Computing Barycentres of Measures for Generic Transport Costs

Eloi Tanguy[1], Julie Delon[1], and Nathaël Gozlan[1]

[1]Université Paris Cité, CNRS, MAP5, F-75006 Paris, France

20th December 2024

Abstract

Wasserstein barycentres represent average distributions between multiple probability measures for the Wasserstein distance. The numerical computation of Wasserstein barycentres is notoriously challenging. A common approach is to use Sinkhorn iterations, where an entropic regularisation term is introduced to make the problem more manageable. Another approach involves using fixed-point methods, akin to those employed for computing Fréchet means on manifolds. The convergence of such methods for 2-Wasserstein barycentres, specifically with a quadratic cost function and absolutely continuous measures, was studied by Alvarez-Esteban et al. in [Álv+16]. In this paper, we delve into the main ideas behind this fixed-point method and explore how it can be generalised to accommodate more diverse transport costs and generic probability measures, thereby extending its applicability to a broader range of problems. We show convergence results for this approach and illustrate its numerical behaviour on several barycentre problems.

Table of Contents

1 Introduction

1.1 Related Works and Motivation

Wasserstein barycentres represent a powerful concept in Optimal Transport theory, enabling the computation of average distributions between multiple probability measures. These barycentres preserve the geometric structure of the underlying distributions, making them particularly suited for machine learning tasks. They have proven useful in numerous applications, including image processing [Rab+12], computer graphics [Sol+15; BPC16], statistics [BCP19], domain adaptation [MM21], generative modelling [Kor+22], fairness in machine learning [Gor+19] or model selection in Bayesian learning [Bac+22]. Wasserstein barycentres are also at the core of clustering methods such as K-means, to define centroids in spaces of probability measures [Ho+17; Mi+18].

The classical notion of barycentre refers to the weighted average of a set of points (x_k) with positive weights (λ_k) summing to 1, in a metric space (E, d). Formally, a barycentre $\bar{x}$ is a point that minimises the weighted sum of (typically squared) distances:

$$\bar{x} \in \underset{x\in E}{\operatorname{argmin}} \sum_{k=1}^{K} \lambda_k d^2(x, x_k).$$

This concept can be extended to the space of probability measures, where d can be replaced for instance by a transportation cost $\mathcal{T}_c$. We remind that for two probability measures μ and ν on metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$, and a cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$, the optimal transport cost between μ and ν for the ground cost c is defined as

$$\mathcal{T}_c(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\pi,$$

where $\Pi(\mu, \nu)$ is the set of probability measures on $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν. Considering K different cost functions c_k, the barycentre problem can be written in this setting as

$$\bar{\mu} \in \underset{\mu}{\operatorname{argmin}} \sum_{k=1}^{K} \lambda_k \mathcal{T}_{c_k}(\mu, \nu_k). \tag{1}$$

When $(\mathcal{X}, d_{\mathcal{X}}) = (\mathcal{Y}, d_{\mathcal{Y}})$ is a Polish space and $c = d_{\mathcal{X}}^p$ with $p \geq 1$, $\mathrm{W}_p(\mu, \nu) := (\mathcal{T}_{d^p}(\mu, \nu))^{\frac{1}{p}}$ defines a distance between probability measures (with finite moment of order p), called p-Wasserstein distance. In this case, the barycentre $\bar{\mu}$ defined above is called a Wasserstein barycentre. Generalisation to a barycentre of a probability measure on $\mathcal{P}(\mathcal{X})$ and the consistency of their discrete approximations is also studied by several authors [AC17].

The theoretical analysis of Wasserstein barycentres begins with the foundational work by Carlier and Ekeland [CE10], who studied the existence, uniqueness and dual formulations for barycentre problems with generic continuous cost functions. Subsequent work by [AC11] re-established the existence and dual formulations of such barycentres for the quadratic Wasserstein distance W_2 on Euclidean spaces, and showed uniqueness under the hypothesis that one of the original measures is absolutely continuous. More recent studies have broadened these results: [CCE24] extended the theoretical analysis to Wasserstein medians (W_1), studying their stability properties, and investigated dual and multi-marginal formulations. [BFR24b] further extended the framework to W_p distances for $p > 1$, proving existence and uniqueness of barycentres for absolutely continuous measures on $\mathbb{R}^d$. A follow-up study by [BFR24a] analysed the general case for strictly convex and $\mathcal{C}^2$ cost functions with non-degenerate Hessian.

From a computational perspective, calculating Wasserstein barycentres is known to be a highly challenging problem, classified as NP-hard. According to [AB21], although polynomial-time algorithms exist for computing Wasserstein barycentres with a fixed number of points, their computational complexity scales exponentially with respect to the dimension of the space, or with respect to the number of marginals. This makes direct computation infeasible for high-dimensional problems or large sets of distributions, which are common in practical applications.

To tackle these computational challenges, several approximate methods have been developed for Wasserstein barycentres. The first paper to propose an algorithmic solution for computing these barycentres was by [Rab+12], which computed Sliced Wasserstein barycentres through a gradient descent approach. This method leveraged the sliced Wasserstein distance to achieve an efficient approximation, significantly simplifying computations.

A natural approach to develop easily computable approximations of such barycentres is to replace transport costs $\mathcal{T}_c$ by regularised versions

$$\mathcal{T}_{c,\varepsilon}(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\pi + \varepsilon\,\mathrm{KL}(\pi|\mu\otimes\nu),$$

as proposed in [CD14]. When the support of the distributions and barycentre is fixed (a grid for instance), the problem can be rewritten as a KL projection problem and the so-called entropic barycentre can be computed efficiently with a modified version of Sinkhorn's algorithm [Ben+15; PC19].

In order to deal with distributions without imposed support a second approach also described in [CD14] relies on a fixed-point algorithm inspired by the computation of Fréchet means on manifolds. Each step of this fixed point approach consists in replacing the current barycentre μ by its image measure by the map $\sum_{k=1}^K \lambda_k T_k$, where the T_k are optimal maps between μ and ν_k (assuming these maps exist). The authors of [Álv+16] were the first to establish a rigorous proof of convergence for this fixed-point approach in the case of absolutely continuous measures ν_k: more precisely, they proved convergence of a subsequence to a fixed point and showed that if the fixed point is unique, it is indeed a barycentre. Their study focuses specifically on the case of W_2 barycentres, with applications demonstrated mainly on Gaussian measures. Although their proof is only provided for absolutely continuous measures, this fixed point approach is frequently used for discrete measures and probably the baseline free-support method provided in numerical optimal transport libraries [Fla+21]. Building on the same ideas as [Álv+16], the author of [Lin23] extends the investigation of the fixed point algorithm for discrete measures on $\mathbb{R}^d$, limited to just one single iteration, and deriving a worst-case error bound in the W_2 and W_1 settings. The iterative solver of [Álv+16] has also been extended in high dimensional settings by [Kor+22], which use a neural solver for computing the optimal maps T_k.

In closely related directions, several other approaches have been proposed to compute Wasserstein barycentres over Riemannian manifolds [KP17], or Gromov-Wasserstein barycentres [BB25; BBS23] and the approach we develop in this paper share similarities with [BB25].

1.2 Contributions and Outline

In this paper, we develop a fixed-point approach to compute barycentres between probability measures for generic transport costs, i.e. solutions of the optimisation problem (1). Our only hypotheses are that we work on compact spaces, and that the ground costs c_k are continuous and such that $\mathrm{argmin}_x \sum_{k=1}^K \lambda_k c_k(x, x_k)$ is uniquely defined. In particular, we do not assume existence of optimal transport maps between μ and the ν_k, and we do not assume anything on the probability measures μ and ν_k. We propose an iterative fixed-point algorithm generalising [Álv+16] in this generic case. We show that the sequences generated by this algorithm have converging sub-sequences, that limits must be fixed-points of a certain mapping G, and that a barycentre for (1) is also a fixed point of G. We show that these results still hold for entropic regularised transport costs.

Numerically, we show that our approach specifically allows to extend the recent definition of generalised Wasserstein barycentres presented in [DGS21], notably by considering non-linear functions between the ambient space and the subspaces of measures ν_k. It also enables efficient computation of barycentres for the mixture Wasserstein metric [DD20], which until now were calculated using their multi-marginal equivalent formulation.

The paper is organised as follows. In Section 2, we introduce a novel notion of Optimal Transport barycentres in a certain space between measures ν_k on potentially different spaces for generic costs c_k. In Section 3, we propose a fixed-point algorithm which generalises [Álv+16] and converges to solutions (in a certain sense). We re-write the problem in a discrete setting in Section 4 and illustrate our method in Section 5 on several numerical examples, providing a publicly available Python toolkit.

2 Lifting Ground Barycentres to Measures

We work with probability measures ν_k on compact metric spaces $(\mathcal{Y}_k, d_{\mathcal{Y}_k})_{k\in[\![1,K]\!]}$, of which we will seek a "barycentre" μ in a compact metric space $(\mathcal{X}, d_{\mathcal{X}})$. To compare a measure $\nu_k \in \mathcal{P}(\mathcal{Y}_k)$ and $\mu \in \mathcal{P}(\mathcal{X})$ we consider continuous cost functions $c_k : \mathcal{X} \times \mathcal{Y}_k \longrightarrow \mathbb{R}_+$. A barycentre will be a minimiser of the sum of the transport costs with respect to the measure ν_k, leading to the following energy for a measure $\mu \in \mathcal{P}(\mathcal{X})$:

$$V(\mu) := \sum_{k=1}^K \mathcal{T}_{c_k}(\mu, \nu_k), \tag{2}$$

hence our minimisation problem reads

$$\underset{\mu\in\mathcal{P}(\mathcal{X})}{\mathrm{argmin}}\ V(\mu). \tag{3}$$

Note that to introduce barycentre weights λ_k, it suffices to replace c_k with $\lambda_k c_k$, which allows us to include weights in the costs and alleviate notation. We summarise our standing assumptions on the spaces and costs in Assumption 1:

Assumption 1. *The metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}_k, d_{\mathcal{Y}_k})$ are compact, and the costs $c_k : \mathcal{X} \times \mathcal{Y}_k \longrightarrow \mathbb{R}_+$ are continuous.*

Existence of solutions for Problem (3) was established by [CE10] Proposition 2 under Assumption 1.

Remark 2.1. *Uniqueness was proven in [CE10] Proposition 4 if, essentially, for at least one k, the problem $\mathcal{T}_{c_k}(\mu, \nu_k)$ has a Monge solution, for which they assume that each ν_k is absolutely continuous on $\mathcal{Y}_k = \overline{\Omega}$ with Ω an open and bounded subset of $\mathbb{R}^d$ with with $\nu_k(\partial\Omega) = 0$. They* *also assume that the costs $c_k(\cdot, y)$ are Lipschitz with a uniform constant L and that c_k verifies the Twist condition: $c_k(\cdot, y)$ is differentiable, with $\partial_x c_k(x, \cdot)$ injective.*

The definition of a barycentre between measures ν_k can be seen as a lifting of a notion of barycentre within $\mathcal{X}$ of points $(y_1, \cdots, y_K) \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K$. To give mathematical meaning to this intuition and to our method, we will make the following assumption throughout the paper:

Assumption 2. *For all $(y_1, \cdots, y_K) \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K$, the set $\underset{x\in\mathcal{X}}{\operatorname{argmin}} \sum_{k=1}^K c_k(x, y_k)$ has a unique element.*

The uniqueness of the optimisation problem in Assumption 2 allows us to introduce the ground barycentre function B:

$$B : \left\{ \begin{array}{ccc} \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K & \longrightarrow & \mathcal{X} \\ (y_1, \cdots, y_K) & \longmapsto & \underset{x\in\mathcal{X}}{\operatorname{argmin}} \sum_{k=1}^K c_k(x, y_k). \end{array} \right. \tag{4}$$

For convenience, we introduce $\mathcal{Y} := \Pi_k \mathcal{Y}_k$, equipped with the product distance, with the notation $Y := (y_1, \cdots, y_K)$ for an element of $\mathcal{Y}$, as well as the total cost function:

$$C := \left\{ \begin{array}{ccc} \mathcal{X} \times \mathcal{Y} & \longrightarrow & \mathbb{R}_+ \\ (x, y_1, \cdots, y_K) & \longmapsto & \sum_{k=1}^K c_k(x, y_k) \end{array} \right. . \tag{5}$$

Equipped with these convenient notations, we can write the multi-marginal formulation of our barycentre problem:

$$\underset{\pi\in\Pi(\nu_1,\cdots,\nu_K)}{\operatorname{argmin}} \int_{\mathcal{Y}} C(B(Y), Y)\mathrm{d}\pi(Y). \tag{6}$$

The barycentre problem defined in Eq. (3) is related to the multi-marginal formulation through the following equation, due to [CE10], Proposition 3.3:

$$\underset{\mu\in\mathcal{P}(\mathcal{X})}{\operatorname{argmin}} V(\mu) = B\# \underset{\pi\in\Pi(\nu_1,\cdots,\nu_K)}{\operatorname{argmin}} \int_{\mathcal{Y}} C(B(Y), Y)\mathrm{d}\pi(Y). \tag{7}$$

The following technical result uses the continuity of the c_k and Assumption 2 to show that B is continuous.

Lemma 2.2. *The function $B : \mathcal{Y} \longrightarrow \mathcal{X}$ defined in Eq. (4) is continuous.*

Proof. The proof uses standard compactness arguments, showing that for $Y_n \xrightarrow[n\longrightarrow+\infty]{} Y \in \mathcal{Y}$, $(B(Y_n))$ can only have $B(Y)$ as a subsequential limit. □

Another important technical result is the regularity of transport costs, which we will use repeatedly. We gather well-known results in Lemma 2.3.

Lemma 2.3. *Consider E, F compact metric spaces and let $c : E \times F \longrightarrow \mathbb{R}_+$ a measurable cost function. The optimal transport cost $\mathcal{T}_c$ has the following regularity for the weak convergence of measures depending on c:*

1. *If c is lower-semi-continuous, then $\mathcal{T}_c$ is lower-semi-continuous.*
2. *If c is continuous, then $\mathcal{T}_c$ is continuous.*
3. *If $E = F$ and c is l.s.c. with $c(x, y) = 0 \Longrightarrow x = y$, then $\mathcal{T}_c(\mu, \nu) = 0 \Longrightarrow \mu = \nu$.*

Proof. Regarding item 1), by [San15] Theorem 1.42, Kantorovich duality holds for c l.s.c. and thus $\mathcal{T}_c$ can be written as a supremum of l.s.c. functions, hence is l.s.c.. For item 2), the result is verbatim [San15] Theorem 1.51. For item 3), if $\mathcal{T}_c(\mu,\nu) = 0$ then there exists $\pi \in \Pi(\mu,\nu)$ such that $\int_{E^2} c(x,y)\mathrm{d}\pi(x,y) = 0$ (existence follows from lower semi-continuity, as in [San15] Theorem 1.5). Thus for π-almost-every (x,y), $c(x,y) = 0$, which by assumption gives $x = y$, hence (using the same technique as in [San15] Proposition 5.1) for any test function $\phi \in \mathcal{C}^0(E,\mathbb{R})$:

$$\int_E \phi(x)\mathrm{d}\mu(x) = \int_{E^2} \phi(x)\mathrm{d}\pi(x,y) = \int_{E^2} \phi(y)\mathrm{d}\pi(x,y) = \int_E \phi(y)\mathrm{d}\nu(y),$$

which shows that $\mu = \nu$. □

3 A Fixed-Point Algorithm

3.1 Algorithm Definition

In this section, we define a sequence $(\mu_t) \in \mathcal{P}(\mathcal{X})^{\mathbb{N}}$ that will approach a barycentre of fixed measures $\nu_k \in \mathcal{P}(\mathcal{Y}_k)$. We propose a modified version of the iterated scheme from [Álv+16] to solve Eq. (3). To define an iteration mapping, for $\mu \in \mathcal{P}(\mathcal{X})$, we consider the set of multi-marginal couplings

$$\Gamma(\mu) := \left\{\gamma \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}_1\times\cdots\times\mathcal{Y}_K)\ :\ \forall k \in [\![1,K]\!],\ \gamma_{0,k} \in \Pi^*_{c_k}(\mu,\nu_k)\right\}, \tag{8}$$

where, for all k, $\gamma_{0,k}$ denotes the $\mathcal{X}\times\mathcal{Y}_k$ marginal of γ and $\Pi^*_{c_k}(\mu,\nu_k)$ denotes the set of all optimal couplings for the transport problem between μ and ν_k associated to the cost function c_k. The existence of such multi-couplings is a consequence of the well-known "gluing lemma" (see [San15] Lemma 5.5). The following multi-coupling provides an explicit element of $\Gamma(\mu)$ given $\pi_k \in \Pi^*_{c_k}(\mu,\nu_k)$:

$$\gamma(\mathrm{d}x,\mathrm{d}y_1,\cdots,\mathrm{d}y_K) := \mu(\mathrm{d}x)\pi_1^x(\mathrm{d}y_1)\cdots\pi_K^x(\mathrm{d}y_K), \tag{9}$$

where we wrote the disintegration of π_k with respect to its first marginal μ as $\pi_k(\mathrm{d}x,\mathrm{d}y_k) = \mu(\mathrm{d}x)\pi_k^x(\mathrm{d}y_k)$. By abuse of notation, we will denote $B\#\gamma := B\#\gamma_{1,\cdots,K}$, where $\gamma_{1,\cdots,K} \in \mathcal{P}(\mathcal{Y}_1\times\cdots\times\mathcal{Y}_k)$ is the marginal of γ with respect to $(y_1,\cdots,y_K)$. In terms of random variables, if $(X,Y_1,\cdots,Y_K)\sim\gamma$, then $B\#\gamma = \mathrm{Law}[B(Y_1,\cdots,Y_K)]$. Denoting $B\#\Gamma(\mu) := \{B\#\gamma,\ \gamma\in\Gamma(\mu)\}$, we define the multi-valued mapping G which maps $\mu \in \mathcal{P}(\mathcal{X})$ to the set of next iterates $G(\mu) \subset \mathcal{P}(\mathcal{X})$:

$$G := \left\{\begin{array}{ccc} \mathcal{P}(\mathcal{X}) & \rightrightarrows & \mathcal{P}(\mathcal{X}) \\ \mu & \mapsto & B\#\Gamma(\mu) \end{array}\right. . \tag{10}$$

Note that this construction is similar to that of [Álv+16], Remark 3.4. Moreover, the candidate barycentre $\overline{\mu} = B\#\gamma_{1,\cdots,K}$ is closely related to the multi-marginal formulation of the barycentre problem (see Eq. (7)). Indeed, set $\pi := \gamma_{1,\cdots,K} \in \Pi(\mu_1,\cdots,\mu_K)$, notice that π is a candidate for the multi-marginal problem of a particular structure induced by the reference measure μ. In the case where the plans $\gamma_{0,k}$ are induced by maps T_k, then this structure is the coupling $(T_1,\cdots,T_K)\#\mu$. In terms of random variables, if $X\sim\mu$, then the chosen coupling is $(T_1(X),\cdots,T_K(X))$.

