# Sliced Optimal Transport Plans

Eloi Tanguy[1], Laetitia Chapel[2], and Julie Delon[1]

[1]Université Paris Cité, CNRS, MAP5, F-75006 Paris, France
[2]Institut Agro Rennes-Angers, IRISA

2nd August 2025

**Abstract**

Since the introduction of the Sliced Wasserstein distance in the literature, its simplicity and efficiency have made it one of the most interesting surrogate for the Wasserstein distance in image processing and machine learning. However, its inability to produce transport plans limits its practical use to applications where only a distance is necessary. Several heuristics have been proposed in the recent years to address this limitation when the probability measures are discrete. In this paper, we propose to study these different propositions by redefining and analysing them rigorously for generic probability measures. Leveraging the $\nu$-based Wasserstein distance and generalised geodesics, we introduce and study the Pivot Sliced Discrepancy, inspired by a recent work by Mahey et al.. We demonstrate its semi-metric properties and its relation to a constrained Kantorovich formulation. In the same way, we generalise and study the recent Expected Sliced plans introduced by Liu et al. for completely generic measures. Our theoretical contributions are supported by numerical experiments on synthetic and real datasets, including colour transfer and shape registration, evaluating the practical relevance of these different solutions.

# Table of Contents

# 1   Introduction

Known for its ability to capture geometric structure in probability distributions, optimal transport has attracted considerable attention in both theoretical and applied fields. Several studies have developed its mathematical foundations in great detail [San15; Vil09], and its practical impact has been demonstrated on a broad spectrum of applications. Originally developed for applications in logistics, economics [Gal17] and fluid mechanics, computational optimal transport has also emerged in the last fifteen years as a central tool in data science. It is used nowadays for a large variety of applications, ranging from image processing, computer vision and computer graphics [RDG09; HHR22; Fey+17; BD23; Pon+21], to domain adaptation [Cou+16; MM21; Fat+21], natural language processing [Che+], generative modelling [ACB17; Gul+17; Sal+18; Ton+24; HCD25], quantum chemistry [BDG12] or biology [Bun+24; NS25], to cite just a few.

In these applications, optimal transport is used to define meaningful discrepancies between probability distributions, taking into account the underlying geometry of the data, but also as a way to define optimal plans or maps between such data, in order to transform a given distribution into another in an optimal way. In the continuous setting, we recall that the 2-Wasserstein distance $W_2$ between two probability measures $\mu$ and $\nu$ on $\mathbb{R}^d$ is defined as:

$$W_2^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_2^2 \, \mathrm{d}\pi(x, y),$$

where $\Pi(\mu, \nu)$ is the set of couplings with marginals $\mu$ and $\nu$. In the discrete case, with empirical measures supported on finite point clouds, this problem becomes a linear program over a polytope. Computing Wasserstein distances between discrete datasets comes with significant computational expense. Classical linear programming solvers used to evaluate the transport cost between two discrete measures of size $n$ typically have a complexity of $\mathcal{O}(n^3 \log n)$ [PC19]. This limitation has motivated the development of computationally lighter surrogates or approximations that preserve key characteristics of optimal transport metrics.

One of these popular and efficient surrogate is the Sliced Wasserstein distance (SW) [Rab+12a; Bon+15]. This approach leverages the fact that in one dimension, the Wasserstein distance has a closed-form solution. The Sliced Wasserstein distance is derived by averaging 1D Wasserstein distances over all directions on the unit sphere, offering a simple alternative to $W_2$:

$$SW_2^2(\mu, \nu) = \int_{\mathbb{S}^{d-1}} W_2^2(P_\theta \# \mu, P_\theta \# \nu) \, \mathrm{d}\theta,$$

where $P_\theta$ denotes the projection onto direction $\theta$. Since evaluating the full integral is intractable in practice, it is approximated by Monte Carlo sampling. One draws $L$ random directions, computes the 1D Wasserstein distance for each, and averages the results. The 1D Wasserstein distance between empirical distributions of $n$ points can be obtained in $\mathcal{O}(n \log n)$, so the approximate $SW_2$ distance can be computed in $\mathcal{O}(Ln \log n)$. This efficiency makes it especially appealing for large values of $n$.

The SW distance remains a true distance on the space of probability measures and retains several fundamental features of Wasserstein distances. For probability measures with compact support, it has been shown to be equivalent to the Wasserstein distance [Bon13]. It also has desirable statistical properties, such as sample complexity bounds and robustness [Nad+20]. Its efficiency has been confirmed in numerous use cases, including domain adaptation [Lee+19], texture generation, colour and style transfer [Hei+21; Bon+15; EW22], statistical inference [KRH18], generative modelling [DZS18; Wu+19; CTV25], auto-encoder regularisation [Kol+18], topological data analysis [SDT25] or shape analysis [Le+24; NNH23]. Extensions to Riemannian settings have also been investigated [BDC25]. Nevertheless, a key limitation of SW is that it does not provide a transport plan or a map between distributions, which limits its use in applications that require correspondences between datasets.

To circumvent this issue, several heuristics have been proposed to extract approximate transport plans from SW. A notable example is the use of stochastic gradient descent (SGD) to minimise the objective

$X \longmapsto \mathrm{SW}(\delta_X, \delta_Y)$, as a way to gradually move points from a source point cloud $X$ to a target point cloud[1] $Y$. This strategy has been first explored for colour transfer and image matching tasks in [Rab+12b; Bon+15], and can provide plausible pointwise correspondences in practice, although theoretical guarantees remain partial [TFD24; CS25; LM25].

More recently, two alternative strategies have been introduced to build transport plans grounded in Sliced Wasserstein distances. The first one, called Sliced Wasserstein Generalised Geodesics (SWGG) [Mah+23; CTV25], defines a map between two discrete distributions $\delta_X = \frac{1}{n} \sum_i \delta_{x_i}$ and $\delta_Y = \frac{1}{n} \sum_i \delta_{y_i}$ as $\tau_\theta \circ \sigma_\theta^{-1}$, where $\sigma_\theta$ is a permutation which sorts $(\theta^\top x_i)_{i=1}^n$ and $\tau_\theta$ a permutation sorting $(\theta^\top y_i)_{i=1}^n$. The Sliced Wasserstein Generalised Geodesic distance ([Mah+23], Equation 8) is then defined as: (see also Fig. 24)

$$\mathrm{SWGG}_2^2(\mu_1, \mu_2, \theta) := \frac{1}{n} \sum_{i=1}^n \|x_{\sigma_\theta(i)} - y_{\tau_\theta(i)}\|_2^2. \tag{1}$$

The second one, called Expected Sliced Transport Plans, was introduced in [Liu+24] (inspired by [Row+19]), also for discrete measures. It aims to construct couplings by averaging the 1D optimal transport plans obtained from projections. Given $\varpi$ a probability measure on the hypersphere, with the same notations as above, the Expected Sliced Transport distance is defined as:

$$\mathbb{E}_{\theta \sim \varpi} \left[ \frac{1}{n} \sum_{i=1}^n \|x_{\sigma_\theta(i)} - y_{\tau_\theta(i)}\|_2^2 \right] \tag{2}$$

and the average transport plan as $\mathbb{E}_{\theta \sim \varpi}[\tau_\theta \circ \sigma_\theta^{-1}]$. This yields a plan between the two $d$-dimensional measures that reflects the averaged behaviour along slices.

These approaches provide practical and interpretable ways to define approximate transport maps. However, they are currently defined only for discrete measures and lack a rigorous theoretical grounding in more general measure spaces. Moreover, even in the discrete setting, it can easily be shown that the RHS quantity in Eq. (1) depends on the choice of the permutations, rendering the quantity ill-defined, as showcased in Appendix A.1.

The goal of this paper is to rigorously define and analyse these different Sliced Optimal Transport Plans for completely generic probability measures. We introduce the Pivot Sliced Discrepancy $\mathrm{PS}_\theta$, a discrepancy measure based on the $\nu$-based generalised geodesics [NP23], and generalising the Sliced Wasserstein Generalised Geodesic distance [Mah+23]. In doing so, we also provide new theoretical insights on the $\nu$-based Wasserstein distance [NP23]. We prove that $\mathrm{PS}_\theta$ is well-defined, symmetric and separates points. We then establish an equivalence between $\mathrm{PS}_\theta$ and a constrained version of the Wasserstein distance, showing that $\mathrm{PS}_\theta$ coincides with the minimal transport cost among plans that preserve the projected coupling. For empirical measures, we provide Monge and Kantorovich formulations of $\mathrm{PS}_\theta$, proving a constrained version of the Birkhoff-von Neumann theorem [Bir46]. Additionally, we study the Min-Pivot Sliced Discrepancy, a variant that matches the true Wasserstein distance for discrete measures when the space dimension is large enough with respect to the number of points. We then study the Expected Sliced Wasserstein Plan [Liu+24], which averages 1D sliced transport plans to obtain high-dimensional (non sparse) couplings. This theoretical study is followed by numerical experiments, illustrating the behaviour of the proposed transport plans on synthetic datasets and shape registration tasks.

The paper is organised as follows. In Section 2, we recall the necessary background on $\nu$-based Wasserstein geodesics, along with some new theoretical results that will serve as building blocks for the rest of the work. Section 3 presents and analyses the Pivot Sliced Discrepancy. In Section 4, we establish a precise connection between $\mathrm{PS}_\theta$ and a constrained Wasserstein discrepancy, showing that both quantities coincide. This correspondence is further developed in Section 5, where we explore Monge and Kantorovich formulations of $\mathrm{PS}_\theta$ for discrete measures. We then study in Section 6 the

---

[1]For a point cloud $X = (x_i)_{i=1}^N$, we write $\delta_X = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$

Min-Pivot Sliced Discrepancy, and show that it recovers the exact Wasserstein distance in certain discrete settings. Section 7 introduces and analyses the concept of Expected Sliced Wasserstein Plans. Finally, Section 8 is dedicated to numerical experiments.

## 2  Reminders and New Results on the $\nu$-based Wasserstein Distance

In this section, we lay some pre-requisites for the objects at play in the paper. We begin by recalling the concept of generalised geodesics in Section 2.1, which allows us to introduce the $\nu$-based Wasserstein distance in Section 2.2. This (semi-)metric was first defined in [AGS05; NP23], and we will sometimes also refer to it as "Pivot Wasserstein", and prove new technical properties that will be useful later. Later in this work, we will consider the Pivot Wasserstein distance using a "Wasserstein Mean" pivot, and to this end we propose some reminders on Wasserstein means in Section 2.3. Finally, in Section 2.4, we revisit a disintegration formulation of the $\nu$-based Wasserstein distance (first proved in [NP23]), which will sometimes be convenient for computations.

### 2.1  Wasserstein Geodesics and Generalised Geodesics

Given two measures $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, we denote by $\Pi^*(\mu_1, \mu_2)$ the set of Optimal Transport plans between $\mu_1$ and $\mu_2$ for the cost $\|x - y\|_2^2$. Using such plans, we can define a notion of shortest path (i.e. geodesic) between $\mu_1$ and $\mu_2$ in the space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$.

> **Definition 1.** A constant-speed geodesic between $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ is a curve $[0,1] \longrightarrow \mathcal{P}_2(\mathbb{R}^d)$ constructed using an optimal transport plan $\gamma \in \Pi^*(\mu_1, \mu_2)$ as follows:
>
> $$\mu_\gamma^{1\to2}(t) := ((1-t)P_1 + tP_2)\,\#\gamma, \tag{3}$$

where $P_1 : (x, y) \longmapsto x$ and $P_2 : (x, y) \longmapsto y$ are the marginal projection operators. Not only is $\mu_\gamma^{1\to2}$ a geodesic for the $W_2$ metric, but all (constant-speed) geodesics between $\mu_1$ and $\mu_2$ are of the form $\mu_\gamma^{1\to2}$ for a suitable $\gamma \in \Pi^*(\mu_1, \mu_2)$ (this is [AGS05], Theorem 7.2.2).

If the chosen optimal transport plan $\gamma$ is induced by a transport map $T$ (which is to say that $\gamma = (I, T)\#\mu_1$), then the geodesic takes the intuitive "displacement" formulation:

$$\mu_\gamma^{1\to2}(t) := ((1-t)I + tT)\,\#\mu_1, \tag{4}$$

with $I$ denoting the identity map of $\mathbb{R}^d$.

A remarkable property of the 2-Wasserstein space is that it is a Positively Curved (according to Alexandrov's metric definition of curvature) space, as proved in [AGS05] Theorem 7.3.2, Equation 7.3.12: for $\mu_1, \mu_2, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $\gamma \in \Pi^*(\mu_1, \mu_2)$ and $t \in [0,1]$, we have

$$W_2^2(\mu_\gamma^{1\to2}(t), \nu) \geq (1-t)W_2^2(\mu_1, \nu) + tW_2^2(\mu_2, \nu) - (1-t)tW_2^2(\mu_1, \mu_2). \tag{5}$$

For $t := \frac{1}{2}$, this can be re-written as

$$W_2^2(\mu_1, \mu_2) \geq 2W_2^2(\mu_1, \nu) + 2W_2^2(\mu_2, \nu) - 4W_2^2(\mu_\gamma^{1\to2}(t), \nu). \tag{6}$$

Unfortunately, the squared distance $W_2^2$ is not $\lambda$-convex along these Wasserstein geodesics ([AGS05], Example 9.1.5), which motivated [AGS05] to introduce other curves, coined "generalised geodesics", that satisfy this desirable property. First, we consider two optimal plans $\gamma_1 \in \Pi^*(\nu, \mu_1)$ and $\gamma_2 \in \Pi^*(\nu, \mu_2)$. To introduce the notion of generalised geodesics, we will require a 3-plan $\rho \in \Pi(\nu, \mu_1, \mu_2) \in \mathcal{P}_2(\mathbb{R}^{3d})$ (i.e. with marginals $\rho_0 = \mu, \rho_1 = \mu_1, \rho_2 = \mu_2$), such that its bi-marginals coincide with the plans $\gamma_1$ and $\gamma_2$: we require $\rho_{0,1} := P_{0,1}\#\rho = \gamma_1$ and $\rho_{0,2} := P_{0,2}\#\rho = \gamma_2$, where $P_{0,i} := (y, x_1, x_2) \longmapsto (y, x_i)$. We introduce the following notation for such 3-plans:

$$\Gamma(\nu, \mu_1, \mu_2) := \left\{ \rho \in \mathcal{P}_2(\mathbb{R}^{3d}) : \rho_{0,1} \in \Pi^*(\nu, \mu_1) \text{ and } \rho_{0,2} \in \Pi^*(\nu, \mu_2) \right\}. \tag{7}$$

**Definition 2.** A generalised geodesic based on $\nu$ between $\mu_1$ and $\mu_2$ is then defined as ([AGS05], Definition 9.2.2), given a $\rho \in \Gamma(\nu, \mu_1, \mu_2)$:

$$\mu_\rho^{1\to2}(t) := ((1-t)P_1 + tP_2)\#\rho. \tag{8}$$

Note that this curve depends on the choice of the 3-plan $\rho$, which itself depends on the optimal plans $\gamma_1$ and $\gamma_2$. The existence of such a $\rho$ can be shown using the gluing lemma (as presented in [San15], Lemma 5.5, for example). As desired, the curvature induced by these curves makes $W_2^2$ convex along these geodesics (in a certain sense, see [AGS05] Definition 9.2.4), namely we have the following inequality ([AGS05], Equation 9.2.7c), which is reversed compared to Eq. (5):

$$W_2^2(\mu_\rho^{1\to2}(t), \nu) \le (1-t)W_2^2(\mu_1, \nu) + tW_2^2(\mu_2, \nu) - (1-t)tW_2^2(\mu_1, \mu_2). \tag{9}$$

Like before, setting $t := \frac{1}{2}$ yields the following inequality:

$$W_2^2(\mu_1, \mu_2) \le 2W_2^2(\mu_1, \nu) + 2W_2^2(\mu_2, \nu) - 4W_2^2(\mu_\rho^{1\to2}(t), \nu). \tag{10}$$

If the optimal transport plans $\gamma_1$ and $\gamma_2$ are induced respectively by transport maps $T_1$ and $T_2$, then the choice of $\rho$ is unique, with $\rho = (I, T_1, T_2)\#\nu$ ([AGS05], Remark 9.2.3, see also Lemma 5.3.2 for a formal proof). This yields the following expression of the generalised geodesic, which is substantially more intuitive:

$$\mu_\rho^{1\to2}(t) = ((1-t)T_1 + tT_2)\#\nu. \tag{11}$$

## 2.2   The $\nu$-based Wasserstein Distance

A closely related concept is the $\nu$-based Wasserstein (semi)-distance, introduced by Nenna and Pass in [NP23]. This time we use a pivot measure $\nu$ to introduce a variant of the Wasserstein distance, yielding the following definition by [NP23] (Definition 3)[2]:

**Definition 3.** For $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, the $\nu$-based Wasserstein (semi)-metric between $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ is defined as:

$$W_\nu^2(\mu_1, \mu_2) := \min_{\rho \in \Gamma(\nu, \mu_1, \mu_2)} \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2). \tag{12}$$

We illustrate the $\nu$-based Wasserstein distance on a simple example in Fig. 1.

The question of whether the infimum defining $W_\nu$ is attained was not addressed by [NP23], we show that it is indeed the case in Proposition 1, using a technical property of the 3-plan set $\Gamma$ defined in Eq. (7). We remind that by Prokhorov's theorem, a subset of $\mathcal{P}_2(\mathbb{R}^d)$ is tight set if and only if it is pre-compact, which means that any sequence of measures in the set has a weakly converging subsequence.

**Lemma 1.**   1. For tight sets $P, Q_1, Q_2 \subset \mathcal{P}_2(\mathbb{R}^d)$, the set

$$\Gamma(P, Q_1, Q_2) := \{\rho \in \Gamma(\nu, \mu_1, \mu_2) : (\nu, \mu_1, \mu_2) \in P \times Q_1 \times Q_2\}$$

   is tight in $\mathcal{P}_2(\mathbb{R}^{3d})$.
2. Consider sequences $\nu^{(n)}, \mu_1^{(n)}, \mu_2^{(n)} \in \mathcal{P}_2(\mathbb{R}^d)^\mathbb{N}$ respectively converging to $\nu, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ for the weak convergence of measures, and a sequence $(\rho_n) \in \mathcal{P}_2(\mathbb{R}^{3d})^\mathbb{N}$ such that $\forall n \in \mathbb{N}, \rho_n \in \Gamma(\nu^{(n)}, \mu_1^{(n)}, \mu_2^{(n)})$ with $\rho_n \xrightarrow[n\to+\infty]{w} \rho \in \mathcal{P}_2(\mathbb{R}^{3d})$. Then $\rho \in \Gamma(\nu, \mu_1, \mu_2)$.
3. For $\nu, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, the set $\Gamma(\nu, \mu_1, \mu_2)$ is compact in $\mathcal{P}_2(\mathbb{R}^{3d})$.

---

[2]Their definition seems to have a typo, with $\Pi^*(\mu_i, \nu)$ instead of $\Pi^*(\nu, \mu_i)$. Furthermore, they work with measures supported on a bounded and convex domain of $\mathbb{R}^d$, but as they remark (footnote 4), and given [AGS05], Chapter 9, generalisation to measures on $\mathbb{R}^d$ with a finite moment of order 2 is perfectly natural.
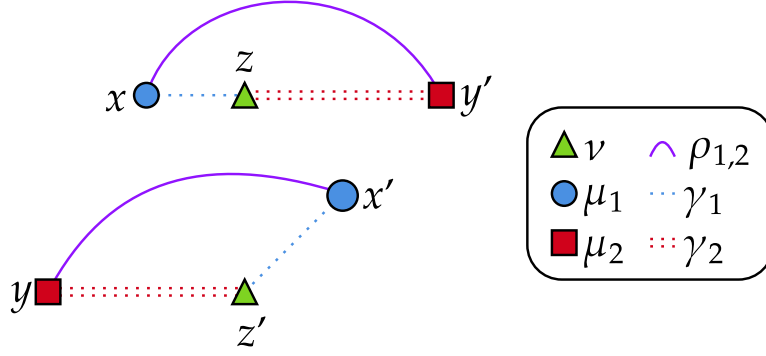
Figure 1: Example of the couplings behind $W_\nu(\mu_1, \mu_2)$ for discrete measures on $\mathbb{R}^2$. The measure $\nu$ is drawn with green triangles, $\mu_1$ with blue circles and $\mu_2$ with red squares. The (unique) OT plan $\gamma_1$ between $\nu$ and $\mu_1$ is drawn with dotted blue lines, the (also unique) OT plan $\gamma_2$ between $\nu$ and $\mu_2$ with red double dotted lines. The plans induce a unique valid 3-plan $\rho \in \Gamma(\nu, \mu_1, \mu_2)$, we represent the coupling $\rho_{1,2}$ between $\mu_1$ and $\mu_2$ with curved purple lines. Notice that the coupling $\rho_{1,2}$ differs from the (unique) OT coupling between $\mu_1$ and $\mu_2$.

*Proof.* For 1. we set $\varepsilon > 0$. By tightness of $P, Q_1, Q_2$ and Prokhorov's theorem, there exists a compact set $\mathcal{K} \subset \mathbb{R}^d$ such that for any $\mu \in P \cup Q_1 \cup Q_2$, $\mu(\mathbb{R}^d \setminus \mathcal{K}) < \varepsilon/3$. It follows that for any $\rho \in \Gamma(P, Q_1, Q_2)$,

$$
\begin{aligned}
\rho(\mathbb{R}^{3d} \setminus \mathcal{K}^3) &\leq \rho\left((\mathbb{R}^d \setminus \mathcal{K}) \times \mathbb{R}^d \times \mathbb{R}^d\right) + \rho\left(\mathbb{R}^d \times (\mathbb{R}^d \setminus \mathcal{K}) \times \mathbb{R}^d\right) + \rho\left(\mathbb{R}^d \times \mathbb{R}^d \times (\mathbb{R}^d \setminus \mathcal{K})\right) \\
&= \nu(\mathbb{R}^d \setminus \mathcal{K}) + \mu_1(\mathbb{R}^d \setminus \mathcal{K}) + \mu_2(\mathbb{R}^d \setminus \mathcal{K}) \\
&< \varepsilon,
\end{aligned}
$$

and thus $\Gamma(P, Q_1, Q_2)$ is tight.

For 2. we observe that for $i \in \{1, 2\}$, $[\rho_n]_{0,i} \in \Pi^*(\nu^{(n)}, \mu_i^{(n)})$. Given that $\nu^{(n)} \xrightarrow[n \to +\infty]{w} \nu$ and $\mu_i^{(n)} \xrightarrow[n \to +\infty]{w} \mu_i$, and that $[\rho_n]_{0,i} \xrightarrow[n \to +\infty]{w} \rho_{0,i}$, [Vil09] Theorem 5.20 shows that $\rho_{0,i} \in \Pi^*(\nu, \mu_i)$ (the result provides the existence of a subsequence converging to an element of $\Pi^*(\nu, \mu_i)$, then uniqueness of the limit shows $\rho_{0,i} \in \Pi^*(\nu, \mu_i)$), and we conclude that $\rho \in \Gamma(\nu, \mu_1, \mu_2)$ by definition.

For 3., take $(\rho_n) \in \Gamma(\nu, \mu_1, \mu_2)^{\mathbb{N}}$. By 1) and tightness of $\{\nu\}, \{\mu_1\}, \{\mu_2\}$, there exists an extraction $\alpha$ such that $\rho_{\alpha(n)} \xrightarrow[n \to +\infty]{w} \rho \in \mathcal{P}_2(\mathbb{R}^{3d})$, then we show that $\rho \in \Gamma(\nu, \mu_1, \mu_2)$ using 2) with $\forall n \in \mathbb{N}$, $\nu^{(n)} := \nu$, $\mu_i^{(n)} := \mu_i$ for $i \in \{1, 2\}$. $\qquad\square$

---

**Proposition 1.** For $\nu, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, it holds

$$
\inf_{\rho \in \Gamma(\nu, \mu_1, \mu_2)} \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2) = \min_{\rho \in \Gamma(\nu, \mu_1, \mu_2)} \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2)
$$

---

*Proof.* By Lemma 1 item 3), $\Gamma(\nu, \mu_1, \mu_2)$ is a compact subset of $\mathcal{P}_2(\mathbb{R}^{3d})$. Then the map $J : \rho \in \mathcal{P}_2(\mathbb{R}^{3d}) \longmapsto \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2)$ is lower semi-continuous with respect to the weak convergence of measures ([San15] Lemma 1.6), hence the infimum is attained. $\qquad\square$

Another consequence of Lemma 1 is that the $\nu$-based Wasserstein distance is lower semi-continuous with respect to the weak convergence of measures, which is a property that was not studied in [NP23].

**Proposition 2.** The map $(\nu, \mu_1, \mu_2) \in \mathcal{P}_2(\mathbb{R}^d)^3 \longmapsto W_\nu(\mu_1, \mu_2)$ is lower semi-continuous with respect to the weak convergence of measures: for any $\nu^{(n)} \xrightarrow[n \longrightarrow +\infty]{w} \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $\mu_i^{(n)} \xrightarrow[n \longrightarrow +\infty]{w}$ $\mu_i \in \mathcal{P}_2(\mathbb{R}^d)$, $i \in \{1, 2\}$, we have:

$$W_\nu(\mu_1, \mu_2) \leq \liminf_{n \longrightarrow +\infty} W_{\nu^{(n)}}(\mu_1^{(n)}, \mu_2^{(n)}). \tag{13}$$

*Proof.* Without loss of generality, we can assume that $W_{\nu^{(n)}}(\mu_1^{(n)}, \mu_2^{(n)}) \xrightarrow[n \longrightarrow +\infty]{} \liminf_{n \longrightarrow +\infty} W_{\nu^{(n)}}(\mu_1^{(n)}, \mu_2^{(n)})$ (up to considering an extraction of all sequences). For $n \in \mathbb{N}$, we can choose $\rho_n \in \Gamma(\nu^{(n)}, \mu_1^{(n)}, \mu_2^{(n)})$ optimal by Proposition 1. By Lemma 1 item 1) and tightness of the sets $\{\nu^{(n)}\}, \{\mu_1^{(n)}\}, \{\mu_2^{(n)}\}$, there exists an extraction $\alpha$ such that $\rho_{\alpha(n)} \xrightarrow[n \longrightarrow +\infty]{w} \rho \in \mathcal{P}_2(\mathbb{R}^{3d})$. and by Lemma 1 item 2) we have $\rho \in \Gamma(\nu, \mu_1, \mu_2)$. By lower semi-continuity of the map $J : \rho \in \mathcal{P}_2(\mathbb{R}^{3d}) \longmapsto \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 d\rho(y, x_1, x_2)$ ([San15] Lemma 1.6), we have:

$$W_\nu^2(\mu_1, \mu_2) \leq J(\rho) \leq \liminf_{n \longrightarrow +\infty} J(\rho_{\alpha(n)}) = \liminf_{n \longrightarrow +\infty} W_{\nu^{(\alpha(n))}}^2(\mu_1^{(\alpha(n))}, \mu_2^{(\alpha(n))}) = \liminf_{n \longrightarrow +\infty} W_{\nu^{(n)}}^2(\mu_1^{(n)}, \mu_2^{(n)}),$$

where the first inequality follows from the definition of $W_\nu$, since $\rho \in \Gamma(\nu, \mu_1, \mu_2)$ is admissible, and the second inequality follows from the lower semi-continuity of $J$. The first equality is due to the optimality of $\rho_{\alpha(n)}$, and the second equality follows from our reduction to the case where $W_{\nu^{(n)}}(\mu_1^{(n)}, \mu_2^{(n)}) \xrightarrow[n \longrightarrow +\infty]{} \liminf_{n \longrightarrow +\infty} W_{\nu^{(n)}}(\mu_1^{(n)}, \mu_2^{(n)})$. $\qquad \square$

Full continuity with respect to the weak convergence of measures is not guaranteed, as shown in Example 1.

**Example 1** ($W_\nu(\cdot, \mu_2)$ is not continuous)**.** Consider the following empirical measures in $\mathbb{R}^2$:

$$\nu := \tfrac{1}{2}(\delta_z + \delta_{z'}), \ z := (0, 1), \ z' := (0, -1);$$
$$\mu_1^{(n)} := \tfrac{1}{2}(\delta_{x_n} + \delta_{x'}), \ x_n := (-1, 2^{-n}), \ x' := (1, 0);$$
$$\mu_2 = \tfrac{1}{2}(\delta_y + \delta_{y'}), \ y := (-2, -1), \ y := (2, 1).$$

For each $n \in \mathbb{N}$, we have $\Pi^*(\nu, \mu_1^{(n)}) = \{\gamma_1^{(n)}\}$ with $\gamma_1^{(n)} := \tfrac{1}{2}(\delta_{(z, x_n)} + \delta_{(z', x')})$. We also have $\Pi^*(\nu, \mu_2) = \{\gamma_2\}$ with $\gamma_2 := \tfrac{1}{2}(\delta_{(z, y')} + \delta_{(z', y)})$. This shows that $\Gamma(\nu, \mu_1^{(n)}, \mu_2) = \{\rho_n\}$ where $\rho_n := \tfrac{1}{2}(\delta_{(z, x_n, y')} + \delta_{(z', x', y)})$, yielding the cost

$$W_\nu^2(\mu_1^{(n)}, \mu_2) = \tfrac{1}{2}\|x_n - y'\|_2^2 + \tfrac{1}{2}\|x' - y\|_2^2 = \tfrac{1}{2}\left(3^2 + (1 - 2^{-n})^2\right) + \tfrac{1}{2}\left(3^2 + 1^2\right) \xrightarrow[n \longrightarrow +\infty]{} 10.$$

However, we have $\mu_1^{(n)} \xrightarrow[n \longrightarrow +\infty]{w} \mu_1 = \tfrac{1}{2}(\delta_x + \delta_{x'})$ with $x := (-1, 0)$. We see that $\Pi^*(\nu, \mu_1) = \Pi(\nu, \mu_2)$, and clearly the choice $\gamma_1 := \tfrac{1}{2}(\delta_{(z, x')} + \delta_{(z', x)})$ will be optimal, such that $\rho := \tfrac{1}{2}(\delta_{(z, x', y')} + \delta_{(z', x, y)})$ is optimal for $W_\nu^2(\mu_1, \mu_2) = 2 < \lim_{n \longrightarrow +\infty} W_\nu^2(\mu_1^{(n)}, \mu_2) = 10$. We illustrate the setting of this example in Fig. 2.

As remarked earlier (again, [AGS05], Remark 9.2.3), if each $\Pi^*(\nu, \mu_i)$ is reduced to a single plan $\gamma_i$ induced by $T_i$ (for $i \in \{1, 2\}$), then the only element $\rho \in \Gamma(\nu, \mu_1, \mu_2)$ is $\rho = (I, T_1, T_2)\#\nu$, yielding the following formulation for the $\nu$-based Wasserstein distance (see also [NP23], Example 9 and the Linear OT framework [Wan+13]):

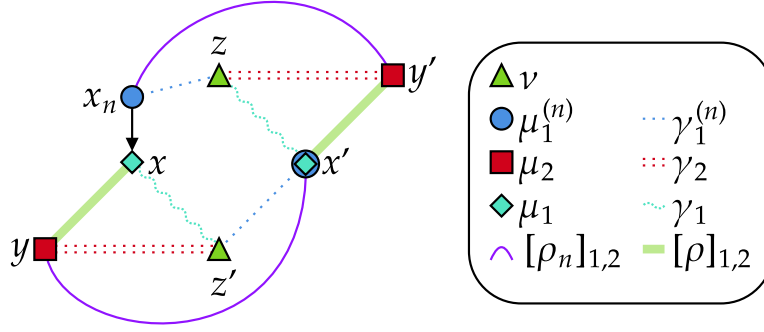$$W_\nu^2(\mu_1, \mu_2) = \int_{\mathbb{R}^d} \|T_1(y) - T_2(y)\|_2^2 d\nu(y). \tag{14}$$

Figure 2: Representation of Example 1. The measure $\nu$ is drawn with green triangles, $\mu_1^{(n)}$ with blue circles, $\mu_2$ with red squares, and the limit $\mu$ with light blue diamonds. The OT plan $\gamma_1^{(n)}$ between $\nu$ and $\mu_1^{(n)}$ is drawn with dotted blue lines, the OT plan $\gamma_2$ between $\nu$ and $\mu_2$ with red double dotted lines, and the induced plan $[\rho_n]_{1,2}$ between $\mu_1^{(n)}$ and $\mu_2$ with curved purple lines. As for the limit, an OT plan $\gamma_1$ between $\nu$ and $\mu_1$ is drawn with curved dashed light blue lines, and the induced plan $[\rho]_{1,2}$ between $\mu_1$ and $\mu_2$ using $\gamma_1$ and $\gamma_2$ is drawn with thick green lines.

