac{1}{S}\sum_{i=1}^m\sum_{j=1}^n a_i b_j\big(\delta_{x_i} - \delta_{y_j}\big).$$

□

Definition 2.3. Let $\mu = \mathcal{F}(M)$ have the molecular representation

$$\mu = \sum_{j=1}^n r_j(\delta_{x_j} - \delta_{y_j}),$$

$r_j > 0$, $x_j, y_j \in M$, for $j = 1, 2, \dots n$, then we define

$$t\big((r_j)_{j=1}^n, (x_j)_{j=1}^n, (y_j)_{j=1}^n\big) := \sum_{j=1}^n r_j d(x_j, y_j), \tag{3}$$

and call it the *transportation costs of that representation.*

Interpretation: Let us assume that transporting a units of a product from x to y costs $a \cdot d(x, y)$.

Let $\mu = \mu^+ - \mu^- \in \tilde{\mathcal{F}}(M)$, where μ^+ is the positive part and a negative part μ^-. We interpret μ^+ as the distribution of the surplus and μ^- as the distribution of the need of the product.

Then, a molecular representation

$$\mu = \sum_{j=1}^l r_j(\delta_{x_j} - \delta_{y_j})$$

can be seen as a *transportation plan* (and will be called as such) to move r_j units from x_j to y_j and thereby balancing the surplus with the need. For such a transportation plan, $t\big((r_j)_{j=1}^n, (x_j)_{j=1}^l(y_j)\big)$ represents the total transportation costs.

We define

$$\|\mu\|_{\mathrm{tc}} = \inf\Big\{t\big((r_j)_{j=1}^n, (x_j)_{j=1}^n(y_j)_{j=1}^n\big) : \mu = \sum_{j=1}^n r_j(\delta_{x_j} - \delta_{y_j})\Big\}$$

We will soon see that $\|\cdot\|_{\mathrm{tc}}$ is a norm. We first want to show that the inf in the definition of $\|\cdot\|_{\mathrm{tc}}$ is attained for $\mu \in \mathcal{F}(M)$. To do that we need the following proposition

Proposition 2.4. *Assume $\mu \in \tilde{\mathcal{F}}(M)$ and*

$$\mu = \sum_{j=1}^n r_i(\delta_{x_j} - \delta_{y_j})$$

is a molecular representation.

Then, there exists a molecular representation

$$\mu = \sum_{j=1}^{n'} r_i'(\delta_{x_j'} - \delta_{y_j'})$$

for which the sets $\{x_j' : j = 1, 2, \dots, n'\}$ *and* $\{y_j' : j = 1, 2, \dots n'\}$ *are disjoint and*

$$t\big((r_j')_{j=1}^n, (x_j')_{j=1}^{n'}, (y_j')_{j=1}^{n'}\big) \le t\big((r_j)_{j=1}^n, (x_j)_{j=1}^n, (y_j)_{j=1}^n\big).$$

In this case, we call

$$\mu = \sum_{j=1}^{n'} r_i'(\delta_{x_j'} - \delta_{y_j'})$$

a disjoint molecular representation.

Proof. Assume $\{x_j : j = 1, 2, \dots, n\}$ and $\{y_j : j = 1, 2, \dots n\}$ are not disjoint and without loss of generality $x_{n-1} = y_n$. We write

$$\mu = \sum_{j=1}^{n-2} r_j(\delta_{x_j} - \delta_{y_j}) + \underbrace{r_{n-1}(\delta_{x_{n-1}} - \delta_{y_{n-1}}) + r_n(\delta_{x_n} - \delta_{y_n})}_{\nu}.$$

Assume $r_{n-1} \ge r_n$ (similar argument if $r_{n-1} < r_n$) and write ν as

$$\nu = (r_{n-1} - r_n)(\delta_{x_{n-1}} - \delta_{y_{n-1}}) + r_n(\delta_{x_n} - \delta_{y_{n-1}})$$

and note that

$$\begin{aligned}(r_{n-1} - r_n)d(x_{n-1}, y_{n-1}) + r_n d(x_n, y_{n-1}) &= r_{n-1}d(x_{n-1}, y_{n-1}) + r_n(d(x_n, y_{n-1}) - d(x_{n-1}, y_{n-1}))\\ &= r_{n-1}d(x_{n-1}, y_{n-1}) + r_l(d(x_l, y_{n-1}) - d(y_n, y_{n-1}))\\ &\le r_{n-1}d(x_{n-1}, y_{n-1}) + r_l(d(x_n, y_n).\end{aligned}$$

Thus

$$\mu = \sum_{j=1}^{n-2} r_j(\delta_{x_j} - \delta_{y_j}) + (r_{n-1} - r_n)(\delta_{x_{n-1}} - \delta_{y_{n-1}}) + r_n(\delta_{x_n} - \delta_{y_{n-1}})$$

is a molecular representation eliminating $x_{n-1} = y_n$ in the intersection without increasing the transportation costs.

We can therefore iterate this procedure until we arrive at a disjoint representation of μ. □

Remark. Let

$$\mu = \sum_{j=1}^{n} r_j(\delta_{x_j} - \delta_{x_j})$$

be a disjoint molecular representation of μ.

Then it follows from the *Jordan Decomposition Theorem* (which in the simple case of finite linear combinations of Dirac measures is trivial)

$$\mu^+ = \sum_{j=1}^n r_j \delta_{x_j} \text{ is the positive part and } \mu^- = \sum_{j=1}^n r_j \delta_{y_j} \text{ is the negative part of } \mu.$$

Put

$$A^+ = \text{supp}(\mu^+) = \{x \in M : \mu^+(x) > 0\} \text{ and } A^- = \text{supp}(\mu^-) = \{x \in M : \mu^-(x) > 0\}.$$

A disjoint molecular decomposition is then always of the following form

$$\mu = \sum_{x \in A^+} \sum_{y \in A^-} \nu(x,y)(\delta_x - \delta_y).$$

For $x \in A^+$ and $y \in A^-$

$$\mu(x) = \mu^+(x) = \sum_{y \in A^-} \nu(x,y) \text{ and } \mu^-(y) = \sum_{x \in A^+} \nu(x,y).$$

Thus ν (put $\nu(x,y) = 0$ if $(x,y) \neq A^+ \times A^-$) can be seen as a (positive) measure on M^2 whose *marginals* are μ^+ and μ^-. Also, note that.

$$\nu(M^2) = \sum_{x,y \in M^2} \nu(x,y) = \sum_{x \in M} \sum_{y \in M} \nu(x,y) = \sum_{x \in A^+} \sum_{y \in A^-} \nu(x,y) = \mu^+(M) = \mu^-(M).$$

Let $t(\nu)$ be the transportation costs of the representation

$$\mu = \sum_{x \in A^+} \sum_{y \in A^-} \nu(x,y)(\delta_x - \delta_y).$$

Then

$$t(\nu) = \sum_{x,y \in M} \nu(x,y) d(x,y) = \int_M \int_M d(x,y)\, d\nu(x,y).$$

and thus

$$\|\mu\|_{\text{tc}} = \inf \left\{ \sum_{(x,y) \in M^2} \nu(x,y) d(x,y) : \begin{array}{c} \nu \text{ measure on } M^2, \\ \mu^+(x) = \sum_{y' \in M} \nu(x,y') \text{ for } x \in M \text{ and} \\ \mu^-(y) = \sum_{x' \in M} \nu(x',y) \text{ for } y \in M \end{array} \right\} \tag{4}$$

$$= \inf \left\{ \sum_{(x,y) \in M^2} \nu(x,y) d(x,y) : \begin{array}{c} \nu \text{ measure on } M^2, \nu(M^2) = \mu^+(M) \\ \text{supp}(\nu) \subset \text{supp}(\mu^+) \times \text{supp}(\mu^-) \\ \mu^+(x) = \sum_{y' \in M} \nu(x,y') \text{ for } x \in M \text{ and} \\ \mu^-(y) = \sum_{x' \in M} \nu(x',y) \text{ for } y \in M \end{array} \right\}.$$

From compactness, we therefore deduce that

Corollary 2.5. *For $\mu \in \tilde{\mathcal{F}}$*

$$\|\mu\|_{\mathrm{tc}} = \inf\Big\{ t\big((r_j)_{j=1}^n, (x_j)_{j=1}^l (y_j)\big) : \mu = \sum_{j=1}^n r_j(\delta_{x_j} - \delta_{y_j}) \Big\}$$

is attained.

We call a representation of μ optimal *if its transportation cost equals to $\|\mu\|_{\mathrm{tc}}$.*

Corollary 2.6. *For $x, y \in M$*

$$\|\delta_x - \delta_y\|_{\mathrm{tc}} = d(x,y).$$

Proof. By above remark $\|\mu\|_{\mathrm{tc}} = t(\nu)$ where ν is a measure on $M \times M$ with $\mathrm{supp}(\nu) = \{(x,y)\}$, and thus $\nu = \delta_x - \delta_y$ is an optimal representation □

Theorem 2.7 (Duality Theorem of Kantorovich [14])**.** *For $\mu \in \tilde{\mathcal{F}}(M)$ it follows that $\|\mu\|_{\mathcal{F}} = \|\mu\|_{\mathrm{tc}}$.*

Proof. It is easy to see that $\|\cdot\|_{\mathrm{tc}}$ is a semi norm on $\tilde{\mathcal{F}}(M)$. By Corollary 2.6 $\|\delta_x - \delta_y\|_{\mathrm{tc}} = d(x,y)$, for $x, y \in M$.

Claim. For every norm $\|\cdot\|$ on $\tilde{\mathcal{F}}(M)$ with $\|\delta_x - \delta_y\| = d(x,y)$, for all $x, y \in M$, it follows that $\|\mu\|_{\mathrm{tc}} \geq \|\mu\|$ for all $\mu \in \tilde{\mathcal{F}}(M)$.

Indeed, let $\mu = \sum_{x,y} \nu(x,y)(\delta_x - \delta_y) \in \tilde{\mathcal{F}}(M)$ be a molecular representation. Then

$$t(\nu) = \sum_{x,y \in M} \nu(x,y) d(x,y) \geq \Big\| \sum_{x,y} \nu(x,y)(\delta_x - \delta_y) \Big\|.$$

So the claim follows by taking the infimum over all representations.

Since $\|\cdot\|_{\mathcal{F}}$ is such a norm, it follows that $\|\cdot\|_{\mathcal{F}} \leq \|\cdot\|_{\mathrm{tc}}$, and in particular that $\|\cdot\|_{\mathrm{tc}}$ is also a norm.

Let X be the completion of $\tilde{\mathcal{F}}(M)$ with respect to $\|\cdot\|_{\mathrm{tc}}$. The map $L : M \to X$, $x \to \delta_x$ is an isometric embedding, which by the extension property of $\mathcal{F}(M)$ can be extended to a linear operator $\bar{L} : \mathcal{F}(M) \to X$ with $\|\bar{L}\|_{\mathcal{F}(M) \to X} = 1$.

This means that $\|\mu\|_{\mathcal{F}} \geq \|\mu\|_{\mathrm{tc}}$, thus $\|\mu\|_{\mathcal{F}} = \|\mu\|_{\mathrm{tc}}$, for $\mu \in \tilde{\mathcal{F}}(M)$. Therefore $X = \mathcal{F}(M)$ and $\|\cdot\|_{\mathrm{tc}} = \|\cdot\|_{\mathcal{F}}$ the norms are the same. □

Corollary 2.8. *Let $\mu \in \tilde{\mathcal{F}}(M)$. Then a representation $\mu = \sum_{j=1}^n r_j(\delta_{x_j} - \delta_{y_j})$ is optimal if and only if there is an $f \in \mathrm{Lip}_0(M)$, with $\|f\|_{\mathrm{Lip}} = 1$, for which*

$$f(x_j) - f(y_j) = d(x_j, y_j), \text{ for } j = 1, 2, \ldots, n. \tag{5}$$

Proof. If the representation $\mu = \sum_{j=1}^n r_j(\delta_{x_j} - \delta_{y_j})$,with $r_j > 0$ and $x_j, y_j \in M$, for $j = 1, 2, \ldots, n$, is optimal, then it follows from the Hahn-Banach theorem and Theorem 2.7, that there is an $f \in \mathrm{Lip}_0(M)$, $\|f\|_{\mathrm{Lip}} = 1$ for which

$$\int f \, d\mu = \sum_{j=1}^n r_j\big(f(x_j) - f(y_j)\big) = \|\mu\|_{tc} = \sum_{j=1}^n r_i d(x_j, (y_j)$$

since $\|f\|_{\rm Lip} = 1$ it follows that $f(x_i) - f(y_i) \le d(x_j, (y_j)$, and, thus, $f(x_i) - f(y_i) = d(x_j, y_j)$, for all $j = 1, 2, \ldots n$.

Conversely, if (5) holds for some $f \in {\rm Lip}_0(M)$, with $\|f\|_{\rm Lip} = 1$, then by Theorem 2.7

$$\int f\,d\mu = \sum_{j=1}^n r_j\big(f(x_j) - f(y_j)\big) \le \|\mu\|_{\mathcal F} = \|\mu\|_{tc}. \tag{6}$$

On the other hand, by definition of $\|\mu\|_{tc}$, it follows that

$$\|\mu\|_{tc} \le \sum_{j=1}^n r_i d(x_j, (y_j) = \sum_{j=1}^n r_i\big(f(x_j) - (y_j)\big),$$

and thus, the inequality in (6) is an equality, and the representation is optimal. □

2.2. **The Extreme Points of $B_{\mathcal F(M)}$.**

Definition 2.9. Let $\mu \in \mathcal F(M)$ and let

$$\mu = \sum_{j=1}^n r_j(\delta_{x_j} - \delta_{y_j})$$

be a molecular representation with $r_j > 0$, $j = 1, 2, \ldots n$. We call a sequence $(j_i)_{i=1}^l$ of pairwise distinct elements in $\{1, 2, \ldots n\}$ *path* for the above representation if $x_{j_{i+1}} = y_{j_i}$, for $i = 1, 2, \ldots, l-1$, and we call it a *circle*, if moreover $y_{j_l} = x_{j_1}$

Not every optimal representation needs to be disjoint. Nevertheless, if it is not disjoint, something special has to happen.

Proposition 2.10. *If for $\mu \in \mathcal F(M)$ the representation*

$$\mu = \sum_{j=1}^n r_j(\delta_{x_j} - \delta_{y_j}). \tag{7}$$

is optimal, then for any path $(j_i)_{i=1}^l$ it follows that

$$d(x_{j_1}, y_{j_l}) = \sum_{i=1}^l d(x_{j_i}, y_{j_i}) = d(x_{i_1}, y_{i_1}) + \sum_{i=2}^l d(y_{j_{i-1}}, y_{j_i}).$$

In particular, it does not contain a circle.

Intuitively the claim is clear.

Proof. Without loss of generality (after reordering), we can assume that $j_i = n - l + i$, for $i = 1, 2, \ldots l$. We put $\varepsilon = \min\{r_j : n - l < j \le n\}$. Then we write

$$\mu = \sum_{j=1}^{n-l} r_j(\delta_{x_j} - \delta_{y_j}) + \sum_{j=n-l+1}^{n} (r_j - \varepsilon)(\delta_{x_j} - \delta_{y_j}) + \varepsilon \underbrace{\sum_{j=n-l+1}^{n} (\delta_{x_j} - \delta_{y_j})}_{=:\nu}$$

$$= \sum_{j=1}^{n-l} r_j(\delta_{x_j} - \delta_{y_j}) + \sum_{j=n-l+1}^{n} (r_j - \varepsilon)(\delta_{x_j} - \delta_{y_j}) + \varepsilon(\delta_{x_{j_1}} - \delta_{y_{j_l}}).$$

From the optimality of the representation (7), it follows that

$$d(x_{j_1}, y_{j_l}) = \sum_{i=1}^{l} d(x_{j_i}, y_{j_i}).$$

□

We deduce

Corollary 2.11. *Assume $x, y \in M$, $x \neq y$, and there is no $z \in M \setminus \{x, y\}$ for which $d(x,y) = d(x,z)+d(z,y)$. Then every optimal representation of $\delta_x - \delta_y$ is disjoint, i.e., by the remark after Proposition* 2.4 *it is of the form*

$$\delta_x - \delta_y = \sum_{j=1}^{n} r_j(\delta_x - \delta_y), \textit{ with } \sum_{j=1}^{n} r_j = d(x,y).$$

Proof. Note that a representation

$$\delta_x - \delta_y = \sum_{j=1}^{n} r_j(\delta_{x_j} - \delta_{y_j}),$$

which is not disjoint, would have a path $(j_i)_{i=1}^{l}$, $l \geq 2$, with $x = x_{j_1}$ and $y = y_{j_l}$, but this would by the assumption on x and y and by Proposition 2.10 mean that this representation is not optimal. □

The following result was shown in the general case (*i.e.,* in the case that M is not finite) by Aliega and Prenecká in [1]. Recall that for a Banach space E, and a subset $C \subset E$ an element $x \in C$ is called *extreme point of* C if x cannot be written as $x = \alpha y + (1-\alpha)z$,with $0 < \alpha < 1$ and $y \neq z$, $z, y \in C \setminus \{x\}$. In other words, if

$$x = \sum_{j=1}^{n} x_j \alpha_j,\ 0 < \alpha_j < 1, \text{ and } x_j \in C,\ j = 1, 2 \ldots n, \quad \text{with } \sum_{j=1}^{n} \alpha_j = 1,$$

then $x_1 = x_2 = \ldots = x_n = x$.

Recall (Krein-Milman Theorem): Every convex and compact subset C of E is the closed convex hull of its extreme points,

Theorem 2.12. *Let (M, d) be a finite metric space, then $\mu \in B_{\mathcal{F}(M)}$ is an extreme point if and only if μ is of the form*

$$\mu = \frac{\delta_x - \delta_y}{d(x,y)}, \textit{ with } x \neq y, \textit{ and there is no } z \in M \setminus \{x,y\} \textit{ for which } d(x,y) = d(x,z) + d(z,y).$$

Proof. (We are using that $\mathcal{F}(M) = \tilde{\mathcal{F}}(M)$) Assume that $\mu \in B_{\mathcal{F}(M)}$ is an extreme point, and let

$$\mu = \sum_{j=1}^{l} a_j \frac{\delta_{x_j} - \delta_{y_j}}{d(x_j, y_j)},$$

be its optimal representation. Thus it follows that

$$1 = \|\mu\|_{\mathcal{F}} = \sum_{j=1}^{l} a_j.$$

Since μ is an extreme point it follows that $x_i = x_j = x$ and all $y_i = y_j = y$ for all $i, j \in \{1, 2, \ldots, l\}$, and thus $\mu = \frac{\delta_x - \delta_y}{d(x,y)}$. There cannot be a $z \in M$ so that $d(x, y) = d(x, z) + d(z, y)$, because otherwise we could write

$$\mu = \frac{d(x,z)}{d(x,y)} \frac{\delta_x - \delta_z}{d(x,z)} + \frac{d(y,z)}{d(x,y)} \frac{\delta_z - \delta_y}{d(y,z)}.$$

and note that this is also an optimal representation.