Taking inspiration from the W_2^2 case, we can see informally the iterate $\overline{\mu} \in G(\mu)$ as a local linearisation of $\mathcal{P}(\mathcal{X})$. To illustrate this intuition, we consider the case $\mathcal{X} = \mathcal{Y}_1 = \cdots = \mathcal{Y}_K$ and assume that for each k, the set of optimal plans $\Pi^*_{c_k}(\mu,\nu_k)$ is reduced to (I,T_k), or in other words, that the Monge problem has a unique solution. Informally, one may see the set of maps $T : \mathcal{X} \longrightarrow \mathcal{X}$ sending μ to a measure $T\#\mu \in \mathcal{P}(\mathcal{X})$ as the tangent space to $\mathcal{P}(\mathcal{X})$ at μ. As a result, the problem of finding a barycentre $\overline{\mu}$ can be seen from the viewpoint of the reference measure μ in the tangent space $T_\mu\mathcal{P}(\mathcal{X})$ as the problem of finding $S \in T_\mu\mathcal{P}(\mathcal{X})$ such that $S\#\mu$ would minimise the cost V. Our approach takes a barycentre of the optimal maps T_k by choosing the candidate $S := B \circ (T_1, \cdots, T_K)$. In the case of the squared-Euclidean cost on the common space $\mathbb{R}^d$, this amounts to $S := \sum_k \lambda_k T_k$, which is exactly the Linearised Optimal Transport barycentre approximation for the reference measure μ, as introduced in [MDC20], Section 4.3. We illustrate this viewpoint schematically in Fig. 1.

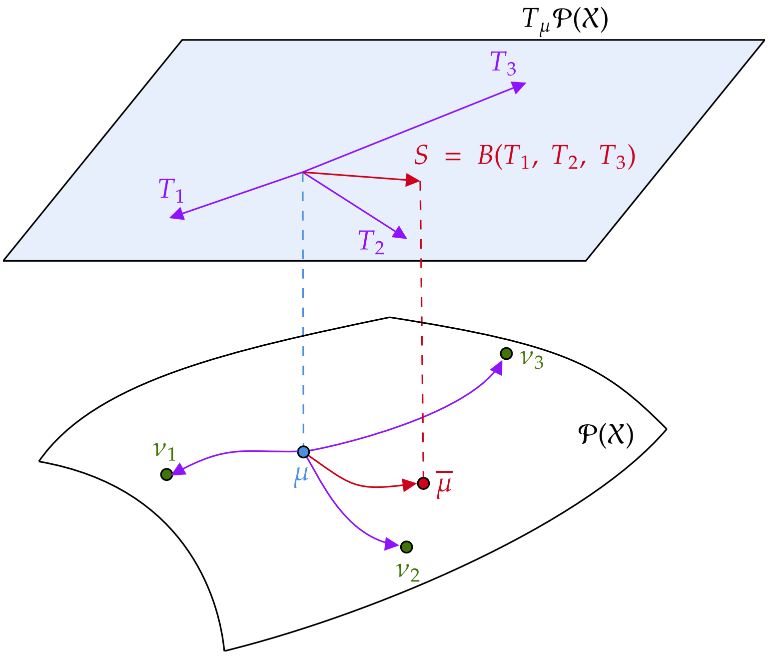


Figure 1: Illustration of the informal linearisation interpretation for the barycentre candidate $\overline{\mu} = B \circ (T_1, \cdots, T_K)\#\mu$.

Starting from a measure $\mu_0 \in \mathcal{P}(\mathcal{X})$, our algorithm consists of choosing iterates through the multi-function G:

$$\forall t \in \mathbb{N},\ \mu_{t+1} \in G(\mu_t).$$

We dedicate the next section to a theoretical study of the convergence of this fixed-point iteration.

3.2 Convergence of Fixed-Point Iterations

We can formulate a regularity result of the multi-valued map G: namely, we will show that G is *upper hemi-continuous*. For the sake of simplicity, we will take the following definition[1]:

Definition 3.1. *A multi-valued function $\varphi : E \rightrightarrows F$ from a compact metric space E to parts of a compact metric space F is said to be upper hemi-continuous (u.h.c.) if for any sequence $(x_n, y_n) \in (E \times F)^{\mathbb{N}}$ such that $y_n \in \varphi(x_n)$ and $x_n \xrightarrow[n\longrightarrow+\infty]{} x \in E$, there exists an extraction such that $y_{\alpha(n)} \xrightarrow[n\longrightarrow+\infty]{} y \in F$ with $y \in \varphi(x)$.*

For more technical reasons, we also need to introduce the notion of *lower hemi-continuity*[2]

[1] We refer to [AB94] Chapter 17 for a more general definition and introduction to these concepts on Polish spaces. We choose a stronger sequential definition from [AB94] Theorem 17.20, which in their vocabulary corresponds to u.h.c multi-functions with compact values.

[2] whose formulation is is equivalent to [AB94], Definition 17.2 by their Theorem 17.21.

Definition 3.2. *A multi-valued function $\varphi : E \rightrightarrows F$ from a compact metric space space E to parts of a compact metric space space F is said to be lower hemi-continuous (l.h.c.) if for any sequence $(x_n) \in E^{\mathbb{N}}$ such that $x_n \xrightarrow[n\longrightarrow+\infty]{} x \in E$, then for any $y \in F$ such that $y \in \varphi(x)$, there exists an extraction α and a sequence $(y_n) \in F^{\mathbb{N}}$ such that $y_n \in \varphi(x_{\alpha(n)})$ and $y_n \xrightarrow[n\longrightarrow+\infty]{} y$.*

To illustrate the technical differences between these two notions, we consider two specific multi-valued functions in Fig. 2. Finally, an hemi-continuous multi-map is one that is both

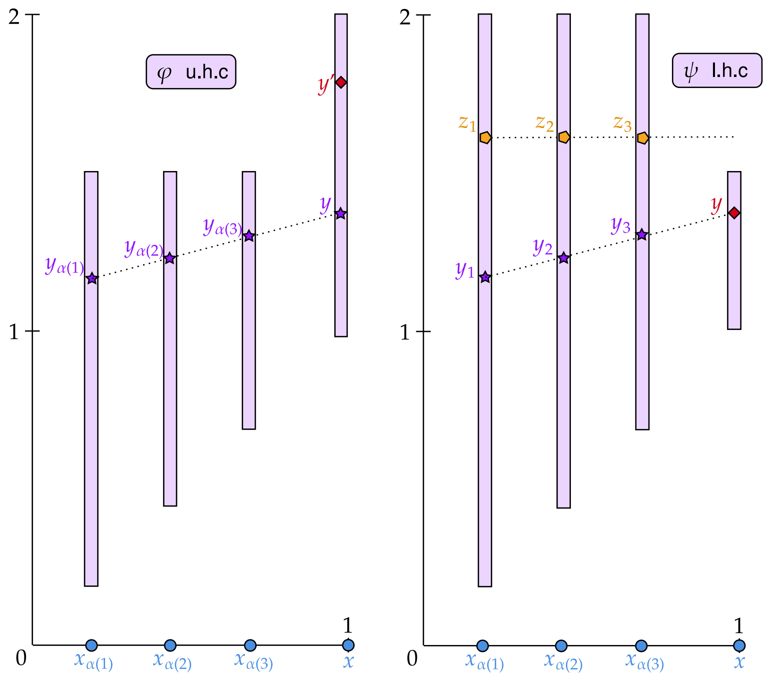


Figure 2: *Left:* the multi-function $\varphi : [0,1] \rightrightarrows [0,2]$ defined by $\forall x \in [0,1),\ \varphi(x) = [x, 3/2]$ and $\varphi(1) = [1,2]$ is u.h.c.. Indeed, taking any sequence (x_n, y_n) such that $y_n \in \varphi(x_n)$ and $x_n \xrightarrow[n\longrightarrow+\infty]{} x$, there exists an extraction α such that $y_{\alpha(n)} \xrightarrow[n\longrightarrow+\infty]{} y \in \varphi(x)$. However, φ is not l.h.c. at 1 since the target $y' := 7/4 \in \varphi(1)$ can never be a limit of a sequence (x_n, y_n) with $x_n \xrightarrow[n\longrightarrow+\infty]{} 1$ and $y_n \in \varphi(x_n)$.
Right: $\psi : [0,1] \rightrightarrows [0,2]$ defined by $\forall x \in [0,1),\ \psi(x) = [x,2]$ and $\psi(1) = [1,3/2]$ is l.h.c.. Take $x_n \xrightarrow[n\longrightarrow+\infty]{} x$ and a target $y \in \psi(x)$. Then there exists an extraction α and a sequence (y_n) such that $y_n \in \psi(x_n)$ and $y_n \xrightarrow[n\longrightarrow+\infty]{} y$. However, ψ is not u.h.c: take $x_n \xrightarrow[n\longrightarrow+\infty]{} 1$ and the sequence $z_n := 5/3$. We have $\forall n \in \mathbb{N},\ z_n \in \psi(x_n)$, however any subsequence of (z_n) converges to $5/3 \notin \psi(1)$.

u.h.c. and l.h.c.:
Definition 3.3. *A multi-valued function $\varphi : E \rightrightarrows F$ from a compact metric space space E to parts of a compact metric space space F is said to be hemi-continuous if it is both u.h.c. (Definition 3.1) and l.h.c. (Definition 3.2).*

We begin with technical lemmas on the hemi-continuity properties of sets of couplings.
Lemma 3.4. *Consider E, F compact metric spaces and $\nu \in \mathcal{P}(F)$. The multi-function*

$$\Pi_\nu := \left\{ \begin{array}{ccc} \mathcal{P}(E) & \rightrightarrows & \mathcal{P}(E \times F) \\ \mu & \mapsto & \Pi(\mu, \nu) \end{array} \right. \tag{11}$$

is hemi-continuous.

Proof. **u.h.c..** We apply Definition 3.1: introduce $\mu_n \xrightarrow[n\longrightarrow+\infty]{w} \mu \in \mathcal{P}(E)$ and $\pi_n \in \Pi(\mu_n, \nu)$. Since $\mathcal{P}(E\times F)$ is compact, we can introduce α an extraction such that $\pi_{\alpha(n)} \xrightarrow[n\longrightarrow+\infty]{w} \pi \in \mathcal{P}(E\times F)$. By continuity of marginalisation, we deduce $\pi \in \Pi(\mu,\nu)$, which shows that Π_ν is u.h.c. by definition.

l.h.c.. We consider W_1, the 1-Wasserstein distance on $\mathcal{P}(E)$ (i.e. $\mathcal{T}_{d_E}$), and use the same notation for the 1-Wasserstein distance on $\mathcal{P}(E^2)$, with the distance $d_{E^2}((x,y),(x',y')) := \max(d_E(x,x'), d_E(y,y'))$, both of which metrise the weak convergence by [Vil09] Corollary 6.13. We apply Definition 3.2: take $\mu_n \xrightarrow[n\longrightarrow+\infty]{w} \mu \in \mathcal{P}(E)$, and let $\pi \in \Pi(\mu,\nu)$. Consider (X,Y) two coupled random variables of law π, and for $n\in\mathbb{N}$, take X_n a random variable such that (X,X_n) is an optimal coupling for $\mathrm{W}_1(\mu,\mu_n)$, and let $\pi_n := \mathrm{Law}(X_n,Y)$. We have

$$\mathrm{W}_1(\pi,\pi_n) \leq \mathbb{E}\left[d_{E^2}\left((X,Y),(X_n,Y)\right)\right] = \mathbb{E}\left[\max(d_E(X,X_n), d_E(Y,Y))\right] = \mathrm{W}_1(\mu,\mu_n),$$

then by metrisation, we get $\mathrm{W}_1(\mu,\mu_n) \xrightarrow[n\longrightarrow+\infty]{} 0$, then $\pi_n \xrightarrow[n\longrightarrow+\infty]{w} \pi$, concluding the proof that Π_ν is l.h.c.. □

We can apply Berge's maximisation theorem to show that the set of *optimal* transport plans is upper hemi-continuous for a continuous cost function:

Lemma 3.5. *Consider E, F compact metric spaces, a continuous cost $c : E\times F \longrightarrow \mathbb{R}_+$ and $\nu\in\mathcal{P}(F)$. The multi-function*

$$[\Pi_c^*]_\nu := \left\{ \begin{array}{ccc} \mathcal{P}(E) & \rightrightarrows & \mathcal{P}(E\times F) \\ \mu & \mapsto & \Pi_c^*(\mu,\nu) \end{array} \right. \tag{12}$$

is upper hemi-continuous.

Proof. By compactness, the map $\pi \longmapsto \int_{E\times F} c\mathrm{d}\pi$ is continuous, and by Lemma 3.4, the multi-map $\mu \rightrightarrows \Pi(\mu,\nu)$ is hemi-continuous (with compact values), hence by Berge's maximisation theorem from [AB94] Theorem 17.31, the map

$$[\Pi_c^*]_\nu : \mu \longmapsto \Pi_c^*(\mu,\nu) = \underset{\pi\in\Pi(\mu,\nu)}{\mathrm{argmin}} \int_{E\times F} c\mathrm{d}\pi$$

is upper hemi-continuous. □

Remark 3.6. *The multifunction $[\Pi_c^*]_\nu$ is not **lower** hemi-continuous. Indeed, take the following points of $\mathbb{R}^2$:*

$$\forall n\in\mathbb{N},\ x_n := (-1, 2^{-n}),\ y_n := (1, -2^{-n}),\ x_\infty := (-1,0),\ y_\infty := (1,0),\ w := (0,1),\ z := (0,-1),$$

and the following discrete measures (see Fig. 3):

$$\forall n\in\mathbb{N},\ \mu_n := \frac{1}{2}(\delta_{x_n} + \delta_{y_n}),\ \mu_\infty := \frac{1}{2}(\delta_{x_\infty} + \delta_{y_\infty}),\ \nu := \frac{1}{2}(\delta_w + \delta_z).$$

We have $\mu_n \xrightarrow[n\longrightarrow+\infty]{w} \mu_\infty$, and a unique OT plan for the cost $c(\cdot,\cdot) := \|\cdot - \cdot\|_2^2$ between μ_n and ν, which sends x_n to w and y_n to z:

$$\forall n\in\mathbb{N},\ \Pi_c^*(\mu_n,\nu) = \{\pi_n\},\ \pi_n := \frac{1}{2}\left(\delta_{x_n,w} + \delta_{y_n,z}\right),$$

with $\pi_n \xrightarrow[n\longrightarrow+\infty]{w} \pi_\infty := \frac{1}{2}(\delta_{x_\infty,w} + \delta_{y_\infty,z})$*. However, the set of optimal plans between the limit* μ_∞ *and* ν *has more than one element, since* $\|x_\infty - w\|_2^2 = \|x_\infty - z\|_2^2$ *and* $\|y_\infty - w\|_2^2 = \|y_\infty - z\|_2^2$*:*

$$\Pi_c^*(\mu, \nu) = \{(1-t)\pi_\infty + t\pi',\ t \in [0,1]\}\,,\ \pi' := \frac{1}{2}(\delta_{x_\infty,z} + \delta_{y,w})\,.$$

We conclude that there does not exist an extraction α *and a sequence* (π_n') *such that* $\forall n \in \mathbb{N},\ \pi_n' \in \Pi_c^*(\mu_{\alpha(n)}, \nu)$ *and* $\pi_n' \xrightarrow[n\longrightarrow+\infty]{w} \pi'$*.*

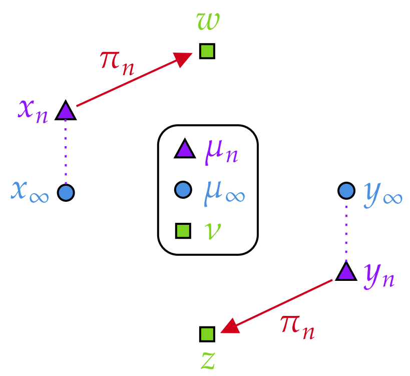


Figure 3: Counter-example from Remark 3.6 showing that $\Pi_c^*(\cdot, \nu)$ is not lower hemi-continuous in general.

A direct corollary of Lemma 3.5 is the upper hemi-continuity of Γ and G. For notational convenience, we introduce $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K$.

Proposition 3.7. *The multi-map*

$$\Gamma := \left\{ \begin{array}{ccc} \mathcal{P}(\mathcal{X}) & \rightrightarrows & \mathcal{P}(\mathcal{Z}) \\ \mu & \mapsto & \Gamma(\mu) \end{array} \right.$$

where $\Gamma(\mu)$ *is defined in Eq. (8) and* G *defined in Eq. (10) are upper hemi-continuous (and compact-valued).*

Proof. Let $\mu \in \mathcal{P}(\mathcal{X})$. To show that $G(\mu)$ and $\Gamma(\mu)$ are compact, it suffices to show that $\Gamma(\mu)$ is closed, since $\mathcal{P}(\mathcal{Z})$ is compact, and $G(\mu) = B\#\Gamma(\mu)$ with B continuous by Lemma 2.2. Take $(\gamma_n) \in \Gamma(\mu)^{\mathbb{N}}$ such that $\gamma_n \xrightarrow[n\longrightarrow+\infty]{} \gamma \in \mathcal{P}(\mathcal{Z})$. We show that $\gamma \in \Gamma(\mu)$. For $k \in [\![1, K]\!]$ and $n \in \mathbb{N}$, we have $\gamma_n \in \Gamma(\mu)$, hence $[\gamma_n]_{0,k} \in \Pi_{c_k}^*(\mu, \nu_k)$. By continuity of marginalisation, we deduce that $\gamma \in \Pi(\mu, \nu_1, \cdots, \nu_K)$. By continuity of $\pi \longmapsto \int_{\mathcal{X}\times\mathcal{Y}_k} c_k \mathrm{d}\pi$ (which holds by compactness), we deduce that $\gamma_{0,k} \in \Pi_{c_k}^*(\mu, \nu_k)$, hence $\gamma \in \Gamma(\mu)$.

For the u.h.c. of Γ, take a sequence $(\mu_n) \in \mathcal{P}(\mathcal{X})^{\mathbb{N}}$ such that $\mu_n \xrightarrow[n\longrightarrow+\infty]{w} \mu \in \mathcal{P}(\mathcal{X})$, and take a sequence $(\gamma_n) \in \mathcal{P}(\mathcal{Z})^{\mathbb{N}}$, with $\gamma_n \in \Gamma(\mu_n)$. Since $\gamma_n \in \mathcal{P}(\mathcal{Z})$ which is compact, take α an extraction such that $\gamma_{\alpha(n)} \xrightarrow[n\longrightarrow+\infty]{w} \gamma \in \mathcal{P}(\mathcal{Z})$. We will show that $\gamma \in \Gamma(\mu)$.

Start with $k := 1$. For $n \in \mathbb{N}$, we have $\gamma_{\alpha(n)} \in \Gamma(\mu_{\alpha(n)})$, hence $\pi_{\alpha(n)}^{(1)} := [\gamma_{\alpha(n)}]_{0,1} \in \Pi_{c_1}^*(\mu_{\alpha(n)}, \nu_1)$. By Lemma 3.5, the map $\mu \longmapsto \Pi_{c_1}^*(\mu, \nu_1)$ is u.h.c., hence by definition, since $\mu_{\alpha(n)} \xrightarrow[n\longrightarrow+\infty]{w} \mu \in \mathcal{P}(\mathcal{X})$ and $\pi_{\alpha(n)}^{(1)} \in \Pi_{c_1}^*(\mu_{\alpha(n)}, \nu_1)$, there exists an extraction α_1 such that $\pi_{\alpha\circ\alpha_1(n)}^{(1)} \xrightarrow[n\longrightarrow+\infty]{w} \pi^{(1)} \in \Pi_{c_1}^*(\mu, \nu_1)$.

Continuing this method for $k \in [\![2, K]\!]$ with successive sub-extractions α_k, setting $\beta := \alpha \circ \alpha_1 \circ \cdots \circ \alpha_K$, we have for any $k \in [\![1, K]\!]$, $[\gamma_{\beta(n)}]_{0,k} = \pi_{\beta(n)}^{(k)} \xrightarrow[n\longrightarrow+\infty]{w} \pi^{(k)} \in \Pi_{c_k}^*(\mu, \nu_k)$. The continuity of marginalisation implies $\gamma_{0,k} = \pi^{(k)}$, and in turn shows that $\gamma \in \Gamma(\mu)$, concluding that Γ is u.h.c.