A result of interest is [AGS05], Lemma 9.2.1 Equation 9.2.7b, which states that for any $\rho \in \Gamma(\nu, \mu_1, \mu_2)$ (see Eq. (7))

$$\mathrm{W}_2^2(\mu_1, \mu_\rho^{1 \to 2}(t)) = (1-t)\mathrm{W}_2^2(\mu_1, \nu) + t\mathrm{W}_2^2(\mu_2, \nu) - (1-t)t \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2). \qquad (15)$$

Taking in particular a 3-plan $\rho^* \in \Gamma(\nu, \mu_1, \mu_2)$ that is optimal for the $\nu$-based Wasserstein distance (Eq. (12)), we obtain

$$\mathrm{W}_2^2(\mu_1, \mu_{\rho^*}^{1 \to 2}(t)) = (1-t)\mathrm{W}_2^2(\mu_1, \nu) + t\mathrm{W}_2^2(\mu_2, \nu) - (1-t)t\mathrm{W}_\nu^2(\mu_1, \mu_2). \qquad (16)$$

## 2.3   Reminders on Wasserstein Means

A natural application of Wasserstein geodesics is the concept of Wasserstein means, which we will require in Section 3. The following result states that Wasserstein means are exactly the middles of Wasserstein geodesics. For the sake of completeness, we provide some reminders on geodesic middles in the Appendix Appendix A.2, wherein we recall and prove an analogous result for geodesic spaces.

**Proposition 3.** For $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, the set of Wasserstein Means between $\mu_1$ and $\mu_2$

$$\mathrm{M}(\mu_1, \mu_2) := \underset{\mu \in \mathcal{P}_2(\mathbb{R}^d)}{\mathrm{argmin}} \ \mathrm{W}_2^2(\mu_1, \mu) + \mathrm{W}_2^2(\mu, \mu_2) \qquad (17)$$

can be expressed using Wasserstein geodesics Eq. (3):

$$\mathrm{M}(\mu_1, \mu_2) = \left\{ \mu_\gamma^{1 \to 2}(\tfrac{1}{2}) : \gamma \in \Pi^*(\mu_1, \mu_2) \right\} = \left\{ \left( \tfrac{1}{2}P_1 + \tfrac{1}{2}P_2 \right) \# \gamma : \gamma \in \Pi^*(\mu_1, \mu_2) \right\}. \qquad (18)$$

*Proof.* The result is an application of Lemma 12 in the geodesic space $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$.   $\square$

## 2.4   Another Formulation of $\mathrm{W}_\nu$ with Measure Disintegration

In this work, we will need a convenient formulation of the $\nu$-based Wasserstein distance which uses the notion of disintegration of measures. We recall this notion in Appendix A.3, and provide a proof in Appendix A.4 of the Theorem by Nenna and Pass ([NP23], Theorem 12 item 1), adapted

to measures in $\mathcal{P}_2(\mathbb{R}^d)$. In Example 2, we illustrate the result on a simple example with discrete measures.

> **Example 2.** We consider measures $\nu, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ as in Fig. 3. We consider two optimal plans $\gamma_1 \in \Pi^*(\nu, \mu_1)$ and $\gamma_2 \in \Pi^*(\nu, \mu_2)$, represented in Fig. 3. Writing the disintegrations as $\gamma_i(\mathrm{d}y, \mathrm{d}x_i) = \nu(\mathrm{d}y)\gamma_i^y(\mathrm{d}x_i)$, we can apply Theorem 1 to compute $\mathrm{W}_\nu^2(\mu_1, \mu_2)$:
>
> $$\mathrm{W}_\nu^2(\mu_1, \mu_2) = \tfrac{1}{2}\mathrm{W}_2^2\left(\gamma_1^{z_1}, \gamma_2^{z_1}\right) + \tfrac{1}{2}\mathrm{W}_2^2\left(\gamma_1^{z_2}, \gamma_2^{z_2}\right)$$
> $$= \tfrac{1}{2}\mathrm{W}_2^2\left(\tfrac{1}{2}\delta_{x_1} + \tfrac{1}{2}\delta_{x_2}, \tfrac{1}{2}\delta_{y_1} + \tfrac{1}{2}\delta_{y_2}\right) + \tfrac{1}{2}\mathrm{W}_2^2\left(\tfrac{1}{2}\delta_{x_2} + \tfrac{1}{2}\delta_{x_3}, \tfrac{1}{2}\delta_{y_2} + \tfrac{1}{2}\delta_{y_3}\right).$$
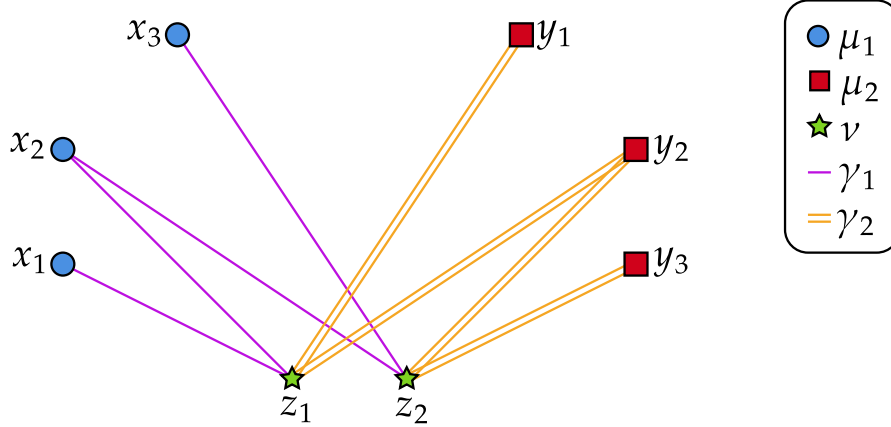


Figure 3: In this example, there is a unique optimal transport plan $\gamma_1$ (purple lines) between $\mu_1$ (blue circles) and the pivot $\nu$ (green stars), and likewise for $\gamma_2$ (orange double lines) between $\mu_2$ (red squares) and $\nu$. The disintegration kernel $\gamma_1^{z_1}$ in the disintegration $\gamma_1(\mathrm{d}z, \mathrm{d}x) = \nu(\mathrm{d}z)\gamma_1^z(\mathrm{d}x)$ is the probability measure $\gamma_1^{z_1} = \tfrac{1}{2}\delta_{x_1} + \tfrac{1}{2}\delta_{x_2}$, and likewise for $\gamma_1^{z_2}, \gamma_2^{z_1}, \gamma_2^{z_2}$.

> **Theorem 1** ([NP23] Theorem 12 item 1). Let $\nu, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$. The following equality holds:
>
> $$\mathrm{W}_\nu^2(\mu_1, \mu_2) = \min_{\gamma_i \in \Pi^*(\nu, \mu_i),\, i \in \{1,2\}} \int_{\mathbb{R}^d} \mathrm{W}_2^2(\gamma_1^y, \gamma_2^y)\mathrm{d}\nu(y), \tag{19}$$
>
> where for $i \in \{1, 2\}$, $\gamma_i^y \in \mathcal{P}_2(\mathbb{R}^d)$ is defined using the disintegration $\gamma_i(\mathrm{d}y, \mathrm{d}x) = \nu(\mathrm{d}y)\gamma_i^y(\mathrm{d}x)$.

*Proof.* We provide a proof in Appendix A.4, which generalises that in [NP23] to measures in $\mathcal{P}_2(\mathbb{R}^d)$, following similar ideas. $\qquad\square$

## 3  The Pivot Sliced Discrepancy

### 3.1  Definition with the $\nu$-based Wasserstein Distance

We introduce a generalised version of SWGG introduced in [Mah+23] for general measures in $\mathcal{P}_2(\mathbb{R}^d)$ (and fixing the ambiguity issues that will be discussed in Example 8), using the $\nu$-based Wasserstein distance (Eq. (12), and see [NP23]), where the base measure $\nu$ is taken as a middle of projected versions of the measures:

> **Definition 4.** Let $\mu_1, \mu_2$ of $\mathcal{P}_2(\mathbb{R}^d)$, take $\mu_\theta \in \mathrm{M}(Q_\theta \# \mu_1, Q_\theta \# \mu_2)$, where $Q_\theta : x \longmapsto (\theta^\top x)\theta$. Then, we define
>
> $$\mathrm{PS}_\theta(\mu_1, \mu_2) := \mathrm{W}_{\mu_\theta}(\mu_1, \mu_2). \tag{20}$$

We remind that we consider to related projection operations: $P_\theta : x \longmapsto \theta^\top x$ and $Q_\theta : x \longmapsto (\theta^\top x)\theta$. The first one is valued in $\mathbb{R}$, while the second is valued in $\mathbb{R}\theta \subset \mathbb{R}^d$. To fix ideas, we illustrate the definition of $\mathrm{PS}_\theta$ in the case of discrete measures without projection ambiguity in Fig. 4.
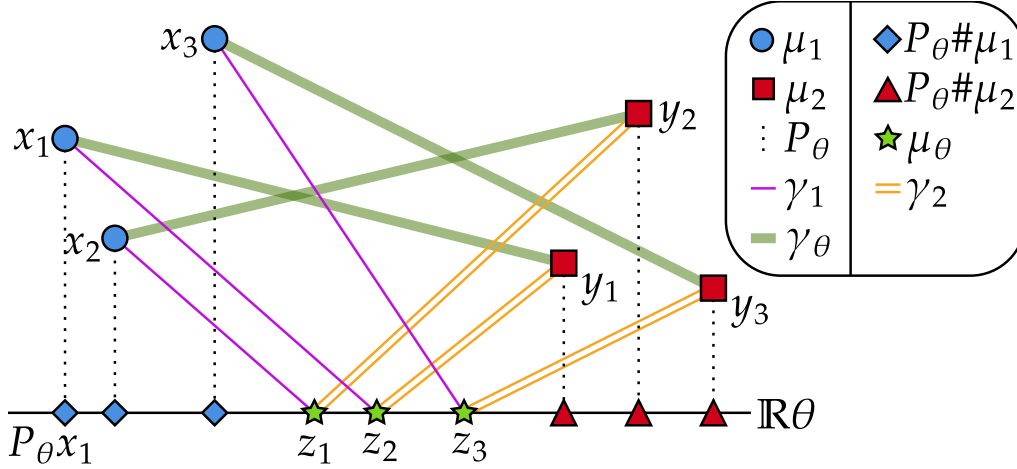


Figure 4: Illustration of the definition of $\mathrm{PS}_\theta$ in the case of discrete measures without projection ambiguity. The measure $\mu_1$ is represented by blue circles, and $\mu_2$ by red squares. The projected measures $Q_\theta\#\mu_1$ and $Q_\theta\#\mu_2$ are represented by blue diamonds and red triangles respectively. The middle $\mu_\theta$ of the projections is represented by green stars. Once this middle is determined, we compute optimal transport plans $\gamma_1, \gamma_2$ between $\mu_\theta$ and $\mu_1, \mu_2$ respectively (in this case, they are unique). We represent $\gamma_1$ by purple lines and $\gamma_2$ by orange double lines. To obtain the coupling corresponding to the cost $\mathrm{PS}_\theta(\mu_1, \mu_2)$, we look at the targets of each point $(z_i)$ of the projected middle $\mu_\theta$: since $z_1$ is mapped to $x_1$ in $\mu_1$ and to $y_1$ in $\mu_2$, the coupling $\gamma_\theta$ maps $x_1$ to $x_2$, and so on. The coupling $\gamma_\theta$ is represented with thick green lines.

The idea of using a pivot measure is to find an optimal manner of correcting projection ambiguities. To illustrate this, we consider a simple pathological example in Fig. 5, where the projections of the points of the support of $\mu_1$ and $\mu_2$ are not distinct.
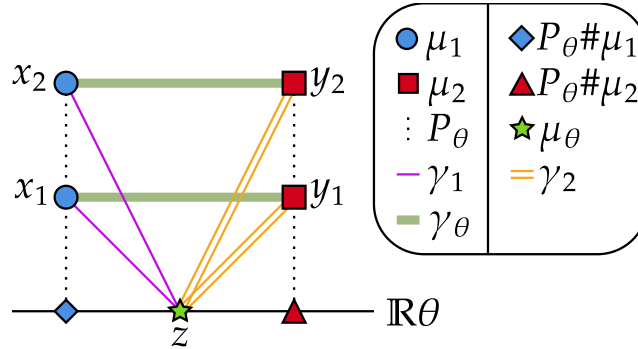


Figure 5: In this example, we notice that $P_\theta\#\mu_1$ and $P_\theta\#\mu_2$ are reduced to Dirac masses, thus their middle $\mu_\theta$ is the middle Dirac mass. The optimal couplings $\gamma_1$ and $\gamma_2$ between $\mu_\theta, \mu_1$ and $\mu_\theta, \mu_2$ are then unique. It is then easy to see that the optimal $\rho \in \Gamma(\mu_\theta, \mu_1, \mu_2)$ is such that $\rho_{1,2} =: \gamma_\theta$ is the OT coupling between $\mu_1$ and $\mu_2$. In this example, $\mathrm{PS}_\theta(\mu_1, \mu_2) = \mathrm{W}_2(\mu_1, \mu_2)$.

In Fig. 6, we illustrate another simple example where the projections of the points of the support of $\mu_1$ are not distinct, but where they are distinct for $\mu_2$.
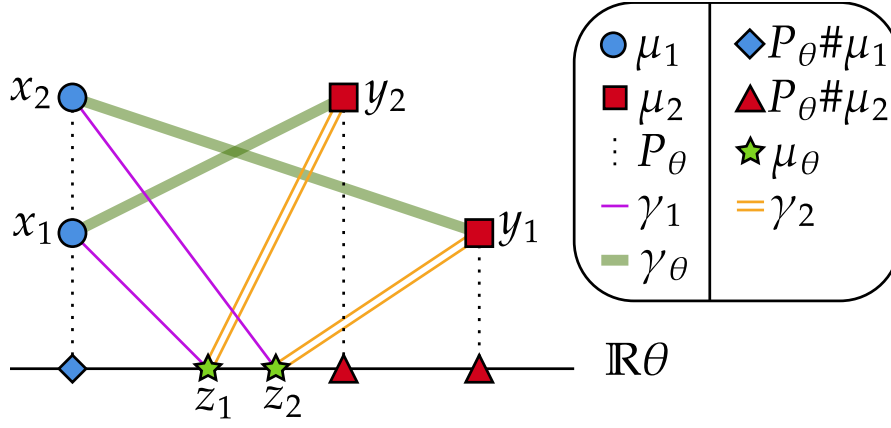
Figure 6: In this illustration, $P_\theta\#\mu_1$ is a Dirac mass but not $P_\theta\#\mu_2$. Since we compare the middle $\mu_\theta$ with $\mu_1$ and not $P_\theta\#\mu_1$, there is in this case a unique optimal plan $\gamma_1$ between $\mu_\theta$ and $\mu_1$. The optimal plan $\gamma_2$ between $\mu_\theta$ and $\mu_2$ is also unique. The constraint $\rho_{0,1} = \gamma_1$, $\rho_{0,2} = \gamma_2$ imposes that $\rho_{1,2} = \frac{1}{2}\delta_{x_1\otimes y_2} + \frac{1}{2}\delta_{x_2\otimes y_1}$ for any $\rho \in \Gamma(\mu_\theta, \mu_1, \mu_2)$, hence there is no choice in the optimisation over $\rho$.

> **Remark 1.** As remarked by [NP23] in Proposition 16, when $\nu$ is absolutely continuous with respect to the one-dimensional Hausdorff on a line, then the $\nu$-based Wasserstein distance equates the *layer-wise Wasserstein metric* introduced by [KPS20]. We will see in Section 4 that $\mathrm{PS}_\theta$ equals another discrepancy that we call $\mathrm{CW}_\theta$, and this equality allows us to show that $\mathrm{PS}_\theta$ satisfies the triangle inequality (and thus is a distance) on the set of measures with atomless projections, which is a stronger result than assuming absolute continuity of the pivot.

Note that this is a generalisation of SWGG introduced [Mah+23] in the sense that they show that their definition of SWGG coincides with the expression of Eq. (20) in Proposition 4.2. To prove that the quantity $\mathrm{PS}_\theta^2$ is well-defined, which is to say that it does not depend on the choice of $\mu_\theta \in \mathrm{M}(Q_\theta\#\mu_1, Q_\theta\#\mu_2)$, we will show that in fact $\mathrm{M}(Q_\theta\#\mu_1, Q_\theta\#\mu_2)$ has only one element.

> **Lemma 2.** Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, and $\theta \in \mathbb{S}^{d-1}$. Then
>
> $$\mathrm{M}(Q_\theta\#\mu_1, Q_\theta\#\mu_2) = \{\mu_\theta[\mu_1, \mu_2]\}, \quad \mu_\theta[\mu_1, \mu_2] := \left[\left(\tfrac{1}{2}F_{\nu_1}^{[-1]} + \tfrac{1}{2}F_{\nu_2}^{[-1]}\right)\theta\right]\#\mathscr{L}_{[0,1]}, \quad (21)$$
>
> where for $i = 1, 2$, the measure $\nu_i$ is defined as $\nu_i = P_\theta\#\mu_i$, with $P_\theta : x \longmapsto \theta^\top x$, $\mathscr{L}_{[0,1]}$ the Lebesgue measure on $[0,1]$, and where $F_\nu^{[-1]}$ for $\nu \in \mathcal{P}(\mathbb{R})$ denotes the pseudo-inverse of its cumulative distribution:
>
> $$\forall t \in [0,1], \quad F_\nu^{[-1]}(t) := \inf\{s \in \mathbb{R} : \nu((-\infty, s]) \geq t\}. \quad (22)$$

*Proof.* First, since the $Q_\theta\#\mu_i, i \in \{1, 2\}$ are supported on $\mathbb{R}\theta$, we have

$$\mathrm{M}(Q_\theta\#\mu_1, Q_\theta\#\mu_2) = \{\theta\#\mu : \mu \in \mathrm{M}(P_\theta\#\mu_1, P_\theta\#\mu_2)\}, \quad (23)$$

which amounts to reducing a problem on a line of direction $\theta$ to a problem on $\mathbb{R}$, then embedding the result onto the line $\mathbb{R}\theta$. We introduce $\nu_i := P_\theta\#\mu_i$ for $i \in \{1, 2\}$ and leverage Proposition 3:

$$\mathrm{M}(\nu_1, \nu_2) = \left\{\left(\tfrac{1}{2}P_1 + \tfrac{1}{2}P_2\right)\#\gamma : \gamma \in \Pi^*(\nu_1, \nu_2)\right\}. \quad (24)$$

Since the $\nu_i$ are measures on $\mathbb{R}$, by [San15] Theorem 2.9, the set of optimal plans $\Pi^*(\nu_1, \nu_2)$ is reduced to the plan $(F_{\nu_1}^{[-1]}, F_{\nu_2}^{[-1]})\#\mathscr{L}_{[0,1]}$. Using Eq. (24) above and the projection embedding from Eq. (23), we obtain the result stated in Eq. (21). $\square$

**Remark 2.** Consider $\mu_1 = \frac{1}{n}\sum_i \delta_{x_i}$, $\mu_2 = \frac{1}{n}\sum_i \delta_{y_i}$, $\theta \in \mathbb{S}^{d-1}$ and $\sigma_\theta, \tau_\theta$ two permutations sorting respectively $(\theta^\top x_i)_i$ and $(\theta^\top y_i)_i$ (they may not be unique if the families $(\theta^\top x_i)_i$ and $(\theta^\top y_i)_i$ are not injective). Then the projected middle (computed using Eq. (21)) is explicit:

$$\mu_\theta[\mu_1, \mu_2] = \frac{1}{n}\sum_{i=1}^{n} \delta\left(\frac{\theta^\top\left(x_{\sigma_\theta(i)} + y_{\tau_\theta(i)}\right)}{2}\theta\right). \tag{25}$$

Note that measure above does not depend on the choice of the sorting permutations $(\sigma_\theta, \tau_\theta)$, since the families $(\theta^\top x_{\sigma_\theta(i)})_i$ and $(\theta^\top y_{\tau_\theta(i)})_i$ do not. This expression is specific to the case where $\mu_1$ and $\mu_2$ are uniform discrete measures with the same amount of atoms.

An interesting property of optimal transport between a measure $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and another measure $\nu$ supported on a line $\mathbb{R}\theta$ is that the set of optimal plans and the cost can be related to the one-dimensional projections of $\mu$ and $\nu$ onto the line $\mathbb{R}\theta$. We remind $Q_\theta : x \longmapsto (\theta^\top x)\theta$, and introduce $Q_{\theta\perp} := I - Q_\theta$. The following result is a generalisation of [Mah+23], Lemma 4.6, which was written in the case of uniform discrete measures. Note that the exponent 2 in the cost is paramount and allows the separation of orthogonal terms.

**Proposition 4.** Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\nu$ is supported on $\mathbb{R}\theta$, where $\theta \in \mathbb{S}^{d-1}$. Then for any plan $\gamma \in \Pi(\nu, \mu)$, we have

$$\int_{\mathbb{R}^{2d}} \|x - y\|_2^2 \mathrm{d}\gamma(y, x) = \int_{\mathbb{R}^{2d}} (\theta^\top(x - y))^2 \mathrm{d}\gamma(y, x) + \int_{\mathbb{R}^d} \|Q_{\theta\perp} x\|_2^2 \mathrm{d}\mu(x), \tag{26}$$

with the alternate expression

$$\int_{\mathbb{R}^{2d}} (\theta^\top(x - y))^2 \mathrm{d}\gamma(y, x) = \int_{\mathbb{R}^2} (s - t)^2 \mathrm{d}(P_\theta, P_\theta)\#\gamma(s, t). \tag{27}$$

This yields the following expression for the OT cost:

$$\mathrm{W}_2^2(\nu, \mu) = \mathrm{W}_2^2(P_\theta\#\nu, P_\theta\#\mu) + \int_{\mathbb{R}^d} \|Q_{\theta\perp} x\|_2^2 \mathrm{d}\mu(x), \tag{28}$$

and the following characterisation of the optimal plans:

$$\Pi^*(\nu, \mu) = \left\{\gamma \in \Pi(\nu, \mu) : (P_\theta, P_\theta)\#\gamma = \left(F_{P_\theta\#\nu}^{[-1]}, F_{P_\theta\#\mu}^{[-1]}\right)\#\mathscr{L}_{[0,1]}\right\}. \tag{29}$$

*Proof.* Let $\gamma \in \Pi(\nu, \mu)$. We have

$$\int_{\mathbb{R}^{2d}} \|x - y\|_2^2 \mathrm{d}\gamma(y, x) = \int_{\mathbb{R}^{2d}} \left(\|Q_\theta(x - y)\|_2^2 + \|Q_{\theta\perp}(x - y)\|_2^2\right) \mathrm{d}\gamma(y, x)$$

$$= \int_{\mathbb{R}^{2d}} (\theta^\top(x - y))^2 \mathrm{d}\gamma(y, x) + \int_{\mathbb{R}^d} \|Q_{\theta\perp} x\|_2^2 \mathrm{d}\mu(x), \tag{30}$$

where the last equality comes from the fact that $\nu$ is supported on $\mathbb{R}\theta$ and that the second marginal of $\gamma$ is $\mu$. Since the second term does not depend on $\gamma$, by taking the infimum in Eq. (26), we obtain Eq. (28), where the equality

$$\inf_{\gamma \in \Pi(\nu, \mu)} \int_{\mathbb{R}^{2d}} (\theta^\top(x - y))^2 \mathrm{d}\gamma(y, x) = \mathrm{W}_2^2(P_\theta\#\nu, P_\theta\#\mu)$$

is justified by [DLV24], Lemma 2. Furthermore, Eq. (26) shows that $(P_\theta, P_\theta)\#\Pi^*(\nu, \mu) \subset \Pi^*(P_\theta\#\nu, P_\theta\#\mu)$. Indeed, take $\gamma \in \Pi^*(\nu, \mu)$, then since $\pi_\theta := (P_\theta, P_\theta)\#\gamma \in \Pi(P_\theta\#\nu, P_\theta\#\mu)$, Eq. (26) yields the optimality of $\pi_\theta$ for the problem $\mathrm{W}_2^2(P_\theta\#\nu, P_\theta\#\mu)$. By [San15] Theorem 2.9,

$$\Pi^*(P_\theta\#\nu, P_\theta\#\mu) = \left\{\left(F_{P_\theta\#\nu}^{[-1]}, F_{P_\theta\#\mu}^{[-1]}\right)\#\mathscr{L}_{[0,1]}\right\} =: \{\pi_\theta^*\},$$

hence we have shown that $(P_\theta, P_\theta)\#\Pi^*(\nu, \mu) = \{\pi_\theta^*\}$. Conversely, take $\gamma \in \Pi(\nu, \mu)$ such that $(P_\theta, P_\theta)\#\gamma = \pi_\theta^*$, where $\pi_\theta^*$ is the unique element of $\Pi^*(P_\theta\#\nu, P_\theta\#\mu)$. Then by plugging $\gamma$ into Eq. (26), we obtain $\gamma \in \Pi^*(\nu, \mu)$ using Eq. (28). We conclude that

$$\Pi^*(\nu, \mu) = \{\gamma \in \Pi(\nu, \mu) : (P_\theta, P_\theta)\#\gamma = \pi_\theta^*\}. \qquad \square$$

## 3.2   Semi-Metric Properties of $\mathrm{PS}_\theta$

We begin by stating straightforward properties of the discrepancy $\mathrm{PS}_\theta$:

> **Proposition 5.** Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, and $\theta \in \mathbb{S}^{d-1}$. Then the following properties hold:
> - (Separation) $\mathrm{PS}_\theta(\mu_1, \mu_2) = 0$ if and only if $\mu_1 = \mu_2$.
> - (Symmetry) $\mathrm{PS}_\theta(\mu_1, \mu_2) = \mathrm{PS}_\theta(\mu_2, \mu_1)$.
> - (Upper-bound of $\mathrm{W}_2$) $\mathrm{PS}_\theta(\mu_1, \mu_2) \geq \mathrm{W}_2(\mu_1, \mu_2)$.

*Proof.* If $\mathrm{PS}_\theta(\mu_1, \mu_2) = 0$, then by Proposition 1, there exists $\rho \in \Gamma(\mu_\theta[\mu_1, \mu_2], \mu_1, \mu_2)$ such that $\int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2) = 0$, then in particular, for $\gamma := \rho_{1,2} \in \Pi(\mu_1, \mu_2)$, we have $\int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\gamma(x_1, x_2) = 0$ and thus $x_1 = x_2$ for $\gamma$-almost-every $(x_1, x_2) \in \mathbb{R}^{2d}$. For a test function $\phi \in \mathcal{C}_b^0$, we compute:

$$\int_{\mathbb{R}^d} \phi(x_1)\mathrm{d}\mu_1(x_1) = \int_{\mathbb{R}^d} \phi(x_1)\mathrm{d}\gamma(x_1, x_2) = \int_{\mathbb{R}^d} \phi(x_2)\mathrm{d}\gamma(x_1, x_2) = \int_{\mathbb{R}^d} \phi(x_2)\mathrm{d}\nu(x_2),$$

and thus $\mu_1 = \mu_2$. The converse is clear, but we detail for completeness. We show that $\mathrm{PS}_\theta(\mu, \mu) = 0$ for $\mu \in \mathcal{P}_2(\mathbb{R}^d)$: notice that $\mu_\theta[\mu, \mu] = Q_\theta\#\mu =: \mu_\theta$, take any $\gamma \in \Pi^*(\mu_\theta, \mu)$ and introduce by disintegration $\rho(\mathrm{d}y, \mathrm{d}x_1, \mathrm{d}x_2) := \mu_\theta(\mathrm{d}y)\gamma^y(\mathrm{d}x_1)\delta_{x_1=x_2}(\mathrm{d}x_1, \mathrm{d}x_2)$. We have $\rho \in \Gamma(\mu_\theta, \mu, \mu)$, and for $\rho$-almost-every $(y, x_1, x_2) \in \mathbb{R}^{3d}$, we have $\|x_1 - x_2\|_2^2 = 0$, hence $\mathrm{PS}_\theta(\mu, \mu) = 0$.

Symmetry is immediate from the definition (Eq. (20)). As for the upper-bound, by Proposition 1 we can take $\rho \in \Gamma(\mu_\theta[\mu_1, \mu_2], \mu_1, \mu_2)$ optimal for $\mathrm{PS}_\theta(\mu_1, \mu_2)$, and we have $\rho_{1,2} \in \Pi(\mu_1, \mu_2)$ and thus:

$$\mathrm{PS}_\theta^2(\mu_1, \mu_2) = \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho_{1,2}(\mathrm{d}x_1, \mathrm{d}x_2) \geq \mathrm{W}_2^2(\mu_1, \mu_2). \qquad \square$$

The triangle inequality is not satisfied in general, as shown in Example 3.

