HAL open science

Wasserstein medians: robustness, PDE characterization and numerics

Guillaume Carlier, Enis Chenchene, Katharina Eichinger

To cite this version:

Guillaume Carlier, Enis Chenchene, Katharina Eichinger. Wasserstein medians: robustness, PDE characterization and numerics. 2023. hal-04154602

HAL Id: hal-04154602

https://hal.science/hal-04154602v1

Preprint submitted on 6 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

arXiv:2307.01765v1 [math.OC] 4 Jul 2023

Wasserstein medians: robustness, PDE characterization and numerics

G. Carlier* E. Chenchene† K. Eichinger‡

July 6, 2023

Abstract

We investigate the notion of Wasserstein median as an alternative to the Wasserstein barycenter, which has become popular but may be sensitive to outliers. In terms of robustness to corrupted data, we indeed show that Wasserstein medians have a breakdown point of approximately $\frac{1}{2}$. We give explicit constructions of Wasserstein medians in dimension one which enable us to obtain L^p estimates (which do not hold in higher dimensions). We also address dual and multimarginal reformulations. In convex subsets of $\mathbb{R}^d$, we connect Wasserstein medians to a minimal (multi) flow problem à la Beckmann and a system of PDEs of Monge–Kantorovich-type, for which we propose a p-Laplacian approximation. Our analysis eventually leads to a new numerical method to compute Wasserstein medians, which is based on a Douglas–Rachford scheme applied to the minimal flow formulation of the problem.

Keywords: Wasserstein medians, optimal transport, duality, Beckmann's problem, p-Laplace system approximation, Douglas–Rachford splitting method.

1 Introduction

The notions of mean and median are well-known to be of variational nature. For instance, the arithmetic mean of a sample composed by N points in $\mathbb{R}^d$ is the minimizer of the sum of the *squared* Euclidean distances to the sample points. Minimizing a weighted sum of distances to the sample points, one gets a notion of weighted medians, which in the literature is commonly referred to as Torricelli–Fermat–Weber points or geometric medians. As pointed out by Maurice Fréchet in his seminal work [29], these definitions can be generalized to any metric space $(\mathcal{X}, d)$, yielding the notion of Fréchet mean and Fréchet median (or in general typical element).

The concept of Wasserstein barycenter, which corresponds to Fréchet means over the Wasserstein space of probability measures with finite second moments and equipped with the quadratic Wasserstein distance, was introduced and extensively studied in [1]. Since then, research on Wasserstein barycenters has expanded in various directions. For instance, investigations have been conducted on Riemannian manifolds [33], population barycenters involving possibly infinitely many measures [35], and Radon spaces [34]. The concept has gained popularity as a valuable tool for meaningful geometric interpolation between probability measures, finding applications in diverse

*CEREMADE, Université Paris Dauphine, PSL, and INRIA-Paris, MOKAPLAN,
email: carlier@ceremade.dauphine.fr
†Department of Mathematics and Scientific Computing, University of Graz, Graz, Austria.
email: enis.chenchene@uni-graz.at
‡CMAP, École Polytechnique
email: katharina.eichinger@polytechnique.edu

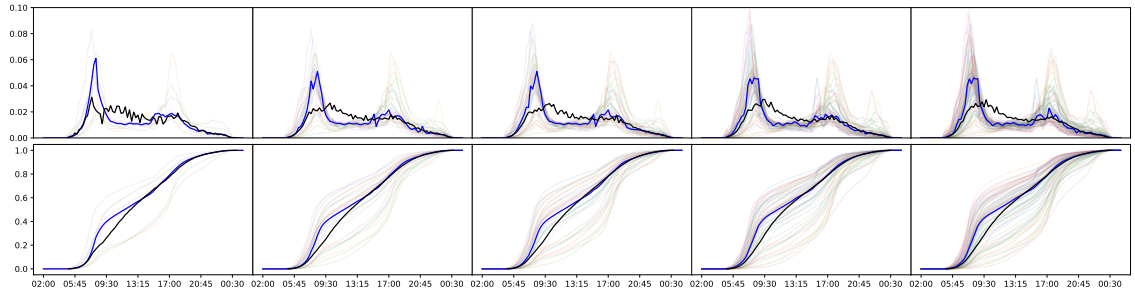


Figure 1: Superposition of a Wasserstein median (blue), a Wasserstein barycenter (black) and the corresponding sample of $N = 9,\ 29,\ 39,\ 59,\ 81$ one-dimensional histograms. Each histogram represents the daily attendance frequency of some London underground stations[1]. Second row: the corresponding cumulative distribution functions.

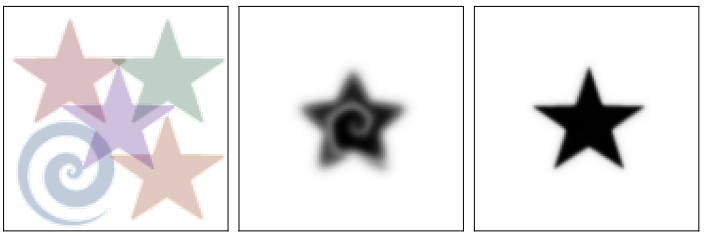
fields such as image synthesis [44], template estimation [10], bayesian learning [6], and statistics [40]. Despite the inherent complexity of computing Wasserstein barycenters [2], numerical methods based on entropic regularization and the Sinkhorn algorithm have demonstrated their efficiency in calculating these interpolations [9, 20, 42].

In the present paper, we investigate a slightly different problem, namely that of Wasserstein medians, which, given weights $\lambda_1, \dots \lambda_N$ and probability measures $\nu_1, \dots, \nu_N$ with finite first moments, over a metric space $\mathcal{X}$, consists in finding a probability measure ν minimizing the dispersion criterion $\sum_{i=1}^N \lambda_i W_1(\nu_i, \nu)$. Following Fréchet's metric viewpoint, this amounts to look for Torricelli–Fermat–Weber or equiprobable points in the Wasserstein space of order one. Our primary motivation for studying these objects comes from the following question: does the well-known *robustness* of geometric medians extend to Wasserstein medians? Consider for instance the problem of averaging the daily attendance frequency of some London underground stations as in Figure 1 or the five pictures on the left of Figure 2. It is pretty clear in these examples that Wasserstein medians show some sort of robustness, and that in general they should behave quite differently from the barycenter.

Our objective is to further explore the notion of Wasserstein median with a first focus on stability and robustness. We also investigate in depth the one-dimensional case where special constructions (which we call vertical and horizontal selections) select medians which inherit properties of the sample measures, in particular, we show that if all the sample measures ν_i are absolutely continuous with densities bounded by some M_i, then there exists a Wasserstein median with a density bounded by $\max_i M_i$, which, as we will show later (Example 5.3), cannot be true in higher dimensions. For more general situations, we present some general tools to study Wasserstein medians, such as multi-marginal and dual formulations for the initial convex minimization problem. To the best of our knowldege, Wasserstein medians have not been very much investigated even in the Euclidean setting with more than two sample measures, however related optimal matching problems (with two sample measures and additional constraints) have been studied in [38], [31] and [15]. In the Euclidean setting in several dimensions, we also characterize medians by a minimal flow problem à la Beckmann [7] and a system of PDEs of Monge–Kantorovich type. This analysis leads to a new numerical method to compute Wasserstein medians, which is based on a Douglas–Rachford scheme applied to the minimal flow formulation of the problem.

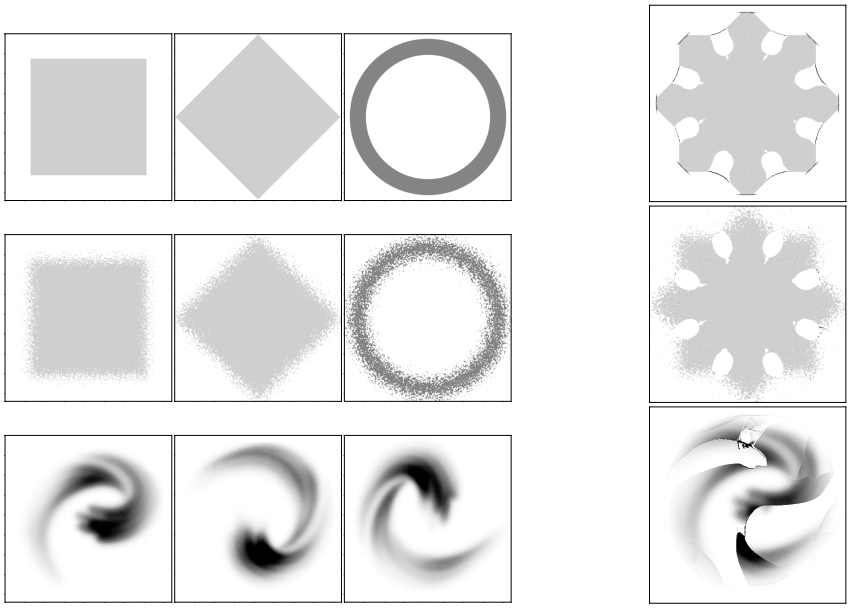
The paper is organized as follows: in Section 2, we introduce the problem, show existence of Wasserstein medians and consider some basic examples. In Section 3, we discuss the stability of the notion subject to perturbations of the sample measures and prove that the *break-down point*

[1] Tfl open data `https://tfl.gov.uk/info-for/open-data-users`, accessed on June 29, 2023.



(a) Sample measures. (b) Barycenter. (c) Median.

Figure 2: Comparison between a Wasserstein barycenter and a Wasserstein median for a sample of five measures computed with Sinkhorn (cf., Section 7) in 1000 iterations.



(a) Sample measures. (b) Wasserstein medians.

Figure 3: Some Wasserstein medians on a 420×420 grid computed with Douglas–Rachford up to 2000 iterations, with a final residual of about $\sim 10^{-7}$, cf., Section 7.

of the Wasserstein median problem with uniform weights is at least $1/2$, i.e. to *drastically corrupt* the estimation of the Wasserstein median one has to modify at least half of the sample measures. In Section 4, we focus on the one-dimensional case and emphasize the properties of medians which we call vertical and horizontal median selections. In Section 5, we present dual and multi-marginal formulations of the problem. In Section 6, we use a minimal flow formulation of the Wasserstein median problem to derive a system of Monge–Kantorovich type PDEs that characterizes medians. We also describe an approximation by a system of p-Laplace equations. We conclude in Section 7 with a brief description of the numerical methods we implemented to obtain the various figures in this paper and present a new one based on a Douglas–Rachford scheme on the flow formulation.

2 Definition, existence and basic examples

Setting. Let $(\mathcal{X}, d)$ be a *proper* metric space, i.e. a metric space in which closed balls are compact. This implies in particular that $(\mathcal{X}, d)$ is *Polish*, i.e. separable and complete. Note that $(\mathcal{X}, d)$ being proper is a natural assumption to define medians by minimization of weighted sums of distances; indeed this implies that for every integer $N \geq 1$, every $(x_1, \ldots, x_N) \in \mathcal{X}^N$ and every $\boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_N)$ in the simplex Δ_N:

$$\Delta_N := \Big\{(\lambda_1, \ldots, \lambda_N) \in \mathbb{R}_+^N \ : \ \sum_{i=1}^N \lambda_i = 1\Big\},$$

the set of medians of $(x_1, \ldots, x_N)$ with weights $\boldsymbol{\lambda}$, defined by

$$\mathsf{M}_{\boldsymbol{\lambda}}(x_1, \ldots, x_N) := \arg\min_{x \in \mathcal{X}} \sum_{i=1}^N \lambda_i d(x_i, x) \tag{1}$$

is a nonempty (and compact) subset of $\mathcal{X}$.

Example 2.1 (**Medians on the real line**). *For $\mathcal{X} = \mathbb{R}$ equipped with the distance associated with the absolute value, $N \geq 1$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N) \in \Delta_N$ and $\boldsymbol{x} := (x_1, \ldots, x_N) \in \mathbb{R}^N$, $\mathsf{M}_{\boldsymbol{\lambda}}(\boldsymbol{x})$ is the set of minimizers of the convex, piecewise affine function $x \mapsto f(x) := \sum_{i=1}^N \lambda_i |x - x_i|$, this function being right and left differentiable at each point with corresponding one-sided derivative given by*

$$f'(x^-) = \sum_{i \,:\, x_i < x} \lambda_i - \sum_{i \,:\, x_i \geq x} \lambda_i = 2 \sum_{i \,:\, x_i < x} \lambda_i - 1, \ f'(x^+) = 2 \sum_{i \,:\, x_i \leq x} \lambda_i - 1.$$

We see that x belongs to the median interval $\mathsf{M}_{\boldsymbol{\lambda}}(\boldsymbol{x})$ if and only if $f'(x^-) \leq 0 \leq f'(x^+)$, i.e.

$$\sum_{i \,:\, x_i < x} \lambda_i \leq \frac{1}{2} \leq \sum_{i \,:\, x_i \leq x} \lambda_i,$$

that is, $\mathsf{M}_{\boldsymbol{\lambda}}(\boldsymbol{x}) = [\mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{x}), \mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{x})]$, where $\mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{x})$ and $\mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{x})$ stand for the lower and upper medians respectively, which are given by:

$$\mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{x}) := \inf\Big\{y \in \mathbb{R} \ : \ \sum_{i \,:\, x_i \leq y} \lambda_i \geq \frac{1}{2}\Big\}, \ \mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{x}) := \sup\Big\{y \in \mathbb{R} \ : \ \sum_{i \,:\, x_i < y} \lambda_i \leq \frac{1}{2}\Big\}. \tag{2}$$

We shall use extensively properties of lower and upper medians when studying Wasserstein medians on $\mathbb{R}$ in Section 4. Obviously, since f is affine in the neighbourhood of each point of $\mathbb{R} \backslash \{x_1, \ldots, x_N\}$, both $\mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{x})$ and $\mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{x})$ belong to the sample $\{x_1, \ldots, x_N\}$:

$$I_\pm(\boldsymbol{x}) := \{i = 1, \ldots, N \ : \ \mathsf{M}_{\boldsymbol{\lambda}}^\pm(\boldsymbol{x}) = x_i\} \neq \emptyset. \tag{3}$$

Note also both $\mathsf{M}_{\boldsymbol{\lambda}}^-$ and $\mathsf{M}_{\boldsymbol{\lambda}}^+$ are positively homogeneous and that setting $\boldsymbol{e} = (1, \ldots, 1)$,

$$\mathsf{M}_{\boldsymbol{\lambda}}^\pm(\boldsymbol{e}) = 1, \ \mathsf{M}_{\boldsymbol{\lambda}}^\pm(\alpha \boldsymbol{x}) = \alpha \mathsf{M}_{\boldsymbol{\lambda}}^\pm(\boldsymbol{x}), \quad \text{for all } \alpha \in \mathbb{R}_+.$$

Of course, in general, medians are highly non-unique. For instance if $N = 2k$ is even, $\lambda_i = 1/N$ and $x_1 < \cdots < x_N$, the median interval is $[x_k, x_{k+1}]$. A mild condition guaranteeing uniqueness i.e. $\mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{x}) = \mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{x})$ for every $\boldsymbol{x}$ is:

$$\text{There is no subset } I \subset \{1, \ldots, N\} \text{ such that: } \sum_{i \in I} \lambda_i = \frac{1}{2}. \tag{4}$$

Despite non-uniqueness, both selections $\mathsf{M}^+_{\boldsymbol{\lambda}}$ *and* $\mathsf{M}^-_{\boldsymbol{\lambda}}$ *enjoy nice properties: obviously they are monotone in each of their argument and invariant by translation, that is, for all* $\boldsymbol{x} \geq \boldsymbol{y}$ *(i.e.* $\boldsymbol{x}-\boldsymbol{y} \in \mathbb{R}^N_+$*) we have* $\mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{x}) \geq \mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{y})$*, and for* $\alpha \in \mathbb{R}$ *it holds* $\mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{x}+\alpha\boldsymbol{e}) = \mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{x}) + \alpha$*. This implies in particular that for every* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *one has*

$$\mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{x}) + \min_{i=1,\dots,N}(y_i - x_i) \leq \mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{y}) \leq \mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{x}) + \max_{i=1,\dots,N}(y_i - x_i), \tag{5}$$

so that $\mathsf{M}^\pm_{\boldsymbol{\lambda}}$ *are Lipschitz continuous. Inequality (5) will be very useful for studying horizontal and vertical Wasserstein median selections on the real line in Section 4. In fact, we will also need to use a slightly refined form of (5), namely: for all* $\boldsymbol{x}$ *there exists* $\varepsilon > 0$ *such that for all* $\boldsymbol{y}$ *with* $\|\boldsymbol{x}-\boldsymbol{y}\|_\infty \leq \varepsilon$ *it holds*

$$\mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{x}) + \min_{i\in I_\pm(\boldsymbol{x})}(y_i - x_i) \leq \mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{y}) \leq \mathsf{M}^\pm_{\boldsymbol{\lambda}}(\boldsymbol{x}) + \max_{i\in I_\pm(\boldsymbol{x})}(y_i - x_i), \tag{6}$$

where we recall that $I_\pm(\boldsymbol{x})$ *are given by (3). The proof of (6) is postponed to the appendix.*

Example 2.2 (Torricelli–Fermat–Weber points). *Consider now* $\mathcal{X} = \mathbb{R}^d$ *equipped with the distance associated with the Euclidean norm* $|\cdot|$*,* $\boldsymbol{\lambda} \in \Delta_N$ *and* $(x_1,\dots,x_N) \in \mathcal{X}^N$*, by definition,* $x \in \mathsf{M}_{\boldsymbol{\lambda}}(x_1,\dots,x_N)$ *if and only if* x *minimizes the convex function* $\sum_{i=1}^N \lambda_i|\cdot - x_i|$ *i.e. satisfies the optimality condition*

$$0 \in \sum_{i=1}^N \lambda_i \partial|\cdot|(x-x_i),$$

where $\partial|\cdot|(x-x_i)$ *is the subdifferential of the Euclidean norm at* $x-x_i$*:*

$$\partial|\cdot|(x-x_i) = \{p \in \mathbb{R}^d \ : \ |p| \leq 1,\ \langle p, x-x_i\rangle = |x-x_i|\} = \begin{cases} B(0,1) & \text{if } x = x_i \\ \frac{x-x_i}{|x-x_i|} & \text{otherwise} \end{cases}$$

where $B(0,1)$ *stands for the closed unit Euclidean ball. Therefore* $x \in \mathsf{M}_{\boldsymbol{\lambda}}(x_1,\dots,x_N)$ *if and only if there exist* $p_1,\dots,p_N$ *such that*

$$|p_i| \leq 1,\ \langle p_i, x-x_i\rangle = |x-x_i|,\ i=1,\dots,N,\ \text{and} \sum_{i=1}^N \lambda_i p_i = 0. \tag{7}$$

Note that for $x \in \mathsf{M}_{\boldsymbol{\lambda}}(x_1,\dots,x_N)$ *either* $x = x_i$ *for some* i *or*

$$\sum_{i=1}^N \lambda_i \frac{x-x_i}{|x-x_i|} = 0$$

so that in any case x *is a convex combination of* $x_1,\dots,x_N$*, we thus have*

$$\mathsf{M}_{\boldsymbol{\lambda}}(x_1,\dots,x_N) \subset \mathrm{co}\{x_1,\dots,x_N\}.$$

Denote by $\mathcal{P}(\mathcal{X})$ the set of Borel probability measures on $\mathcal{X}$. Recall that on $\mathcal{P}(\mathcal{X})$ the narrow topology is the coarsest topology making $\mu \in \mathcal{P}(\mathcal{X}) \mapsto \int_{\mathcal{X}} f \mathrm{d}\mu$ continuous for every continuous and bounded function f on $\mathcal{X}$ and that this topology is metrizable on $\mathcal{P}(\mathcal{X})$ (so that there is no need to distinguish between narrow compactness and narrow sequential compactness). We denote by $\mathcal{P}_1(\mathcal{X})$ the set of Borel probability measures with finite first moment i.e. the set of $\mu \in \mathcal{P}(\mathcal{X})$ for which for some (equivalently for all) $x_0 \in \mathcal{X}$, $d(x_0,\cdot) \in L^1(\mathcal{X},\mu)$. We endow $\mathcal{P}_1(\mathcal{X})$ with the *Wasserstein* distance of order one:

$$W_1(\mu,\nu) := \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathcal{X}^2} d(x,y)\mathrm{d}\gamma(x,y),$$

where $\Pi(\mu,\nu)$ is the set of transport plans between μ and ν i.e. the set of Borel probability measures on $\mathcal{X}^2$ with marginals μ and ν. With this choice, the metric space $(\mathcal{P}_1(\mathcal{X}), W_1)$ is a Polish (but not necessarily proper) space. Let us recall the Kantorovich–Rubinstein duality formula which expresses $W_1(\mu,\nu)$ as

$$W_1(\mu,\nu) := \sup\Big\{ \int_{\mathcal{X}} u \mathrm{d}\mu - \int_{\mathcal{X}} u \mathrm{d}\nu \ : \ u : \mathcal{X} \to \mathbb{R}, \text{ 1-Lipschitz}\Big\}, \tag{8}$$

in particular, $W_1(\mu,\nu)$ is the dual Lipschitz semi-norm of $\mu - \nu$ and the linear interpolation $\mu_t := (1-t)\mu + t\nu$ for $t \in [0,1]$, is obviously a geodesic between μ and ν i.e.:

$$W_1(\mu_t, \mu_s) = |t-s| W_1(\mu,\nu), \qquad \text{for all } (t,s) \in [0,1]^2. \tag{9}$$

Note that convergence in $\mathcal{P}_1(\mathcal{X})$ for W_1 implies convergence for the narrow topology but is stronger unless $\mathcal{X}$ is compact. For proofs of these classical facts and more on Wasserstein distances, we refer to the textbooks [4, 46].