For G, the fact that $G(\mu) = B\#\Gamma(\mu)$ and the continuity of B prove that G is u.h.c. using the u.h.c. of Γ by [AB94] Theorem 17.23. □

In order to study the energy of iterates of G, we first require a technical result on the error of sub-optimal ground barycentres for B. We introduce a radius constant $R := \max_{(x,Y)\in\mathcal{X}\times\mathcal{Y}} d_{\mathcal{X}}(x, B(Y))$, which is finite since $\mathcal{X}$ and $\mathcal{Y}$ are compact, and B is continuous. We need to make a trivial assumption to ensure that $R > 0$:

Assumption 3. *There exists $x \in \mathcal{X}$ and $Y \in \mathcal{Y}$ such that $x \neq B(Y)$.*

Lemma 3.8 is a generalisation of the following elementary Euclidean property in $\mathbb{R}^d$ for the cost $\|\cdot - \cdot\|_2^2$, for which $B(y_1, \cdots, y_K) = \sum_{k=1}^K \lambda_k y_k$ verifies the following identity:

$$\forall x \in \mathbb{R}^d,\ \forall (y_1, \cdots, y_K) \in (\mathbb{R}^d)^K,\ \overline{x} := \sum_{k=1}^K \lambda_k y_k :\ \sum_{k=1}^K \lambda_k \|x - y_k\|_2^2 = \sum_{k=1}^K \lambda_k \|\overline{x} - y_k\|_2^2 + \|x - \overline{x}\|_2^2.$$

Lemma 3.8. *There exists a function $\delta = \eta \circ d_{\mathcal{X}}$, with $\eta : [0, R] \longrightarrow \mathbb{R}_+$ lower-semi-continuous, non-decreasing and verifying $\eta(s) = 0 \iff s = 0$, such that*

$$\forall (x, Y) \in \mathcal{X} \times \mathcal{Y},\ C(x, Y) \geq C(B(Y), Y) + \delta(x, B(Y)). \tag{13}$$

Proof. — *Step 1*: Definition of η. First, for $(x, Y) \in \mathcal{X} \times \mathcal{Y}$, let $\Delta(x, Y) := C(x, Y) - C(B(Y), Y)$. By definition of B, $\Delta(x, Y) \geq 0$, and $\Delta(x, Y) = 0 \iff x = B(Y)$. By assumption, B and C are continuous, which implies that Δ is also continuous.

We now introduce $S := \max_{(x,Y)\in\mathcal{X}\times\mathcal{Y}} \Delta(x, Y)$. Assumption 3 ensures $S > 0$. Define now the function η:

$$\eta := \left\{ \begin{array}{ccc} [0, R] & \longrightarrow & [0, S] \\ u & \longmapsto & \min\limits_{(x,Y)\in\mathcal{X}\times\mathcal{Y}} \{\Delta(x, Y)\ :\ d_{\mathcal{X}}(x, B(Y)) \geq u\} \end{array} \right. . \tag{14}$$

We show that for $u \in [0, R]$, the infimum is attained. First, let $f := (x, Y) \longmapsto d_{\mathcal{X}}(x, B(Y))$, we remark that

$$\forall (x, Y) \in \mathcal{X} \times \mathcal{Y},\ d_{\mathcal{X}}(x, B(Y)) \geq u \iff (x, Y) \in f^{-1}([u, R]).$$

By continuity of f and compactness of $\mathcal{X} \times \mathcal{Y}$, $\mathcal{K}_u := f^{-1}([u, R])$ is a compact subset of $\mathcal{X} \times \mathcal{Y}$. $\mathcal{K}_u$ is not empty since there exists $(x_R, Y_R) \in \mathcal{X} \times \mathcal{Y}$ such that $d_{\mathcal{X}}(x_R, B(Y_R)) = R$ (by continuity, compactness and definition of R).

— *Step 2*: Proof of Eq. (13). Let $(x, Y) \in \mathcal{X} \times \mathcal{Y}$, and $u := d_{\mathcal{X}}(x, B(Y))$. By definition, $(x, Y) \in \mathcal{K}_u$, hence $\eta(u) \leq \Delta(x, Y)$, which is equivalent to Eq. (13).

— *Step 3*: Lower semi-continuity of η. Let $u_n \xrightarrow[n\longrightarrow+\infty]{} u \in [0, R]$, and for $n \in \mathbb{N}$ introduce $(x_n, Y_n) \in \mathcal{K}_{u_n}$ such that $\eta(u_n) = \Delta(x_n, Y_n)$. Since $(\eta(u_n)) \in [0, S]^{\mathbb{N}}$, consider an extraction α such that $\eta(u_{\alpha(n)}) \xrightarrow[n\longrightarrow+\infty]{} a_\alpha \in [0, S]$. By compactness of $\mathcal{X} \times \mathcal{Y}$, we can extract from $(x_{\alpha(n)}, Y_{\alpha(n)})_n$ a subsequence such that $(x_{\alpha\circ\beta(n)}, Y_{\alpha\circ\beta(n)}) \xrightarrow[n\longrightarrow+\infty]{} (x_{\alpha,\beta}, Y_{\alpha,\beta}) \in \mathcal{X} \times \mathcal{Y}$. By construction of the sequence $(x_n, Y_n)_n$, we have

$$\forall n \in \mathbb{N},\ d_{\mathcal{X}}(x_{\alpha\circ\beta(n)}, B(Y_{\alpha\circ\beta(n)})) \geq u_{\alpha\circ\beta(n)}, \tag{15}$$

since $(x_{\alpha\circ\beta(n)}, Y_{\alpha\circ\beta(n)}) \in \mathcal{K}_{u_{\alpha\circ\beta(n)}}$. Taking the limit in Eq. (15) yields $d_{\mathcal{X}}(x_{\alpha,\beta}, B(Y_{\alpha,\beta})) \geq u$, by continuity of B, Lemma 2.2. This shows that $(x_{\alpha,\beta}, Y_{\alpha,\beta}) \in \mathcal{K}_u$, hence $\eta(u) \leq \Delta(x_{\alpha,\beta}, Y_{\alpha,\beta})$. However, by continuity of Δ, and since $\Delta(x_{\alpha(n)}, Y_{\alpha(n)}) \xrightarrow[n\longrightarrow+\infty]{} a_\alpha$, it follows that $\Delta(x_{\alpha,\beta}, Y_{\alpha,\beta}) = a_\alpha$. Since the subsequential limit a_α was chosen arbitrarily, we conclude that $\eta(u) \leq \liminf_{n\longrightarrow+\infty} \eta(u_n)$, hence η is lower semi-continuous.

— *Step 4*: η is non-decreasing. Let $0 \leq u \leq v \leq R$, we have $\mathcal{K}_v \subset \mathcal{K}_u$, hence

$$\eta(u) = \min_{(x,Y)\in\mathcal{K}_u} \Delta(x, Y) \leq \min_{(x,Y)\in\mathcal{K}_v} \Delta(x, Y) = \eta(v).$$

— *Step 5*: Separation property. Let $u \in [0, R]$ such that $\eta(u) = 0$. This implies that there exists $(x, Y) \in \mathcal{X} \times \mathcal{Y}$ such that $\Delta(x, Y) = 0$ and $d_{\mathcal{X}}(x, B(Y)) \geq u$. Now by Step 1 this implies $x = B(Y)$, thus $d_{\mathcal{X}}(x, B(Y)) = 0$ and finally $u = 0$. □

Given the inequality in Eq. (13), we can now find an informative inequality between $V(\overline{\mu})$ and $V(\mu)$ for any $\overline{\mu} \in G(\mu)$. Applying Proposition 3.9 to the W_2 case for absolutely continuous measures yields [Álv+16] Proposition 4.3, wherein the cost $\mathcal{T}_\delta$ is simply W_2^2. This decrease was also studied by [Lin23] (Proposition 4.4) in the discrete setting W_p^p.

Proposition 3.9. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\overline{\mu} \in G(\mu)$. Then $V(\mu) \geq V(\overline{\mu}) + \mathcal{T}_\delta(\mu, \overline{\mu})$. If μ^* is a barycentre, then $G(\mu^*) = \{\mu^*\}$.*

Proof. Let $\overline{\mu} = B\#\gamma \in G(\mu)$ with $\gamma \in \Gamma(\mu)$, by definition of $\mathcal{T}_{c_k}$ and by optimality of the bi-marginals $\gamma_{0,k}$ of γ:

$$\begin{aligned}
\sum_{k=1}^K \mathcal{T}_{c_k}(\mu, \nu_k) &= \int_{\mathcal{X}\times\mathcal{Y}} C(x, Y)\mathrm{d}\gamma(x, Y) &(16)\\
&\geq \int_{\mathcal{X}\times\mathcal{Y}} (C(B(Y), Y) + \delta(x, B(Y)))\mathrm{d}\gamma(x, Y) &(17)\\
&\geq \sum_{k=1}^K \mathcal{T}_{c_k}(B\#\gamma, \nu_k) + \mathcal{T}_\delta(\mu, B\#\gamma) &(18)\\
&= V(\overline{\mu}) + \mathcal{T}_\delta(\mu, \overline{\mu}). &(19)
\end{aligned}$$

The inequality in Eq. (17) comes from Lemma 3.8, and the inequality in Eq. (18) comes from the definition of $\Gamma(\mu)$ (Eq. (8)), which allows us to write for $k \in [\![1, K]\!]$:

$$\int_{\mathcal{X}\times\mathcal{Y}} c_k(B(Y), y_k)\mathrm{d}\gamma(x, Y) = \int_{\mathcal{X}\times\mathcal{Y}_k} c_k \mathrm{d}\pi_k,$$

where we introduce the coupling $\pi_k := (B, P_k)\#[\gamma_{1,\cdots,K}]$, with $P_k(y_1, \cdots, y_K) = y_k$. The first marginal of π is $B\#[\gamma_{1,\cdots,K}]$ (which we write $B\#\gamma$ for legibility), and the second marginal is ν_k. Similarly,

$$\int_{\mathcal{X}\times\mathcal{Y}} \delta(x, B(Y))\mathrm{d}\gamma(x, Y) = \int_{\mathcal{X}\times\mathcal{X}} \delta \mathrm{d}[(I, B)\#\gamma] \geq \mathcal{T}_\delta(\mu, B\#\gamma).$$

If μ^* is a barycentre, then by definition for any $\overline{\mu} \in G(\mu)$, we have $V(\overline{\mu}) \geq V(\mu^*)$, thus Eqs. (17) and (18) are equalities, and $\mathcal{T}_\delta(\mu^*, \overline{\mu}) = 0$. By Lemmas 2.3 and 3.8, the cost δ guarantees the separation property of the transport cost $\mathcal{T}_\delta$, hence $\mu^* = \overline{\mu}$. □

The inequality in Proposition 3.9 shows that the amount of decrease in the energy between two iterations is lower-bounded by a transport discrepancy $\mathcal{T}_\delta$ (we remind that in the squared-Euclidean case, $\mathcal{T}_\delta = \mathrm{W}_2^2$). We can now show convergence of iterates of G, in the sense that any weakly converging subsequence converges towards a fixed point of G.

Theorem 3.10. *For any $\mu_0 \in \mathcal{P}(\mathcal{X})$, let (μ_t) verifying $\mu_{t+1} \in G(\mu_t)$. Then (μ_t) has converging subsequences, and any weakly converging subsequence necessarily converges towards a $\mu \in \mathcal{P}(\mathcal{X})$ such that $\mu \in G(\mu)$.*

Proof. Fix a sequence (μ_t) such that $\mu_{t+1} \in G(\mu_t)$ and write $\mu_{t+1} = B\#[\gamma_t]_{1,\cdots,K}$ with $\gamma_t \in \Gamma(\mu_t)$. Since $\mathcal{X}$ is compact, the space $\mathcal{P}(\mathcal{X})$ is also compact, and so the sequence (μ_t) is tight. Consider an extraction α such that $\mu_{\alpha(t)} \xrightarrow[t\longrightarrow+\infty]{w} \mu \in \mathcal{P}(\mathcal{X})$. By u.h.c. of Γ (Proposition 3.7), there exists an extraction β such that $\gamma_{\alpha\circ\beta(t)} \xrightarrow[t\longrightarrow+\infty]{w} \gamma \in \Gamma(\mu)$.

By Proposition 3.9, the sequence $(V(\mu_t))$ is non-increasing and non-negative, hence it is convergent, imposing $\lim\limits_{t\longrightarrow+\infty} \left[V(\mu_{\alpha\circ\beta(t)}) - V(\mu_{\alpha\circ\beta(t)+1})\right] = 0$. Using the lower-bound in Proposition 3.9 we obtain:

$$\forall t \in \mathbb{N},\ 0 \leq \mathcal{T}_\delta(\mu_{\alpha\circ\beta(t)}, \mu_{\alpha\circ\beta(t)+1}) \leq V(\mu_{\alpha\circ\beta(t)}) - V(\mu_{\alpha\circ\beta(t)+1}),$$

and take the limit inferior:

$$0 \leq \liminf_{t\longrightarrow+\infty} \mathcal{T}_\delta(\mu_{\alpha\circ\beta(t)}, \mu_{\alpha\circ\beta(t)+1}) \leq 0. \tag{20}$$

We remind that $(\mu_{\alpha\circ\beta(t)+1})_t$ is a sequence in $\mathcal{P}(\mathcal{X})$ which is compact, and take $\rho \in \mathcal{P}(\mathcal{X})$ a subsequential limit of $(\mu_{\alpha\circ\beta(t)+1})_t$. By lower-semi-continuity of $\mathcal{T}_\delta$ (which holds by applying Lemma 2.3 item 1) with Lemma 3.8), Eq. (20) provides $\mathcal{T}_\delta(\mu, \rho) = 0$. By Lemma 2.3 item 3), we obtain that $\rho = \mu$, thus any subsequential limit of $(\mu_{\alpha\circ\beta(t)+1})_t$ is μ, which proves that it converges weakly to μ.

Writing abusively $B\#\gamma$ for $B\#\gamma_{1,\cdots,K}$, we conclude:

$$\begin{array}{ccc} \mu_{\alpha\circ\beta(t)+1} & \xrightarrow[t\longrightarrow+\infty]{w} & \mu \\ \| & & \| \\ B\#\gamma_{\alpha\circ\beta(t)} & \xrightarrow[t\longrightarrow+\infty]{w} & B\#\gamma \end{array}$$

hence we have found $\gamma \in \Gamma(\mu)$ such that $\mu = B\#\gamma$, proving $\mu \in G(\mu)$. □

Fixed-points of G may not be unique and may not be barycentres, as shown the the following example. Take the following measures:

$$\mu := \frac{1}{2}\left(\delta_{(0,1)} + \delta_{(0,-1)}\right),\ \nu_1 := \frac{1}{2}\left(\delta_{(-2,1)} + \delta_{(2,-1)}\right),\ \nu_2 := \frac{1}{2}\left(\delta_{(-2,-1)} + \delta_{(2,1)}\right).$$

Between μ and ν_k, the unique OT plan for the squared-Euclidean cost is given by a permutation, with

$$\pi_1^* = \frac{1}{2}\left(\delta_{(0,1)\otimes(-2,1)} + \delta_{(0,-1)\otimes(2,-1)}\right),$$

and likewise for π_2^*. The next iterate of G and H are both equal to μ itself, which is distinct from the unique barycenter $\mu^* = \frac{1}{2}(\delta_{(-2,0)} + \delta_{(2,0)})$. We show this example in Fig. 4.

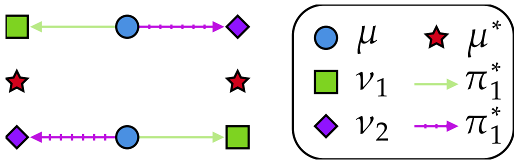


Figure 4: Example showing non-barycentre measure μ which is a fixed-point of G and H.

3.3 Expression of the Iterates when the Plans are Maps

In some cases, the plans introduced in $\Gamma(\mu)$ (Eq. (8)) are induced by maps, which is to say that they are each supported on a set of the form $(x, T_k(x))$. This is the case in the specific setting chosen by [Álv+16], which is to say that all measures are absolutely continuous on $\mathbb{R}^d$ and the costs are all $c(x,y) = \|x-y\|_2^2$. By Brenier's Theorem (as stated in [San15], Theorem 1.22, for example), this implies that optimal transport couplings are supported on the graph of a map. This property holds under the weaker condition that the costs verify the Twist condition (see [Vil09] Theorem 10.28 for example). In this case, each set optimal transport plans $\Pi_{c_k}^*(\mu, \nu_k)$ is composed of one element $(I, T_k)\#\mu$, and as a result, the expression of $G(\mu)$ becomes substantially simpler, namely $G(\mu) = \{B \circ (T_1, \cdots, T_K)\#\mu\}$. In the linearisation interpretation (Fig. 1), this expression can be understood as taking the ground barycentre of the maps T_k using the ground map B.

Drawing inspiration from this observation, we can define an alternative iteration consisting in choosing a map T_k as the barycentric projection of the coupling $\gamma_{0,k} \in \Pi_{c_k}^*(\mu, \nu_k)$ for $\gamma \in \Gamma(\mu)$: see Definition 3.11 and Fig. 5.

Definition 3.11. *The **barycentric projection** of a coupling $\pi \in \Pi(\mu, \nu)$ for $\mu \in \mathcal{P}(E)$ and $\nu \in \mathcal{P}(F)$ is the map $\overline{\pi} : E \longrightarrow F$, which is defined for μ-almost-every $x \in E$ as:*

$$\overline{\pi}(x) = \int_F y \pi_x(\mathrm{d}y),$$

where we wrote the disintegration $\pi(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x)\pi_x(\mathrm{d}y)$. In terms of random variables, one may write this expression as:

$$\overline{\pi}(x) = \underset{(X,Y)\sim\pi}{\mathbb{E}}[Y \mid X = x].$$

Note that for this expression to be well-defined, the target space F must be a *convex space*, i.e. a space where one may define convex combinations of points (or, more precisely, expectations of probability measures). In the case $\mathcal{X} = \mathcal{Y}_1 = \cdots = \mathcal{Y}_K$, a meaningful choice of convex combination is the ground barycentre B. We can apply this barycentric projection idea to define an alternate multi-mapping $H : \mathcal{P}(\mathcal{X}) \rightrightarrows \mathcal{P}(\mathcal{X})$:

$$\forall \mu \in \mathcal{P}(\mathcal{X}),\ H(\mu) := \left\{B \circ (\overline{\gamma_{0,1}}, \cdots, \overline{\gamma_{0,K}})\#\mu,\ \gamma \in \Gamma(\mu)\right\}. \tag{21}$$

In general, for $\pi \in \Pi(\mu, \nu)$, $\overline{\pi}\#\mu \neq \nu$, hence one does not necessarily have $\forall \widetilde{\mu} \in H(\mu),\ V(\widetilde{\mu}) \leq V(\mu)$. However, if each $\Pi_{c_k}^*(\mu, \nu_k)$ are composed of plans supported by maps, then $H(\mu) = G(\mu)$. In the case of discrete measures and for the squared Euclidean cost, the iterations of H correspond to the approach proposed in [CD14], Algorithm 2.