> **Example 3** ($\mathrm{PS}_\theta$ does not verify the triangle inequality). We represent the counter-example in Fig. 7. Consider $\theta := (1, 0)$ and:
>
> $$x_1 := (-1, 0), \; x_2 := (1, 5), \; \mu_1 := \tfrac{1}{2}\delta_{x_1} + \tfrac{1}{2}\delta_{x_2},$$
>
> $$y_1 := (-1, 5), \; y_2 := (1, 0), \; \mu_2 := \tfrac{1}{2}\delta_{y_1} + \tfrac{1}{2}\delta_{y_2},$$
>
> $$z_1 := (0, 0), \; z_2 := (0, 5), \; \mu_3 := \tfrac{1}{2}\delta_{z_1} + \tfrac{1}{2}\delta_{z_2}.$$
>
> First, we compute $\mathrm{PS}_\theta(\mu_1, \mu_2)$: we have
>
> $$u_1 := (-1, 0), \; u_2 := (1, 0), \; \mu_\theta[\mu_1, \mu_2] = \tfrac{1}{2}\delta_{u_1} + \tfrac{1}{2}\delta_{u_2},$$
>
> and then we see that there are unique optimal plans between $\mu_\theta[\mu_1, \mu_2]$ and each $\mu_1, \mu_2$:
>
> $$\Pi^*(\mu_\theta[\mu_1, \mu_2], \mu_1) = \left\{\gamma_{121} := \tfrac{1}{2}\delta_{u_1 \otimes x_1} + \tfrac{1}{2}\delta_{u_2 \otimes x_2}\right\}, \; \Pi^*(\mu_\theta[\mu_1, \mu_2], \mu_2) = \left\{\gamma_{122} := \tfrac{1}{2}\delta_{u_1 \otimes y_1} + \tfrac{1}{2}\delta_{u_2 \otimes y_2}\right\}.$$
>
> Using Theorem 1, we compute:
>
> $$\mathrm{PS}_\theta(\mu_1, \mu_2) = \sqrt{\tfrac{1}{2}\|x_1 - y_1\|_2^2 + \tfrac{1}{2}\|x_2 - y_2\|_2^2} = 5.$$

We now turn to $\text{PS}_\theta(\mu_1, \mu_3)$. This time, we have

$$v_1 := (-\tfrac{1}{2}, 0), \ v_2 := (\tfrac{1}{2}, 0), \ \mu_\theta[\mu_1, \mu_3] = \tfrac{1}{2}\delta_{v_1} + \tfrac{1}{2}\delta_{v_2}.$$

There is a unique optimal transport plan between $\mu_\theta[\mu_1, \mu_3]$ and $\mu_1$:

$$\Pi^*(\mu_\theta[\mu_1, \mu_3], \mu_1) = \left\{\gamma_{131} := \tfrac{1}{2}\delta_{v_1 \otimes x_1} + \tfrac{1}{2}\delta_{v_2 \otimes x_2}\right\}.$$

On the other hand, there are an infinite number of OT between $\mu_\theta[\mu_1, \mu_3]$ and $\mu_3$, which are convex combinations of two extremal plans (which correspond to the two permutations of $\{1, 2\}$):

$$\Pi^*(\mu_\theta[\mu_1, \mu_3], \mu_3) = \left\{\gamma_{133}(t) := (1-t)\left(\tfrac{1}{2}\delta_{v_1 \otimes z_1} + \tfrac{1}{2}\delta_{v_2 \otimes z_2}\right) + t\left(\tfrac{1}{2}\delta_{v_2 \otimes z_1} + \tfrac{1}{2}\delta_{v_1 \otimes z_2}\right), \ t \in [0, 1]\right\}.$$

Following Theorem 1, we have

$$\text{PS}_\theta(\mu_1, \mu_3) = \min_{t \in [0,1]} \tfrac{1}{2}\text{W}_2^2\left(\delta_{x_1}, \tfrac{1-t}{2}\delta_{z_1} + \tfrac{t}{2}\delta_{z_2}\right) + \tfrac{1}{2}\text{W}_2^2\left(\delta_{x_2}, \tfrac{1-t}{2}\delta_{z_2} + \tfrac{t}{2}\delta_{z_1}\right),$$

which is clearly minimal at $t = 0$, yielding

$$\text{PS}_\theta(\mu_1, \mu_3) = \sqrt{\tfrac{1}{2}\|x_1 - z_1\|_2^2 + \tfrac{1}{2}\|x_2 - z_2\|_2^2} = 1,$$

and by symmetry, $\text{PS}_\theta(\mu_2, \mu_3) = \text{PS}_\theta(\mu_1, \mu_3) = 1$. We conclude that the triangle inequality does not hold:

$$\text{PS}_\theta(\mu_1, \mu_2) = 5 > \text{PS}_\theta(\mu_2, \mu_3) + \text{PS}_\theta(\mu_1, \mu_3) = 2.$$
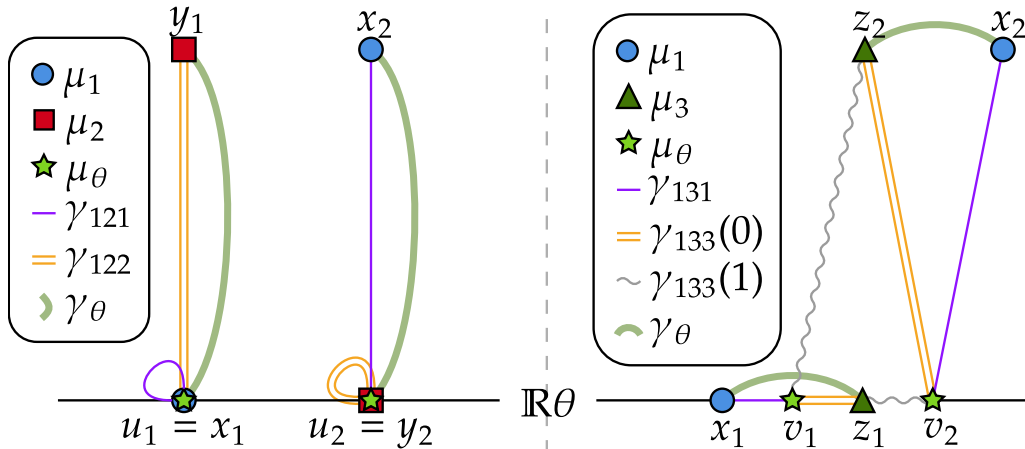


Figure 7: Counter-example from Example 3 to the triangle inequality for $\text{PS}_\theta$. Left: illustration of the couplings for $\text{PS}_\theta(\mu_1, \mu_2)$, with the optimal coupling $\gamma_\theta$ between $\mu_1$ and $\mu_2$ for $\text{PS}_\theta(\mu_1, \mu_2)$ represented with thick green lines. Right: illustration of the couplings for $\text{PS}_\theta(\mu_1, \mu_3)$. The optimal coupling $\gamma_\theta$ for $\text{PS}_\theta(\mu_1, \mu_3)$ corresponds to gluing $\gamma_{131}$ and $\gamma_{133}(0)$.

In the following, we show that $\text{PS}_\theta$ is lower semi-continuous with respect to the weak convergence of measures, along with a result on continuity of the middle $\mu_\theta[\mu_1, \mu_2]$. We speak of continuity with respect to the Euclidean topology on $\mathbb{S}^{d-1}$, and the weak topology on $\mathcal{P}_2(\mathbb{R}^d)$.

**Proposition 6.** The map $(\theta, \mu_1, \mu_2) \longmapsto \mu_\theta[\mu_1, \mu_2]$ is continuous, and $(\theta, \mu_1, \mu_2) \longmapsto \text{PS}_\theta(\mu_1, \mu_2)$ is lower semi-continuous.

*Proof.* Take measure sequences $\mu_1^{(n)} \xrightarrow[n \to +\infty]{w} \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mu_2^{(n)} \xrightarrow[n \to]{w} \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and a sequence

of projections $\theta_n \xrightarrow[n\longrightarrow+\infty]{} \theta \in \mathbb{S}^{d-1}$. By Lemma 12, we have

$$\mu_{\theta_n}[\mu_1^{(n)}, \mu_2^{(n)}] = \left[\left(\tfrac{1}{2}F^{[-1]}_{P_{\theta_n}\#\mu_1^{(n)}} + \tfrac{1}{2}F^{[-1]}_{P_{\theta_n}\#\mu_2^{(n)}}\right)\theta_n\right]\#\mathscr{L}_{[0,1]} =: f_n\#\mathscr{L}_{[0,1]},$$

and:

$$\mu_\theta[\mu_1, \mu_2] = \left[\left(\tfrac{1}{2}F^{[-1]}_{P_\theta\#\mu_1} + \tfrac{1}{2}F^{[-1]}_{P_\theta\#\mu_2}\right)\theta\right]\#\mathscr{L}_{[0,1]} =: f\#\mathscr{L}_{[0,1]}.$$

Let $i \in \{1,2\}$ and $g_n := F^{[-1]}_{P_{\theta_n}\#\mu_i^{(n)}}$, we show that $(g_n)$ converges pointwise $\mathscr{L}_{[0,1]}$-almost-everywhere to $g := F^{[-1]}_{P_\theta\#\mu_i}$. Since $P_{\theta_n}\#\mu_i^{(n)} \xrightarrow[n\longrightarrow+\infty]{w} P_\theta\#\mu_i$, we have by [Van00] Lemma 21.2 that for all $p \in [0,1]$ such that $g$ is continuous at $p$, $g_n(p) \xrightarrow[n\longrightarrow+\infty]{} g(p)$. Since $g$ is non-decreasing, it is continuous $\mathscr{L}_{[0,1]}$-almost-everywhere, and thus the convergence happens $\mathscr{L}_{[0,1]}$-almost-everywhere. Having shown that $f_n$ converges pointwise to $f$ $\mathscr{L}_{[0,1]}$-almost-everywhere, we deduce that

$$\mu_{\theta_n}[\mu_1^{(n)}, \mu_2^{(n)}] \xrightarrow[n\longrightarrow+\infty]{w} \mu_\theta[\mu_1, \mu_2].$$

By Proposition 2, we deduce that $(\theta, \mu_1, \mu_2) \longmapsto \mathrm{PS}_\theta(\mu_1, \mu_2)$ is lower semi-continuous. $\qquad\square$

We show in Example 4 that full continuity does not holds.

> **Example 4** ($\mathrm{PS}_\theta$ is not continuous with respect to the weak convergence)**.** Consider
>
> $$x_n := (-1 - 2^{-n}, 5), \ x' := (-1, 0), \ \mu_n := \tfrac{1}{2}\delta_{x_n} + \tfrac{1}{2}\delta_{x'},$$
>
> $$y := (1, 0), \ y' := (2, 5), \ \nu := \tfrac{1}{2}\delta_y + \tfrac{1}{2}\delta_{y'}.$$
>
> Obviously, $\mu_n \xrightarrow[n\longrightarrow+\infty]{w} \mu$, with $\mu = \tfrac{1}{2}\delta_{(-1,5)} + \tfrac{1}{2}\delta_{x'}$. For $n \in \mathbb{N}$, we compute easily that:
>
> $$\mathrm{PS}_\theta^2(\mu_n, \nu) = \tfrac{1}{2}\|x_n - y'\|_2^2 + \tfrac{1}{2}\|x' - y\|_2^2 = 36 + 3.2^{-n} + \tfrac{4^{-n}}{2} \xrightarrow[n\longrightarrow+\infty]{} 36.$$
>
> The limit does not coincide with $\mathrm{PS}_\theta^2(\mu, \nu) = \tfrac{13}{2}$. We summarise this counter-example in Fig. 8.
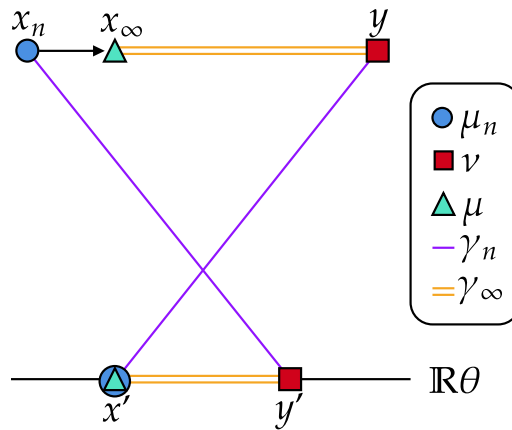


Figure 8: Representation of Example 4, showing a counter-example to the continuity of $\mathrm{PS}_\theta^2(\cdot, \nu)$ with respect to the weak convergence of measures. At each $n \in \mathbb{N}$, the coupling $\gamma_n$ associated to $\mathrm{PS}_\theta^2(\mu_n, \nu)$ between $\mu_n$ and $\nu$ (represented by purple lines) is imposed to assign $x_n$ to $y'$ and $x'$ to $y$. However, the coupling $\gamma_\infty$ associated to $\mathrm{PS}_\theta^2(\mu, \nu)$ represented by orange double lines has more freedom due to the fact that $P_\theta x_\infty = P_\theta x'$, and therefore can perform the less costly assignment of $x_\infty$ to $y$ and $x'$ to $y'$.

# 4    Correspondence of Pivot-Sliced and a Constrained Wasserstein Discrepancy

In this section, we will compare the quantity $\mathrm{PS}_\theta(\mu_1, \mu_2)$ defined in Eq. (20) with a particular lifting of the 1D sliced plan between $\mu_1$ and $\mu_2$. Namely, we will compare the two quantities:

$$
\mathrm{PS}_\theta^2(\mu_1, \mu_2) := \min_{\rho \in \Gamma(\mu_\theta[\mu_1, \mu_2], \mu_1, \mu_2)} \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2)
$$
$$
\overset{?}{=} \mathrm{CW}_\theta^2(\mu_1, \mu_2) := \min_{\substack{\omega \in \Pi(\mu_1, \mu_2) \\ (P_\theta, P_\theta)\#\omega = \pi_\theta[\mu_1, \mu_2]}} \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\omega(x_1, x_2),
\tag{31}
$$

where $\pi_\theta[\mu_1, \mu_2]$ is the unique optimal plan between $P_\theta\#\mu_1$ and $P_\theta\#\mu_2$. We introduce the following notation for the set of admissible plans $\omega$ for $\mathrm{CW}_\theta(\mu_1, \mu_2)$:

$$
\Omega_\theta(\mu_1, \mu_2) := \{\omega \in \Pi(\mu_1, \mu_2) : (P_\theta, P_\theta)\#\omega = \pi_\theta[\mu_1, \mu_2]\}.
\tag{32}
$$

Note that by compactness of $\Pi(\mu_1, \mu_2)$, continuity of $\omega \longmapsto (P_\theta, P_\theta)\#\omega$ and lower semi-continuity of $J := \omega \longmapsto \int_{\mathbb{R}^{2d}} \|\cdot - \cdot\|_2^2 \mathrm{d}\omega$, the infimum in $\mathrm{CW}_\theta^2$ is attained.

To draw a correspondence between $\mathrm{PS}_\theta$ and $\mathrm{CW}_\theta$, we will compare their optimisation sets, and to this end, we introduce the set $\Gamma_{\theta,1,2} \subset \Pi(\mu_1, \mu_2)$ defined as:

$$
\Gamma_{\theta,1,2}(\mu_1, \mu_2) := \{\rho_{1,2} : \rho \in \Gamma(\mu_\theta[\mu_1, \mu_2], \mu_1, \mu_2)\}.
\tag{33}
$$

We have by definition:

$$
\mathrm{CW}_\theta^2(\mu_1, \mu_2) = \min_{\omega \in \Omega_\theta(\mu_1, \mu_2)} J(\omega); \quad \mathrm{PS}_\theta^2(\mu_1, \mu_2) = \min_{\gamma \in \Gamma_{\theta,1,2}(\mu_1, \mu_2)} J(\gamma).
\tag{34}
$$

## 4.1    First Inequality: $\mathrm{PS}_\theta \leq \mathrm{CW}_\theta$

To prove a first inequality between $\mathrm{PS}_\theta$ and $\mathrm{CW}_\theta$, we will show that $\Omega_\theta(\mu_1, \mu_2) \subset \Gamma_{\theta,1,2}(\mu_1, \mu_2)$ (these sets are defined in Eq. (32) and Eq. (33)). We start with two Lemmas on Wasserstein means. The first result provides an explicit optimal coupling between a Wasserstein mean and the two measures.

> **Lemma 3.** Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and an optimal coupling $\gamma \in \Pi^*(\mu_1, \mu_2)$.
> Then $\mu_{\frac{1}{2}} := ((1-t)P_1 + tP_2)\#\gamma = \mathrm{Law}_{(X_1, X_2) \sim \gamma}\left[\frac{X_1 + X_2}{2}\right]$ belongs to $\mathrm{M}(\mu_1, \mu_2)$, and furthermore
> the coupling $\gamma_{\frac{1}{2}} := \mathrm{Law}_{(X_1, X_2) \sim \gamma}\left[\left(\frac{X_1 + X_2}{2}, X_1\right)\right]$ belongs to $\Pi^*(\mu_{\frac{1}{2}}, \mu_1)$.

*Proof.* By Proposition 3 we have $\mu_{\frac{1}{2}} \in M(\mu_1, \mu_2)$, and $\mathrm{W}_2^2(\mu_{\frac{1}{2}}, \mu_1) = \frac{1}{4}\mathrm{W}_2^2(\mu_1, \mu_2)$. We compute:

$$
\mathrm{W}_2^2(\mu_{\frac{1}{2}}, \mu_1) \leq \mathop{\mathbb{E}}_{(X_1, X_2) \sim \gamma}\left[\left\|\frac{X_1 + X_2}{2} - X_1\right\|_2^2\right] = \frac{1}{4}\mathop{\mathbb{E}}_{(X_1, X_2) \sim \gamma}\left[\|X_1 - X_2\|_2^2\right] = \frac{1}{4}\mathrm{W}_2^2(\mu_1, \mu_2),
$$

showing optimality of the coupling $\gamma_{\frac{1}{2}}$, since by Proposition 3, $\mathrm{W}_2^2(\mu_{\frac{1}{2}}, \mu_1) = \frac{1}{4}\mathrm{W}_2^2(\mu_1, \mu_2)$.    □

Note that Lemma 3 is also a consequence of [AGS05] Lemma 7.2.1 (which states a stronger result with more abstract language). The following second lemma relates an admissible plan $\omega \in \Omega_\theta(\mu_1, \mu_2)$ for $\mathrm{CW}_\theta(\mu_1, \mu_2)$ to an explicit optimal coupling between the projected middle $\mu_\theta[\mu_1, \mu_2]$ and the measures $\mu_1, \mu_2$, which will be useful to construct an admissible 3-plan for $\mathrm{PS}_\theta(\mu_1, \mu_2)$.

> **Lemma 4.** Let $\omega \in \Pi(\mu_1, \mu_2)$ such that $(P_\theta, P_\theta)\#\omega = \Pi^*(P_\theta\#\mu_1, P_\theta\#\mu_2)$. Let $(X_1, X_2) \sim \omega$
> and $Y := \frac{P_\theta X_1 + P_\theta X_2}{2}\theta$. Then $\mathrm{Law}[Y] = \mu_\theta[\mu_1, \mu_2]$, and $\mathrm{Law}[(Y, X_i)] \in \Pi^*(\mu_\theta[\mu_1, \mu_2], \mu_i), \; i \in$

$\{1, 2\}$.

*Proof.* First, we apply Lemma 3 to the optimal coupling $(P_\theta X_1, P_\theta X_2)$, which shows that $\mathrm{Law}[P_\theta Y] = P_\theta \# \mu_\theta[\mu_1, \mu_2]$, thus that $\mathrm{Law}[Y] = \mu_\theta[\mu_1, \mu_2]$. Lemma 3 also shows that $(P_\theta Y, P_\theta X_1)$ is the optimal coupling between $P_\theta \# \mu_\theta[\mu_1, \mu_2]$ and $P_\theta \# \mu_1$. Then by Eq. (29) in Proposition 4, it follows that $\mathrm{Law}[(Y, X_1)] \in \Pi^*(\mu_\theta[\mu_1, \mu_2], \mu_1)$, and the same reasoning applies to $\mathrm{Law}[(Y, X_2)] \in \Pi^*(\mu_\theta[\mu_1, \mu_2], \mu_2)$. $\qquad\square$

Using Lemma 3 and Lemma 4, we can now show an inequality between $\mathrm{PS}_\theta$ and $\mathrm{CW}_\theta$:

> **Proposition 7.** Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, and $\theta \in \mathbb{S}^{d-1}$. The two sets defined in Eq. (32) and Eq. (33) verify $\Omega_\theta(\mu_1, \mu_2) \subset \Gamma_{\theta, 1, 2}(\mu_1, \mu_2)$, and the two quantities defined in Eq. (31) verify $\mathrm{PS}_\theta(\mu_1, \mu_2) \le \mathrm{CW}_\theta(\mu_1, \mu_2)$.

*Proof.* Let $\omega \in \Pi(\mu_1, \mu_2)$ such that $(P_\theta, P_\theta) \# \omega = \Pi^*(P_\theta \# \mu_1, P_\theta \# \mu_2)$ optimal for $\mathrm{CW}_\theta(\mu_1, \mu_2)$. Consider $(X_1, X_2) \sim \omega$ and $Y := \frac{P_\theta X_1 + P_\theta X_2}{2} \theta$. By Lemma 4, we have $\mathrm{Law}[Y] = \mu_\theta[\mu_1, \mu_2]$, and $\mathrm{Law}[(Y, X_i)] \in \Pi^*(\mu_\theta[\mu_1, \mu_2], \mu_i)$, $i \in \{1, 2\}$. By definition the 3-plan $\rho$ defined by $\rho := \mathrm{Law}[(Y, X_1, X_2)]$ belongs to $\Gamma(\mu_\theta[\mu_1, \mu_2], \mu_1, \mu_2)$, thus $\omega \in \Gamma_{\theta, 1, 2}(\mu_1, \mu_2)$. We compute:

$$\mathrm{PS}_\theta^2(\mu_1, \mu_2) \le \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2) = \int_{\mathbb{R}^{2d}} \|x_2 - x_2\|_2^2 \mathrm{d}\omega(x_1, x_2) = \mathrm{CW}_\theta^2(\mu_1, \mu_2). \qquad\square$$

## 4.2 Converse Inequality: $\mathrm{PS}_\theta \ge \mathrm{CW}_\theta$

To show the converse inequality $\mathrm{CW}_\theta(\mu_1, \mu_2) \le \mathrm{PS}_\theta(\mu_1, \mu_2)$, we will use more technical arguments from [AGS05] Lemma 7.2.1, which will show that (denoting $\mu_\theta := \mu_\theta[\mu_1, \mu_2]$,) for $i \in \{1, 2\}$, the unique optimal plan $\pi_i$ between $P_\theta \# \mu_\theta$ and $P_\theta \# \mu_i$ is induced by a transport map $T_i$, i.e. $\pi_i = (I, T_i) \# P_\theta \# \mu_\theta$. This is a consequence of the fact that $\mu_\theta$ is chosen as the middle of the geodesic between $P_\theta \# \mu_1$ and $P_\theta \# \mu_2$, and remarkably holds without atomless assumptions on the $P_\theta \# \mu_i$.

> **Theorem 2.** Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\theta \in \mathbb{S}^{d-1}$. Then the two sets defined in Eq. (32) and Eq. (33) verify $\Gamma_{\theta, 1, 2}(\mu_1, \mu_2) = \Omega_\theta(\mu_1, \mu_2)$, and the two quantities defined in Eq. (31) verify $\mathrm{PS}_\theta(\mu_1, \mu_2) = \mathrm{CW}_\theta(\mu_1, \mu_2)$.

*Proof.* We have already shown that $\mathrm{PS}_\theta(\mu_1, \mu_2) \le \mathrm{CW}_\theta(\mu_1, \mu_2)$ in Proposition 7. We now show that $\Gamma_{\theta, 1, 2}(\mu_1, \mu_2) \subset \Omega_\theta(\mu_1, \mu_2)$ (we write $\mu_\theta := \mu_\theta[\mu_1, \mu_2]$, and $\pi_\theta$ the unique element of $\Pi^*(P_\theta \# \mu_1, P_\theta \# \mu_2)$). Let $\rho \in \Gamma(\mu_\theta, \mu_1, \mu_2)$. We introduce $\eta := (P_\theta, P_\theta, P_\theta) \# \rho \in \Pi(\nu_{\frac{1}{2}}, \nu_1, \nu_2)$, where for convenience we write $\nu_i := P_\theta \# \mu_i$ for $i \in \{1, 2\}$ and $\nu_{\frac{1}{2}} := P_\theta \# \mu_\theta$. By Eq. (29) in Proposition 4, we have $\eta_{0,i} \in \Pi^*(\nu_{\frac{1}{2}}, \nu_i)$.

We now write $\eta$ using OT maps. By [AGS05] Lemma 7.2.1, since $\nu_{\frac{1}{2}} = \mathrm{M}(\nu_1, \nu_2)$ (i.e. it is the middle of the constant-speed geodesic between $\nu_1$ and $\nu_2$, which is unique since the measures are one-dimensional), for $i \in \{1, 2\}$ the transport plan $\eta_{0,i} \in \Pi^*(\nu_{\frac{1}{2}}, \nu_i)$ is induced by a non-decreasing transport map $T_i$, which is to say that $\eta_{0,i} = (I, T_i) \# \nu_{\frac{1}{2}}$. It follows that for $\eta$-almost-every $(t, s_1, s_2) \in \mathbb{R}^3$, we have $s_1 = T_1(t)$ and $s_2 = T_2(t)$.

We now verify that $\eta_{1,2} \in \Pi^*(\nu_1, \nu_2)$ using the cyclical monotonicity criterion: Let $(s_1, s_2), (s_1', s_2') \in \mathrm{supp}\, \eta_{1,2}$ such that $s_1 < s_1'$. Our earlier considerations on $\eta$ show that there exists $t, t' \in \mathbb{R}$ verifying $s_1 = T_1(t)$, $s_2 = T_2(t)$ and $s_1' = T_1(t')$, $s_2' = T_2(t')$. Since $s_1 = T_1(t) < T_1(t') = s_1'$ and $T_1$ is non-decreasing, we deduce $t < t'$. Now since $T_2$ is non-decreasing, $t < t'$ implies that $s_2 = T_2(t) \le T_2(t') = s_2'$. We have shown the following property of $\eta_{1,2}$:

$$\forall (s_1, s_2), (s_1', s_2') \in \mathrm{supp}\, \eta_{1,2}, \; s_1 < s_1' \implies s_2 \le s_2'. \tag{35}$$

By [San15] Lemma 2.8, Eq. (35) implies that $\eta_{1,2}$ is the co-monotone plan between $\nu_1$ and $\nu_2$, and by [San15] Theorem 2.9, we conclude that $\eta_{1,2} \in \Pi^*(\nu_1, \nu_2)$.

Having shown that $\eta_{1,2} \in \Pi^*(\nu_1, \nu_2)$, we conclude that $(P_\theta, P_\theta)\#\rho_{1,2} \in \Pi^*(P_\theta\#\mu_1, P_\theta\#\mu_2)$, and by definition we conclude $\rho_{1,2} \in \Omega_\theta(\mu_1, \mu_2)$, which shows the inclusion $\Gamma_{\theta,1,2}(\mu_1, \mu_2) \subset \Omega_\theta(\mu_1, \mu_2)$, and equality is obtained by combining with Proposition 7. By Eq. (34) we conclude that $\mathrm{CW}_\theta(\mu_1, \mu_2) = \mathrm{PS}_\theta(\mu_1, \mu_2)$. □

## 4.3 Triangle Inequality for $\mathrm{PS}_\theta$ for Projection-Atomless Measures

Using Theorem 2 and the following technical lemma on one-dimensional 3-plans, we will show that $\mathrm{PS}_\theta$ is a metric on the set of probability measures whose projections on $\mathbb{R}\theta$ are atomless. We show that in the one-dimensional atomless case, 3-plans with two optimal bi-marginals automatically verify that *all* their bi-marginals are optimal. In terms of random variables, Lemma 5 states that if $(X_1, X_2)$ and $(X_1, X_3)$ are optimal couplings, then so is $(X_2, X_3)$.

**Lemma 5.** Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_2(\mathbb{R})$ such that $\mu_1$ and $\mu_2$ are atomless, and let $\rho \in \Pi(\mu_1, \mu_2, \mu_3)$ be a 3-plan such that $\rho_{1,2} \in \Pi^*(\mu_1, \mu_2)$ and $\rho_{1,3} \in \Pi^*(\mu_1, \mu_3)$. Then $\rho_{2,3} \in \Pi^*(\mu_2, \mu_3)$.

*Proof.* For $i \in \{1, 2, 3\}$, introduce the c.d.f. $F_i$ of $\mu_i$. Take $(X_1, X_2, X_3) \sim \rho$. By [San15] Theorem 2.9, since $\mu_1$ is atomless and $(X_1, X_3)$ is optimal, we have almost-surely $X_3 = F_3^{[-1]} \circ F_1(X_1)$. Likewise, since $\mu_2$ is atomless and $(X_2, X_1)$ is optimal, we have almost-surely $X_1 = F_1^{[-1]} \circ F_2(X_2)$. Combining these equalities yields almost-surely:

$$X_3 = F_3^{[-1]} \circ F_1(X_1) = F_3^{[-1]} \circ F_1 \circ F_1^{[-1]} \circ F_2(X_2).$$

By continuity of $F_1$ (since $\mu_1$ is atomless) and defining $F_1(-\infty) := 0$ and $F_1(+\infty) := 1$, we have $F_1(\mathbb{R}) \cup \{F_1(-\infty)\} \cup \{F_1(+\infty)\} = [0, 1]$, allowing us to apply [EH13] Proposition 2.3 item 4), which yields $F_1 \circ F_1^{[-1]} = I_{[0,1]}$.

We have shown that almost-surely $X_3 = F_3^{[-1]} \circ F_2(X_2)$, which shows by [San15] Theorem 2.9 that $(X_2, X_3)$ is the optimal coupling between $\mu_2$ and $\mu_3$. □

We can now show that $\mathrm{PS}_\theta$ verifies the triangle inequality on the set of probability measures with atomless projections. Combining this statement with Proposition 5 shows that $\mathrm{PS}_\theta$ is a metric this subset of $\mathcal{P}_2(\mathbb{R}^d)$.

**Proposition 8.** Let $\theta \in \mathbb{S}^{d-1}$ and $\mathcal{P}_{2,a}(\mathbb{R}^d, \theta)$ be the set of probability measures $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $P_\theta\#\mu$ is atomless. The quantity $\mathrm{PS}_\theta$ is a metric on $\mathcal{P}_{2,a}(\mathbb{R}^d, \theta)$.

*Proof.* First, by Proposition 5, it only remains to show the triangle inequality. Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_{2,a}(\mathbb{R}^d, \theta)$ and let $\omega_{1,2} \in \Omega_\theta(\mu_1, \mu_2)$ be optimal for $\mathrm{CW}_\theta(\mu_1, \mu_2)$, and likewise let $\omega_{2,3} \in \Omega_\theta(\mu_2, \mu_3)$ be optimal for $\mathrm{CW}_\theta(\mu_2, \mu_3)$. We apply the standard gluing technique (see for example [San15] Lemma 5.5): the second marginal of $\omega_{1,2}$ and the first marginal of $\omega_{2,3}$ are both $\mu_2$, hence we can write their disintegrations with respect to $\mu_2$ as:

$$\omega_{1,2}(\mathrm{d}x_1, \mathrm{d}x_2) = \mu_2(\mathrm{d}x_2)\omega_{1,2}^{x_2}(\mathrm{d}x_1), \quad \omega_{2,3}(\mathrm{d}x_2, \mathrm{d}x_3) = \mu_2(\mathrm{d}x_2)\omega_{2,3}^{x_2}(\mathrm{d}x_3).$$

We now introduce the "composition" 3-plan $\rho \in \Pi(\mu_1, \mu_2, \mu_3)$ as:

$$\rho(\mathrm{d}x_1, \mathrm{d}x_2, \mathrm{d}x_3) := \mu_2(\mathrm{d}x_2)\omega_{1,2}^{x_2}(\mathrm{d}x_1)\omega_{2,3}^{x_2}(\mathrm{d}x_3).$$

Writing $\rho_\theta := (P_\theta, P_\theta, P_\theta)\#\rho$, by definition of $\omega_{1,2}$ and $\omega_{2,3}$, we have $[\rho_\theta]_{1,2} = \pi_\theta[\mu_1, \mu_2]$ and $[\rho_\theta]_{2,3} = \pi_\theta[\mu_2, \mu_3]$. By Lemma 5, we deduce that $\rho_{1,3} = \pi_\theta[\mu_1, \mu_3]$, since each $P_\theta\#\mu_i$ is atomless. This shows

that $\rho_{1,3} \in \Omega_\theta(\mu_1, \mu_2)$. Denoting $\phi_i := (x_1, x_2, x_3) \longmapsto x_i$ for $i \in \{1, 2, 3\}$, we have:

$$
\begin{aligned}
\mathrm{CW}_\theta(\mu_1, \mu_3) &\leq \|\phi_1 - \phi_3\|_{L^2(\rho_{1,3})} = \|\phi_1 - \phi_3\|_{L^2(\rho)} \\
&\leq \|\phi_1 - \phi_2\|_{L^2(\rho)} + \|\phi_2 - \phi_3\|_{L^2(\rho)} = \|\phi_1 - \phi_2\|_{L^2(\omega_{1,2})} + \|\phi_2 - \phi_3\|_{L^2(\omega_{2,3})} \\
&= \mathrm{CW}_\theta(\mu_1, \mu_2) + \mathrm{CW}_\theta(\mu_2, \mu_3).
\end{aligned}
$$

Using Theorem 2, we deduce the triangle inequality for $\mathrm{PS}_\theta$. $\qquad\square$

# 5   A Monge Formulation of $\mathrm{PS}_\theta$ Between Point Clouds

## 5.1   The Case of Non-Ambiguous Projections

A direct consequence of Theorem 1 is that, in the case of point clouds with non-ambiguous projections, the computation of $\mathrm{PS}_\theta$ can be done simply by sorting the projections and taking the associated plan between the projected measures.

> **Corollary 1.** Let $(x_1, \cdots, x_n) \in (\mathbb{R}^d)^n$ and $(y_1, \cdots, y_n) \in (\mathbb{R}^d)^n$, and $\theta \in \mathbb{S}^{d-1}$ such that the families $(P_\theta x_i)_i$ and $(P_\theta y_i)_i$ are injective. Then for the measures $\mu_1 := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$, $\mu_2 := \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$, it holds
> $$ \mathrm{PS}_\theta^2(\mu_1, \mu_2) = \frac{1}{n} \sum_{i=1}^n \|x_{\sigma_\theta(i)} - y_{\tau_\theta(i)}\|_2^2, $$
> where $\mathrm{PS}_\theta$ is introduced in Definition 4 and where $\sigma_\theta, \tau_\theta$ are the (unique) permutations sorting $(P_\theta x_i)$ and $(P_\theta y_i)$. For injective families $(x_i)$ and $(y_i)$, the injectivity assumptions holds for $\mathcal{U}(\mathbb{S}^{d-1})$-almost-every $\theta \in \mathbb{S}^{d-1}$.