Conversely, assume that $x \neq y$ are in M, and that there is no $z \in M$ for which $d(x, y) = d(x, z) + d(z, y)$, and assume we write

$$\frac{\delta_x - \delta_y}{d(x,y)} = \alpha\mu + (1 - \alpha)\nu, \text{ with } \mu, \nu \in B_{\mathcal{F}}, \text{ and } 0 < \alpha < 1.$$

Write the optimal decompositions of μ and ν as

$$\mu = \sum_{j=1}^{l} a_j \frac{\delta_{x_j} - \delta_{y_j}}{d(x_j, y_j)} \text{ and } \nu = \sum_{j=l+1}^{m+l} a_j \frac{\delta_{x_j} - \delta_{y_j}}{d(x_j, y_j)},$$

with $a_j \geq 0$, $j = 1, 2 \ldots, l + m$. By Corollary 2.11

$$1 = \left\| \frac{\delta_x - \delta_y}{d(x,y)} \right\|_{\mathcal{F}} \leq \alpha\|\mu\|_{\mathcal{F}} + (1 - \alpha)\|\nu\|_{\mathcal{F}} = \alpha \sum_{j=1}^{l} a_j + (1 - \alpha) \sum_{j=l+1}^{l+m} a_j = 1.$$

This implies that

$$\frac{\delta_x - \delta_y}{d(x,y)} = \sum_{j=1}^{m+l} b_j \frac{\delta_{x_j} - \delta_{y_j}}{d(x_j, y_j)}$$

with $b_j = \alpha a_j$, if $j = 1, 2 \ldots, l$ and $b_j = (1 - \alpha)a_j$, if $j = l + 1, l + 2, \ldots, l_{l+m}$, is an optimal representation of $\frac{\delta_x - \delta_y}{d(x,y)}$, and thus, from Corollary 2.11 it follows that $x_j = x$ and $y_j = y$, for $j = 1, 2, \ldots, , m + l$. This implies that $\frac{\delta_x - \delta_y}{d(x,y)}$, is an extreme point of $B_{\mathcal{F}(M)}$. □

2.3. **Some Notational Remarks.** In the literature there are several names for the space $\mathcal{F}(M)$: Other than the *Lipschitz free space over M*, $\mathcal{F}(M)$ is also called

- *Transportation Cost Space*,
- *Wasserstein Space* or more precisely *Wasserstein 1-Space*,
- *Arens Eals* (denoted by Æ),
- *Earthmover Space.*

Let us introduce some more notation

Denote the set of measures on M, with finite support, by $\mathcal{M}$ *i.e.,*

$$\mathcal{M}(M) = \Big\{ \sum_{j=1}^{n} a_j \delta_{x_j} : n \in \mathbb{N}, a_j \in \mathbb{R}, x_j \in M, \text{ for } j = 1, 2, \dots n \Big\}$$

(actually $\mathcal{M}(M) = \tilde{\mathcal{F}}(M)$). Let $\mathcal{M}^+(M)$ denote the positive measures and $\mathcal{P}(M)$ denote probabilities on M, with finite support. For $\sigma, \tau \in \mathcal{P}(M)$ define the *Wasserstein distance of σ and τ* by

$$d_{\mathrm{Wa}}(\sigma, \tau) = \|\sigma - \tau\|_{\mathcal{F}} = \|\sigma - \tau\|_{\mathrm{tc}}$$

Thus if we let $\mu = \sigma - \tau$, it follows by the Remark after Proposition 2.4 that

$$d_{\mathrm{Wa}}(\sigma, \tau) = \inf \left\{ \sum_{(x,y)\in M^2} \nu(x, y) d(x, y) : \begin{array}{c} \nu \in \mathcal{M}(M^2), \\ \mu^+(x) = \sum_{y'\in M} \nu(x, y') \text{ for } x \in M \text{ and} \\ \mu^-(y) = \sum_{x'\in M} \nu(x', y) \text{ for } y \in M \end{array} \right\}$$

We claim that

$$d_{\mathrm{Wa}}(\sigma, \tau) = \inf \left\{ \sum_{(x,y)\in M^2} \pi(x, y) d(x, y) : \begin{array}{c} \pi \in \mathcal{P}(M^2), \\ \sigma(x) = \sum_{y'\in M} \pi(x, y') \text{ for } x \in M \text{ and} \\ \tau(y) = \sum_{x'\in M} \pi(x', y) \text{ for } y \in M \end{array} \right\}.$$

(which is the usual definition of the Wasserstein distance).

Indeed $\sigma - \tau = \mu = \mu^+ - \mu^-$ and thus, for every $x \in M$

$$\rho(x) := \sigma(x) - \mu^+(x) = \tau(x) - \mu^-(x)$$

Let $\nu \in \mathcal{M}^+(M^2)$ be such that

$$\mu^+(x) = \sum_{y'\in M} \nu(x, y'), \text{ for } x \in M \text{ and } \mu^-(y) = \sum_{x'\in M} \nu(x', y), \text{ for } y \in M. \tag{8}$$

Then define $\pi = \nu + \sum_{x\in M} \delta_{(x,x)} \rho(x)$ and note that

$$\sigma(x) = \sum_{y'\in M} \pi(x, y'), \text{ for } x \in M \text{ and } \tau(y) = \sum_{x'\in M} \pi(x', y), \text{ for } y \in M,$$

and thus $\pi \in \mathcal{P}(M^2)$. Moreover, we have

$$\sum_{(x,y)\in M^2} \nu(x,y)d(x,y) = \sum_{(x,y)\in M^2} \pi(x,y)d(x,y), \tag{9}$$

Similarly it follows that if $\pi \in \mathcal{P}(M^2)$ satisfies (9) then $\nu = \pi - \sum_{x\in M} \delta_{(x,x)}\pi(x,x)$ satisfies (8). We define for $\sigma, \tau \in \mathcal{P}(M)$ the *Transition Probabilities from* σ *to* τ

$$\mathcal{P}(\sigma,\tau) = \Big\{\pi \in \mathcal{P}(M^2) : \sigma(x) = \sum_{y'\in M} \pi(x,y'), \tau(y) = \sum_{x'\in M} \pi(x',y), \text{ for } x,y \in M\Big\}$$

and can rewrite $d_{\mathrm{Wa}}(\sigma,\tau)$ as

$$d_{\mathrm{Wa}}(\sigma,\tau) = \min\Big\{\sum_{x,y\in M} \pi(x,y)d(x,y) : \pi \in \mathcal{P}(\sigma,\tau)\Big\}. \tag{10}$$

Definition 2.13. We call $\mathcal{P}(M)$ together with the metric $d_{\mathrm{Wa}}(\cdot,\cdot)$ *Wasserstein space*, and denote it by $\mathrm{Wa}(M)$.

More generally if $\mu_1, \mu_2 \in \mathcal{M}^+(M)$, with $\mu_1(M) = \mu_2(M)$, we put

$$\mathcal{M}^+(\mu_1,\mu_2) = \Big\{\nu \in \mathcal{M}^+(M^2), : \mu_1(x) = \sum_{y'\in M} \nu(x,y'), \mu_2(y) = \sum_{x'\in M} \nu(x',y), \text{ for } x,y\in M\Big\}$$

and conclude that

$$\|\mu_1 - \mu_2\|_{\mathcal{F}} = \min\Big\{\underbrace{\sum_{x,y\in M} \nu(x,y)d(x,y)}_{=\int_{M^2} d(x,y)\,d\nu(x,y)} : \nu \in \mathcal{M}^+(\mu_1,\mu_2)\Big\}. \tag{11}$$

2.4. **Uniform Distibutions.** For $A \subset M$ we denote the uniform distribution on A by μ_A,*i.e.*, $\mu_A(x) = \frac{1}{|A|}\chi_A(x)$.

Proposition 2.14. *If* $A, B \subset M$ *with* $n = |A| = |B|$, *then there exist a bijection* $f : A \to B$ *So that*

$$d_{\mathrm{Wa}}(\mu_A, \mu_B) = \frac{1}{n}\sum_{x\in M} d(x, f(x))$$

In other words, the representation

$$\mu_A - \mu_B = \frac{1}{n}\sum_{x\in A} \big(\delta_x - \delta_{f(x)}\big)$$

is optimal.

The proof is a Corollary of the following theorem by Birkhoff (see appendix for a proof)

Theorem 2.15. *(Birkhoff) Assume $n \in \mathbb{N}$ and that $A = (a_{i,j})_{i,j=1}^n$ is a doubly stochastic matrix, i.e.,*

$$0 \le a_{i,j} \le 1 \textit{ for all } 1 \le i,j \le n,$$
$$\sum_{j=1}^n a_{i,j} = 1 \textit{ for } i = 1,2,\dots,n, \textit{ and } \sum_{i=1}^n a_{i,j} = 1 \textit{ for } j = 1,2,\dots,n.$$

Then A is a convex combination of permutation matrices, i.e.,matrices which have in each row and each column exactly one entry whose value is 1 and vanish elsewhere.

Proof of Proposition 2.14. Let $A = \{x_1, x_2, \dots x_n\}$ and $B = \{y_1, y_2, \dots, y_n\}$. We note that for every $\pi \in \mathcal{P}(\sigma,\tau)$ the matrix
$M = (n\pi(x_i, y_j) : 1 \le i,j \le n)$ is a doubly stochastic matrix (since $\sum_{x\in A} \pi(x,y) = \tau(y) = \frac{1}{|B|} = \frac{1}{|A|} = \sigma(x) = \sum_{y\in B} \pi(x,y)$). Thus by (10)

$$d_{\mathrm{Wa}}(\mu_A, \mu_B) = \frac{1}{n} \min\Big\{ \sum_{i,j=1}^n M_{i,j} d(x_i, y_j) : M \in \mathrm{DS}_n \Big\}$$

Since the map

$$\mathrm{DS} \to [0,\infty), \quad M \mapsto \sum_{i,j=1}^n M_{i,j} d(x_i, y_j)$$

is linear, it achieves its minimum on an extreme point; our claim follows from Theorem 2.15 □

3. Embeddings of Transportation cost spaces over trees into L_1

3.1. **Distortion.** Let (M,d) and (M',d') be two metric spaces. For $f:M\to M'$ the *distortion of* f is defined by

$$\mathrm{dist}(f)=\underbrace{\sup_{x\neq y,x,y\in M}\frac{d'(f(x),f(y))}{d(x,y)}}_{\mathrm{Lip}(f)}\underbrace{\sup_{x\neq y,x,y\in M}\frac{d(x,y)}{d'(f(x),f(y))}}_{\mathrm{Lip}(f^{-1})}$$

where f^{-1} is defined on $f(M)$, if f is injective, and otherwise $\mathrm{dist}(f):=\infty$.

We define the M'- *distortion of* M by

$$c_{M'}(M)=\inf\big\{\mathrm{dist}(f)|\, f:M\to M'\big\}.$$

Let $\mathcal{M}'$ be a family of metric spaces. We define the $\mathcal{M}'$*-distortion of* M by

$$c_{\mathcal{M}'}(M)=\inf_{M'\in\mathcal{M}'}c_{M'}(M).$$

The main question we want to address:

Problem. Let (M,d) be a finite metric space. Find upper and lower estimates of

- $c_{(\ell_1^n,n\in\mathbb{N})}(\mathcal{F}(M))$,
- $c_{(\ell_1^n,n\in\mathbb{N})}(\mathrm{Wa}(M))$,
- $\inf\big\{\|T\|_{\mathcal{F}(M)\to L_1}\cdot\|T^{-1}\|_{T(L_1)\to\mathcal{F}(M)}:T:\mathcal{F}\to L_1 \text{ linear and bounded}\big\}$.

Our families of metric spaces are usually closed under scaling, *i.e.,* if $(M',d')\in\mathcal{M}'$ and $\lambda>0$, then also $(\mathcal{M}',\lambda\cdot d')\in\mathcal{M}'$. In that case

$$c_{\mathcal{M}'}(M)=\inf_{M'\in\mathcal{M}'}\{\|f\|_{\mathrm{Lip}}:f:M\to M' \text{ is expansive}\}$$

where $f:(M,d)\to(M',d')$ is called *expansive*, if

$$d'(f(x),f(z))\geq d(x,z),\ \text{for } x,z\in M.$$

3.2. **Lipschitz embeddings of $\mathrm{Wa}(M)$ into L_1 imply linear embeddings of $\mathcal{F}(M)$ into $\ell_1(N)$.** Throughout this subsection, we assume that (M,d) is a finite metric space and let $n=|M|$.

Theorem 3.1. *Assume that*

$$F:\mathcal{P}(M)\to L_1[0,1]$$

has the property that for some $L\geq 1$

$$d_{\mathrm{Wa}}(\sigma,\tau)\leq\|F(\sigma)-F(\tau)\|_1\leq Ld_{\mathrm{Wa}}(\sigma,\tau) \text{ for } \sigma,\tau\in\mathcal{P}(M).$$

Then, there exists a Lipschitz map

$$H:\mathcal{F}(M)\to L_1[0,1]$$

for which

$$\|\mu-\nu\|_{\mathcal{F}}\leq\|H(\mu)-H(\nu)\|_1\leq 3L\|\mu-\nu\|_{\mathcal{F}},\ \text{for } \mu,\nu\in\mathcal{F}(M).$$

Remark. Using more sophisticated tools, we can obtain a linear bounded operator

$$T : \mathcal{F}(M) \to L_1[0,1]$$

for which

$$\|\mu\|_{\mathcal{F}} \le \|T(\mu)\|_1 \le L\|\mu\|_1.$$

Here, we are following a more elementary but also more technical proof by Naor and Schechtman [17].

Proof. After scaling we can assume that $d(u,v) \ge 1$ for all $u \ne v$ in M, and secondly we can assume that the image of the uniform distribution, $\mu_0 = \frac{1}{n}\sum_{x\in M}\delta_x$, under F vanishes. Note that for $\mu \in \mathcal{F}(M)$

$$\|\mu\|_\infty = \max_{x\in M}|\mu(x)| \le \|\mu\|_{\mathcal{F}}.$$

Indeed, if $\nu \in \mathcal{M}^+(\mu^+,\mu^-)$, then

$$\int_{M\times M} d(x,y)\,d\nu(x,y) \ge \int_{M\times M} 1\,d\nu(x,y) = \int_M 1\,d\mu^+(x) = \mu^+(M) = \mu^-(M) \ge \|\mu\|_\infty,$$

and thus, taking the infimum over all $\nu \in \mathcal{M}^+(\mu^+,\mu^-)$, we deduce our claim.

We put

$$B = \{\mu \in \mathcal{F}(M), \|\mu\|_\infty \le 1\}.$$

$$\Psi : B \to \mathcal{P}(M), \quad \mu \mapsto \sum_{x\in M}\frac{1+\mu(x)}{n}\delta_x \in \mathcal{P}(M).$$

For any $f \in \mathrm{Lip}_0(M)$, $\mu,\nu \in B$

$$\int_M f(x)d(n\Psi(\mu) - n\Psi(\nu)) = \sum_{x\in M} f(x)(\mu(x)-\nu(x)) = \int_M f(x)d(\mu-\nu),$$

and thus

$$\|\mu-\nu\|_{\mathcal{F}(M)} = n\|\Psi(\mu)-\Psi(\nu)\|_{\mathcal{F}(M)} = nd_{\mathrm{Wa}}\big(\Psi(\mu),\Psi(\nu)\big).$$

Then define

$$h : B \to L_1[0,1] \quad \mu \mapsto nF\circ\Psi(\mu).$$

Then $h(0) = 0$ (because the image of the uniform distribution under F was assumed to vanish) and for $\mu,\nu \in B$

$$\begin{aligned}
(12)\qquad &\|h(\mu)-h(\nu)\|_1\\
&= n\|F(\Psi(\mu)) - F(\Psi(\nu))\|_1\\
&= n\Big\|F\Big(\sum_{x\in M}\frac{1}{n}(1+\mu(x))\delta_x\Big) - F\Big(\sum_{x\in M}\frac{1}{n}(1+\nu(x))\delta_x\Big)\Big\|_1\\
&\ge n\Big\|\sum_{x\in M}\frac{1}{n}(1+\mu(x))\delta_x - \sum_{x\in M}\frac{1}{n}(1+\nu(x))\delta_x\Big\|_{\mathcal{F}} \quad \text{(expansiveness)}\\
&= \|\mu-\nu\|_{\mathcal{F}},
\end{aligned}$$

and

$$\|h(\mu)-h(\nu)\|_1 \tag{13}$$
$$= n\Big\|F\Big(\sum_{x\in M}\frac{1}{n}(1+\mu(x))\delta_x\Big)-F\Big(\sum_{x\in M}\frac{1}{n}(1+\nu(x))\delta_x\Big)\Big\|_1$$
$$\le nLd_{\mathrm{Wa}}\Big(\sum_{x\in M}\frac{1}{n}(1+\mu(x))\delta_x,\sum_{x\in M}\frac{1}{n}(1+\nu(x))\delta_x\Big)=\|\mu-\nu\|_{\mathcal F}. \tag{14}$$

We define $\chi(f):[0,1]\times\mathbb{R}\to\{-1,0,1\}$, for $f\in L_1[0,1]$, by

$$\chi(f)(s,t)=\operatorname{sign}\big(f(s)\big)1_{[0,|f(s)|]}(t)=\begin{cases}1 & \text{if } f(s)>0 \text{ and } 0\le t\le f(s),\\ -1 & \text{if } f(s)<0 \text{ and } 0\le t\le -f(s),\\ 0 & \text{else.}\end{cases}$$

Note that (by Fubini's Theorem)

$$\|\chi(f)-\chi(g)\|_{L_1([0,1]\times\mathbb{R})}=\|f-g\|_{L_1[0,1]},\ \text{ for } f,g\in L_1[0,1]. \tag{15}$$

Indeed, note that $f(s)\ge g(s) \iff \chi(f)(s,t)\ge\chi(g)(s,t)$ and that, since $\chi(f)$ and $\chi(g)$ are $-1,0,1$ valued, we have

$$\int_0^1\int_{-\infty}^{\infty}|\chi(f)(s,t)-\chi(g)(s,t)|\,dtds$$
$$=\int_{\{s:f(s)>g(s)\}}\int_{-\infty}^{+\infty}\chi(f)(s,t)-\chi(g)(s,t)\,dtds$$
$$+\int_{\{s:f(s)<g(s)\}}\int_{-\infty}^{+\infty}\chi(g)(s,t)-\chi(f)(s,t)\,dtds$$
$$=\int_{\{s:f(s)>g(s)\}}f(s)-g(s)ds+\int_{\{s:f(s)<g(s)\}}g(s)-f(s)ds=\|f-g\|_{L_1[0,1]}.$$

Now we put (recall that $\mu/\|\mu\|_{\mathcal F}\in B$, for $\mu\in\mathcal{F}(M)$)

$$H:\mathcal{F}(M)\to L_1([0,1]\times\mathbb{R}),\quad \mu\mapsto\|\mu\|_{\mathcal F}\cdot\chi\circ h\Big(\frac{\mu}{\|\mu\|_{\mathcal F}}\Big)\text{if } \mu\neq0, \text{ and } H(0)=0.$$