3.4 Extension to the Entropic Case

In this section, we explain how our results from Section 3.2 extend to Entropic-Regularised Optimal transport, wherein we introduce for a regularisation $\varepsilon > 0$:

$$\mathcal{T}_{c,\varepsilon}(\mu, \nu) := \min_{\pi \in \Pi(\mu,\nu)} \int_{E\times F} c\mathrm{d}\pi + \varepsilon\, \mathrm{KL}(\pi|\mu \otimes \nu), \quad V_\varepsilon := \sum_{k=1}^K \mathcal{T}_{c_k,\varepsilon}(\mu, \nu_k). \tag{22}$$

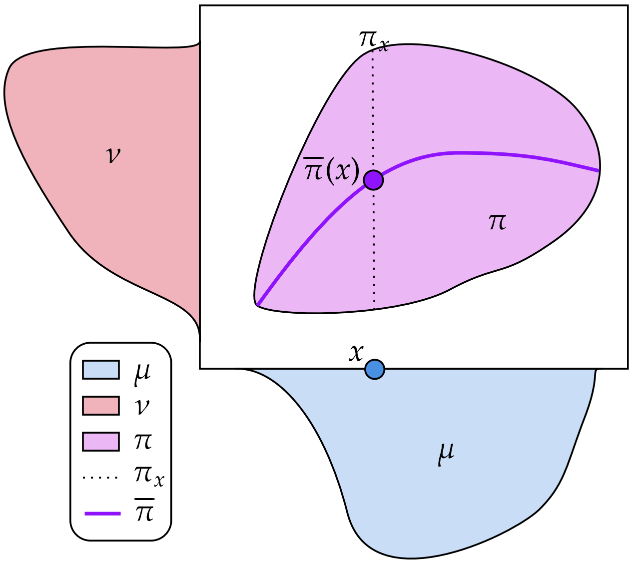


Figure 5: Illustration of a barycentric projection. The disintegration of the coupling π with respect to its first marginal μ at x is the measure π_x concentrated on the dotted line. The barycentric projection of π evaluated at x is the mean of the measure π_x.

Strict convexity of the KL divergence yields existence and uniqueness of entropic optimal transport plans (denoted $\Pi^*_{c,\varepsilon}(\mu,\nu)$), and by [GNB22] Theorem 1.4, the (single-valued) map $\mu \longmapsto \Pi^*_{c,\varepsilon}(\mu,\nu)$ is continuous for the weak convergence, provided that the cost c is continuous. Akin to the OT case, we define the map $\Gamma_\varepsilon : \mathcal{P}(\mathcal{X}) \longrightarrow \mathcal{P}(\mathcal{Z})$ by:

$$\Gamma_\varepsilon(\mu) := \mu(\mathrm{d}x)\pi^x_{1,\varepsilon}(\mathrm{d}y_1)\cdots\pi^x_{K,\varepsilon}(\mathrm{d}y_K),\ \forall k \in [\![1,K]\!],\ \pi_{k,\varepsilon} := \Pi^*_{c_k,\varepsilon}(\mu,\nu_k), \tag{23}$$

and the iteration functional $G_\varepsilon(\mu) := B\#\Gamma_\varepsilon(\mu)$. Using Lemma 3.8 and some technical manipulations of the KL divergence, we adapt Proposition 3.9 to this entropic case.

Proposition 3.12. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\overline{\mu} := G_\varepsilon(\mu)$. Then $V_\varepsilon(\mu) \geq V_\varepsilon(\overline{\mu}) + \mathcal{T}_\delta(\mu,\overline{\mu})$. If μ^* is a barycentre, then $G_\varepsilon(\mu^*) = \mu^*$.*

Proof. We begin as in Proposition 3.9, with $\gamma := \Gamma_\varepsilon(\mu)$:

$$\sum_{k=1}^K \mathcal{T}_{c_k,\varepsilon}(\mu,\nu_k) = \int_{\mathcal{X}\times\mathcal{Y}} C(x,Y)\mathrm{d}\gamma(x,Y) + \varepsilon\sum_{k=1}^K \mathrm{KL}(\gamma_{0,k}|\mu\otimes\nu_k) \tag{24}$$

$$\geq \int_{\mathcal{X}\times\mathcal{Y}} (C(B(Y),Y) + \delta(x,B(Y)))\mathrm{d}\gamma(x,Y) + \varepsilon\sum_{k=1}^K \mathrm{KL}(\gamma_{0,k}|\mu\otimes\nu_k). \tag{25}$$

For convenience, write $\gamma_\otimes := \mu\otimes\nu_1\otimes\cdots\otimes\nu_K$. Using the notation from Eq. (23), notice that $\dfrac{\mathrm{d}\gamma}{\mathrm{d}\gamma_\otimes} = \prod_{k=1}^K \dfrac{\mathrm{d}\pi_{k,\varepsilon}}{\mathrm{d}(\mu\otimes\nu_k)}$, which implies that $\sum_k \mathrm{KL}(\gamma_{0,k}|\mu\otimes\nu_k) = \mathrm{KL}(\gamma|\gamma_\otimes)$. Putting this with Eq. (25) yields

$$V_\varepsilon(\mu) \geq \sum_{k=1}^K \int_{\mathcal{X}\times\mathcal{Y}_k} c_k \mathrm{d}(B,P_k)\#\gamma + \varepsilon\,\mathrm{KL}(\gamma|\gamma_\otimes) + \int_{\mathcal{X}^2} \delta \mathrm{d}(I,B)\#\gamma. \tag{26}$$

Now let $f : \mathcal{X}\times\mathcal{Y}_1\times\cdots\times\mathcal{Y}_K \longrightarrow \mathcal{X}\times\mathcal{Y}_1\times\cdots\times\mathcal{Y}_K$ the continuous function defined by $f(x,y_1,\cdots,y_K) := (B(y_1,\cdots,y_K),y_1,\cdots,y_K)$. We apply the data processing inequality (use [PW14] Theorem 3.5, or apply [AGS05] Lemma 9.4.5): $\mathrm{KL}(\gamma|\gamma_\otimes) \geq \mathrm{KL}(f\#\gamma|f\#\gamma_\otimes)$. Now we use the disintegration formula and the change-of-reference formula for KL. Notice that the first marginals of $f\#\gamma$ and $f\#\gamma_\otimes$ are both equal to $\overline{\mu}$, and that $(f\#\gamma)_{1,\cdots,K} \in \Pi(\nu_1, \cdots, \nu_K)$ and $(f\#\gamma_\otimes)_{1,\cdots,K} = \nu_1 \otimes \cdots \otimes \nu_K$.

$$\begin{aligned}\mathrm{KL}(f\#\gamma|f\#\gamma_\otimes) &= \mathrm{KL}((f\#\gamma)_0|(f\#\gamma_\otimes)_0) + \int_{\mathcal{X}} \mathrm{KL}((f\#\gamma)^x|\nu_1 \otimes \cdots \otimes \nu_K)\mathrm{d}\overline{\mu}(x) \\ &= 0 + \int_{\mathcal{X}} \mathrm{KL}((f\#\gamma)^x|(f\#\gamma)_1^x \otimes \cdots \otimes (f\#\gamma)_K^x)\mathrm{d}\overline{\mu}(x) \\ &+ \mathrm{KL}((f\#\gamma)_1^x \otimes \cdots \otimes (f\#\gamma)_K^x|\nu_1 \otimes \cdots \otimes \nu_K)\mathrm{d}\overline{\mu}(x) \\ &\geq \sum_{k=1}^K \int_{\mathcal{X}} \mathrm{KL}((f\#\gamma)_k^x|\nu_k)\mathrm{d}\overline{\mu}(x) = \sum_{k=1}^K \mathrm{KL}((f\#\gamma)_{0,k})|\mu \otimes \nu_k).\end{aligned}$$

Now we notice that $(f\#\gamma)_{0,k} = (B, P_k)\#\gamma \in \Pi(\overline{\mu}, \nu_k)$, which with Eq. (26) provides:

$$\begin{aligned}V_\varepsilon(\mu) &\geq \sum_{k=1}^K \left(\int_{\mathcal{X}\times\mathcal{Y}_k} c_k \mathrm{d}(B, P_k)\#\gamma + \varepsilon\, \mathrm{KL}((B, P_k)\#\gamma|\mu \otimes \nu_k)\right) + \int_{\mathcal{X}^2} \delta \mathrm{d}(I, B)\#\gamma \\ &\geq \sum_{k=1}^K \mathcal{T}_{c,\varepsilon}(\overline{\mu}, \nu_k) + \mathcal{T}_\delta(\mu, \overline{\mu}) = V_\varepsilon(\overline{\mu}) + \mathcal{T}_\delta(\mu, \overline{\mu}).\end{aligned}$$

The rest of the proof follows as in Proposition 3.9. □

From Proposition 3.12, we deduce an adaptation of Theorem 3.10 to the entropic case.

Theorem 3.13. *For any $\mu_0 \in \mathcal{P}(\mathcal{X})$, let (μ_t) verifying $\mu_{t+1} = G_\varepsilon(\mu_t)$. Then (μ_t) has converging subsequences, and any weakly converging subsequence necessarily converges towards a $\mu \in \mathcal{P}(\mathcal{X})$ such that $\mu = G_\varepsilon(\mu)$.*

Proof. The proof can be adapted from Theorem 3.10 without difficulty, in particular given the fact that each $\mu \longmapsto \Pi^*_{c_k,\varepsilon}(\mu, \nu_k)$ is continuous with respect to the weak convergence of measures, which ensures that Γ_ε is also continuous. □

3.5 The Particular Case of Conditionally Independent Couplings

In Eq. (8), we chose all possible multi-couplings with optimal bi-marginals. It is possible to restrict the set of couplings to the smaller set of multi-couplings with conditionally independent marginals, i.e. multi-couplings $\gamma \in \Pi(\mu, \nu_1, \cdots, \nu_K)$ such that there exists $\pi_k \in \Pi^*_{c_k}(\mu, \nu_k)$ for $k \in [\![1, K]\!]$ such that $\gamma_{0,k} = \pi_k$ and specifically:

$$\gamma(\mathrm{d}x, \mathrm{d}y_1, \cdots, \mathrm{d}y_K) := \mu(\mathrm{d}x)\ \pi_1^x(\mathrm{d}y_1) \cdots \pi_K^x(\mathrm{d}y_K),$$

as in Eq. (9). In terms of random variables, this corresponds to the choice of $(X, Y_1, \cdots, Y_K)$ such that $(X, Y_k) \sim \pi_k$ and conditionally to X, the variables $Y_1, \cdots, Y_K$ are independent. We denote by $\Gamma_\otimes(\mu)$ the set of such couplings, and consider the associated multi-map $G_\otimes := B\#\Gamma_\otimes$ as in Eq. (10). It is clear that $\forall \mu \in \mathcal{P}(\mathcal{X})$, $G_\otimes(\mu) \subset G(\mu)$. In particular, this implies subsequential convergence converges of iterates $\mu_{t+1} = G_\otimes(\mu_t)$ to a fixed-point of G. In Proposition 3.14, we show that the convergence is to a fixed-point of $G_\otimes$ in the discrete case (measures with finite support). First, we emphasise that with a discrete initialisation measure and discrete measures (ν_k), the support of the sequence (μ_t) is finite and always contained in:

$$\{B(y_1, \cdots, y_K),\ \forall k \in [\![1, K]\!],\ y_k \in \mathrm{supp}(\nu_k)\},$$

which ensures that iterates remain discrete.

Proposition 3.14. *Take $\mu_0 \in \mathcal{P}(\mathcal{X})$ a discrete measure and $\nu_1, \cdots \nu_k \in \mathcal{P}(\mathcal{Y}_1) \times \cdots \times \mathcal{P}(\mathcal{Y}_K)$ discrete measures. Then any sub-sequential limit $\mu \in \mathcal{P}(\mathcal{X})$ of the sequence (μ_t) defined by $\mu_{t+1} \in G_\otimes(\mu_t)$ verifies $\mu \in G_\otimes(\mu)$.*

Proof. We follow a technique used in the proof of Theorem 9.6 in [Goz+17], specifically page 65. As commented before the statement of the result, the sequence (μ_t) remains discrete. Write for $t \in \mathbb{N}$, $\mu_{t+1} = B\#\gamma_t$ with $\gamma_t \in \Gamma_\otimes(\mu_t)$, and take an extraction α such that $\mu_{\alpha(t)} \xrightarrow[t\longrightarrow+\infty]{w} \mu$. As done in Theorem 3.10, the u.h.c. property of Γ allows us to extract a subsequence β such that $\gamma_{\alpha\circ\beta(t)} \xrightarrow[t\longrightarrow+\infty]{w} \gamma \in \Gamma(\mu)$, since we have the (point-wise) inclusion $\Gamma_\otimes \subset \Gamma$. As shown in the proof of Theorem 3.10, the sequence $(\mu_{\alpha\circ\beta(t)})_t$ weakly converges to μ, hence we now want to show that $\gamma \in \Gamma_\otimes(\mu)$, which would allow to conclude $\mu \in G_\otimes(\mu)$.

For $t \in \mathbb{N}$ and $k \in [\![1, K]\!]$, introduce $\pi_{\alpha\circ\beta(t)}^{(k)} := [\gamma_{\alpha\circ\beta(t)}]_{0,k}$, and its disintegration with respect to $\mu_{\alpha\circ\beta(t)}$ as

$$\pi_{\alpha\circ\beta(t)}^{(k)}(\mathrm{d}x, \mathrm{d}y_k) = \mu_{\alpha\circ\beta(t)}(\mathrm{d}x)\ [\pi_{\alpha\circ\beta(t)}^{(k)}]^x(\mathrm{d}y_k).$$

As argued above the statement of the proposition, the sequence $(\mu_t)_t$ remains discrete with a support contained in $B(\prod_k \mathrm{supp}(\nu_k))$ and thus (γ_t) also remains discrete, and its first marginal μ_t has a finite support of size at most $n := \prod_k \#\,\mathrm{supp}(\nu_k)$ on fixed points $(x_1, \cdots, x_n)$. For simplicity, we will see the measures (μ_t) as supported on $\mathcal{X}_n := \{x_1, \cdots, x_n\}$ with possibly zero mass at some of these points, and in such cases, we define $[\pi_{\alpha\circ\beta(t)}^{(k)}]^x$ as the null measure $\mathcal{M}_0$. Since $\gamma_{\alpha\circ\beta(t)} \in \Gamma_\otimes(\mu_{\alpha\circ\beta(t)})$, by definition we can write its disintegration with respect to $\mu_{\alpha\circ\beta(t)}$ as:

$$\gamma_{\alpha\circ\beta(t)}(\mathrm{d}x, \mathrm{d}y_1, \cdots, \mathrm{d}y_K) = \mu_{\alpha\circ\beta(t)}(\mathrm{d}x)\ [\pi_{\alpha\circ\beta(t)}^{(1)}]^x(\mathrm{d}y_1) \cdots [\pi_{\alpha\circ\beta(t)}^{(K)}]^x(\mathrm{d}y_K).$$

For $i \in [\![1, n]\!]$ and $k \in [\![1, K]\!]$, there exists an extraction $\chi_{i,k}$ such that the sequence $\left([\pi_{\alpha\circ\beta\circ\chi_{i,k}(t)}^{(k)}]^{x_i}\right)_{t\in\mathbb{N}}$ converges weakly to a $[\pi^{(k)}]^{x_i} \in \mathcal{P}(\mathcal{X}) \times \{\mathcal{M}_0\}$. We choose the extractions as successive sub-extractions, such that $\chi_{1,2}$ is a sub-extraction of $\chi_{1,1}$, until $\chi_{n,K}$ which is a sub-extraction of all previous extractions. We then define $\chi := \chi_{n,K}$. The extraction χ is such that for $i \in [\![1, n]\!]$ and $k \in [\![1, K]\!]$, the sequence $\left([\pi_{\alpha\circ\beta\circ\chi(t)}^{(k)}]^{x_i}\right)_{t\in\mathbb{N}}$ converges weakly to $[\pi^{(k)}]^{x_i}$. By verifying against test functions, we deduce the following disintegration holds for γ:

$$\gamma(\mathrm{d}x, \mathrm{d}y_1, \cdots, \mathrm{d}y_K) = \mu(\mathrm{d}x)\ [\pi^{(1)}]^x(\mathrm{d}y_1) \cdots [\pi^{(K)}]^x(\mathrm{d}y_K),$$

which shows that $\gamma \in \Gamma_\otimes(\mu)$, and thus $\mu \in G_\otimes(\mu)$. □

Remark 3.15. *The proof of Proposition 3.14 can also be written for discrete measures with at-most-countable supports through a diagonal extraction argument, we kept to finite supports for legibility.*

4 Focus on the Discrete Case

In this section, we will formulate the fixed-point algorithm in the discrete case, and discuss some algorithmic aspects.

4.1 Discrete Expression and Algorithms

Consider discrete measures $\nu_k := \sum_{i=1}^{n_k} b_{k,i}\delta_{y_{k,i}} \in \mathcal{P}(\mathbb{R}^{d_k})$ where $\forall k \in [\![1, K]\!]$, $\forall i \in [\![1, n_k]\!]$, $y_{k,i} \in \mathbb{R}^{d_k}$. We stack the support of ν_k into $Y_k \in \mathbb{R}^{n_k \times d_k}$ such that $[Y_k]_{i,\cdot} = y_{k,i}$, and similarly introduce $b_k := (b_{k,i})_{i=1}^{n_k} \in \Delta_{n_k}$.

First, our objective is to re-write the iteration Eq. (10) in this discrete setting, with an initial measure $\mu = \sum_{i=1}^n a_i \delta_{x_i} \in \mathcal{P}(\mathbb{R}^d)$. For each k, we choose $\pi_k \in \mathbb{R}_+^{n \times n_k}$ an optimal transport plan, which is to say a solution of the Kantorovich linear program:

$$\underset{\Pi(a,b_k)}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{n_k} c_k(x_i, y_{k,j}) \pi_{i,j},$$

where $\Pi(a, b_k) := \left\{\pi \in \mathbb{R}_+^{n\times n_k} \ : \ \pi \mathbf{1} = a, \ \pi^T \mathbf{1} = b_k\right\}$. Seeing multi-couplings $\gamma \in \Gamma(\mu)$ as a tensors $\gamma \in \mathbb{R}^{n\times n_1 \times \cdots, n_k}$, the discrete expression of G reads:

$$G(\mu) = \left\{ \sum_{j_1,\cdots,j_K} \left(\sum_{i=1}^{n} \gamma_{1,j_1,\cdots,j_K} \right) \delta\left(B(y_{1,j_1}, \cdots, y_{K,j_K})\right), \ \gamma \in \Gamma(\mu) \right\}. \tag{27}$$

A visualisation of Eq. (27) using the multi-coupling from Eq. (9) is provided in Fig. 6.

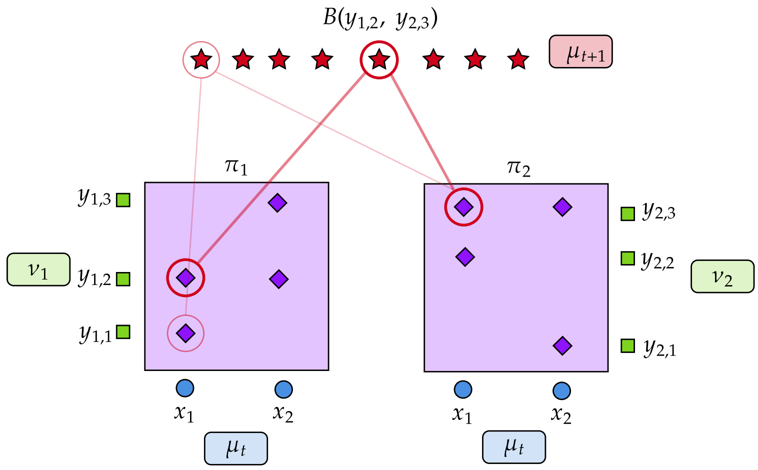


Figure 6: Visual explanation of the discrete fixed-point iteration G. For each point x_i in the support of μ_t, we look at all the points $(y_{1,j_1}, \cdots, y_{K,j_K})$ which are assigned from x_i by the multi-coupling γ, then the ground barycentre $B(y_{1,j_1}, \cdots, y_{K,j_K})$ is taken on all these tuples, with weights given by the multi-coupling.