*Proof.* We begin under the injectivity assumptions, which allows us to define $\sigma_\theta, \tau_\theta$ as the (unique) permutations sorting $(P_\theta x_i)$ and $(P_\theta y_i)$ respectively, and for $i \in [\![1, n]\!]$, let $z_i := \frac{1}{2} P_\theta(x_{\sigma_\theta(i)} + y_{\tau_\theta(i)})$. We remark that the family $(z_i)$ is increasing by construction (we provide further details at the end of the proof for almost-sure injectivity) and denote $\nu := \frac{1}{n} \sum_i \delta_{z_i}$. Let $\gamma_1 \in \Pi^*(\nu, \mu_1)$, and write

$$ \gamma_1 = \sum_{i,j} A_{i,j} \delta_{(z_i, x_{\sigma_\theta(j)})}. $$

By Proposition 4 and the injectivity assumptions, we have $(P_\theta, P_\theta) \# \gamma_1 = \frac{1}{n} \sum_i \delta_{(P_\theta z_i, P_\theta x_{\sigma_\theta(i)})}$, and thus injectivity allows us to identify the coefficients $A_{i,j}$, yielding $\gamma_1 = \frac{1}{n} \sum_i \delta_{(z_i, x_{\sigma_\theta(i)})}$, and in particular, for any $i \in [\![1, n]\!]$, $\gamma_1^{z_i} = \delta_{x_{\sigma_\theta(i)}}$. The same reasoning applies to $\gamma_2 \in \Pi^*(\nu, \mu_2)$, and thus Theorem 1 yields

$$ \mathrm{PS}_\theta^2(\mu_1, \mu_2) = \mathrm{W}_\nu^2(\mu_1, \mu_2) = \frac{1}{n} \sum_{i=1}^n \|x_{\sigma_\theta(i)} - y_{\tau_\theta(i)}\|_2^2 $$

Regarding the almost-sure injectivity claim, assume now that the families $(x_i)$ and $(y_i)$ are injective, and take $\theta \sim \mathcal{U}(\mathbb{S}^{d-1})$. Then $(P_\theta x_i)$ is almost-surely injective, since $\mathbb{P}(P_\theta x_i = P_\theta x_j) = \mathbb{P}(\theta \in (x_i - x_j)^\top)$. The same reasoning applies to $(P_\theta y_i)$, and the injectivity of $(z_i)$ comes from the fact that almost-surely, for $i < j$, we have $P_\theta x_{\sigma_\theta(i)} < P_\theta x_{\sigma_\theta(j)}$, and $P_\theta y_{\tau_\theta(i)} < P_\theta x_{\tau_\theta(j)}$, hence by sum $z_i < z_j$, almost-surely. $\qquad\square$

## 5.2   Problem Formulation and Reduction to Sorted Projections

The natural question that arises is the impact of projection ambiguity, i.e. non-injectivity of $(P_\theta x_i)$ or $(P_\theta y_j)$. In this section, we will start from the equality $\mathrm{PS}_\theta = \mathrm{CW}_\theta$ from Theorem 2, to provide the following Monge formulation of $\mathrm{PS}_\theta$ between point clouds (without injectivity assumptions), that

we will prove in Theorem 4:

$$\mathrm{CW}_\theta^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{j=1}^n \delta_{y_j}\right) = \min_{(\sigma,\tau)\in\mathfrak{S}_\theta(X,Y)} \frac{1}{n}\sum_{i=1}^n \|x_{\sigma(i)} - y_{\tau(i)}\|_2^2,$$

where $\mathfrak{S}_\theta(X,Y)$ is the set of pairs of permutations $(\sigma,\tau)$ such that $\sigma$ sorts $(P_\theta x_i)_{i=1}^n$ and $\tau$ sorts $(P_\theta y_i)_{i=1}^n$, for given $X := (x_1,\cdots,x_n) \in \mathbb{R}^{n\times d}$ and $Y := (y_1,\cdots,y_n) \in \mathbb{R}^{n\times d}$:

$$\mathfrak{S}_\theta(X,Y) := \mathfrak{S}_\theta(X)\times\mathfrak{S}_\theta(Y),\ \mathfrak{S}_\theta(X) := \left\{\sigma\in\mathfrak{S}_n : \forall i \in [\![1,n-1]\!],\ P_\theta x_{\sigma(i)} \leq P_\theta x_{\sigma(i+1)},\right\}, \quad (36)$$

with an analogous definition for $\mathfrak{S}_\theta(Y)$. We will reduce to the case where the identity permutation sorts the projections $(P_\theta x_i)_{i=1}^n$ and $(P_\theta y_i)_{i=1}^n$, which will greatly simplify notation and proofs. The following Lemma states that re-labelling the points does not change the value of $\mathrm{CW}_\theta$ and of the Monge formulation.

> **Lemma 6.** Let $X,Y \in \mathbb{R}^{n\times d}$ and $\sigma_0,\tau_0 \in \mathfrak{S}_n$. Denote by $X\circ\sigma_0 := (x_{\sigma_0(1)},\cdots,x_{\sigma_0(n)})_{i=1}^n$ and likewise $Y\circ\tau_0 := (y_{\tau_0(1)},\cdots,y_{\tau_0(n)})_{i=1}^n$. Then we have:
>
> $$\mathrm{CW}_\theta^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{j=1}^n \delta_{y_j}\right) = \mathrm{CW}_\theta^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_{\sigma_0(i)}}, \frac{1}{n}\sum_{j=1}^n \delta_{y_{\tau_0(j)}}\right),$$
>
> and for the Monge formulation, we have the following cost equality:
>
> $$\min_{(\sigma,\tau)\in\mathfrak{S}_\theta(X,Y)} \frac{1}{n}\sum_{i=1}^n \|x_{\sigma(i)} - y_{\tau(i)}\|_2^2 = \min_{(\sigma,\tau)\in\mathfrak{S}_\theta(X\circ\sigma_0,Y\circ\tau_0)} \frac{1}{n}\sum_{i=1}^n \|x_{\sigma_0\circ\sigma(i)} - y_{\tau_0\circ\tau(i)}\|_2^2. \quad (37)$$

*Proof.* The first equality is simply a consequence of the equality between measures:

$$\frac{1}{n}\sum_{i=1}^n \delta_{x_i} = \frac{1}{n}\sum_{i=1}^n \delta_{x_{\sigma_0(i)}}, \quad \frac{1}{n}\sum_{j=1}^n \delta_{y_j} = \frac{1}{n}\sum_{j=1}^n \delta_{y_{\tau_0(j)}}.$$

For the second equality, notice that a permutation $\sigma\in\mathfrak{S}_n$ sorts $(P_\theta x_{\sigma_0(i)})_{i=1}^n$ if and only if $P_\theta x_{\sigma_0\circ\sigma(1)} \leq \cdots \leq P_\theta x_{\sigma_0\circ\sigma(n)}$ if and only if $\sigma_0\circ\sigma$ sorts $(P_\theta x_i)_{i=1}^n$, thus we obtain:

$$\mathfrak{S}_\theta(X\circ\sigma_0,Y\circ\tau_0) = \left\{(\sigma_0^{-1}\circ\sigma, \tau_0^{-1}\circ\tau),\ (\sigma,\tau)\in\mathfrak{S}_\theta(X,Y)\right\}.$$

Eq. (37) follows by change of variables.                                                                                □

Thanks to Lemma 6, we can assume without loss of generality (for the cost values) that the identity permutation sorts the projections $(P_\theta x_i)_{i=1}^n$ and $(P_\theta y_i)_{i=1}^n$. We formulate this assumption as follows:

> **Assumption 1.** The points $X,Y \in \mathbb{R}^{n\times d}$ and $\theta\in\mathbb{S}^{d-1}$ are such that:
>
> $$P_\theta x_1 \leq \cdots \leq P_\theta x_n \text{ and } P_\theta y_1 \leq \cdots \leq P_\theta y_n.$$

Assumption 1 holds up to relabelling the points $(x_i)$ and $(y_j)$: taking $\sigma\in\mathfrak{S}_n$ sorting $(P_\theta x_i)$ and $\tau\in\mathfrak{S}_n$ sorting $(P_\theta y_j)$, the relabelled points $\widetilde{X} := (x_{\sigma(i)}) =: X\circ\sigma$ and $\widetilde{Y} := (y_{\tau(j)}) =: Y\circ\tau$ verify the condition.

## 5.3   A Kantorovich Formulation of $\mathrm{CW}_\theta$ Between Point Clouds

We begin by a characterisation of $\mathfrak{S}_\theta(X,Y)$, which states that a pair $(\sigma,\tau)$ belongs to $\mathfrak{S}_\theta(X,Y)$ if and only if each "ambiguity" set $\{i : P_\theta x_i = t\}$ is stable by $\sigma$ and likewise for $\tau$.

**Lemma 7.** Let $X := (x_1, \cdots, x_n) \in \mathbb{R}^{n \times d}$, $Y := (y_1, \cdots, y_n) \in \mathbb{R}^{n \times d}$ and $\theta \in \mathbb{S}^{d-1}$ verifying Assumption 1. Let $A := \#\{P_\theta x_i\}_{i=1}^n$ and $B := \#\{P_\theta y_j\}_{j=1}^n$. Write $\{P_\theta x_i\}_{i=1}^n = (s_a)_{a=1}^A$ where $s_1 < \cdots < s_A$ and likewise $\{P_\theta y_j\}_{j=1}^n = (t_b)_{b=1}^B$ where $t_1 < \cdots < t_B$. Introduce the "ambiguity group" sets:

$$\forall a \in [\![1, A]\!],\ I_a := \{i \in [\![1, n]\!] : P_\theta x_i = s_a\}, \quad \forall b \in [\![1, B]\!],\ J_b := \{j \in [\![1, n]\!] : P_\theta y_j = t_b\}. \tag{38}$$

Then the set $\mathfrak{S}_\theta(X, Y)$ can be re-written as follows:

$$\mathfrak{S}_\theta(X, Y) = \left\{(\sigma, \tau) \in \mathfrak{S}_n^2 : \forall a \in [\![1, A]\!],\ \sigma(I_a) = I_a,\ \forall b \in [\![1, B]\!],\ \tau(J_b) = J_b\right\}. \tag{39}$$

*Proof.* We show the property $\mathfrak{S}_\theta(X) = \widetilde{\mathfrak{S}} := \{\sigma \in \mathfrak{S}_n : \forall a \in [\![1, A]\!],\ \sigma(I_a) = I_a\}$ by double inclusion. First, since $s_1 < \cdots < s_A$, the inclusion $\widetilde{\mathfrak{S}} \subset \mathfrak{S}_\theta(X)$ is clear. For the converse inclusion, take $\sigma \in \mathfrak{S}_\theta(X)$. By definition and by Assumption 1, we have:

$$P_\theta x_1 \leq \cdots \leq P_\theta x_n; \quad P_\theta x_{\sigma(1)} \leq \cdots \leq P_\theta x_{\sigma(n)},$$

which implies that $\forall i \in [\![1, n]\!]$, $P_\theta x_i = P_\theta x_{\sigma(i)}$. Take now $a$ the unique element of $[\![1, A]\!]$ such that $i \in I_a$. We have $s_a = P_\theta x_i = P_\theta x_{\sigma(i)}$ and thus $\sigma(i) \in I_a$, and we conclude that $\sigma \in \widetilde{\mathfrak{S}}$. The proof for $\mathfrak{S}_\theta(Y)$ follows verbatim, and Eq. (39) follows from the definition (see Eq. (36)). $\qquad\square$

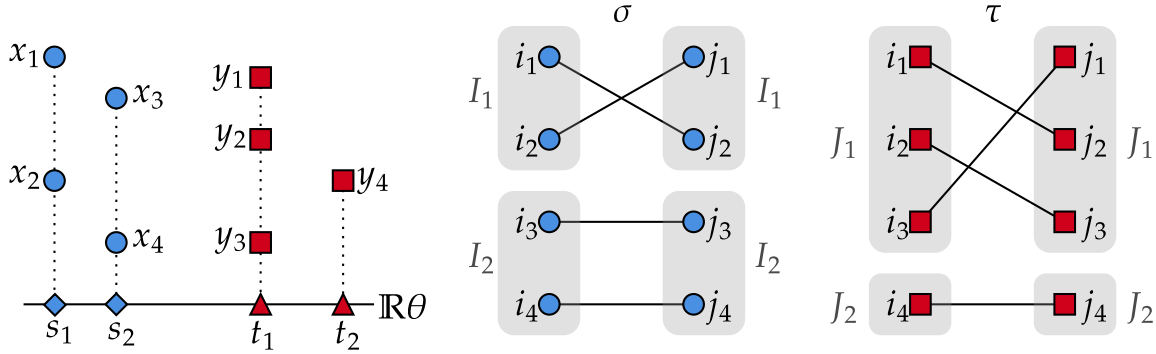To illustrate Lemma 7, we consider an example with projection ambiguities in Fig. 9.



Figure 9: In this example we consider two discrete uniform measures with $n := 4$ points with projection ambiguity: $s_1 := P_\theta x_1 = P_\theta x_2 < s_2 := P_\theta x_3 = P_\theta x_4$ and $t_1 := P_\theta y_1 = P_\theta y_2 = P_\theta y_3 < t_2 := P_\theta x_4$. In the notation of Lemma 7, we have $A = B = 2$ and $I_1 = \{i_1, i_2\}$, $I_2 = \{i_3, i_4\}$, $J_1 = \{j_1, j_2, j_3\}$ and $J_2 = \{j_4\}$. We consider a permutation pair $(\sigma, \tau) \in \mathfrak{S}_\theta(X, Y)$, specifically $\sigma := (2, 1, 3, 4)$ and $\tau := (2, 3, 1, 4)$. We see that $\sigma$ sorts the sequence $(P_\theta x_i)_{i=1}^n$ and that $I_1$ and $I_2$ are stable by $\sigma$, and likewise for $\tau$.

We now write a discrete Kantorovich formulation of $\mathrm{CW}_\theta$ between point clouds, whose expression we will be able to simplify later. The main idea is that transport plans $P$ are constrained to exchange exactly as much mass between $I_a$ and $J_b$ as the one-dimensional OT plan $\pi_\theta$ between $P_\theta \# \mu$ and $P_\theta \# \nu$ sends from $s_a$ to $t_b$, as illustrated in Fig. 10.

**Proposition 9.** Under Assumption 1, let $\mu := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $\nu := \frac{1}{n} \sum_{j=1}^n \delta_{y_j}$ be empirical measures. Then the $\mathrm{CW}_\theta$ discrepancy introduced in Eq. (31) has the following expression:

$$\mathrm{CW}_\theta^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{j=1}^n \delta_{y_j}\right) = \min_{P \in \mathbb{U} \cap \mathcal{P}_\theta(X, Y)} \sum_{i=1}^n \sum_{j=1}^n \|x_i - y_j\|_2^2 P_{i,j}, \tag{40}$$

$$\mathbb{U} := \{P \in \mathbb{R}_+^{n \times n} : P\mathbf{1} = \tfrac{1}{n}\mathbf{1},\ P^\top\mathbf{1} = \tfrac{1}{n}\mathbf{1}\}, \tag{41}$$

$$\mathcal{P}_\theta(X,Y) := \left\{ P \in \mathbb{R}^{n \times n} : \forall (a,b) \in [\![1,A]\!] \times [\![1,B]\!], \sum_{(i,j) \in I_a \times J_b} P_{i,j} = \frac{1}{n} \#(I_a \cap J_b) \right\}. \tag{42}$$

*Proof.* Fix $\omega \in \Omega_\theta(\mu,\nu)$ (see Eq. (32)). By the marginal constraints, we have $\operatorname{supp}\omega \subset \{(x_i,y_j)\}_{i,j}$, allowing us to define $P \in \mathbb{R}_+^{n \times n}$ by $\forall (i,j) \in [\![1,n]\!]^2$, $P_{i,j} = \omega(\{(x_i,y_j)\})$. Since $\omega \in \Pi(\mu,\nu)$, we verify immediately that $P \in \mathbb{U}$. As for the constraint $(P_\theta,P_\theta)\#\omega = \pi_\theta[\mu,\nu] =: \pi_\theta$, notice that by construction (see the notations in Lemma 7), we can write by [San15] Theorem 2.9 for *any* $(\sigma,\tau) \in \mathfrak{S}_\theta(X,Y)$ that $\pi_\theta = \frac{1}{n}\sum_i \delta_{(P_\theta x_{\sigma(i)}, P_\theta y_{\tau(i)})}$. Since $\{P_\theta x_i\}_{i=1}^n = (s_a)_{a=1}^A$ and $\{P_\theta y_j\}_{j=1}^n = (t_b)_{b=1}^B$, it follows that for any $(a,b) \in [\![1,A]\!] \times [\![1,B]\!]$:

$$\pi_\theta(\{(s_a,t_b)\}) = \sum_{k=1}^n \frac{1}{n}\mathbb{1}(\sigma(k) \in I_a)\mathbb{1}(\tau(k) \in J_b) = \sum_{k=1}^n \frac{1}{n}\mathbb{1}(k \in I_a \cap J_b) = \frac{1}{n}\#(I_a \cap J_b),$$

where we have used that $\sigma(I_a) = I_a$ and $\tau(J_b) = J_b$, which holds by Lemma 7. We can now show that $P \in \mathcal{P}_\theta(X,Y)$ using the constraint $(P_\theta,P_\theta)\#\omega = \pi_\theta$:

$$\frac{1}{n}\#(I_a \cap J_b) = \pi_\theta(\{(s_a,t_b)\}) = (P_\theta,P_\theta)\#\omega(\{(s_a,t_b)\}) = \sum_{(i,j) \in I_a \times J_b} P_{i,j}.$$

The cost $\int_{\mathbb{R}^{2d}} \|x-y\|_2^2 \mathrm{d}\omega(x,y)$ writes $\sum_{i,j} \|x_i-y_j\|_2^2 P_{i,j}$ by definition of $P$. Conversely, it can readily be checked with the same computations that for any $P \in \mathbb{U} \cap \mathcal{P}_\theta(X,Y)$, the coupling $\omega := \sum_{i,j} P_{i,j}\delta_{(x_i,y_j)}$ belongs to $\Omega_\theta(\mu,\nu)$, and yields the same transportation cost. We conclude that the equality in Eq. (40) holds. □
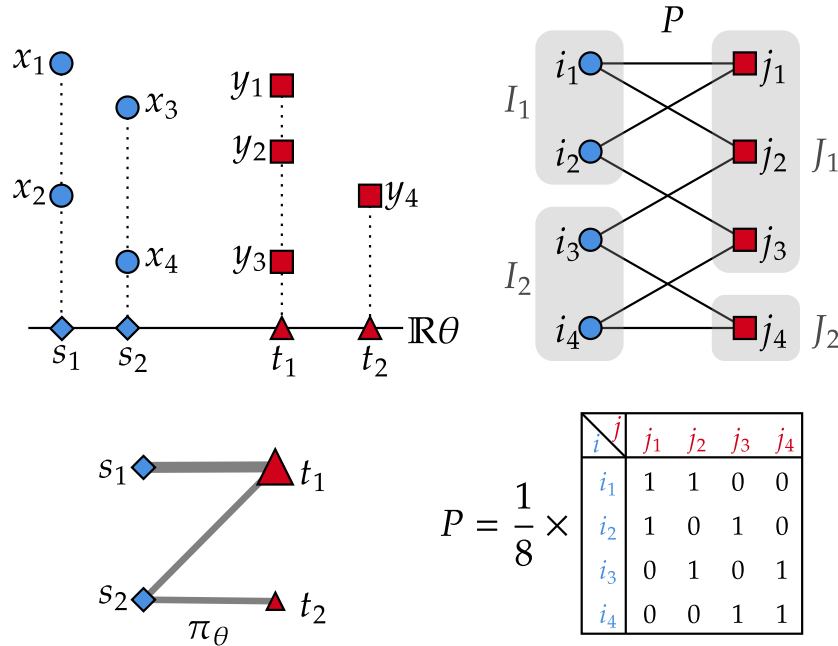


Figure 10: We continue with the example from Fig. 9 and illustrate the unique optimal transport plan $\pi_\theta = \frac{1}{2}\delta_{(s_1,t_1)} + \frac{1}{4}\delta_{(s_2,t_1)} + \frac{1}{4}\delta_{(s_2,t_2)}$ between $P_\theta\#\mu = \frac{1}{2}\delta_{s_1} + \frac{1}{2}\delta_{s_2}$ and $P_\theta\#\nu = \frac{3}{4}\delta_{t_1} + \frac{1}{4}\delta_{t_2}$. We show a particular transport plan $P \in \mathbb{U} \cap \mathcal{P}_\theta(X,Y)$ which is not a permutation matrix. For the constraints, notice for example that $\pi_\theta(\{(s_1,t_1)\}) = \frac{\#(I_1 \cap J_1)}{4} = \frac{1}{2} = \sum_{i=1}^2 \sum_{j=1}^3 P_{i,j}$.

The discrete problem in Eq. (40) can be seen as a constrained Kantorovich problem. Our goal is now to show that it admits a constrained Monge formulation, which is to say a minimisation over the constrained set of permutations $\mathfrak{S}_\theta(X,Y)$. To show this, we will adapt the proof of the Birkhoff Von Neumann Theorem [Bir46] (see also [Pey19] Theorem 2 and [Hur08] for other proofs which inspired

our method). Our objective is now to build up definitions and technical lemmas to adapt Birkhoff's Theorem, and prove the generalisation stated in Theorem 3. We will consider particular elements of $\mathbb{U}$ called permutation matrices:

$$\forall (\alpha, \beta) \in \mathfrak{S}_n^2, \ P^{\alpha, \beta} := \left[ \frac{1}{n} \mathbb{1}(\alpha(i) = \beta(j)) \right]_{i,j}. \tag{43}$$

This method of writing permutation matrices differs from the usual $P_{i,j}^\sigma := \frac{1}{n} \mathbb{1}(\sigma(i) = j)$, and will be more convenient for our purposes. An elementary property of permutation matrices is that:

$$\forall (\alpha, \beta, \varphi) \in \mathfrak{S}_n^3, \ P^{\alpha, \beta} = P^{\varphi \circ \alpha, \varphi \circ \beta}, \tag{44}$$

since $\varphi \circ \alpha(i) = \varphi \circ \beta(j) \iff \alpha(i) = \beta(j)$. For $S \subset \mathfrak{S}_n^2$, we will write $\mathcal{P}^S := \{P^{\alpha, \beta} : (\alpha, \beta) \in S\}$. The Birkhoff Von Neumann Theorem [Bir46] states that $\mathrm{Extr}\, \mathbb{U} = \mathcal{P}^{\mathfrak{S}_n^2}$, where the set of extreme points of a convex set is defined as:

> **Definition 5.** The set of extreme points $\mathrm{Extr}\, C$ of a convex set $C$ is the set of points $c \in C$ that cannot be written $c = \frac{1}{2}a + \frac{1}{2}b$ for some $a, b \in C$.

Our objective is to show that $\mathrm{Extr}(\mathbb{U} \cap \mathcal{P}_\theta(X, Y)) = \mathcal{P}^{\mathfrak{S}_\theta(X, Y)}$. We begin with a Lemma showing a condition for $P^{\alpha, \beta}$ to belong to $\mathcal{P}_\theta(X, Y)$.

> **Lemma 8.** Under Assumption 1, for any $(\alpha, \beta) \in \mathfrak{S}_n^2$, we have
>
> $$P^{\alpha, \beta} \in \mathcal{P}_\theta(X, Y) \iff \exists \varphi \in \mathfrak{S}_n : (\varphi \circ \alpha, \varphi \circ \beta) \in \mathfrak{S}_\theta(X, Y).$$
>
> In other words, $P^{\alpha, \beta} \in \mathcal{P}_\theta(X, Y)$ if and only if $P^{\alpha, \beta} = P^{\sigma, \tau}$ for some $(\sigma, \tau) \in \mathfrak{S}_\theta(X, Y)$.

*Proof.* Suppose that $P^{\alpha, \beta} \in \mathcal{P}_\theta(X, Y)$. Applying the definition of $\mathcal{P}_\theta(X, Y)$ from Eq. (42), we see that (using the notation of Lemma 7):

$$\forall (a, b) \in [\![1, A]\!] \times [\![1, B]\!], \quad \sum_{(i,j) \in I_a \times J_b} \frac{\mathbb{1}(\alpha(i) = \beta(j))}{n} = \frac{\#(I_a \cap J_b)}{n}, \text{ thus } \#(\alpha(I_a) \cap \beta(J_b)) = \#(I_a \cap J_b).$$

Let $E := \{(a, b) : I_a \cap J_b \neq \varnothing\}$. For any $(a, b) \in E$, we have $\#(\alpha(I_a) \cap \beta(J_b)) = \#(I_a \cap J_b)$, and thus we can introduce a bijection $\varphi_{a,b} : \alpha(I_a) \cap \beta(J_b) \longrightarrow I_a \cap J_b$. We have the partition $[\![1, n]\!] = \cup_{(a,b) \in E} I_a \cap J_b$ where the union is disjoint and the elements are non-empty. Since $\alpha, \beta$ are permutations and by the property $\#(\alpha(I_a) \cap \beta(J_b)) = \#(I_a \cap J_b)$, we have the partition $[\![1, n]\!] = \cup_{(a,b) \in E} \alpha(I_a) \cap \beta(J_b)$, again with disjoint unions and non-empty terms. We can define $\psi : [\![1, n]\!] \longrightarrow E$ a map such that $\forall i \in [\![1, n]\!]$, $i \in \alpha(I_a) \cap \beta(J_b)$ where $\psi(i) = (a, b)$. The map $\varphi := i \longmapsto \varphi_{\psi(i)}(i)$ is therefore well-defined, we verify easily that it is a permutation of $[\![1, n]\!]$ using the partition $[\![1, n]\!] = \cup_{(a,b) \in E} \alpha(I_a) \cap \beta(J_b)$.

We now fix $a \in [\![1, A]\!]$ and show that $\varphi \circ \alpha(I_a) = I_a$. Let $i \in I_a$, we have $\alpha(i) \in \alpha(I_a)$, and there exists (a unique) $b \in [\![1, B]\!]$ such that $\alpha(i) \in \alpha(I_a) \cap \beta(J_b)$. By definition, we get $\psi(\alpha(i)) = (a, b)$, and thus $\varphi(\alpha(i)) = \varphi_{a,b}(\alpha(i)) \in I_a \cap J_b \subset I_a$. We conclude that $\varphi \circ \alpha(I_a) \subset I_a$ and similarly that $\varphi \circ \beta(J_b) \subset J_b$ for any $b \in [\![1, B]\!]$. By Lemma 7, we conclude that $(\varphi \circ \alpha, \varphi \circ \beta) \in \mathfrak{S}_\theta(X, Y)$, concluding the "left to right" implication.

Conversely, let $\varphi \in \mathfrak{S}_n$ and $(\sigma, \tau) \in \mathfrak{S}_\theta(X, Y)$. Notice that $P^{\sigma, \tau} = P^{\varphi \circ \sigma, \varphi \circ \tau}$ by Eq. (44). We check that $P^{\sigma, \tau} \in \mathcal{P}_\theta(X, Y)$ by applying the definition: let $(a, b) \in [\![1, A]\!] \times [\![1, B]\!]$, we have:

$$\sum_{(i,j) \in I_a \times J_b} P_{i,j}^{\sigma, \tau} = \frac{\#(\sigma(I_a) \cap \tau(J_b))}{n} = \frac{\#(I_a \cap J_b)}{n},$$

where we used that $\sigma(I_a) = I_a$ and $\tau(J_b) = J_b$, which is a consequence of Lemma 7. $\qquad \square$

## 5.4 Technical Lemmas on Bipartite Graphs Associated to Couplings

To study the extreme points of $\mathbb{U} \cap \mathcal{P}_\theta(X, Y)$ we will adapt the techniques from [Pey19; Hur08] and consider the bipartite graph associated to a matrix in $P \in \mathbb{R}_+^{n \times m}$, which we define in Definition 6.

> **Definition 6.** The bipartite directed graph $G_P$ associated to a matrix $P \in \mathbb{R}_+^{n \times m}$ is the graph with vertices $V_P := \{i_k, \ k \in [\![1, n]\!]\} \cup \{j_l, \ \ell \in [\![1, m]\!]\}$ and directed edges:
>
> $$E_P := \{(i_k, j_l), \ (k, \ell) \in [\![1, n]\!] \times [\![1, m]\!], \ : P_{i_k, j_l} > 0\} \cup \{(j_l, i_k), \ (k, \ell) \in [\![1, n]\!] \times [\![1, m]\!], \ P_{i_k, j_l} > 0\}.$$

By slight abuse of notation, we will often write $\{i_k, \ k \in [\![1, n]\!]\}$ as simply $[\![1, n]\!]$ and $\{j_l, \ \ell \in [\![1, m]\!]\}$ as $[\![1, m]\!]$, seeing them as disjoint sets of labels. The $i$'s will be called "left" vertices, and the $j$'s "right" vertices. Edges $(i, j)$ being directed from left to right, we will call them "right" edges, and likewise edges $(j, i)$ will be referred to as "left" edges. We continue the example from Fig. 10 in Fig. 11 showing the bipartite graph associated to the matrix $P$.
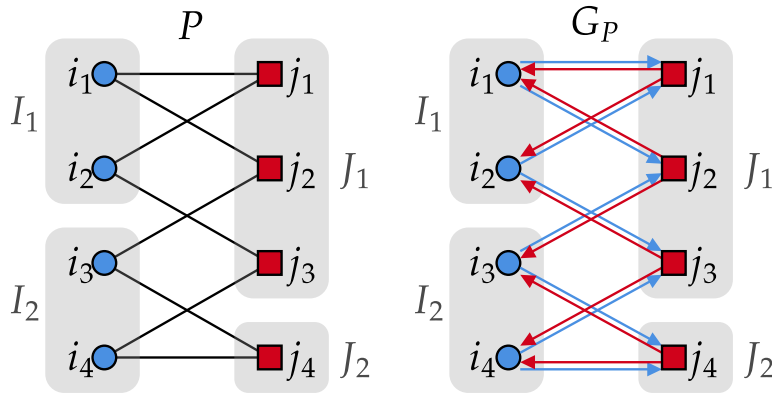


Figure 11: We consider the matrix $P$ from Fig. 10 and show the associated bipartite graph $G_P$. The "right" edges from an $i \in [\![1, n]\!]$ on the left to a $j \in [\![1, n]\!]$ on the right are represented in blue, and the "left edges" are represented in red. Note that by construction, for each $(i, j)$ such that $P_{i,j} > 0$, there is both a left edge $(i, j)$ and a right edge $(j, i)$ in $G_P$.

In the following, we will extract a particular cycle $i_1 \to j_1 \to i_2 \to \cdots \to i_{p+1} = i_1$ from the graph $G_P$ of an element $P \in \mathbb{U} \setminus \mathcal{P}^{\mathfrak{S}_n^2}$. In proofs of Birkhoff's theorem, this is commonly used to show that $P$ is not an extreme point of $\mathbb{U}$. In our setting, we will also make use of this property, in addition to strategies specific to $\mathcal{P}_\theta(X, Y)$.