It follows for $\mu,\nu\in\mathcal{F}(M)$, with $\|\mu\|_{\mathcal F}\ge\|\nu\|_{\mathcal F}$ (w.l.o.g.), that

$$\big|H(\mu)-H(\nu)\big|=\Big|\|\mu\|_{\mathcal F}\cdot\chi\circ h\Big(\frac{\mu}{\|\mu\|_{\mathcal F}}\Big)-\|\nu\|_{\mathcal F}\cdot\chi\circ h\Big(\frac{\nu}{\|\nu\|_{\mathcal F}}\Big)\Big|$$
$$=\Big|\|\nu\|_{\mathcal F}\Big(\chi\circ h\Big(\frac{\mu}{\|\mu\|_{\mathcal F}}\Big)-\chi\circ h\Big(\frac{\nu}{\|\nu\|_{\mathcal F}}\Big)\Big)$$
$$+(\|\mu\|_{\mathcal F}-\|\nu\|_{\mathcal F})\cdot\chi\circ h\Big(\frac{\mu}{\|\mu\|_{\mathcal F}}\Big)\Big|$$

$$= \|\nu\|_{\mathcal{F}}\Big|\chi\circ h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big)-\chi\circ h\Big(\frac{\nu}{\|\nu\|_{\mathcal{F}}}\Big)\Big|$$
$$+ (\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}})\Big|\chi\circ h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big)\Big|$$
$$\Big(\text{Check possible values of } \chi\circ h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big), \chi\circ h\Big(\frac{\nu}{\|\nu\|_{\mathcal{F}}}\Big)\in\{-1,0,1\}\Big).$$

Thus

$$\begin{aligned}\big\|H(\mu)-H(\nu)\big\|_{L_1([0,1]\times\mathbb{R})} &= \|\nu\|_{\mathcal{F}}\Big\|\chi\circ h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big)-\chi\circ h\Big(\frac{\nu}{\|\nu\|_{\mathcal{F}}}\Big)\Big\|_1\\
&\qquad + (\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}})\Big\|h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big)\Big\|_1\\
&= \|\nu\|_{\mathcal{F}}\Big\|h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big)-h\Big(\frac{\nu}{\|\nu\|_{\mathcal{F}}}\Big)\Big\|_1 \text{ (By (15))}\\
&\qquad + (\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}})\Big\|h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big)\Big\|_1\\
&\geq \|\nu\|_{\mathcal{F}}\Big\|\frac{\mu}{\|\mu\|_{\mathcal{F}}}-\frac{\nu}{\|\nu\|_{\mathcal{F}}}\Big\|_{\mathcal{F}}+\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}} \text{ (By (12))}\\
&\geq \|\nu-\mu\|_{\mathcal{F}}-\Big\|\mu\Big(1-\frac{\|\nu\|_{\mathcal{F}}}{\|\mu\|_{\mathcal{F}}}\Big)\Big\|+\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}}. = \|\mu-\nu\|_{\mathcal{F}}\end{aligned}$$

We also get

$$\begin{aligned}\big\|H(\mu)-H(\nu)\big\|_{L_1([0,1]\times\mathbb{R})} &= \|\nu\|_{\mathcal{F}}\Big\|h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big)-h\Big(\frac{\nu}{\|\nu\|_{\mathcal{F}}}\Big)\Big\|_1\\
&\qquad + (\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}})\Big\|h\Big(\frac{\mu}{\|\mu\|_{\mathcal{F}}}\Big)\Big\|_1\\
&\leq L\|\nu\|_{\mathcal{F}}\Big\|\frac{\mu}{\|\mu\|_{\mathcal{F}}}-\frac{\nu}{\|\nu\|_{\mathcal{F}}}\Big\|_1\\
&\qquad + L(\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}}) \text{ (By (13))}\\
&= L\Big\|\frac{\|\nu\|_{\mathcal{F}}}{\|\mu\|_{\mathcal{F}}}\mu-\nu\Big\|_{\mathcal{F}}+L(\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}})\\
&\leq L\Big(\Big\|\frac{\|\nu\|_{\mathcal{F}}}{\|\mu\|_{\mathcal{F}}}\mu-\mu\Big\|_{\mathcal{F}}+\|\mu-\nu\|_{\mathcal{F}}\Big)+L(\|\mu-\nu\|_{\mathcal{F}})\\
&= L\Big(\|\mu\|_{\mathcal{F}}-\|\nu\|_{\mathcal{F}}\Big)+2L(\|\mu-\nu\|_{\mathcal{F}})\leq 3L(\|\mu-\nu\|_{\mathcal{F}}).\end{aligned}$$

Thus H is a bi- Lipschitz map from $\mathcal{F}(M)$ to $L_1([0,1]\times\mathbb{R})\equiv L_1[0,1]$ of distortion not larger than $3L$. □

Remark. If $F:\mathcal{P}(M)\to L_1[0,1]$ is Lipschitz, we could perturb F a bit, to a map $\tilde{F}$ having a finite-dimensional image and almost the same distortion. If we could then produce a Lipschitz map

$H : \mathcal{F}(M) \to L_1[0,1]$, which also has a finite-dimensional image we would only need Rademacher's Theorem to linearize H,

The following result is a generalization of Rademacher's Theorem by Heinrich and Mankiewicz.

Let X and Z be Banach spaces and let $f : X \to Z^*$ be a Lipschitz map. We say that f is w^* differentiable at a point x_0 if for all $x \in X$

$$(D^* f)_{x_0}(x) = w^* - \lim_{\lambda \to 0} \frac{f(x_0 + \lambda x) - f(x_0)}{\lambda} \text{ for all } x \in X \text{ exists.}$$

We call $(D^* f)_{x_0}$ the w^*-derivative of f at x_0.

Theorem 3.2. [12, Theorem 3.2] *Let X and Z be Banach spaces, Z being separable, and $f : X \to Z^*$ be a Lipschitz map. Then there is a dense set $D \subset X$ so that for all $x_0 \in D$ the w^* derivative $(D^* f)_{x_0}$ exists.*

Moreover:

1. *For every $x \in D$, $(D^* f)_{x_0}$ is bounded linear operator from X to Z^* and $\|(D^* f)_{x_0}\| \le \|f\|_{\mathrm{Lip}}$.*
2. *If, moreover, f is bi-Lipschitz, then $(D^* f)_{x_0}$ is an isomorphic embedding and $\|(D^* f)^{-1}_{x_0}\| \le \|f^{-1}\|_{\mathrm{Lip}}$.*

Corollary 3.3. *Assume that*

$$F : \mathcal{P}(M) \to L_1[0,1]$$

has the property that

$$d_{\mathrm{Wa}}(\sigma, \tau) \le \|F(\sigma) - F(\tau)\|_1 \le L d_{\mathrm{Wa}}(\sigma, \tau), \text{ for } \sigma, \tau \in \mathcal{P}(M) \tag{16}$$

and let $\varepsilon > 0$

Then there exists an $N \in \mathbb{N}$, and an linear embedding T of $\mathcal{F}(M)$ into ℓ_1^N so that

$$\|\mu\|_{\mathcal{F}} \le \|T(\mu)\|_1 \le (3L + \varepsilon)\|\mu\|_{\mathcal{F}}.$$

Proof. Let $H : \mathcal{F}(M) \to L_1[0,1]$ be defined as in Theorem 3.1. Since $L_1[0,1]$ is isometrically a subspace of $C^*[0,1]$ we can apply Theorem 3.2 and obtain an isomorphic embedding S of $\mathcal{F}(M)$ into $C^*[0,1]$, with

$$\|\mu\|_{\mathcal{F}} \le \|S(\mu)\|_1 \le 3L\|\mu\|_{\mathcal{F}}.$$

Since $\mathcal{F}(M)$ is finite-dimensional $S(\mathcal{F}(M))$ is also finite-dimensional. $C^*[0,1]$ is a $\mathcal{L}_1$-space of constant 1, which means that for every finite dimensional subspace F of $C^*[0,1]$ there is a finite dimensional subspace G of $C^*[0,1]$ which contains F, and which is $(1+\varepsilon)$-isomorphic to ℓ_1^N, for some N and $(1+\varepsilon)$-complemented in $C^*[0,1]$ (the complementation is not needed). We deduce from this our claim. □

Exercise 3.4. Prove that $C^*[0,1]$ is a $\mathcal{L}_1$-space.

3.3. **Geodesic Graphs.** An undirected graph G is a pair $G = (V(G), E(G))$ with

$$E(G) \subset [V(G)]^2 = \{e \subset V(G) : |e| = 2\}.$$

For v we call

$$\deg(v) = \big\{e \in E(G) : v \in e\big\}$$

the *degree of* v.

A *walk* in a graph G is a graph $W = (V(W), E(W))$ with $V(W) \subset V(G)$, and $E(W) \subset E(G)$, and $V(W)$ can be ordered into $\{x_j : j = 0, 1, 2, \dots n\}$ (where the x_j are not necessarily distinct) so that $E(W) = \big\{\{x_{j-1}, x_j\} : j = 1, 2 \dots n\big\}$. In that case we call W *a walk from* x_0 *to* x_n, and also write $W = (x_j)_{j=0}^n$.

In the case that $x_j \neq x_i$, if $i \neq j$ in $\{0, 1, 2, \dots, n\}$ we call W a path. If $x_0 = x_n$, $x_j \neq x_i$, unless $\{i, j\} = \{0, n\}$, we call W a *cycle*.

We call a graph $G = (V(G), E(G))$ *connected* if for each $u, v \in V(G)$ there is a walk (and thus also a path) from u to v. A connected graph that does not contain a cycle is called a *tree*. In that case, a unique path exists between any two vertices u and v. We denote that unique path by $[u, v]_G$.

Proposition 3.5. *Let $G = (V(G), E(G))$. The following statements are equivalent:*

1. *G is a tree,*
2. *G is minimal connected graph, i.e., for every $e \in E(G)$, the graph $G' = (V(G), E(G) \setminus \{e\})$ is not connected,*
3. *G is a maximal graph without a cycle, i.e., for every $e \in [V(G)]^2 \setminus E(G)$, the graph $\tilde{G} = (V(G), E(G) \bigcup \{e\})$ has a cycle,*

and in the case that $n = |V(G)| < \infty$

4. *G is connected and $|E(G)| = n - 1$,*

Proof. Exercise. □

Definition. Let $T = (V(T), E(T))$ be a tree we call $v \in V(T)$ a *leaf of* T if $\deg(v) = 1$,

Exercise 3.6. Every finite tree has leaves.

If G is a connected graph and d_G is a metric on $V(G)$, we call d_G a *geodesic metric on* G if

$$d_G(u, v) = \min\{\text{length}_{d_G}(p) : p \text{ is a path from } u \text{ to } v\}, \text{ for } u, v \in V,$$

where for a path $p = (x_j)_{j=0}^n$ in G we define the length of p by

$$\text{length}_{d_G}(p) = \sum_{j=1}^{n} d_G(x_{j-1}, x_j).$$

In that case we call the pair (G, d_G) a *geodesic graph*. For $e = \{u, v\} \in E(G)$ we put $d_G(e) = d_G(u, v)$. If G is a cycle, a path, or a tree, we call it a *geodesic cycle*, *geodesic path*, or a *geodesic tree*, respectively.

Assume that $w : E(G) \to \mathbb{R}^+$ is a function. Define for $u, v \in V(G)$

$$d_G(u,v) := \min\Big\{ \sum_{j=1}^n w(\{x_{i-1},x_i\}) : (x_j)_{j=0}^n \text{ is a path from } u \text{ to } v\Big\}.$$

Then d_G is a geodesic metric on G, and we call it the *metric generated by the weight function w*.

This is why geodesic graphs are often referred to as *weighted graphs*.

3.4. **Transportation cost spaces over geodesic trees.** Let (M,d) be a metric space. Recall the following notation from 2.3: for $\mu = \mu^+ - \mu^- \in \tilde{\mathcal{F}}(M)$,

$$\mathcal{M}^+(\mu_1,\mu_2) = \Big\{ \nu \in \mathcal{M}^+(M^2) : \mu_1(x) = \sum_{y'\in M} \nu(x,y'), \mu_2(y) = \sum_{x'\in M} \nu(x',y), \text{ for } x,y \in M\Big\}.$$

We showed

$$\|\mu\|_{\mathcal{F}} = \|\mu\|_{\mathrm{tc}} = \min\Big\{ \sum_{x,y\in M} \underbrace{d(x,y)\, d\nu(x,y)}_{=\int_{M^2} d(x,y)\,\nu(x,y)} : \nu \in \mathcal{M}^+(\mu^+,\mu^-)\Big\}.$$

Throughout this subsection $T = (V(T), E(T))$ is a, possibly infinite, tree, and d_T is a geodesic distance. Let $w : E(T) \to (0,\infty)$, $e \mapsto d(e)$ be the corresponding weight function.

Our goal is to show that $\mathcal{F}(V(T), d_T)$ is isometrically isomorphic to the *weighted ℓ_1-space* $\ell_1(E(T), w)$, with the norm

$$\|x\|_1 = \sum_{e\in E(T)} |x_e| w(e) \text{ for } x = (x_e : e \in E(T)) \subset \mathbb{R}.$$

We first introduce some notation.

Recall that for $u, v \in V(T)$, $[u,v]_T$ denotes the unique path from u to v. We choose a *root* $v_0 \in V$ and note that for any $e = \{u,v\} \in E(T)$

$$\text{Either } u \in [v_0,v]_T \text{ or } v \in [v_0,u]_T.$$

We assign to $e \in \{u,v\} \in E(T)$ an orientation, and choose (u,v) iff $u \in [v_0,v]_T$, which is equivalent to $d(v_0,u) < d(v_0,v)$ and define

$$E_d(T) = \big\{(u,v) : \{u,v\} \in E(T) \text{ and } d(v_0,u) < d(v_0,v\big\},$$

which means that if $e = \{u,v\} \in E(T)$, and v is further awy from v_0 than u then (u,v) is the orientation of e. For $e = (u,v) \in E_d(T)$ we put $e^- = u$ and $e^+ = v$. We define a partial order in $V(T)$ by

$$u \preceq v :\iff [v_0,u]_T \subset [v_0,v]_T.$$

From the uniqueness of paths in T, we deduce the following facts

Exercise 3.7. (1) For every $v \in V(T)$, we have $\{u \in V(t) : u \preceq v\} = [v_0,v]_T$, and thus is linearly ordered with respect to $\preceq$.

(2) If $v \in V(T)$, $v \neq v_0$, there is a unique $e \in E(T)$ with $e^+ = v$. We call, in that case, e^- the *immediate predecessor of v*.

(3) (Tripot property of trees) If $u, v \in V(T)$ then there exists a (unique) $z \in V(T)$ for which $[v_0, z]_T = [v_0, u]_T \cap [v_0, v]_T$ which we denote by $\min(u, v)$. Moreover, it follows in this case that $d_T(u, v) = d_T(u, z) + d_T(z, v)$.

Proposition 3.8. *The map $f : (V(T), d_T) \to \ell_1(E(T), w)$, $v \mapsto 1_{E([v_0,v]_T}$ is an isometry.*
Here $1_{E([v_0,v]_T)} : E(T) \to \mathbb{R}$ is seen as element of $\ell_1(E(T), w)$ defined by

$$1_{E([v_0,v]_T)}(e) := \begin{cases} 1 & \textit{if } e \in E([v_0, v]_T), \\ 0 & \textit{else.} \end{cases}$$

Proof. Let $u, v \in V(T)$ and let $z = \min(v, u)$. Since $[v, z]_T \cup [z, u]_T$ is the (unique!) path from u to v, it follows that

$$\begin{aligned} d_T(u, v) &= \sum_{e \in E([u,z])} w(e) + \sum_{e \in E([z,v])} w(e) \\ &= \sum_{E \in E(T)} |1_{E([u,z])}(e) - 1_{E([z,v])}(e)| w(e) \\ &= \sum_{E \in E(T)} |1_{E([v_0,u])}(e) - 1_{E([v_0,v])}(e)| w(e) = \|f(u) - f(v)\|_1. \end{aligned}$$

□

We use now the Extension property of $\mathcal{F}(V(T), d_T)$ and denote the unique linear extension of f by F, and recall that $\|F\|_{\mathcal{F}(V(T),d_T) \to \ell_1(E(T),w)} = 1$. By linearity of $F : \mathcal{F}(M) \to \ell_1(E(T), w)$ we deduce that for any $\mu = \mu^+ - \mu^-$ and any $\nu \in \mathcal{M}^+(\mu^+, \mu^-)$, that

$$F(\mu) = F\Big(\sum_{x,y \in V(T)} \nu(x, y)(\delta_x - \delta_y)\Big) = \sum_{x,y \in V(T)} \nu(x, y)(f(x) - f(y))$$

and thus

$$\|F(\mu)\|_1 \le \|\mu\|_{\mathcal{F}} = \|(\mu)\|_{\text{tc}} = \inf_{\nu \in \mathcal{M}^+(\mu^+,\mu^-)} \Big\| \sum_{x,y \in V(T)} \nu(x, y)(f(x) - f(y)) \Big\|_1. \tag{17}$$

For our next step, we introduce another notation: Let $e = (e^-, e^+) \in E_d(T)$, and put

$$V_e = \{v \in V(T) : e^+ \preceq v\},$$

We note that $T_e = (V_e, E(T) \cap [V_e]^2)$ and its complement $T_e^c = (V(T) \setminus V_e, E(T) \cap [V(T) \setminus V_e]^2)$ are subtrees of T.

Proposition 3.9. *For any $\mu \in F(V(T), d_T)$ it follows that*

$$\|F(\mu)\|_1 = \sum_{e \in E} w_e |\mu(V(T_e))|.$$

Remark. Since for $e \in E(T)$ the identicator function 1_{T_e}, as function on $V(T)$, is a Lipschitz function whose Lipschitz norm is $\frac{1}{w(e)}$ the term $\mu(T_e)$ is therefore well defined, even if T is not a finite tree.

Proof. W.lo.g. $\mu \in \tilde{\mathcal{F}}(M)$.

By linearity

$$F(\mu) = \sum_{v\in V(T)} \mu(v) f(v) = \sum_{v\in V(T)} \mu(v) 1_{E([v_0,v])}$$

and thus

$$\|F(\mu)\|_1 = \sum_{e\in E(T)} w(e) \Big| \sum_{v\in V(T)} \mu(v) \underbrace{1_{E([v_0,v])}(e)}_{=1 \iff v\succeq e^+ \iff v\in T_e} \Big| = \sum_{e\in E(T)} w(e)|\mu(T_e)|.$$

□

We are now able to provide an explicit formula for $\|\mu\|_{\mathrm{tc}}$, for $\mu \in \mathcal{F}(V(T), d_T)$

Theorem 3.10. *For* $\mu \in \mathcal{F}(M)$

$$\|\mu\|_{tc} = \sum_{e\in E} w_e |\mu(T_e))|. \tag{18}$$

Proof. It is enough to show (18) for $\mu \in \tilde{\mathcal{F}}(V(T), d_T)$.

From the inequality (17) and Proposition 3.9 we deduce

$$\sum_{e\in E} w_e |\mu(V(T_e))| = \|F(\mu)\|_1 \le \|\mu\|_{\mathrm{tc}}.$$

We need to show that there is a specific transportation plan $\nu \in \mathcal{M}^+(\mu^+, \mu^-)$ for which

$$t(\nu) = \sum_{x,y\in V(T)} \nu(x,y) d_T(x,y) = \sum_{e\in E} w_e |\mu(V(T_e))|.$$

We put

$$S_\mu = \bigcup \{[v_0, v]_T : v \in \mathrm{supp}(\mu)\}$$

Note that $T_\mu = (S_\mu, E(T) \cap [S_\mu]^2)$ is a finite subtree of T, we put $n_\mu = |S_\mu|$ and we will verify our claim by induction for all values of n_ν.