As in [BB25], we can use a generalisation of the North-West Corner (NWC) method to compute $\gamma \in \Gamma(\mu)$ with prescribed bi-marginals $\pi_k \in \Pi(a, b_k)$. In Algorithm 1, we present the NWC strategy. The idea is to fill the entries of γ greedily using entries $\pi_{i,j_k}^{(k)}$ for increasing $i, j_1, \cdots, j_K$ (see [PC19] Section 3.4.2 for a presentation of the method in the standard setting).

Noticing that Eq. (27) only requires the $n_1 \times \cdots \times n_K$-tensor $\rho := \gamma_{1,\cdots,K}$, it is possible to only store the indices $(j_1, \cdots, j_K)$ such that $\rho_{j_1,\cdots,j_K} > 0$, as well as the corresponding weights $\rho_{j_1,\cdots,j_K}$. This avoids the prohibitive memory cost of storing the full tensor γ, and takes advantage of the sparsity of the multi-coupling γ: if each π_k is an extremal point of $\Pi(a, b_k)$, we conjecture that $\mathrm{NWC}(\pi_1, \cdots, \pi_K)$ is an extremal point of $\Pi(a, b_1, \cdots, b_K)$, and thus $\#\operatorname{supp}\gamma \leq n + \sum_k n_k - K$ (adapting techniques from [ABM16] Theorem 2).

Thanks to Eq. (27) we formalise the fixed-point iterations in the discrete case in Algorithm 2. Given our considerations on the support of NWC gluing, we expect (without formal proof) the upper bound $\#\operatorname{supp}\mu_T \leq n + T(\sum_k n_k) - TK$. This is the same conclusion as [BB25], which they also state without proof about their Gromov-Wasserstein fixed-point iteration (Algorithm 5.2). From a memory perspective, the algorithm does not require the storage of each $\gamma \in \mathbb{R}^{N_t \times n_1 \times \cdots \times n_K}$, as remarked for the NWC algorithm.

Algorithm 1: North-West Corner Gluing.

Data: For $k \in [\![1, K]\!]$, transport plan $\pi_k \in \Pi(a, b_k)$, with $a \in \Delta_n$ and $b_k \in \Delta_{n_k}$.

Result: Gluing $\mathrm{NWC}(\pi_1, \cdots, \pi_K) = \gamma \in \Pi(a, b_1, \cdots, b_K)$ such that each $\gamma_{0,k} = \pi_k$.

1 **Initialisation:** $\gamma = 0_{n\times n_1 \times \cdots \times n_K}$ and for $k \in [\![1, K]\!]$, $P_k = \pi_k$.

2 **for** $i \in [\![1, n]\!]$ **do**

3   Set $(j_1, \cdots, j_K) = (1, \cdots, 1)$ and $u = a_i$;

4   **while** $u > 0$ **do**

5   Compute $v = \min(P^{(1)}_{i,j_1}, \cdots, P^{(K)}_{i,j_K})$;

6   Assign $\gamma_{i,j_1,\cdots,j_K} = v$ and decrease $u \leftarrow u - v$;

7   **for** $k \in [\![1, K]\!]$ **do**

8   Decrease $P^{(k)}_{i,j_k} \leftarrow P^{(k)}_{i,j_k} - v$;

9   **if** $P^{(k)}_{i,j_k} = 0$ **then**

10   Increment $j_k \leftarrow j_k + 1$;

11   **end**

12   **end**

13   **end**

14 **end**

In some specific cases, the expression in Eq. (27) becomes simpler. If the weights a and b_k are all uniform and $n = n_1 = \cdots = n_K$, then the Birkhoff-von-Neumann Theorem allows the choice of each transport plan π_k as permutation assignments $[\pi_k]_{i,j} = \frac{1}{n}\mathbb{1}(\sigma_k(i) = j)$. In this case, the expression of $G(\mu)$ becomes:

$$G(\mu) = \frac{1}{n}\sum_{i=1}^{n} \delta\left(B(y_{1,\sigma_1(i)}, \cdots, y_{K,\sigma_K(i)})\right). \tag{28}$$

If one takes the barycentric projections of the OT plans $\pi^{(k)}$ in Eq. (27), one obtains a discrete expression of H (from Eq. (21)) written in Eq. (29) and visualised in Fig. 7.

$$H(\mu) = \left\{\sum_{i=1}^{n} a_i\delta\left[B\left((1/a_i)\sum_{j=1}^{n_1}\pi^{(1)}_{i,j}y_{1,j}, \cdots, (1/a_i)\sum_{j=1}^{n_K}\pi^{(K)}_{i,j}y_{K,j}\right)\right],\ \pi^{(k)} \in \Pi^*_{c_k}(\mu, \nu_k)\right\}. \tag{29}$$

Contrary to G, for H the number of points in the support of μ_t remains the same, and the weights a remain fixed. In this setting, the optimisation is done solely on the positions, which can be seen as a Lagrangian formulation. Note that in the squared-Euclidean case, Eq. (29) is the formula proposed in [CD14] (Equation 8) and currently implemented in the Python OT library [Fla+21]. A technical difference is that [CD14] also proposes an optimisation over the barycentre weights (by sub-gradient descent), while the fixed-point approach by [Álv+16] and ours do not. Furthermore, [CD14] suggests a computational simplification by using barycentric projections of *entropic* plans (as in Section 3.4), for which, as for H, there are no theoretical guarantees (to our knowledge).

The practical advantage of the map-supported expressions in Eqs. (28) and (29) over Eq. (27) is computational: since the support size of μ_t cannot increase, the cost of computing the OT plans at Line 4 is smaller. We shall see in Section 4.3 that in some cases, Kantorovich solutions are almost-surely permutations for random supports. While convenient, this expression only holds when all the measures have the same amount of points, in contrast to the barycentric expression Eq. (29).

We present the iteration of H as a cost-effective alternative to G, which is in some sense a simplification of the Block-Coordinate Descent (BCD) method, wherein the update with

Algorithm 2: Discrete iteration of G.

Data: barycentre coefficients $(\lambda_k) \in \Delta_K$, for $k \in [\![1, K]\!]$, support of ν_k: $Y_k \in \mathbb{R}^{n_k \times d_k}$, weights of ν_k: $b_k \in \Delta_{n_k}$ and cost function $c_k : \mathbb{R}^d \times \mathbb{R}^{d_k} \longrightarrow \mathbb{R}_+$. Number of iterations T, initial size $n \geq 1$ and stopping criterion $\alpha \geq 0$.

Result: Barycentre $\mu_T = \sum_{i=1}^{N_t} a_i^{(T)} \delta_{x_i^{(T)}}$.

Initialisation: Choose $\mu_0 = \sum_{i=1}^{n} a_i^{(0)} \delta_{x_i^{(0)}}$ with $a^{(0)} \in \Delta_n$ and $X^{(0)} \in \mathbb{R}^{n \times d}$.

for $t \in [\![0, T-1]\!]$ **do**
  for $k \in [\![1, K]\!]$ **do**
    Solve the OT problem: $\pi^{(k)} \in \underset{\pi \in \Pi(a^{(t)}, b_k)}{\text{argmin}} \sum_{i,j} \pi_{i,j} c_k(x_i^{(t)}, y_{k,j})$;
  end
  Compute $\gamma = \mathrm{NWC}(\pi^{(1)}, \cdots, \pi^{(K)})$;
  Compute $\rho = \gamma_{1,\cdots,K} = [\sum_i \gamma_{j_1,\cdots,j_K}]_{j_1,\cdots,j_K}$ and write $\operatorname{supp} \rho = \left((j_1^{(i)}, \cdots, j_K^{(i)})\right)_{i=1}^{N_t}$;
  for $i \in [\![1, N_t]\!]$ **do**
    Compute $x_i^{(t+1)} = B(y_{1,j_1^{(i)}}, \cdots, y_{K,j_K^{(i)}})$ and $a_i^{(t+1)} = \rho_{j_1,\cdots,j_K}$;
  end
  if $\mathrm{W}_2^2(\mu_{t+1}, \mu_t) < \frac{\alpha}{N_t} \|X^{(t)}\|_2^2$ **then**
    Declare convergence and terminate.
  end
end
return $a^{(T)}, X^{(T)}$

respect to the support $X \in \mathbb{R}^{n \times d}$ with transport plans (π_k) fixed is done by computing:

$$X^* \in \underset{X \in \mathbb{R}^{n \times d}}{\text{argmin}} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n_k} \pi_{i,j}^{(k)} c_k(x_i, y_{k,j}). \tag{30}$$

In practice, apart from the case of the squared Euclidean cost, the optimisation in Eq. (30) is not tractable, and one must resort to Gradient Descent (GD) methods. BCD methods with GD for the update of X can be seen as a variant of the full GD method which minimises $X \longmapsto V(\frac{1}{n}\sum_i \delta_{x_i})$, and we leave their study for future work.

4.2 Correspondence of Gradient Descent with Fixed-Point Iterations

The fixed-point method of [Álv+16] applied to Bures-Wasserstein barycentres also corresponds to a gradient descent algorithm with a specific step size, as remarked by [Alt+21]. This also holds for discrete measures. Indeed, writing $X = \{x_1, \cdots, x_n\}$ and assuming $\mu_X = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, an alternative to fixed-point iterations would be to apply a gradient descent directly on the non convex functional $F : X \longmapsto \sum_{k=1}^K \lambda_k \mathcal{T}_{c_k}(\mu_X, \nu_k)$. For differentiable costs c_k, assuming that $\nu_k = \frac{1}{n}\sum_{i=1}^n \delta_{y_{k,i}}$, one step of such a gradient descent writes

$$\forall i \in [\![1, n]\!],\ x_i^{(t+1)} = x_i^{(t)} - \alpha \sum_{k=1}^{K} \lambda_k \nabla_x c_k(x_i^{(t)}, y_{k,\sigma_k^{(t)}(i)}), \tag{31}$$

where we choose an element of $\Pi_{c_k}^*(\mu_{X^{(t)}}, \nu_k)$ induced by a permutation $\sigma_k^{(t)}$ between $\{x_1^{(t)}, \cdots, x_n^{(t)}\}$ and $\{y_{k,1}, \cdots, y_{k,n}\}$. The whole optimisation algorithm consists in alternating such gradient

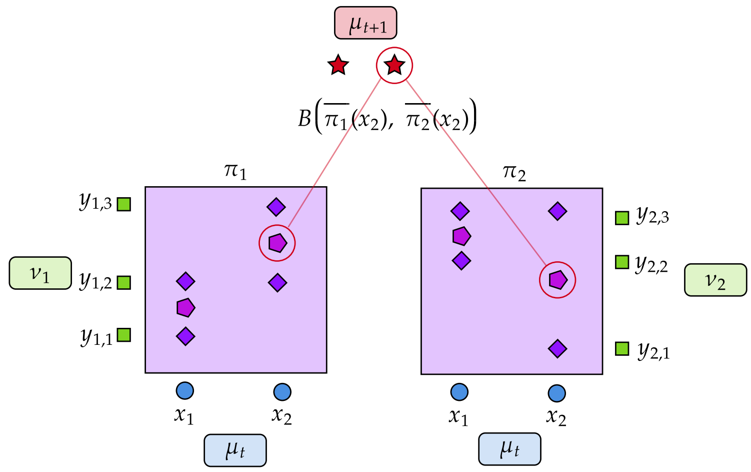


Figure 7: Visual explanation of the discrete fixed-point iteration H. For each point x_i in the support of μ_t and $k \in [\![1, K]\!]$, we look at all the points (y_{k,j_k}) which are assigned from x_i by the OT plan $\pi^{(k)}$ between μ and ν_k, then take their barycenter $\overline{\pi}_k(x_i)$ (pentagons on the figure). The point x_i in the support of μ_t is then sent to the point $B(\overline{\pi}_1(x_i), \cdots, \overline{\pi}_K(x_i))$ in μ_{t+1}.

steps on X with updates of the optimal assignments $\{\sigma_k^{(t)}\}$, depending on the new point positions. In the fixed-point approach, this gradient step on each $x_i^{(t)}$ is replaced by the computation of $B(y_{1,\sigma_1^{(t)}(i)}, \cdots, y_{K,\sigma_K^{(t)}(i)})$, which corresponds to a full descent on X for a given configuration of assignments before updating the said assignments (in other words, alternate minimisation). For generic costs c_k, one may also use a gradient descent strategy to compute barycentres $B(y_{1,\sigma_1^{(t)}(i)}, \cdots, y_{K,\sigma_K^{(t)}(i)})$, that is gradient descents on the K functionals $x \longmapsto \sum_{k=1}^K c_k(x, y_{k,\sigma_k^{(t)}(i)})$, and such descents write exactly as Eq. (31). In this case, the only difference between both approaches is that the fixed point algorithm applies the whole descent on X before updating assignments, while gradient descent on F alternates steps of gradient descent on X with updates of the assignments.

When $c_k = \| \cdot - \cdot \|_2^2$, both approaches are equivalent if the gradient step is chosen as $\alpha = \frac{1}{2}$. Indeed, a gradient iteration on F writes

$$\forall i \in [\![1, n]\!],\ x_i^{(t+1)} = (1 - 2\alpha)x_i^{(t)} + 2\alpha \sum_{k=1}^K \lambda_k y_{k,\sigma_k^{(t)}(i)} = \sum_{k=1}^K \lambda_k y_{k,\sigma_k^{(t)}(i)}.$$

It follows that for $\alpha = \frac{1}{2}$, one step of gradient descent computes directly the barycentre for the current configuration of assignments $\{\sigma_k^{(t)}\}$, which is precisely one iteration of the fixed-point algorithm. For different cost functions, similar optimal steps may be formulated, but the step may depend on i and $x_i^{(t)}$.

Choosing the best strategy between the fixed point approach and the gradient descent surely depends on the set of costs. When B is easily computable (more efficiently than by gradient descent), the fixed point algorithm moves the points faster than gradient descent. However, it is not obvious what should be the better option for complex costs c_k in practice. More generally, one could wonder if updating assignments more often (which is the case for the gradient descent on F) might not help avoiding local minima of the whole functional which is non convex in X. We did not observe this behaviour in practice in our experiments and therefore recommand the fixed point approach as the default choice.

Algorithm 3: Discrete iteration of H.

Data: barycentre coefficients $(\lambda_k) \in \Delta_K$, for $k \in [\![1, K]\!]$, support of ν_k: $Y_k \in \mathbb{R}^{n_k \times d_k}$, weights of ν_k: $b_k \in \Delta_{n_k}$ and cost function $c_k : \mathbb{R}^d \times \mathbb{R}^{d_k} \longrightarrow \mathbb{R}_+$. Number of iterations T, barycentre size $n \geq 1$, weights $a \in \Delta_n$ and stopping criterion $\alpha \geq 0$.

Result: Barycentre $\mu_T = \sum_{i=1}^n a_i \delta_{x_i^{(T)}}$.

Initialisation: Choose $\mu_0 = \sum_{i=1}^n a_i \delta_{x_i^{(0)}}$ with $X^{(0)} \in \mathbb{R}^{n \times d}$.

for $t \in [\![0, T-1]\!]$ **do**
    for $k \in [\![1, K]\!]$ **do**
        Solve the OT problem: $\pi^{(k)} \in \underset{\pi \in \Pi(a, b_k)}{\mathrm{argmin}} \ \sum_{i,j} \pi_{i,j} c_k(x_i^{(t)}, y_{k,j})$;
    end
    for $i \in [\![1, n]\!]$ **do**
        Compute $x_i^{(t+1)} = B\left((1/a_i)\sum_{j=1}^{n_1} \pi_{i,j}^{(1)} y_{1,j}, \cdots, (1/a_i)\sum_{j=1}^{n_K} \pi_{i,j}^{(K)} y_{K,j}\right)$;
    end
    if $\mathrm{W}_2^2(\mu_{t+1}, \mu_t) < \frac{\alpha}{n}\|X^{(t)}\|_2^2$ **then**
        Declare convergence and terminate.
    end
end
return $X^{(T)}$

4.3 Discrete Uniqueness Discussion

In this section, we investigate conditions to have uniqueness in the discrete Kantorovich problem between measures $\mu = \sum_{i=1}^{n_x} a_i \delta_{x_i} \in \mathcal{P}(\mathbb{R}^{d_x})$ and $\nu = \sum_{j=1}^{n_y} b_j \delta_{y_j} \in \mathcal{P}(\mathbb{R}^{d_y})$:

$$\min_{\pi \in \Pi(a,b)} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \pi_{i,j} c(x_i, y_j). \tag{32}$$

For convenience, we introduce $X := (x_1, \cdots, x_{n_x}) \in \mathbb{R}^{n_x \times d_x}$ and $Y := (y_1, \cdots, y_{n_y}) \in \mathbb{R}^{n_y \times d_y}$. The following result shows that if the cost matrix $M := (X, Y) \longmapsto (c(x_i, y_j))_{i,j} \in \mathbb{R}^{n_x \times n_y}$ is not orthogonal to a face of the transportation polytope, then the discrete Kantorovich problem has a unique solution. For convenience, we write $\pi \cdot M := \sum_{i,j} \pi_{i,j} M_{i,j}$.

Proposition 4.1. *Let $a \in \Delta_{n_x}$ and $b \in \Delta_{n_y}$ be fixed weights and $c : \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \longrightarrow \mathbb{R}_+$ a cost function. Consider the cost matrix function*

$$M := \left\{ \begin{array}{ccc} \mathbb{R}^{n_x \times d_x} \times \mathbb{R}^{n_y \times d_y} & \longrightarrow & \mathbb{R}^{n_x \times n_y} \\ (X, Y) & \longmapsto & (c(x_i, y_j))_{i,j} \end{array} \right. ,$$

and let $(X, Y) \in \mathbb{R}^{n_x \times d_x} \times \mathbb{R}^{n_y \times d_y}$. Denote by $\mathrm{Extr}\,\Pi(a,b)$ the (finite) set of extremal points of the transportation polytope $\Pi(a,b)$.

$$\min_{\pi \in \Pi(a,b)} \pi \cdot M(X,Y) \text{ has a unique solution} \iff M(X,Y) \notin \bigcup_{\pi_1 \neq \pi_2 \in \mathrm{Extr}\,\Pi(a,b)} (\pi_1 - \pi_2)^\perp. \tag{33}$$

Proof. Since $\Pi(a,b)$ is convex and compact in $\mathbb{R}^{n_x \times n_y}$, by the Krein-Milman theorem, it is the convex hull of the set of its extreme points, denoted $\mathrm{Extr}\,\Pi(a,b)$. With the definition

$$\Pi(a,b) = \left\{ \pi \in \mathbb{R}^{n_x \times n_y} : \pi \geq 0,\ \pi \mathbf{1} = a,\ \pi^T \mathbf{1} = b \right\},$$

we see that $\Pi(a,b)$ is a polytope, and thus $\operatorname{Extr}\Pi(a,b)$ is finite. Since the Kantorovich problem is a linear problem, the set of optimal solutions is exactly the set of convex combinations of optimal extremal points. As a result, we have non-uniqueness in Eq. (32) if and only if there exists $\pi_1 \neq \pi_2 \in \operatorname{Extr}\Pi(a,b) : \pi_1 \cdot M(X,Y) = \pi_2 \cdot M(X,Y)$. We conclude that uniqueness holds if and only if $\forall \pi_1 \neq \pi_2 \in \operatorname{Extr}\Pi(a,b) : M(X,Y) \notin (\pi_1 - \pi_2)^\perp$. □

A consequence of Proposition 4.1 is that if $M\#\mathscr{L}^{n_x\times d_x+n_y\times d_y}$ does not give mass to hyperplanes of $\mathbb{R}^{n_x\times n_y}$, then the Kantorovich problem has a unique solution for $\mathscr{L}^{n_x\times d_x+n_y\times d_y}$-almost-every (X,Y). Furthermore, if the measures have the same amount of points ($n_x = n_y$) and the weights are uniform, then the extreme points of $\Pi(a,b)$ are permutations, which provides a theoretical justification for the convenient expression in Eq. (28).