> **Lemma 9.** Assume $n \geq 2$ and let $P \in \mathbb{U} \setminus \mathcal{P}^{\mathfrak{S}_n^2}$.
> Then there exists a cycle $(i_1, j_1, \cdots, i_p, j_p, i_{p+1}) \in [\![1, n]\!]^{2p+1}$ in $G_P$ with $p \geq 2$ verifying:
>
> $$\begin{aligned} &i_{p+1} = i_1; \ (i_k)_{k=1}^p \text{ and } (j_k)_{k=1}^p \text{ are injective;} \\ &\text{and } \forall k \in [\![1, p]\!], \ P_{i_k, j_k} \in (0, \tfrac{1}{n}), \ P_{i_{k+1}, j_k} \in (0, \tfrac{1}{n}). \end{aligned} \tag{45}$$

*Proof.* First, we show a weaker result:

$$\begin{aligned} &\exists p \geq 2, \ \exists (i_1, j_1, \cdots, i_p, j_p, i_{p+1}) \in [\![1, n]\!]^{2p+1} \\ &\text{such that } i_{p+1} = i_1; \ \forall k \in [\![1, p]\!], \ i_k \neq i_{k+1}, \ \forall k \in [\![1, p-1]\!], \ j_k \neq j_{k+1}; \\ &\text{and } \forall k \in [\![1, p]\!], \ P_{i_k, j_k} \in (0, \tfrac{1}{n}), \ P_{i_{k+1}, j_k} \in (0, \tfrac{1}{n}). \end{aligned} \tag{46}$$

Since $P \in \mathbb{U} \setminus \mathcal{P}^{\mathfrak{S}_n^2}$, there exists $(i_1, j_1) \in [\![1, n]\!]^2$ such that $P_{i_1, j_1} \in (0, \tfrac{1}{n})$. Since $0 < P_{i_1, j_1} < \sum_i P_{i, j_1} = \tfrac{1}{n}$, there exists $i_2 \neq i_1$ such that $P_{i_2, j_1} \in (0, \tfrac{1}{n})$. Likewise, since $0 < P_{i_2, j_1} < \sum_j P_{i_2, j} = \tfrac{1}{n}$, there exists $j_2 \neq j_1$ such that $P_{i_2, j_2} \in (0, \tfrac{1}{n})$. We continue and show the existence of $i_3 \neq i_2$ such

that $P_{i_3,j_2} \in (0, \frac{1}{n})$. So far, we have built a chain $i_1 \to j_1 \to i_2 \to j_2 \to i_3$. If $i_3 = i_1$ then we have shown Eq. (46). Otherwise we continue the process up to $i_k$, $k \geq 4$ while $i_k \neq i_1$, and there are two exclusive possibilities:

1. The process terminates with $(i_1, j_1, \cdots, i_p, j_p, i_{p+1})$ such that $i_{p+1} = i_1$, and by construction the cycle verifies the conditions of Eq. (46);

2. The process continues at least up to $k = n + 1$, yielding $(i_1, j_1, \cdots, i_n, j_n, i_{n+1})$ verifying the conditions of Eq. (46) except $i_{n+1} \neq i_1$. Then by the pigeonhole principle, there exists $k_1 < k_2 \in [\![1, n+1]\!]^2$ such that $i_{k_1} = i_{k_2}$. Consider the cycle $(i_{k_1}, j_{k_1}, i_{k_1+1}, j_{k_1+1}, \cdots, i_{k_2})$, it verifies Eq. (46) (the length is sufficient since by construction $i_{k_1} \neq i_{k_1+1}$).

Now that we have shown Eq. (46), we deduce Eq. (45) by taking $p \geq 2$ minimal in Eq. (46).     $\square$

As an illustration, in Fig. 11, by following the edges of $G_P$ starting from the edge $(i_1, j_1)$, we observe the cycle $(i_1, j_1, i_2, j_3, i_4, j_4, i_3, j_2, i_1)$ which satisfies the criteria of Eq. (45).

We will also require the following technical result about extracting injective cycles from (possibly) redundant cycles in a graph. For a set $S$ and $n, m \in \mathbb{N}$, we say that two families $(s_i)_{i=1}^n \in S^n$ and $(t_j)_{j=1}^n \in S^n$ are equipotent if $n = m$ and there exists a permutation $\varphi \in \mathfrak{S}_n$ such that $\forall i \in [\![1, n]\!]$, $s_{\varphi(i)} = t_i$. We write this property $(s_i) \simeq (t_j)$. This concept is particularly useful when the families are not injective, which will sometimes be the case in the following.

> **Lemma 10.** Let $G := (V := \mathcal{A} \cup \mathcal{B}, E)$ be a directed bipartite graph, set $p \geq 1$ and consider a cycle written $(a_1, b_1, \cdots, a_p, b_p, a_{p+1}) \in V^{2p+1}$ with $\forall k \in [\![1, p]\!]$, $(a_k, b_k) \in E$, $(b_k, a_{k+1}) \in E$. Then there exists $L \geq 1$ cycles of $G$ of the form $(a_1^\ell, b_1^\ell, \cdots, a_{p_\ell}^\ell, b_{p_\ell}^\ell, a_{p_\ell+1}^\ell)$ (with $a_1^\ell = a_{p_\ell+1}^\ell$ and each $(a_k^\ell, b_k^\ell), (b_k^\ell, a_{k+1}^\ell) \in E$) whose combined elements (without the last vertex) are exactly the elements of $(a_1, b_1, \cdots, a_p, b_p)$:
>
> $$(a_1, b_1, \cdots, a_p, b_p) \simeq (a_1^1, b_1^1, \cdots, a_{p_1}^1, b_{p_1}^1, \cdots\cdots, a_1^L, b_1^L, \cdots, a_{p_L}^L, b_{p_L}^L), \qquad (47)$$
>
> and such that for each $\ell \in [\![1, L]\!]$, the families of edges $((a_k^\ell, b_k^\ell))_{k=1}^{p_\ell}$ and $((b_k^\ell, a_{k+1}^\ell))_{k=1}^{p}$ are injective.

*Proof.* Given such a cycle $\mathcal{C} := (a_1, b_1, \cdots, a_p, b_p, a_{p+1})$ we consider the two following "splitting" operators:

- $\mathrm{Split}_R$ takes the first pair $i < j \in [\![1, p]\!]^2$ (for the lexicographic order) such that $(a_i, b_i) = (a_j, b_j)$ if such a pair $(i, j)$ exists (if not, $\mathrm{Split}_R(\mathcal{C})$ returns $\mathcal{C}$). $\mathrm{Split}_R(\mathcal{C})$ then returns the two following sub-cycles:

$$\mathcal{C}_1 := (a_1, b_1, \cdots, a_{i-1}, b_{i-1}, a_j, b_j, \cdots, a_p, b_p, a_{p+1}), \quad \mathcal{C}_2 := (a_i, b_i, \cdots, , a_{j-1}, b_{j-1}, a_j).$$

Obviously, their concatenation without endpoints is exactly $\mathcal{C}$ without its endpoint:

$$(a_1, b_1, \cdots, a_{i-1}, b_{i-1}, a_j, b_j, \cdots, a_p, b_p, a_i, b_i, \cdots, , a_{j-1}, b_{j-1}) \simeq (a_1, b_1, \cdots, a_p, b_p).$$

- $\mathrm{Split}_L$ takes the first pair $i < j \in [\![1, p]\!]^2$ such that $(b_i, a_{i+1}) = (b_j, a_{j+1})$ if such a pair $(i, j)$ exists (if not, $\mathrm{Split}_L(\mathcal{C})$ returns $\mathcal{C}$). $\mathrm{Split}_L(\mathcal{C})$ then returns the two following sub-cycles, (which verify the equipotence condition):

$$\mathcal{C}_1 := (a_1, b_1, \cdots, a_i, b_i, a_{j+1}, b_{j+1}, \cdots, a_p, b_p, a_{p+1}), \quad \mathcal{C}_2 := (a_{i+1}, b_{i+1}, \cdots, a_j, b_j, a_{j+1}).$$

To split an initial $\mathcal{C}$, we construct a family $(\mathcal{C}_\ell)$ of cycles iteratively starting with $(\mathcal{C})$ by applying $\mathrm{Split}_R$ and $\mathrm{Split}_L$ to the cycles to the family $\mathcal{C}_\ell$ until no cycle can be split. This process terminates since each iteration increases the number of cycles (they are non-empty), which is bounded because $\mathcal{C}$

is finite and the concatenation of the cycles $(\mathcal{C}_\ell)$ without endpoints is exactly $\mathcal{C}$ without its endpoint. At the end of the process, the equipotence condition remains and each cycle $\mathcal{C}_\ell$ has injective edges $((a_k^\ell, b_k^\ell))_k, ((b_k^\ell, a_{k+1}^\ell))_k$ since the splitting process could not continue. □

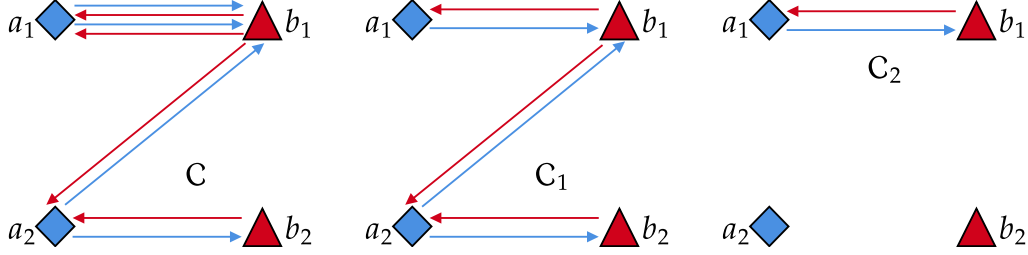In Fig. 12 we illustrate the splitting process of Lemma 10.



Figure 12: Extracting two cycles $\mathcal{C}_1, \mathcal{C}_2$ from the cycle $\mathcal{C}$ such that each cycle $\mathcal{C}_\ell$ has distinct (directed) edges.

The cycle from Fig. 12 is a cycle of $G_{\overline{P}}$ (where $\overline{P}$) is the OT plan matrix between the measures $P_\theta \# \mu$ and $P_\theta \# \nu$, constructed using $P$ from Fig. 11 with $a_1 \in [\![1, A]\!]$ where $i \in I_{a_1}$, then $b_1 \in [\![1, B]\!]$ such that $j_1 \in I_{b_1}$ and so one. This example case is paramount since it will be the use case of Lemma 10 in the proof of Theorem 3.

The following lemma is in essence a cyclical monotonicity property, and concerns a property of cycles in the bipartite graph associated to the matrix $\overline{P} \in \mathbb{R}_+^{A \times B}$ which is the unique optimal transport plan matrix between the one-dimensional measures $P_\theta \# \mu = \sum_{a=1}^A \frac{\# I_a}{n} \delta_{s_a}$ and $P_\theta \# \nu = \sum_{b=1}^B \frac{\# J_b}{n} \delta_{t_b}$. The idea is that by monotonicity of $\overline{P}$, no edges of $G_{\overline{P}}$ can cross one another, which constrains cycles to have a left edge $(b, a)$ corresponding to each right edge $(a, b)$. We remind that by assumption $s_1 < \cdots < s_A$ and $t_1 < \cdots < t_B$ (the notation was introduced in Lemma 7).

**Lemma 11.** Let $\overline{P} \in \mathbb{R}_+^{A \times B}$ be the OT matrix between $\sum_{a=1}^A \frac{\# I_a}{n} \delta_{s_a}$ and $\sum_{b=1}^B \frac{\# J_b}{n} \delta_{t_b}$. If $\mathcal{C} := (a_1, b_1, \cdots, a_p, b_p, a_{p+1})$ is a cycle in $G_{\overline{P}}$ (i.e. $(a_k)_{k=1}^{p+1} \in [\![1, A]\!]^{p+1}$, $(b_k)_{k=1}^p \in [\![1, B]\!]^p$, $a_{p+1} = a_1$ and $\forall k \in [\![1, p]\!]$, $\overline{P}_{a_k, b_k} > 0$, $\overline{P}_{a_{k+1}, b_k} > 0$) such that the families of edges $((a_k, b_k))_{k=1}^p$ and $((b_k, a_{k+1}))_{k=1}^p$ are injective, then $((a_k, b_k))_{k=1}^p \simeq ((a_{k+1}, b_k))_{k=1}^p$.

*Proof.* First, by optimality of $\overline{P}$, by [San15] Lemma 2.8, $\overline{P}$ is monotone in the sense that:

$$\forall (a, b), (a', b') \in [\![1, A]\!] \times [\![1, B]\!], \text{ such that } \overline{P}_{a,b} > 0, \overline{P}_{a',b'} > 0, \ a < a' \implies b \leq b'.$$

Note that the contrapositive yields the symmetrical property that if $b < b'$ then $a \leq a'$. Furthermore, we remind that since each $(a_k, b_k)$ and $(b_k, a_{k+1})$ are edges of the graph $G_{\overline{P}}$, we have $\overline{P}_{a_k, b_k} > 0$ and $\overline{P}_{a_{k+1}, b_k} > 0$. We can understand the monotonicity property as the fact that the edges of the cycle cannot cross one another.

By injectivity of the edge families, to show the equipotence result, it suffices to show that $\forall k \in [\![1, p]\!]$, $\exists k' \in [\![1, p]\!]$ such that $(a_k, b_k) = (a_{k'+1}, b_{k'})$. Since the vertices $a_k$ and $b_k$ are part of the cycle $a_1 \to b_1 \to a_2 \to \cdots \to a_{p+1} = a_1$, there exists a sub-cycle $b_k \to a_1' \to b_1' \to \cdots \to b_q' \to a_k$, which is to say that there exists, for some $q \geq 0$, $(b_k, a_1') \in \mathcal{C}$, $\forall k' \in [\![1, q]\!]$, $(a_{k'}', b_{k'}') \in \mathcal{C}$, $(b_{k'}', a_{k'+1}') \in \mathcal{C}$ (writing $a_{q+1}' := a_k$), and we now take $q \geq 0$ minimal. We will show that $q = 0$ by contradiction: assume $q \geq 1$, which implies that $a_1' \neq a_k$ by minimality. Assume that $a_1' < a_k$ (the case $a_1' > a_k$ is analogous). By monotonicity of $\overline{P}$, we deduce $b_1' \leq b_k$, and even $b_1' < b_k$ since $b_1' = b_k$ would violate the minimality of $q$. By monotonicity, we deduce that $a_2' \leq a_1'$ and again, even $a_2' < a_1'$ by minimality of $a$. Continuing this process we find that $a_{q+1}' < a_k$ contradicting $a_{q+1}' = a_k$. We conclude that the edge $(b_k, a_k)$ belongs to the cycle, which is to say that there exists $k' \in [\![1, p]\!]$ such that $(a_k, b_k) = (a_{k'+1}, b_{k'})$, finishing the proof. □

In Fig. 13 we illustrate the result of Lemma 11 in the use case of the proof of Theorem 3, regrouping the continued example from Figs. 9 to 12.
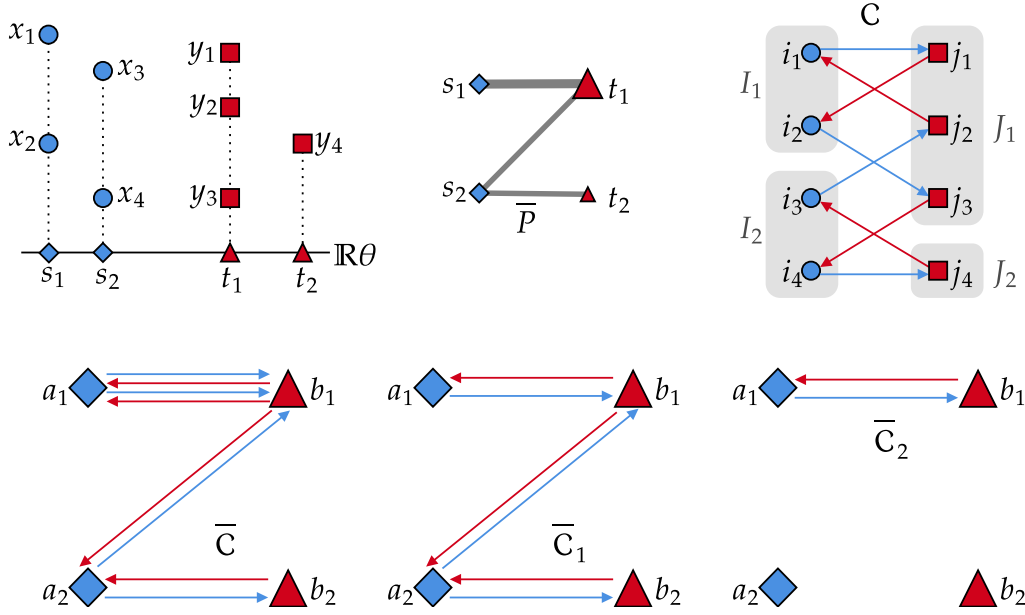


Figure 13: We take two discrete uniform measures $\mu := \frac{1}{n}\sum_i \delta_{x_i}$ and $\nu := \frac{1}{n}\sum_j \delta_{y_j}$ such that $s_1 := P_\theta x_1 = P_\theta x_2 < s_2 := P_\theta x_3 = P_\theta x_4$ and $t_1 := P_\theta y_1 = P_\theta y_2 = P_\theta y_3 < t_2 := P_\theta y_4$. We represent the OT matrix $\overline{P}$ between the measures $P_\theta\#\mu$ and $P_\theta\#\nu$ and consider the bipartite graph $G_P$ associated to a coupling $P \in \mathbb{U} \cap \mathcal{P}_\theta(X,Y)$ (presented in Fig. 11). In this case the graph $G_P$ (top-right) contains the cycle $\mathcal{C} := (i_1, j_1, i_2, j_3, i_4, j_4, i_3, j_2, i_1)$. We consider for each $k$ the "right" edge $(a_k, b_k)$ in $G_{\overline{P}}$ such that $i_k \in I_{a_k}$ and $j_k \in J_{b_k}$, and the "left" edge $(b_k, a_{k+1})$ such that $j_k \in J_{b_k}$ and $i_{k+1} \in I_{a_{k+1}}$. This defines the cycle $\overline{\mathcal{C}}$ in $G_{\overline{P}}$, that we decompose into cycles with distinct edges $(\overline{\mathcal{C}}_\ell)$ using Lemma 10. Lemma 11 then applies to each $\overline{\mathcal{C}}_\ell$ and we observe indeed that in $\overline{\mathcal{C}}_\ell$, each "left" edge $(b, a)$ has a corresponding "right" edge $(a, b)$ in the cycle.

## 5.5   A Constrained Version of the Birkhoff von Neumann Theorem

We are now ready to prove a constrained version of the Birkhoff von Neumann Theorem [Bir46]. We remind that $\mathcal{P}_\theta(X,Y)$ is defined in Eq. (42), $\mathbb{U}$ in Eq. (41) and $\mathcal{P}^{\mathfrak{S}_\theta(X,Y)} = \{P^{\sigma,\tau}, \ (\sigma,\tau) \in \mathfrak{S}_\theta(X,Y)\}$, with $P^{\sigma,\tau}$ the permutation matrix introduced in Eq. (43) and $\mathfrak{S}_\theta(X,Y)$ defined in Eq. (36). Finally, the notion of extreme points is defined in Definition 5.

> **Theorem 3.** Let $(X,Y) \in \mathbb{R}^{n\times d}$ and $\theta \in \mathbb{S}^{d-1}$ verifying Assumption 1. Then
>
> $$\mathrm{Extr}(\mathbb{U} \cap \mathcal{P}_\theta(X,Y)) = \mathcal{P}^{\mathfrak{S}_\theta(X,Y)}.$$

*Proof.* — *Step 1*: $\mathcal{P}^{\mathfrak{S}_\theta(X,Y)} \subset \mathrm{Extr}(\mathbb{U} \cap \mathcal{P}_\theta(X,Y))$

First, for $(\sigma,\tau) \in \mathfrak{S}_\theta(X,Y)$, we have $P^{\sigma,\tau} \in \mathcal{P}_\theta(X,Y)$ by Lemma 8, which shows that $P^{\sigma,\tau} \in \mathbb{U} \cap \mathcal{P}_\theta(X,Y)$. Now if $P^{\sigma,\tau} = \frac{1}{2}Q + \frac{1}{2}R$ for some $(Q,R) \in \mathbb{U} \cap \mathcal{P}_\theta(X,Y)$, then for any $(i,j) \in [\![1,n]\!]^2$, we have $P^{\sigma,\tau}_{i,j} \in \{0, \frac{1}{n}\}$, thus $\frac{1}{2}Q_{i,j} + \frac{1}{2}R_{i,j} = P^{\sigma,\tau}_{i,j}$ implies that $Q_{i,j} = R_{i,j}$ since $Q_{i,j}$ and $R_{i,j}$ are both in $[0, \frac{1}{n}]$ (since they belong to $\mathbb{U}$). This shows that $P^{\sigma,\tau} \in \mathrm{Extr}(\mathbb{U} \cap \mathcal{P}_\theta(X,Y))$.

— *Step 2*: Writing $P \in (\mathbb{U} \cap \mathcal{P}_\theta(X,Y)) \setminus \mathcal{P}^{\mathfrak{S}_\theta(X,Y)}$ as $P = (Q+R)/2$ with $Q, R \in [0, \frac{1}{n}]^{n\times n}$

To show $\mathrm{Extr}(\mathbb{U} \cap \mathcal{P}_\theta(X,Y)) \subset \mathcal{P}^{\mathfrak{S}_\theta(X,Y)}$, we will show that

$$(\mathbb{U} \cap \mathcal{P}_\theta(X,Y)) \setminus \mathcal{P}^{\mathfrak{S}_\theta(X,Y)} \subset (\mathbb{U} \cap \mathcal{P}_\theta(X,Y)) \setminus \mathrm{Extr}(\mathbb{U} \cap \mathcal{P}_\theta(X,Y)).$$

Note that when $n = 1$ the entire Theorem is trivial, and in the following we assume $n \geq 2$. We take $P \in (\mathbb{U} \cap \mathcal{P}_\theta(X,Y)) \setminus \mathcal{P}^{\mathfrak{S}_\theta(X,Y)}$ and apply Lemma 9, allowing us to introduce $(i_1, j_1, \cdots, i_p, j_p, i_{p+1}) \in \llbracket 1, n \rrbracket^{2p+1}$ such that $i_{p+1} = i_1$, the families $(i_k)_{k=1}^p$ and $(j_k)_{k=1}^p$ are injective and $\forall k \in \llbracket 1, p \rrbracket$, $P_{i_k, j_k} \in (0, \frac{1}{n})$, $P_{i_{k+1}, j_k} \in (0, \frac{1}{n})$. We consider the set of "right edges" $E_R := ((i_k, j_k))_{k=1}^p$ and "left edges": $E_L := ((i_{k+1}, j_k))_{k=1}^p$. By injectivity, we have $E_R \cap E_L = \varnothing$ and that $E_R$ and $E_L$ are themselves injective. Note that our cycle construction is illustrated on an example in Fig. 13. We take the smallest margin $\varepsilon > 0$ that $P$ has to be in $\mathbb{U}$ (within the cycle):

$$\varepsilon := \min_{k \in \llbracket 1, p \rrbracket} \{ P_{i_k, j_k}, 1 - P_{i_k, j_k}, P_{i_{k+1}, j_k}, 1 - P_{i_{k+1}, j_k} \} \in (0, \tfrac{1}{n}),$$

and introduce the matrices $Q, R \in \mathbb{R}^{n \times n}$ defined as, for $(i,j) \in \llbracket 1, n \rrbracket^2$:

$$Q_{i,j} := \begin{cases} P_{i,j} & \text{if} \quad (i,j) \notin E_R \cup E_L \\ P_{i,j} + \varepsilon & \text{if} \quad (i,j) \in E_R \\ P_{i,j} - \varepsilon & \text{if} \quad (i,j) \in E_L \end{cases} \quad, \quad R_{i,j} := \begin{cases} P_{i,j} & \text{if} \quad (i,j) \notin E_R \cup E_L \\ P_{i,j} - \varepsilon & \text{if} \quad (i,j) \in E_R \\ P_{i,j} + \varepsilon & \text{if} \quad (i,j) \in E_L \end{cases} \quad.$$

For visualisation purposes, in the example of Fig. 13, we represent "left" edges of the cycle in blue (on these edges, we add $+\varepsilon$ in $Q$ and $-\varepsilon$ in $R$) and "right" edges in red (on which we do the opposite). By definition of $\varepsilon$, we have $Q, R \in [0, \frac{1}{n}]^{n \times n}$. By construction, we also have $P = \frac{1}{2}(Q + R)$.

— *Step 3*: Showing that $Q, R \in \mathbb{U}$

Fix $j \in \llbracket 1, n \rrbracket$, we show that $\sum_i Q_{i,j} = \frac{1}{n}$. Since $E_L$ and $E_R$ are injective and disjoint, we compute

$$\sum_{i=1}^n Q_{i,j} = \sum_{i:(i,j)\notin E_R \cup E_L} P_{i,j} + \sum_{i:(i,j)\in E_R} (P_{i,j} + \varepsilon) + \sum_{i:(i,j)\in E_L} (P_{i,j} - \varepsilon) = \frac{1}{n} + \varepsilon(\#I_R(j) - \#I_L(j)),$$

where $I_R(j) := \{i \in \llbracket 1, n \rrbracket : (i,j) \in E_R\}$ and $I_L(j) := \{i \in \llbracket 1, n \rrbracket : (i,j) \in E_L\}$. Since $E_R$ and $E_L$ are injective, we deduce that $I_R$ and $I_L$ are also injective. Take $i \in I_R(j)$ and write $(i,j) = (i_k, j_k) \in E_L$ for some $k \in \llbracket 1, p \rrbracket$. We notice that $(i_k, j_{k-1}) \in E_R$ where if $k = 1$ we write $j_{k-1} := j_p$. We conclude that $\#I_R(j) = \#I_L(j)$ and thus that $\sum_i Q_{i,j} = \frac{1}{n}$. The same reasoning shows that $\sum_j Q_{i,j} = \frac{1}{n}$ for all $i \in \llbracket 1, n \rrbracket$, and we conclude that $Q \in \mathbb{U}$. The same computations show that $R \in \mathbb{U}$ as well.

— *Step 4*: Showing that $Q, R \in \mathcal{P}_\theta(X,Y)$

We now show that $Q, R \in \mathcal{P}_\theta(X, Y)$ using the definition (Eq. (42)). Take $(a, b) \in \llbracket 1, A \rrbracket \times \llbracket 1, B \rrbracket$. We have:

$$\sum_{(i,j) \in I_a \times J_b} Q_{i,j} = \sum_{(i,j) \in (I_a \times J_b) \cap E_R^c \cap E_L^c} P_{i,j} + \sum_{(i,j) \in (I_a \times J_b) \cap E_R} (P_{i,j} + \varepsilon) + \sum_{(i,j) \in (I_a \times J_b) \cap E_L} (P_{i,j} - \varepsilon)$$

$$= \frac{\#I_a \cap J_b}{n} + \varepsilon \left( \#((I_a \times J_b) \cap E_R) - \#((I_a \times J_b) \cap E_L) \right). \tag{48}$$

Let $\overline{P} \in \mathbb{R}_+^{A \times B}$ be the OT matrix between $\sum_{a=1}^A \frac{\#I_a}{n} \delta_{s_a}$ and $\sum_{b=1}^B \frac{\#J_b}{n} \delta_{t_b}$. Consider the family $\overline{\mathcal{C}} := (a_1, b_1, \cdots, a_p, b_p, a_{p+1})$ defined by the condition $\forall k \in \llbracket 1, p \rrbracket$, $i_k \in I_{a_k}$, $j_k \in J_{b_k}$ and $a_{p+1} := a_1$. Since $\mathcal{C} := (i_1, i_2, \cdots, i_p, j_p, i_{p+1})$ is a cycle in $G_P$, it follows that $\overline{\mathcal{C}}$ is a cycle in $G_{\overline{P}}$ since the condition $P \in \mathcal{P}_\theta(X, Y)$ implies:

$$\forall (a,b) \in \llbracket 1, A \rrbracket \times \llbracket 1, B \rrbracket, \quad \sum_{(i,j) \in I_a \times J_b} P_{i,j} = \overline{P}_{a,b},$$

thus if $P_{i,j} > 0$ for some $(i,j) \in I_a \times J_b$ then $P_{a,b} > 0$. See also Fig. 13 for an example. We now apply Lemma 10 to show that $\overline{\mathcal{C}}$ is the "concatenation" of $L \geq 1$ cycles $\overline{\mathcal{C}}_\ell$ of $G_{\overline{P}}$ of the form:

$$\overline{\mathcal{C}}_\ell := (a_1^\ell, b_1^\ell, \cdots, a_{p_\ell}^\ell, b_{p_\ell}^\ell, a_{p_\ell+1}^\ell),$$

where "concatenation" means that Eq. (47) holds with the same notation, and where each $\overline{\mathcal{C}}_\ell$ is such that the edge families $((a_k^\ell, b_k^\ell))_{k=1}^{p_\ell}$ and $((b_k^\ell, a_{k+1}^\ell))_{k=1}^{p_\ell}$ are injective. For each $\ell \in \llbracket 1, L \rrbracket$, we apply Lemma 11, which shows in particular that for any $(a, b) \in \llbracket 1, A \rrbracket \times \llbracket 1, B \rrbracket$:

$$\# \left\{ k \in \llbracket 1, p_\ell \rrbracket : (a_k^\ell, b_k^\ell) = (a, b) \right\} = \# \left\{ k \in \llbracket 1, p_\ell \rrbracket : (a_{k+1}^\ell, b_k^\ell) = (a, b) \right\}. \tag{49}$$

We understand Eq. (49) as the fact that for any left edge from group $a$ to group $b$ in $\mathcal{C}_\ell$, there corresponds exactly as many right edges from group $b$ to group $a$. This will allow us to show that the terms in $+\varepsilon$ and $-\varepsilon$ are in the same number. We now re-write the sets from the condition on $Q$ (Eq. (48)) for a fixed $(a,b) \in [\![1, A]\!] \times [\![1, B]\!]$:

$$
\begin{aligned}
\# \left( (I_a \times J_b) \cap E_R \right) &= \{ (i_k, j_k),\ k \in [\![1, p]\!],\ (i_k, j_k) \in I_a \times J_b \} \\
&= \# \left( (a_k, b_k),\ k \in [\![1, p]\!],\ (a_k, b_k) = (a, b) \right) \\
&= \sum_{\ell=1}^{L} \# \left( (a_k^\ell, b_k^\ell),\ k \in [\![1, p_\ell]\!],\ (a_k^\ell, b_k^\ell) = (a, b) \right) \\
&= \sum_{\ell=1}^{L} \# \left( (b_k^\ell, a_{k+1}^\ell),\ k \in [\![1, p_\ell]\!],\ (a_{k+1}^\ell, b_k^\ell) = (a, b) \right) \\
&= \# \left( (I_a \times J_b) \cap E_L \right),
\end{aligned}
$$

where the first equality uses the definition of $E_R$, the second inequality comes from associating to each pair $(i_k, j_k)$ its group pair $(a_k, b_k)$ and counting the group pairs *with repetition*, the third equality the concatenation property of the cycles $\mathcal{C}_\ell$ (Eq. (47)), the fourth equality from Eq. (49) and the last inequality from the definition of $E_L$ (doing the same computations as for $E_R$ in reverse order).