If $n_\nu = 0$, it follows that $\mu = 0$; thus, our claim is trivial. Assume that $n_\mu = n + 1$ and that our claim is true as long as $n_\mu \le n$.

We choose a leaf v of T_μ. Note $v \ne v_0$ (otherwise μ would be non zero multiple of δ_{v_0}) and let u be its immediate predecessor. Then we define $\mu' = \mu - \mu(v)\delta_v + \mu(v)\delta_u$. It follows that

(1) $S_{\mu'} \subset S_\mu \setminus \{v\}$, and thus $n_{\mu'} < n_\mu$,
(2) $\mu(T_{(u,v)}) = \mu(v)$ and $\mu'(T_{(u,v)}) = 0$,
(3) $\mu(T_e) = \mu'(T_e)$ for all $e \in E(T) \setminus \{(u, v)\}$.

Using the induction hypothesis, let $\nu' \in \mathcal{M}^+(\mu'^+, \mu'^-)$ be such that

$$\|\mu'\|_{\mathrm{tc}} = \sum_{x,y} \nu'(x,y) d(x,y) = \sum_{e\in E(T)} w_e |\mu'(T_e)| = \sum_{e\in E(T), e\ne(u,v)} w_e |\mu(T_e)|.$$

If $\mu(v) > 0$ then $\nu := \nu' + \mu(v)(\delta_v - \delta_u) \in \mathcal{M}^+(\mu^+, \mu^-)$ and if $\mu(v) < 0$ then $\nu := \nu' + |\mu(v)|(\delta_u - \delta_v) \in \mathcal{M}^+(\mu^+, \mu^-)$, and in both cases

$$\sum_{x,y \in V(T)} \nu(x,y) d(x,y) = \sum_{x,y \in V(T)} \nu'(x,y) d(x,y) + |\mu(v)| d(u,v) = \sum_{e \in T} w(e) |\mu(T_e|.$$

□

Corollary 3.11. *$\mathcal{F}(V(T), d_T)$ is isometrically isomorphic to $\ell_1(E(T), w)$ via the operator*

$$I : \mathcal{F}(V(T), d_T) \to \ell_1(E(T), w), \quad \sigma \mapsto \big(w_e \sigma(T_e) : e \in E(T)\big).$$

Proof. For $\mu \in \mathcal{F}(V(T), d_T)$ we have, by Theorem 3.10, $\|I(\mu)\| = \sum_{e \in E(T)} w_e |\mu(T_e)| = \|\mu\|_{\text{tc}}$. Since

$$I\Big(\frac{\delta_{e^+} - \delta_{e^-}}{d(e)} : e \in E(T)\Big)$$

is the unit vector basis of $\ell_1(E(T), w)$ I is also surjective. □

Remark. Alain Godard [8] proved Corollary 3.11 for more general trees ($\mathbb{R}$-trees), we followed a simplified version of his arguments. The most general version of Corollary 3.11 can be found in [4]. In [8], Godard also proved a converse, namely that every metric space (M, d), for which $\mathcal{F}(M, d)$ isometrically to L_1, ℓ_1 or ℓ_1^N , is isometrically equivalent to an $\mathbb{R}$-tree. We will prove that fact for a finite metric space (in which case the proof is much simpler).

Proposition 3.12. *Assume (M, d) is a finite metric space and $\mathcal{F}(M)$ is isometrically equivalent to ℓ_1^n for some $n \in \mathbb{N}$.*

Then (M, d) is isometrically equivalent to a geodesic tree, meaning there is a $E \subset [M]^2$ so that $T = (M, E)$ is a tree and d is a geodesic metric for T.

We will use the following fact:

Proposition 3.13. *A Banach space E of dimension $n \in \mathbb{N}$ is isometric to ℓ_1^n if and only if there are $x_1, x_2, \ldots, x_n \in S_E$ for which $\{\pm x_j : j = 1, 2, \ldots n\}$ are the extreme points of B_E.*

Proof of Proposition 3.12. Define

$$E = \big\{\{x, y\} \in [M]^2 : \text{ there does not exist } z \in M \setminus \{x, y\} \text{ with } d(x, y) = d(x, z) + d(z, y)\big\}.$$

Then $G = (M, E)$ is a connected graph and d is a geodesic metric with respect to G. Thus, by Theorem 2.12 the extreme points of $B_{\mathcal{F}(M)}$ are

$$\Big\{ \pm \frac{\delta_u - \delta_v}{d(u, v)} : \{u, v\} \in E\Big\}.$$

Assume that $\mathcal{F}(M)$ is isometrically equivalent to ℓ_1^n. Thus, $\dim(\mathcal{F}(M)) = n$, which implies that $|M| = n + 1$, and $B_{\mathcal{F}(M)}$ has $2n$ extreme points, which means that the cardinality of E must be n. So (M, E) is a connected graph with $n + 1$ vertices and n edges. This implies by Proposition 3.5 that (M, E) is a tree. □

4. Stochastic Embeddings of metric spaces into trees

4.1. Definition of Stochastic Embeddings, Examples.

Definition. Let $\mathcal{M}$ be a class of metric spaces, and let (M, d_M) be a metric space. A family $(f_i)_{i=1}^n$ of maps $f_i : M \to M_i$, with $(M_i, d_i) \in \mathcal{M}$, together with numbers $\mathbb{P} = (p_i)_{i=1}^\infty \subset (0, 1]$, with $\sum_{i=}^\infty p_i = 1$ is called a *D-stochastic embedding of M into elements of the class* $\mathcal{M}$ if for all $x, y \in M$ and $i = 1, 2 \ldots,$

$$d_M(x, y) \le d_i(f_i(x), f_i(y)) \text{ (expansiveness)}, \tag{19}$$

$$\mathbb{E}_\mathbb{P}\big(d_i(f_i(x), f_i(y)\big) = \sum_{i=1}^\infty p_i d_i\big(f_i(x), f_i(y)\big) \le D d_M(x, y). \tag{20}$$

In that case we say that (M, d_M) is *D-stochastically embeds into* $\mathcal{M}$. If moreover the maps $f_i : M \to M_i$ are bijections we say that $(f_i)_{i=1}^n$ together with $(p_i)_{i=1}^n$ is a *bijective D-stochastic embedding of M into elements of the class* $\mathcal{M}$.

Remark. Of course if $(f_i)_{i=1}^\infty$, $f_i : M \to M_i$, $i = 1, 2, 3, \ldots$ together with numbers $(p_i)_{i=1}^\infty$ is a bijective D-stochastic embedding of M into elements of the class $\mathcal{M}$, we can assume that as sets $M_i = M$, and that d_i is a metric on M.

We will mainly be interested in how a finite metric graph can be stochastically embedded into trees.

Example 4.1. Let $C_n = (V(C_n), E(C_n))$ be a cycle of length n. We can write $V(C_n)$ and $E(C_n)$ as $V(C_n) = \mathbb{Z}/nZ$ and $E(C_n) = \big\{\{j-1, j\} : j = 1, 2, \ldots n\big\}$ and let d be the geodesic metric generated by the constant weight function 1.

Consider for $j_0 = 1, 2, \ldots n$, the path P_{j_0} defined by $V(P_j) = V(C_n) = \{0, 1, 2, \ldots, n-1\}$ and

$$E(P_{j_0}) = E(C_n) \setminus \big\{\{j_0 - 1, j_0\}\big\}.$$

We consider on P_{j_0} the (usual *path distance*) generated by the weight function $w(e) = 1$, for $e \in P_{j_0}(e)$, and denote it by d_{j_0}

It follows that

$$d_{j_0}(j_0 - 1, j_0) = n - 1$$

and thus it follows for the identity $Id : (V(C_n), d_{C_n}) \to (V(C_n), d_{j_0})$ that $\text{dist}(Id) = n - 1$.

Let $\mathcal{T}$ be the set of all metric trees. It can be shown that the $\mathcal{T}$-distortion of a cycle C_n of length n is of the order n, ie $c_T(C_n) \ge c{\cdot}n$. In other words, there are no embeddings of cycles into trees with sublinear (with respect to n) distortion.

Nevertheless, the distortion of stochastic embedding of C_n into trees is not larger than 2:

$I = \{1, 2, \ldots n\}$, $p_i = \frac{1}{n}$, and let for $i \in I$, P_i be the above introduced path. Then $d_{C_n}(u, v) \le d_i(u, v)$, for $u, v \in V(C_n)$, and for $e = \{j-1, j\} \in E(C_n)$ it follows that

$$\sum_{i=1}^n p_i d_i(j-1, j) = \frac{1}{n}(2(n-1)) < 2.$$

4.2. **Stochastic Embeddings into Trees: the Theorem of Fakcharoenphol, Rao, and Talwar.**

Theorem 4.2. [6] *Let M be a metric space with $n \in \mathbb{N}$ elements. Then there is a $O(\log n)$ stochastic embedding of M into the class of weighted trees.*

A complete proof of Theorem 4.2 is given in the Appendix. Here, we only want to define the stochastic embedding.

Let $M = \{x_1, x_2, \dots, x_n\}$. After rescaling, we can assume that $d(x, y) > 1$ for all $x \neq y$ in M.

We introduce the following notation.

(1) $B_r(x) = \{z \in M : d(z, x) \leq r\}$ for $x \in X$ and $r > 0$.
(2) For $A \subset M$, $\text{diam}(A) = \max_{x,y \in A} d(x, y)$.
We choose $k \in \mathbb{N}$ so that $2^{k-1} < \text{diam}(M) \leq 2^k$.
(3) $\mathcal{P}_M$ denotes the sets of all partitions of M. Let $P = \{A_1, A_2, \dots A_l\} \in \mathcal{P}_M$.
P is called an r-Partition if $\text{diam}(A_j) \leq r$, $j = 1, 2 \dots l$
for $B \subset M$, let $P|_B = (A_1 \cap B, A_2 \cap B, \dots, A_l \cap B)$
(4) For two partitions $P = \{A_1, A_2, \dots A_m\}$ and $Q = \{B_1, B_2, \dots, B_n\}$ we say Q subdivides P and write $Q \succeq P$, if for each $i = 1, 2, \dots, m$ there is a $j = 1, 2 \dots, n$ with $B_j \subset A_i$
(5) Let Ω be the set of all sequences $(P^{(j)})_{j=0}^k$ so that: $P^{(j)}$ is a 2^{k-j}-partition of M, with $P^{(0)} = (M)$ and $P^{(j)} \succeq P^{(j-1)}$
Note: $P^{(k)}$ are singletons.
(6) Every $(P^{(j)})_{j=0}^k \in \Omega$ defines a tree $T = (V(T), E(T))$ as follows:
$V(T) = \{(j, A) : j = 0, 1, 2, \dots k, \text{ and } A \in P^{(j)}\}$, and
$E(T) = \big\{\{(j, B), (j-1, A)\} : j = 1, 2, \dots k \text{ and } A \subset B\}\big\}$.
The weight function is defined by $w(\{(j, B), (j-1, A)\}) = 2^{k-j}$ if $\{(j, B), (j-1, A)\} \in E(T)$

For each such tree T, we can define a map from M to the leaves of T by assigning each x the element $(k, \{x\}) \in V(T)$.

We now have to define a Probability on Ω.

We first define for each R a probability $\mathbb{P}^R$ on R-bounded partitions: Let Π_n be the set of all permutations on $\{1, 2, \dots n\}$

We consider on $\Pi \times [R/4, R/2]$ the product of the uniform distribution on Π and the uniform distribution on $[R/4, R/2]$. each pair $(\pi, r) \in \Pi \times [R/4, R/2]$ defines the following r-partition $(C_j(\pi, r))_{j=1}^l$:
$\tilde{C}_1(\pi, r) = B_r(x_{\pi(1)})$, and assuming $(\tilde{C}_1(\pi, r), \tilde{C}_2(\pi, r)), \dots, \tilde{C}_{j-1}(\pi, r)$ are defined we put $\tilde{C}_j(\pi, r) = B_r(x_{\pi(j)}) \setminus \bigcup_{i=1}^{j-1} \tilde{C}_i(\pi, r)$. Then let $\{C_j(\pi, r) : j = 1, 2 \dots l\}$ be the non empty elements of $(\tilde{C}_j(\pi, r))_{j=1}^n$ and let $\mathbb{P}^R$ be the image distribution of that mapping $(\pi, r) \mapsto C(\pi, r)$. For $B \subset M$ let $\mathbb{P}^{(R,B)}$ be the image distribution of the mapping $(\pi, r) \mapsto C(\pi, r)|_B$

Let $\mathbb{P}$ be the probability on Ω uniquely defined by

$$\mathbb{P}\big((P^{(j)})_{j=0}^k : P^{(0)} = (0, \{M\})\big) = 1, \tag{21}$$

for $j = 1, 2 \dots n$, $B \subset M$, $\mathrm{diam}(B) \le 2^{k-(i-1)}$, and any $\mathcal{A} \subset \mathcal{P}_M$

$$\mathbb{P}(P^{(i)}|_B \in \mathcal{A} | (i-1, B) \in P^{(i-1)}) = \mathbb{P}^{(2^{k-i}, B)}(\mathcal{A}). \tag{22}$$

4.3. **Bijective embeddings onto trees: The Restriction Theorem by Gupta.** Our goal is to show that for some universal constant c, every metric space with n elements can be bijectively $c\log(n)$-stochastic embedded into trees. This will follow from Theorem 4.2 and the following result by Gupta.

Theorem 4.3. [10, Theorem 1.1] *Let $T = (V, E, W)$ be a weighted tree and $V' \subset V$. Then there is $E \subset [V']^2$ and $W' : E(\mathbb{G}') \to [0, \infty)$ so that $T' = (V', E(\mathbb{G}'), W')$ is a tree*

$$\frac{1}{4} \le \frac{d_{T'}(x, y)}{d_T(x, y)} \le 2, \text{ for } x, y \in V'. \tag{23}$$

A proof of Theorem 4.3 is given in the appendix.

Corollary 4.4. *If a finite metric space (M, d) D stochastically embeds into geodesic trees, it bijectively 8-stochastically embeds into geodesic trees.*

In particular, by 4.2 is a constant $c > 0$ so that every metric space (M, d) with n elements bijectively $c\log(n)$-stochastically embeds into a weighted tree.

Exercise 4.5. The estimate in Corollary 4.4 is optimal in the following sense:

Let $\mathbb{G}_n = (\mathbb{Z}/n\mathbb{Z})^2$ is the n-t discrete torus, then there is constant $c > 0$ so that for all $n \in \mathbb{N}$ the following holds:

If T is a tree and $f : V(\mathbb{G}_n) \to V(T)$ is in an injective map so that

$$1 \le d_T(f(x), f(y)) \text{ for all } x, y \in \mathbb{G}_n, \text{ with } \{x, y\} \in E(\mathbb{G})$$

then

$$\frac{1}{|E(\mathbb{G})|} \sum_{e=\{x,y\} \in E(\mathbb{G})} d_T(f(x), f(y)) \ge c\log(n)$$

The same is true for the family of diamond graphs $(D_n)_{n\in\mathbb{N}}$ (see [11, Theorem 5.6]).

Remark. Assume that (G, d_G) is a finite geodesic graph, and that the geodesic trees (T_i, d_i), expansive maps $f_i : V(T_i) \to V(G)$, $i = 1, 2, \dots n$, together with the probability $\mathbb{P} = (p_i)_{i=1}^n$ form a bijective D-stochastic embedding of $V(G)$. Then we can assume that actually $V(T) = V(G)$, and thus $E(T_i) \subset [V(G)]^2$, and that f_i is the identity.

For $i = 1, 2, \dots n$ and $e = \{x, y\} \in E(T_i)$ (which does not need to be in E(G))

$$w'_i(e) = d_G(x, y) = \min\big\{\mathrm{length}_{d_G}(P) : P \text{ is a path in } G \text{ from } x \text{ to } y\big\} \le d_i(e).$$

and let d'_i be the geodesic metric on $V(T_i) = V(G)$ generated by w'_i.

We note that $d'_i(x, y) \le d_i(x, y)$, for $x, y \in V(G)$, but that the identity $(V(G), d_G) \to (V(G), d'_i)$ is still expansive. It follows therefore that if we replace d_i by d'_i we do not increase the stochastic distortion, may only reduce it.

4.4. Application: Embedding Transportation Cost Spaces into ℓ_1.

Theorem 4.6. [3, 16] *Let (M,d) be a countable metric space which D-stochastically embeds into geodesic trees. Then, there is an isomorphic embedding*

$$\Phi : \mathcal{F}(M) \to \ell_1^\infty$$

with

$$\|\mu\|_{\mathcal{F}} \le \|\Phi(\mu)\|_1 \le D\|\mu\|_{\mathcal{F}} \text{ for all } \mu \in \mathcal{F}(\mathcal{M}).$$

We will use the following easy fact: Let $\mathbb{P}$ be a probability on a finite or countable infinite set I, with $\mathbb{P}(i) > 0$, for $i \in I$. Put

$$L_1(\mathbb{P}, \ell_1) = \{f : I \to \ell_1\}$$

and for $f \in L_1(\mathbb{P}, \ell_1)$, we write $f(i) = \sum_{j=1}^\infty e_j f(i,j)$, where (e_j) denotes the unit vector basis of ℓ_1, and put

$$\|f\|_{L_1(\mathbb{P},\ell_1)} = \int_I \|f\|_1 d\mathbb{P} = \sum_{i\in I} \|f(i)\|_1 \mathbb{P}(i) = \sum_{i\in I}\sum_{j=1}^\infty |f(i,j)|$$

Then

$$L_1(\mathbb{P}, \ell_1) \to \ell_1^{I\times\mathbb{N}}, \quad f \mapsto \Big(\frac{f(i,j)}{\mathbb{P}(i)} : i \in I, j \in \mathbb{N}\Big).$$

is an onto isometry.

Proof of Theorem 4.6. For $i \in I$ let (T_i, d_i) let d_i be a geodesic metric on T_i, $\mathbb{P} = (p_i)_{i\in I}$ a (strictly positive) probability on I, and let $\phi_i : M \to V(T_i)$ so that

$$d_i(\phi_i(x), \phi(y)) \ge d(x,y) \text{ for } i\in I \text{ and } x,y\in M, \text{ and } \sum_{i\in I} p_i d_i(\phi_i(x), \phi_i(y)) \le Dd(x,y) \text{ for } x,y\in M.$$

We can assume that T_i are countable trees. By Corollary 3.11 $\mathcal{F}(V(T_i), d_i)$ is isometrically isomorphic to $\ell_1^{(E(T_i))}$, for $i \in I$, and thus, there are isometric embeddings $E_i : \mathcal{F}(V(T_i,)d_i) \to \ell_1$, for $i \in I$.