Wasserstein medians. As mentioned in the introduction, on $(\mathcal{P}_1(\mathcal{X}), W_1)$, one can naturally define medians in the Fréchet sense as follows. Given $N \geq 1$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \mathcal{P}_1(\mathcal{X})^N$ and $\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_N) \in \Delta_N$, consider the weighted dispersion functional

$$\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}(\mu) := \sum_{i=1}^N \lambda_i W_1(\nu_i, \mu), \quad \text{for all } \mu \in \mathcal{P}_1(\mathcal{X}) \tag{10}$$

then Wasserstein medians are defined as minimizers of this dispersion functional:

Definition 2.3 (**Wasserstein medians**). *For $N \geq 1$, let $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \mathcal{P}_1(\mathcal{X})^N$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N) \in \Delta_N$, defining $\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}(\nu)$ by (10), we call Wasserstein median of $(\nu_1, \dots, \nu_N)$ with weights $\boldsymbol{\lambda}$ any solution of the following (convex) problem*

$$v(\boldsymbol{\lambda}, \boldsymbol{\nu}) := \min_{\mu \in \mathcal{P}_1(\mathcal{X})} \mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}(\mu) \tag{11}$$

We denote by $\mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \dots, \nu_N)$ the set of all Wasserstein medians of $\boldsymbol{\nu}$ with weights $\boldsymbol{\lambda}$.

The existence of a solution of (11) follows from the direct method:

Lemma 2.4 (**Existence of Wasserstein medians**). *Let $N \geq 1$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \mathcal{P}_1(\mathcal{X})^N$ and $\boldsymbol{\lambda} \in \Delta_N$, then there exists a minimizer of (11) and the set $\mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \dots, \nu_N)$ is a convex and narrowly compact subset of $\mathcal{P}_1(\mathcal{X})$.*

Proof. The functional $\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}$ is l.s.c. for the narrow topology (this follows at once from (8)) and by the triangle inequality for every $x_0 \in \mathcal{X}$ and every $\mu \in \mathcal{P}_1(\mathcal{X})$ one has

$$W_1(\delta_{x_0}, \mu) = \int_{\mathcal{X}} d(x_0, x) \mathrm{d}\mu(x) \leq \mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}(\mu) + \mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}(\delta_{x_0}),$$

which implies that the first moment is uniformly bounded on sublevel sets of $\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}$. Since $(\mathcal{X}, d)$ is proper, this implies that sublevel sets of $\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}$ are tight hence narrowly relatively compact by Prokhorov's theorem. This implies nonemptiness and narrow compactness of $\mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \dots, \nu_N)$, convexity follows from the convexity of $\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}$. □

Let us end this section with some simple explicit examples.

Example 2.5 (**Medians of Dirac masses**). *If $\nu_i = \delta_{x_i}$ is a Dirac mass for all $i = 1, \dots, N$, then $\mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \dots, \nu_N)$ is nothing but the set of probability measures supported on $\mathsf{M}_{\boldsymbol{\lambda}}(x_1, \dots, x_N)$. In particular if $\mathcal{X} = \mathbb{R}$, and $x_1 \le x_2$ and $\boldsymbol{\lambda} = (\lambda_1, 1-\lambda_1)$ with $\lambda_1 \in (0,1)$, $\mathsf{M}_{\boldsymbol{\lambda}}(x_1, x_2) = [x_1, x_2]$ so that $\mathsf{Med}_{\boldsymbol{\lambda}}(\delta_{x_1}, \delta_{x_2})$ is the set of* all *probability measures supported on $[x_1, x_2]$.*

Example 2.6 (**Threshold effect**). *Suppose that there is $J \subset \{1, \dots, N\}$ with $\sum_{j \in J} \lambda_j \ge \frac{1}{2}$ and $\boldsymbol{\nu} := (\nu_1, \dots, \nu_N)$ with $\nu_j = \rho$ for $j \in J$ for some $\rho \in \mathcal{P}_1(\mathcal{X})$. Then a Wasserstein median of $\boldsymbol{\nu}$ is given by ρ since for any $\tilde{\rho} \in \mathcal{P}_1(\mathcal{X})$, denoting $J^c := \{1, \dots, N\} \setminus J$*

$$\begin{aligned}
\sum_{i=1}^N \lambda_i W_1(\nu_i, \rho) &= \sum_{i \in J^c} \lambda_i W_1(\nu_i, \rho) \\
&\le \sum_{i=1}^N \lambda_i W_1(\nu_i, \tilde{\rho}) + \sum_{i \in J^c} \lambda_i W_1(\tilde{\rho}, \rho) - \sum_{i \in J} \lambda_i W_1(\nu_i, \tilde{\rho}) \\
&= \sum_{i=1}^N \lambda_i W_1(\nu_i, \tilde{\rho}) + \underbrace{\left(\sum_{i \in J^c} \lambda_i - \sum_{i \in J} \lambda_i \right)}_{\le 0} W_1(\tilde{\rho}, \rho).
\end{aligned}$$

Note that if $\sum_{j \in J} \lambda_j > \frac{1}{2}$, this also proves that the Wasserstein median is unique and equal to ρ. Note that this threshold effect is not specific to Wasserstein medians but holds for Fréchet medians in any metric space.

Example 2.7 (**Medians of two measures**). *If $N = 2$, $\nu_1 \ne \nu_2$, it follows from the previous example that when $\lambda_1 \in (1/2, 1)$ (respectively $\lambda_1 \in (0, 1/2)$) the median of (ν_1, ν_2) with weights $(\lambda_1, 1-\lambda_1)$ is ν_1 (respectively ν_2), when $\lambda_1 = \lambda_2 = 1/2$, by the triangle inequality any interpolate $(1-t)\nu_1 + t\nu_2$, $t \in [0,1]$ belongs to $\mathsf{Med}_{1/2,1/2}(\nu_1, \nu_2)$.*

Example 2.8 (**Medians of translated measures**). *Consider $\mathcal{X} = \mathbb{R}^d$ endowed with the Euclidean distance, $\mu \in \mathcal{P}_1(\mathcal{X})$, $(x_1, \dots, x_N) \in \mathcal{X}^N$ and let $\nu_i := \tau_{v_i \#}\mu$ be the translation of μ by v_i (i.e. $\tau_{v_i \#}\mu(A) = \mu(A - v_i)$, for every Borel subset A of $\mathbb{R}^d$). We claim that whenever $x \in \mathsf{M}_{\boldsymbol{\lambda}}(x_1, \dots, x_N)$ one has $\tau_{x\#}\mu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\tau_{x_1 \#}\mu, \dots, \tau_{x_N \#}\mu)$. To see this, let $(p_1, \dots, p_N)$ satisfy the optimality condition (7), then we first have*

$$\sum_{i=1}^N \lambda_i W_1(\tau_{x_i \#}, \tau_{x\#}\mu) \le \sum_{i=1}^N \lambda_i |x - x_i| = \sum_{i=1}^N \lambda_i \langle p_i, x - x_i \rangle = -\sum_{i=1}^N \lambda_i \langle p_i, x_i \rangle.$$

Let now $\nu \in \mathcal{P}_1(\mathcal{X})$, since $p_i \in B(0,1)$ the affine function $u_i(y) := \langle p_i, y + x - x_i \rangle$ is 1-Lipschitz so that by the Kantorovich–Rubinstein formula

$$\begin{aligned}
W_1(\tau_{x_i \#}\mu, \nu) &\ge \big\langle p_i, \int_{\mathbb{R}^d} (y - x_i + x) \mathrm{d}\nu(y) \big\rangle - \big\langle p_i, \int_{\mathbb{R}^d} (y + x) \mathrm{d}\mu(y) \big\rangle \\
&= \big\langle p_i, \int_{\mathbb{R}^d} (y - x_i) \mathrm{d}\nu(y) \big\rangle - \big\langle p_i, \int_{\mathbb{R}^d} y \mathrm{d}\mu(y) \big\rangle.
\end{aligned}$$

Multiplying by λ_i, summing and using (7), we obtain

$$\sum_{i=1}^N \lambda_i W_1(\tau_{x_i \#}\mu, \nu) \ge -\sum_{i=1}^N \lambda_i \langle p_i, x_i \rangle \ge \sum_{i=1}^N \lambda_i W_1(\tau_{x_i \#}, \tau_{x\#}\mu)$$

which shows that $\tau_{x\#}\mu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\tau_{x_1 \#}\mu, \dots, \tau_{x_N \#}\mu)$.

3 Stability and robustness

The stability with respect to perturbations of the sample measures is a crucial property for any location estimator especially when the underlying space $\mathcal{X}$ is unbounded. This is why, in this section, we will first investigate some stability properties of Wasserstein medians (improving the easy narrow stability to the stability in W_1-distance), note that Theorem 5.5 in [34] establishes strong consistency results in a much more general framework. We will then show robustness to outliers by showing that the breakdown point of Wasserstein medians on an unbounded $\mathcal{X}$ is at least 1/2, the proof will be an easy adaptation of [37] revealing that the argument is in fact quite general and actually carries over to Fréchet medians on geodesic metric spaces.

3.1 Compactness in W_1 distance and stability with respect to data

Let $N \geq 1$, $(\boldsymbol{\lambda}, \boldsymbol{\nu}) = (\lambda_1, \ldots, \lambda_N, \nu_1, \ldots, \nu_N)$ and $(\boldsymbol{\lambda}', \boldsymbol{\nu}') = (\lambda'_1, \ldots, \lambda'_N, \nu'_1, \ldots, \nu'_N)$ in $\Delta_N \times \mathcal{P}_1(\mathcal{X})^N$, an obvious consequence of the triangle inequality is the fact that for any $\mu \in \mathcal{P}_1(\mathcal{X})$, one has

$$\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}(\mu) \leq \mathcal{F}_{\boldsymbol{\lambda}',\boldsymbol{\nu}'}(\mu) + \max_{i=1,\ldots,N} W_1(\nu_i, , \nu'_i) + \sum_{i=1}^{N} |\lambda_i - \lambda'_i| \max_{i=1,\ldots,N} W_1(\nu'_i, \mu). \tag{12}$$

This pointwise inequality for the dispersions corresponding to $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ and $(\boldsymbol{\lambda}', \boldsymbol{\nu}')$, implies in particular that $\mathcal{F}_{\boldsymbol{\lambda}',\boldsymbol{\nu}'}$ converges to $\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}$ uniformly on W_1 balls as

$$\sum_{i=1}^{N} |\lambda_i - \lambda'_i| + \max_{i=1,\ldots,N} W_1(\nu_i, , \nu'_i) \to 0.$$

Let us also observe that for every $x_0 \in \mathcal{X}$, again by the triangle inequality, one also has the moment bound

$$\sup_{\mu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})} \int_{\mathcal{X}} d(x_0, x) \mathrm{d}\mu(x) \leq 2 \max_{i=1,\ldots,N} \int_{\mathcal{X}} d(x_0, x_i) \mathrm{d}\nu_i(x). \tag{13}$$

Recalling the definition of $v(\boldsymbol{\lambda}, \boldsymbol{\nu})$ from (11), (12) and (13) show that v is locally Lipschitz continuous for W_1. Combining the previous pointwise convergence with the narrow lower semicontinuity of W_1, (13) and the narrow compactness of measures with bounded first moments, we straightforwardly get:

Lemma 3.1. *Let $N \geq 1$, $(\boldsymbol{\lambda}, \boldsymbol{\nu}) = (\lambda_1, \ldots, \lambda_N, \nu_1, \ldots, \nu_N) \in \Delta_N \times \mathcal{P}_1(\mathcal{X})^N$ and $(\boldsymbol{\lambda}^n, \boldsymbol{\nu}^n)_{n\in\mathbb{N}} = (\lambda^n_1, \ldots, \lambda^n_N, \nu^n_1, \ldots, \nu^n_N)_{n\in\mathbb{N}}$ be a sequence in $\Delta_N \times \mathcal{P}_1(\mathcal{X})^N$ such that*

$$\sum_{i=1}^{N} |\lambda^n_i - \lambda_i| + \max_{i=1,\ldots,N} W_1(\nu^n_i, \nu_i) \to 0 \text{ as } n \to \infty. \tag{14}$$

Then, $\mathcal{F}_{\boldsymbol{\lambda}^n,\boldsymbol{\nu}^n}$ Γ-converges to $\mathcal{F}_{\boldsymbol{\lambda},\boldsymbol{\nu}}$ for the narrow topology, in particular if $\mu^n \in \mathsf{Med}_{\boldsymbol{\lambda}^n}(\boldsymbol{\nu}^n)$ for all $n \in \mathbb{N}$, narrow cluster points of $(\mu^n)_{n\in\mathbb{N}}$ belong to $\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$.

One can improve the previous (elementary and expected) result by stability in W_1 distance as follows (for more general results of this type, we refer the reader to [34]):

Theorem 3.2 (Stronger stability of Wasserstein medians). *Let $N \geq 1$, $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \Delta_N \times \mathcal{P}_1(\mathcal{X})^N$, $(\boldsymbol{\lambda}^n, \boldsymbol{\nu}^n)_{n\in\mathbb{N}}$ be a sequence in $\Delta_N \times \mathcal{P}_1(\mathcal{X})^N$ such that (14) holds and let $\mu^n \in \mathsf{Med}_{\boldsymbol{\lambda}^n}(\boldsymbol{\nu}^n)$ for all $n \in \mathbb{N}$, then $(\mu^n)_{n\in\mathbb{N}}$ admits a subsequence that converges for W_1 to some $\mu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$. In particular $\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ is compact and the set-valued map $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \Delta_N \times \mathcal{P}_1(\mathcal{X})^N \mapsto \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu}) \subset \mathcal{P}_1(\mathcal{X})$ has a closed graph for the W_1 distance.*

Proof. We already know from Lemma 3.1 that $(\mu^n)_{n\in\mathbb{N}}$ has a (not relabeled) subsequence that converges narrowly to some μ which belongs to $\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$. To improve narrow to W_1 convergence, it follows from Proposition 7.1.5 of [4], that it is enough to show that (some subsequence of) $(\mu^n)_{n\in\mathbb{N}}$ has uniformly integrable moments. More precisely, fixing $x_0 \in \mathcal{X}$ and for $R > 0$ denoting by $B(x_0, R)$ the open ball of radius R, we have to show that (passing to a subsequence if necessary)

$$\lim_{R\to+\infty} \sup_n \int_{\mathcal{X}\setminus B(x_0,R)} d(x_0, x)\mathrm{d}\mu^n(x) = 0. \tag{15}$$

Let $\gamma_i^n \in \Pi(\nu_i^n, \mu^n)$ such that $\int_{\mathcal{X}^2} d(x_i, x)\mathrm{d}\gamma_i^n(x_i, x) = W_1(\nu_i^n, \mu^n)$, since both sequences $(\nu_i^n)_{n\in\mathbb{N}}$ and $(\mu^n)_{n\in\mathbb{N}}$ are tight so is $(\gamma_i^n)_{n\in\mathbb{N}}$, passing to subsequences if necessary, we may thus assume that $(\gamma_i^n)_{n\in\mathbb{N}}$ converges narrowly to some $\gamma_i \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$. Of course $\gamma_i \in \Pi(\nu_i, \mu)$ and then

$$W_1(\nu_i, \mu) \le \int_{\mathcal{X}^2} d(x_i, x)\mathrm{d}\gamma_i(x_i, x) \le \liminf_n \int_{\mathcal{X}^2} d(x_i, x)\mathrm{d}\gamma_i^n(x_i, x) = \liminf_n W_1(\nu_i^n, \mu^n).$$

We deduce from Lemma 3.1 and the fact that $(\nu_i^n)_{n\in\mathbb{N}}$ and $(\mu^n)_{n\in\mathbb{N}}$ have uniformly bounded moments

$$\sum_{i=1}^N \lambda_i W_1(\nu_i, \mu) = \lim_n \sum_{i=1}^N \lambda_i^n W_1(\nu_i^n, \mu^n) \ge \sum_{i=1}^N \liminf_n \lambda_i^n W_1(\nu_i^n, \mu^n) = \sum_{i=1}^N \lambda_i \liminf_n W_1(\nu_i^n, \mu^n).$$

Hence, for every i for which $\lambda_i > 0$, one has $W_1(\nu_i, \mu) = \liminf_n W_1(\nu_i^n, \mu^n)$. Assuming without loss of generality that $\lambda_1 > 0$, we thus have

$$W_1(\nu_1, \mu) = \int_{\mathcal{X}^2} d(x_1, x)\mathrm{d}\gamma_1(x_1, x) = \liminf_n \int_{\mathcal{X}^2} d(x_1, x)\mathrm{d}\gamma_1^n(x_1, x).$$

Passing to a subsequence if necessary, we may assume that the liminf of the right hand side above is a true limit and then, using Lemma 5.1.7 of [4], we deduce that

$$\lim_{R\to+\infty} \sup_n \int_{\{(x_1,x)\in\mathcal{X}^2\,:\,d(x_1,x)\ge R\}} d(x_1, x)\mathrm{d}\gamma_1^n(x_1, x) = 0. \tag{16}$$

Note also that since $(\nu_1^n)_{n\in\mathbb{N}}$ converges in W_1 we also have

$$\lim_{R\to+\infty} \sup_n \int_{\mathcal{X}\setminus B(x_0,R)} d(x_0, x_1)\mathrm{d}\nu_1^n(x_1) = 0. \tag{17}$$

Defining for $R > 0$ and $t \ge 0$,

$$\Phi_R(t) := \begin{cases} t & \text{if } t \ge R, \\ 0 & \text{else,} \end{cases}$$

note that Φ_R is non decreasing and

$$\Phi_R(t + s) \le 2\Big(\Phi_{\frac{R}{2}}(t) + \Phi_{\frac{R}{2}}(s)\Big),$$

so by the triangle inequality for every $(x, x_1) \in \mathcal{X}^2$, we have

$$\Phi_R(d(x_0, x)) \le 2\Big(\Phi_{\frac{R}{2}}(d(x_0, x_1)) + \Phi_{\frac{R}{2}}(d(x_1, x))\Big).$$

Integrating with respect to γ_1^n which has marginals ν_1^n and μ^n yields

$$\begin{aligned}\int_{\mathcal{X}\setminus B(x_0,R)} d(x_0,x)\mathrm{d}\mu^n(x) &= \int_{\mathcal{X}} \Phi_R(d(x_0,x))\mathrm{d}\mu^n(x) = \int_{\mathcal{X}} \Phi_R(d(x_0,x))\mathrm{d}\gamma_1^n(x_1,x) \\ &\leq 2\int_{\mathcal{X}} \Phi_{\frac{R}{2}}(d(x_0,x_1))\mathrm{d}\nu_1^n(x_1) + 2\int_{\mathcal{X}^2} \Phi_{\frac{R}{2}}(d(x_1,x))\mathrm{d}\gamma_1^n(x_1,x) \\ &= 2\int_{\mathcal{X}\setminus B(x_0,\frac{R}{2})} d(x_0,x_1)\mathrm{d}\nu_1^n(x_1) \\ &\quad + 2\int_{\{(x_1,x)\in\mathcal{X}^2\,:\,d(x_1,x)\geq\frac{R}{2}\}} d(x_1,x)\mathrm{d}\gamma_1^n(x_1,x).\end{aligned}$$

Then, (15) readily follows from (16) and (17).

□

3.2 Robustness of Wasserstein medians

In statistics, a popular robustness index is the so-called *break-down point*. Roughly speaking, it is the largest fraction of the input data which could be corrupted (i.e. changed arbitrarily) without moving the estimation too far from the original estimation for the non-corrupted data. It is well known that the break-down point of geometric medians with uniform weights is approximately $\frac{1}{2}$, see, e.g. Theorem 2.1 and 2.2 [37], so that even corrupting about half of the data, we can stay rather confident on the output. In this section, we prove a similar result for Wasserstein medians. To do so, we first recall some basic facts about break-down points, starting with a definition of the break-down point adapted to the case of a non-unique estimator.