Combining with Eq. (48) shows that $\sum_{(i,j) \in I_a \times J_b} Q_{i,j} = \frac{\#(I_a \cap J_b)}{n}$ and thus that $Q \in \mathcal{P}_\theta(X, Y)$. Likewise we show $R \in \mathcal{P}_\theta(X, Y)$ and thus we have found $Q, R \in \mathbb{U} \cap \mathcal{P}_\theta(X, Y)$ such that $P = \frac{1}{2}(Q+R)$, and we conclude that $P$ does not belong to $\mathrm{Extr}(\mathbb{U} \cap \mathcal{P}_\theta(X, Y))$, finishing the proof. □

From Theorem 3 we deduce the following theorem, which is a Monge formulation of the constrained Kantorovich problem in $\mathrm{CW}_\theta$:

---

**Theorem 4.** Let $(X, Y) \in \mathbb{R}^{n \times d}$ and $\theta \in \mathbb{S}^{d-1}$, we have:

$$
\mathrm{CW}_\theta^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \frac{1}{n} \sum_{j=1}^n \delta_{y_j} \right) = \min_{(\sigma, \tau) \in \mathfrak{S}_\theta(X,Y)} \frac{1}{n} \sum_{i=1}^n \| x_{\sigma(i)} - y_{\tau(i)} \|_2^2, \tag{50}
$$

---

*Proof.* Beginning under Assumption 1, we combine Theorem 3 with the expression of $\mathrm{CW}_\theta^2$ from Proposition 9:

$$
\mathrm{CW}_\theta^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \frac{1}{n} \sum_{j=1}^n \delta_{y_j} \right) = \min_{P \in \mathbb{U} \cap \mathcal{P}_\theta(X,Y)} \sum_{i,j} \| x_i - y_j \|_2^2 P_{i,j} = \min_{P \in \mathrm{Extr}(\mathbb{U} \cap \mathcal{P}_\theta(X,Y))} \sum_{i,j} \| x_i - y_j \|_2^2 P_{i,j},
$$

since the solution of a linear program over a non-empty convex compact set is attained at an extreme point ([BT97] Theorem 2.7), and we conclude that the expression in Eq. (50) holds thanks to Theorem 3. For the general case without Assumption 1, we use Lemma 6. □

# 6   Min-Pivot Sliced

## 6.1   Min-Pivot Sliced Discrepancy: Definition

A specificity of the Pivot Sliced Wasserstein discrepancy is the dependence on the axis $\theta \in \mathbb{S}^{d-1}$, which can overly constrain the choice of transport plans. In this section, we study the Min-Sliced min-pivot sliced Discrepancy which minimises $\mathrm{PS}_\theta$ over $\theta \in \mathbb{S}^{d-1}$. This object was first introduced in [Mah+23] on the set of discrete uniform measures with $n$ points.

$$
\min \mathrm{PS}^2(\mu_1, \mu_2) := \min_{\theta \in \mathbb{S}^{d-1}} \mathrm{PS}_\theta^2(\mu_1, \mu_2) = \min_{\substack{\theta \in \mathbb{S}^{d-1} \\ \omega \in \Omega_\theta(\mu_1, \mu_2)}} \int_{\mathbb{R}^{2d}} \| x_1 - x_2 \|_2^2 \mathrm{d}\omega(x_1, x_2), \tag{51}
$$

where we used the notation $\Omega_\theta(\mu_1, \mu_2)$ defined in Eq. (32), and Theorem 2. We show below that the infimum is attained:

**Proposition 10.** Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$. Then the minimum in Eq. (51) is attained.

*Proof.* Take a sequence $(\theta_n)_{n \in \mathbb{N}} \in \mathbb{S}^{d-1}$ such that $\mathrm{PS}_{\theta_n}(\mu_1, \mu_2) \xrightarrow[n \longrightarrow +\infty]{} \min \mathrm{PS}(\mu_1, \mu_2)$. By compactness of $\mathbb{S}^{d-1}$, we can extract a converging subsequence of $(\theta_n)$: up to extraction we can assume that $\theta_n \xrightarrow[n \longrightarrow +\infty]{} \theta \in \mathbb{S}^{d-1}$. Denoting $\mu_{\theta_n} := \mu_{\theta_n}[\mu_1, \mu_2]$, by Proposition 1, for each $n \in \mathbb{N}$ we can choose $\rho_n \in \Gamma(\mu_{\theta_n}, \mu_1, \mu_2)$ optimal for $\mathrm{PS}_{\theta_n}(\mu_1, \mu_2)$. By Proposition 6, we have $\mu_{\theta_n} \xrightarrow[n \longrightarrow +\infty]{w} \mu_\theta$. Using Lemma 1 item 1) we obtain that the set of $\Gamma(\{\mu_{\theta_n}\}, \mu_1, \mu_2)$ is tight in $\mathcal{P}_2(\mathbb{R}^{3d})$, and since $(\rho_n) \in \Gamma(\{\mu_{\theta_n}\}, \mu_1, \mu_2)^{\mathbb{N}}$, there exists an extraction $\alpha$ such that $\rho_{\alpha(n)} \xrightarrow[n \longrightarrow +\infty]{w} \rho \in \mathcal{P}_2(\mathbb{R}^{3d})$. Applying Lemma 1 item 2) shows that $\rho \in \Gamma(\mu_\theta, \mu_1, \mu_2)$.

The cost function $J := \rho \in \mathcal{P}_2(\mathbb{R}^{3d}) \longmapsto \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho_{1,2}(y, x_1, x_2)$ is lower semi-continuous by [San15] Lemma 1.6, which provides the following inequality:

$$\mathrm{PS}_\theta^2(\mu_1, \mu_2) \leq J(\rho) \leq \liminf_{n \longrightarrow +\infty} J(\rho_{\alpha(n)}) = \lim_{n \longrightarrow +\infty} S_{\theta_{\alpha(n)}}^2(\mu_1, \mu_2) = \min \mathrm{PS}^2(\mu_1, \mu_2),$$

where the first inequality holds by the property $\rho \in \Gamma(\mu_\theta, \mu_1, \mu_2)$, the second inequality comes from the lower semi-continuity of $J$ and the first equality comes from the fact that $\forall n \in \mathbb{N}$, $J(\rho_n) = \mathrm{PS}_{\theta_n}^2(\mu_1, \mu_2)$. We conclude that $\min \mathrm{PS}(\mu_1, \mu_2) = \mathrm{PS}_\theta(\mu_1, \mu_2)$ and thus the infimum is attained. $\square$

From Proposition 10, we conclude that the properties of $\mathrm{PS}_\theta$ stated in Proposition 5 are inherited by $\min \mathrm{PS}$. In Example 5, we show an example which numerically contradicts the triangle inequality.

**Example 5** ($\min \mathrm{PS}$ does not verify the triangle inequality)**.** We consider a setting with three measures $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(\mathbb{R}^2)$ with 10 points each, obtained with five rotations of the example from Example 3, which we represent in Fig. 14. Extensive numerical approximation with $L := 10^5$ directions yields the following violation of the triangle inequality:

$$\min \mathrm{PS}(\mu_1, \mu_3) + \min \mathrm{PS}(\mu_3, \mu_2) - \min \mathrm{PS}(\mu_1, \mu_2) \approx -0.612.$$

While the expression are not tractable in closed form, this numerical experiment strongly suggests that the triangle inequality does not hold for $\min \mathrm{PS}$.

## 6.2 Equality with the Wasserstein Distance for Certain Discrete Measures

In [Mah+23] (Proposition 3.2), the authors show (proof in [Mah+23] Section 11.1) that the Min-Sliced Discrepancy $\min \mathrm{PS}$ equals $W_2$ on the set $\mathcal{P}^n(\mathbb{R}^d)$ of uniform discrete measures with $n$ points under a condition on $n$ and $d$. Their proof relies on an application of [Cov67] which requires the points to be in general position (see Definition 7), however the condition is not stated in [Mah+23]. For the sake of clarity, we restate the result and provide a detailed proof. First, we remind the notion of points of $\mathbb{R}^d$ in general position in Definition 7.

**Definition 7.** Let $x_1, \cdots, x_n \in \mathbb{R}^d$. We say that the points are in general position if for all $k \in [\![1, d]\!]$, there is no subset $I \subset [\![1, n]\!]$ with $k + 2$ elements such that $\{x_i\}_{i \in I}$ is contained in a $k$-dimensional affine subspace of $\mathbb{R}^d$.

**Proposition 11.** Let $\mu := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $\nu := \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$ such that the union of supports
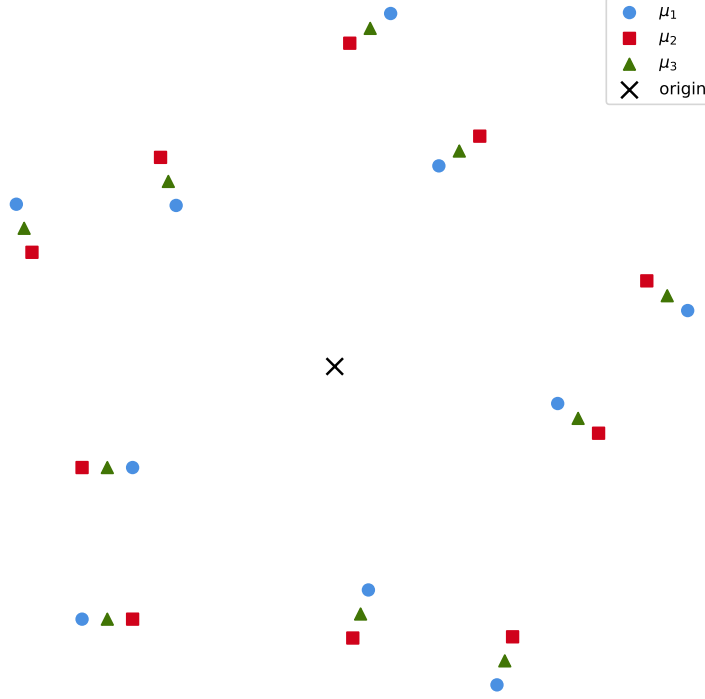
Figure 14: Counter-example from Example 5 to the triangle inequality for $\min \mathrm{PS}$.

$(x_i) \cup (y_j)$ is in general position. If $2n \leq d+1$, then:

$$\left\{(\sigma, \tau) \in \mathfrak{S}_n^2 : \exists \theta \in \mathbb{S}^{d-1} : \forall i \in [\![1, n-1]\!],\ P_\theta x_{\sigma(i)} < P_\theta x_{\sigma(i+1)},\ P_\theta y_{\tau(i)} < P_\theta y_{\tau(i+1)}\right\} = \mathfrak{S}_n^2.$$
(52)

As a result, $\min \mathrm{PS}(\mu, \nu) = \mathrm{W}_2(\mu, \nu)$.

*Proof.* By the Theorem in [Cov67] Section 2 and Equation (12) in Section 3, since $(x_1, \cdots, x_n, y_1, \cdots, y_n)$ are in general position and $d \geq 2n - 1$, we have

$$\left\{\alpha \in \mathfrak{S}_{2n} : \exists \theta \in \mathbb{S}^{d-1} : \forall k \in [\![1, 2n-1]\!],\ P_\theta z_{\alpha(i)} < P_\theta z_{\alpha(i+1)}\right\} = \mathfrak{S}_{2n},$$
(53)

where $(z_1, \cdots, z_{2n}) := (x_1, \cdots, x_n, y_1, \cdots, y_n)$. Take now $(\sigma, \tau) \in \mathfrak{S}_n^2$ and define $\alpha \in \mathfrak{S}_{2n}$ by:

$$\forall i \in [\![1, n]\!],\ \alpha(i) := \sigma(i), \quad \forall j \in [\![1, n]\!],\ \alpha(n+j) := \tau(j).$$

By Eq. (53), there exists $\theta \in \mathbb{S}^{d-1}$ such that $\forall k \in [\![1, 2n-1]\!]$, $P_\theta z_{\alpha(i)} < P_\theta z_{\alpha(i+1)}$, showing Eq. (52).

Now by definition of $\mathfrak{S}_\theta(X, Y)$, the RHS term of Eq. (52) is a subset of $\cup_{\theta \in \mathbb{S}^{d-1}} \mathfrak{S}_\theta(X, Y)$, which shows that $\cup_{\theta \in \mathbb{S}^{d-1}} \mathfrak{S}_\theta(X, Y) = \mathfrak{S}_n^2$. Using Theorem 2 and Theorem 4 we conclude:

$$\min \mathrm{PS}(\mu, \nu) = \min_{(\sigma, \tau) \in \cup_{\theta \in \mathbb{S}^{d-1}} \mathfrak{S}_\theta(X, Y)} \frac{1}{n} \sum_{i=1}^n \|x_{\sigma(i)} - y_{\tau(i)}\|_2^2 = \min_{(\sigma, \tau) \in \mathfrak{S}_n^2} \frac{1}{n} \sum_{i=1}^n \|x_{\sigma(i)} - y_{\tau(i)}\|_2^2 = \mathrm{W}_2^2(\mu, \nu),$$

where the last equality comes from the Monge formulation of $\mathrm{W}_2$ in the case of uniform measures with the same number of points (see [PC19] Proposition 2.1 for instance). □

## 7   Expected Sliced Wasserstein

In [Liu+24], Liu et al. present a variant of the Sliced Wasserstein distance, consisting in taking the transport cost of a coupling that is an average of lifted sliced couplings. In this section, we will explain how to define these notions for general measures of $\mathcal{P}_2(\mathbb{R}^d)$ instead of discrete measures ones.

## 7.1   Lifting Sliced Plans

To lift a 1D transport plan onto $\mathbb{R}^d$, we will require the notion of disintegration of measures with respect to a map reminded in Section 2.4. Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\theta \in \mathbb{S}^{d-1}$. Consider the disintegration of $\mu_1$ with respect to $P_\theta := x \longmapsto x \cdot \theta$ as in Definition 10: $\mu_1(\mathrm{d}x) = (P_\theta \# \mu_1)(P_\theta \mathrm{d}x) \mu_1^{P_\theta x}(\mathrm{d}x)$. The kernel $\mu_1^{P_\theta x}$ is a measure on $\mathbb{R}^d$ supported on the slice $\{x' \in \mathbb{R}^d \mid x' \cdot \theta = x \cdot \theta\} = x + \theta^\perp$. Denoting similarly the disintegration of $\mu_2$ by $\mu_2(\mathrm{d}y) = (P_\theta \# \mu_2)(P_\theta \mathrm{d}y) \mu_2^{P_\theta y}(\mathrm{d}y)$, we first notice that the disintegration of $\mu_1 \otimes \mu_2$ with respect to $(P_\theta, P_\theta)$ writes simply as a product:

$$\mu_1 \otimes \mu_2(\mathrm{d}x, \mathrm{d}y) = (P_\theta \# \mu_1)(P_\theta \mathrm{d}x)(P_\theta \# \mu_2)(P_\theta \mathrm{d}y) \mu_1^{P_\theta x}(\mathrm{d}x) \mu_2^{P_\theta y}(\mathrm{d}y), \tag{54}$$

noticing that $(P_\theta, P_\theta) \# (\mu_1 \otimes \mu_2) = (P_\theta \# \mu_1) \otimes (P_\theta \# \mu_2)$.

Take now the 1D OT plan $\pi_\theta := \pi_\theta[\mu_1, \mu_2] \in \Pi^*(P_\theta \# \mu_1, P_\theta \# \mu_2)$, the idea behind the lift is to replace the independent coupling $(P_\theta, P_\theta) \# (\mu_1 \otimes \mu_2) = (P_\theta \# \mu_1) \otimes (P_\theta \# \mu_2)$ in Eq. (54) by $\pi_\theta$: we define the lifted plan through its disintegration as:

$$\gamma_\theta(\mathrm{d}x, \mathrm{d}y) := \pi_\theta(P_\theta \mathrm{d}x, P_\theta \mathrm{d}y) \mu_1^{P_\theta x}(\mathrm{d}x) \mu_2^{P_\theta y}(\mathrm{d}y). \tag{55}$$

More formally, we can define $\gamma_\theta$ using test functions $\phi \in \mathcal{C}_b^0(\mathbb{R}^d \times \mathbb{R}^d)$:

$$\int_{\mathbb{R}^{2d}} \phi(x, y) \mathrm{d}\gamma_\theta(x, y) = \int_{\mathbb{R}^2} \left( \int_{P_\theta^{-1}(s) \times P_\theta^{-1}(t)} \phi(x, y) \mathrm{d}\mu_1^s(x) \mathrm{d}\mu_2^t(y) \right) \mathrm{d}\pi_\theta(s, t). \tag{56}$$

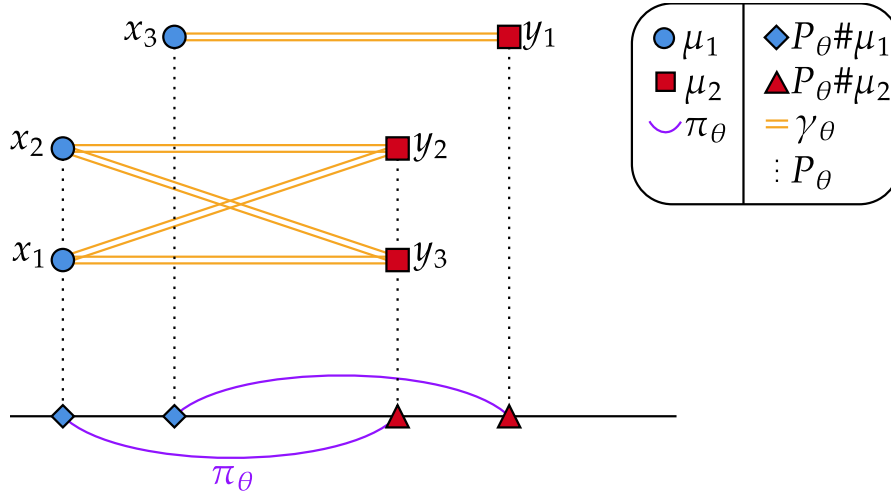We illustrate the definition of the lifted plan on a simple example in Fig. 15.



Figure 15: Example of the lifted plan $\gamma_\theta$ between two measures $\mu_1$ and $\mu_2$. In this case, we notice that $P_\theta x_1 = P_\theta x_2$ and $P_\theta y_2 = P_\theta y_3$. As a result, the optimal plan $\pi_\theta$ between $P_\theta \# \mu_1$ and $P_\theta \# \mu_2$ does not allow us to deduce an assignment between $(x_1, x_2)$ and $(y_2, y_3)$. The lifted coupling $\gamma_\theta$ chooses the independent coupling: $x_1$ is assigned uniformly to $(y_2, y_3)$ and likewise for $x_2$. As for $x_3$ and $y_1$, the coupling $\pi_\theta$ assigns $P_\theta x_3$ to $P_\theta y_1$ which imposes that $\gamma_\theta$ send $x_3$ to $y_1$.

In Proposition 12 we show that the lifted plan is a valid coupling between $\mu_1$ and $\mu_2$. We also provide an explicit expression for discrete measures $\mu_1, \mu_2$, which coincides with the expression in [Liu+24] Equation 9, which serves as their definition of lifted plans.

**Proposition 12.** Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, $\theta \in \mathbb{S}^{d-1}$ and $\gamma_\theta := \gamma_\theta[\mu_1, \mu_2]$ the lifted plan defined in Eq. (56). Then:
  1. $\gamma_\theta \in \Pi(\mu_1, \mu_2)$.
  2. If $\mu_1 = \sum_{i=1}^n a_i \delta_{x_i}$ and $\mu_2 = \sum_{j=1}^m b_j \delta_{y_j}$, let $\pi_\theta \in \Pi^*(P_\theta \# \mu_1, P_\theta \# \mu_2)$. For $(i, j) \in [\![1, n]\!] \times$

$[\![1, m]\!]$, we define $Q_{i,j} := \pi_\theta(\{(P_\theta x_i, P_\theta y_j)\})$, which allows us to see $\pi_\theta$ as a matrix of size $n \times m$.

$$\gamma_\theta = \sum_{i=1}^n \sum_{j=1}^m \frac{a_i b_j}{A_i B_j} Q_{i,j} \delta_{(x_i, y_j)}, \quad A_i := \sum_{i': x_{i'} \cdot \theta = x_i \cdot \theta} a_{i'}, \ B_j := \sum_{j': y_{j'} \cdot \theta = y_j \cdot \theta} b_{j'}. \tag{57}$$

*Proof.* For 1. we verify the property using a test function $\phi \in \mathcal{C}_b^0(\mathbb{R}^d)$ with Eq. (56):

$$\int_{\mathbb{R}^{2d}} \phi(x) \mathrm{d}\gamma_\theta(x, y) = \int_{\mathbb{R}^2} \left( \int_{P_\theta^{-1}(s) \times P_\theta^{-1}(t)} \phi(x) \mathrm{d}\mu_1^s(x) \mathrm{d}\mu_2^t(y) \right) \mathrm{d}\pi_\theta(s, t)$$

$$= \int_{\mathbb{R}^2} \left( \int_{P_\theta^{-1}(s)} \phi(x) \mathrm{d}\mu_1^s(x) \right) \underbrace{\left( \int_{P_\theta^{-1}(t)} \mathrm{d}\mu_2^t(y) \right)}_{=1} \mathrm{d}\pi_\theta(s, t)$$

$$= \int_{\mathbb{R}} \left( \int_{P_\theta^{-1}(s)} \phi(x) \mathrm{d}\mu_1^s(x) \right) \mathrm{d}(P_\theta \# \mu_1)(s)$$

$$= \int_{\mathbb{R}^d} \phi(x) \mathrm{d}\mu_1(x),$$

where we use the fact that the first marginal of $\pi_\theta$ is $P_\theta \# \mu_1$, and then used disintegration of $\mu_1$ with respect to $P_\theta$. The same method shows that the second marginal of $\gamma_\theta$ is $\mu_2$, concluding $\gamma_\theta \in \Pi(\mu_1, \mu_2)$.

For 2. we begin by writing explicitly the disintegration of $\mu_1$ with respect to $P_\theta$. For $\mu_1 = \sum_{i=1}^n a_i \delta_{x_i}$, we have $P_\theta \# \mu = \sum_i a_i \delta_{P_\theta x_i}$, and for $s = P_\theta x_i \in \mathrm{supp}(P_\theta \# \mu)$, we have $\mu_1^s = A_i^{-1} \sum_{i': x_{i'}' = s} a_{i'} \delta_{x_{i'}}$. We establish Eq. (57) by testing on $\phi \in \mathcal{C}_b^0(\mathbb{R}^{2d})$. The support of $\pi_\theta$ is (at most) the family of pairs $((x_i \cdot \theta, y_j \cdot \theta))_{i,j}$. We choose $I \subset [\![1, n]\!] \times [\![1, m]\!]$ such that the support of $\pi_\theta$ is the injective family $((x_i \cdot \theta, y_j \cdot \theta))_{(i,j) \in I}$. We then have:

$$\int_{\mathbb{R}^{2d}} \phi(x, y) \mathrm{d}\gamma_\theta(x, y) = \int_{\mathbb{R}^2} \left( \int_{P_\theta^{-1}(s) \times P_\theta^{-1}(t)} \phi(x, y) \mathrm{d}\mu_1^s(x) \mathrm{d}\mu_2^t(y) \right) \mathrm{d}\pi_\theta(s, t)$$

$$= \sum_{(i,j) \in I} \left( \sum_{\substack{i': P_\theta x_{i'} = P_\theta x_i \\ j': P_\theta y_{j'} = P_\theta y_j}} \phi(x_{i'}, y_{j'}) \frac{a_{i'} b_{j'}}{A_i B_j} \right) \pi_\theta(\{(P_\theta x_i, P_\theta y_j)\})$$

$$= \sum_{i=1}^n \sum_{j=1}^m \phi(x_i, y_j) \frac{a_i b_j}{A_i B_j} Q_{i,j},$$

where we use the fact that for $(i, j) \in I$ and $(i', j')$ such that $P_\theta x_{i'} = P_\theta x_i$ and $P_\theta y_{j'} = P_\theta y_j$, it holds that $A_i = A_{i'}$, $B_j = B_{j'}$ and $Q_{i,j} = Q_{i',j'}$. $\qquad\square$

The discrete expression in Eq. (57) shows that the definition of listed plans in Eq. (56) is a generalisation of the plan lift from [Liu+24] (Equation 9). We now study the transport cost associated to the lifted plan $\gamma_\theta$:

**Definition 8.** For $\theta \in \mathbb{S}^{d-1}$ and $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$. With $\gamma_\theta[\mu_1, \mu_2]$ the lifted plan defined in Eq. (56), we define the lifted cost as:

$$\mathrm{LS}_\theta^2(\mu_1, \mu_2) := \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\gamma_\theta[\mu_1, \mu_2](x_1, x_2).$$

We will see that $\mathrm{LS}_\theta$ defines a discrepancy on $\mathcal{P}_2(\mathbb{R}^d)$ that is almost a distance.

**Proposition 13.** Fix $\theta \in \mathbb{S}^{d-1}$. The quantity $\mathrm{LS}_\theta$ is non-negative, symmetric, verifies the triangle inequality, and if $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ verify $\mathrm{LS}_\theta(\mu_1, \mu_2) = 0$ then $\mu_1 = \mu_2$. Furthermore, we have the inequality $\mathrm{LS}_\theta \geq \mathrm{W}_2$.

*Proof.* Non-negativity and symmetry are immediate. For $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, by Proposition 12, we have $\gamma_\theta[\mu_1, \mu_2] \in \Pi(\mu_1, \mu_2)$, hence $\mathrm{LS}_\theta(\mu_1, \mu_2) \geq \mathrm{W}_2(\mu_1, \mu_2)$. Suppose now that $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ are such that $\mathrm{LS}_\theta(\mu_1, \mu_2) = 0$, then $\mathrm{W}_2(\mu_1, \mu_2) = 0$ and therefore $\mu_1 = \mu_2$.

We now show the triangle inequality: let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_2(\mathbb{R}^d)$. We consider the sliced 3-plan $\eta_\theta$ defined by:

$$\eta_\theta := \left( F_{P_\theta \# \mu_1}^{[-1]}, F_{P_\theta \# \mu_2}^{[-1]}, F_{P_\theta \# \mu_3}^{[-1]} \right) \# \mathscr{L}_{[0,1]}.$$

For $i < j \in \{1, 2, 3\}$, introduce $\pi_\theta^{(i,j)}$ the unique optimal transport plan between $P_\theta \# \mu_i$ and $P_\theta \# \mu_j$. By [San15] Theorem 2.9, we see that $[\eta_\theta]_{i,j} = \pi_\theta^{(i,j)}$. We now lift the sliced plan $\eta_\theta$ in the same manner as in Eq. (56), defining a plan $\rho_\theta \in \mathcal{P}_2(\mathbb{R}^{3d})$ by disintegration:

$$\rho_\theta(\mathrm{d}x_1, \mathrm{d}x_2, \mathrm{d}x_3) := \eta_\theta(P_\theta \mathrm{d}x_1, P_\theta \mathrm{d}x_2, P_\theta \mathrm{d}x_3) \mu_1^{P_\theta x_1}(\mathrm{d}x_1) \mu_2^{P_\theta x_2}(\mathrm{d}x_2) \mu_3^{P_\theta x_3}(\mathrm{d}x_3).$$

By computing the expectation against test functions, $\forall i < j \in \{1, 2, 3\}$, $[\rho_\theta]_{i,j} = \gamma_\theta[\mu_i, \mu_j]$. We now use the classical gluing method (as in [San15] Lemma 5.5) to show the triangle inequality, introducing the functions $\phi_i := (x_1, x_2, x_3) \longmapsto x_i$ for $i \in \{1, 2, 3\}$:

$$
\begin{aligned}
\mathrm{LS}_\theta(\mu_1, \mu_3) &= \sqrt{\int_{\mathbb{R}^{2d}} \|x_1 - x_3\|_2^2 \, \mathrm{d}\gamma_\theta[\mu_1, \mu_3](x_1, x_3)} \\
&= \sqrt{\int_{\mathbb{R}^{3d}} \|x_1 - x_3\|_2^2 \, \mathrm{d}\rho_\theta(x_1, x_2, x_3)} \\
&= \|\phi_1 - \phi_3\|_{L^2(\rho_\theta)} \\
&\leq \|\phi_1 - \phi_2\|_{L^2(\rho_\theta)} + \|\phi_2 - \phi_3\|_{L^2(\rho_\theta)} \\
&= \sqrt{\int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \, \mathrm{d}\gamma_\theta[\mu_1, \mu_2](x_1, x_2)} + \sqrt{\int_{\mathbb{R}^{2d}} \|x_2 - x_3\|_2^2 \, \mathrm{d}\gamma_\theta[\mu_2, \mu_3](x_1, x_2)} \\
&= \mathrm{LS}_\theta(\mu_1, \mu_2) + \mathrm{LS}_\theta(\mu_2, \mu_3).
\end{aligned}
$$

$\square$

The discrepancy $\mathrm{LS}_\theta$ is not a distance on $\mathcal{P}_2(\mathbb{R}^d)$: in Example 6, we introduce a particular case in dimension two where $\mathrm{LS}_\theta(\mu, \mu) > 0$.

**Example 6** ($\mathrm{LS}_\theta(\mu, \mu)$ can be non-zero). Take $\theta := (1, 0)$ and $\mu := \frac{1}{2}(\delta_{x_0} + \delta_{x_1})$, $x_0 := (0, 0)$, $x_1 := (0, 1)$. We have $P_\theta \# \mu = \delta_0$ and thus $\gamma_\theta[\mu, \mu] = \mu \otimes \mu$. The lifted cost is then:

$$\mathrm{LS}_\theta^2(\mu, \mu) = \frac{1}{4} \left( \|x_0 - x_0\|_2^2 + \|x_0 - x_1\|_2^2 + \|x_1 - x_0\|_2^2 + \|x_1 - x_1\|_2^2 \right) = \frac{1}{2} > 0.$$

For probability measures $\mu$ with countable support, for almost-every $\theta \in \mathbb{S}^{d-1}$, there is no ambiguity in the projections, and thus $\mathrm{LS}_\theta(\mu, \mu) = 0$, as shown in Proposition 14. To state the result, we introduce the following notation for the set of probability measures with countable[3] support:

$$\mathcal{P}_{\mathrm{DC}}(\mathbb{R}^d) := \left\{ \mu = \sum_{x \in X} a_x \delta_x : X \subset \mathbb{R}^d \text{ countable}, \ (a_x)_{x \in X} \in (0, 1]^d, \ \sum_{x \in X} a_x = 1 \right\}. \tag{58}$$

---

[3]by "countable", we mean a set that is either finite or equipotent to $\mathbb{N}$.

**Proposition 14.** Consider $\mu \in \mathcal{P}_{\mathrm{DC}}(\mathbb{R}^d)$, then for almost-every $\theta \in \mathbb{S}^{d-1}$, we have $\mathrm{LS}_\theta(\mu, \mu) = 0$.

*Proof.* Denoting $\sigma_u$ the uniform measure on the unit sphere $\mathbb{S}^{d-1}$, we have by countable additivity $\mathbb{P}_{\theta \sim \sigma_u}(\exists x \neq y \in X^2 : P_\theta x = P_\theta y) \leq \sum_{x \neq y \in X^2} \sigma_u((x-y)^\perp) = 0$. We now fix $\Theta \subset \mathbb{S}^{d-1}$ the set $\Theta := \{\theta \in \mathbb{S}^{d-1} : \forall x \neq y \in X^2, \theta \notin (x-y)^\perp\}$, we have shown that $\sigma_u(\Theta) = 1$. For $\theta \in \Theta$, the family $(P_\theta x)_{x \in X}$ is injective, and the disintegration kernel $\mu$ with respect to $P_\theta$ at $P_\theta x$ is simply $\delta_{Q_{\theta^\perp} x}$, therefore the lifted plan $\gamma_\theta[\mu, \mu]$ is $\sum_{x \in X} a_x \delta_{(x,x)}$. It follows from the definition that $\mathrm{LS}_\theta(\mu, \mu) = 0$ for any $\theta \in \Theta$, concluding the proof. $\square$

## 7.2 Averaging Lifted Plans

Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\theta \in \mathbb{S}^{d-1}$. We have constructed a lifted plan $\gamma_\theta \in \Pi(\mu_1, \mu_2)$ (see Eq. (56) and Proposition 12). We now define the expected lifted plan as the "average" of lifted plans over all directions $\theta \in \mathbb{S}^{d-1}$ through a probability measure $\sigma \in \mathcal{P}(\mathbb{S}^{d-1})$. We define $\overline{\gamma}[\mu_1, \mu_2, \sigma]$ by duality on test functions $\phi \in \mathcal{C}_b^0(\mathbb{R}^d \times \mathbb{R}^d)$:

$$\int_{\mathbb{R}^{2d}} \phi(x, y) \mathrm{d}\overline{\gamma}[\mu_1, \mu_2, \sigma](x, y) := \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^{2d}} \phi(x, y) \mathrm{d}\gamma_\theta[\mu_1, \mu_2](x, y) \mathrm{d}\sigma(\theta). \tag{59}$$

Having defined the expected lifted plan, we can now define the expected sliced discrepancy:

**Definition 9.** Let $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\sigma \in \mathcal{P}(\mathbb{S}^{d-1})$. The expected sliced discrepancy between $\mu_1$ and $\mu_2$ is defined as:

$$\begin{aligned}
\mathrm{ES}_\sigma^2(\mu_1, \mu_2) &:= \int_{\mathbb{R}^{2d}} \|x - y\|_2^2 \mathrm{d}\overline{\gamma}[\mu_1, \mu_2, \sigma](x, y) \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^{2d}} \|x - y\|_2^2 \mathrm{d}\gamma_\theta[\mu_1, \mu_2](x, y) \mathrm{d}\sigma(\theta) \\
&= \int_{\mathbb{S}^{d-1}} \mathrm{LS}_\theta^2(\mu_1, \mu_2) \mathrm{d}\sigma(\theta).
\end{aligned}$$

where $\overline{\gamma}[\mu_1, \mu_2, \sigma]$ is the expected lifted plan between $\mu_1$ and $\mu_2$ for the measure $\sigma$ on $\mathbb{S}^{d-1}$, defined in Eq. (59), and $\gamma_\theta[\mu_1, \mu_2]$ is the lifted plan defined in Eq. (56).