We define

$$\Psi_i : \tilde{\mathcal{F}}(M,d) \to \mathcal{F}(V(T_i), d_i), \quad \sum_{j=1}^n r_i(\delta_{x_i} - \delta_{y_i}) \mapsto \sum_{j=1}^n r_i(\delta_{\phi_i(x_i)} - \delta_{\phi(y_i)}).$$

Note that if $\mu \in \tilde{\mathcal{F}}(M,d)$ is represented by $\mu = \sum_{j=1}^n r_i(\delta_{x_i} - \delta_{y_i})$ then $\Psi_i(\mu)$ is represented by $\Psi_i(\mu) = \sum_{j=1}^n r_j(\delta_{\phi_i(x_j)} - \delta_{\phi_i(y_j)})$, and thus

$$\begin{aligned}\|\Psi_i(\mu)\|_{tc} &= \inf\Big\{\sum_{j=1}^n r_j d_i(u_j, v_j) : \Psi_i(\mu) = \sum_{j=1}^n r_j(\delta_{u_j} - \delta_{v_j}),\ (r_j)_{j=1}^n \subset \mathbb{R}^+\Big\}\\ &= \inf\Big\{\sum_{j=1}^n r_j d_i(\phi_i(x_j), \phi_i(y_j)) : \mu = \sum_{j=1}^n r_i(\delta_{x_j} - \delta_{y_j}),\ (r_j)_{j=1}^n \subset \mathbb{R}^+\Big\} \ge \|\mu\|_{tc}.\end{aligned}$$

For the second equality note that "$\leq$" is clear and that "$\geq$" follows from the fact (see Proposition 2.4) that among the optimal representations of $\Psi_i(\mu)$, there always must be one whose support is the support of $\Psi_i(\mu)$, which is a subset of $\phi_i(M)$.

We define

$$\Psi : \tilde{\mathcal{F}}(M,d) \to L_1(\mathbb{P},\ell_1) \equiv \ell_1, \quad \mu \mapsto \Psi(\mu) : I \ni i \to E_i \circ \Psi_i(\mu) \in \ell_1.$$

Then it follows for $\mu \in \tilde{\mathcal{F}}(M,d)$, that

$$\|\Psi(\mu)\|_{L_1(\mathbb{P},\ell_1)} = \sum_{i\in I} \pi \|\Psi_i(\mu)\|_{tc} \geq \|\mu\|_{tc}$$

and on the other hand if $\mu = \sum_{j=1}^n r_j(\delta_{x_j} - \delta_{y_j})$ is an optimal representation of μ then

$$\begin{aligned}\|\Psi(\mu)\|_{L_1(\mathbb{P},\ell_1)} &= \sum_{i\in I} p_i \|\Psi_i(\mu)\|_{tc} \leq \sum_{i\in I} p_i \sum_{j=1}^n r_j d(\phi_i(x_j),\phi_i(y_j))\\ &\leq D \sum_{i\in I} p_i \sum_{j=1}^n r_j d(x_j,y_j)) = D\|\mu\|_{tc}.\end{aligned}$$

□

4.5. **Application: Extension of Integral Operators from a Conservative Field to the whole Vector Field, and the Embedding of Transportation Cost Spaces into L_1 complementably.**

We recall some notation from *discrete Calculus*. We are given a finite graph $G = (V(G), E(G))$ with a geodesic metric d_G. We put $\bar{E}(G) = \big\{(x,y),(y,x) : \{x,y\} \in E(G)\big\}$.

A map $f : \bar{E}(G) \to \mathbb{R}$ is called a *vector field on* G if $f(x,y) = -f(y,x)$ for all $\{x,y\} \in E(G)$, and we put

$$\|f\|_\infty = \sup_{e=(x,y)\in \bar{E}(G)} |f(x,y)|.$$

We denote the space of vector fields on G by $\mathrm{VF}(G)$. Together with this norm $\mathrm{VF}(G) \equiv \ell_\infty(E(G))$

If $W = (x_j)_{j=0}^n$ is a walk in G and $f \in \mathrm{VF}(G)$, we call the *integral of f along W* to be

$$\int_W f(e) d_G(e) = \sum_{j=1}^n f(x_{j-1},x_j) d_G(x_{j-1},x_j).$$

A vector field f on G is called *conservative* if the integral along any cycle C vanishes, *i.e.,* if $C = (x_j)_{j=1}^n$ is a walk with $x_n = x_0$, then

$$\int_C f(e) d_G(e) = \sum_{j=1}^n f(x_{j-1},x_j) d_G(x_{j-1},x_j) = 0.$$

Equivalently, if for any $x, y \in V(G)$ and any two paths P and Q from x to y it follows that

$$\int_P f(e) d_G(e) = \int_Q f(e) d_G(e).$$

We denote the space of conservative vector fields on G by $\mathrm{CVF}(G)$.

Remark. If T is a tree, then $\mathrm{VF}(T) = \mathrm{CVF}(T)$.

For a map $F : V(G) \to \mathbb{R}$ we define the *gradient of* F by

$$\nabla F = \nabla_d F : \bar{E} \to \mathbb{R}, \quad (x, y) \mapsto \frac{f(y) - f(x)}{d_G(x, y)}.$$

Proposition 4.7. *For* $F : V(G) \to \mathbb{R}$

$$\|F\|_{\mathrm{Lip}} = \|\nabla F\|_\infty.$$

Proof. We need to observe that

$$\sup_{x,y \in V(G), x \neq y} \frac{|f(y) - f(x)|}{d_G(x, y)} = \sup_{\{x,y\} \in E(G)} \frac{|f(y) - f(x)|}{d_G(x, y)}.$$

Indeed, let $x, y \in V(G)$, $x \neq y$ and $P = (x_j)_{j=0}^n$ path of shortest metric length from x to y then

$$\begin{aligned}
\frac{|f(y) - f(x)|}{d_G(x, y)} &\leq \frac{\left| \sum_{j=1}^n f(x_j) - f(x_{j-1}) \right|}{\sum_{j=1}^n d_G(x_{j-1}, x_j)} \\
&\leq \frac{\sum_{j=1}^n |f(x_j) - f(x_{j-1})|}{\sum_{j=1}^n d_G(x_{j-1}, x_j)} \leq \max_{j=1,2\ldots} \frac{|f(x_j) - f(x_{j-1})|}{d_G(x_{j-1}, x_j)},
\end{aligned}$$

where the last inequality follows from iteratively applying the inequality

$$\frac{a + b}{c + d} \leq \max\left(\frac{a}{c}, \frac{b}{d}\right).$$

□

Proposition 4.8. *Fix a point* $0 \in V(G)$. *Then for every* $F \in \mathrm{Lip}_0(V(G))$, *it follows that* $\nabla F \in \mathrm{CVF}(G)$ *and*

$$\|F\|_{\mathrm{Lip}} = \|\nabla F\|_\infty.$$

Moreover

$$\nabla : \mathrm{Lip}_0(V(G)) \to \mathrm{CVF}(G)$$

is a surjective isometry whose inverse is defined by the Integral operator

$$I(f)(x) = \int_P f(e) d(e), \ \textit{for } x \in V(G).$$

where P *is any path from* 0 *to* x, *for* $x \in V(G)$.

Question: How well is $\mathrm{CVF}(G)$ complemented in $\mathrm{VF}(G)$ Equivalently, what is the smallest constant $C \geq 1$, so that the operator

$$I : \mathrm{CVF}(G) \to \mathrm{Lip}_0(V(G), d)$$

can be extended to an operator

$$\tilde{I} : \mathrm{VF}(G) \to \mathrm{Lip}_0(V(G), d).$$

Theorem 4.9. *Assume that $(V(G), d_G)$ is finite D-stochastically embeds into geodesic trees. Then the integral operator $I : \mathrm{CVF}(G) \to \mathrm{Lip}_0(V(G), d_G)$ can be extended to an operator $\tilde{I} : \mathrm{VF}(G) \to \mathrm{Lip}_0(V(G), d)$, with $\|\tilde{I}\| \le 8D$.*

Proof. Theorem 4.3 and the assumption imply that $(V(G), d_G)$ bijectively $8D$-stochastically embeds into geodesic trees. We can find $n \in \mathbb{N}$, geodesic trees (T_i, d_i), with $V(T_i) = V(G)$, for $i = 1, 2 \ldots, n$, and a probability $\mathbb{P} = (p_j)_{j=1}^n$, so that

$$d_G(x, y) \le d_i(x, y) \text{ for all } i = 1, 2 \ldots, n, \text{ and } x, y \in V(G), \text{ and} \tag{24}$$

$$\sum_{j=1}^n p_i d_i(x, y) \le 8Dd(x, y), \text{ for all } x, y \in V(G). \tag{25}$$

By the Remark after Theorem 3.3. we can assume that for each $i = 1, 2, \ldots, n$ and any $e = \{x, y\} \in E(T_i)$, it follows that $d_i(x, y) = d_G(x, y)$ and we will choose a path P_e in G from x to y, so that

$$d_i(x, y) = \mathrm{length}_{d_G}(P_e) = d_G(x, y).$$

For $f \in VF(G)$, and $i = 1, 2 \ldots, n$ we define $f^{(i)} \in \mathrm{VF}(T_i)$ (=CVF(T_i)) by

$$f^{(i)}(e) = \frac{1}{d_i(x, y)} \int_{P_e} f(e')d_G(e') = \frac{1}{d_G(x, y)} \int_{P_e} f(e')d_G(e') \text{ for } e \in E(T_i).$$

From (24) we deduce that

$$|f^{(i)}(e)| \le \|f\|_\infty \frac{d_G(x, y)}{d_i(x, y)} \le \|f\|_\infty \text{ for all } i = 1, 2, \ldots, n \text{ and } e \in E(T_i). \tag{26}$$

Denote the (unique) path from x to y in T_i by $[x, y]_i$. We define $\tilde{I}(f) \in \mathrm{Lip}_0(V(G), d_G)$ by

$$\begin{aligned}\tilde{I}(f)(x) &:= \sum_{i=1}^n p_i \int_{[0,x]_i} f^{(i)}(e)d_i(e) \\ &= \sum_{i=1}^n p_i \sum_{e \in E([0,x]_i)} f^{(i)}(e)d_i(e) = \sum_{i=1}^n p_i \sum_{e \in E([0,x]_i)} \int_{P_e} f(e')d_G(e').\end{aligned}$$

We note:

- If f is a conservative field, then $\tilde{I}(f) = I(f)$. Indeed, denote for $i = 1, 2, \ldots, n$ and $x \in V(G)$ the walk in G from 0 to x, obtained by concatenating the paths P_e, $e \in E([0, x]_i)$, by W_i Then

$$\tilde{I}(f)(x) = \sum_{i \in I} p_i \sum_{e \in [0,x]_i} \int_{p_e} f d_G(e) = \sum_{i \in I} p_i \int_{W_i} f(e)d_G(e) = I(f)(x).$$

- For $e = \{x, y\} \in E$, it follows from (26) that

$$|\tilde{I}(f)(y) - \tilde{I}(f)(x)| = \left| \sum_{i \in I} p_i \int_{[x,y]_i} f^{(i)}(e)d_i(e) \right|$$

$$\le \sum_{i\in I} p_i \left| \int_{[x,y]_i} f^{(i)}(e) d_i(e) \right|$$
$$\le \|f\|_\infty \sum_{i\in I} p_i \sum_{e\in[x,y]_i} d_i(e)$$
$$= \|f\|_\infty \sum_{i\in I} p_i d_i(x,y) \le 8D d_G(x,y) \|f\|_\infty,$$

and, thus, $\|\tilde{I}(f)\|_\infty \le 8D\|f\|_\infty$. □

Corollary 4.10. *If $(V(G), d_G)$ is a finite graph which D-stochastically embeds into geodesic trees, then $\mathcal{F}(V(G), d_G)$ is $8D$-complemented in $\ell_1(E(G))$.*

Proof. By Theorem 4.2 $\mathrm{Lip}_0(V(G), d_G) \equiv \mathrm{CVF}(G)$ is $8D$ complemented in $\mathrm{VF}(G) \equiv \ell_\infty(E(G))$. Passing to the dual we obtain that $\mathcal{F}(V(G), d_G)$ is $8D$ complemented in $\ell_\infty(E(G))$. □

Now assume that $G = (V(G), E(G))$ is a countable graph with geodesic metric d_G, which is D stochastically embeddable into trees. We cannot use Gupta's result Theorem 4.3. Therefore, we must also assume that (G, d_G) is bijectively D stochastically embeddable into trees. So let $I \subset \mathbb{N}$, $\mathbb{P} = (p_i)_{i\in I} \subset (0,1]$ with $\sum_{i\in I} p_i = 1$, and for $i \in I$ let $T_i = (V(T_i), E(T_i))$ be a tree with $V(T_i) = V(T)$ with a geodesic metric d_i, with $d_i(e) = d_G(e)$, if $e \in E(T_i)$ so that:

$$d_i\big(x,y)\big) \ge d_G(x,y), \text{ for } i \in I, \text{ and } x,y \in V(G), \tag{27}$$

$$\mathbb{E}_P\Big(d_i\big(f_i(x), f_i(y)\big)\Big) = \sum_{i\in I} p_i d_i\big(f_i(x), f_i(y)\big) \le D d_G(x,y), \text{ for } x,y \in V(G). \tag{28}$$

Choose for each $i \in I$ a root $v_i \in V(T_i)$ and define the Tree T, by *gluing the trees T_i together at v_i, i.e.,* put $V(T) = \bigcup_{i\in I}\{i\} \times V(T_i)$, and identify $(i,0)$ with $(j,0)$ for $i,j \in I$, and denote this point by 0 (the root of T) and $E(T) = \bigcup_{i\in I}\big\{\{(i,x),(i,y)\} : \{x,y\} \in E(T_i)\big\}$. We put $w_e = d_i(x,y)$, for $e = \{(i,x),(i,y)\} \in E(T)$ (and thus $\{x,y\} \in E(T_i)$). This defines a weight function on $E(T)$, which generates a geodesic metric on d_T, which has the property that $d_T(e) = d_i(x,y)$, for $e = \{(i,x),(i,y)\} \in T$.

We direct the edges of $E(T)$ by choosing the orientation $((i,x),(i,y))$ for $\{(i,x),(i,y)\} \in T$ if $d(0,(i,x)) < d(0,(i,y))$.

We now consider the maps $f_i : V(G) \to V(T)$, $x \mapsto (i,x)$ which satisfy for $x,y \in V(G)$

$$d_T\big(f_i(x), f_i(y)\big) \ge d_G(x,y) \text{ and } \mathbb{E}_P\Big(d_T\big(f_i(x), f_i(y)\big)\Big) \le D d_G(x,y). \tag{29}$$

Again, as in Theorem 4.9, we can choose for each $e \in E_d(T)$, say $e = \big((i,x),(i,y)\big)$, with $i \in I$, a path in G from x to y, which we denote by P_e, with $\mathrm{length}(P_e) = d_T(e) = d_i(e)$

For $f \in VF(G)$ and $i \in I$, and $e = \{(i,x),(i,y)\} \in E(T)$ we define

$$f^{(i)}(e) = \frac{1}{d_T(e)} \int_{P_e} f(e')\, d_G(e') = \frac{1}{d_G(e)} \int_{P_e} f(e')\, d_G(e')$$

From (29) we deduce that

$$|f^{(i)}(e)| \le \|f\|_\infty \frac{d_G(x,y)}{d_T(e)} \le \|f\|_\infty \text{ for all } i\in I \text{ and } e=((i,x),(i,y))\in E_d(T). \tag{30}$$

For $x, y \in V(G)$ and $i \in I$ denote the (unique) path from (i,x) to (i,y) in T by $[x,y]_i$. We define $\tilde{I}(f) \in \mathrm{Lip}_0(V(G), d_G)$ by

$$\begin{aligned}\tilde{I}(f)(x) &:= \sum_{i\in I} p_i \int_{[0,x]_i} f^{(i)}(e)d_T(e)\\ &= \sum_{i\in I} p_i \sum_{e\in E([0,x]_i)} f^{(i)}(e)d_i(e) = \sum_{i\in I} p_i \sum_{e\in E([0,x]_i)} \int_{P_e} f(e')d_G(e').\end{aligned}$$

We note:

- If f is a conservative field, then $\tilde{I}(f) = I(f)$. Indeed, denote for $i \in I$ and $x \in V(G)$ the walk in G from 0 to x, obtained by concatenating the paths P_e, $e \in E([0,x]_i)$, by W_i. Then
$$\tilde{I}(f)(x) = \sum_{i\in I} p_i \sum_{e\in[0,x]_i} \int_{p_e} f d_G(e) = \sum_{i\in I} p_i \int_{W_i} f(e)d_G(e) = I(f)(x).$$
- For $e = \{x,y\} \in E$, it follows from (30) that
$$\begin{aligned}|\tilde{I}(f)(y) - \tilde{I}(f)(x)| &= \left|\sum_{i\in I} p_i \int_{[x,y]_i} f^{(i)}(e)d_i(e)\right|\\ &\le \sum_{i\in I} p_i \left|\int_{[x,y]_i} f^{(i)}(e)d_i(e)\right|\\ &\le \|f\|_\infty \sum_{i\in I} p_i \sum_{e\in[x,y]_i} d_i(e)\\ &= \|f\|_\infty \sum_{i\in I} p_i d_i(x,y) \le 8D d_G(x,y)\|f\|_\infty,\end{aligned}$$

and, thus, $\|\tilde{I}(f)\|_\infty \le 8D\|f\|_\infty$

5. Lower estimates for embeddings of $\mathcal{F}(M)$ into L_1

In this last section, we want to formulate a criterion on geodesic graphs (G,d) which implies that the distortion of embeddings of $\mathcal{F}(V(G),d)$ has to satisfy lower estimates.

This will lead to sequences of geodesic graphs (G_n, d_n), for which

$$c_{L_1}(\mathcal{F}(V(G_n), d_n) \ge C\sqrt{\log(V(G_n))}.$$

Among these sequences are, for example the sequence of discrete tori $\big((\mathbb{Z}/n\mathbb{Z})^2 : n \in \mathbb{N}\big)$ [17] and the sequence of diamond graphs $(D_n : n \in \mathbb{N})$, [2]. The idea of the proof goes back to a result of

Kislyakov from 1975 [15]. This result of Kislyakov implies, for example that $\mathcal{F}(\mathbb{R}^2)$ is not isomorphic to $\mathcal{F}(\mathbb{R})$ (which is isometrically isomorphic to $L_1(\mathbb{R})$).