4.4 Application to Gaussian Mixture Model Barycentres

In this section, we explain how our fixed-point algorithm can be applied to compute barycentres between Gaussian Mixture Models (GMMs), providing a new numerical method for the GMM barycentre notion introduced in [DD20] (Section 5). The notation $S_d^{++}(\mathbb{R})$ will refer to the cone of positive definite symmetric $d\times d$ matrices.

We consider the case where the measures are Gaussian Mixture Models, seen as discrete measures over the space of Gaussian measures on $\mathbb{R}^d$: $\mathcal{X} := \mathcal{N} := \left\{\mathcal{N}(m,S) : m\in\mathbb{R}^d,\ S\in S_d^{++}(\mathbb{R})\right\}$, equipped with the 2-Wasserstein distance, which has a specific expression called the *Bures-Wasserstein distance*:

$$\mathrm{W}_2^2(\mathcal{N}(m_1,S_1),\mathcal{N}(m_2,S_2)) = \|m_1-m_2\|_2^2 + \underbrace{\operatorname{Tr}\left(S_1+S_2-2(S_1^{1/2}S_2S_1^{1/2})^{1/2}\right)}_{d^2_{\mathrm{BW}}(S_1,S_2):=}. \tag{34}$$

Alternatively, one could see the same problem differently, setting $\mathcal{X} := \mathbb{R}^d\times S_d^{++}(\mathbb{R})$ equipped with the distance defined in Eq. (34). To remind the definition of barycentres between Gaussian mixture models from [DD20], we will consider measures that lie on the same space of Gaussian measures: $\mathcal{X} = \mathcal{Y}_1 = \cdots = \mathcal{Y}_K = \mathcal{N}$. Next, we choose cost functions c_k on $\mathcal{N}$ as the squared Bures-Wasserstein distance W_2^2 scaled by λ_k. Given mixture models $\mu,\nu\in\mathcal{P}(\mathcal{N})$ of the form

$$\mu = \sum_{i=1}^n a_i\delta_{\mathcal{N}(m_i,S_i)},\ \nu = \sum_{j=1}^m b_j\delta_{\mathcal{N}(m_j',S_j')},$$

the Optimal Transport cost $\mathcal{T}_{\mathrm{W}_2^2}(\mu,\nu)$ is the value of a discrete problem, which is precisely the Mixed Wasserstein Distance introduced in [DD20] (as per their Proposition 4):

$$\mathcal{T}_{\mathrm{W}_2^2}(\mu,\nu) = \min_{\pi\in\Pi(a,b)} \sum_{i,j}\pi_{i,j}\mathrm{W}_2^2(\mathcal{N}(m_i,S_i),\mathcal{N}(m_j',S_j')). \tag{35}$$

Consider K GMM measures ν_k written as:

$$\nu_k = \sum_{j=1}^{n_k} b_{k,j}\delta_{\mathcal{N}(m_{k,j},S_{k,j})}\in\mathcal{P}(\mathcal{N}),$$

their GMM barycentre cost with weights (λ_k) for $\mu = \sum_{i=1}^n a_i\delta_{\mathcal{N}(m_i,S_i)}\in\mathcal{P}(\mathcal{N})$ reads:

$$V(\mu) = \sum_{k=1}^K \lambda_k \min_{\pi_k\in\Pi(a,b_k)} \sum_{i,j}\pi_{i,j}\left(\|m_i-m_{k,j}\|_2^2 + d^2_{\mathrm{BW}}(S_i,S_{k,j})\right). \tag{36}$$

We now turn to the expression of the ground barycentre function $B : \mathcal{N}^K \longrightarrow \mathcal{N}$. This corresponds to a 2-Wasserstein barycentre problem in the Gaussian case, which was first studied by [AC11] (showing existence and uniqueness in Theorem 6.1):

$$B(\mathcal{N}(m_1, S_1), \cdots, \mathcal{N}(m_K, S_K)) = \mathcal{N}(\overline{m}, \overline{S}),\ \overline{m} := \sum_{k=1}^K \lambda_k m_k,\ \overline{S} := \underset{S \in S_d^{++}(\mathbb{R})}{\operatorname{argmin}} \sum_{k=1}^K \lambda_k d_{\mathrm{BW}}^2(S, S_k).$$

A fixed-point formulation of this problem is presented in [Álv+16] as a particular case of their study of the fixed-point algorithm for the ground cost $\|\cdot - \cdot\|_2^2$ and absolutely continuous measures. This problem is presented again in [BJL17], were they prove additional convergence guarantees. We recall from [Álv+16; BJL17] the fixed-point algorithm to compute the barycentre of K Gaussians $(\mathcal{N}(m_k, S_k))$ and weights $(\lambda_1, \cdots, \lambda_K)$, which consists in iterating the function $G_{\mathcal{N}} : S_d^{++}(\mathbb{R}) \longrightarrow S_d^{++}(\mathbb{R})$:

$$G_{\mathcal{N}}(S) = S^{-1/2}\left(\sum_{k=1}^K \lambda_k (S^{1/2} S_k S^{1/2})^{1/2}\right)^2 S^{-1/2}. \tag{37}$$

Now that we have defined the ground barycentre map B, we can apply our fixed-point algorithm to compute a barycentre. Given a reference GMM with n components $\mu = \sum_{i=1}^n a_i \delta_{\mathcal{N}(m_i, S_i)}$, for $k \in [\![1, K]\!]$, solve the discrete Kantorovich problem between μ and ν_k (Eq. (35)) and choose $\pi_k \in \Pi^*_{\mathrm{W}_2^2}(\mu, \nu_k)$. The GMM of $G(\mu)$ associated to the choice of plans $\pi_k \in \Pi(a, b_k)$ in the iteration scheme is the GMM $\overline{\mu}$ defined by:

$$\overline{\mu} = \sum_{j_1, \cdots, j_K} \sum_{i=1}^n \frac{1}{a_i^{K-1}} \pi_{i,j_1}^{(1)} \times \cdots \times \pi_{i,j_K}^{(K)} \delta[B(\mathcal{N}(m_{1,j_1}, S_{1,j_1}), \cdots, \mathcal{N}(m_{K,j_K}, S_{K,j_K}))].$$

As we argued in Section 4.1, it is computationally wise to consider a variant of the fixed-point iterations which use the barycentric projections of the couplings π_k (see Eq. (21)). To use this in the case of the space $\mathcal{N}$, we need to choose a notion of convex combination in $\mathcal{N}$ to be able to compute the images of the barycentric projections. The most meaningful choice is a Wasserstein Gaussian barycentre, which corresponds to using the ground barycentre map B (this time with weights given by the disintegration of the coupling in question).
Remark 4.2. *The metric space $(\mathcal{N}, \mathrm{W}_2)$ is not compact, however we consider discrete measures (GMMs). We will show how one can restrict $\mathcal{N}$ to a compact subset containing all barycentres. Combining [DD20] Corollary 3 and [Álv+16] Theorem 4.2 (Equations 20 and 21), shows that the barycentre is within a certain compact subset of $\mathcal{P}(\mathcal{N})$ of measures supported on Gaussians with covariances whose eigenvalues are in a segment $[r, R]$, where $0 < r < R$ are explicit constants depending on the covariances of the components of $\nu_1, \cdots, \nu_K$. As for the means, they can be constrained to the convex hull of the means of the components of the mixtures ν_k.*

5 Numerical Illustrations

In this section, we provide numerical experiments to illustrate the fixed-point method (specifically its barycentric variant presented in Algorithm 3) on various toy datasets. All code from this section is available in our companion Python toolkit. A numerical implementation of Algorithm 2, which allows flexible support sizes, is also possible, but computationally much less appealing than Algorithm 3.

5.1 Toy Example for Barycentre Computation

We begin with a simple example of barycentre computation in $\mathbb{R}^2$ of two discrete uniform measures with different support sizes and for the square-Euclidean cost $c_k(x, y) = \|x - y\|_2^2$. We observe convergence to the true barycentre in two iterations in Fig. 8. The support size increases from 10 to 19 at the first iteration and remains at 19 at the final iteration.

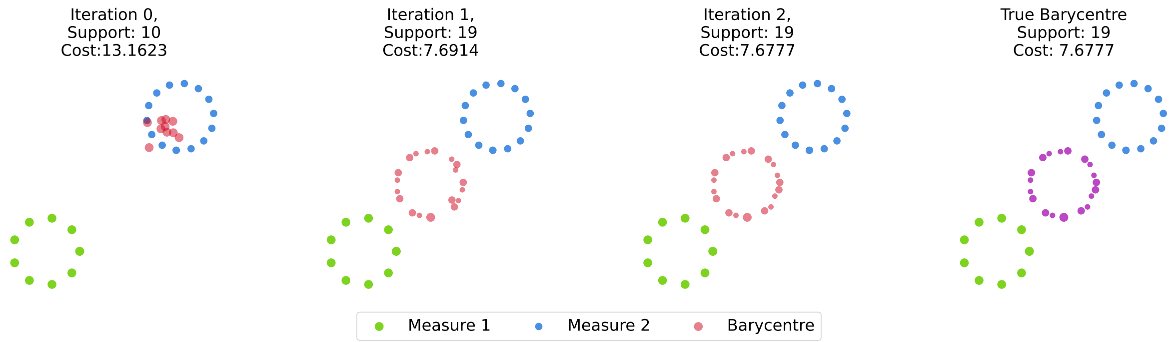


Figure 8: Iterations of Algorithm 2 for the square-Euclidean cost, and comparison with the true W_2^2 barycentre.

5.2 Illustration with Norm Powers

We consider discrete measures in $\mathbb{R}^2$ for costs $c_k(x, y) = \|x - y\|_p^q$, as illustrated in Fig. 9.

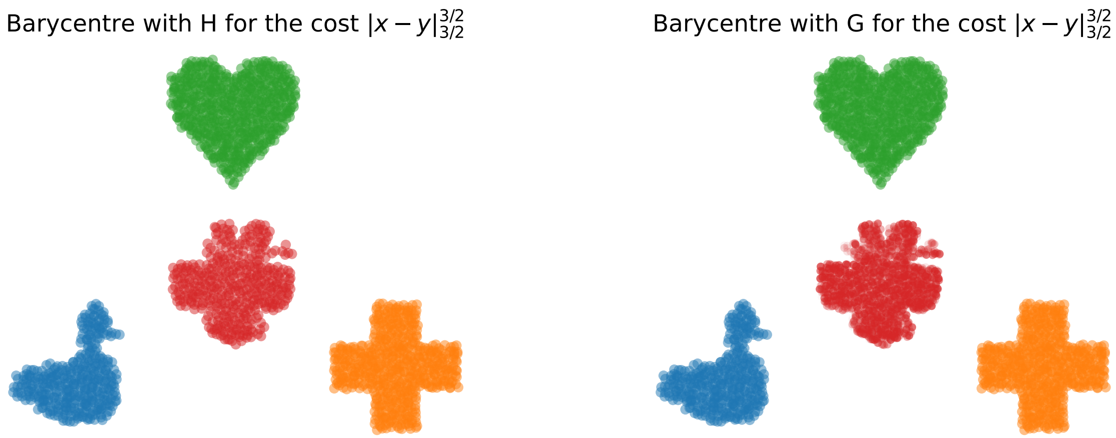
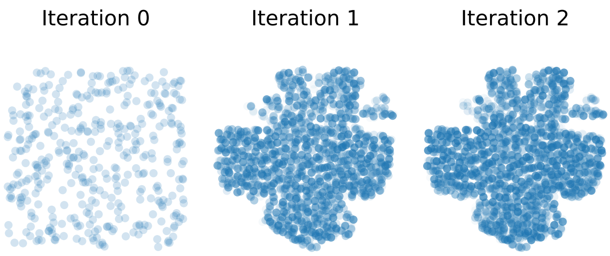
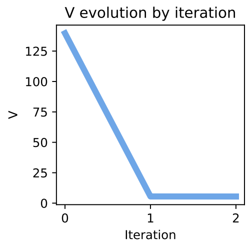


Figure 9: Barycentres with initial support size $n = 400$ for $(p, q) = (\frac{3}{2}, \frac{3}{2})$ of three measures with sizes 561, 382, 629.

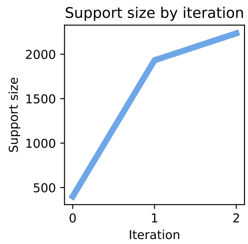
In Fig. 10, we observe that for $(p, q) = (\frac{3}{2}, \frac{3}{2})$, the iterates of G (Algorithm 2) have an energy that converges in one iteration, but the support size continues to grow at iteration 2. As for H (Algorithm 3), we observe in Fig. 11 convergence in one iteration. In Fig. 12, we present barycentres for various pairs (p, q) using iterates of H.



(a) Fixed-point iterations for $(p, q) = (\frac{3}{2}, \frac{3}{2})$.

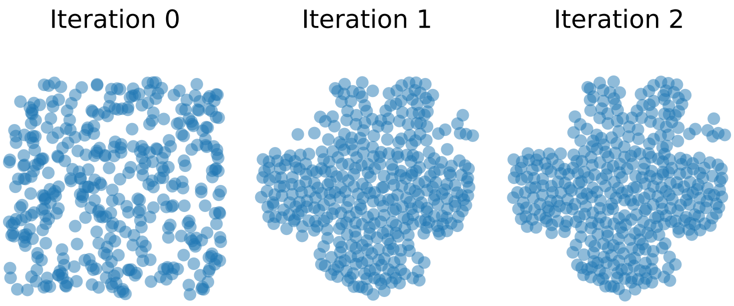


(b) Energy V.

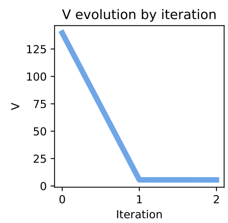


(c) Support size.

Figure 10: Convergence of the iterations of G (Algorithm 2).



(a) Fixed-point iterations for $(p, q) = (\frac{3}{2}, \frac{3}{2})$.



(b) Barycentre energy V of the iterations.

Figure 11: Convergence of the iterations of H (Algorithm 3).

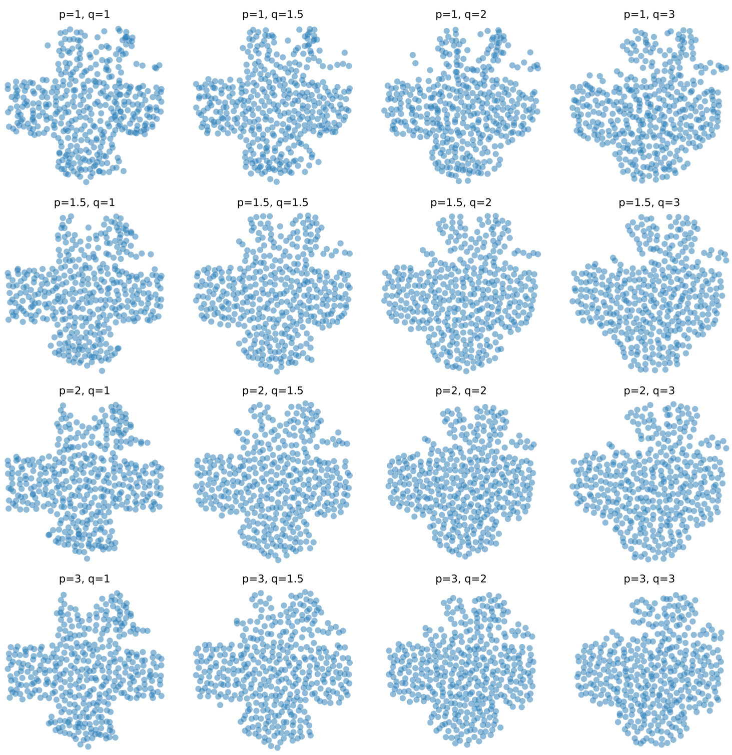


Figure 12: Barycentres for the cost $\|x - y\|_p^q$ for different values of (p, q).

In the following, we consider a different setting where two of the three target measures are identical, and with a different third target. This will allow us to study the robustness properties of the associated barycentre, seeing the third different measure as an outlier. We represent the target measures and a barycentre in Fig. 13, and compare different barycentres varying the parameters (p, q) of the cost $\| \cdot - \cdot \|_p^q$ in Fig. 14. We observe that the barycentre obtained for $q = 1$ always takes the shape of the duck, as this power allows for greater robustness to outliers (here the heart-shaped cloud), regardless of the norm. The influence of the third point cloud becomes increasingly evident as p and q grow.

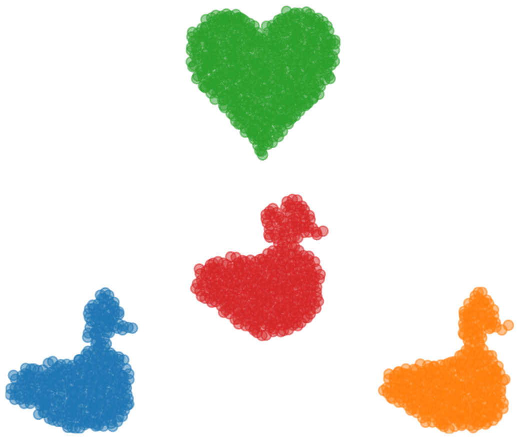


Figure 13: Barycentre of three point clouds for the cost $\| \cdot - \cdot \|_{3/2}^{3/2}$.

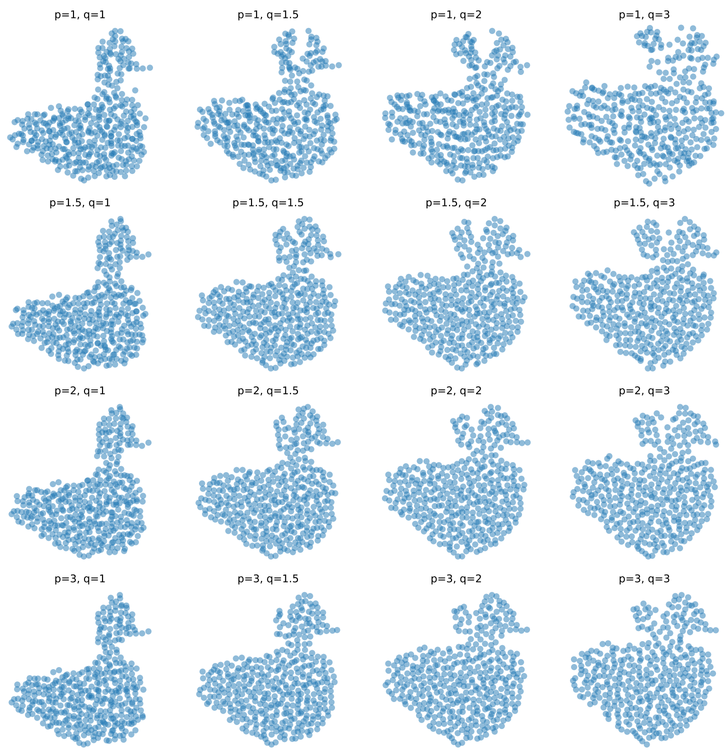
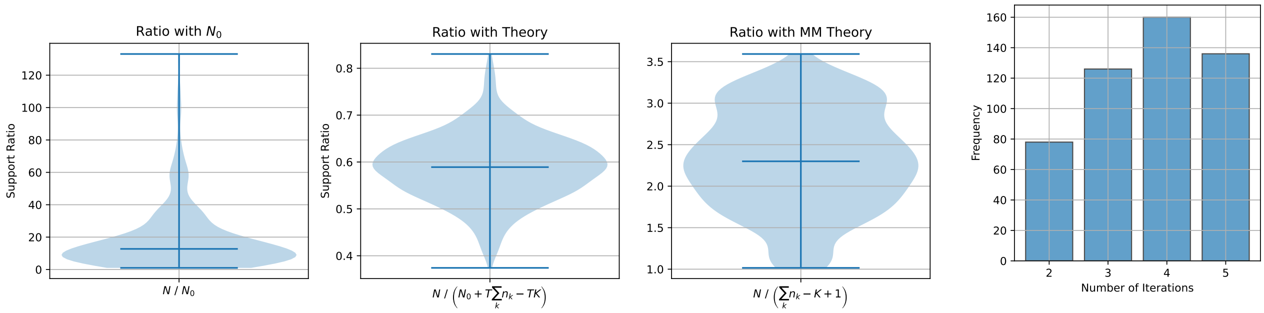


Figure 14: Barycentres for the measures of figure 13, for norms p raised to the power q.