The properties of $\mathrm{LS}_\theta$ are passed on to $\mathrm{ES}_\sigma$ by integration.

**Corollary 2.** For any probability measure $\sigma$ on $\mathbb{S}^{d-1}$, the quantity $\mathrm{ES}_\sigma$ is non-negative, symmetric, verifies the triangle inequality, and if $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ verify $\mathrm{ES}_\sigma(\mu_1, \mu_2) = 0$ then $\mu_1 = \mu_2$. Furthermore, we have the inequality $\mathrm{ES}_\sigma \geq \mathrm{W}_2$.

*Proof.* Non-negativity, symmetry and the property $\mathrm{ES}_\sigma \geq \mathrm{W}_2$ are immediate by applying the definition and Proposition 13. For the triangle inequality, let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_2(\mathbb{R}^d)$ and for $i < j \in \{1, 2, 3\}$, introduce $f_{i,j} := \theta \longmapsto \mathrm{LS}_\theta(\mu_i, \mu_j)$. By Proposition 13 we have $0 \leq f_{1,3} \leq f_{1,2} + f_{2,3}$. We write:

$$\mathrm{ES}_\sigma(\mu_1, \mu_3) = \|f_{1,3}\|_{L^2(\sigma)} \leq \|f_{1,2} + f_{2,3}\|_{L^2(\sigma)} \leq \|f_{1,2}\|_{L^2(\sigma)} + \|f_{2,3}\|_{L^2(\sigma)} = \mathrm{ES}_\sigma(\mu_1, \mu_2) + \mathrm{ES}_\sigma(\mu_2, \mu_3).$$

$\square$

The discrepancy $\mathrm{ES}_\sigma$ is not a distance on $\mathcal{P}_2(\mathbb{R}^d)$. First, if $d = 2$ and $\sigma = \delta_{(1,0)}$, $\mathrm{ES}_\sigma = \mathrm{LS}_\theta$ and the counter-example from Example 6 earlier with $\mu := \frac{1}{2}\delta_{(0,0)} + \frac{1}{2}\delta_{(0,1)}$ yields $\mathrm{ES}_\sigma(\mu, \mu) > 0$.

Even for probability measures $\sigma$ that are absolutely continuous with respect to the uniform measure on $\mathbb{S}^{d-1}$, we can find examples where $\mathrm{ES}_\sigma(\mu, \mu) > 0$, as presented in Example 7.

**Example 7** (Case where $\mathrm{ES}_\sigma(\mu, \mu) > 0$ for any $\sigma$). Take $\sigma$ any probability measure on $\mathbb{S}^1$ and $\mu := \mathcal{U}(B_{\mathbb{R}^2}(0,1))$ the uniform measure on the Euclidean unit ball of $\mathbb{R}^2$. We have for any $\theta \in \mathbb{S}^1$, $P_\theta\mu = \nu$, where $\nu(\mathrm{d}t) = \frac{2\sqrt{1-t^2}}{\pi}\mathbb{1}_{[-1,1]}(t)\mathrm{d}t$. The disintegration of $\mu$ with respect to $P_\theta$ is covariant with respect to $\theta$, and the disintegration kernel at $t = P_\theta x$ is $\mu^t = \mathcal{U}\left(\left\{t\theta + v\theta_\perp, \; v \in [-\sqrt{1-t^2}, \sqrt{1-t^2}]\right\}\right)$, where we have fixed $\theta_\perp$ a unit orthogonal vector to $\theta$. The disintegration kernel $\mu^t$ is the uniform measure on the ball slice of $B_{\mathbb{R}^2}(0,1) \cap (t\theta + \theta^\perp)$ (with $\theta^\perp := \{x \in \mathbb{R}^d : \theta \cdot x = 0\}$), and can simply be understood as the uniform measure on the segment $[-\sqrt{1-t^2}, \sqrt{1-t^2}]$, cast into $\mathbb{R}^2$. The optimal transport plan between $\nu$ and itself is $\pi_\theta := (I, I)\#\nu$, and it follows that the lifted plan between $\mu$ and itself is (denoting $t := P_\theta x_1$ for legibility):

$$\gamma_\theta(\mathrm{d}x_1, \mathrm{d}x_2) = \delta_{P_\theta x_1 = P_\theta x_2}(\mathrm{d}P_\theta x_1, \mathrm{d}P_\theta x_2)\, \nu(\mathrm{d}P_\theta x_1)$$
$$\otimes \mathcal{U}\left(\left\{(t\theta + v_1\theta_\perp, t\theta + v_2\theta_\perp), \; (v_1, v_2) \in [-\sqrt{1-t^2}, \sqrt{1-t^2}]^2\right\}\right)(\mathrm{d}x_1, \mathrm{d}x_2).$$

We provide a visualisation of the disintegration $\mu^t$ and the coupling $\gamma_\theta$ in Fig. 16. By symmetry $\mathrm{LS}_\theta(\mu, \mu)$ does not depend on $\theta$, we compute it for $\theta := (1, 0)$:

$$\mathrm{LS}_\theta^2(\mu, \mu) = \int_{\mathbb{R}^3} \|(u, v_1) - (u, v_2)\|_2^2 \mathbb{1}_{[-1,1]}(u) \frac{2\sqrt{1-u^2}}{\pi} \mathbb{1}_{[-\sqrt{1-u^2}, \sqrt{1-u^2}]^2}(v_1, v_2) \left(\frac{1}{2\sqrt{1-u^2}}\right)^2 \mathrm{d}v_1\mathrm{d}v_2\mathrm{d}u$$

$$= \int_{u=-1}^{u=1} \frac{\sqrt{1-u^2}}{2\pi} \int_{v_1=-\sqrt{1-u^2}}^{v_1=\sqrt{1-u^2}} \int_{v_2=-\sqrt{1-u^2}}^{v_2=\sqrt{1-u^2}} (v_1 - v_2)^2 \mathrm{d}v_1\mathrm{d}v_2\mathrm{d}u$$

$$= \frac{5\pi}{12} > 0.$$

We conclude that $\mathrm{ES}_\sigma(\mu, \mu) > 0$ (for any $\sigma$), and thus $\mathrm{ES}_\sigma$ is not a distance on $\mathcal{P}_2(\mathbb{R}^d)$.
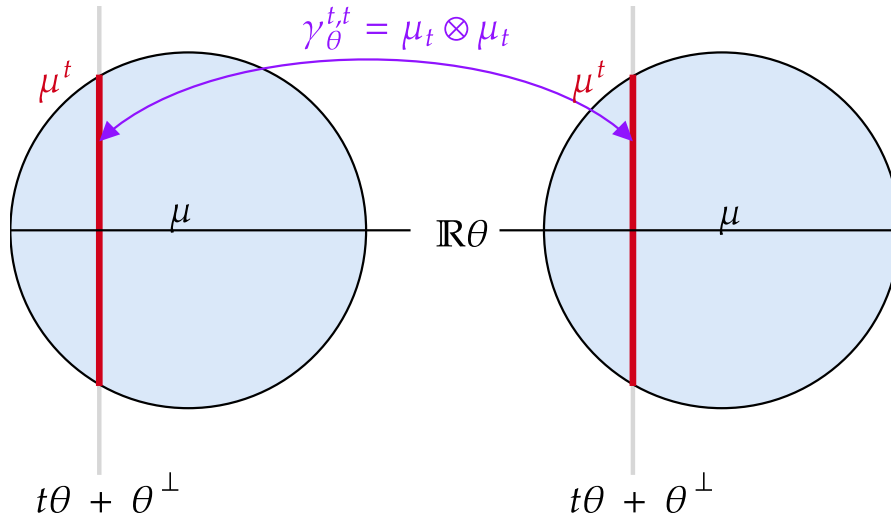


Figure 16: Illustration of the lifted plan $\gamma_\theta$ from Example 7 between $\mu$ the uniform measure on the unit Euclidean ball of $\mathbb{R}^2$, and itself. The plan is defined by disintegration: the coupling between $P_\theta\#\mu$ and $P_\theta\#\mu$ is simply $(I, I)$, the coupling induced by the identity map. As for the orthogonal part, the disintegration kernel of $\mu$ at $t\theta$ is $\mu^t$, the uniform measure on the ball slice $B_{\mathbb{R}^2}(0,1) \cap (t\theta + \theta^\perp)$, represented as a thick red vertical line. The lifted plan couples the disintegration kernel $\mu^t$ with itself with the independent coupling: writing $\gamma_\theta^{t,t}$ as the disintegration kernel of $\gamma_\theta$ at $(t, t) \in [-1, 1]^2$, we have $\gamma_\theta^{(t,t)} = \mu_t \otimes \mu_t$, which corresponds to the uniform measure on the square $\{(t\theta + v_1\theta_\perp, t\theta + v_2\theta_\perp), \; (v_1, v_2) \in [-\sqrt{1-t^2}, \sqrt{1-t^2}]^2\} \subset \mathbb{R}^4$.

Using Proposition 14, we can show that the expected sliced distance is a distance on the set of

"countably discrete" probability measures defined in Eq. (58).

> **Corollary 3.** For any $\varpi$ a probability measure on $\mathbb{S}^{d-1}$ that is absolutely continuous with respect to $\varpi_u$, the quantity $\mathrm{ES}_\varpi$ is a distance on $\mathcal{P}_{\mathrm{DC}}(\mathbb{R}^d)$.

*Proof.* Thanks to Corollary 2, the only axiom to verify to show that $\mathrm{ES}_\varpi$ is a distance on $\mathcal{P}_{\mathrm{DC}}(\mathbb{R}^d)$ is to show that $\forall \mu \in \mathcal{P}_{\mathrm{DC}}(\mathbb{R}^d)$, $\mathrm{ES}_\varpi(\mu, \mu) = 0$. We now fix $\mu \in \mathcal{P}_{\mathrm{DC}}(\mathbb{R}^d)$. Since $\varpi \ll \varpi_u$, we have by Proposition 14 that for $\varpi$-almost-every $\theta \in \mathbb{S}^{d-1}$, $\mathrm{LS}_\theta(\mu, \mu) = 0$. We conclude $\mathrm{ES}_\varpi^2(\mu, \mu) = \int_{\mathbb{S}^{d-1}} \mathrm{LS}_\theta^2(\mu, \mu) \mathrm{d}\varpi(\theta) = 0$. $\qquad\square$

# 8 Numerics

In this section, we evaluate the efficiency and practicability of the sliced-based transport plans, namely min-Pivot Sliced Wasserstein (min PS) and expected Sliced Wasserstein (ES), in both synthetic and real-world scenarios. We begin by presenting quantitative and qualitative results on toy datasets, evaluating their ability to generate meaningful transport plans and costs across various settings. We continue with a colour transfer task, which is simple to assess qualitatively yet can be computationally challenging in classic OT due to the large sample size ($n \geq 500^2$). We finish with a more complex task that involves large-scale datasets where a transport plan is required, namely point cloud registration. For these experiments, we employ the POT toolbox [Fla+21]. Note that we report experimental results only in the context of distributions with the same number of samples but that the results can be easily extended to the case of different number of samples. All experiments were run on CPU on a MacBook Pro with an M1 chip.

## 8.1 Evaluation of the Transport Losses and Plans

### 8.1.1 Gradient Flows

We perform a gradient flow on the support of a discrete source distribution $\mu$, aiming to minimise the (Sliced) Wasserstein distance with respect to a discrete target distribution $\nu$: $\min_\mu \{\mathcal{F}^\nu(\mu)\}$, following the setting of [CTV25]. This procedure yields a flow $(\mu_t)_t$ that decreases the functional $\mathcal{F}^\nu(\mu)$ over time $0 \leq t \leq 1$. We consider here several functionals: the Wasserstein distance $\mathrm{W}_2^2$, the Sliced Wasserstein distance $\mathrm{SW}_2^2$, $\min \mathrm{PS}^2$ and $\mathrm{ES}^2$. For min PS and ES, at each step we draw randomly $L$ directions $\theta_\ell \in \mathbb{S}^{d-1}$ and compute $\min \mathrm{PS}^2 \approx \min_\ell \mathrm{PS}_{\theta_\ell}^2$ and $\mathrm{ES} \approx \frac{1}{L} \sum_{\ell=1}^L \mathrm{LS}_{\theta_\ell}^2$. Moreover, for min PS, we use an optimisation scheme described in [CTV25] to obtain an approximation $\hat{\theta}^\star$ of an optimal direction $\theta^\star \in \mathrm{argmin}_{\theta \in \mathbb{S}^{d-1}} \mathrm{PS}_\theta^2$. In what follows, it is denoted $\mathrm{PS}_{\hat{\theta}^\star}^2$.

We consider several target distributions of $n = 50$ samples, shown in the first and third columns of Fig. 17: a Gaussian distribution (in 2 and 500 dimensions), a spiral, two moons, a circle and eight Gaussians of different means. The source distribution is chosen to be a uniform distribution. We use Adam as an optimisation scheme, with a learning rate of 0.02 for all methods, and consider $L = 50$ directions for the sliced approaches. We report the 2-Wasserstein distance between $\mu_t$ and $\nu$ at each iteration of the optimisation procedure, and repeat each experiment 10 times.

One can observe that the Expected Sliced discrepancy does not converge in any setting. This finding is consistent with the one of [Liu+24] (section 3.4). In contrast, all other methods enable convergence to the target distribution, i.e. $\mu_t \to \nu$ as $t \to 1$, when working in two dimensions. When considering a 500-dimensional Gaussian distribution, only Wasserstein and $\mathrm{PS}_{\hat{\theta}^\star}$ achieve convergence: with a fixed number of samples $n$, we suspect that the required number of directions to obtain a good approximation must grow exponentially with the dimension, making min PS (with $L = 50$ directions), ES and Sliced Wasserstein inadequate for this context. Using optimisation techniques in min PS provides a single meaningful direction $\hat{\theta}^\star$, even when $n$ is small compared to the dimension. One can notice that Wasserstein and $\mathrm{PS}_{\hat{\theta}^\star}$ have very similar behaviours, which is backed by Proposition 11 which states that $\mathrm{PS}_{\hat{\theta}^\star}$ is equal to the 2-Wasserstein distance when $\hat{\theta}^\star$ is an optimal direction and

when $d \geq 2n - 1$. This encourages the use of the minimisation method proposed in [CTV25], which outperforms the search over $L$ random projections.
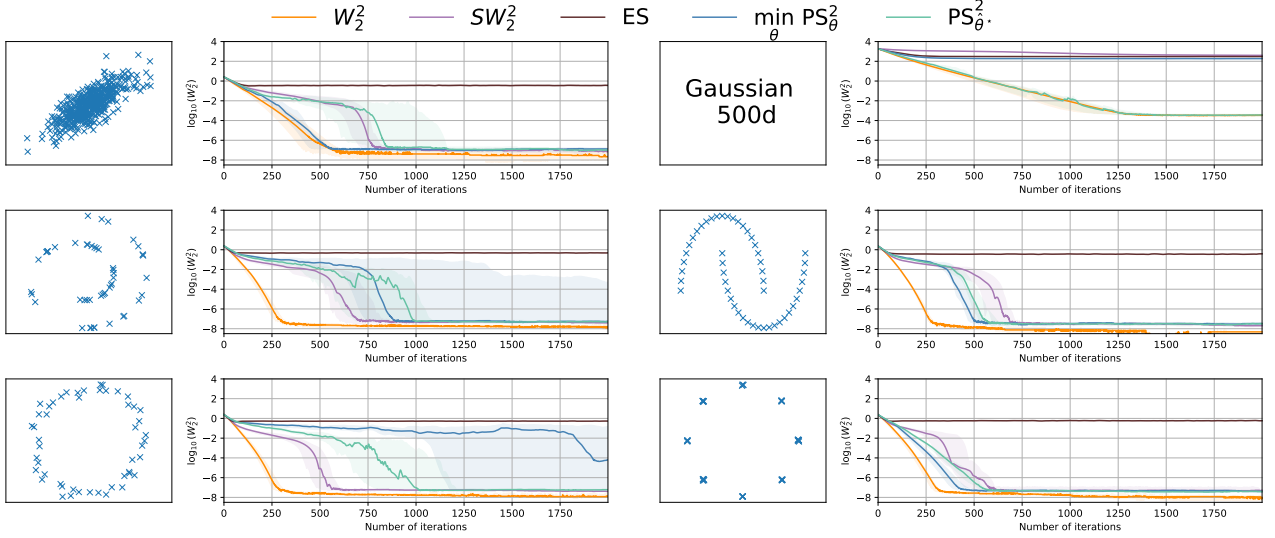


Figure 17: Log 2-Wasserstein distance measured between a source and different target distributions as a function of number of iterations. Plain lines represent the median over 10 iterations while shaded regions indicate 0.25 and 0.75 quantiles.

### 8.1.2   Comparison of Transport Plans and Discrepancies

We now provide a quantitative assessment of the transport plans that can be estimated from sliced-based methods.

**Qualitative assessment of the transport plans.**   We illustrate some transport plans in several two-dimensional settings. The first one corresponds to transporting samples of a source Gaussian distribution to samples of a target Gaussian distribution with different parameters. The second one considers two distributions sampled on circles of the same centre and different radii, with $n = 24$ samples. The last one considers a more challenging and non-linear setting, in which the source distribution is composed of 8 Gaussians of several means and the target is composed of two moons.

Fig. 18 presents the plans obtained with 2-Wasserstein, Expected Sliced and min-Pivot Sliced, together with the associated discrepancy. We choose $L = 50$ directions, and fix $n = 10$ samples for the first scenario and $n = 24$ otherwise. On can notice that, in the simple case of 2 Gaussians as source and target distributions (first line), the transport cost is close to the 2-Wasserstein one. Min-Pivot Sliced provides a plan that is close to the OT one; Expected Sliced provides a highly non-deterministic coupling, associated each source point to numerous targets. When it comes to non-linear settings (third and fifth lines), one can notice that the sliced estimated costs deviate from their OT counterpart: as min PS and ES rely on plans obtained by projecting on a line then lifted to the original space, and because none of these projections capture the true matching, the approximation is quite poor, with spurious matchings between the two parts of the moon. Dedicated variants of Sliced Wasserstein have been proposed in this non-linear setting, for instance *generalised* versions in which the data are projected onto a non linear surface, e.g. [Kol+19], and *augmented* ones [CYL20] that first embeds the data into a higher dimensional space in which a linear surface better captures the distances. These variants are out of the scope of this paper, but note that a non-linear variant of min PS has been proposed in [CTV25].

**Comparing plans obtained by flows.**   To avoid relying on one single direction and to better take into account the non linearities on the distributions, we propose here to build on *flows*, for which different directions can be chosen at each iteration. The second, fourth and sixth lines of

Fig. 18 present trajectories obtained when considering such flows, with an SGD optimiser and a fixed learning rate equal to 2 (as recommended by [Bon+15] under Equation 44, we take a learning rate equal to the dimension). If the flow has converged after 200 steps (that is to say, when the Wasserstein distance between two consecutive step is less than $10^{-6}$), we infer a transport plan as the map linking the source and the target sample reached by the flow. This strategy also allows considering Sliced Wasserstein to obtain a plan, as proposed in Section 3.3 of [Rab+12a]. One can notice that, as expected, 2-Wasserstein flows plan recover the transport plan and that Sliced Wasserstein based plan is close to the actual one. As observed in Section 8.1.1, even in the simple case of 2 Gaussians, Expected Sliced does not converge. When considering min PS, flow-based transport allows enhancing the approximation of the plan, avoiding spurious couplings between the two moons. One further notices that this strategy comes with an extra computational cost as several iterations for computing the flow are needed to obtain the approximation. We present this method to highlight the benefits of stochastic algorithms when using sliced-based methods.
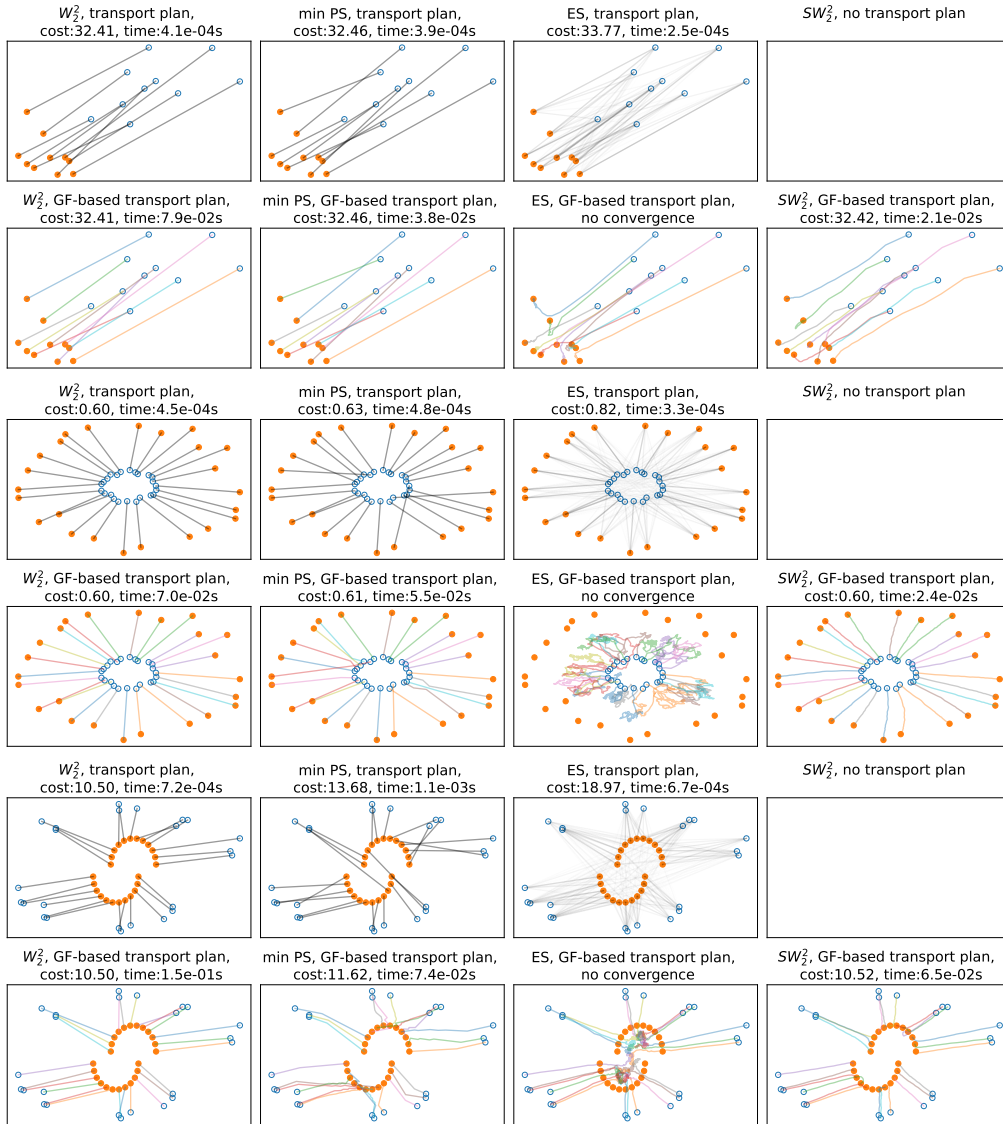


Figure 18: Comparison of the plans obtained by sliced plans methods and 2-Wasserstein between a source (blue samples) and a target (orange samples) distributions. First, third and fifth lines: transport plans obtained by solving Wasserstein, min-Pivot Sliced and Expected Sliced. Second, fourth and sixth lines: trajectories obtained by solving a gradient flow for Wasserstein, min-Pivot Sliced, Expected Sliced and Sliced Wasserstein. In that case, the associated cost is computed by mapping the source sample to the target sample that is reached by the flow.

**Timings**   We report some timings for the different methods, in order to assess their computational efficiency. We consider the same settings as the first scenario of the previous section (two Gaussians as a source and target distribution). For the Sliced Wasserstein flow, we perform 10 steps, with an extra complexity linear with the number of steps. We vary the number of samples from $n = 10$ to $n = 10^7$, and present the results in Fig. 19. One can notice that sliced-based method are significantly faster to compute when $n$ grows. Note that Wasserstein fails to be computed for $n \geq 10^5$ due to memory issues, as it requires to store the full cost matrix $C \in \mathbb{R}^{n \times n}$ of size $n^2$; there is no need to store $C$ for sliced-based methods, that require a memory of $2n$ for $\min \mathrm{PS}$ and at most of $2Ln$ for ES. The time complexities of all flow variants are proportional to the number of flow steps, and we notice that all sliced methods have comparable complexities in $\mathcal{O}(Lnd + Ln \log(n))$, which is substantially advantageous compared to the $\mathcal{O}(n^3 \log n)$ complexity of standard OT.
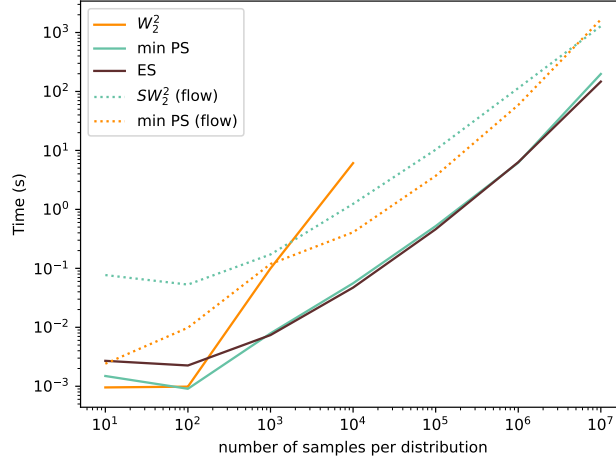


Figure 19: Running time comparisons of different methods for varying number of samples $n$.

## 8.2   Illustration on Colour Transfer

Colour transfer consists in transferring the colour distribution of a source image onto a target image, while preserving the structure of the source. We see an RGB image $I \in \mathbb{R}^{w \times h \times 3}$ as the uniform measure of its pixels in the RGB space $\mu_I := \frac{1}{wh} \sum_{i=1}^{w} \sum_{j=1}^{h} \delta_{I_{i,j,\cdot}} \in \mathcal{P}(\mathbb{R}^3)$. Given a source image $I$ and a target image $J$ of same size, our objective is to match (in a certain sense) each pixel $(i, j)$ of $I$ to a pixel $(i', j')$ of $J$. We consider three different approaches: first, we compute a permutation which is (approximately) optimal for the $\min \mathrm{PS}$ discrepancy, approximated by searching over $L = 50$ directions. Using this permutation, we replace each pixel of $I$ by its corresponding pixel in $J$. Second, we approximate the Expected Sliced plan by averaging over $L = 50$ directions. Since this does not yield a permutation but only a transport plan $\overline{\gamma}$, we use the barycentric projection (i.e. conditional expectation) of $\overline{\gamma}$, which provides only an approximate matching to $\mu_J$. Finally, we compare these methods with the Sliced Wasserstein (SW) flow proposed in [Rab+12a], which operates 10 steps of Stochastic Gradient Descent with a learning rate of 1 on $X \longmapsto \mathrm{SW}_2^2(\mu_X, \mu_J)$ initialised at $X_0 := I$ and sampling a batch of 3 orthonormal directions at each step. Note that while the final iteration is expected to verify $\mu_X \approx \mu_J$, it may not be the case in practice depending on the hyperparameter choices. We report our results on three different image pairs in Figs. 20 to 22.



Figure 20: Colour transfer example on images of size $1000 \times 669$.

Figure 21: Colour transfer example on images of size $1280 \times 1024$.



Figure 22: Colour transfer example on images of size $500 \times 500$.

In Fig. 20, the source and target images are relatively monochrome, which makes the colour transfer task easier. We observe that the Pivot-Sliced and SW methods are comparable, while the Expected Sliced results in duller colours. Contrastingly, in Fig. 21, the colour palettes are more diverse and Pivot-Sliced yields a visually worse result than SW, while SW matches the colour distributions less faithfully, with some artifacts in the sky. As for Expected Sliced, the results are again duller and quite different to the target colour distribution. Finally, in Fig. 22, only the SW method produces visually consistent results, the matching provided by min PS and ES fail to preserve sufficient spatial structure, in particular in the green colours. Overall, while the plan associated to min PS can suffice in practice, it appears that iterative methods such as the SW flow are better suited for this task. Our experiments suggest that the barycentric projection of the Expected Sliced plan does not provide a sound transportation.

## 8.3   Experiments on a Shape Registration Task

We now consider a shape registration task, with a rigid transformation that involves a translation and a rotation. Most approaches to solve this problem are concerned with finding the right correspondences between the points. For instance, the Iterative Closest Point (ICP) algorithm [BM92] relies on nearest neighbour correspondences, considering the Euclidean distance between points. Optimal transport is now a workhorse for this task, as it provides a principled way to find correspondences between two point clouds; see [BD23] for a review of OT-based methods for point cloud registration. We here evaluate the performance of the sliced-based methods, namely min PS and ES, in this context. We compare them to the 2-Wasserstein distance, which is a standard benchmark for point cloud registration, and also to Sliced Wasserstein, using a gradient flow as described in Section 8.1.1 to get an approximated transport plan. Note that Expected Sliced does not provide a one-to-one correspondence but they can be inferred from the blurred transport plan [Sol+15].

We consider two point clouds of 3D shapes, which are subsampled from the *bunny* and *armadillo* shapes of the *open3d* library [ZPK18]. We first subsample both shapes with $n = 500$ points, and then we apply 10 different rigid transformations to the source shape to get the target shape. We then run the ICP algorithm, with several alignment methods: the nearest neighbour correspondence, Wasserstein, Sliced Wasserstein, min-Pivot Sliced and Expected Sliced, to realign the two shapes. Fig. 23 presents the two shapes, subsampled with $n = 2000$ points for visualization purposes. The second line presents the Wasserstein distance between the (registered) source and target point clouds along the iterations. One can notice that Min-Pivot Sliced yields the best registration among all

methods: this conclusion was also reached by [Mah+23] who conjecture that it allows exiting local minima of the ICP algorithm by finding an approximated matching.

We also consider the case where the shapes are not subsampled, which is a more computationally challenging setup, especially for the armadillo shape. One can drawn similar conclusions, with min PS yielding the best registration, with little variation around the different repetitions of the experiment.
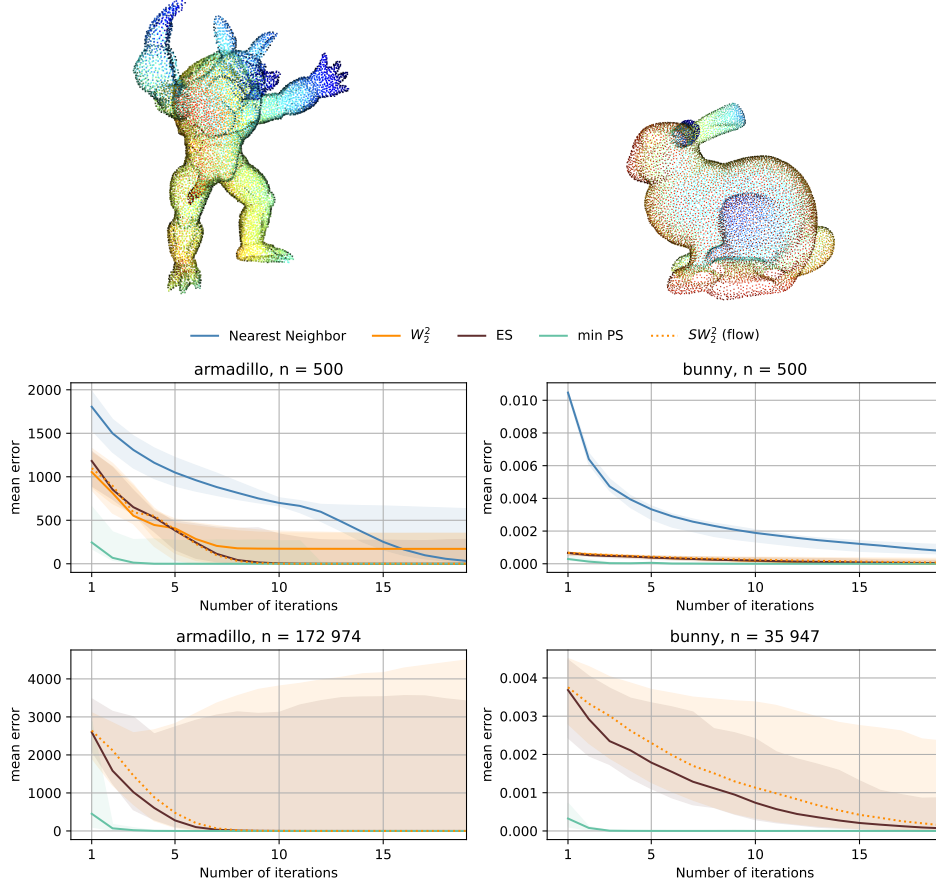


Figure 23: Evolution of the loss along the iterations of the ICP algorithm. The loss is computed as the mean square distance between each target sample and the registered source. The first column corresponds to the results for the armadillo shape, while the second column corresponds to the bunny shape.