Throughout this section, (G,d) is a geodesic finite graph and ν a probability on $E(G)$, whose support is all of $E(G)$. We define the probability on $V(G)$, *induced by* ν as follows

$$\mu(v)=\mu_\nu(v)=\frac{1}{2}\sum_{e\in E(G),v\in e}\nu(e)v\in V(G).$$

Note that

$$\sum_{v\in V(G)}\mu(v)=\sum_{v\in V(G)}\frac{1}{2}\sum_{e\in E(G),v\in e}\nu(e)=\frac{1}{2}\sum_{e\in E(G)}\sum_{v\in e}\nu(e)=1,$$

which shows that μ is indeed a probability on μ

5.1. **Isoperimetric dimension and the Sobolev inequality.** Let (G,d) be a geodesic finite graph and let ν be a probability on $E(G)$. We define the probability on $V(G)$, *induced by* ν as follows

$$\mu(v)=\mu_\nu(v)=\frac{1}{2}\sum_{e\in E(G),v\in e}\nu(e)v\in V(G).$$

Note that

$$\sum_{v\in V(G)}\mu(v)=\sum_{v\in V(G)}\frac{1}{2}\sum_{e\in E(G),v\in e}\nu(e)=\frac{1}{2}\sum_{e\in E(G)}\sum_{v\in e}\nu(e)=1.$$

For $A\subset V(G)$ we define the *boundary* of A to be

$$\partial_G A:=\big\{\{x,y\}\in E(G)\colon |\{x,y\}\cap A|=1\big\}$$

and the perimeter of A to be

$$\mathrm{Per}_{\nu,d}(A):=\sum_{e\in\partial_G A}\frac{\nu(e)}{d(e)}.$$

Definition 5.1 (Isoperimetric dimension)**.** For $\delta\in[1,\infty)$, and $C_{iso}\in(0,\infty)$, μ we say that (G,d) has *ν-isoperimetric dimension δ* with constant C_{iso} if for every $A\subset V(G)$

$$\min\{\mu(A),\mu(A^c)\}^{\frac{\delta-1}{\delta}}\le C_{iso}\mathrm{Per}_{\nu,d}(A), \tag{31}$$

Definition 5.2 (Sobolev Norm)**.** For $f\colon(V(G),d)\to\mathbb{R}$ and $p\in[1,\infty]$, we define the $(1,p)$-Sobolev norm (with respect to ν and d) of f by

$$\begin{aligned}\|f\|_{W^{1,p}(\nu,d)}&=\|\nabla_d f\|_{L_p(\nu)}=\mathbb{E}_\nu[|\nabla_d f|^p]^{1/p}\\&=\left[\int_{E(G)}|\nabla_d f(e)|^p d\nu(e)\right]^{1/p}=\left[\sum_{e=\{u,v\}\in E(G)}\frac{|f(u)-f(v)|^p}{d(u,v)^p}\nu(e)\right]^{1/p},\end{aligned}$$

with the usual convention when $p=\infty$.

Note that if $\nu(e) > 0$ for all $e \in E(G)$, then

$$\|f\|_{W^{1,\infty}(\nu,d)} = \max_{e=\{u,v\}\in E(G)} \frac{|f(u)-f(v)|}{d(u,v)} = \|f\|_{\mathrm{Lip}}.$$

Theorem 5.3 (Sobolev inequality from isoperimetric inequality)**.** *Assume that (G,d) has ν -isoperimetric dimension δ with constant C, then for every map $f : (V(G),d) \to \mathbb{R}$,*

$$\|f - \mathbb{E}_\mu f\|_{L_{\delta'}(\mu)} \le 2C\|f\|_{W^{1,1}(\nu,d)}, \tag{32}$$

where $\mathbb{E}_\mu f = \int_{V(G)} f(x)d\mu(x)$, and δ' is the Hölder conjugate exponent of δ, i.e. $\frac{1}{\delta} + \frac{1}{\delta'} = 1$.

Proof. Exercise. Hint: (32) follows imediately from (31) if $f = 1_A$. □

5.2. **Lipschitz- spectral profile of a graph.** Before we define what we mean by a "Lipschitz-spectral profile of a graph" we want to motivate it with an example:

Example 5.4. Consider the finite abelian group $G = (\mathbb{Z}/n\mathbb{Z})^2$, with the metric

$$d\big((v_1,v_2),(u_1,u_2)\big) = \frac{1}{n}\max\big(|v_1-u_1|,|v_2-u_2|\big).$$

The *characters* of G are the group homomorphisms $\chi : G \to T = \{e^{i2\pi x} : 0 \le x \le 1\}$. These characters can be represented as follows: $\chi : G \to T$ is a character if and only if for some $k, m \in \{0,1,2,\dots n-1\}$

$$\chi = \chi_{(k,m)} : (\mathbb{Z}/n\mathbb{Z})^2 \to T, \quad (x,y) = e^{\frac{2\pi i}{n}(xk+ym)}.$$

Note the following properties of $(\chi_{(k,m)} : 0 \le k,m \le n)$

- $(\chi_{(k,m)} : 0 \le k,m \le n)$ is an orthoromal basis in $L_2((\mathbb{Z}/n\mathbb{Z})^2,\mu)$ where μ is the uniform distribution.
- $\|\chi_{(k,m)}\|_{L_\infty(\mu)} = \|\chi_{(k,m)}\|_{L_1(\mu)} = 1$, for $0 \le k,m \le n$,
- $\|\chi_{k,m}\| \le C\max(k,m)$ and thus

$$\big|\{(k,m) \in (\mathbb{Z}/n\mathbb{Z})^2 : \|\chi_{(k,m)}\|_{\mathrm{Lip}} \le L\big\} \ge cL^2, \text{ for } L = 1,2\dots n.$$

Definition 5.5 (Lipschitz-spectral profile)**.** Let $\delta \in [1,\infty)$, and $\beta \in [1,\infty)$, and $C \ge 1$. We say that (G,d) has (μ,d)-*Lipschitz-spectral profile of dimension δ and bandwidth β with constant C* if there exists a collection of functions $F = \{f_i \colon V(G) \to \mathbb{R}\}_{i\in I}$ satisfying:

(1) $C^{-1} \le \inf_{i\in I}\|f_i\|_{L_1(\mu)} \le \sup_{i\in I}\|f_i\|_{L_\infty(\mu)} \le C$,
(2) $\{f_i\}_{i\in I}$ is an orthogonal family in $L_2(\mu)$, and
(3) for every $s \in [1,\beta]$, $|\{i \in I \colon \mathrm{Lip}(f_i) \le s\}| \ge C^{-1}s^\delta$.

5.3. **Main result.**

Theorem 5.6. *Let $\delta_{iso} \in [2,\infty)$, $\delta_{spec} \in [1,\infty)$ and $C \ge 1$. If G has (ν,d)-isoperimetric dimension δ_{iso} with constant C, and Lipschitz-spectral profile of dimension δ_{spec}, bandwidth β, with constants C,* *then any D-isomorphic embedding from the Lipschitz-free space $\mathcal{F}(V(G),d)$ into a finite-dimensional L_1-space ℓ_1^N satisfies*

$$D \geq \frac{1}{2C^5}\left(\int_1^\beta s^{\delta_{spec}-\delta_{iso}-1}ds\right)^{\frac{1}{\delta_{iso}}}. \tag{33}$$

Sketch. Assume that $T:\mathcal{F}(V(G),d)\to\ell_1^N$, is such that $\|\mu\|_{\mathcal{F}}\leq\|T(\mu)\|_1\leq D\|\mu\|_{\mathcal{F}}$ for all $\mu\in\mathcal{F}(V(G),d)$. We need to find a lower estimate for D. Passing to the adjoint $T^*:\ell_\infty^N\searrow \mathrm{Lip}_0(V(G),d)$, which is a surjection, it follows that

$$B_{\mathrm{Lip}_0(V(G),d)}\subset T^*(\ell_\infty^N) \text{ and } \|T^*\|\leq D. \tag{34}$$

We define several operators:

- $\iota_1:\mathrm{Lip}_0(V(G),d)\to W^{1,1}(V(G),d,\nu)$ identity, 1-summing operator $\pi_1(\iota_1)=1$,
  (Note that $\nabla_d:\mathrm{Lip}_0(V(G),d)\to\ell_\infty(E(G))\equiv L_\infty(E(G),\nu)$, $\nabla_d: W^{(1,1)}(V(G),d,\nu_G)\to L_1(\nu_G)$ are isometric embeddings.)
- $\iota_2:W^{1,1}(d,\nu)\to L_{\delta'_{iso}}(\mu)$ identity. By Sobolev inequality $\|\iota_2\|\leq 2C$.
- Let $F=\{f_j:j\in J\}$ be the set of orthogonal function: $f_j:V(G)\to\mathbb{R}$, as required by the spectral profile, and define the *Fourier Transform*

  $$FT: L_2(\mu)\to\ell_2(J),\quad g\mapsto(\mathbb{E}_\mu(g\cdot f_j):j\in J).$$

  By the assumed orthogonality we deduce that $\|FT\|_{L_2(\mu)\to\ell_2(J)}=1$. From the assumption on the L_1 - and L_∞-norms of $(f_j)_{j\in J}$

  $$\max_{j\in J}\big|\mathbb{E}_\mu(g\cdot f_j)\big\|\leq C\|g\|_{L_1(\mu)}, \text{ for all } g\in L_1(\mu).$$

  and thus $\|FT\|_{L_1(\mu)\to\ell_\infty(J)}\leq C$.

  We deduce, therefore from the Riesz-Thorin Interpolation Theorem (Recall that $2\leq\delta_{iso}<\infty$, and thus $1<\delta'_{iso}\leq 2$) that $\|FT\|_{L_{\delta'_{iso}}(\mu)\to\ell_{iso}(J)}\leq C$.

Then we consider the product of all these operators

$$R:\ell_\infty^N\overset{T^*}{\searrow}\mathrm{Lip}_0(d)\overset{\iota_1}{\to}W^{1,1}(d,\nu)\overset{\iota_2}{\to}L_{\delta'_{iso}}(\mu)\overset{FT}{\to}\ell_{\delta_{iso}}(J)$$

Since the summing norm π_1 is an *ideal norm*, we deduce that

$$\pi_1(R)=\pi_1(\iota_1)\|T^*\|\cdot\|\iota_2\|\cdot\|\cdot\|FT\|\leq 2DC^2.$$

An important property of 1-summing operators is that 1-summing operators between two Banach lattices are *Lattice bounded*. In our case, this means the following:

Let $R_j:\ell_\infty^N:\ell_\infty\to\mathbb{R}$ be the j-componenet of R, *i.e.*,$R(x)=\sum_{j\in J}R_j(x)e_j$, where e_j is the j-th unit vector basis of $\ell_{\delta_{iso}}(J)$. Then there exists a $b\in\ell^+_{\delta_{iso}}(J)$ with $\|b\|_{\delta_{iso}}\leq\pi_1(R)\leq 2DC^2$ so that for every $x\in\ell_\infty^N$

$$|R_j(x)\leq b_j\|x\|_\infty.$$

Now, using that $B_{\mathrm{Lip}_0(V(G),d)} \subset T^*(B_{\ell_\infty^N})$ choose for every $j \in J$ and $x_j \in B_{\ell_\infty^N}$ so that

$$T^*(x_j) = \frac{f_j}{\|f_j\|_{\mathrm{Lip}}}, \text{ for all } j \in J.$$

It follows that

$$R(x_j) = \frac{R(f_j)}{\|f_j\|_{\mathrm{Lip}}} = \frac{1}{\|f_j\|_{\mathrm{Lip}}} e_j \text{ and } |R_j(x_j)| \le b_j \text{ for all } j \in J.$$

Therefore we obtain

$$\begin{aligned} 2C^2 D \ge \|b\|_{\ell_{\delta_{iso}}} &= \Big(\sum_{j\in J} b_j^{\delta_{\mathrm{iso}}}\Big)^{1/\delta_{\mathrm{iso}}} \ge \Big(\sum_{j\in J} |R_j(x_j)|^{\delta_{\mathrm{iso}}}\Big)^{1/\delta_{\mathrm{iso}}} \\ &= \Big(\sum_{j\in J} \big(\|\mathcal{F}(f_j)\|_{L_2}^2/\|f_j\|_{\mathrm{Lip}}\big)^{\delta_{\mathrm{iso}}}\Big)^{1/\delta_{\mathrm{iso}}} \ge C^{-2}\Big(\sum_{j\in J}\Big(\frac{1}{\|f_j\|_{\mathrm{Lip}}}\Big)^{\delta_{\mathrm{iso}}}\Big)^{1/\delta_{\mathrm{iso}}} \end{aligned}$$

Thus

$$D \ge \frac{1}{2C^4}\Big(\sum_{j\in J}\Big(\frac{1}{\|f_j\|_{\mathrm{Lip}}}\Big)^{\delta_{\mathrm{iso}}}\Big)^{1/\delta_{\mathrm{iso}}}. \tag{35}$$

From here, we calculate the sum by applying the classical formula

$$\int_\Omega |h|^p d\sigma = p\int_0^\infty t^{p-1}\sigma(\{|h| > t\})dt$$

with $\Omega = J$ and σ the counting measure:

$$\begin{aligned} \sum_{j\in J} \frac{1}{\mathrm{Lip}(f_j)^{\delta_{iso}}} &= \delta_{iso}\int_0^\infty t^{\delta_{iso}-1}\Big|\Big\{j \in J\colon \frac{1}{\mathrm{Lip}(f_j)} > t\Big\}\Big|dt \\ &= \delta_{iso}\int_0^\infty \frac{1}{s^{\delta_{iso}-1}}\Big|\Big\{j \in J\colon \frac{1}{\mathrm{Lip}(f_j)} > \frac{1}{s}\Big\}\Big|\frac{1}{s^2}ds \\ &= \delta_{iso}\int_0^\infty \frac{1}{s^{\delta_{iso}+1}}\Big|\Big\{j \in J\colon \mathrm{Lip}(f_j) < s\Big\}\Big|ds \\ &\ge \delta_{iso}\int_1^\beta \frac{1}{s^{\delta_{iso}+1}}\frac{s^{\delta_{spec}}}{C}ds. \end{aligned} \tag{36}$$

(where we used (3) of the Lipschitz spectral profile in the last inequality) from which, together with (33), we deduce our claim (35). □

5.4. **Examples satisfying Theorem 5.6.**

6. Appendix

6.1. Proof of Birkhoff's Theorem 2.15.

Theorem 6.1. *(Birkhoff) Assume $n \in \mathbb{N}$ and that $A = (a_{i,j})_{i,j=1}^n$ is a doubly stochastic matrix, i.e.,*

$$0 \le a_{i,j} \le 1 \text{ for all } 1 \le i, j \le n,$$
$$\sum_{j=1}^n a_{i,j} = 1 \text{ for } i = 1, 2, \ldots, n, \text{ and } \sum_{i=1}^n a_{i,j} = 1 \text{ for } j = 1, 2, \ldots, n.$$

Then A is a convex combination of permutation matrices, i.e.,matrices which have in each row and each column exactly one entry whose value is 1 and vanishes elsewhere.

Proof. Clearly the set DS_n of all doubly stochastic $n$by$n$ matrices is bounded, convex and compact in $\mathbb{R}^{n\times n}$, and is thus the convex hull of its extreme points. It is, therefore, enough to show that a matrix $A \in \mathrm{DS}$, which has entries that are not integers, cannot be an extreme point of DS.

So assume that for some r_1 and s_1 in $\{1, 2, \ldots, n\}$ we have $0 < a_{r_1,s_1} < 1$ and since the row r_1 adds up to 1 there must be an $s_2 \neq s_1$ in $\{1, 2, \ldots, n\}$, with $0 < a_{r_1,s_2} < 1$. Again, since the column s_2 adds to up to 1 there must be an $r_2 \neq r_1$ in $\{1, 2, \ldots, n\}$ with $0 < a_{r_2,s_2} < 1$. We continue this way and eventually for some k, either $\{r_k, s_k\}$ or $\{r_k, s_{k+1}\}$ must be among the previously chosen pairs (r_j, s_j) or (r_j, s_{j+1}).

Possibly by changing the starting point, relabeling, and exchanging rows with columns, we obtain a cycle which is either of the form

$$(r_1, s_1), (r_1, s_2), (r_2, s_2), \ldots, (r_{k-1}, s_k), (r_k, s_k) = (r_1, s_1)$$

(implying the cycle is of even length) or of the form

$$(r_1, s_1), (r_1, s_2), (r_2, s_2), \ldots, (r_k, s_k), (r_k, s_{k+1}) = (r_1, s_1),$$

(implying the cycle is of odd length)
so that $r_j \neq r_{j+1}$ and $s_j \neq s_{j+1}$ (if the cycle is of the first form, we put $s_i = s_{i \bmod (k-1)}$, if $I > k-1$ and if it is of the second form we put $s_i = s_{i \bmod (k)}$, if $i > k$) and $0 < a_{r_j,s_j}, a_{r_j,s_{j+1}} < 1$. Let us assume we have chosen a cycle of minimal length. Then we claim it must be of the first form, *i.e.*, it must be of even length.

Indeed, assume it is of the second form, then $(r_k, s_{k+1}) = (r_1, s_1)$ and (r_1, s_2) are in the same row and therefore

$$(r_2, s_2), (r_2, s_3), (r_3, s_3), \ldots, (\underbrace{r_k}_{=r_1}, s_k), (r_1, s_2),$$

is a shorter cycle, which is a contradiction; thus, our shortest cycle is of the first form. Let now $0 < \varepsilon < 1$ small enough so that for all $j \le k$

$$\varepsilon < \min(a_{r_j,s_j}, a_{r_j,s_{j+1}}, 1 - a_{r_j,s_j}, 1 - a_{r_j,s_{j+1}})$$

Then define

$$b^{(1)}_{s,t} = \begin{cases} a_{r_j,s_j} + \varepsilon & \text{if } (s,t) = (r_j, s_j) \text{ for some } j \\ a_{r_j,s_{j+1}} - \varepsilon & \text{if } (s,t) = (r_j, s_{j+1}) \text{ for some } j \\ a_{s,t} & \text{otherwise,} \end{cases}$$

and

$$b^{(2)}_{s,t} = \begin{cases} a_{r_j,s_j} - \varepsilon & \text{if } (s,t) = (r_j, s_j) \text{ for some } j \\ a_{r_j,s_{j+1}} + \varepsilon & \text{if } (s,t) = (r_j, s_{j+1}) \text{ for some } j \\ a_{s,t} & \text{otherwise.} \end{cases}$$

It follows that $B^{(1)} = (b^{(1)})^n_{s,t=1}$ and $B^{(2)} = (b^{(2)})^n_{s,t=1}$ are in DS and

$$A = \frac{1}{2}\big(B^{(1)} + B^{(2)}\big),$$

which implies that A cannot be an extreme point. □

Proof of Proposition 2.14. Let $A = \{x_1, x_2, \dots x_n\}$ and $B = \{y_1, y_2, \dots, y_n\}$. We note that for every $\pi \in \mathcal{P}(\sigma, \tau)$ the matrix
$M = (n\pi(x_i, y_j) : 1 \le i, j \le n)$ is a doubly stochastic matrix (since $\sum_{x\in A} \pi(x,y) = \tau(y) = \frac{1}{|B|} = \frac{1}{|A|} = \sigma(x) = \sum_{y\in B} \pi(x,y)$). Thus by (10)

$$d_{\mathrm{Wa}}(\mu_A, \mu_B) = \frac{1}{n} \min\Big\{ \sum_{i,j=1}^n M_{i,j} d(x_i, y_j) : M \in \mathrm{DS}_n \Big\}$$

Since the map

$$\mathrm{DS} \to [0, \infty), \quad M \mapsto \sum_{i,j=1}^n M_{i,j} d(x_i, y_j)$$

is linear, it achieves its minimum on an extreme point, our claim follows from Theorem 2.15 □

6.2. **Proof of Theorem 4.2 on stochastic embeddings of finite metric spaces into trees.**

Theorem 6.2. [6] *Let M be a metric space with $n \in \mathbb{N}$ elements. Then, there is a $O(\log n)$ stochastic embedding of M into the class of weighted trees.*

We need some notation:
We fix a metric space (M, d).