Definition 3.3 (**Break-down point**). *Let $(\mathcal{X},d)$ be a metric space. Let $N\geq 2$ and $\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_N)\in\Delta_N$. For a set-valued map $t_{\boldsymbol{\lambda}}:\mathcal{X}^N\to 2^{\mathcal{X}}$ with nonempty values, we define its break-down point associated to the weights $\boldsymbol{\lambda}$ at $\boldsymbol{x}=(x_1,\dots,x_N)\in\mathcal{X}^N$ by*

$$b(t_{\boldsymbol{\lambda}}(\boldsymbol{x})) := \min\Big\{\sum_{i\in I}\lambda_i\,:\;I\subset\{1,\dots,N\},\;\sup_{\substack{\boldsymbol{y}^I\in\mathcal{X}^N\\ \boldsymbol{y}^I_j=\boldsymbol{x}_j\;\forall j\notin I}}\big\{d(y,x)\;:\;y\in t_{\boldsymbol{\lambda}}(\boldsymbol{y}^I),\;x\in t_{\boldsymbol{\lambda}}(\boldsymbol{x})\big\}=+\infty\Big\}.$$

We now state the main theorem for Wasserstein medians, where the reference metric space is $\mathcal{P}_1(\mathcal{X})$ equipped with the W_1 distance. The proof is a slight generalization of Theorem 2.2. in [37].

Theorem 3.4 (**Break-down point of Wasserstein medians**). *Suppose the metric space $(\mathcal{X},d)$ is proper and unbounded. Let $N\geq 2$, $\boldsymbol{\nu}:=(\nu_1,\dots,\nu_N)\in\mathcal{P}_1(\mathcal{X})^N$ and $\boldsymbol{\lambda}:=(\lambda_1,\dots,\lambda_N)\in\Delta_N$. Then the break-down point of $\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ is given by*

$$b(\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})) = \min\Big\{\sum_{j\in J}\lambda_j\;:\;J\subset\{1,\dots,N\},\;\sum_{j\in J}\lambda_j\geq\frac{1}{2}\Big\}. \tag{18}$$

Proof. For future reference, let us denote by B the right hand-side of (18). Let us take $\nu\in\mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1,\dots,\nu_N)$ and $I\subset\{1,\dots,N\}$ such that $\sum_{i\in I}\lambda_i<\frac{1}{2}$. Denote by $\boldsymbol{\mu}:=(\mu_1,\dots,\mu_N)\in\mathcal{P}_1(\mathcal{X})^N$ a corrupted collection of $\boldsymbol{\nu}:=(\nu_1,\dots,\nu_N)$, i.e. such that $\mu_j=\nu_j$ for all $j\notin I$. Let

$$C := \max_{\rho\in\mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1,\dots,\nu_N)}\;\max_{1\leq i\leq N} W_1(\rho,\nu_i),\;\delta:=\max\Big\{\sum_{j\in J}\lambda_j\;:\;J\subset\{1,\dots,N\},\;\sum_{j\in J}\lambda_j<\frac{1}{2}\Big\}. \tag{19}$$

Let $\mu\in\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\mu})$, let us first prove by contradiction that

$$W_1(\nu,\mu)\leq\frac{2C\delta}{1-2\delta}+2C.$$

In order to do so, let $\boldsymbol{\mathcal{B}} = B_{2C}(\nu)$ be the ball with center ν and radius $2C$ with respect to the W_1 distance. Further, let

$$\xi := \mathsf{Dist}(\mu, \boldsymbol{\mathcal{B}}) := \inf_{\rho \in \boldsymbol{\mathcal{B}}} W_1(\mu, \rho).$$

Then by the triangle inequality $W_1(\mu, \nu) \leq \xi + 2C$, so that for all $j = 1, \dots, N$

$$W_1(\mu_j, \mu) \geq W_1(\mu_j, \nu) - W_1(\nu, \mu) \geq W_1(\mu_j, \nu) - (\xi + 2C). \tag{20}$$

Now suppose by contradiction that $\xi > 2C\delta/(1-2\delta)$, which in particular implies that $W_1(\mu, \nu) > 2C$. Using the fact that in $(\mathcal{P}_1(\mathcal{X}), W_1)$, line segments are geodesics (recall (9)), defining for $j = 1, \dots, N$ the interpolation $\nu_j^t := (1-t)\nu_j + t\mu$, $t \in [0,1]$, we have

$$W_1(\nu_j, \mu) = W_1(\nu_j, \nu_j^t) + W_1(\nu_j^t, \mu) \text{ for } t \in [0,1].$$

Since $W_1(\nu, \nu_j^0) = W_1(\nu, \nu_j) \leq C$ and $W_1(\nu, \nu_j^1) = W_1(\nu, \mu) > 2C$, there exists $\bar{t} \in [0,1]$ such that $W_1(\nu, \nu_j^{\bar{t}}) = 2C$. In particular $W_1(\nu_j^{\bar{t}}, \mu) \geq \xi$ and $W_1(\nu_j, \nu_j^{\bar{t}}) \geq W_1(\nu, \nu_j^{\bar{t}}) - W_1(\nu_j, \nu) = 2C - W_1(\nu_j, \nu) \geq W_1(\nu_j, \nu)$ so that

$$W_1(\nu_j, \mu) = W_1(\nu_j, \nu_j^{\bar{t}}) + W_1(\nu_j^{\bar{t}}, \mu) \geq W_1(\nu_j, \nu) + \xi. \tag{21}$$

Putting together (20) and (21) yields

$$\begin{aligned}
\sum_{j=1}^N \lambda_j W_1(\mu_j, \mu) &\geq \sum_{j \in I} \lambda_j (W_1(\mu_j, \nu) - (\xi + 2C)) + \sum_{j \notin I} \lambda_j (W_1(\mu_j, \nu) + \xi) \\
&= \sum_{j=1}^N \lambda_j W_1(\mu_j, \nu) + \xi(\sum_{j \notin I} \lambda_j - \sum_{j \in I} \lambda_j) - 2C \sum_{j \in I} \lambda_j \\
&\geq \sum_{j=1}^N \lambda_j W_1(\mu_j, \nu) + \xi(1 - 2\delta) - 2C\delta \\
&> \sum_{j=1}^N \lambda_j W_1(\mu_j, \nu),
\end{aligned}$$

which contradicts μ being a Wasserstein median for the corrupted collection $\boldsymbol{\mu}$. We thus have

$$\xi \leq \frac{2C\delta}{1-\delta} \text{ and } W_1(\nu, \mu) \leq \xi + 2C \leq \frac{2C\delta}{1-2\delta} + 2C, \tag{22}$$

and $b(\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})) > \delta$, yielding $b(\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})) \geq B$. To obtain the exact value of the breakdown point, take now $J \subset \{1, \dots, N\}$ with $\sum_{j \in J} \lambda_j \geq \frac{1}{2}$ and consider a sequence $(x_n)_{n \in \mathbb{N}}$ in $\mathcal{X}$ such that $d(x_n, x_0) \to +\infty$ as $n \to \infty$. Then, by using the sequence of corrupted collections defined by $\boldsymbol{\mu}^n := (\mu_1^n, \dots, \mu_N^n)$ with $\mu_j^n = \delta_{x_n}$ for $j \in J$ and $\mu_k^n = \nu_k$ for $k \notin J$ we have $\delta_{x_n} \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\mu}^n)$ as we have observed in Example 2.6 and

$$W_1(\delta_{x_n}, \nu) \geq d(x_n, x_0) - W_1(\delta_{x_0}, \nu) \to +\infty \quad \text{as } n \to \infty,$$

implying $b(\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})) \leq B$ and concluding the proof. □

Note that in the case of uniform weights, i.e. with $\boldsymbol{\lambda} := (1/N, \dots, 1/N)$, (18) turns into the classical estimate $b(\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})) = \left\lfloor \frac{N+1}{2} \right\rfloor / N$. Let us finally emphasize that the proof of Theorem 3.4 actually works for Fréchet medians on any geodesic metric space.

4 One dimensional Wasserstein medians

In this section, we study the case of Wasserstein medians on $\mathcal{X} = \mathbb{R}$ with distance d induced by the absolute value. Since the Wasserstein distance of order 1 is equal to the L^1 distance between cumulative or quantile distribution functions, the problem becomes more explicit. This will in particular enable us to find different explicit constructions of Wasserstein medians. In this section for all $\nu \in \mathcal{P}_1(\mathbb{R})$ we denote by F_ν its associated cumulative distribution function (cdf), which is defined by $F_\nu(x) = \nu((-\infty, x])$ for all $x \in \mathbb{R}$. We also denote by $Q_\nu : [0,1] \to \bar{\mathbb{R}}$ its *pseudo-inverse* or quantile distribution function (qdf), which is defined by

$$Q_\nu(t) := \inf\{x \in \mathbb{R} \ : \ F_\nu(x) \geq t\}.$$

Denoting by $\mathcal{L}$ the Lebesgue measure on $[0,1]$, it is well-known that one recovers ν from its qdf Q_ν through $Q_{\nu\#}\mathcal{L} = \nu$, that is Q_ν is the *monotone transport* between $\mathcal{L}$ and ν. We first recall that in one dimension, both maps $\nu \in \mathcal{P}_1(\mathbb{R}) \mapsto F_\nu$ and $\nu \in \mathcal{P}_1(\mathbb{R}) \mapsto Q_\nu$ map isometrically, for the L^1 distance, the Wasserstein space $(\mathcal{P}_1(\mathbb{R}), W_1)$ to the set of cdf's of probabilities in $\mathcal{P}_1(\mathbb{R})$ (i.e. the set of nondecreasing, right-continuous function $F : \mathbb{R} \to [0,1]$ such that $(1-F) \in L^1((0,+\infty))$, $F \in L^1((-\infty, 0))$, $F(+\infty) = 1$ and $F(-\infty) = 0$) and the set of qdf's (i.e. the set of $L^1((0,1), \mathcal{L})$ non-decreasing left-continuous functions) respectively. More precisely, for $(\mu, \nu) \in \mathcal{P}_1(\mathbb{R})^2$ we have the following convenient expressions for the 1-Wasserstein distance between μ and ν (see Theorem 2.9 in [46]):

$$W_1(\mu, \nu) = \int_0^1 |Q_\nu(t) - Q_\mu(t)| \mathrm{d}t = \|Q_\nu - Q_\mu\|_{L^1([0,1])} \tag{23}$$

$$= \int_{\mathbb{R}} |F_\mu(x) - F_\nu(x)| \mathrm{d}x = \|F_\nu - F_\mu\|_{L^1(\mathbb{R})}. \tag{24}$$

This enables us to reformulate the Wasserstein median problem as

$$\min (11) = \min_{\nu \in \mathcal{P}_1(\mathbb{R})} \int_{\mathbb{R}} \sum_{i=1}^N \lambda_i |F_\nu(t) - F_{\nu_i}(t)| \mathrm{d}t \tag{25}$$

$$= \min_{\nu \in \mathcal{P}_1(\mathbb{R})} \int_0^1 \sum_{i=1}^N \lambda_i |Q_\nu(t) - Q_{\nu_i}(t)| \mathrm{d}t, \tag{26}$$

which will be referred as *vertical* (25) and *horizontal* (26) *formulations*. The terminology will become clear in the sequel. Note that, in this way, the problem is equivalent to performing a proper selection of a weighted median of all cumulative or quantile distribution functions, the lower and upper median maps $\mathsf{M}^+_{\boldsymbol{\lambda}}$ and $\mathsf{M}^-_{\boldsymbol{\lambda}}$ defined in (2) in Example 2 and their regularity properties will be particularly useful in this setting.

Proposition 4.1. *Let $\boldsymbol{\lambda} \in \Delta_N$, $\boldsymbol{\nu} := (\nu_1, \ldots, \nu_N) \in \mathcal{P}_1(\mathbb{R})^N$ and $\nu \in \mathcal{P}_1(\mathbb{R})$, then the following statements are equivalent*

1. $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$,
2. $F_\nu(x) \in \mathsf{M}_{\boldsymbol{\lambda}}(F_{\nu_1}(x), \ldots, F_{\nu_N}(x))$ *for all* $x \in \mathbb{R}$,
3. $Q_\nu(t) \in \mathsf{M}_{\boldsymbol{\lambda}}(Q_{\nu_1}(t), \ldots, Q_{\nu_N}(t))$ *for all* $t \in [0,1]$.

In particular, if (4) *holds, there exists a unique Wasserstein median.*

Proof. The fact that 2 implies 1 obviously follows from the definition of $\mathsf{M}_{\boldsymbol{\lambda}}$ and the expression in (24) for the Wasserstein distance. Assume now that $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$, then we should have for a.e. $x \in \mathbb{R}$,

$$\mathsf{M}^-_{\boldsymbol{\lambda}}(F_{\nu_1}(x), \dots, F_{\nu_N}(x)) \le F_\nu(x) \le \mathsf{M}^+_{\boldsymbol{\lambda}}(F_{\nu_1}(x), \dots, F_{\nu_N}(x)). \tag{27}$$

Hence for $x \in \mathbb{R}$, there exists a sequence $\varepsilon_n > 0$, $\varepsilon_n \to 0$ such that the previous inequality holds at $x + \varepsilon_n$, by the right continuity of F_ν, $(F_{\nu_1}(\cdot), \dots, F_{\nu_N}(\cdot))$ at x and the continuity of $\mathsf{M}^\pm_{\boldsymbol{\lambda}}$, we easily get that (27) actually holds at x hence everywhere, proving the equivalence between 1 and 2. The equivalence between 1 and 3 follows the same lines (using left-continuity of qdf's). □

This suggests to define

$$F^-(x) := \mathsf{M}^-_{\boldsymbol{\lambda}}(F_{\nu_1}(x), \dots, F_{\nu_N}(x)),\ F^+(x) := \mathsf{M}^+_{\boldsymbol{\lambda}}(F_{\nu_1}(x), \dots, F_{\nu_N}(x)), \quad \text{for all } x \in \mathbb{R},$$

as well as for $\theta \in [0,1]$,

$$F_\theta(x) := (1-\theta)F^-(x) + \theta F^+(x), \quad \text{for all } x \in \mathbb{R}. \tag{28}$$

Thanks to the properties of $\mathsf{M}^+_{\boldsymbol{\lambda}}$ and $\mathsf{M}^-_{\boldsymbol{\lambda}}$ we saw in Example 2 and the fact that the F_{ν_i}'s are the cdf's of probability measures with finite first moments, F^+ and F^- are also the cdf's of measures with finite first moments and then so is F_θ. Thanks to Proposition 4.1, F_θ is the cdf of ν^θ which belongs to the set of Wasserstein medians $\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$, we call these measures ν^θ a vertical median selections:

Definition 4.2 (**Vertical median selections**). *For every $\theta \in [0,1]$, the measure ν^θ whose cdf is F_θ given by (28) is called the vertical median selection of $\boldsymbol{\nu}$ with weights $\boldsymbol{\lambda}$ and interpolation parameter θ and simply denoted $\mathsf{VMed}_{\boldsymbol{\lambda}}(\theta, \boldsymbol{\nu})$.*

Let us also define

$$Q^-(t) := \mathsf{M}^-_{\boldsymbol{\lambda}}(Q_{\nu_1}(t), \dots, Q_{\nu_N}(t)),\ Q^+(t) := \mathsf{M}^+_{\boldsymbol{\lambda}}(Q_{\nu_1}(x), \dots, Q_{\nu_N}(t)), \quad \text{for all } t \in (0,1),$$

as well as for $\theta \in [0,1]$,

$$Q_\theta(t) := (1-\theta)Q^-(t) + \theta Q^+(t), \quad \text{for all } t \in (0,1). \tag{29}$$

It is easy to see that Q_θ is nondecreasing, left-continuous and in $L^1((0,1), \mathcal{L})$; it is therefore the qdf of a median $\mu^\theta \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ which we call an horizontal median selection:

Definition 4.3 (**Horizontal median selections**). *For every $\theta \in [0,1]$, the measure μ^θ whose qdf is Q_θ given by (28) is called the horizontal median selection of $\boldsymbol{\nu}$ with weights $\boldsymbol{\lambda}$ and interpolation parameter θ and simply denoted $\mathsf{HMed}_{\boldsymbol{\lambda}}(\theta, \boldsymbol{\nu})$.*

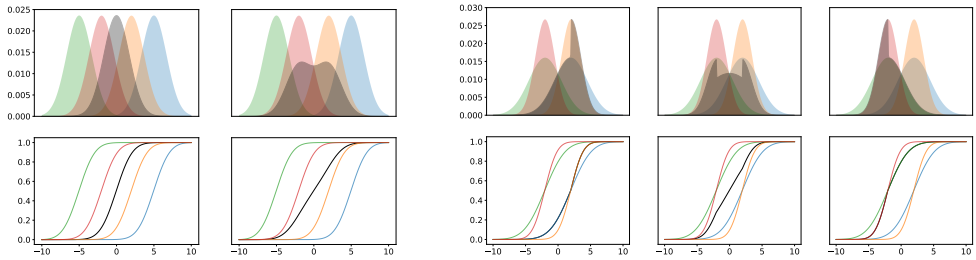
A first nice feature of both vertical and horizontal median selections is that it selects medians in a Lipschitz continuous way with respect to the sample measures:

Lemma 4.4. *Let $\boldsymbol{\lambda} \in \Delta_N$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \mathcal{P}_1(\mathbb{R})^N$, $\widetilde{\boldsymbol{\nu}} = (\widetilde{\nu}_1, \dots, \widetilde{\nu}_N) \in \mathcal{P}_1(\mathbb{R})^N$, $\theta \in [0,1]$ then*

$$W_1(\mathsf{VMed}_{\boldsymbol{\lambda}}(\theta, \boldsymbol{\nu}), \mathsf{VMed}_{\boldsymbol{\lambda}}(\theta, \widetilde{\boldsymbol{\nu}})) \le \sum_{i=1}^N W_1(\nu_i, \widetilde{\nu}_i), \tag{30}$$

and

$$W_1(\mathsf{HMed}_{\boldsymbol{\lambda}}(\theta, \boldsymbol{\nu}), \mathsf{HMed}_{\boldsymbol{\lambda}}(\theta, \widetilde{\boldsymbol{\nu}})) \le \sum_{i=1}^N W_1(\nu_i, \widetilde{\nu}_i). \tag{31}$$



(a) Horizontal selection compared to the vertical selection with $\theta = 1/2$.

(b) Three vertical selections for $\theta = 0,\ 1/2,\ 1$.

Figure 4: Comparison between different Wasserstein median selections. In the second line we displayed the corresponding cumulative distribution functions.

Proof. From (5), we have for every $x \in \mathbb{R}$:

$$\begin{aligned}&|\mathsf{M}^{\pm}_{\boldsymbol{\lambda}}(F_{\nu_1}(x), \dots, F_{\nu_N}(x)) - \mathsf{M}^{\pm}_{\boldsymbol{\lambda}}(F_{\widetilde{\nu}_1}(x), \dots, F_{\widetilde{\nu}_N}(x))| \\ &\leq \max_{i=1,\dots,N} |F_{\nu_i}(x) - F_{\widetilde{\nu}_i}(x)| \leq \sum_{i=1}^{N} |F_{\nu_i}(x) - F_{\widetilde{\nu}_i}(x)|.\end{aligned}$$

Integrating and recalling the cdf expression (24) for the Wasserstein distance, we readily get (30) for $\theta = 0$ and $\theta = 1$, the general case $\theta \in [0,1]$ follows by the triangle inequality. The proof of (31) is similar using the expression of W_1 in terms of quantiles as in (23). □

One may wonder whether some medians inherit properties of the sample measures and in particular whether samples consisting of probabilities with an L^p density with respect to the Lebesgue measure have medians with the same property. As we will shortly see, vertical and horizontal medians will enable us to answer these questions by the positive.