## Acknowledgements

# References

[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.

[ACB17]    Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks". In: *International conference on machine learning.* PMLR. 2017, pp. 214–223.

[BT97]     Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization.* Vol. 6. Athena scientific Belmont, MA, 1997.

[BM92]     Paul J Besl and Neil D McKay. "Method for registration of 3-D shapes". In: *Sensor fusion IV: control paradigms and data structures.* Vol. 1611. Spie. 1992, pp. 586–606.

[Bir46]    Garrett Birkhoff. "Three observations on linear algebra". In: *Univ. Nac. Tacuman, Rev. Ser. A* 5 (1946), pp. 147–151.

[BDC25]    Clément Bonet, Lucas Drumetz, and Nicolas Courty. "Sliced-Wasserstein distances and flows on Cartan-Hadamard manifolds". In: *Journal of Machine Learning Research* 26.32 (2025), pp. 1–76.

[BD23]     Nicolas Bonneel and Julie Digne. "A survey of optimal transport for computer graphics and computer vision". In: *Computer Graphics Forum.* Vol. 42. 2. Wiley Online Library. 2023, pp. 439–460.

[Bon+15]   Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. "Sliced and Radon Wasserstein barycenters of measures". In: *Journal of Mathematical Imaging and Vision* 51.1 (2015), pp. 22–45.

[Bon13]    Nicolas Bonnotte. "Unidimensional and Evolution Methods for Optimal Transportation." In: *PhD Thesis, Paris 11* (2013).

[Bun+24]   Charlotte Bunne, Geoffrey Schiebinger, Andreas Krause, Aviv Regev, and Marco Cuturi. "Optimal transport for single-cell and spatial omics". In: *Nature Reviews Methods Primers* 4.1 (2024), p. 58.

[BDG12]    Giuseppe Buttazzo, Luigi De Pascale, and Paola Gori-Giorgi. "Optimal-transport formulation of electronic density-functional theory". In: *Physical Review A—Atomic, Molecular, and Optical Physics* 85.6 (2012), p. 062502.

[CTV25]    Laetitia Chapel, Romain Tavenard, and Samuel Vaiter. *Differentiable Generalized Sliced Wasserstein Plans.* 2025. arXiv: 2505.22049 [cs.LG].

[Che+]     Guangyi Chen, Weiran Yao, Xiangchen Song, Xinyue Li, Yongming Rao, and Kun Zhang. "PLOT: Prompt Learning with Optimal Transport for Vision-Language Models". In: *The Eleventh International Conference on Learning Representations.*

[CYL20]    Xiongjie Chen, Yongxin Yang, and Yunpeng Li. "Augmented Sliced Wasserstein Distances". In: *International Conference on Learning Representations.* 2020.

[Cou+16]   Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. "Optimal transport for domain adaptation". In: *IEEE transactions on pattern analysis and machine intelligence* 39.9 (2016), pp. 1853–1865.

[Cov67]    Thomas M Cover. "The number of linearly inducible orderings of points in d-space". In: *SIAM Journal on Applied Mathematics* 15.2 (1967), pp. 434–439.

[CS25]     Giacomo Cozzi and Filippo Santambrogio. "Long-time asymptotics of the sliced-wasserstein flow". In: *SIAM Journal on Imaging Sciences* 18.1 (2025), pp. 1–19.

[DZS18]    Ishan Deshpande, Ziyu Zhang, and Alexander G Schwing. "Generative modeling using the sliced wasserstein distance". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2018, pp. 3483–3491.

[DLV24]    Théo Dumont, Théo Lacombe, and François-Xavier Vialard. "On the existence of Monge maps for the Gromov–Wasserstein problem". In: *Foundations of Computational Mathematics* (2024), pp. 1–48.

[EW22]     Ariel Elnekave and Yair Weiss. "Generating natural images with direct patch distributions matching". In: *European Conference on Computer Vision.* Springer. 2022, pp. 544–560.

[EH13]       Paul Embrechts and Marius Hofert. "A note on generalized inverses". In: *Mathematical Methods of Operations Research* 77 (2013), pp. 423–432.

[Fat+21]     Kilian Fatras, Thibault Séjourné, Rémi Flamary, and Nicolas Courty. "Unbalanced mini-batch optimal transport; applications to domain adaptation". In: *International conference on machine learning*. PMLR. 2021, pp. 3186–3197.

[Fey+17]     Jean Feydy, Benjamin Charlier, François-Xavier Vialard, and Gabriel Peyré. "Optimal transport for diffeomorphic registration". In: *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*. Springer. 2017, pp. 291–299.

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

[Gal17]      Alfred Galichon. "A survey of some recent applications of optimal transport methods to econometrics". In: *The Econometrics Journal* 20.2 (2017), pp. C1–C11.

[Gul+17]     Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. "Improved training of Wasserstein GANs". In: *Advances in neural information processing systems* 30 (2017).

[Hei+21]     Eric Heitz, Kenneth Vanhoey, Thomas Chambon, and Laurent Belcour. "A sliced Wasserstein loss for neural texture synthesis". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 9412–9420.

[HCD25]      Johannes Hertrich, Antonin Chambolle, and Julie Delon. *On the Relation between Rectified Flows and Optimal Transport*. 2025. arXiv: 2505.19712 [cs.LG].

[HHR22]      Johannes Hertrich, Antoine Houdard, and Claudia Redenbach. "Wasserstein patch prior for image superresolution". In: *IEEE Transactions on Computational Imaging* 8 (2022), pp. 693–704.

[Hur08]      Glenn Hurlbert. "A short proof of the birkhoff-von neumann theorem". In: *preprint (unpublished)* (2008).

[KPS20]      Young-Heon Kim, Brendan Pass, and David J Schneider. "Optimal transport and barycenters for dendritic measures". In: *Pure and Applied Analysis* 2.3 (2020), pp. 581–601.

[Kol+19]     Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde. "Generalized sliced wasserstein distances". In: *Advances in neural information processing systems* 32 (2019).

[Kol+18]     Soheil Kolouri, Phillip E Pope, Charles E Martin, and Gustavo K Rohde. "Sliced Wasserstein auto-encoders". In: *International Conference on Learning Representations*. 2018.

[KRH18]      Soheil Kolouri, Gustavo K. Rohde, and Heiko Hoffmann. "Sliced Wasserstein Distance for Learning Gaussian Mixture Models". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2018.

[Le+24]      Tung Le, Khai Nguyen, Shanlin Sun, Nhat Ho, and Xiaohui Xie. "Integrating efficient optimal transport and functional maps for unsupervised shape correspondence learning". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 23188–23198.

[Lee+19]     Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. "Sliced wasserstein discrepancy for unsupervised domain adaptation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 10285–10295.

[LM25]       Shiying Li and Caroline Moosmueller. *Measure transfer via stochastic slicing and matching*. 2025. arXiv: 2307.05705 [math.NA].

[Liu+24]     Xinran Liu, Rocio Diaz Martin, Yikun Bai, Ashkan Shahbazi, Matthew Thorpe, Akram Aldroubi, and Soheil Kolouri. "Expected Sliced Transport Plans". In: *The Thirteenth International Conference on Learning Representations*. 2024.

[Mah+23]   Guillaume Mahey, Laetitia Chapel, Gilles Gasso, Clément Bonet, and Nicolas Courty. "Fast Optimal Transport through Sliced Generalized Wasserstein Geodesics". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine. Vol. 36. Curran Associates, Inc., 2023, pp. 35350–35385.

[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.

[NS25]     Navid NaderiAlizadeh and Rohit Singh. "Aggregating residue-level protein language model embeddings with optimal transport". In: *Bioinformatics Advances* 5.1 (2025), vbaf060.

[Nad+20]   Kimia Nadjahi, Alain Durmus, Lénaïc Chizat, Soheil Kolouri, Shahin Shahrampour, and Umut Simsekli. "Statistical and Topological Properties of Sliced Probability Divergences". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020, pp. 20802–20812.

[NP23]     Luca Nenna and Brendan Pass. "Transport type metrics on the space of probability measures involving singular base measures". In: *Applied Mathematics & Optimization* 87.2 (2023), p. 28.

[NNH23]    Khai Nguyen, Dang Nguyen, and Nhat Ho. "Self-attention amortized distributional projection optimization for sliced wasserstein point-cloud reconstruction". In: *International Conference on Machine Learning*. PMLR. 2023, pp. 26008–26030.

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

[Pey19]    Gabriel Peyré. "Course notes on Computational Optimal Transport". In: *Couse Notes* (2019).

[Pon+21]   Mathieu Pont, Jules Vidal, Julie Delon, and Julien Tierny. "Wasserstein distances, geodesics and barycenters of merge trees". In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021), pp. 291–301.

[RDG09]    Julien Rabin, Julie Delon, and Yann Gousseau. "A statistical approach to the matching of local features". In: *SIAM Journal on Imaging Sciences* 2.3 (2009), pp. 931–958.

[Rab+12a]  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.

[Rab+12b]  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.

[Row+19]   Mark Rowland, Jiri Hron, Yunhao Tang, Krzysztof Choromanski, Tamas Sarlos, and Adrian Weller. "Orthogonal Estimation of Wasserstein Distances". In: (Mar. 2019).

[Sal+18]   Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas. "Improving GANs Using Optimal Transport". In: *International Conference on Learning Representations*. 2018.

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

[SDT25]    Keanu Sisouk, Julie Delon, and Julien Tierny. *A User's Guide to Sampling Strategies for Sliced Optimal Transport*. 2025. arXiv: 2502.02275 [cs.LG].

[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.

[TFD24]   Eloi Tanguy, Rémi Flamary, and Julie Delon. "Properties of Discrete Sliced Wasserstein Losses". In: *Mathematics of Computation* (June 2024).

[Ton+24]  Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. "Improving and generalizing flow-based generative models with minibatch optimal transport". In: *Transactions on Machine Learning Research* (2024), pp. 1–34.

[Van00]   Aad W Van der Vaart. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.

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

[Wan+13]  Wei Wang, Dejan Slepčev, Saurav Basu, John A Ozolek, and Gustavo K Rohde. "A linear optimal transportation framework for quantifying and visualizing variations in sets of images". In: *International journal of computer vision* 101 (2013), pp. 254–269.

[Wu+19]   Jiqing Wu, Zhiwu Huang, Dinesh Acharya, Wen Li, Janine Thoma, Danda Pani Paudel, and Luc Van Gool. "Sliced wasserstein generative models". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 3713–3722.

[ZPK18]   Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. *Open3D: A Modern Library for 3D Data Processing*. 2018. arXiv: 1801.09847 [cs.CV].

# A    Appendix

## A.1    Ambiguity in SWGG from [Mah+23]

Let $\mu_1 = \frac{1}{n} \sum_i \delta_{x_i}$, $\mu_2 = \frac{1}{n} \sum_i \delta_{y_i}$, and $\theta \in \mathbb{S}^{d-1}$. Consider $\sigma_\theta$ a permutation which sorts $(\theta^\top x_i)_{i=1}^n$ and $\tau_\theta$ sorting $(\theta^\top y_i)_{i=1}^n$. The Sliced Wasserstein Generalised Geodesic distance ([Mah+23], Equation 8) is defined as

$$\mathrm{SWGG}_2^2(\mu_1, \mu_2, \theta) := \frac{1}{n} \sum_{i=1}^n \|x_{\sigma_\theta(i)} - y_{\tau_\theta(i)}\|_2^2. \tag{60}$$

We illustrate the coupling induced by $\mathrm{SWGG}_2^2(\mu_1, \mu_2, \theta)$ in Fig. 24:
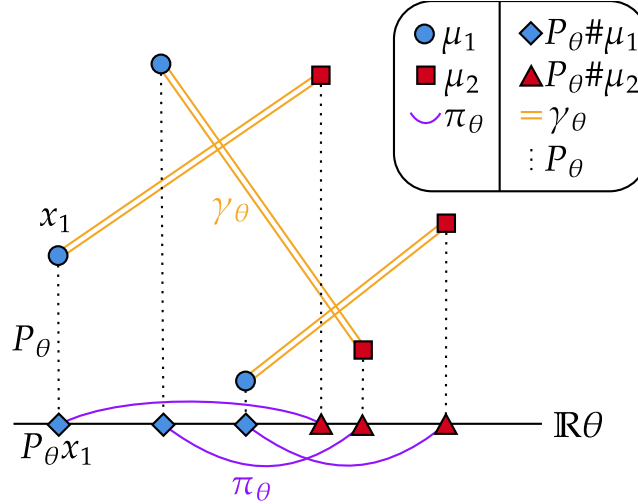


Figure 24: Coupling $\gamma_\theta \in \Pi(\mu_1, \mu_2)$ induced by $\mathrm{SWGG}_2^2(\mu_1, \mu_2, \theta)$ for $d = 2$, $n = 3$, $\theta = (1, 0)$. The support of the measure $\mu_1$ is represented by blue circles, and the support of $\mu_2$ with red squares. The projected measures $P_\theta\#\mu_1$ and $P_\theta\#\mu_2$ are represented by the blue diamonds and triangles respectively. The optimal coupling between $P_\theta\#\mu_1$ and $P_\theta\#\mu_2$ is drawn with purple curves, and the associated coupling $\gamma_\theta$ between $\mu_1$ and $\mu_2$ is represented by the orange double lines. In this example, the projections of the points of the support of $\mu_1$ are distinct (as for $\mu_2$), thus the coupling $\pi_\theta$ determines uniquely the coupling $\gamma_\theta$, there is no ambiguity.

Unfortunately, the RHS quantity in Eq. (60) depends on the choice of the permutations, rendering the quantity ill-defined, as showcased in Example 8.

**Example 8** (Ambiguity in SWGG). Consider $d = 2$, $n = 2$, the points $x_1 = (0, 1)$, $x_2 = (0, 0)$, $y_1 = (0, 0)$, $y_2 = (0, 1)$, the line $\theta = (1, 0)$ and the measures $\mu_1 = \frac{1}{2}(\delta_{x_1} + \delta_{x_2})$, $\mu_2 = \frac{1}{2}(\delta_{y_1} + \delta_{y_2})$. We have $\mu_1 = \mu_2$, and $\theta^\top u = 0$ for all points $u \in \{x_1, x_2, y_1, y_2\}$, hence any choice of permutations $(\sigma_\theta, \tau_\theta)$ sorts the respective points $(\theta^\top x_i)$ and $(\theta^\top y_i)$. Choosing $(\sigma_\theta, \tau_\theta) = (I, I)$, we obtain

$$\mathrm{SWGG}_2^2(\mu_1, \mu_2, \theta) = \tfrac{1}{2}(\|x_1 - y_1\|_2^2 + \|x_2 - y_2\|_2^2) = 1,$$

which in particular in non-zero, which shows that $\mathrm{SWGG}_2(\cdot, \cdot, \theta)$ is not a distance. Another possible choice $(\sigma_\theta, \tau_\theta) = ((2, 1), (2, 1))$ yields a value of 0.

One could consider the following "fix" to the permutation choice issue:

$$\widetilde{\mathrm{SWGG}}_2^2(\mu_1, \mu_2, \theta) := \min_{(\sigma_\theta, \tau_\theta) \in \mathfrak{S}_\theta(X,Y)} \frac{1}{n} \sum_{i=1}^n \|x_{\sigma_\theta(i)} - y_{\tau_\theta(i)}\|_2^2, \tag{61}$$

where $\mathfrak{S}_\theta(X, Y)$ is the set of pairs of permutations $(\sigma_\theta, \tau_\theta)$ that sort $(\theta^\top x_i)_{i=1}^n$ and $(\theta^\top y_i)_{i=1}^n$ respectively. We illustrate this idea in Fig. 25.
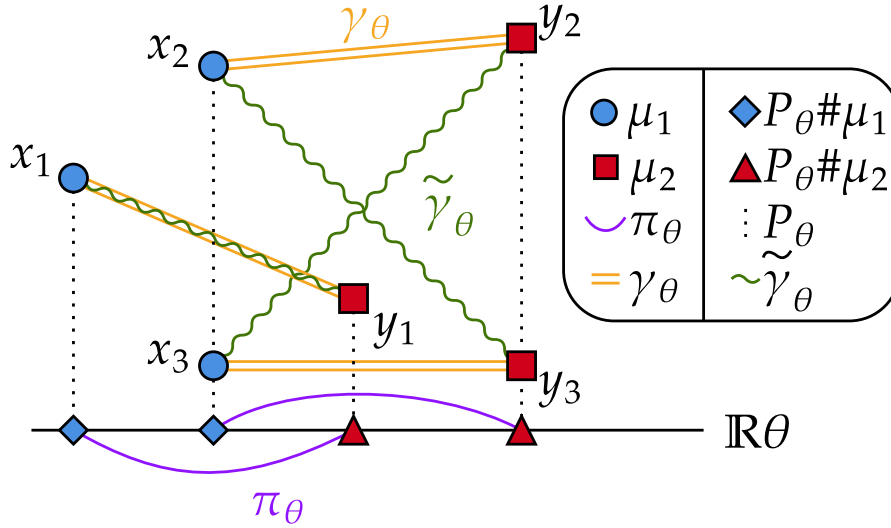
Figure 25: In this example, the projections sometimes coincide, and the optimal coupling $\pi_\theta$ between $P_\theta\#\mu_1$ and $P_\theta\#\mu_2$ does not determine the coupling between $(x_2, x_3)$ and $(y_2, y_3)$. In terms of permutations, there are two possibilities: $\gamma_\theta := \frac{1}{3}(\delta_{x_1\otimes y_1} + \delta_{x_2\otimes y_2} + \delta_{x_3\otimes y_3})$ displayed with orange double lines, and $\widetilde{\gamma}_\theta := \frac{1}{3}(\delta_{x_1\otimes y_1} + \delta_{x_2\otimes y_3} + \delta_{x_3\otimes y_2})$ represented by green squiggly lines. Here, the cost of $\gamma_\theta$ is lower, so we would choose it.

## A.2   Midpoints are Geodesic Middles

In the following, we remind a well-known simple result about geodesic spaces, which we apply to show that Wasserstein means are middles of Wasserstein geodesics (see Proposition 3). We consider a geodesic space $(\mathcal{X}, d)$, which is to say that $d$ is a distance on $\mathcal{X}$ such that for any $(x_1, x_2) \in \mathcal{X}^2$ there exists a curve $\gamma : [0, 1] \longrightarrow \mathcal{X}$ with $\gamma(0) = x_1$ and $\gamma(1) = x_2$ such that $d(\gamma(t), \gamma(s)) = |t - s| d(x_1, d_2)$. Such a curve is called a geodesic between $x_1$ and $x_2$.

**Lemma 12.** Let $(\mathcal{X}, d)$ be a geodesic space, let $x_1, x_2 \in \mathcal{X}$ and consider the set $\mathrm{M}(x_1, x_2)$ of Midpoints:

$$\mathrm{M}(x_1, x_2) = \operatorname*{argmin}_{y \in \mathcal{X}} d(x_1, y)^2 + d(y, x_2)^2. \tag{62}$$

This set is in fact exactly the set of middles of geodesics:

$$\mathrm{M}(x_1, x_2) = \left\{ \gamma(\tfrac{1}{2}) \mid \gamma \text{ is a geodesic between } x_1 \text{ and } x_2 \right\}. \tag{63}$$

*Proof.* Denote by $\mathrm{M}'(x_1, x_2)$ the RHS of Eq. (63), first we show $\mathrm{M}'(x_1, x_2) \subset \mathrm{M}(x_1, x_2)$ and compute the optimal value of Eq. (62). Let $\gamma$ a constant-speed geodesic between $x_1$ and $x_2$, we have

$$d(x_1, \gamma(\tfrac{1}{2}))^2 + d(\gamma(\tfrac{1}{2}), x_2)^2 = d(\gamma(0), \gamma(\tfrac{1}{2}))^2 + d(\gamma(\tfrac{1}{2}), \gamma(1))^2 = d(x_1, x_2)^2/2.$$

Now take any $y \in \mathcal{X}$, we have (by convexity of $t \longmapsto t^2$, then by the triangle inequality for $d$)

$$d(x_1, y)^2 + d(y, x_2)^2 = 2(d(x_1, y)^2/2 + d(y, x_2)^2/2) \tag{64}$$

$$\geq 2(d(x_1, y)/2 + d(y, x_2)/2)^2 \tag{65}$$

$$\geq d(x_1, x_2)^2/2. \tag{66}$$

This shows that any such $\gamma(\tfrac{1}{2})$ is solution of the optimisation problem which defines $\mathrm{M}(x_1, x_2)$, and thus $\mathrm{M}'(x_1, x_2) \subset \mathrm{M}(x_1, x_2)$. The value of the minimisation problem from Eq. (62) is $d(x_1, x_2)^2/2$.

Let $y^* \in \mathrm{M}(x_1, x_2)$, we now show that $d(x_1, y^*) = d(y^*, x_2) = d(x_1, x_2)/2$. Since $y^*$ is optimal and that the optimal value is $d(x_1, x_2)^2$, the inequalities Eq. (65) and Eq. (66) are equalities for $y := y^*$. First, Eq. (65) yields $d(x_1, y^*) = d(y^*, x_2)$, then Eq. (66) yields $d(x_1, y^*) = d(x_1, x_2)/2$.

We now show that $M(x_1, x_2) \subset M'(x_1, x_2)$: let $y^* \in M(x_1, x_2)$, consider $\gamma_1$ a geodesic from $x_1$ to $y^*$, and $\gamma_2$ a geodesic from $y^*$ to $x_2$. We introduce the curve

$$\gamma : \begin{cases} [0,1] & \longrightarrow & \mathcal{X} \\ t & \longmapsto & \begin{cases} \gamma_1(2t) & \text{if } t \in [0, \frac{1}{2}]; \\ \gamma_2(2t-1) & \text{if } t \in [\frac{1}{2}, 1]. \end{cases} \end{cases}$$

Our objective is to show that $\gamma$ is a geodesic from $x_1$ to $x_2$ (since $\gamma(\frac{1}{2}) = y^*$, this will show that $y^* \in M'(x_1, x_2)$). By construction $\gamma(0) = x_1$, $\gamma(1) = x_2$. Let $(t, s) \in [0, 1]^2$ with $t \leq s$, we want to prove $d(\gamma(t), \gamma(s)) = |s - t| d(x_1, x_2)$.

Firstly, we consider the case $(t, s) \in [0, \frac{1}{2}]^2$. In this case,

$$d(\gamma(t), \gamma(s)) = d(\gamma_1(2t), \gamma_1(2s)) = (2s - 2t)d(x_1, y^*) = (s - t)d(x_1, x_2),$$

where we used $d(x_1, y^*) = d(x_1, x_2)/2$, which we proved earlier for any optimal $y^*$. The case $(t, s) \in [\frac{1}{2}, 1]^2$ can be treated similarly.

Secondly, we assume $t \in [0, \frac{1}{2}]$ and $s \in [\frac{1}{2}, 1]$. We first prove $d(\gamma(t), \gamma(s)) \leq (s - t)d(x_1, x_2)$ using the triangle inequality and $d(x_i, y^*) = d(x_1, x_2)/2$ for $i \in \{1, 2\}$:

$$\begin{aligned} d(\gamma(t), \gamma(s)) &\leq d(\gamma(t), y^*) + d(y^*, \gamma(s)) \\ &= d(\gamma_1(2t), \gamma_1(1)) + d(\gamma_2(0), \gamma_2(2s - 1)) \\ &= (1 - 2t)d(x_1, y^*) + (2s - 1)d(y^*, x_2) \\ &= (s - t)d(x_1, x_2). \end{aligned}$$

For the converse inequality $d(\gamma(t), \gamma(s)) \geq (s - t)d(x_1, x_2)$, we apply the triangle inequality:

$$d(x_1, x_2) \leq d(x_1, \gamma(t)) + d(\gamma(t), \gamma(s)) + d(\gamma(s), x_2),$$

which yields:

$$\begin{aligned} d(\gamma(t), \gamma(s)) &\geq d(x_1, x_2) - d(\gamma_1(0), \gamma_1(2t)) - d(\gamma_2(2s - 1), x_2) \\ &= (1 - t - (1 - s))d(x_1, x_2) = (s - t)d(x_1, x_2). \end{aligned}$$

The case $s \in [0, \frac{1}{2}]$ and $t \in [\frac{1}{2}, 1]$ is done symmetrically and thus $d(\gamma(t), \gamma(s)) = |s - t| d(x_1, x_2)$, which shows that $y^* \in M'(x_1, x_2)$. We conclude that $M'(x_1, x_2) = M(x_1, x_2)$. $\square$

### A.3   Reminders on Disintegration of Measures

In Definition 10, we recall the definition of disintegration of measures with respect to a map (taken from [AGS05], Theorem 5.3.1). By slight abuse of notation, we will write $P^{-1}(y) := P^{-1}(\{y\})$ for a map $P : \mathcal{X} \longrightarrow \mathcal{Y}$ that need not be injective and $y \in \mathcal{Y}$.

**Definition 10.** Consider a Borel map $P : \mathcal{X} \longrightarrow \mathcal{Y}$ between Polish spaces $\mathcal{X}, \mathcal{Y}$ and $\mu \in \mathcal{P}(\mathcal{X})$. There exists a $P\#\mu$-almost-everywhere unique Borel family $(\mu^y)_{y \in \mathcal{Y}} \subset \mathcal{P}(\mathcal{X})$ of measures verifying $\mu^y(\mathcal{X} \setminus P^{-1}(y)) = 0$, and verifying the following identity against test functions $\phi \in \mathcal{C}_b^0(\mathcal{X})$:

$$\int_{\mathcal{X}} \phi(x) d\mu(x) = \int_{\mathcal{Y}} \left( \int_{P^{-1}(y)} \phi(x) d\mu^y(x) \right) d(P\#\mu)(y). \tag{67}$$

We will write Eq. (67) symbolically as:

$$\mu(dx) = (P\#\mu)(P(dx)) \, \mu^{P(x)}(dx). \tag{68}$$

For example, in the case $\mathcal{X} = \mathbb{R}^d \times \mathbb{R}^d$ and $P(y, x) = y$, the disintegration corresponds to the disintegration with respect to the first marginal $\nu$ of a coupling $\gamma \in \Pi(\nu, \mu)$. In this case, each measure $\gamma^y$ is a measure of $\mathcal{P}(\mathbb{R}^{2d})$ concentrated on the slice $\{y\} \times \mathbb{R}^d$, which is routinely identified as a measure on $\mathbb{R}^d$ in literature. This disintegration is written symbolically as $\gamma(dy, dx) = \nu(dy)\gamma^y(dx)$.

### A.4  Proof of the Disintegration Formula for $\nu$-based Wasserstein

In this section, we provide a proof to Theorem 1, and use the notation from the statement. Let $\rho \in \Gamma(\nu, \mu_1, \mu_2)$ (see Eq. (7)), we have

$$
\int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2) = \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho^y(x_1, x_2) \right) \mathrm{d}\nu(y)
$$
$$
\geq \int_{\mathbb{R}^d} \mathrm{W}_2^2(P_1 \# \rho^y, P_2 \# \rho^y) \mathrm{d}\nu(y), \tag{69}
$$

where we wrote the disintegration $\rho(\mathrm{d}y, \mathrm{d}x_1, \mathrm{d}x_2) = \nu(\mathrm{d}y)\rho^y(\mathrm{d}x_1, \mathrm{d}x_2)$. Note that by [AGS05] Lemma 12.4.7, the map $y \longmapsto \mathrm{W}_2^2(P_1 \# \rho^y, P_2 \# \rho^y)$ is Borel.

Now since $\rho \in \Gamma(\nu, \mu_1, \mu_2)$, we can write $P_{1,2}\#\rho =: \gamma_1 \in \Pi^*(\nu, \mu_1)$ and $P_{1,3}\#\rho =: \gamma_2 \in \Pi^*(\nu, \mu_2)$. It follows that for $\nu$-almost every $y \in \mathbb{R}^d$, we have for $i \in \{1, 2\}$ that $P_i \# \rho^y = \gamma_i^y$, where we disintegrated $\gamma_i(\mathrm{d}y, \mathrm{d}x) = \nu(\mathrm{d}y)\gamma_i^y(\mathrm{d}x)$ (for example by [AGS05] Lemma 5.3.2). Taking the infimum on $\rho$ on both sides yields

$$
\mathrm{W}_\nu^2(\mu_1, \mu_2) \geq \inf_{\gamma_i \in \Pi^*(\nu, \mu_i),\, i \in \{1, 2\}} \int_{\mathbb{R}^d} \mathrm{W}_2^2(\gamma_1^y, \gamma_2^y) \mathrm{d}\nu(y). \tag{70}
$$

Fixing $\gamma_i \in \Pi^*(\nu, \mu_i)$ for $i \in \{1, 2\}$, we now construct a 3-plan $\rho \in \Gamma(\nu, \mu_1, \mu_2)$ which attains the lower bound in Eq. (69). Consider the disintegrations $\gamma_i(\mathrm{d}y, \mathrm{d}x) = \nu(\mathrm{d}y)\gamma_i^y(\mathrm{d}x)$ for $i \in \{1, 2\}$. The two families $(\gamma_i^y)_{y \in \mathbb{R}^d}$ are Borel in $\mathcal{P}_2(\mathbb{R}^d)$, hence by [AGS05] Lemma 12.4.7, there exists a Borel family $(\rho^y)_{y \in \mathbb{R}^d}$ in $\mathcal{P}_2(\mathbb{R}^{2d})$ such that for all $y \in \mathbb{R}^d$, $\rho^y \in \Pi^*(\gamma_1^y, \gamma_2^y)$. Setting $\rho(\mathrm{d}y, \mathrm{d}x_1, \mathrm{d}x_2) := \nu(\mathrm{d}y)\rho^y(\mathrm{d}x_1, \mathrm{d}x_2)$ yields the desired 3-plan, since for $\nu$-almost every $y \in \mathbb{R}^d$, $\rho^y$ is an optimal transport plan between $\gamma_1^y$ and $\gamma_2^y$. We have shown that

$$
\forall \gamma_i \in \Pi^*(\nu, \mu_i),\ i \in \{1, 2\},\ \mathrm{W}_\nu^2(\mu_1, \mu_2) \leq \int_{\mathbb{R}^d} \mathrm{W}_2^2(\gamma_1^y, \gamma_2^y) \mathrm{d}\nu(y), \tag{71}
$$

which shows the equality in Eq. (19).

We finish by showing that the infimum in Eq. (19) is indeed attained. Note that having the weak convergence of plans $(\gamma_n) \in \Pi(\nu, \mu_1)$ does not yield the $\nu$-almost-everywhere convergence of the disintegrations $\gamma_n^y$ in general. Thankfully, we can leverage the existence of a solution of the original formulation from Eq. (12) by Proposition 1. Using the fact that the two problems have the same value, we can take a solution $\rho$ of Eq. (12) and construct a solution of Eq. (19) by disintegration.