- $B_r(x) = \{y \in M : d(x,y) \le r\}$, for $x \in M$ and $r > 0$.
- For $A \subset M$ the *diameter of* A is $\mathrm{diam}(A) = \sup_{x,y\in A} d(x,y)$.
- The set of partitions of M is denoted by $\mathcal{P}_M$. The elements of a partition P are called *clusters of* P. Let $P = (A_1, A_2, \dots A_n) \in \mathcal{P}_M$. For $r \ge 0$, P is called *r-bounded* if $\mathrm{diam}(A_j) = \max_{x,z\in A_j} d(x,z) < r$, for all $j = 1, 2, \dots, n$. For $x \in M$ and a partition $P = (A_1, A_2, \dots A_n)$ of M we denote the unique $A_j \in P$ which contains x by P_x.
- A *stochastic decomposition* is a probability measure $\mathbb{P}$ on $\mathcal{P}_M$, and its *support* is given by $\mathrm{supp}(\mathbb{P}) = \{P \in \mathcal{P} : \mathbb{P}(P) > 0\}$.

Lemma 6.3. *Let $R > 0$. There is a Probability measure $\mathbb{P}$ on $\mathcal{P}_M$ so that*

$$\text{supp}(\mathbb{P}) \subset \big\{P \in \mathcal{P}_M : P \textit{ is } R\textit{-bounded}\big\} \textit{ and} \tag{37}$$

$$\mathbb{P}\big(P\in\mathcal{P}_M : B_t(x)\subset P_x\big) \geq \Big(\frac{|B_{R/8}(x)|}{|B_R(x)|}\Big)^{8t/R} \textit{ for all } 0 < t \leq R/8 \textit{ and } x\in M. \tag{38}$$

Proof. Let $M = \{x_1, x_2, \ldots, x_n\}$, let π be a permutation on $\{1, 2, \ldots, n\}$ and $r \in \big[\frac{R}{4}, \frac{R}{2}\big)$. We define a partition $\tilde{P}(\pi, r) = \big(\tilde{C}_i(\pi, r)\big)_{i=1}^n$ as follows:

$\tilde{C}_1(\pi, r) = B_r(x_{\pi(1)})$ and assuming $\tilde{C}_1(\pi, r)$, $\tilde{C}_2(\pi, r), \ldots, \tilde{C}_{j-1}(\pi, r)$, let $\tilde{C}_j(\pi, r) = B_r(x_{\pi(j)}) \setminus \bigcup_{i=1}^{j-1} B_r(x_{\pi(i)})$. Some of the $\tilde{C}_j(\pi, r)$, $j = 1, 2 \ldots n$ could be empty and we let $P(\pi, r) = \big(C_j(\pi, r)\big)_{j=1}^m$, with $1 \leq m \leq n$ be the non empty members of $\big(\tilde{C}_j(\pi, r)\big)_{j=1}^n$, in the order inherited from $\tilde{P}(\pi, r)$.

Let μ be the uniform distribution on Π_n the set of all permutation on $\{1, 2, \ldots, n\}$ (and thus $\mu(\pi) = \frac{1}{n!}$, for $\pi \in \Pi_n$) and let ν be the uniform distribution on $\big[\frac{R}{4}, \frac{R}{2}\big)$. Finally let $\mathbb{P}$ be the image distribution of $\mu \otimes \nu$ under the map $(\pi, r) \mapsto \tilde{P}(\pi, r)$, and thus

$$\mathbb{P}(A) = \mu \otimes \nu\big(\{(\pi, r) : \tilde{P}(\pi, r) \in A\}\big) \text{ for } A \subset \mathcal{P}_M.$$

It follows from Fubini's Theorem that

$$\begin{aligned} \mathbb{P}(A) &= \int_{R/4}^{R/2} \int_{\Pi_n} 1_{\{(\pi', r') : \tilde{P}(\pi', r') \in A\}}(\pi, r)\, d\mu(\pi)\, d\nu(r) \\ &= \int_{R/4}^{R/2} \mu(\{\pi\in\Pi_n : \tilde{P}(\pi, r)\in A\})\, d\nu(r) \\ &= \frac{4}{R} \int_{R/4}^{R/2} \mu(\{\pi\in\Pi_n : \tilde{P}(\pi, r)\in A\})\, d(r). \end{aligned} \tag{39}$$

Claim. For $\frac{R}{4} \leq r < \frac{R}{2}$, $0 < t \leq r$ and $x \in M$ it follows that

$$\mu\big(\{\pi \in \Pi_n : B_t(x) \subset \tilde{P}_x(\pi, r)\big) \geq \frac{|B_{r-t}(x)|}{|B_{r+t}(x)|}. \tag{40}$$

In order to prove the claim we order $B_{r+t}(x)$ into $(y_i)_{i=1}^a$ and $(y_i)_{i=a+1}^b$ so that

$$0 = d(x, y_1) \leq d(x, y_2) \leq \ldots \leq d(x, y_a) \leq r - t < d(x, y_{a+1}) \leq \ldots \leq d(x, y_b) \leq r + t.$$

Thus $|B_{r-t}(x)| = a$, $|B_{r+t}(x)| = b$, and $y_i \in B_{r-t}(x)$ if $1 \leq i \leq a$, and $y_i \in B_{r+t}(x) \setminus B_{r-t}(x)$ if $a < i \leq b$. Define

$$E_i := \left\{\pi \in \Pi_n : \begin{matrix} y_{\pi(s)} \notin \{y_1, y_2, \ldots y_b\} \text{ for } s = 1, 2 \ldots, i-1 \\ y_{\pi(i)} \in \{y_1, y_2, \ldots y_a\} \end{matrix}\right\}.$$

In other words E_i is the event that i is the smallest $j \leq n$ for which $y_{\pi(j)} \in B_{r+t}(x)$ intersected with the event that $y_{\pi(i)} \in B_{r-t}(x)$.

Since for $\pi \in E_i$ and $s < i$ we have $d(y_{\pi(s)}, x) > r + t$ and $d(y_{\pi(i)}, x) < r - t$ it follows that $B_t(x) \subset \tilde{P}_x(\pi, r)$. For $i = 1, 2, \dots n$ let

$$A_i = \left\{ \pi \in \Pi_n : \begin{matrix} y_{\pi(s)} \notin \{y_1, y_2, \dots y_b\} \text{ for } s = 1, 2 \dots, i-1 \\ y_{\pi(i)} \in \{y_1, y_2, \dots y_b\} \end{matrix} \right\}.$$

Then the sets $(A_i)_{i=1}^n$ are a partition of Π_n, and $E_i \subset A_i$, and moreover $\mu(E_i|A_i) = \frac{a}{b}$ (since $\mu(E_i|A_i)$ is the probability that $\pi(i) \le a$ assuming that $\pi(i) \le b$). Thus

$$\mu(\{\pi \in \Pi_n : B_t(x) \subset \tilde{P}_x(\pi, r)\}) \ge \mu\Big(\bigcup_{i=1}^n E_i\Big) = \sum_{i=1}^n \mu(A_i)\mu(E_i|A_i) = \frac{a}{b} = \frac{|B_{r-t}(x)|}{|B_{r+t}(x)|},$$

which proves our claim.

To finish the proof of the Lemma we deduce from (39) and (40) for $0 < t \le R/8$ that

$$\begin{aligned}
\mathbb{P}(\{\pi \in \Pi_n : B_t(x) \subset P_x(\pi, r)\}) &= \frac{4}{R} \int_{R/4}^{R/2} \mu\big(\{\pi \in \Pi_n : B_t(x) \subset P_x(\pi, r)\}\big)\, dr \\
&\ge \frac{4}{R} \int_{R/4}^{R/2} \frac{|B_{r-t}(x)|}{|B_{r+t}(x)|}\, dr \\
&= \frac{4}{R} \int_{R/4}^{R/2} e^{h(r-t) - h(r+t)}\, dr \\
&\quad \text{where } h(s) = \log\big(|B_s(x)|\big) \\
&\ge \exp\Big(\frac{4}{R} \int_{R/4}^{R/2} h(r-t) - h(r+t)\, dr\Big) \\
&\quad \text{(By Jensen's inequality)} \\
&= \exp\Big(\frac{4}{R} \int_{\frac{R}{4}-t}^{\frac{R}{2}-t} h(s)\, ds - \frac{4}{R} \int_{\frac{R}{4}+t}^{\frac{R}{2}+t} h(s)\, ds\Big) \\
&= \exp\Big(\frac{4}{R} \int_{\frac{R}{4}-t}^{\frac{R}{4}+t} h(s)\, ds - \frac{4}{R} \int_{\frac{R}{2}-t}^{\frac{R}{2}+t} h(s)\, ds\Big) \\
&\ge \exp\Big(\frac{8t}{R}\Big(h\Big(\frac{R}{4} - t\Big) - h\Big(\frac{R}{2} + t\Big)\Big)\Big) \\
&\ge \exp\Big(\frac{8t}{R}\Big(h\Big(\frac{R}{8}\Big) - h(R)\Big)\Big) = \left(\frac{|B_{R/8}(x)|}{|B_R(x)|}\right)^{\frac{8t}{R}}.
\end{aligned}$$

□

Corollary 6.4. *For the probability measure* $\mathbb{P}$ *defined in Lemma* 6.3 *it follows that*

$$\mathbb{P}\big(\{P \in \mathcal{P}_M : B_t(x) \not\subset P_x\}\big) \le \frac{8}{R} t \log\Big(\frac{|B_R(x)|}{|B_{R/8}(x)|}\Big), \text{ for all } t \in (0, R/8).$$

Proof. We deduce from Lemma 6.3 that

$$\begin{aligned}\mathbb{P}\big(\{P\in\mathcal{P}_M : B_t(x)\not\subset P_x\}\big) &= 1-\mathbb{P}\big(\{P\in\mathcal{P}_M : B_t(x)\subset P_x\}\big)\\ &\le 1-\Big(\frac{|B_{R/8}(x)|}{|B_R(x)|}\Big)^{\frac{8t}{R}}\\ &\le \log\left(\Big(\frac{|B_R(x)|}{|B_{R/8}(x)|}\Big)^{\frac{8t}{R}}\right) = \frac{8}{R}t\log\Big(\frac{|B_R(x)|}{|B_{R/8}(x)|}\Big),\end{aligned}$$

where the last inequality follows from the fact that for $z\ge 1$

$$1-\frac{1}{z}\le\log(z).$$

□

Proof of Theorem 6.2. After rescaling we can assume that $d(x,y)\ge 1$ for all $x\ne y$, in M, and choose $k\in\mathbb{N}$ so that $\mathrm{diam}(M)\in[2^{k-1},2^k)$. We will first introduce some notation:

For two partitions P and Q of the same set S, we say P *subdivides* Q and write $P\succeq Q$ if every cluster $A\in Q$ is a union of clusters in P. If $A\subset S$ then the *restriction of* P *onto* A, is the

$$P|_A=\{B\cap A : B\in P\}\setminus\{\emptyset\}.$$

For a metric space M and any $R>0$, we denote the probability measure on the set $\mathcal{P}$ on R-bounded Partitions constructed in Lemma 6.3 by $\mathbb{P}^{(M,R)}$.

Our probability will be defined on the set

$$\Omega=\big\{(P^{(j)})_{j=0}^k\subset\mathcal{P}_M : P^{(j)}\text{ is }2^{k-j}\text{ bounded and }P^{(j)}\succeq P^{(j-1)},\text{ for }j=1,2,\dots,k\big\}.$$

We let $\mathbb{P}$ be the probability measure on the subsets of Ω, uniquely defined by the following properties:

$$\mathbb{P}\big(\{\bar P=(P^{(j)})_{j=0}^k\in\Omega : P^{(0)}=M\big\}\big)=1 \tag{41}$$

for $i=1,2,\dots n$, $B\subset M$, $\mathrm{diam}(B)\le 2^{k-(i-1)}$ and $\mathcal{A}\subset\mathcal{P}_B$ we have

$$\mathbb{P}\big(P^{(i)}|_B\in\mathcal{A}\,\big|B\text{ is cluster of }P^{(i-1)}\big)=\mathbb{P}^{(B,2^{k-i})}(\mathcal{A}). \tag{42}$$

Condition (42) means the following: under the condition that $P^{(i-1)}$ is given and contains a cluster $B\subset M$, the distribution of $P^{(i)}$ restricted to B is $\mathbb{P}^{(B,2^{k-i})}$. So we can think of $(P^{(j)})_{j=0}^k$ as a stochastic process with values in $\mathcal{P}_M$, whose distribution is determined by transition probabilities: $P^{(0)}=M$ (the trivial partition and given $P^{(i-1)}$, we consider each cluster B of $P^{(i-1)}$, and randomly divide B according to the distribution $\mathbb{P}^{(B,2^{k-i})}$. Since $d(x,y)\ge 1$ for $x\ne y$, and since $\mathrm{diam}(M)<2^k$, it follows that $P^{(k)}$ is the *finest partition, i.e.,* $P^{(k)}=\big\{\{x\}:x\in M\big\}$.

By induction on $k\in\mathbb{N}$ it can easily be seen that the probability $\mathbb{P}$ exists and is uniquely defined by the above property.

For each $\bar{P} = (P^{(j)})_{j=0}^k \in \Omega$ we define the following weighted tree $T = T(\bar{P}) = (V(\bar{P}), E(\bar{P}), W(\bar{P}))$, where

$$\begin{aligned} V &= \bigcup_{j=0}^k V_j \text{ with } V_j = \{(j,B) : B \text{ is cluster of } P^{(j)}\}, \text{ for } j = 0,1,\ldots,n \\ E &= \bigcup_{j=1}^k E_j \text{ with } E_j = \big\{\{(j,A),(j-1,B)\} : A \in V_j, B \in V_{j-1}, \text{ and } A \subset B\big\} \\ &\qquad\qquad \text{for } j = 1,\ldots,n, \\ W &: E \to \mathbb{R}, \quad e \mapsto 2^{k-j} \text{ if } e \in E_j. \end{aligned}$$

One might ask why we did not simply define the vertices of T as the set of all the clusters of the $P^{(j)}$. The problem is that the partitions $P^{(j)}$ and $P^{(j-1)}$ could share the same clusters, and we need to distinguish them.

We note that T is a tree, and that then $(k,\{x\})$, $x \in M$ are the leaves of T.

For $\bar{P}$ we define

$$f_{\bar{P}} : M \to T_{\mathbb{P}}, \quad x \mapsto (k,\{x\}).$$

We claim that for some universal constant D, $(f_{\bar{P}})_{\bar{P}\in\Omega}$ with the coefficients $(\mathbb{P}(\bar{B}))_{\bar{B}\in\Omega}$ is a $D\log(n)$ stochastic embedding.

Let us first show the lower bound. Assume that $\bar{P} = (P^{(j)})_{j=0}^k \in \Omega$ and $x \neq y$ are in M. Then there is an $i \in \{1,2,\ldots,k\}$ so that x,y are in the same cluster of $P^{(i-1)}$ but in two different clusters of $P^{(i)}$. This implies that $d(x,y) \le 2^{k-(i-1)} = 2^{k+1-i}$ and and for the tree metric of $T_{\bar{P}}$, it follows that

$$d_{T_{\bar{P}}}(f_{\bar{P}}(x), f_{\bar{P}}(y)) = 2\sum_{j=i}^k 2^{k-j} = 2\sum_{j=0}^{k-i} 2^j = 2(2^{k+1-i}-1) \ge d(x,y).$$

To get the upper estimate, first note that from Corollary 6.4 it follows for $i = 1,2,\ldots k$, and $x,y \in M$, if $d(x,y) < 2^{k-i-3}$

$$\begin{aligned} (43) \qquad \mathbb{P}(P_x^{(i-1)} = P_y^{(i-1)} &\text{ and } P_x^{(i)} \neq P_y^{(i)}) \\ &= \mathbb{P}(P_x^{(i)} \neq P_y^{(i)} \big| P_x^{(i-1)} = P_y^{(i-1)})\mathbb{P}(P_x^{(i-1)} = P_y^{(i-1)}) \\ &\le \mathbb{P}(P_x^{(i)} \neq P_y^{(i)} \big| P_x^{(i-1)} = P_y^{(i-1)}) \\ &= \mathbb{P}^{(P_x^{i-1}, 2^{k-i})}(P \in \mathcal{P}_M : y \notin P_x) \\ &\text{(Apply (42) to } B = P_x^{i-1} = P_y^{i-1}) \\ &\le \mathbb{P}^{(P_x^{i-1}, 2^{k-i})}(P \in \mathcal{P}_M : B_{d(x,y)} \not\subset P_x) \\ &\le \frac{8}{2^{k-i}} \log\Big(\frac{|B_{2^{k-i}}(x)|}{|B_{2^{k-i-3}}(x)|}\Big) d(x,y). \end{aligned}$$

It is also true that if $d(x,y) > 2^{k-i+1}$ then

$$\mathbb{P}(P_x^{(i-1)} = P_y^{(i-1)} \text{ and } P_x^{(i)} \neq P_y^{(i)}) \leq \mathbb{P}(P_x^{(i-1)} = P_y^{(i-1)}) = 0.$$

Secondly, note that for $\bar{P} \in \{(P^{(j)})_{j=0}^k \in \Omega : \{P_x^{(i-1)} = P_y^{(i-1)} \text{ and } P_x^{(i)} \neq P_y^{(i)}\}$ it follows that

$$d_{T(\bar{P})}(f(x), f(y)) = 2\sum_{j=0}^{k-i} 2^j = 2(2^{k-i+1} - 1) < 2^{k+2-i}.$$

Let $s \leq k$ so that $2^{s-1} < d(x,y) \leq 2^s$. We compute

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}(d_{T(\bar{P})}(f(x), f(y))) &\leq \sum_{i=1}^{k} 2^{k+2-i}\mathbb{P}(P_x^{(i-1)} = P_y^{(i-1)} \text{ and } P_x^{(i)} \neq P_y^{(i)}) \\
&= \underbrace{\sum_{i=1}^{k-s-3} 2^{k+2-i}\mathbb{P}(P_x^{(i-1)} = P_y^{(i-1)} \text{ and } P_x^{(i)} \neq P_y^{(i)})}_{=\Sigma_1} \\
&\quad + \underbrace{\sum_{i=k-s-2}^{k-s} 2^{k+2-i}\mathbb{P}(P_x^{(i-1)} = P_y^{(i-1)} \text{ and } P_x^{(i)} \neq P_y^{(i)})}_{=\Sigma_2} \\
&\quad + \underbrace{\sum_{i=k-s+1}^{k} 2^{k+2-i}\mathbb{P}(P_x^{(i-1)} = P_y^{(i-1)} \text{ and } P_x^{(i)} \neq P_y^{(i)})}_{=\Sigma_3}.
\end{aligned}$$

Since $d(x,y) > 2^{s-1}$ it follows that $\Sigma_3 = 0$. Secondly

$$\Sigma_2 \leq 3 \cdot 2^{k+2-(k-s-2)} = 3 \cdot 2^{s+4} \leq 3 \cdot 2^5 d(x,y)$$

To estimate Σ_1 we are able to use (43) since $d(x,y) \leq 2^s = \frac{1}{8}2^{k-(k-s-3)}$

$$\begin{aligned}
\Sigma_1 &\leq \sum_{i=1}^{k} 2^{k+2-i}\frac{8}{2^{k-i}} \log\Big(\frac{|B_{2^{k-i}}(x)|}{|B_{2^{k-i-3}}(x)|}\Big) d(x,y) \\
&= 32 \log\Big(\prod_{i=1}^{k} \frac{|B_{2^{k-i}}(x)|}{|B_{2^{k-i-3}}(x)|}\Big) d(x,y) \\
&\leq 32\log(n^3) d(x,y) = 96 \log(n) d(x,y)
\end{aligned}$$

And thus

$$\mathbb{E}_{\mathbb{P}}(d_{T(\bar{P})}(f(x), f(y))) \leq 96\log(n) d(x,y) + 3 \cdot 2^5 d(x,y).$$

□

6.3. **Proof of Theorem 4.3.** Our goal is to show that for some universal constant c, every metric space with n elements can be bijectively $c\log(n)$-stochastically embedded into trees. This will follow from Theorem 4.2 and the following result by Gupta.