Lemma 4.5. *Let $\boldsymbol{\lambda} \in \Delta_N$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \mathcal{P}_1(\mathbb{R})^N$, $\theta \in [0,1]$, and $\nu^\theta := \mathsf{VMed}_{\boldsymbol{\lambda}}(\theta, \boldsymbol{\nu})$, $\mu^\theta := \mathsf{HMed}_{\boldsymbol{\lambda}}(\theta, \boldsymbol{\nu})$, then*

1. *if $\nu_1, \dots, \nu_N$ are atomless, then so are ν^θ and μ^θ,*
2. *if $\nu_1, \dots, \nu_N$ have connected supports, then so does μ^θ.*

Proof. Recall that for a probability measure, being atomless is equivalent to having a continuous cdf as well as to having a strictly increasing qdf, see, e.g., Proposition 1 in [27]. Let us denote by F_θ the cdf (see (28)) of ν^θ and by Q_θ the qdf of μ^θ (see (29)). If $\nu_1, \dots, \nu_N$ are atomless then F_θ is continuous by continuity of F_{ν_i} so that ν^θ is atomless. On the other hand, (5) entails

$$Q_\theta(t) - Q_\theta(s) \geq \min_{1 \leq i \leq N} (Q_{\nu_i}(t) - Q_{\nu_i}(s)), \text{ for all } (t,s) \in (0,1)^2,$$

so that Q_θ is strictly increasing whenever each Q_{ν_i} is. Let us assume now that $\nu_1, \dots, \nu_N$ have connected supports then each Q_{ν_i} is continuous and so is Q_θ. Thus, μ^θ has a connected support (again by Proposition 1 in [27]). □

Considering absolute continuity of medians, we first discuss the easier case of vertical selections:

Theorem 4.6 (Vertical selections: absolute continuity). *Let $\boldsymbol{\lambda} \in \Delta_N$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \mathcal{P}_1(\mathbb{R})^N$, $\theta \in [0,1]$, and $\nu^\theta := \mathsf{VMed}_{\boldsymbol{\lambda}}(\theta, \boldsymbol{\nu})$. If $\nu_1, \dots, \nu_N$ are all absolutely continuous (with respect*

to the Lebesgue measure on $\mathbb{R}$) with densities $f_1, \dots, f_N \in L^1(\mathbb{R})$ then ν^θ is absolutely continuous with a density $f_{\nu^\theta} \in L^1(\mathbb{R})$ which satisfies

$$\min_{1\le i\le N} f_i \le f_{\nu^\theta} \le \max_{1\le i\le N} f_i, \quad \text{a.e. on } \mathbb{R}. \tag{32}$$

In particular, if, for some $p \in [1,\infty]$, $f_i \in L^p(\mathbb{R})$ for $i = 1, \dots, N$, then $f_{\nu^\theta} \in L^p(\mathbb{R})$ and

$$\| \min_{1\le i\le N} f_i\|_{L^p(\mathbb{R})} \le \|f_{\nu^\theta}\|_{L^p(\mathbb{R})} \le \| \max_{1\le i\le N} f_i\|_{L^p(\mathbb{R})} \le \sum_{i=1}^N \|f_i\|_{L^p(\mathbb{R})}. \tag{33}$$

Proof. Let $x \in \mathbb{R}$ and $h \ge 0$, it follows from (5) and the definition of the cdf F_θ that

$$0 \le F_\theta(x+h) - F_\theta(x) \le \max_{1\le i\le N}\{F_{\nu_i}(x+h) - F_{\nu_i}(x)\} = \max_{1\le i\le N} \int_x^{x+h} f_i \le \int_x^{x+h} \max_{1\le i\le N} f_i,$$

which yields absolute continuity of F_θ, i.e. ν^θ is absolutely continuous with respect to Lebesgue's measure, and the upper bound in (32). In a similar fashion,

$$F_\theta(x+h) - F_\theta(x) \ge \int_x^{x+h} \min_{1\le i\le N} f_i,$$

which shows shows the lower bound in (32) and concludes the proof. □

In particular, in dimension one, vertical medians automatically select medians which inherit integrability properties of the sample measures, with simple explicit pointwise bounds.

Let us now turn our attention to the case horizontal selections which is slightly more involved.

Theorem 4.7 (**Horizontal selections: absolute continuity**). *Let $\boldsymbol{\lambda} \in \Delta_N$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \mathcal{P}_1(\mathbb{R})^N$, $\theta \in [0,1]$, and $\mu^\theta := \mathsf{HMed}_{\boldsymbol{\lambda}}(\theta, \boldsymbol{\nu})$. If $\nu_1, \dots, \nu_N$ are all absolutely continuous (with respect to the Lebesgue measure on $\mathbb{R}$) with densities $f_1, \dots, f_N \in L^1(\mathbb{R})$ then:*

- *μ^0 and μ^1 are absolutely continuous with densities f_{μ^0}, f_{μ^1} which satisfy*

$$\min_{1\le i\le N} f_i \le \min(f_{\mu^0}, f_{\mu^1}) \le \max(f_{\mu^0}, f_{\mu^1}) \le \max_{1\le i\le N} f_i, \quad \text{a.e. on } \mathbb{R}, \tag{34}$$

- *for every $\theta \in [0,1]$, μ^θ is absolutely continuous, we denote its density f_{μ^θ},*
- *if, for some $p \in [1,\infty]$, $f_i \in L^p(\mathbb{R})$ for $i = 1, \dots, N$, then $f_{\mu^\theta} \in L^p(\mathbb{R})$ and*

$$\|f_{\mu^\theta}\|_{L^p(\mathbb{R})} \le \| \max_{1\le i\le N} f_i\|_{L^p(\mathbb{R})} \le \sum_{i=1}^N \|f_i\|_{L^p(\mathbb{R})}. \tag{35}$$

Proof. We shall proceed in three steps.

Step 1: Let us show (34), under the extra assumption that each f_i satisfies

$$f_i \in L^\infty(\mathbb{R}), \ \frac{1}{f_i} \in L^\infty_{\text{loc}}(\mathbb{R}). \tag{36}$$

Recall that by construction

$$\mu^0 := Q^-_\# \mathcal{L}, \ \mu^1 := Q^+_\# \mathcal{L} \text{ with } Q^\pm := \mathsf{M}^\pm_{\boldsymbol{\lambda}}(Q_{\nu_1}, \dots, Q_{\nu_N}),$$

and (36) ensures that each F_{ν_i} is Lipschitz (with Lipschitz constant $\|f_i\|_{L^\infty(\mathbb{R})}$) with inverse Q_{ν_i}, which is locally Lipschitz on $(0,1)$, with Q'_{ν_i} satisfying

$$f_i(Q_{\nu_i})Q'_{\nu_i} = 1, \text{ hence } Q'_{\nu_i} \geq \frac{1}{M} \text{ with } M := \max_j \|f_j\|_{L^\infty(\mathbb{R})} \text{ a.e. on } (0,1). \tag{37}$$

Hence, $Q^\pm$ are locally Lipschitz on $(0,1)$, and it follows from (5) that for $0 < t < s < 1$, one has

$$Q^\pm(s) - Q^\pm(t) \geq \frac{1}{M}(s-t),$$

which implies that $Q^- = Q_{\mu^0}$ and $Q^+ = Q_{\mu^1}$ have M-Lipschitz inverses which are the cdf's F_{μ^0} and F_{μ^1}. Thus, μ^0 and μ^1 are absolutely continuous with bounded positive densities f_{μ^0}, f_{μ^1}, and

$$f_{\mu^0}(Q_{\mu^0})Q'_{\mu^0} = 1,\ f_{\mu^1}(Q_{\mu^1})Q'_{\mu^1} = 1 \text{ a.e. on } (0,1). \tag{38}$$

Using (6), we also have for $0 < t < s < 1$ with $|t-s|$ small enough

$$\max_{i\in I_-(t)}(Q_{\nu_i}(s) - Q_{\nu_i}(t)) \geq Q_{\mu^0}(s) - Q_{\mu^0}(t) \geq \min_{i\in I_-(t)}(Q_{\nu_i}(s) - Q_{\nu_i}(t)),$$

where $I_-(t) := \{i \ : \ Q^-(t) = Q_{\nu_i}(t)\}$. If we choose t a point where all qdf's Q_{μ^0}, Q_{ν_i} are differentiable and the change of variable formulas (37) and (38) hold dividing the previous inequality by $(s-t)$ and letting $s \to t^+$ yields

$$\begin{aligned}\max_{i\in I_-(t)} Q'_{\nu_i}(t) = \max_{i\in I_-(t)} \frac{1}{f_i(Q_{\mu^0}(t))} \geq Q'_{\mu^0}(t) &= \frac{1}{f_{\mu^0}(Q_{\mu^0}(t))} \\ \geq \min_{i\in I_-(t)} Q'_{\nu_i}(t) &= \min_{i\in I_-(t)} \frac{1}{f_i(Q_{\mu^0}(t))},\end{aligned}$$

so that

$$\min_{1\leq i\leq N} f_i(Q_{\mu^0}(t)) \leq f_{\mu^0}(Q_{\mu^0}(t)) \leq \max_{1\leq i\leq N} f_i(Q_{\mu^0}(t)) \text{ for a.e. } t \in (0,1).$$

But since $\mu^0 = Q_{\mu^0}{}_\#\mathcal{L}$ has a positive and bounded density, it has the same null sets as $\mathcal{L}$ hence the previous inequality can be simply reformulated as

$$\min_{1\leq i\leq N} f_i \leq f_{\mu^0} \leq \max_{1\leq i\leq N} f_i \text{ a.e. on } \mathbb{R}.$$

The fact that f_{μ^1} obeys the same inequality can be proved in a similar way using (6) for $\mathsf{M}^+_{\boldsymbol{\lambda}}$ instead of $\mathsf{M}^-_{\boldsymbol{\lambda}}$, we thus have shown (34) under (36). Note also that the L^p bound (35) follows for $\theta \in \{0,1\}$.

Step 2: Again assuming (36), let us show absolute continuity of μ^θ and the L^p bound (35) for $\theta \in (0,1)$. We shall proceed by a displacement convexity argument which is reminiscent of McCann's seminal work [39]. Let us recall that F_{μ^0} is Lipschitz with locally Lipschitz inverse Q^- so that $F_{\mu^0}{}_\#\mu^0 = \mathcal{L}$ and

$$\mu^\theta = ((1-\theta)Q^- + \theta Q^+)_\#\mathcal{L} = ((1-\theta)Q^- + \theta Q^+)_\#(F_{\mu^0}{}_\#\mu^0) = ((1-\theta)\mathrm{id} + \theta T)_\#\mu^0,$$

where $T := Q^+ \circ F_{\mu^0}$ is the monotone transport from μ^0 to μ^1 so that μ^θ is the *displacement interpolation* between μ^0 and μ^1 as defined by McCann in [39] (in the more general and involved multi-dimensional setting). Since the (locally Lipschitz) map $(1-\theta)\mathrm{id} + \theta T$ has a Lipschitz inverse and μ^0 is absolutely continuous, μ^θ is absolutely continuous, we then denote by f_{μ^θ} its density.

Recalling that the qdf of μ^θ, $Q_\theta = (1-\theta)Q^- + Q^+$ is locally Lipschitz, differentiable with a strictly positive derivative a.e. and we have the change of variable formula:

$$f_{\mu^\theta}(Q_\theta)Q'_\theta = 1 \text{ a.e.}$$

Let $V : \mathbb{R} \to \mathbb{R}$ with $V(0) = 0$ be convex, then the function $\alpha > 0 \mapsto \Phi(\alpha) := \alpha V(\alpha^{-1})$ is convex as well and then we have

$$\begin{aligned}\int_{\mathbb{R}} V(f_{\mu^\theta}(x))\mathrm{d}x &= \int_0^1 V\left(\frac{1}{{Q_\theta}'(t)}\right){Q_\theta}'(t)\mathrm{d}t = \int_0^1 \Phi((1-\theta)Q'_{\mu^0}(t) + \theta Q'_{\mu^1}(t)))\mathrm{d}t \\ &\leq (1-\theta)\int_0^1 \Phi(Q'_{\mu^0}(t))\mathrm{d}t + \theta\int_0^1 \Phi(Q'_{\mu^1}(t))\mathrm{d}t \\ &= (1-\theta)\int_{\mathbb{R}} V(f_{\mu^0}(x))\mathrm{d}x + \theta\int_{\mathbb{R}} V(f_{\mu^1}(x))\mathrm{d}x.\end{aligned}$$

Taking $V(\alpha) = |\alpha|^p$, recalling (34) we in particular get

$$\int_{\mathbb{R}} (f_{\mu^\theta}(x))^p\mathrm{d}x \leq (1-\theta)\int_{\mathbb{R}} (f_{\mu^0}(x))^p\mathrm{d}x + \theta\int_{\mathbb{R}} (f_{\mu^1}(x))^p\mathrm{d}x \leq \|\max_{1\leq i\leq N} f_i\|^p_{L^p(\mathbb{R})},$$

which gives (35).

Step 3: general case by Lemma 4.4. To get rid of the extra assumption (36), let g be the density of a standard Gaussian measure and for $\varepsilon > 0$ set

$$f_i^\varepsilon := \frac{\min((1-\varepsilon)f_i + \varepsilon g, \varepsilon^{-1})}{\int_{\mathbb{R}} \min((1-\varepsilon)f_i + \varepsilon g, \varepsilon^{-1})}.$$

Applying the previous steps to $\mu^\theta_\varepsilon := \mathsf{HMed}_{\boldsymbol{\lambda}}(\theta, f_1^\varepsilon, \ldots, f_N^\varepsilon)$, we get

$$\min_{1\leq i\leq N} f_i^\varepsilon \leq \min(f_{\mu^0_\varepsilon}, f_{\mu^1_\varepsilon}) \leq \max(f_{\mu^0_\varepsilon}, f_{\mu^1_\varepsilon}) \leq \max_{1\leq i\leq N} f_i^\varepsilon,$$

and for every $\theta \in [0,1]$,

$$\|f_{\mu^\theta_\varepsilon}\|_{L^p(\mathbb{R})} \leq \|\max_{1\leq i\leq N} f_i^\varepsilon\|_{L^p(\mathbb{R})}.$$

Since f_i^ε converges to f_i in L^p and μ^θ_ε converges to μ^θ in Wasserstein distance thanks to Lemma 4.4, we can pass to the limit $\varepsilon \to 0^+$ in these bounds, obtaining (34) and (35). □

Remark 4.8. *For Wasserstein barycenters (in any dimension), the fact that one sample measure with positive weight is L^p implies that the barycenter is L^p as well (see [1]). For Wasserstein medians in one dimension, we really need all sample measures to be L^p to find an L^p median. To see this, recall that the median of $\nu_1 := \delta_x$ with weight $2/3$ and any probability $\nu_2 \in \mathcal{P}_1(\mathbb{R})$ (with the smoothest density one can think of) with weight $1/3$ is δ_x. Note also that due to the fact that $\mathsf{M}^\pm_{\boldsymbol{\lambda}}$ are Lipschitz but nonsmooth, vertical and horizontal median selections of sample measures with smooth (or Sobolev) densities do not have a continuous density in general.*

5 Multi-marginal and dual formulations

5.1 Multi-marginal formulation

Given the proper metric space $(\mathcal{X}, d)$, $\boldsymbol{\lambda} \in \Delta_N$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N) \in \mathcal{P}_1(\mathcal{X})^N$, the Wasserstein median problem (11) is, like the Wasserstein barycenter problem, a special instance of the *matching*

for teams problem [16] and, as such, admits linear reformulations which take the form of multi-marginal optimal transport problems. Let us now recall this reformulation in the Wasserstein median context. For $\boldsymbol{x} := (x, x_1, \dots, x_N) \in \mathcal{X}^{N+1}$, let us define:

$$f_{\boldsymbol{\lambda}}(x, x_1, \dots, x_N) := \sum_{i=1}^{N} \lambda_i d(x_i, x),\ c_{\boldsymbol{\lambda}}(x_1, \dots, x_N) := \min_{y \in \mathcal{X}} f_{\boldsymbol{\lambda}}(y, x_1, \dots, x_N), \tag{39}$$

and the projections:

$$\pi_0(\boldsymbol{x}) = x,\ \pi_j(\boldsymbol{x}) = x_j,\ 1 \le j \le N,\ \pi_{0,j}(\boldsymbol{x}) = (x, x_j),\ \pi_{1,\dots,N}(\boldsymbol{x}) = (x_1, \dots, x_N).$$

We denote by $\Pi(\nu_1, \dots, \nu_N)$ the set of Borel probability measures on $\mathcal{X}^N$ having ν_i as i-th marginal and the linear multi-marginal problems

$$\inf \left\{ \int_{\mathcal{X}^{N+1}} f_{\boldsymbol{\lambda}} \mathrm{d}\theta \ :\ \theta \in \mathcal{P}_1(X^{N+1}),\ \pi_{1,\dots,N\#}\theta \in \Pi(\nu_1, \dots, \nu_N) \right\}, \tag{40}$$

and

$$\inf_{\gamma \in \Pi(\nu_1, \dots, \nu_N)} \int_{\mathcal{X}^N} c_{\boldsymbol{\lambda}} \mathrm{d}\gamma. \tag{41}$$

Since $(\mathcal{X}, d)$ is Polish, it follows from the disintegration theorem (see paragraph 5.3 in [4]) that if θ is admissible for (40) it can be disintegrated with respect to its marginal $\gamma := \pi_{1,\dots,N\#}\theta$ as

$$\theta = \theta^{x_1,\dots,x_N} \otimes \gamma,$$

for a Borel family of conditional probability measures $\theta^{x_1,\dots,x_N}$ on $\mathcal{X}$. For fixed marginal $\gamma := \pi_{1,\dots,N\#}\theta$, minimizing with respect to the conditional probability $\theta^{x_1,\dots,x_N}$ the integral of $f_{\boldsymbol{\lambda}}$ obviously amounts to choose it supported on $\mathsf{M}_{\boldsymbol{\lambda}}(x_1, \dots, x_N)$ so that it is easy to see that (40) and (41) are equivalent in the sense that they have the same value and that θ solves (40) if and only if $\gamma := \pi_{1,\dots,N\#}\theta$ solves (41) and θ is supported by the set of $(x, x_1, \dots, x_N)$ such that $x \in \mathsf{M}_{\boldsymbol{\lambda}}(x_1, \dots, x_N)$. The fact that $\Pi(\nu_1, \dots, \nu_N)$ is tight and the properness of $(\mathcal{X}, d)$ ensure that the infimum in both (40) and (41) is attained. The connection with the Wasserstein median problem (11) and its solutions $\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ is summarized by:

Theorem 5.1. *The following hold:*

1. $\min$ (11) $=$ $\min$ (40) $=$ $\min$ (41).
2. *If $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ i.e. ν solves (11), then, there exists θ solving (40) such that $\nu = \pi_{0\#}\theta$ and, conversely, if θ solves (40), then $\pi_{0\#}\theta \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$.*
3. *If θ solves (40) and $\nu = \pi_{0\#}\theta$, then for every j such that $\lambda_j > 0$, $\gamma_j := \pi_{0,j\#}\theta$ is an optimal transport plan between ν and ν_j.*
4. *If $\mathsf{m}_\lambda : \mathcal{X}^N \to \mathcal{X}$ is a Borel selection of $\mathsf{M}_{\boldsymbol{\lambda}}$ and γ solves (41), then $(\mathsf{m}_{\boldsymbol{\lambda}})_\#\gamma \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$.*
5. *$\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ if and only if there exists γ solving (41) and a Borel family of probability measures $\theta^{x_1,\dots,x_N}$ such that $\theta^{x_1,\dots,x_N}$ is supported on $\mathsf{M}_{\boldsymbol{\lambda}}(x_1, \dots, x_N)$ for γ-a.e. $(x_1, \dots, x_N)$ and $\nu = \pi_{0\#}(\theta^{x_1,\dots,x_N} \otimes \gamma)$.*

The proof of similar results can be found in [16] and therefore omitted. Even though [16] consider a compact setting (with general costs), the same proof easily adapts to the present setting of a proper metric space with the distance as cost. Note that, as in [16], one can deduce from a Wasserstein median $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ a solution of (40) with first marginal ν as follows: let γ_i be

an optimal plan between ν and ν_i disintegrated with respect to ν as $\gamma_i = \nu \otimes \gamma_i^x$ and define θ by gluing i.e.:

$$\int_{\mathcal{X}} \phi(x, x_1, \dots, x_N) \mathrm{d}\theta(x, x_1, \dots, x_N) := \int_{\mathcal{X}} \Big(\int_{\mathcal{X}^N} \phi(x, x_1, \dots, x_N) \mathrm{d}\gamma_1^x(x_1) \dots \mathrm{d}\gamma_N^x(x_N) \Big) \mathrm{d}\nu(x)$$

for every $\phi \in C_b(\mathcal{X}^{N+1})$. Then, θ solves (40) and by construction $\pi_{0\#}\theta = \nu$. Note that pushing forward by a selection of $\mathsf{M}_{\boldsymbol{\lambda}}$ a solution of the multi-marginal (41) as in 3 above corresponds to special medians for which, using the notation of 4, $\theta^{x_1,\dots,x_N} = \delta_{\mathsf{m}_\lambda(x_1,\dots,x_N)}$ is a Dirac mass. Since $\mathsf{M}_{\boldsymbol{\lambda}}$ is in general not single-valued, not all medians are of this form. Consider for instance $\mathcal{X} = [-1, 1]$ equipped with the usual Euclidean distance and let $\nu_1 = \delta_{-1/2}$ and $\nu_2 = \delta_{1/2}$ with uniform weights. Then $\mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \nu_2)$ is the set of all probability measures supported on $[-1/2, 1/2]$ whereas 3 only selects Dirac masses.