5.3 Study of the Support Size of Iterates of G

In this section, we study the support size N of the final iteration of G (Algorithm 2). As discussed in Section 4.1, we expect (without formal proof) that the support size after T

iterations is upper-bounded by $N_0 + T\sum_k n_k - TK$, where N_0 is the initial support size and n_k is the size of the k-th marginal. We verify this hypothesis on numerical experiments on numerous configurations varying $N_0, (n_k), d, (d_k)$ with measure points and weights generated randomly and for the square-Euclidean cost in Fig. 15. We observe that the upper-bound is indeed respected, and that the algorithm attains convergence in a small number of iterations.



(a) Ratios of the final support size N with various support sizes.

(b) Iterations until convergence.

Figure 15: Numerical study of the support size N of iterates of G. We ran 500 samples, each drawing a random number of measures $K \in [\![2, 10]\!]$, random dimensions $d, d_1, \cdots, d_K \in [\![1, 20]\!]$, random initial support sizes $N_0 \in [\![10, 100]\!]$ and random measure sizes $n_k \in [\![10, 100]\!]$.

Running the same experiment as in Fig. 15 with $N_0 = n_1 = \cdots = n_K$ and uniform measure weights, we obtain, as expected in Eq. (28) that the support size N_t remains constant.

5.4 Comparison with the Multi-Marginal Formulation

Following Eq. (7), the discrete OT barycentre problem has a multi-marginal formulation, which can be written as follows, given measures $\nu_k = \sum_{j=1}^{m_k} b_{k,j}\delta_{y_{k,j}}$:

$$\underset{\pi \in \Pi(b_1,\cdots,b_K)}{\operatorname{argmin}} \sum_{j_1,\cdots,j_K} \pi_{j_1,\cdots,j_K} \sum_{k=1}^{K} c_k(B(y_{1,j_1},\cdots,y_{K,j_K}), y_{k,j_k}). \tag{38}$$

Numerical solvers for Eq. (38), while slow, allow the computation of the exact solution of the barycentre problem. Comparing this solution to the output of our algorithm is technical, since the barycentric version of our algorithm imposes the size of the support of the barycentre in addition to imposing the weights, which introduces bias. We aim to illustrate that the speed of the barycentric algorithm, with a quantitative study of the error with respect to the multi-marginal "ground truth". Note that even in this square-euclidean experiment, there is no widespread multi-marginal solver, which is why we also contribute an implementation.

The experimental setup is the following: the K measures ν_k are all uniform measures with n points in $\mathbb{R}^d$ drawn independently from $\mathcal{N}(0, 1)$. For the fixed-point algorithm, the initial measure is also taken as a uniform measure over n points with $\mathcal{N}(0, 1)$ samples. We compare different numbers of iterations of the fixed-point algorithm and different choices of n, d, K. The plots show the ratios of the energy V and computation times for our algorithm divided by a Linear Programming multi-marginal solver, plotting 30% and 70% quantiles across 10 samples for each configuration. As expected in Eq. (28), since in this case the measures are uniform with a common support size, the iterates of H and G are identical in this setting.

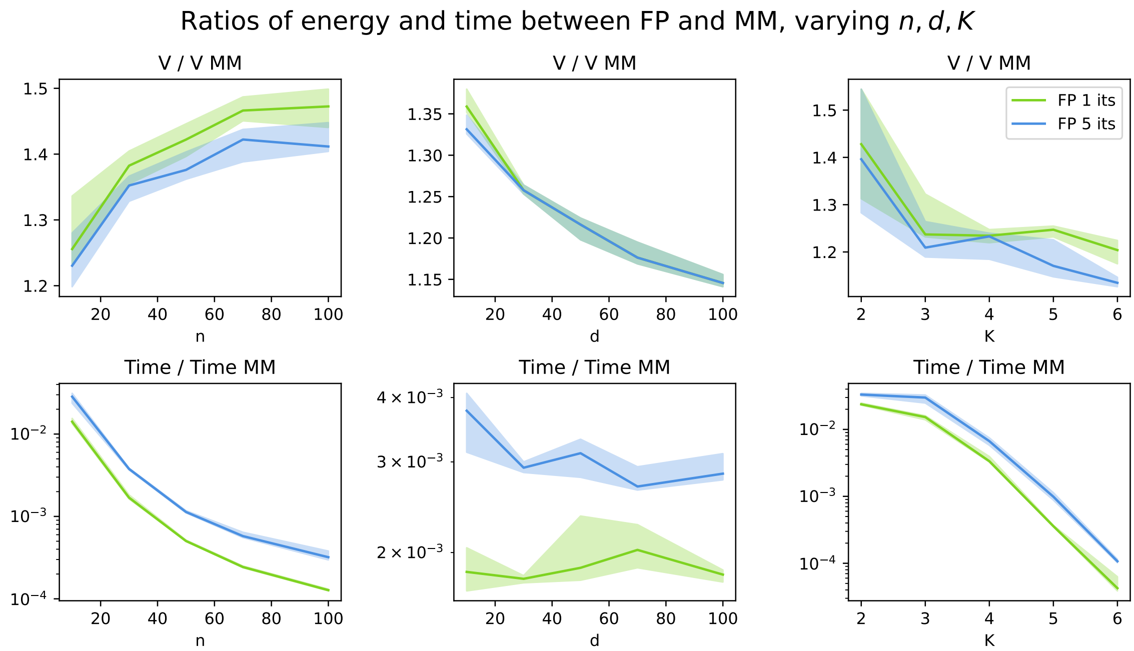


Figure 16: Comparing the fixed-point solver with a linear programming multi-marginal solver. From left to right columns: varying n with $d = 10$ and $K = 3$; varying d with $n = 30$ and $K = 3$; varying K with $n = 10$ and $d = 10$. The comparison is made by dividing the energy value V (resp. computation time) of the fixed-point solution by the multi-marginal solution. The different curves correspond to $T = 1, 5, 10$ iterations (legend in the top-right).

From the results presented in Fig. 16, it appears that the fixed-point algorithm converges in very few iterations, has an energy at most 50% worse than the exact multi-marginal solution, and is orders of magnitude faster, especially for larger measure sizes n and for greater numbers of marginals K. Note that for $n \geq 10$ and $K \geq 10$ for example, the multi-marginal problem is computationally intractable.

To compare with similar barycentre support sizes, in Fig. 17 we experiment with fixed-point barycentres using H (Algorithm 3) with $N_{\text{FP}} = (n-1)K+1$ points. The rationale behind this choice stems from the fact that discrete measures with $n_1, \cdots, n_K$ points have a barycentre with at most $\sum_k n_k - K + 1$ points ([ABM16] Theorem 2[3]).

[3] whose techniques are in fact not specific to the cost $\|\cdot - \cdot\|_2^2$

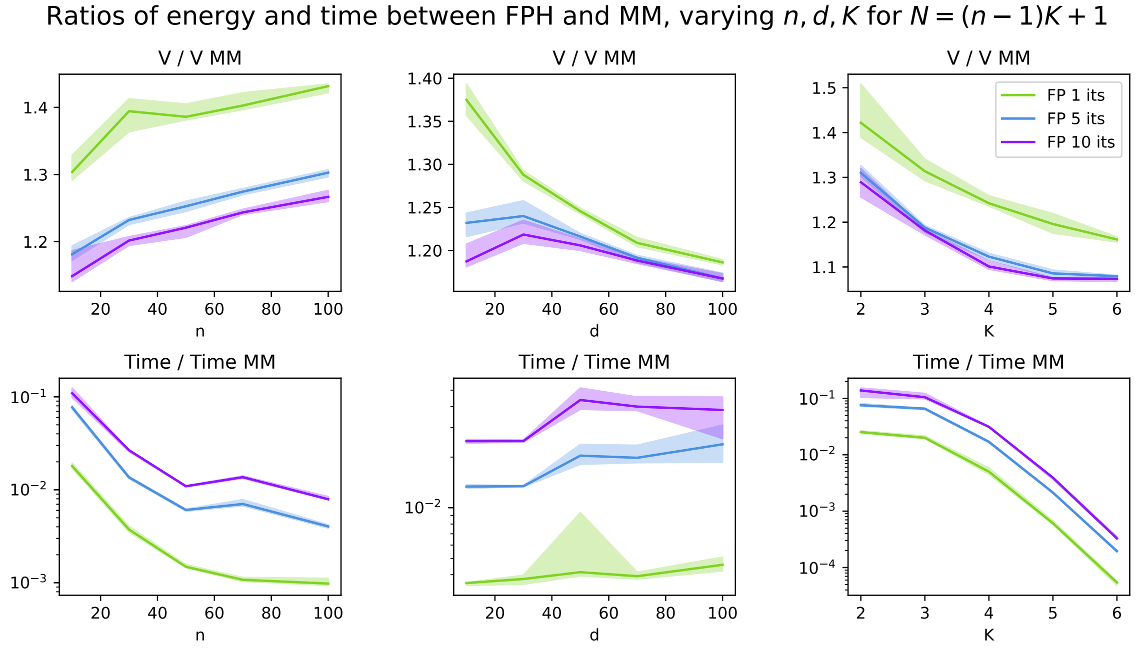


Figure 17: Comparing the fixed-point solver from Algorithm 3 for $N_{\text{FP}} = (n-1)K + 1$ and the same setup as in Fig. 16.

We now focus on the iterations of G (Algorithm 2) in the case of uniform measures where the initialisation is taken with n points and the target measures have even spaced sizes $n_1 = \frac{n}{2} \cdots n_K = 2n$. This ensures that iterates of G differ from iterates of H, and we present the results in Fig. 18.

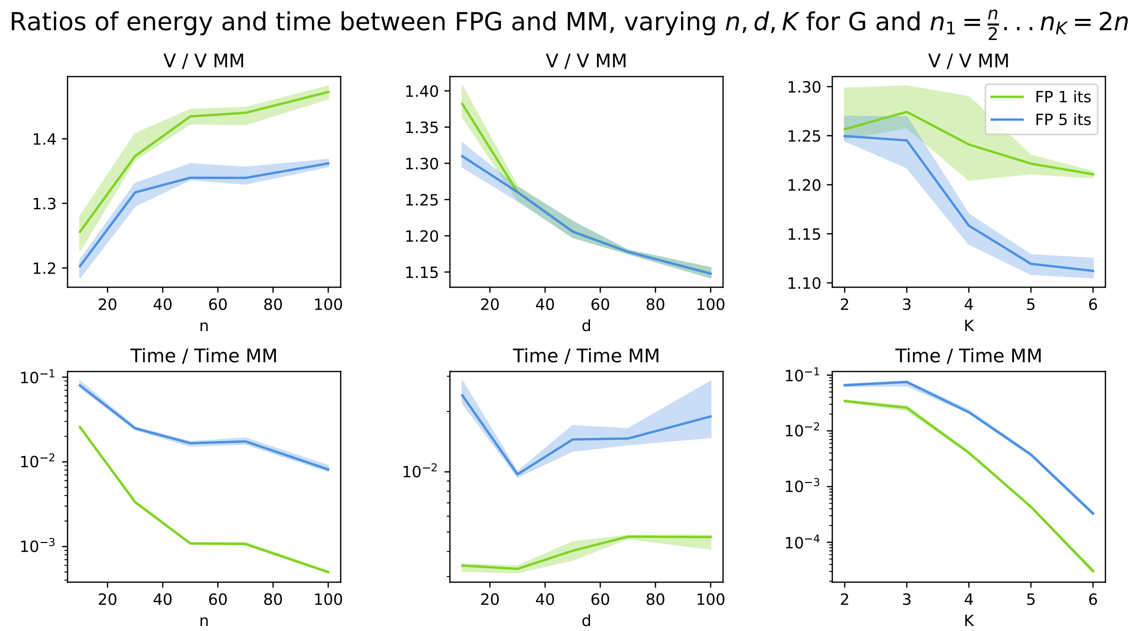
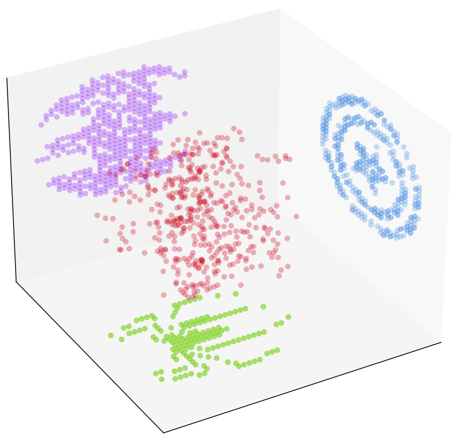


Figure 18: Comparing the fixed-point solver from Algorithm 2 with the MM solver, for an initialisation with n points and target measures with different sizes $n_1 = \frac{n}{2} \cdots n_K = 2n$.

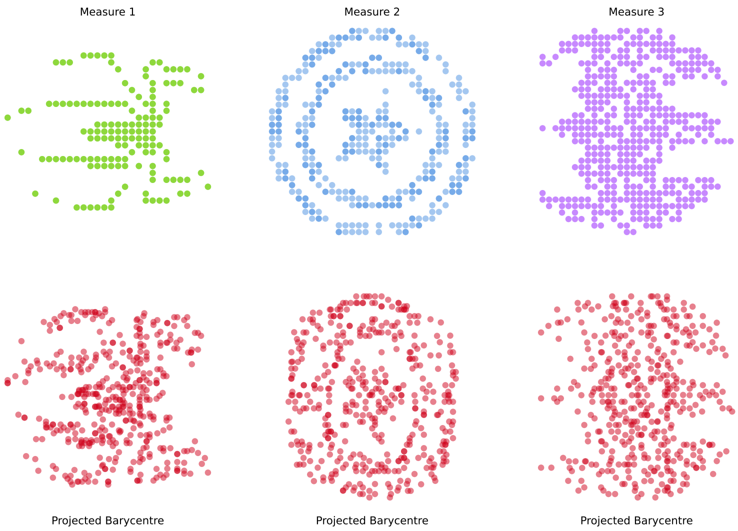
Figs. 16 to 18 suggests that the fixed-point methods proposed in Algorithms 2 and 3a re useful as a fast approximate solvers for the barycentre problem, and that settings with larger barycentre supports may require more iterations to converge. The main takeaway is that our methods remain competitive for large supports and number of target measures, yet its convergence speed and overall advantages are more pronounced for smaller supports.

5.5 Generalised Wasserstein Barycentre Computation

In Fig. 19a, we illustrate the case where $c_k(x, y) = \|P_k x - y\|_2$, where $P_k : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ is an orthogonal projection. The problem finds a 3D measure whose projections attempt to match the reference 2D measures, which we compare in Fig. 19b. This is a modification of the exponent 2 from Generalised Wasserstein Barycentres [DGS21].



(a) Barycentre in $\mathbb{R}^3$ with immersed measures $\nu_k \in \mathcal{P}(\mathbb{R}^2)$.



(b) Projections of the barycentre into $\mathbb{R}^2$.

Figure 19: Barycenter (using Algorithm 3) with costs $c_k(x, y) = \|P_k x - y\|_2$, where P_k are orthogonal projections from $\mathbb{R}^3$ to the three axes-aligned planes of the orthonormal basis. We provide an animation in the companion code.

5.6 Non-linear Generalised Wasserstein Barycentre Computation

In this illustration, we look for a barycentre in $\mathbb{R}^2$ whose projections onto different circles match measures on these circles. We choose the costs $c_k(x, y) = \|P_k(x) - y\|_2^2$, where P_k is the projection onto the circle k. Since P_k is not linear, this is a direct generalisation of [DGS21].

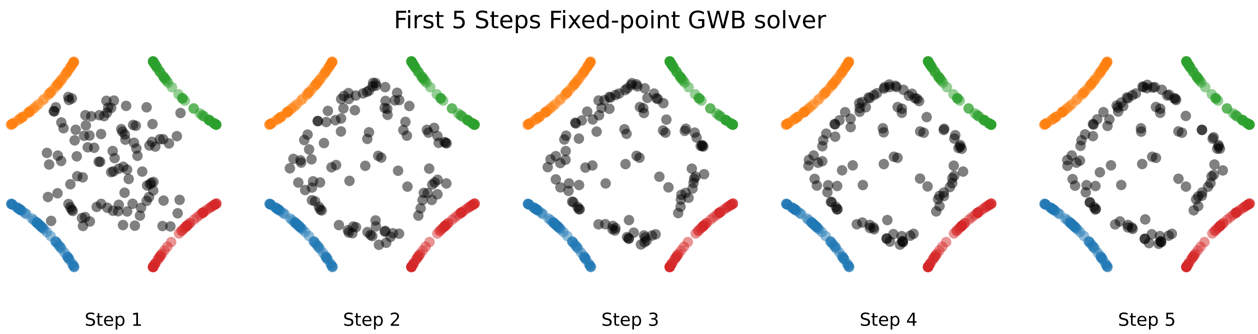


Figure 20: First 5 iterations of the fixed-point algorithm (Algorithm 3) for costs $c_k(x, y) = \|P_k(x) - y\|_2^2$, where P_k are projections onto four different circles on which the ν_k are supported (plotted in colour).

In this instance, convergence happens quickly, but a stationary point is only reached after about 5 iterations, as observed on the steps in Fig. 20 and on the energy curve in Fig. 21.

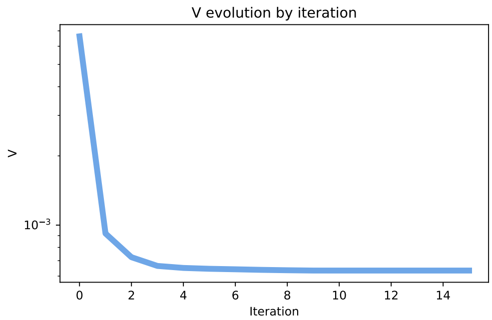


Figure 21: Barycentre energy V of the fixed-point algorithm for H across iterations.

5.7 Gaussian Mixture Model Barycentres

We illustrate numerical solutions of the GMM Barycentre method introduced in Section 4.4. In Fig. 22, we compare the multi-marginal solution with the output of our algorithm (we use Algorithm 3).

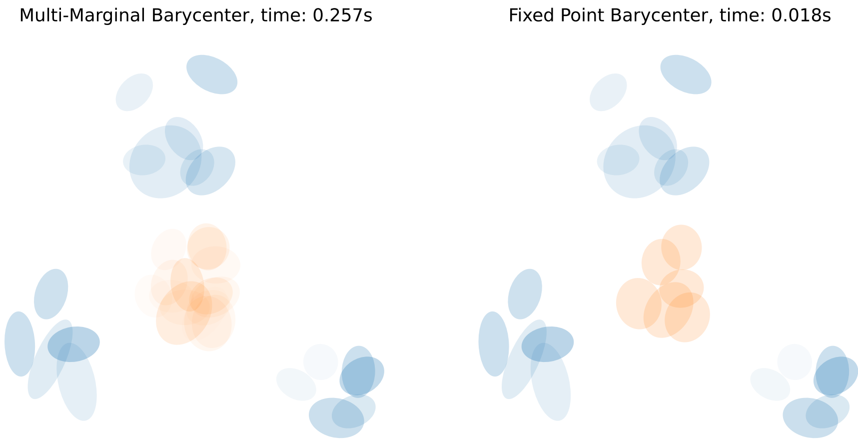


Figure 22: Left: multi-marginal solution for the GMM barycentre problem. Right: fixed-point solution for $n = 6$ components.