Theorem 6.5. [10, Theorem 1.1] *Let $T=(V,E,W)$ be a weighted finite tree and $V'\subset V$. Then there is $E\subset[V']^2$ and $W':E(\mathbb{G}')\to[0,\infty)$ so that $T'=(V',E(\mathbb{G}'),W')$ is a tree*

$$\frac{1}{4}\le\frac{d_{T'}(x,y)}{d_T(x,y)}\le 2,\ \textit{for } x,y\in V'. \tag{44}$$

We first show the claim of Theorem 6.5 in the special case that V' consists of leaves of T. Recall that in a tree $T=(V,E)$

$$\text{Leaf }(T)=\{v\in V:\deg_T(v)=1\}.$$

Lemma 6.6. *Let $T=(V,E,W)$ be a weighted finite tree and $V'\subset\text{Leaf }(T)$, and let d_T be the geodesic metric generated by $W:E\to(0,\infty)$. Then there is $E(\mathbb{G}')\subset[V']^2$ and $W':E'\to(0,\infty)$ so that $T'=(V',E',W')$ is a tree and*

$$\frac{1}{4}\le\frac{d_{T'}(x,y)}{d_T(x,y)}\le 2,\ \textit{for } x,y\in V'.$$

Proof. We start with a general weighted tree $T=(V,E,W)$ and a subset V' of V. By eliminating successively every leaf of T which is not in V', we can assume that V' are all the leaves. Unless $V'=V$ (in which case we are done), there must be an element $x_0\in V\setminus V'$, and the degree of this element must be at least 2. Denote the partial order defined by letting x_0 be the root of T by $\succeq$. For $x,y\in V$ we denote by $x\wedge y$ the *minimum of x and y* meaning the maximal vertex z with respect to $\succeq$ for which $x\succeq z$ and $y\succeq z$.

Let $v_0\in V'$ for which $r_0:=d_T(x_0,v_0)$ is minimal. Let $\tilde{E}\subset E$ be the set of edges $e=\{a,b\}$ in E for which $d_T(x_0,a)<r_0/2\le d_T(x_0,b)$. Order $\tilde{E}$ into $\{e_1,e_2,\dots,e_n\}$, $e_i=\{a_i,b_i\}$ with $d_T(x_0,a_i)<r_0/2\le d_T(x_0,b_i)$. One of the edges in $\tilde{E}$ must be contained in $[v_0,x_0]$ and assume that $\tilde{E}$ was ordered so that $e_1\subset[v_0,x_0]$.

We now define new trees $T_1,T_2,\dots,T_n$, which are, up to possibly one additional element subtrees of T.

If $r_0/2=d_T(x_0,b_j)$ we let T_j be the subtree $T_j=(V_j,E_j,W_j)$ with

$$V_j=\{x\in V:x\succeq b_j\},\quad E_j=E\cap[V_j]^2\text{ and }W_j=W|_{E_j}.$$

In that case we put $x_j=b_j$.

If $r_0/2<d_T(x_0,b_j)$ then $T_j=(V_j,E_j,W_j)$, with $V_j=\{x\in V:x\succeq b_j\}\cup\{x_j\}$ where x_j is an element not in V, and distinct from all the other x_i and we let

$$E_j=E\cap[V_j]^2\cup\big\{\{x_j,b_j\}\big\}\text{ and }W_j(e)=\begin{cases}W(e)&\text{if }e\in E\cap[V_j]^2,\\ d(x_0,b_j)-r_0/2&\text{if }e=\{b_j,x_j\}.\end{cases}$$

We also define the following tree $\bar{T}=(\bar{V},\bar{E},\bar{W})$ which contains T, and T_1, $T_2,\dots T_n$ isometrically:

$$\bar{V}=V\cup\{x_j:j=1,2\dots,n\}=V\dot\cup\{x_j:j=1,2\dots,n,d_T(x_0,b_j)>r_0/2\}$$

$$\bar{E} = (E \setminus \tilde{E}) \cup \big\{\{a_j, x_j\} : j = 1, 2 \ldots n\big\} \cup \{\{b_j, x_j\} : j = 1, 2 \ldots n, b_j \neq x_j\}$$

$$\bar{W} : \bar{E} \to (0, \infty, \quad e \mapsto \begin{cases} W(e) & \text{if } e \in E \setminus E', \\ r_0/2 - d_T(a_j, x_0) & \text{if } e = \{a_j, x_j\}, \\ d_T(b_j, x_0) - r_0/2 & \text{if } e = \{b_j, x_j\} \text{ and } b_j \neq x_j. \end{cases}$$

We note that the inclusion $(V, d_T) \subset (\bar{V}, d_{\bar{T}})$, $(V_i, d_{T_i}) \subset (\bar{V}, d_{\bar{T}})$ are isometric embeddings.

Let us make some observations:

(1) Since for $j = 1, \ldots n$ the vertex a_j lies on the path $[x_0, b_j]$ connecting x_0 and b_j, it follows if $b_i = b_j$ then also $a_i = a_j$ and thus $i = j$. Thus, all the b_j, $j = 1, \ldots n$, are pairwise distinct, and the V_j are pairwise disjoint.

(2) We claim that n is at least 2. Indeed, let z_1 and z_2 be two direct successors of x_0 (the degree of x_0 is at least 2) and let w_1 be a leaf in $\{x \in V : x \succeq z_1\}$ and w_2 be a leaf in $\{x \in V : x \succeq z_2\}$, and let for $s = 1, 2$ $e_{i_s} = \{a_{i_s}, b_{i_s}\}$ be the edge in $[x_0, w_s]$ for which $d_T(x_0, a_{i_s}) < r_0/2 \leq d_T(x_0, b_{i_s})$ (such edges exist because of the minimality of r_0). Since the path from w_1 to w_2 must go through x_0 it follows that $b_{i_1} \neq b_{i_2}$ (it could be possible that $a_{i_1} = a_{i_2} = x_0$!)

(3) Put $V'_j = V' \cap V_j$ for $j = 1, \ldots n$. Since for all $w \in V'$ $d(x_0, w) \geq r_0$, it follows that $V' = \bigcup_{j=1}^n V'_j$, and that the V'_j are the leafs of T_j.

(4) For $i \neq j$ and $v \in V'_i$ and $w \in V'_j$, we observe that a_i and a_j have to lie on the path connecting v with w and thus

$$d_T(v, w) \geq d_T(v, a_i) + d_T(w, a_j)$$

For $j = 1, 2, \ldots n$ put $T_j = (V_j, E_j, W_j)$ with $E_j = \bar{E} \cap [V_j]^2$ and $W_j = \bar{W}|_{E_j}$, and V'_j is of strictly less cardinality that V'. Moreover for $x, y \in V_j$ we have $d_{\bar{T}}(x, y) = d_{T_j}(x, y)$.

For $j = 1, \ldots n$, put $r_j = \min_{v \in V'_j} d_{T_j}(x_j, v)$, and choose $v_j \in V_j$ so that $d_{T_j}(v_j, x_j) = r_j$. Note that $r_1 = r_0/2 = d_{T_j}(v_0, x_1)$ and thus we can assume that $v_1 = v_0$.

Then we can continue to decompose $T_1, T_2, \ldots T_n$ until we arrive at trees that consist of one single element of V' by applying inductively the following claim.

We claim the following:

Claim: Assume that for $j = 1, 2 \ldots n$, T_j satisfies the following condition: There is a tree $T'_j = (V'_j, E'_j, W'_j)$ on V'_j so that

$$d_{T'_j}(x, v_j) \leq 2 d_{T_j}(x, x_j) - r_j, \text{ for } x \in V'_j \tag{45}$$

$$\frac{1}{4} \leq \frac{d_{T'_j}(x, y)}{d_{T_j}(x, y)} \leq 2, \text{ for } x, y \in V'_j \tag{46}$$

Then construct $T' = (V', E', W')$ by *glueing* $T'_1, T'_2, \ldots T'_n$ together, connecting $v_1 = v_0$ to the other v_j. More precisely, we put

$$E' = \bigcup_{j=1}^n E'_j \cup \big\{\{v_1, v_j\} : 2 \leq j \leq n\big\}$$

and use the weight

$$W'(e) = W'(e) \text{ if } e \in \bigcup_{j=1}^{n} E'_j \text{ and } W(\{v_1, v_j\}) = d(x_0, v_j)) = r_j.$$

We claim that it follows that

$$d_{T'}(x, v_0) \le 2d_T(x, x_0) - r_0, \text{ for } x \in V', \tag{47}$$

$$\frac{1}{4} \le \frac{d_{T'}(x, y)}{d_T(x, y)} \le 2, \text{ for } x, y \in V'. \tag{48}$$

Since (47) and (48) are clearly satisfied if V' is a singleton, the Theorem follows by induction from the claim.

We first verify the first inequality in (48): For $x, y \in V'$, either x, y lie both in V'_j for some j, we deduce our claim from the induction hypothesis or $x \in V_i$ and $y \in V_j$, $1 \le i, \le n$, $i \ne j$, and so without loss of generality $j \ne 1$ Then

$$\begin{aligned}
d_T(x, y) &\le d_T(x, v_i) + d_T(v_i, x_0) + d_T(v_j, x_0) + d_T(v_j, w) \\
&\le 4d_{T'}(x, v_i) + 4d_{T'}(v_j, y) + d_{T_i}(v_i, x_i) + d_{T_j}(v_j, x_j) + d_{\bar{T}}(x_i, x_0) + d_{\bar{T}}(x_j, x_0) \\
&\quad \text{(By the induction Hypothesis)} \\
&= 4d_{T'}(x, v_i) + 4d_{T'}(v_j, y) + r_i + r_j + r_0 \\
&= 4(d_{T'}(x, v_i) + d_{T'}(v_j, y)) + r_i + r_j + 2r_1 \\
&= 4(d_{T'}(x, v_i) + d_{T'}(v_j, y)) + \begin{cases} 3r_1 + r_j & \text{if } i = 1 \\ 2r_i + 2r_j & \text{if } i \ne 1 \end{cases} \\
&\le 4d_{T'}(x, y).
\end{aligned}$$

Secondly, we verify the second inequality in (48). Let $x, y \in V'$. If x, y lie both in V'_j for some j; we deduce our claim again from the induction hypothesis. So assume that $x \in V_i$ and $y \in V_j$, $1 \le i, \le n$, $i \ne j$, Also assume that $j \ne 1$.

We deduce that

$$\begin{aligned}
d_{T'}(x, y) &= d_{T'_i}(x, v_i) + d_{T'}(v_i, v_j) + d_{T'_j}(y, v_j) \\
&\le d_{T'_i}(x, v_i) + d_{T'_j}(y, v_j) + r_i + r_j \\
&\le 2d_{T_i}(x, x_i) - r_i + 2d_{T_j}(y, x_j) - r_j + r_i + r_j \\
&= 2d_{T_i}(x, x_i) + 2d_{T_j}(y, x_j) \\
&\le 2d_T(x, a_i) + 2d_T(y, a_j) \le 2d_T(x, y).
\end{aligned}$$

We finally verify (47). Let $x \in V'$ and thus $x \in V'_j$ for some $j = 1, 2 \dots n$.

Case 1: $j = 1$. Thus x and $v_1 = v_0$ are in T_1 and it follows that

$$\begin{aligned}
d_{T'}(x, v_1) &= d_{T'_1}(x, v_1) \text{ by definition of } T' \; [x, v_1] \subset V'_1 \\
&\le 2d_{\bar{T}}(x, x_1) - r_1 \text{ (by (45))}
\end{aligned}$$

$$\begin{aligned}&\le 2(d_T(x,x_0)-r_0/2)-r_1/2\\&=2d_T(x,x_0)-r_0.\end{aligned}$$

Case 2: $j>1$. Then

$$\begin{aligned}d_{T'}(x,v_1)&=d_{T'_j}(x,v_j)+d_{T'}(v_j,v_1)\\&\le 2d_{T_j}(x,x_j)-r_j+r_j \text{ (by (45))}\\&=2d_{T_j}(x,x_j)\\&=2(d_T(x,x_0)-r_0/2)=2d_T(x,x_0)-r_0.\end{aligned}$$

□

Proof of Theorem 6.5. As in the proof of Lemma 6.6, we can assume that Leaf $(T)\subset V'$. We will prove the result by the induction of the cardinality of the set $V'\setminus$ Leaf (T) Lemma 6.6 handles the case that $V'\subset$ Leaf (T).

Assume $a\in V'\setminus$ Leaf (T). We choose a as the root and denote the partial order of T, if by $\succeq$. Let n be the degree of a and $a_1,a_2,\ldots a_n$ be the neighbors of a. We define trees $T_j=(V_j,E_j,W_j)$ and $V'_j\subset V_j$, for $j=1,2,\ldots n$ by

$$V_j=\{x\in V: x\succeq a_j\}\cup\{a\}, E_j=E\cap[V_j]^2, W_j=W|_{E_j} \text{ and } V'_j=V'\cap V_j$$

Then the cardinality of $V'_j\setminus$ Leaf (T_j) is strictly smaller than the cardinality of $V'\setminus$ Leaf (T), and we can apply the inductive hypothesis to obtain for each $j=1,2\ldots n$ a tree $T'_j=(V'_j,E'_j,W'_j)$ satisfying (44).

Thus since the sets $V'_j\setminus\{a\}$, are pair wise disjoint and $a\in V_j$, for $j=1,2\ldots,n$ $T'=(V',E',W')$, with $E'=\cup_{j=1}^n E'_1$, $W':E'\to(0,\infty), e\mapsto W'_j(e)$, if $e\in E'_j$ is also a tree. Since every path from an element in V_i to and element in V_j has to contain a it follows that $T'=(V',E',W')$ satisfies (44). □

Using Theorem 6.2 and Corollary 6.5 we obtain the following results

Corollary 6.7. *Assume that (M,d) embeds D-stochastically into trees, then (M,d) embeds bijectively $8D$-stochastically int trees.*

Thus there are n, so that for $i=1,2,\ldots$ there is a set $E_i\subset[M]^2$, a metric d_i, and numbers $p_i\in(0,1]$, so that $\sum_{i=1}^n p_i=1$, $T_i=(M,E_i)$ is a tree, and d_i a geodesic distance on M with respect to T_i and so that for all $x,y\in M$

$$d(x,y)\le d_i(x,y), \text{ for } i=1,2,\ldots,n, \tag{49}$$

$$\sum_{j=1}^n p_i d_i(x,y)\le Dd(x,y). \tag{50}$$

Moreover we can assume that for $i=1,2,\ldots n$ actually d_i is the geodesic metric generated by the weight function

$$w_i: E_i\to[0,\infty), \quad \text{with } w_i(e)=d(u,v) \text{ if } e=\{u,v\}\in E_i.$$

Proof. The existence of n sets $E_i \subset [M]^2$, a metrics d_i, and numbers $p_i \in (0,1]$, for $i = 1,2,\dots,n$ follows from Theorem 6.2 and Corollary 6.5. To see the moreover part, let us denote for $i = 1,2,\dots,n$ the geodesic distance on M generated by the above defined weight function w_i by $\tilde{d}_i$. From the triangle inequality, it follows that $\tilde{d}_i(x,y) \geq d(x,y)$. We note that from (49) it follows that for every $i = 1,2,\dots,n$, any $e = \{u,v\} \in E_i$, we have that $d(u,v) \leq d_i(u,v)$, and thus $d(x,y) \leq \tilde{d}_i(x,y) \leq d_i(x,y)$, for any $x,y \in M$, which implies (50). □

References

[1] R. Aliaga and E. Pernecká, *Supports and extreme points in Lipschitz-free spaces*, Rev. Mat. Iberoam. **36** (2020), no. 7, 2073–2089, DOI 10.4171/rmi/1191.

[2] F. Baudier, C. Gartland, and Th. Schlumprecht, *L_1-distortion of Wasserstein metrics: a tale of two dimensions*, Transactions of the American Mathematical Society **10** (2023), 1077–1118.

[3] F. Baudier, P. Motakis, Th. Schlumprecht, and A. Zsák, *Stochastic approximation of lamplighter metrics*, Bull. Lond. Math. Soc. **54** (2022), no. 5, 1804–1826, DOI 10.1112/blms.12657.

[4] A. Dalet, P. Kaufmann, and Antonín Procházka, *Characterization of metric spaces whose free space is isometric to ℓ_1*, Bull. Belg. Math. Soc. Simon Stevin **23** (2016), no. 3, 391–400.

[5] J. Dixmier, *Sur un théoreème de Banach*, Duke Mathematical Journal **15** (1948), 1057 – 71.

[6] 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.

[7] G. B. Folland, *Real analysis*, Second, Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[8] A. Godard, *Tree metrics and their Lipschitz-free spaces*, Proc. Amer. Math. Soc. **138** (2010), 4311 – 4320.

[9] G. Godefroy and N. J. Kalton, *Lipschitz-free Banach spaces*, Studia Math. **159** (2003), 121–141.

[10] A. Gupta, *Steiner points in tree metrics don't (really) help*, Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms (Washington, DC, 2001), 2001, pp. 220–227.

[11] A. Gupta, I. Newman, Y. Rabinovich, and Al. Sinclair, *Cuts, trees and l_1-embeddings of graphs*, Combinatorica **24** (2004), no. 2, 233–269.

[12] S. Heinrich and P. Mankiewicz, *Applications of ultrapowers to the uniform and Lipschitz classification of Banach spaces*, Studia Math. **73** (1982), no. 3, 225–251.

[13] N. J. Kalton, *Spaces of Lipschitz and Hölder functions and their applications*, Collect. Math. **55** (2004), 171–217.

[14] L. V. Kantorovich, *On the translocation of masses*, Dokl. Akad. Nauk SSSR **37** (1942), 227–229.

[15] S. V. Kislyakov, *Sobolev imbedding operators, and the nonisomorphism of certain Banach spaces*, Funkcional. Anal. i Priložen. **9** (1975), no. 4, 22–27. MR0627173

[16] M. Mathey-Prevot and A. Valette, *Wasserstein distance and metric trees* **69** (2023), no. 3-4, 315–333.

[17] A. Naor and G. Schechtman, *Planar earthmover is not in L_1*, SIAM J. Comput. **37** (2007), 804–826 (electronic).

[18] K. Ng, *On a theorem of Dixmier*, Math.Scand. **29** (1071), 279 – 280.

[19] N. Weaver, *Lipschitz algebras*, First Edition, World Scientific Publishing Co. Inc., River Edge, NJ, 1999.

Department of Mathematics, Texas A&M University, College Station, TX 77843, USA and Faculty of Electrical Engineering, Czech Technical University in Prague, Zikova 4, 166 27, Prague, Czech Republic