Let us now give an application of Theorem 5.1:

Corollary 5.2 (**Moment bounds**). *Let $\mathcal{X} = \mathbb{R}^d$ be equipped with the Euclidean distance. If all the sample measures are supported on a closed convex subset $\mathcal{K} \subset \mathbb{R}^d$, then every Wasserstein median $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ is supported on $\mathcal{K}$ as well. Moreover, if $V : \mathbb{R}^d \to \mathbb{R}_+$ is quasiconvex (i.e. $\{V \le t\}$ is convex for every $t \ge 0$) then*

$$\int_{\mathbb{R}^d} V \mathrm{d}\nu \le \sum_{i=1}^N \int_{\mathbb{R}^d} V \mathrm{d}\nu_i.$$

In particular, for any $p \in (0, +\infty)$ we have the following bound on the p-moments of ν:

$$\int_{\mathbb{R}^d} |x|^p \mathrm{d}\nu \le \sum_{i=1}^N \int_{\mathbb{R}^d} |x|^p \mathrm{d}\nu_i. \tag{42}$$

Proof. We know from point 4 of Theorem 5.1, that there exists $\gamma \in \Pi(\nu_1, \dots \nu_N)$ and a family of probability $\theta^{x_1,\dots,x_N}$ supported by $\mathsf{M}_{\boldsymbol{\lambda}}(x_1, \dots, x_N)$ such that for every continuous and bounded (or more generally Borel) function $f : \mathbb{R}^d \to \mathbb{R}$ one has

$$\int_{\mathbb{R}^d} f(x) \mathrm{d}\nu(x) = \int_{(\mathbb{R}^d)^N} \Big(\int_{\mathbb{R}^d} f(x) \mathrm{d}\theta^{x_1,\dots,x_N}(x) \Big) \mathrm{d}\gamma(x_1, \dots, x_N),$$

but since $\theta^{x_1,\dots,x_N}$ is supported by $\mathsf{M}_{\boldsymbol{\lambda}}(x_1, \dots, x_N) \subset \mathrm{co}\{x_1, \dots, x_N\}$ (as we have seen in Example 2), if all the ν_i's are supported by the closed convex set $\mathcal{K}$ then so is $\theta^{x_1,\dots,x_N}$ for γ-a.e. $(x_1, \dots, x_N)$ and then $\nu(\mathcal{K}) = 1$. Likewise, if V is nonnegative and quasiconvex, then for $\theta^{x_1,\dots,x_N}$ a.e. x we have

$$V(x) \le \max_{1 \le i \le N} V(x_i) \le \sum_{i=1}^N V(x_i),$$

and integrating this inequality with respect to $\theta^{x_1,\dots,x_N}$ first and then with respect to γ in $\Pi(\nu_1, \dots, \nu_N)$ we obtain the announced moment bounds. □

A counterexample to linear L^∞ density bounds in several dimensions. We have seen in Theorems 4.6 and 4.7 that when $\mathcal{X} = \mathbb{R}$, and the sample measures have densities uniformly bounded by some M, vertical and horizontal median selections enable to find Wasserstein medians with a density which is bounded by the same bound M. In other words, in dimension one, it is possible to have a linear control on the L^∞ norm of some well-chosen Wasserstein median in terms of L^∞ bounds of the sample measures. The situation seems to be more intricate in higher dimensions. The following example shows that a linear L^∞-bound cannot hold in two dimensions.

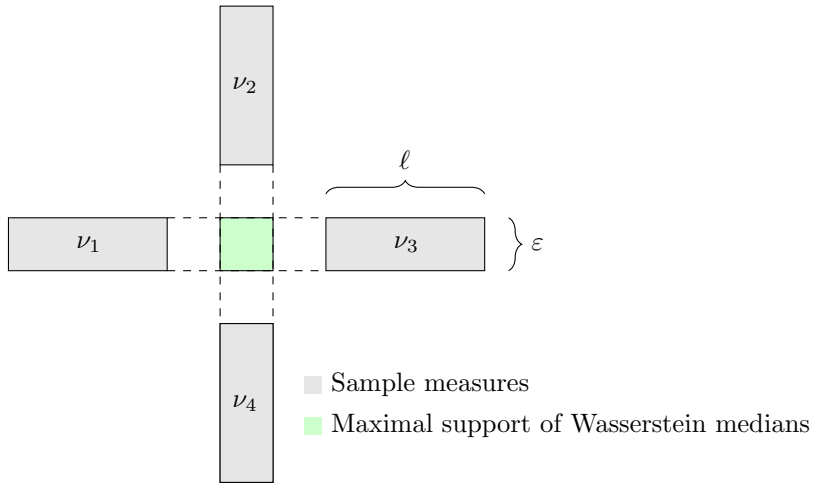


Figure 5: Counterexample to linear L^∞ bounds in dimension two. The support of the four given uniformly distributed measures are indicated in gray. The support of any Wasserstein median is contained in the green area. Confer Example 5.3 for further details.

Example 5.3. *For $0 < \varepsilon < 1$ let ν_1 be a uniform measure supported on the rectangle $[-1-\ell, -1] \times [-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]$ and let ν_2, ν_3 and ν_4 be obtained by successive rotations by $90°$ of ν_1 as in Figure 5. Consider uniform weights $\lambda_i = \frac{1}{4}$, for $i = 1, \dots, 4$, and let $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \dots, \nu_4)$. We know from Theorem 5.1 that one can write $\nu := \pi_{0\#}\theta$ where $\pi_{i\#}\theta = \nu_i$ for $i = 1, \dots, 4$ and x is a geometric median of $(x_1, \dots, x_4)$ for θ-a.e. $(x, x_1, \dots, x_4)$. Now note that, with this construction, four points $x_i \in \operatorname{spt} \nu_i$ always form a convex quadrilateral, and as shown in Theorem 1 in [43], their unique median is the intersection of the two segments $[x_1, x_3]$ and $[x_2, x_4]$. In particular such geometric medians belong to the square $[-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]^2$, which therefore supports any $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \dots, \nu_4)$. This shows that the L^∞ norm of ν is at least ε^{-2}: it cannot be bounded from above uniformly in ε by a multiple of $\max_{i=1,\dots,4} \|\nu_i\|_{L^\infty} = \ell^{-1}\varepsilon^{-1}$.*

5.2 Dual Formulation

To introduce a dual formulation à la Kantorovich of (11), we fix a point $x_0 \in \mathcal{X}$ and define the spaces

$$Y_0 := \left\{ f \in C(\mathcal{X}) \ : \ \lim_{d(x,x_0)\to\infty} \frac{f(x)}{1 + d(x, x_0)} = 0 \right\}, \quad Y_b := \left\{ f \in C(\mathcal{X}) \ : \ \sup_{x\in\mathcal{X}} \frac{|f(x)|}{1 + d(x, x_0)} < \infty \right\}.$$

Note that these spaces are independent of the choice of x_0 and that the dual of Y_0 may be identified with the space of signed measures with finite first moment

$$(Y_0)^* = \{\mu \in \mathcal{M}(\mathcal{X}) \ : \ (1 + d(x, x_0))\mu \in \mathcal{M}(\mathcal{X})\}.$$

We will also assume here without loss of generality that all the weights λ_i are strictly positive in the Wasserstein median problem (11) and define for $\lambda > 0$:

$$\mathsf{Lip}_\lambda(\mathcal{X}) := \{v : \mathcal{X} \to \mathbb{R}, \ |v(x) - v(y)| \le \lambda d(x, y), \ \text{ for all } (x, y) \in \mathcal{X}^2\}.$$

Setting $c_i := \lambda_i d$, the c_i-transform of a function $u : \mathcal{X} \to \mathbb{R}$, denoted u^{c_i}, is by definition given by

$$u^{c_i}(x) := \inf_{y\in\mathcal{X}} \{\lambda_i d(x, y) - u(y)\}, \text{ for all } x \in \mathcal{X},$$

note that, by the triangle inequality u^{c_i} is either everywhere $-\infty$ or a λ_i-Lipschitz function. It is also a classical fact (see, e.g. Proposition 3.1 in [46]) that $u \in \mathsf{Lip}_{\lambda_i}(\mathcal{X})$ if and only if $u^{c_i} = -u$.

Following [1], let us now consider the concave maximization problem

$$\sup \left\{ \sum_{i=1}^{N} \int_{\mathcal{X}} u_i^{c_i} d\nu_i \ : \ u_i \in Y_0, \ \sum_{i=1}^{N} u_i = 0 \right\}, \tag{43}$$

and its relaxed version

$$\sup \ \left\{ \sum_{i=1}^{N} \int_{\mathcal{X}} u_i^{c_i} \mathrm{d}\nu_i \ : \ u_i \in Y_b, \ \sum_{i=1}^{N} u_i = 0 \right\}. \tag{44}$$

By definition of the c_i-transform, it is easy to check the weak duality relation

$$\min (11) \geq \sup (44) \geq \sup (43).$$

Using convex duality by proceeding exactly as in the proof of Propositions 2.2 and 2.3 in [1] for the Wasserstein barycenter case, one can show that (11) is the dual of (43) and that strong duality holds i.e.: $\min (11) = \sup (44) = \sup (43)$. It will be convenient in the sequel to consider yet another formulation of (44):

$$\sup \left\{ \sum_{i=1}^{N} \int_{\mathcal{X}} u_i \mathrm{d}\nu_i \ : \ \ u_i \in \mathsf{Lip}_{\lambda_i}(\mathcal{X}), \ i = 1, \ldots, N, \ \sum_{i=1}^{N} u_i \leq 0 \right\}. \tag{45}$$

Proposition 5.4 (**Lipschitz formulation of the dual problem**). *Let $(\nu_1, \ldots, \nu_N) \in \mathcal{P}_1(\mathcal{X})^N$ and $\boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_N) \in \Delta_N$ with each λ_i strictly positive. Then we have*

$$\min (11) = \sup (44) = \max (45), \tag{46}$$

where we have written max (45) to emphasize that the supremum in (45) is attained.

Proof. Recall that $\min (11) = \sup (44)$.

Step 1: $\sup (44) \geq \sup (45)$. Let $(u_1, \ldots, u_N)$ be admissible for (45), take $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_N)$, with $\psi_i = -u_i$ for all $i = 1, \ldots, N-1$ and $\psi_N = u_1 + \cdots + u_N$. Since Lipschitz functions belong to Y_b, $\boldsymbol{\psi}$ is admissible for (44) and we have:

$$\sum_{i=1}^{N} \int_{\mathcal{X}} u_i \mathrm{d}\nu_i = \sum_{i=1}^{N-1} \int_{\mathcal{X}} (-\psi_i) \mathrm{d}\nu_i + \int_{\mathcal{X}} u_N \mathrm{d}\nu_N.$$

For $i = 1, \ldots, N-1$, since $\psi_i \in \mathsf{Lip}_{\lambda_i}(\mathcal{X})$, we have $-\psi_i = \psi_i^{c_i}$. Moreover, $\psi_N = u_1 + \cdots + u_{N-1} \leq -u_N$, hence $\psi_N^{c_N} \geq u_N$, yielding

$$\sum_{i=1}^{N} \int_{\mathcal{X}} u_i \mathrm{d}\nu_i \leq \sum_{i=1}^{N} \int_{\mathcal{X}} \psi_i^{c_i} \mathrm{d}\nu_i \leq \sup (44).$$

Step 2: $\sup (45) \geq \sup (44)$. Let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_N)$ be admissible for (44). Consider $\boldsymbol{u} = (u_1, \ldots, u_N) = (\psi_1^{c_1}, \ldots, \psi_N^{c_N})$. By construction, each u_i is λ_i-Lipschitz and to see that $\boldsymbol{u}$ is admissible for (45) we observe that for every $x \in \mathcal{X}$:

$$\sum_{i=1}^{N} u_i(x) = \sum_{i=1}^{N} \psi_i^{c_i}(x) = \sum_{i=1}^{N} \inf_y \{\lambda_i d(x, y) - \psi_i(y)\} \leq -\sum_{i=1}^{N} \psi_i(x) = 0,$$

and, then,

$$\sum_{i=1}^{N}\int_{\mathcal{X}}\psi_i^{c_i}\mathrm{d}\nu_i=\sum_{i=1}^{N}\int_{\mathcal{X}}u_i\mathrm{d}\nu_i\leq\sup\,(45).$$

Step 3: the supremum is attained in sup (45). We note that both constraints and the objective function in (45) are unchanged when one replaces u_i by $u_i+\alpha_i$ where the α_i's are constant that sum to 0, we may therefore restrict the maximization in sup (45) to the smaller admissible set of potentials $(u_1,\dots,u_N)$ such that

$$u_i\in\mathsf{Lip}_{\lambda_i}(\mathcal{X}),\ \sum_{i=1}^{N}u_i\leq 0,\ \text{and}\ \int_{\mathcal{X}}u_i\mathrm{d}\nu_i=0,\quad\text{for } i=1,\dots,N-1. \tag{47}$$

Since this set contains $(0,\dots,0)$ we can reduce it even further this set by considering only potentials for which the objective is positive:

$$\int_{\mathcal{X}}u_N\mathrm{d}\nu_N\geq 0. \tag{48}$$

If we denote by K the set of potentials that satisfy (47) and (48), we observe that if $(u_1,\dots,u_N)\in K$ then for $i=1,\dots,N-1$ and $x\in\mathcal{X}$, since u_i is λ_i-Lipschitz, one has

$$u_i(x)\leq\int_{\mathcal{X}}u_i\mathrm{d}\nu_i+\lambda_i\int_{\mathcal{X}}d(x,y)\mathrm{d}\nu_i(y)\leq\lambda_i d(x,x_0)+m_i,\ m_i:=\lambda_i\int_{\mathcal{X}}d(x_0,y)\mathrm{d}\nu_i(y).$$

Reasoning in a similar way for $-u_i$, we get bounds with linear growth, namely $|u_i|\leq\lambda_i d(\cdot,x_0)+m_i$ for $i=1,\dots,N-1$. Since $u_N\leq-\sum_{i=1}^{N}u_i$ we get a similar upper bound with linear growth for u_N, and, for a lower bound, we use (48) which, together with the fact that u_N is λ_N-Lipschitz gives

$$u_N\geq-\lambda_N d(\cdot,x_0)-\lambda_N\int_{\mathcal{X}}d(x_0,y)\mathrm{d}\nu_N(y).$$

Let us now take a maximizing sequence in K for (45). The above linear bounds and Ascoli–Arzelà's theorem guarantee that this sequence converges locally uniformly to some $\boldsymbol{u}$, and again by these linear bounds, the fact that $\nu_i\in\mathcal{P}_1(\mathcal{X})$ for all $i=1,\dots,N$, and Lebesgue's dominated convergence theorem, one deduces that $\boldsymbol{u}\in K$ and $\boldsymbol{u}$ actually solves (45). □

We may derive from the primal-dual relations between (11) and (45) a characterization of Wasserstein medians in terms of Kantorovich potentials

Theorem 5.5 (**Optimality conditions for Wasserstein medians**). *Let $\boldsymbol{\nu}=(\nu_1,\dots,\nu_N)\in\mathcal{P}_1(\mathcal{X})^N$, $\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_N)\in\Delta_N$ with $\lambda_i>0$ and let $\nu\in\mathcal{P}_1(\mathcal{X})$. Then $\nu\in\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ if and only if there exist $\psi_1,\dots,\psi_N$ such that*

1. *for $i=1,\dots,N$, $\psi_i\in\mathsf{Lip}_1(\mathcal{X})$ is a Kantorovich potential between ν_i and ν, i.e.*

$$W_1(\nu_i,\nu)=\int_{\mathcal{X}}\psi_i\mathrm{d}\nu_i-\int_{\mathcal{X}}\psi_i\mathrm{d}\nu,$$

2. *there holds*

$$\sum_{i=1}^{N}\lambda_i\psi_i\leq 0\ \text{on}\ \mathcal{X},\ \text{and}\ \sum_{i=1}^{N}\lambda_i\psi_i=0\ \text{on}\ \mathrm{spt}(\nu).$$

Proof. It follows from the duality result of proposition 5.4 that $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ if and only if there exists $(u_1, \dots, u_N)$ admissible for (45) such that

$$\sum_{i=1}^{N} \lambda_i W_1(\nu_i, \nu) = \sum_{i=1}^{N} \int_{\mathcal{X}} u_i \mathrm{d}\nu_i \tag{49}$$

(in which case $(u_1, \dots, u_N)$ automatically solves (45)). Setting $\psi_i = u_i/\lambda_i$ we thus have $\psi_i \in \mathsf{Lip}_1(\mathcal{X})$ and $\sum_{i=1}^{N} \lambda_i \psi_i \leq 0$ on $\mathcal{X}$. By the Kantorovich–Rubinstein duality formula (8), we have

$$W_1(\nu_i, \nu) \geq \int_{\mathcal{X}} \psi_i \mathrm{d}\nu_i - \int_{\mathcal{X}} \psi_i \mathrm{d}\nu. \tag{50}$$

Multiplying by λ_i summing and using the fact that ν is a nonnegative measure and $\sum_{i=1}^{N} \lambda_i \psi_i \leq 0$ thus yields

$$\begin{aligned}\sum_{i=1}^{N} \lambda_i W_1(\nu_i, \nu) &\geq \sum_{i=1}^{N} \lambda_i \int_{\mathcal{X}} \psi_i \mathrm{d}\nu_i - \sum_{i=1}^{N} \lambda_i \int_{\mathcal{X}} \psi_i \mathrm{d}\nu \\ &\geq \sum_{i=1}^{N} \lambda_i \int_{\mathcal{X}} \psi_i \mathrm{d}\nu_i = \sum_{i=1}^{N} \int_{\mathcal{X}} u_i \mathrm{d}\nu_i,\end{aligned}$$

so that (49) holds if and only if each inequality (50) is an equality, i.e. ψ_i is a Kantorovich potential between ν_i and ν and

$$\int_{\mathcal{X}} \Big(\sum_{i=1}^{N} \lambda_i \psi_i \Big) \mathrm{d}\nu = 0,$$

i.e. $\sum_{i=1}^{N} \lambda_i \psi_i = 0$ on $\mathrm{spt}(\nu)$ since each ψ_i is continuous. □

6 Beckmann minimal flow formulation

In this section, we consider the Wasserstein median problem on a convex compact subset Ω of $\mathbb{R}^d$, with non empty interior (which is not really restrictive) equipped with the Euclidean distance. In this setting, we will see that, taking advantage of the so-called Beckmann's minimal flow formulation of Monge's problem, one can derive a system of PDEs that characterize Wasserstein medians. We are given $\boldsymbol{\lambda} \in \Delta_N$ with $\lambda_i > 0$ for all $i = 1, \dots, N$, and $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \mathcal{P}(\Omega)^N$, we know from Corollary 5.2 that any measure in $\mathsf{Med}_{\boldsymbol{\lambda}}(\boldsymbol{\nu})$ is supported on Ω.

6.1 The Beckmann problem

We denote by $\mathcal{M}(\Omega, \mathbb{R}^d)$ the set of vector valued measures on Ω. For such a measure σ, we denote by $|\sigma| \in \mathcal{M}_+(\Omega)$ its total variation measure and recall that one can write $\mathrm{d}\sigma = \hat{\sigma} \mathrm{d}|\sigma|$ for some Borel map $\hat{\sigma}$ such that $|\hat{\sigma}| = 1$, $|\sigma|$-a.e.; for every test-function $\phi \in C(\Omega, \mathbb{R}^d)$, one can therefore write

$$\int_{\Omega} \phi \cdot \ \mathrm{d}\sigma = \int_{\Omega} \phi(x) \cdot \hat{\sigma}(x) \ \mathrm{d}|\sigma|(x).$$

Let us denote by $\mathcal{M}_{\mathrm{div}}(\Omega, \mathbb{R}^d)$ the set of vector valued measures σ whose divergence $\nabla \cdot \sigma$ is a finite measure, where $\nabla \cdot \sigma$ is defined in the sense of distributions. Given $i = 1, \dots, N$ and $\nu \in \mathcal{P}(\Omega)$, a vector-valued measure $\sigma_i \in \mathcal{M}_{\mathrm{div}}(\Omega, \mathbb{R}^d)$ is an admissible flow between ν_i and ν if it solves

$$\nabla \cdot \sigma_i + \nu_i = \nu$$

in the weak sense, i.e.

$$\int_\Omega \nabla\phi \cdot \ \mathrm{d}\sigma_i = \int_\Omega \phi \ \mathrm{d}(\nu_i - \nu), \text{ for all } \phi \in C^1(\Omega).$$

Beckmann's formulation of the optimal transport problem with distance cost between ν_i and ν consists in finding an admissible flow with minimal total variation, it thus reads

$$\inf_{\sigma_i \in \mathcal{M}_{\mathrm{div}}(\Omega,\mathbb{R}^d)} \{|\sigma_i|(\Omega) \ : \ \nabla\cdot\sigma_i + \nu_i = \nu\} \tag{51}$$

where $|\sigma_i|(\Omega)$ denotes the total variation of σ_i. This problem was introduced by Beckmann in the 1950's [7] and its connections with the optimal transport problem $W_1(\nu,\nu_i)$ is well-known, as we shall recall now, referring the reader to [46] and [3] for detailed statements and proofs. First of all, let us recall that the value of (51) coincides with the Wasserstein distance $W_1(\nu,\nu_i)$ so recalling the Kantorovich–Rubinstein formula, we have (and we write min and max on purpose to emphasize the existence of solutions):

$$W_1(\nu_i,\nu) = \min_{\sigma_i \in \mathcal{M}_{\mathrm{div}}(\Omega,\mathbb{R}^d)} \{|\sigma_i|(\Omega) \ : \ \nabla\cdot\sigma_i + \nu_i = \nu\} = \max_{u_i \in \mathsf{Lip}_1(\Omega)} \int_\Omega u_i \ \mathrm{d}(\nu_i - \nu). \tag{52}$$

Following the seminal work of [11, 12], the sharp connection between optimal flows, i.e. solutions of (51) and Kantorovich potentials is captured by the Monge–Kantorovich PDE system which we now recall.