Finally, in Fig. 24 we illustrate barycentres between 4 GMMs shown in Fig. 23 with different weights.

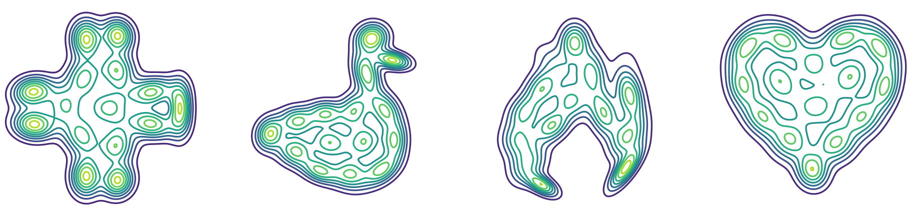


Figure 23: Four GMMs of which we will compute barycentres in Fig. 24.

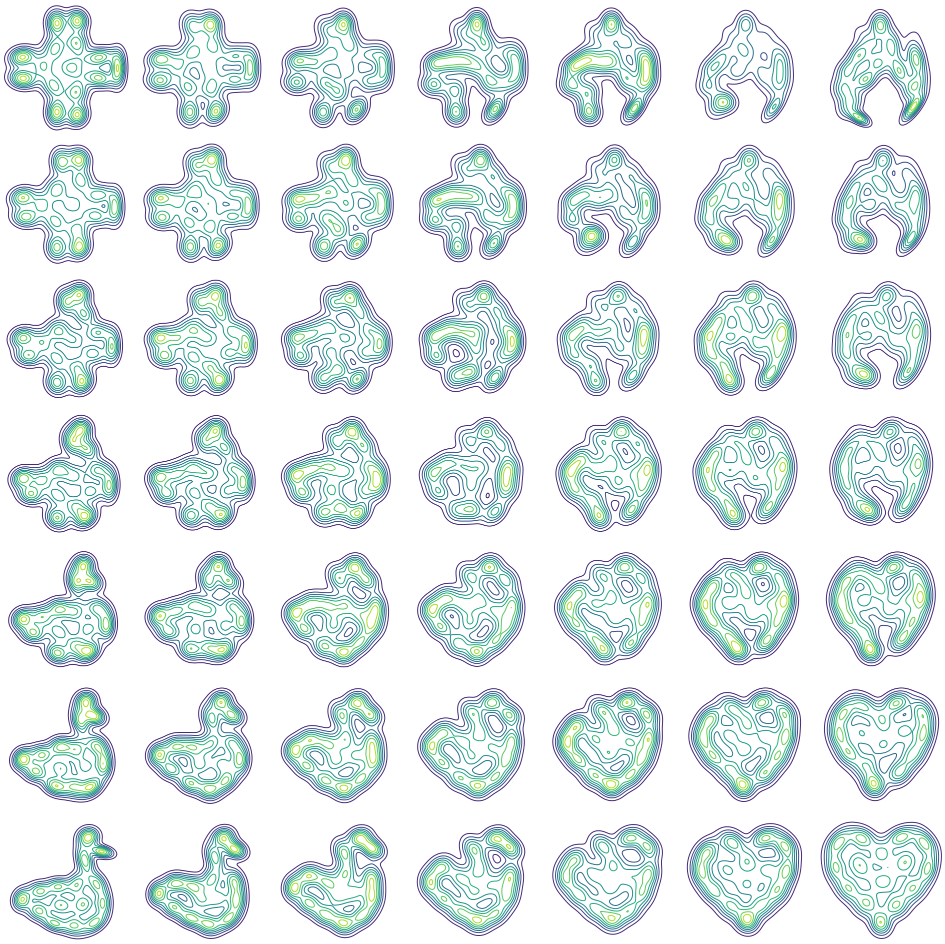


Figure 24: GMM barycentres between the four corner GMMs computed with the fixed-point solver with $n = 15$ components. The GMMs are represented by the contours of their densities on $\mathbb{R}^2$.

5.8 Colour Transfer on a Barycentre of Colour Distributions

In this final experiment, we consider a colour transfer problem. The goal is to compute the barycentre of the colour distributions of several (here three) source images, some of which contain outlier colours, and then use this barycentre as a target measure to modify the colours of a new image (referred to as the *input* here). Figure 25 shows the source images, the *input* image, and the same *input* image after transferring its colour distribution to that of the colour barycentre of the source images. The barycentre is computed either for a W_1 cost or for a W_2^2 cost. This transfer is evaluated on downsampled images, with the RGB matching of a colour c in the high-resolution image subsequently chosen as $c + \tau$, where τ is the colour translation obtained for the closest colour to c in the downsampled image (this amounts to viewing the matching as a piecewise constant translation field). Figure 26 shows the colour distributions of the images in the RGB space. We observe that the W_1 cost enjoys greater robustness to the colour outliers compared to the usual W_2^2 cost.

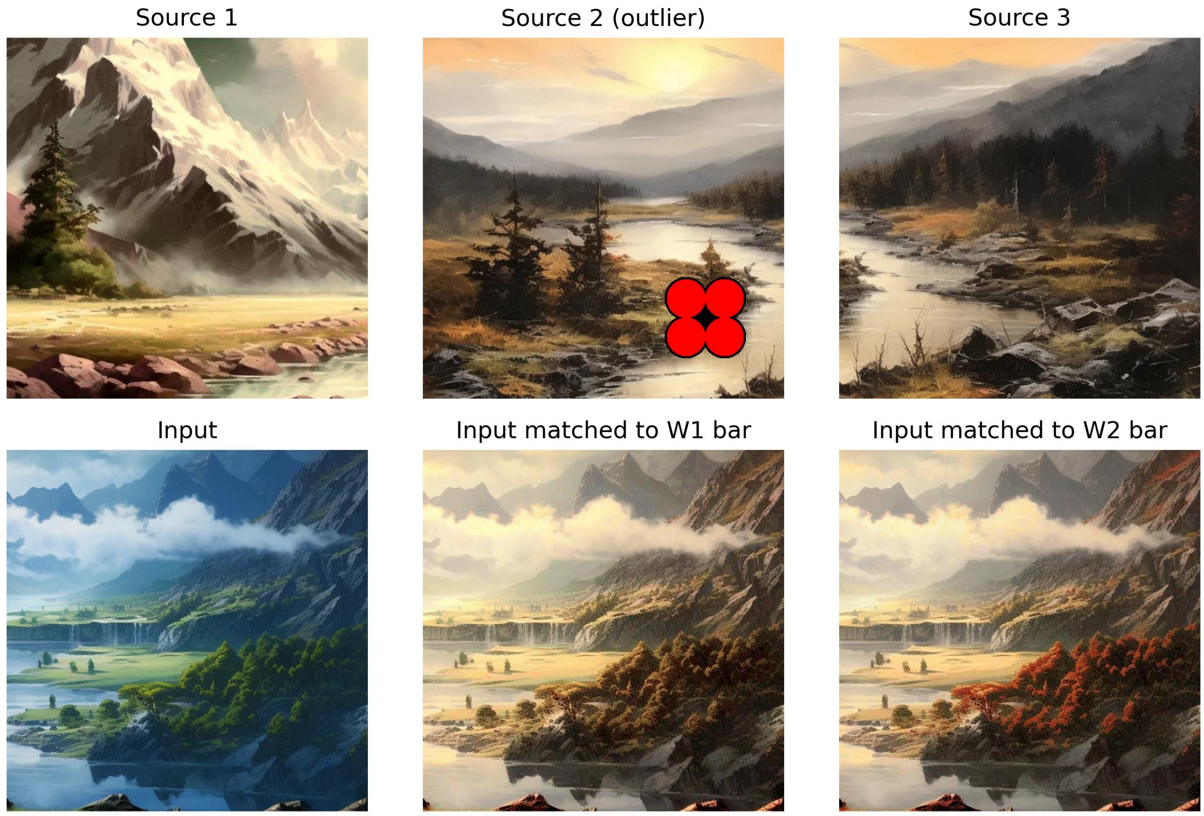


Figure 25: Colour transfer applied to the input image towards the colour barycentre of the source images, for the costs W_1 and W_2^2. One of the source images contains unwanted colour artifacts, which we see as outliers.

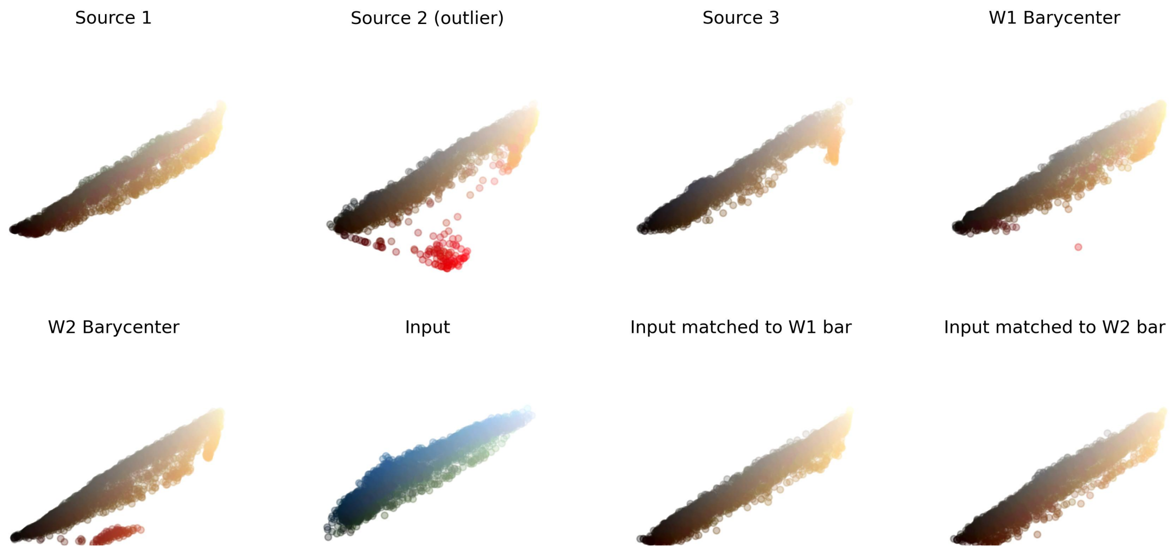


Figure 26: Colour distributions of the different images from figure 25 as well as the W_2^2 and W_1 barycentres of the source images.

Future Directions

There are numerous directions for future research. To begin with, in Theorems 3.10 and 3.13, we show subsequential convergence to fixed-points of G (resp. G_ε), which may not be barycentres. In cases where barycentres and fixed points may not be unique such as the discrete setting, it remains unclear if there exists fixed points that are not barycentres.

The barycentric fixed-point algorithm (iterating Eq. (21)) has no theoretical guarantees of convergence. Given its computational advantages and its current use in practice for the squared Euclidean cost ([CD14], [Fla+21]), this is a timely question.

In Section 3.3, we required a notion of barycentric projection for couplings $\pi \in \Pi^*_{c_k}(\mu, \nu_k)$. In $\mathbb{R}^d$, the underlying convex combinations are performed using the usual linear structure, however this does not generalise to arbitrary metric spaces. To consider these object more formally on generic (compact) metric spaces, it would be necessary to discuss in more detail the meaning of expectation in a space without a linear structure.

Throughout this work, we relied heavily on Assumption 2, but in practice this can be difficult to verify for costs c_k: beyond the case $c_k = h(x-y)$ with h strictly convex, it is difficult to provide large classes of costs that yield this property on B (other examples include $c_k(x, y) = \|P_k x - y\|_2^2$ as in [DGS21] or W_2^2 for absolutely continuous measures). One could alternatively investigate a theoretical framework where B is a multi-function.

In the absolutely continuous case, the Twist condition can ensure uniqueness of the barycentre, as explained in Remark 2.1. A natural question concerns almost-sure uniqueness in the discrete case, as was partially explored in Section 4.3.

From a numerical standpoint, it has been observed that the fixed-point algorithm converges in very few iterations. A theoretical work extending the discrete Wasserstein case from [Lin23] would bridge a significant gap between theory and practical observation.

Acknowledgements

We would like to thank Christophe Gaillac for the initial discussions that motivated the introduction of barycentres with generic costs. This research was funded in part by the Agence nationale de la recherche (ANR), Grant ANR-23-CE40-0017 and by the France 2030 program, with the reference ANR-23-PEIA-0004.

References

[AC11] Martial Agueh and Guillaume Carlier. "Barycenters in the Wasserstein Space". In: *SIAM Journal on Mathematical Analysis* 43.2 (2011), pp. 904–924. eprint: https://doi.org/10.1137/100805741.

[AC17] Martial Agueh and Guillaume Carlier. "Vers un théorème de la limite centrale dans l'espace de Wasserstein?" In: *Comptes Rendus. Mathématique* 355.7 (2017), pp. 812–818.

[AB94] Charalambos D. Aliprantis and Kim C. Border. "Correspondences". In: *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 458–520.

[Alt+21] Jason Altschuler, Sinho Chewi, Patrik R Gerber, and Austin Stromme. "Averaging on the Bures-Wasserstein manifold: dimension-free convergence of gradient descent". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 22132–22145.

[AB21] Jason M. Altschuler and Enric Boix-Adsera. *Wasserstein barycenters are NP-hard to compute*. 2021.

[Álv+16] Pedro C Álvarez-Esteban, E Del Barrio, JA Cuesta-Albertos, and C Matrán. "A fixed-point approach to barycenters in Wasserstein space". In: *Journal of Mathematical Analysis and Applications* 441.2 (2016), pp. 744–762.

[AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.

[ABM16] Ethan Anderes, Steffen Borgwardt, and Jacob Miller. "Discrete Wasserstein barycenters: Optimal transport for discrete data". In: *Mathematical Methods of Operations Research* 84 (2016), pp. 389–409.

[Bac+22] Julio Backhoff-Veraguas, Joaquin Fontbona, Gonzalo Rios, and Felipe Tobar. "Bayesian learning with Wasserstein barycenters". In: *ESAIM: Probability and Statistics* 26 (2022), pp. 436–472.

[BB25] Florian Beier and Robert Beinert. "Tangential fixpoint iterations for Gromov–Wasserstein barycenters". In: *SIAM Journal on Imaging Sciences* 18.2 (2025), pp. 1058–1100.

[BBS23] Florian Beier, Robert Beinert, and Gabriele Steidl. *Multi-Marginal Gromov-Wasserstein Transport and Barycenters*. 2023. arXiv: 2205.06725 [math.OC].

[Ben+15] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. "Iterative Bregman projections for regularized transportation problems". In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138.

[BJL17] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. "On the Bures-Wasserstein distance between positive definite matrices". In: *arXiv* (Dec. 2017). eprint: 1712.01504.

[BCP19] Jérémie Bigot, Elsa Cazelles, and Nicolas Papadakis. "Penalization of barycenters in the Wasserstein space". In: *SIAM Journal on Mathematical Analysis* 51.3 (2019), pp. 2261–2285.

[BPC16] Nicolas Bonneel, Gabriel Peyré, and Marco Cuturi. "Wasserstein barycentric coordinates: histogram regression using optimal transport." In: *ACM Trans. Graph.* 35.4 (2016), pp. 71–1.

[BFR24a] Camilla Brizzi, Gero Friesecke, and Tobias Ried. *h-Wasserstein barycenters*. 2024. arXiv: 2402.13176 [math.AP].

[BFR24b] Camilla Brizzi, Gero Friesecke, and Tobias Ried. "p-Wasserstein barycenters". In: *arXiv preprint arXiv:2405.09381* (2024).

[CCE24] Guillaume Carlier, Enis Chenchene, and Katharina Eichinger. "Wasserstein Medians: Robustness, PDE Characterization, and Numerics". In: *SIAM Journal on Mathematical Analysis* 56.5 (2024), pp. 6483–6520.

[CE10] Guillaume Carlier and Ivar Ekeland. "Matching for teams". In: *Economic theory* 42 (2010), pp. 397–418.

[CD14] Marco Cuturi and Arnaud Doucet. "Fast Computation of Wasserstein Barycenters". In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P. Xing and Tony Jebara. Vol. 32. Proceedings of Machine Learning Research. Bejing, China: PMLR, June 2014, pp. 685–693.

[DD20] Julie Delon and Agnes Desolneux. "A Wasserstein-type distance in the space of Gaussian mixture models". In: *SIAM Journal on Imaging Sciences* 13.2 (2020), pp. 936–970.

[DGS21] Julie Delon, Nathaël Gozlan, and Alexandre Saint-Dizier. *Generalized Wasserstein barycenters between probability measures living on different subspaces*. 2021.

[Fla+21] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain

Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. "POT: Python Optimal Transport". In: *Journal of Machine Learning Research* 22.78 (2021), pp. 1–8.

[GNB22] Promit Ghosal, Marcel Nutz, and Espen Bernton. "Stability of entropic optimal transport and Schrödinger bridges". In: *Journal of Functional Analysis* 283.9 (2022), p. 109622.

[Gor+19] Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. "Obtaining fairness using optimal transport theory". In: *International conference on machine learning*. PMLR. 2019, pp. 2357–2365.

[Goz+17] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali. "Kantorovich duality for general transport costs and applications". In: *Journal of Functional Analysis* 273.11 (2017), pp. 3327–3405.

[Ho+17] Nhat Ho, XuanLong Nguyen, Mikhail Yurochkin, Hung Hai Bui, Viet Huynh, and Dinh Phung. "Multilevel clustering via Wasserstein means". In: *International conference on machine learning*. PMLR. 2017, pp. 1501–1509.

[KP17] Young-Heon Kim and Brendan Pass. "Wasserstein barycenters over Riemannian manifolds". In: *Adv. Math.* 307 (2017), pp. 640–683.

[Kor+22] Alexander Korotin, Vage Egiazarian, Lingxiao Li, and Evgeny Burnaev. "Wasserstein iterative networks for barycenter estimation". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 15672–15686.

[Lin23] Johannes von Lindheim. "Simple approximative algorithms for free-support Wasserstein barycenters". In: *Computational Optimization and Applications* 85.1 (2023), pp. 213–246.

[MDC20] Quentin Mérigot, Alex Delalande, and Frederic Chazal. "Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 3186–3196.

[Mi+18] Liang Mi, Wen Zhang, Xianfeng Gu, and Yalin Wang. "Variational wasserstein clustering". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 322–337.

[MM21] Eduardo Fernandes Montesuma and Fred Maurice Ngole Mboula. "Wasserstein barycenter for multi-source domain adaptation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 16785–16793.

[PC19] G. Peyré and M. Cuturi. "Computational Optimal Transport". In: *Foundations and Trends in Machine Learning* 51.1 (2019), pp. 1–44.

[PW14] Yury Polyanskiy and Yihong Wu. "Lecture notes on information theory". In: *Lecture Notes for ECE563 (UIUC) and* 6.2012-2016 (2014), p. 7.

[Rab+12] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. "Wasserstein barycenter and its application to texture mixing". In: *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*. Springer. 2012, pp. 435–446.

[San15] Filippo Santambrogio. "Optimal transport for applied mathematicians". In: *Birkäuser, NY* 55.58-63 (2015), p. 94.

[Sol+15] Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. "Convolutional wasserstein distances: Efficient optimal transportation on geometric domains". In: *ACM Transactions on Graphics (ToG)* 34.4 (2015), pp. 1–11.

[Vil09] Cédric Villani. *Optimal transport : old and new / Cédric Villani*. eng. Grundlehren der mathematischen Wissenschaften. Berlin: Springer, 2009.