Definition 6.1 (**Monge–Kantorovich PDE**). *A pair $(u_i,\rho_i) \in \mathsf{Lip}_1(\Omega) \times \mathcal{M}_+(\Omega)$ solves the Monge–Kantorovich system between ν_i and ν:*

$$\nabla\cdot(\rho_i \nabla_{\rho_i} u_i) + \nu_i = \nu, \ |\nabla_{\rho_i} u_i| = 1 \ \rho_i\text{-a.e.} \tag{53}$$

if there exists $(u_i^\varepsilon)_{\varepsilon>0} \in C^1(\Omega) \cap \mathsf{Lip}_1(\Omega)$ converging uniformly to u_i as $\varepsilon \to 0$, such that ∇u_i^ε converges in $L^2(\rho_i)$ to some $\hat\sigma_i$ (so that $|\hat\sigma_i| \le 1$) and

$$\nabla\cdot(\rho_i\hat\sigma_i) + \nu_i = \nu, \ |\hat\sigma_i| = 1 \ \rho_i\text{-a.e.}. \tag{54}$$

Assume that $(u_i,\rho_i) \in \mathsf{Lip}_1(\Omega) \times \mathcal{M}_+(\Omega)$ solves the Monge–Kantorovich system between ν_i and ν, and let $(u_i^\varepsilon)_{\varepsilon>0} \in C^1(\Omega) \cap \mathsf{Lip}_1(\Omega)$ converge uniformly to u_i as $\varepsilon \to 0$, and be such that ∇u_i^ε converges in $L^2(\rho_i)$ to some $\hat\sigma_i$ which satisfies (54), then using the fact that $\sigma_i := \rho_i\hat\sigma_i$ is admissible for (51) we deduce from (52) and (54):

$$\begin{aligned} W_1(\nu_i,\nu) \ge \int_\Omega u_i \ \mathrm{d}(\nu_i - \nu) &= \lim_{\varepsilon\to 0} \int_\Omega u_i^\varepsilon \ \mathrm{d}(\nu_i - \nu) = \lim_{\varepsilon\to 0} \int_\Omega \nabla u_i^\varepsilon \cdot \hat\sigma_i \ \mathrm{d}\rho_i \\ &= \int_\Omega |\hat\sigma_i|^2 \ \mathrm{d}\rho_i = \rho_i(\Omega) = |\sigma_i|(\Omega) \ge W_1(\nu_i,\nu) \end{aligned}$$

which proves that u_i is a Kantorovich potential and $\sigma_i := \rho_i\hat\sigma_i$ is an optimal flow:

$$W_1(\nu_i,\nu) = \int_\Omega u_i \ \mathrm{d}(\nu_i - \nu) = |\sigma_i|(\Omega).$$

This also enables one to define unambiguously the $L^2(\rho_i)$-limit of ∇v_i^ε for any *any* approximation[2] of u_i by $C^1(\Omega) \cap \mathsf{Lip}_1(\Omega)$, indeed if $(v_i^\varepsilon)_{\varepsilon>0}$ is a sequence of such approximations, using again (54), we have:

$$\|\nabla v_i^\varepsilon - \hat\sigma_i\|^2_{L^2(\rho_i)} \le 2|\sigma_i|(\Omega) - 2\int_\Omega \nabla v_i^\varepsilon \cdot \hat\sigma_i \mathrm{d}\rho_i = 2W_1(\nu_i,\nu) - 2\int_\Omega v_i^\varepsilon \mathrm{d}(\nu_i - \nu) \to 0 \text{ as } \varepsilon \to 0.$$

[2]Note that such approximations can easily be performed by first extending u_i to a 1-Lipschitz function to the whole of $\mathbb{R}^d$ and then mollifying by convolution this extension.

In other words, in definition 6.1, the direction $\hat{\sigma}_i \in L^2(\rho_i)$ only depends on ρ_i and u_i and not on the approximation of u_i and it is legitimate to set $\nabla_{\rho_i} u_i = \hat{\sigma}_i$ and to call it the tangential gradient of u_i with respect to ρ_i (and justify a posteriori the notation $\nabla_{\rho_i} u_i$). We have seen that solutions of the Monge–Kantorovich system yield optimal flows and optimal potentials, but the converse is easy to check. Indeed, let $u_i \in \mathsf{Lip}_1(\Omega)$ and $\sigma_i \in \mathcal{M}_{\mathrm{div}}(\Omega, \mathbb{R}^d)$ be such that

$$W_1(\nu_i, \nu) = \int_\Omega u_i \, \mathrm{d}(\nu_i - \nu) = |\sigma_i|(\Omega)$$

setting $\rho_i := |\sigma_i|$ and $\hat{\sigma}_i$ such that $|\hat{\sigma}_i| = 1$ ρ_i-a.e. and $\mathrm{d}\sigma_i = \hat{\sigma}_i \mathrm{d}\rho_i$, then (54) holds and if $(u_i^\varepsilon)_{\varepsilon>0}$ is a sequence of $C^1 \cap \mathsf{Lip}_1$ approximations of u_i then

$$\|\nabla u_i^\varepsilon - \hat{\sigma}_i\|^2_{L^2(\rho_i)} \leq 2|\sigma_i|(\Omega) - 2\int_\Omega \nabla u_i^\varepsilon \cdot \hat{\sigma}_i \mathrm{d}\rho_i = 2W_1(\nu_i, \nu) - 2\int_\Omega u_i^\varepsilon \mathrm{d}(\nu_i - \nu) \to 0 \text{ as } \varepsilon \to 0$$

so that (u_i, ρ_i) solves the Monge–Kantorovich system (53) which therefore fully characterizes the primal-dual extremality relations in (52).

Note that if ρ_i is absolutely continuous with respect to the Lebesgue measure $\rho_i \in L^1(\Omega)$, then whenever $u_i \in \mathsf{Lip}_1(\Omega)$, $\rho_i \nabla u_i$ belongs to $L^1(\Omega)$ so $\nabla \cdot (\rho_i \nabla u_i)$ is well defined in the sense of distributions and (53) simplifies to

$$\nabla \cdot (\rho_i \nabla u_i) + \nu_i = \nu, \ |\nabla u_i| = 1 \ \rho_i\text{-a.e.} \tag{55}$$

In Monge–Kantorovich theory, $\rho_i = |\sigma_i|$ where σ_i is an optimal flow is called the transport density and the study of integral estimates for transport densities has been the object of an intensive stream of research [21–23, 25, 45]. In particular, if ν_i is absolutely continuous with respect to the Lebesgue measure (and ν is an arbitrary probability measure) then the solution σ_i of (51) is unique (Theorem 4.14 and Corollary 4.15 in [46]) and absolutely continuous as well (Theorem 4.16 in [46]) so that the transport density ρ_i is in L^1 and the Monge–Kantorovich PDE can be understood as in (55) without using the notion of tangential gradient. Higher integrability results can be found in Theorem 4.20 in [46].

The connection between optimal flows, transport densities and optimal plans, is also well-known, namely given $\gamma_i \in \Pi(\nu, \nu_i)$ optimal i.e. such that $W_1(\nu_i, \nu) = \int_{\Omega\times\Omega} |x - x_i| \mathrm{d}\gamma_i(x, x_i)$, define the vector valued measure σ_{γ_i} by

$$\int_\Omega \phi \cdot \mathrm{d}\sigma_{\gamma_i} = \int_{\Omega\times\Omega} \int_0^1 \phi(x + t(x_i - x)) \cdot (x_i - x) \, \mathrm{d}t \, \mathrm{d}\gamma_i(x, x_i), \text{ for all } \phi \in C(\Omega, \mathbb{R}^d). \tag{56}$$

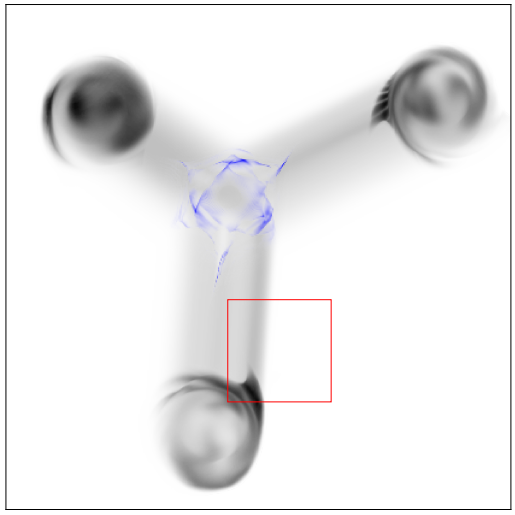
Then, $\nabla \cdot \sigma_{\gamma_i} + \nu_i = \nu$ and σ_{γ_i} is an optimal flow i.e. solves (51), moreover (see Theorem 4.13 in [46]), any σ_i solving (51) is of the form σ_{γ_i} for some optimal plan γ_i. We also refer to in [46] and [3] for more on the subject and in particular connections between optimal flows and the directions of the so-called transport rays.

6.2 A system of PDEs for Wasserstein medians

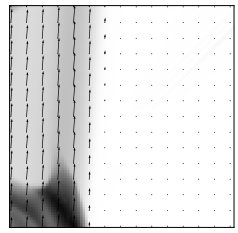
We now rewrite the Wasserstein median problem (11) in terms of a multi-flow minimization:

$$\inf_{(\sigma_1,\ldots,\sigma_N,\nu)\in\mathcal{M}_{\mathrm{div}}(\overline{\Omega},\mathbb{R}^d)^N\times\mathcal{P}(\Omega)} \left\{ \sum_{j=1}^N \lambda_j |\sigma_j|(\Omega) : \nabla \cdot \sigma_j + \nu_j = \nu, \ j = 1, \ldots, N \right\}, \tag{57}$$

and observe that ν solves (11) if and only if there exist $(\sigma_1, \ldots, \sigma_N) \in \mathcal{M}_{\mathrm{div}}(\overline{\Omega}, \mathbb{R}^d)^N$ such that $(\sigma_1, \ldots, \sigma_N, \nu)$ solves (57). Since we have assumed $\lambda_i > 0$ we can perform the change of unknown



(a) Superimposition of Wasserstein median, transport densities, and sample measures.



(b) Zoom of the red square in (a), the arrows represent the transport flow in the corresponding point.

Figure 6: An approximate Wasserstein median (blue) of three sample measures (black) and the three approximately optimal transport densities (in gray) computed via Douglas–Rachford, with step-size $\tau = 10^{-1}$ and relaxation parameters $\theta_k = 1$ for all $k \in \mathbb{N}$, on a 420×420 grid. The figure shows the results after 10000 iteration, with a residual of $7 \cdot 10^{-8}$.

$u_i \to u_i/\lambda_i$ in (45) and rewrite it as

$$\sup\Big\{\sum_{i=1}^{N} \lambda_i \int_\Omega u_i \mathrm{d}\nu_i \ : \ u_i \in \mathsf{Lip}_1(\Omega),\ i = 1, \dots, N,\ \sum_{i=1}^{N} \lambda_i u_i \le 0\Big\}. \tag{58}$$

We may deduce from what we have recalled in the previous paragraph, a characterization of Wasserstein medians as well as optimal flows in (57) and optimal potentials in (58) by a system of PDEs of Monge–Kantorovich type. Note that if a median $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \dots, \nu_N)$ was known, the problem of finding the corresponding optimal flows would be decoupled into N Monge–Kantorovich PDEs in the sense of definition 6.1, but to determine ν, we should take into account the obstacle constraint $\sum_{i=1}^N \lambda_i u_i \le 0$ from (58) and the optimality condition from Theorem 5.5 that requires $\sum_{i=1}^N \lambda_i u_i$ to vanish on the support of ν. All this can be summarized as:

Theorem 6.2 (**A Monge–Kantorovich system of PDEs for medians**). *Let $\nu \in \mathcal{P}(\Omega)$ then $\nu \in \mathsf{Med}_{\boldsymbol{\lambda}}(\nu_1, \dots, \nu_N)$ if and only if there exist $(u_1, \dots, u_N) \in \mathsf{Lip}_1(\Omega)^N$ and $(\rho_1, \dots, \rho_N) \in \mathcal{M}_+(\Omega)^N$ such that, for $i = 1, \dots, N$*

$$\nabla \cdot (\rho_i \nabla_{\rho_i} u_i) + \nu_i = \nu,\ |\nabla_{\rho_i} u_i| = 1 \ \rho_i\text{-a.e.}, \tag{59}$$

coupled with the obstacle conditions

$$\sum_{i=1}^{N} \lambda_i u_i \le 0 \ \ \text{on } \Omega,\ \sum_{i=1}^{N} \lambda_i u_i = 0 \ \ \text{on } \mathrm{spt}(\nu) \tag{60}$$

Moreover in this case $(u_1, \dots, u_N)$ solves (58) and $(\rho_1 \nabla_{\rho_1} u_1, \dots, \rho_N \nabla_{\rho_N} u_N, \nu)$ solves (57).

Remark 6.3 ($d = 1$, **Ω is an interval**). *In dimension* 1*, one can integrate the equation* $\nabla\cdot\sigma_i+\nu_i = \nu$ *and in this case,* (57) *appears as the vertical formulation of the median problem* (25)*. One can therefore interpret* (57) *in higher dimensions as a multidimensional extension of* (25)*.*

Remark 6.4 (**Case of absolutely continuous sample measures**). *If* ν_i *is in* $L^1(\Omega)$*, then the corresponding optimal flow* σ_i *and transport density* ρ_i *are also in* $L^1(\Omega)$ *(even though medians need not be absolutely continuous) and one can replace the tangential gradient* $\nabla_{\rho_i} u_i$ *by* ∇u_i *in the Monge–Kantorovich PDE* (59)*.*

Remark 6.5 (**Connection with the multi-marginal formulation**). *If* θ *solves the multi-marginal problem* (40)*, then we know that* $\nu := \pi_{0\#}\theta$ *is a median and we can recover the corresponding flows as in* (56) *i.e. by defining:*

$$\int_\Omega \phi \cdot \mathrm{d}\sigma_i^\theta = \int_{\Omega^{N+1}} \int_0^1 \phi(x + t(x_i - x)) \cdot (x_i - x)\, \mathrm{d}t\, \mathrm{d}\theta(x, x_1, \dots, x_N), \text{ for all } \phi \in C(\Omega, \mathbb{R}^d),$$

with this construction $(\sigma_1^\theta, \dots, \sigma_N^\theta, \pi_{0\#}\theta)$ *is a solution of* (57)*. In fact, invoking Theorem 4.13 in [46], any solution of* (57) *can be obtained in this way from an optimal multi marginal plan* θ*.*

6.3 Approximation by a system of p-Laplace equations

We shall now see how to approximate a median, as well as dual potentials and Beckmann flows by a single system of p-Laplace equations (with p large as in the seminal work of Evans and Gangbo [28], also see [38] for a similar strategy for a matching problem involving two sample measures). Given $\varepsilon > 0$, we are given an exponent $p_\varepsilon \geq 2d$, and assume these exponents satisfy

$$\lim_{\varepsilon \to 0^+} p_\varepsilon = +\infty. \tag{61}$$

We then consider the functional, J_ε defined for $u = (u_1, \dots, u_N) \in W^{1,p_\varepsilon}(\Omega)^N$ by

$$J_\varepsilon(u) := \frac{1}{p_\varepsilon} \sum_{i=1}^N \int_\Omega |\nabla u_i|^{p_\varepsilon} + \frac{1}{2\varepsilon} \int_\Omega \Big(\sum_{j=1}^N \lambda_j u_j \Big)_+^2 - \sum_{i=1}^N \lambda_i \int_\Omega u_i \mathrm{d}\nu_i,$$

observing that $J_\varepsilon(u) = J_\varepsilon(u + \alpha)$ if α_i's are constants that sum to 0, we can add the normalizing constraint

$$\int_\Omega u_i = 0,\ i = 1, \dots, N-1. \tag{62}$$

With this normalization at hand we can prove the following.

Proposition 6.6. *Let* $\varepsilon > 0$*,* $p_\varepsilon > d$*, then*

$$\inf_{u \in W^{1,p_\varepsilon}(\Omega)^N} J_\varepsilon(u)$$

admits a unique solution which satisfies the normalization (62)*.*

Proof. **Existence.** First note that for $i = 1, \dots, N-1$, $u_i \in W^{1,p_\varepsilon}(\Omega)$ with $\int_\Omega u_i = 0$, using successively Poincaré–Wirtinger's, Morrey's and Young's inequalities, we have

$$\begin{aligned}
& \int_\Omega |\nabla u_i|^{p_\varepsilon} - \lambda_i \int_\Omega u_i \mathrm{d}\nu_i \\
\geq\ & \frac{C_\varepsilon}{2} \|u_i\|^{p_\varepsilon}_{W^{1,p_\varepsilon}(\Omega)} - \lambda_i \|u_i\|_{L^\infty(\Omega)} \\
\geq\ & \frac{C_\varepsilon}{4} \|u_i\|^{p_\varepsilon}_{W^{1,p_\varepsilon}(\Omega)} + C'_\varepsilon \|u_i\|^{p_\varepsilon}_{L^\infty(\Omega)} - \frac{\delta}{p_\varepsilon} \|u_i\|^{p_\varepsilon}_{L^\infty(\Omega)} - \frac{1}{\delta^{\frac{q}{p_\varepsilon}} q} (\lambda_i)^q,
\end{aligned}$$

where C_ε, $C'_\varepsilon > 0$ are constants (independent of u_i), $\delta > 0$ and $q = \frac{p_\varepsilon}{p_\varepsilon - 1}$ the conjugate exponent.

To treat the N-th component, let $u_N \in W^{1,p_\varepsilon}(\Omega)$ and define $a_N := \fint_\Omega u_N \mathrm{d}x$, then similarly as before

$$\begin{aligned}
&\int_\Omega |\nabla u_N|^{p_\varepsilon} - \lambda_N \int_\Omega u_N \mathrm{d}\nu_N \\
\geq\ & \frac{C_\varepsilon}{2} \|u_N - a_N\|^{p_\varepsilon}_{W^{1,p_\varepsilon}(\Omega)} - \lambda_N \|u_N - a_N\|_{L^\infty(\Omega)} - \lambda_N a_N \\
\geq\ & \frac{C_\varepsilon}{4} \|u_N - a_N\|^{p_\varepsilon}_{W^{1,p_\varepsilon}(\Omega)} + C'_\varepsilon \|u_N - a_N\|^{p_\varepsilon}_{L^\infty(\Omega)} - \frac{\delta}{p_\varepsilon} \|u_N - a_N\|^{p_\varepsilon}_{L^\infty(\Omega)} - \frac{1}{\delta^{\frac{q}{p_\varepsilon}} q} (\lambda_N)^q - \lambda_N a_N.
\end{aligned}$$

By choosing $\delta > 0$ small enough, we obtain altogether

$$J_\varepsilon(u) \geq \frac{C_\varepsilon}{4} \sum_{i=1}^{N-1} \|u_i\|^{p_\varepsilon}_{W^{1,p_\varepsilon}(\Omega)} + \frac{C_\varepsilon}{4} \|u_N - a_N\|^{p_\varepsilon}_{W^{1,p_\varepsilon}(\Omega)} + C - \lambda_N a_N + \frac{1}{2\varepsilon} \int_\Omega \Big(\sum_{j=1}^N \lambda_j u_j \Big)_+^2, \quad (63)$$

where C is a constant only depending on p_ε, λ_i $(i = 1, \dots, N)$ and C'_ε.

Now let $(u^n)_{n\in\mathbb{N}} = (u^n_1, \dots, u^n_N)_{n\in\mathbb{N}} \in \big(W^{1,p_\varepsilon}(\Omega)^N\big)^{\mathbb{N}}$ be a minimizing sequence of J_ε satisfying our normalization. In order to conclude that $(u^n)_{n\in\mathbb{N}}$ is bounded in $W^{1,p_\varepsilon}(\Omega)$, it is enough to find an upper bound on $a^n_N = \fint_\Omega u^n_N \mathrm{d}x$. Assume by contradiction, that (up to a not relabeled subsequence) $a^n_N \to +\infty$ as $n \to \infty$, then, by (63) there are constants K, $\tilde{C}_\varepsilon > 0$ (independent of n) such that for $i = 1, \dots, N-1$

$$\left(\frac{K + \lambda_N a^n_N}{\tilde{C}_\varepsilon} \right)^{\frac{1}{p_\varepsilon}} \geq \|u^n_i\|_{L^\infty(\Omega)}, \text{ and } \left(\frac{K + \lambda_N a^n_N}{\tilde{C}_\varepsilon} \right)^{\frac{1}{p_\varepsilon}} \geq \|u^n_N - a^n_N\|_{L^\infty(\Omega)}, \quad (64)$$

for all $n \in \mathbb{N}$. But then, denoting $K^n_\varepsilon := \left(\frac{K + \lambda_N a^n_N}{\tilde{C}_\varepsilon} \right)^{\frac{1}{p_\varepsilon}}$

$$\begin{aligned}
&\frac{1}{2\varepsilon} \int_\Omega \Big(\sum_{j=1}^N \lambda_j u_j \Big)_+^2 - \lambda_N a^n_N \\
\geq\ & \frac{1}{2\varepsilon} \int_\Omega (\lambda_N a^n_N - K^n_\varepsilon)_+^2 - \lambda_N a^n_N \\
\geq\ & \frac{1}{2\varepsilon} \int_\Omega (\lambda_N a^n_N (1 - o(1)))_+^2 - \lambda_N a^n_N \to +\infty \quad \text{as } n \to \infty,
\end{aligned}$$

contradicting $(u^n)_{n\in\mathbb{N}}$ being a minimizing sequence. This implies that $(a^n_N)_{n\in\mathbb{N}}$ is bounded hence $(u^n_N)_{n\in\mathbb{N}}$ is bounded in $W^{1,p_\varepsilon}(\Omega)$. Since $(u^n_i)_{n\in\mathbb{N}}$ is bounded in $W^{1,p_\varepsilon}(\Omega)$ for $i = 1, \dots, N$, it has a subsequence that converges weakly in $W^{1,p_\varepsilon}(\Omega)$, by the weak lower semi continuity of J_ε, the weak limit of this subsequence is indeed a minimizer of J_ε.

Uniqueness. Let $u, \bar{u}$ be minimizers of J_ε. Then by strict convexity of $|\cdot|^{p_\varepsilon}$ and $\cdot^2$ we have

$$\nabla u_i = \nabla \bar{u}_i \quad \mathcal{L}^d\text{-a.e. for } i = 1, \dots, N,$$

$$\Big(\sum_{j=1}^N \lambda_j u_j \Big)_+^2 = \Big(\sum_{j=1}^N \lambda_j \bar{u}_j \Big)_+^2 \quad \mathcal{L}^d\text{-a.e.}$$

By the normalization (62) we then get $u_i = \bar{u}_i$ for $i = 1, \dots, N-1$, and there is $c_N \in \mathbb{R}$ such that $u_N = \bar{u}_N + c_N$. But then

$$0 = J_\varepsilon(u) - J_\varepsilon(\bar{u}) = \lambda_N \int_\Omega u_N \mathrm{d}\nu_N - \lambda_N \int_\Omega (u_N - c_N) \mathrm{d}\nu_N = \lambda_N c_N,$$

which is only possible if $c_N = 0$. □

The unique minimizer of J_ε under the normalization (62), $u^\varepsilon = (u_1^\varepsilon, \dots, u_N^\varepsilon)$ is characterized by the system of PDEs

$$-\nabla \cdot \Big(|\nabla u_i^\varepsilon|^{p_\varepsilon - 2} \nabla u_i^\varepsilon \Big) + \lambda_i \Big(\frac{\sum_{j=1}^N \lambda_j u_j^\varepsilon}{\varepsilon} \Big)_+ = \lambda_i \nu_i, \ i = 1, \dots, N \tag{65}$$

with Neumann boundary conditions, in the weak sense which means that, for every i and every $\varphi \in W^{1,p_\varepsilon}(\Omega)$, one has

$$\int_\Omega |\nabla u_i^\varepsilon|^{p_\varepsilon - 2} \nabla u_i^\varepsilon \cdot \nabla \varphi + \lambda_i \int_\Omega \Big(\frac{\sum_{j=1}^N \lambda_j u_j^\varepsilon}{\varepsilon} \Big)_+ \varphi = \lambda_i \int_\Omega \varphi \mathrm{d}\nu_i,$$

of course, supplemented by the normalization (62). To shorten notations and for further use, let us define

$$\sigma_i^\varepsilon := \frac{|\nabla u_i^\varepsilon|^{p_\varepsilon - 2} \nabla u_i^\varepsilon}{\lambda_i}, \ \nu^\varepsilon := \frac{1}{\varepsilon} \Big(\sum_{j=1}^N \lambda_j u_j^\varepsilon \Big)_+. \tag{66}$$

So that the optimality system (65) can be rewritten as

$$-\nabla \cdot \sigma_i^\varepsilon + \nu^\varepsilon = \nu_i, \ i = 1, \dots, N. \tag{67}$$

In particular (testing the N-th equation against a constant) ν^ε which is a nonnegative continuous (at least 1/2-Hölder when $p_\varepsilon \geq 2d$) function, is a probability density on Ω.

Then we have, the following convergence result:

Proposition 6.7. *Up to extracting a vanishing (not explicitly written) sequence $\varepsilon_n \to 0$ as $n \to \infty$, one may assume that*

- *$(u^\varepsilon)_{\varepsilon > 0}$ converges uniformly to some $u = (u_1, \dots, u_N)$ which is a vector of optimal dual potentials, i.e. solves (58)¡,*
- *for each i, $(\sigma_i^\varepsilon)_{\varepsilon>0}$ converges weakly $*$ to some vector-valued measure σ_i, $(\nu^\varepsilon)_{\varepsilon>0}$ converges weakly $*$ to some probability measure ν and $(\sigma_1, \dots, \sigma_N, \nu)$ solves the Beckmann problem (57). In particular, ν is a Wasserstein median.*

Proof. **Step 1: bounds on** u^ε. Multiplying (65) by u_i^ε first yields

$$\|\nabla u_i^\varepsilon\|_{L^{p_\varepsilon}}^{p_\varepsilon} + \lambda_i \int_\Omega u_i^\varepsilon \nu^\varepsilon = \lambda_i \int_\Omega u_i^\varepsilon \mathrm{d}\nu_i, \ i = 1, \dots, N. \tag{68}$$

Summing over i we thus get

$$\sum_{i=1}^N \|\nabla u_i^\varepsilon\|_{L^{p_\varepsilon}}^{p_\varepsilon} + \frac{1}{\varepsilon} \int_\Omega \Big(\sum_{j=1}^N \lambda_j u_j^\varepsilon \Big)_+^2 = \sum_{i=1}^N \lambda_i \int_\Omega u_i^\varepsilon \mathrm{d}\nu_i. \tag{69}$$

By Morrey's and Hölder's inequalities, $p_\varepsilon \geq 2d$ and the fact that u_i^ε has zero mean for $i = 1, \dots, N-1$, we have for positive constant C and C' depending only on Ω (but possibly changing from one line to another)

$$\|u_i^\varepsilon\|_\infty \leq C \|\nabla u_i^\varepsilon\|_{L^{2d}} \leq C |\Omega|^{\frac{1}{2d} - \frac{1}{p_\varepsilon}} \|\nabla u_i^\varepsilon\|_{L^{p_\varepsilon}} \leq C' \|\nabla u_i^\varepsilon\|_{L^{p_\varepsilon}},$$

which together with (68) and the fact that both ν^ε and ν_i are probability measures gives

$$\max_{i=1,\dots,N-1} \|u_i^\varepsilon\|_\infty \leq C, \ \max_{i=1,\dots,N-1} \|\nabla u_i^\varepsilon\|_{L^{p_\varepsilon}}^{p_\varepsilon} \leq C. \tag{70}$$

Let us now get similar bounds on u_N^ε, using (68) with $i = N$ and using the fact that ν^ε and ν_i are probability measures and then again Morrey's inequality (applied to $u_N - \fint_\Omega u_N$), we get

$$\|\nabla u_N^\varepsilon\|_{L^{p_\varepsilon}}^{p_\varepsilon} = \lambda_N \int (u_N^\varepsilon - \min_{\overline{\Omega}} u_N^\varepsilon)(\nu_i - \nu^\varepsilon) \leq \lambda_N \mathrm{osc}_{\overline{\Omega}}(u_N^\varepsilon) \leq C\|\nabla u_N^\varepsilon\|_{L^{p_\varepsilon}},$$

which gives

$$\|\nabla u_N^\varepsilon\|_{L^{p_\varepsilon}}^{p_\varepsilon} \leq C, \ \mathrm{osc}_{\overline{\Omega}}(u_N^\varepsilon) \leq C.$$

With (70) and (69) and the bound on $\mathrm{osc}_{\overline{\Omega}}(u_N^\varepsilon)$, we thus get taking $C' \geq \sum_{i=1}^{N-1} \lambda_i u_i^\varepsilon$

$$0 \leq \frac{1}{\varepsilon} \int_\Omega (\lambda_N \max_{\overline{\Omega}} u_N^\varepsilon - C')_+^2 \leq C + \lambda_N \int u_N^\varepsilon \nu_N \leq C + \lambda_N \max_{\overline{\Omega}} u_N^\varepsilon$$

from which one readily deduces that $\max_{\overline{\Omega}} u_N^\varepsilon$ is bounded uniformly in ε, hence $(u_N^\varepsilon)_{\varepsilon>0}$ is bounded in L^∞ because of the bound on $\mathrm{osc}_{\overline{\Omega}}(u_N^\varepsilon)$. Finally, we have shown that

$$\max_{i=1,\dots,N} \|u_i^\varepsilon\|_\infty \leq C, \ \max_{i=1,\dots,N} \|\nabla u_i^\varepsilon\|_{L^{p_\varepsilon}}^{p_\varepsilon} \leq C, \tag{71}$$

which implies also $C^{0,\frac{1}{2}}$ bounds so extracting a vanishing (not explicitly written) sequence $\varepsilon_n \to 0$ as $n \to \infty$, thanks to Ascoli–Arzelá's theorem, one may assume that $(u^\varepsilon)_{\varepsilon>0}$ converges uniformly to some u with $u \in W^{1,q}(\Omega)$ for every $q \in (1, +\infty)$. And since $(\nabla u^\varepsilon)_{\varepsilon>0}$ is bounded in every L^q, we may also assume that for every $q \in (1, +\infty)$, $(\nabla u^\varepsilon)_{\varepsilon>0}$ converges weakly to ∇u in $L^q(\Omega)$. Of course, we may also assume that $(\nu^\varepsilon)_{\varepsilon>0}$ converges weakly $*$ to some probability measure ν and that the (bounded in L^1, thanks to (71) and the definition of σ_i^ε) sequence $(\sigma_i^\varepsilon)_{\varepsilon>0}$ converges weakly $*$ to some vector-valued measure σ_i.

Step 2: u satisfies the constraints of the dual. By (69) and (71), we have

$$\int_\Omega \Big(\sum_{j=1}^N \lambda_j u_j^\varepsilon\Big)_+^2 \leq C\varepsilon.$$

so that, letting $\varepsilon \to 0^+$, we get

$$\sum_{j=1}^N \lambda_j u_j \leq 0.$$

Let us now prove that each u_i is 1-Lipschitz as a consequence of (71) and (61). First fix q and let ε be small enough so that $p_\varepsilon \geq q$, then

$$\|\nabla u_i^\varepsilon\|_{L^q} \leq |\Omega|^{\frac{1}{q} - \frac{1}{p_\varepsilon}} C^{\frac{1}{p_\varepsilon}}.$$

So letting $\varepsilon \to 0$, we get with (61)

$$\|\nabla u_i\|_{L^q} \leq |\Omega|^{\frac{1}{q}}, \quad \text{for all } q \in (1, +\infty).$$

So letting now $q \to +\infty$ we obtain

$$\|\nabla u_i\|_{L^\infty} \leq 1$$

which implies that each u_i is 1-Lipschitz by convexity of Ω.

Step 3: optimality of the limits. We already know that ν is a probability measure. Passing to the limit in (67), we get

$$-\nabla \cdot \sigma_i + \nu = \nu_i, \ i = 1, \dots, N,$$

which is the constraint in Beckmann problem (57). Since u is admissible in the dual, to conclude, by weak duality, it is enough to show that

$$\sum_{i=1}^{N} \lambda_i |\sigma_i|(\Omega) \leq \sum_{i=1}^{N} \lambda_i \int_\Omega u_i \mathrm{d}\nu_i. \tag{72}$$

First observe that (69) entails

$$\sum_{i=1}^{N} \lambda_i \int u_i^\varepsilon \mathrm{d}\nu_i \geq \sum_{i=1}^{N} \int_\Omega |\nabla u_i^\varepsilon|^{p_\varepsilon} = \sum_{i=1}^{N} \int_\Omega |\lambda_i \sigma_i^\varepsilon|^{\frac{p_\varepsilon}{p_\varepsilon - 1}}.$$

Note then that, by Hölder's inequality we have

$$\int_\Omega |\sigma_i^\varepsilon|^{\frac{p_\varepsilon}{p_\varepsilon-1}} \geq \|\sigma_i^\varepsilon\|_{L^1}^{\frac{p_\varepsilon}{p_\varepsilon-1}} |\Omega|^{-\frac{1}{p_\varepsilon-1}}$$

so that

$$\liminf_{\varepsilon\to 0^+} \int_\Omega |\sigma_i^\varepsilon|^{\frac{p_\varepsilon}{p_\varepsilon-1}} \geq \liminf_{\varepsilon\to 0^+} \|\sigma_i^\varepsilon\|_{L^1} \|\sigma_i^\varepsilon\|_{L^1}^{\frac{1}{p_\varepsilon-1}} \geq \liminf_{\varepsilon\to 0^+} \|\sigma_i^\varepsilon\|_{L^1} \geq |\sigma_i|(\Omega)$$

where the second inequality is obtained by distinguishing the (obvious) case where (after a suitable extraction) $(\sigma_i^\varepsilon)_{\varepsilon>0}$ converges strongly to 0 in L^1 and the case where $\|\sigma_i^\varepsilon\|_{L^1}$ remains bounded away from 0 and the last inequality follows from the weak $*$ convergence of $(\sigma_i^\varepsilon)_{\varepsilon>0}$ to σ_i. We thus get

$$\sum_{i=1}^{N} \lambda_i \int_\Omega u_i \nu_i = \liminf_{\varepsilon\to 0^+} \sum_{i=1}^{N} \lambda_i \int_\Omega u_i^\varepsilon \nu_i \geq \liminf_{\varepsilon\to 0^+} \sum_{i=1}^{N} \lambda_i^{\frac{p_\varepsilon}{p_\varepsilon-1}} \int_\Omega |\sigma_i^\varepsilon|^{\frac{p_\varepsilon}{p_\varepsilon-1}} \geq \sum_{i=1}^{N} \lambda_i |\sigma_i|(\Omega),$$

which proves (72) and ends the proof. □

7 Numerics

In this section, we briefly mention the numerical methods we employed to generate the figures in the paper and present a new one based on a Douglas–Rachford scheme for the multi-flow formulation (57). All the experiments are performed in Python on a Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz and 8 Gb of RAM and are available for reproducibility at https://github.com/TraDE-OPT/wasserstein-medians.

7.1 Sorting, Linear Programming, Sinkhorn

Recall from Section 4, that in the one dimensional case, the Wasserstein median problem admits an almost-closed form solution, which can be computed directly with simple sorting procedures. We implemented these well-known schemes to generate Figure 1. Here we rather focus on the case $\mathcal{X} \subset \mathbb{R}^2$, which is more relevant e.g. for imaging.

Wasserstein median problems on a fixed grid of size $n = p^2$ for a sample of size N can be tackled either via Linear Programming methods, taking advantage of the minimum-cost flow nature of the problem [5], or via Sinkhorn-like methods on an entropy-regularized finite dimensional variant of (40) [19], see also [9, 20]. The latter represents the most popular approach. We employed the Sinkhorn method to generate Figure 2. Despite their well-known advantages, Sinkhorn algorithm and entropic regularization methods can lead to severe computational issues, such as blurred outputs, important numerical instabilities, and memory issues to store the so-called *kernel matrix* [47]. It is worth mentioning that several efforts have been made to develop Sinkhorn-like methods

that address these limitations, including log-space tricks for stability [9], de-biased variants for blurring artifacts [32], and truncation strategies for memory and speed improvements [47].

In the next paragraph, we present a new approach which targets (57) and benefits from low memory requirements, fast convergence behaviour, and produces non-blurred approximate medians. Note, however, that this is approach, well-suited for Wasserstein medians, cannot be easily generalized to approximate Wasserstein barycenters .

7.2 Douglas–Rachford on the Beckmann formulation.

Given a square domain Ω, and $N \geq 2$ measures $(\nu_1, \dots, \nu_N) \in \mathcal{P}(\Omega)^N$, consider the Beckmann minimal flow formulation of the Wasserstein median problem (57). To discretize (57), we introduce the square grid $\mathcal{G}_h := \{hi \ : \ i = 1, \dots, p\}^2$ with step-length $h := 1/p$, and the discrete spaces $\mathcal{M}_h := \{\mu : \mathcal{G} \to \mathbb{R}\}$ and $\boldsymbol{\mathcal{S}}_h := \{\boldsymbol{\sigma} : \mathcal{G} \to \mathbb{R}^2\}$. Note that $\mathcal{M}_h$ and $\boldsymbol{\mathcal{S}}_h$ are finite dimensional vector spaces which can be identified with $\mathbb{R}^n$ and $\mathbb{R}^{n\times 2}$, respectively, where $n := p^2$. Thus, we often treat elements in $\mathcal{M}_h$ and $\boldsymbol{\mathcal{S}}_h$ as vectors. We consider the usual discretization of the gradient $\nabla_h : \mathcal{M}_h \to \boldsymbol{\mathcal{S}}_h$ defined via forward differences with homogeneous Neumann boundary conditions as in [17, Section 6.1]. The discrete divergence operator, which we denote by $\mathrm{div}_h = -\nabla_h^*$, is the opposite adjoint of ∇_h, with respect to the scalar products $\langle \cdot, \cdot \rangle_{\mathcal{M}_h}$ and $\langle \cdot, \cdot \rangle_{\boldsymbol{\mathcal{S}}_h}$ (i.e. the usual ℓ^2 scalar products on $\mathbb{R}^n$ and $\mathbb{R}^{n\times 2}$, respectively). Now, let

$$\mathcal{F}_h := \big\{ (\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_N, \nu) \in \boldsymbol{\mathcal{S}}_h^N \times \mathcal{M}_h \ : \ \mathrm{div}_h \boldsymbol{\sigma}_k + \nu_k = \nu \text{ for all } k = 1, \dots, N \big\},$$

where $(\nu_1, \dots, \nu_N) \in \mathcal{M}_h^N$ are suitable (not renamed) discretizations of $\nu_1, \dots, \nu_N$ on the grid $\mathcal{G}_h$. With this notation, let us consider the discretized version of (57):

$$\min_{(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_N, \nu) \in \mathcal{F}_h} \sum_{k=1}^N \lambda_k \|\boldsymbol{\sigma}_k\|_{1,2} + \mathbb{I}_\Delta(\nu) \tag{73}$$

Where Δ is the unit simplex, and $\|\cdot\|_{1,2}$ is the $\ell_{1,2}$ norm on $\boldsymbol{\mathcal{S}}_h$, also known as *group-Lasso* penalty, which is defined for all $\boldsymbol{\sigma} \in \boldsymbol{\mathcal{S}}_h$ by $\|\boldsymbol{\sigma}\|_{1,2} := \sum_{i=1}^n \|\boldsymbol{\sigma}(x_i)\|$, where $\|\cdot\|$ is the usual ℓ_2 norm on $\mathbb{R}^n$.

To solve (73), we apply a Douglas–Rachford method to

$$\min_{(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_N, \nu) \in \boldsymbol{\mathcal{S}}_h^N \times \mathcal{M}_h} \underbrace{\sum_{k=1}^N \lambda_k \|\boldsymbol{\sigma}_k\|_{1,2} + \mathbb{I}_\Delta(\nu)}_{:= g_1(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_N, \nu)} + \underbrace{\mathbb{I}_{\mathcal{F}_h}(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_N, \nu)}_{:= g_2(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_N, \nu)}.$$

The Douglas–Rachford method [24, 36] is an instance of the proximal point algorithm [13, 26], which can be employed to solve a minimization problem consisting of the sum of two convex lower semicontinuous functions which are accessible through evaluation of their proximity operators. In our case, the proximity operator of g_1, which is *separable*, consists in a projection onto the unit simplex, denoted by P_Δ, for the discrete measure ν and on the application of the proximity operator of the group-Lasso penalty, denoted by Shrink_τ, where $\tau > 0$, on each component $\boldsymbol{\sigma}_i$, which can be computed in closed form [18].

The proximity operator of g_2, i.e. the projection onto the affine subspace $\mathcal{F}_h$, is more delicate. Recall from optimality conditions that, formally, the projection onto the solution set of a linear system of the form $Ax = b$ is given, for all y, by $Py = y - A^*\xi$ where ξ is any element that solves $AA^*\xi = Ay - b$. In our case, we have $b = -[\nu_1, \dots, \nu_N]^T$ and the linear operators A and AA^* can

be written in block form as

$$A := \begin{bmatrix} \mathrm{div}_h & & & -I \\ & \ddots & & \vdots \\ & & \mathrm{div}_h & -I \end{bmatrix}, \quad AA^* = \begin{bmatrix} -\Delta_h + I & I & \cdots & I \\ I & \ddots & & \vdots \\ \vdots & & & I \\ I & I & \cdots & -\Delta_h + I \end{bmatrix}, \tag{74}$$

where $\Delta_h : \mathcal{M}_h \to \mathcal{M}_h$ is the discrete Laplacian operator, namely $\Delta_h = \mathrm{div}_h \nabla_h$.

Proposition 7.1. *Let $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_N) \in \boldsymbol{\mathcal{S}}_h^N$ and $\nu \in \mathcal{M}_h \cap \Delta$, and let $\nabla_h : \mathcal{M}_h \to \boldsymbol{\mathcal{S}}_h$ be the discrete gradient operator defined via forward differences with homogeneous Neumann boundary conditions. Then the projection $(\widetilde{\boldsymbol{\sigma}}, \widetilde{\nu})$ of $(\boldsymbol{\sigma}, \nu)$ onto $\mathcal{F}_h$ is given by*

$$\widetilde{\boldsymbol{\sigma}}_i := \boldsymbol{\sigma}_i + \nabla_h \xi_i, \quad \widetilde{\nu} := \nu + \xi_1 + \cdots + \xi_N,$$

where $\xi_i := \xi_i' - (I - \frac{1}{N}\Delta_h)^{-1}\left(\frac{1}{N}\sum_{j=1}^N \xi_j'\right)$ and ξ_i' is any solution to

$$-\Delta_h \xi_i' = \mathrm{div}_h \boldsymbol{\sigma}_i + \nu_i - \nu \quad \textit{for all } i = 1, \dots, N. \tag{75}$$

Proof. First, let $i = 1, \dots, N$, let $\mathbf{1} \in \mathcal{M}_h$ be constantly equal to 1, and note that, by definition of the scalar products, since $\nu_i,\ \nu \in \Delta$ and $\ker \nabla_h = \mathrm{span}\{\mathbf{1}\}$, we get

$$\langle \mathrm{div}_h \boldsymbol{\sigma}_i + \nu_i - \nu, \mathbf{1} \rangle_{\mathcal{M}_h} = \langle \mathrm{div}_h \boldsymbol{\sigma}_i, \mathbf{1} \rangle_{\mathcal{M}_h} = -\langle \boldsymbol{\sigma}_i, \nabla_h \mathbf{1} \rangle_{\boldsymbol{\mathcal{S}}_h} = 0.$$

Hence, $\mathrm{div}_h \boldsymbol{\sigma}_i + \nu_i - \nu \in (\ker \nabla_h)^\perp = (\ker \Delta_h)^\perp = \mathrm{Im}\, \Delta_h$ for all $i = 1, \dots, N$, and, thus, (75) actually admits a solution. From optimality conditions, we only need to show that $A(\widetilde{\boldsymbol{\sigma}}, \widetilde{\nu}) = b$ and that $\boldsymbol{\xi} := (\xi_1, \dots, \xi_N)$ solves $AA^*\boldsymbol{\xi} = A(\boldsymbol{\sigma}, \nu) - b$ where A and AA^* are defined in (74). Let us start with the latter. Denoting $\bar{\xi}' := \frac{1}{N}\sum_{i=1}^N \xi_i'$, we have for all $i = 1, \dots, N$ that

$$\begin{aligned} -\Delta_h \xi_i + \xi_1 + \cdots + \xi_N &= -\Delta_h \xi_i' + \Delta_h \left(I - \tfrac{1}{N}\Delta_h\right)^{-1} \bar{\xi}' + N\bar{\xi}' - N\left(I - \tfrac{1}{N}\Delta_h\right)^{-1} \bar{\xi}' \\ &= \mathrm{div}_h \boldsymbol{\sigma}_i + \nu_i - \nu + N\bar{\xi}' - N\left(I - \tfrac{1}{N}\Delta_h\right)\left(I - \tfrac{1}{N}\Delta_h\right)^{-1} \bar{\xi}' \\ &= \mathrm{div}_h \boldsymbol{\sigma}_i + \nu_i - \nu. \end{aligned}$$

Hence $AA^*\boldsymbol{\xi} = A(\boldsymbol{\sigma}, \nu) - b$. Regarding $A(\widetilde{\boldsymbol{\sigma}}, \widetilde{\nu}) = b$, we have

$$\begin{aligned} \mathrm{div}_h \widetilde{\boldsymbol{\sigma}}_i + \nu_i &= \mathrm{div}_h \boldsymbol{\sigma}_i + \Delta_h \xi_i + \nu_i = \mathrm{div}_h \boldsymbol{\sigma}_i + \Delta_h \xi_i' - \Delta_h \left(I - \tfrac{1}{N}\Delta_h\right)^{-1} \bar{\xi}' + \nu_i \\ &= \nu - \Delta_h \left(I - \tfrac{1}{N}\Delta_h\right)^{-1} \bar{\xi}' = \nu + N\bar{\xi}' - N\left(I - \tfrac{1}{N}\Delta_h\right)^{-1} \bar{\xi}' = \widetilde{\nu}, \end{aligned}$$

which concludes the proof. □

Proposition 7.1 allows us to implement a Douglas–Rachford scheme on (73), which we summarize in Algorithm 1. Note that, in Algorithm 1, we are required to solve two *sparse* (elliptic) linear systems, which we tackle with generic sparse linear solvers provided by standard Python libraries. However, one should put adequate care when trying to solve the first Laplacian system. Indeed, if the projection onto the simplex is not computed sufficiently well, the right-hand side can lie out of the range of the Laplacian. For this reason, in our numerical implementation, we smoothed out all possible numerical errors with a further projection of the right-hand side onto the set of discrete measures with a total mass equal to one.

The computational cost required to solve the aforementioned linear systems is overall balanced with a very fast iteration-wise convergence behaviour. Remarkably, there is no need to store dense $n \times n$ matrices. This makes the proposed method suitable for highly large-scale instances, see e.g. Figures 3 and 6.

Convergence. The Douglas–Rachford splitting method benefits from robust convergence guarantees, without any condition neither on the starting point nor on the step-size $\tau > 0$ [13, 36]. In particular, we have that if $(\boldsymbol{\sigma}_q^k)_{k\in\mathbb{N}}$, $(\nu^k)_{k\in\mathbb{N}}$, $(\boldsymbol{\eta}^k)_{k\in\mathbb{N}}$ and $(\mu^k)_{k\in\mathbb{N}}$ are the sequences generated by Algorithm 1, then for each $q = 1, \dots, N$, we have $\boldsymbol{\sigma}_q^k \to \boldsymbol{\sigma}_q^*$ and $\nu^k \to \nu^*$ and $(\boldsymbol{\sigma}_1^*, \dots, \boldsymbol{\sigma}_N^*, \nu^*)$ solves (73). As a stopping criterion, we measure the residual $r^k := \sum_{q=1}^N \|\boldsymbol{\eta}_q^{k+1} - \boldsymbol{\eta}_q^k\|^2_{\boldsymbol{\mathcal{S}}_h} + \|\mu^{k+1} - \mu^k\|^2_{\mathcal{M}_h}$, which is guaranteed to converge to zero with a $o(k^{-1})$ worst-case rate, and we stop the iterations as soon as the residual drops below a prescribed tolerance.

Comments. Note that to solve (73), we also implemented the Primal Dual Hybrid Gradient method by Chambolle and Pock, with different step-size selection strategies, such as *backtracking* and *adaptive* schemes [30], and several different *fixed* step-sizes choices, which, however, always provided very slow behaviours, and therefore, we chose not to discuss it further. Note that for OT-like problems, the Douglas–Rachford splitting method has been employed first in its dual formulation (ADMM) in [41], then in [8] and more recently in [14]. Its extension to the Wasserstein median case proposed in the present paper has been surprisingly overlooked.

Acknowledgments: E.C. would like to thank Stefano Gualandi for the helpful discussions during the preparation of this work. G.C. acknowledges the support of the Lagrange Mathematics and Computing Research Center. E.C. has received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Skłodowska-Curie Grant Agreement No. 861137. The Department of Mathematics and Scientific Computing, to which E.C. is affiliated, is a member of NAWI Graz (https://www.nawigraz.at/). K.E. acknowledges that this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754362.

Appendix

Proof of (6). Of course, if all the x_i's are equal $I_\pm(\boldsymbol{x}) = \{1, \dots, N\}$ and (6) is nothing else than (5). We may therefore assume that

$$\Delta := \min\{|x_i - x_j| \ : \ x_i \neq x_j\} > 0.$$

Data: A collection of discrete probability measures $\nu_1, \dots, \nu_N \in \mathcal{M}_h$, a step-size $\tau > 0$ and relaxation parameters $(\theta_k)_{k\in\mathbb{N}}$ in $[0, 2)$ such that $\sum_{k=0}^\infty \theta_k(2 - \theta_k) = +\infty$
Result: $\nu^* = \lim_{k\to+\infty} \nu^k$, $\boldsymbol{\sigma}_q^* = \lim_{k\to+\infty} \boldsymbol{\sigma}_q^k$ for $q = 1, \dots, N$ solution to (73)
Initialize: $\boldsymbol{\eta}_1^0, \dots, \boldsymbol{\eta}_N^0 \in \boldsymbol{\mathcal{S}}_h$ and $\mu^0 \in \mathcal{M}_h \cap \Delta$
while *not convergent* **do**
  $\boldsymbol{\sigma}_q^{k+1} = \mathrm{Shrink}_\tau\left(\boldsymbol{\eta}_q^k\right)$ for all $q = 1, \dots, N$
  $\nu^{k+1} = P_\Delta(\mu^k)$
  for $q = 1, \dots, N$ **do**
    Solve: $-\Delta_h \xi_q' = \mathrm{div}_h(2\boldsymbol{\sigma}_q^{k+1} - \boldsymbol{\eta}_q^k) + \nu_q - 2\nu^{k+1} + \mu^k$
    $\xi_q = \xi_q' - \left(I - \frac{1}{N}\Delta_h\right)^{-1}\left(\frac{1}{N}\sum_{j=1}^N \xi_j'\right)$
  end
  $\boldsymbol{\eta}_q^{k+1} = (1 - \theta_k)\boldsymbol{\eta}_q^k + \theta_k\left(\boldsymbol{\sigma}_q^{k+1} + \nabla_h \xi_q\right)$ for all $q = 1, \dots, N$
  $\mu^{k+1} = (1 - \theta_k)\mu^k + \theta_k\left(\nu^{k+1} + \xi_1 + \dots + \xi_N\right)$
end

Algorithm 1: Douglas–Rachford for the Wasserstein median problem

Then, setting

$$\delta_i^+ := \max_k \{y_k - x_k \ : \ x_k = x_i\}, \ \delta_i^- := \min_k \{y_k - x_k \ : \ x_k = x_i\},$$

and $\boldsymbol{\delta}^\pm := (\delta_1^\pm, \dots, \delta_N^\pm)$ we have $\boldsymbol{x} + \boldsymbol{\delta}^+ \geq \boldsymbol{y} \geq \boldsymbol{x} + \boldsymbol{\delta}^-$ and then by monotonicity

$$\mathsf{M}_{\boldsymbol{\lambda}}^\pm(\boldsymbol{x} + \boldsymbol{\delta}^+) \geq \mathsf{M}_{\boldsymbol{\lambda}}^\pm(\boldsymbol{y}) \geq \mathsf{M}_{\boldsymbol{\lambda}}^\pm(\boldsymbol{x} + \boldsymbol{\delta}^-),$$

but if we choose $\boldsymbol{y}$ close enough to $\boldsymbol{x}$, namely such that

$$\max_{i,j} |\delta_i^\pm - \delta_j^\pm| \leq \frac{\Delta}{2}$$

this, together with the definition of Δ, implies that the components of $\boldsymbol{x}$ and $\boldsymbol{x} + \boldsymbol{\delta}^\pm$ are ordered in the same way, i.e. $x_j < x_i$ if and only if $x_j + \delta_j^\pm < x_i + \delta_i^\pm$. Thus, for $i \in I_+(x)$, $x_i = \mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{x})$ and

$$\sum_{j \,:\, x_j + \delta_j^- < x_i + \delta_i^-} \lambda_j = \sum_{j \,:\, x_j < x_i} \lambda_j \leq \frac{1}{2},$$

so that

$$\mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{y}) \geq \mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{x} + \boldsymbol{\delta}^-) \geq x_i + \boldsymbol{\delta}_i^- \geq \mathsf{M}_{\boldsymbol{\lambda}}^+(\boldsymbol{x}) + \min_{k \in I_+(\boldsymbol{x})} (y_k - x_k).$$

In a similar way for $i \in I_-(x)$, $x_i = \mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{x})$ and

$$\sum_{j \,:\, x_j + \delta_j^- \leq x_i + \delta_i^-} \lambda_j = \sum_{j \,:\, x_j \leq x_i} \lambda_j \geq \frac{1}{2},$$

so that

$$\mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{y}) \geq \mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{x} + \boldsymbol{\delta}^-) \geq x_i + \boldsymbol{\delta}_i^- \geq \mathsf{M}_{\boldsymbol{\lambda}}^-(\boldsymbol{x}) + \min_{k \in I_-(\boldsymbol{x})} (y_k - x_k).$$

This proves the rightmost inequalites in (6). The proof of the leftmost inequalities in (6) is similar and thus omitted.

References

[1] Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein Space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[2] Jason M. Altschuler and Enric Boix-Adserà. Wasserstein barycenters are NP-hard to compute. *SIAM Journal on Mathematics of Data Science*, 4(1):179–203, 2022.

[3] Luigi Ambrosio. *Lecture Notes on Optimal Transport Problems*, pages 1–52. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[4] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Lectures in mathematics ETH Zürich. Birkhäuser, Boston, 2005.

[5] Gennaro Auricchio, Federico Bassetti, Stefano Gualandi, and Marco Veneroni. Computing Wasserstein barycenters via linear programming. In *Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pages 355–363, 2019.

[6] Julio Backhoff-Veraguas, Joaquin Fontbona, Gonzalo Rios, and Felipe Tobar. Bayesian learning with Wasserstein barycenters. *ESAIM: Probability and Statistics*, 26:436–472, 2022.

[7] Martin Beckmann. A continuous model of transportation. *Econometrica. Journal of the Econometric Society*, 20:643–660, 1952.

[8] Jean-David Benamou and Guillaume Carlier. Augmented lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *Journal of Optimization Theory and Applications*, 167(1):1–26, 2015.

[9] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[10] Emmanuel Boissard, Thibaut Le Gouic, and Jean-Michel Loubes. Distribution's template estimate with Wasserstein metrics. *Bernoulli*, 21(2):740–759, 2015.

[11] Guy Bouchitté and Giuseppe Buttazzo. Characterization of optimal shapes and masses through Monge–Kantorovich equation. *Journal of the European Mathematical Society*, 3(2):139–168, 2001.

[12] Guy Bouchitté, Giuseppe Buttazzo, and Pierre Seppecher. Shape optimization solutions via Monge–Kantorovich equation. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 324(10):1185–1191, 1997.

[13] Kristian Bredies, Enis Chenchene, Dirk A. Lorenz, and Emanuele Naldi. Degenerate preconditioned proximal point algorithms. *SIAM Journal on Optimization*, 32(3):2376–2401, 2022.

[14] Kristian Bredies, Enis Chenchene, and Emanuele Naldi. Graph and distributed extensions of the Douglas–Rachford method. *ArXiv:2211.04782*, 2022.

[15] Giuseppe Buttazzo, Guillaume Carlier, and Katharina Eichinger. Wasserstein interpolation with constraints and application to a parking problem. *arXiv:2207.14261*, 2022.

[16] Guillaume Carlier and Ivar Ekeland. Matching for Teams. *Economic Theory*, 42(2):397–418, 2010.

[17] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[18] Patrick L. Combettes and Valérie R. Wajs. Signal recovery by proximal forward–backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[19] Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

[20] Marco Cuturi and Arnaud Doucet. Fast Computation of Wasserstein Barycenters. In *Proceedings of the 31st International Conference on Machine Learning*, pages 685–693, 2014.

[21] L. De Pascale, L. C. Evans, and A. Pratelli. Integral estimates for transport densities. *The Bulletin of the London Mathematical Society*, 36(3):383–395, 2004.

[22] Luigi De Pascale and Aldo Pratelli. Regularity properties for Monge transport density and for solutions of some shape optimization problem. *Calculus of Variations and Partial Differential Equations*, 14(3):249–274, 2002.

[23] Luigi De Pascale and Aldo Pratelli. Sharp summability for Monge transport density via interpolation. *ESAIM: Control, Optimisation and Calculus of Variations*, 10(4):549–552, 2004.

[24] Jim Douglas, Jr. and H. H. Rachford, Jr. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82:421–439, 1956.

[25] Samer Dweik. $L^{p,q}$ estimates on the transport density. *Communications on Pure and Applied Analysis*, 18(6):3001–3009, 2019.

[26] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(3, Ser. A):293–318, 1992.

[27] Paul Embrechts and Marius Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013.

[28] L. C. Evans and W. Gangbo. Differential equations methods for the Monge–Kantorovich mass transfer problem. *Memoirs of the American Mathematical Society*, 137(653):viii+66, 1999.

[29] Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10:215–310, 1948.

[30] Tom Goldstein, Min Li, and Xiaoming Yuan. Adaptive primal-dual splitting methods for statistical learning and image processing. In *Advances in Neural Information Processing Systems*, volume 28, pages 2089—2097, 2015.

[31] Noureddine Igbida, Van Thanh Nguyen, and Julián Toledo. On the uniqueness and numerical approximations for a matching problem. *SIAM Journal on Optimization*, 27(4):2459–2480, 2017.

[32] Hicham Janati, Marco Cuturi, and Alexandre Gramfort. Debiased Sinkhorn barycenters. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 4692–4701, 2020.

[33] Young-Heon Kim and Brendan Pass. Wasserstein barycenters over Riemannian manifolds. *Advances in Mathematics*, 307:640–683, 2017.

[34] Alexey Kroshnin. Fréchet barycenters in the Monge–Kantorovich spaces. *Journal of Convex Analysis*, 25(4):1371–1395, 2018.

[35] Thibaut Le Gouic and Jean-Michel Loubes. Existence and consistency of Wasserstein barycenters. *Probability Theory and Related Fields*, 168(3-4):901–917, 2017.

[36] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[37] Hendrik P. Lopuhaa and Peter J. Rousseeuw. Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices. *The Annals of Statistics*, 19(1):229–248, 1991.

[38] José M. Mazón, Julio D. Rossi, and Julián Toledo. An optimal matching problem for the Euclidean distance. *SIAM Journal on Mathematical Analysis*, 46(1):233–255, 2014.

[39] Robert J. McCann. A convexity principle for interacting gases. *Advances in Mathematics*, 128(1):153–179, 1997.

[40] Victor M. Panaretos and Yoav Zemel. *An invitation to statistics in Wasserstein space.* SpringerBriefs in Probability and Mathematical Statistics. Springer, Cham, 2020.

[41] Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[42] Gabriel Peyré and Marco Cuturi. Computational Optimal Transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–206, 2018.

[43] Frank Plastria. Four-point Fermat location problems revisited. New proofs and extensions of old results. *IMA Journal of Management Mathematics*, 17(4):387–396, 2006.

[44] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *SSVM*, volume 6667 of *Lecture Notes in Computer Science*, pages 435–446, 2011.

[45] Filippo Santambrogio. Absolute continuity and summability of transport densities: simpler proofs and new estimates. *Calculus of Variations and Partial Differential Equations*, 36(3):343–354, 2009.

[46] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling.* Birkhäuser, 2015.

[47] Bernhard Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM Journal on Scientific Computing*, 41(3):A1443–A1481, 2019.