ESAIM: COCV 32 (2026) 19
https://doi.org/10.1051/cocv/2026004

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

A WASSERSTEIN-TYPE METRIC FOR GENERIC MIXTURE MODELS, INCLUDING LOCATION-SCATTER AND GROUP INVARIANT MEASURES

Geneviève Dusson[1,*], Virginie Ehrlacher[2] and Nathalie Nouaime[3]

Abstract. In this article, we study Wasserstein-type metrics and corresponding barycenters for mixtures of a chosen subset of probability measures called atoms hereafter. In particular, this works extends what was proposed by Delon and Desolneux [*SIAM J. Imaging Sci.* **13** (2020) 936–970] for mixtures of Gaussian measures to other mixtures. We first prove in a general setting that for a set of atoms equipped with a metric that defines a geodesic space, the set of mixtures based on this set of atoms is a also geodesic space for the defined modified Wasserstein metric. We then focus on two particular cases of sets of atoms: (i) the set of location-scatter atoms and (ii) the set of measures that are invariant with respect to some symmetry group. Both cases are particularly relevant for various applications among which electronic structure calculations. Along the way, we also prove some sparsity and symmetry properties of optimal transport plans between measures that are invariant under some well-chosen symmetries.

Mathematics Subject Classification. 65D05, 65K10, 41A05, 41A63, 46G99, 46T12, 60B05, 47N50.

Received November 14, 2025. Accepted January 7, 2026.

1. Introduction

The original motivation of this work stems from electronic structure calculations in quantum chemistry, which are widely used for the simulation of molecular and material systems [1, 2]. Permutation-invariant probability measures naturally arise in this context, as the square modulus of wavefunctions and their marginals satisfy such invariance property, due to the indistinguishability of bosonic and fermionic particles, *e.g.* electrons. In general, these objects are moreover parametrized by the positions of the nuclei in the molecular system, and efficient ways to interpolate between such objects when the positions of the nuclei are changing would be extremely beneficial to this field, where the computations at stake are numerically very expensive, as they involve the resolution of high-dimensional and/or nonlinear eigenvalue partial differential equations. Due to the localized nature of the considered probability measures, optimal transport seems a natural way to interpolate between

Keywords and phrases: Optimal transport, mixture, Wasserstein distance, Wasserstein barycenters

[1] Université Marie et Louis Pasteur, CNRS, LmB (UMR 6623), F-25000 Besançon, France.
[2] CERMICS, École des Ponts & Inria Paris, France.
[3] CEA Saclay, DEN/DM2S/STMF/LMSF, France.

* Corresponding author: genevieve.dusson@math.cnrs.fr

© The authors. Published by EDP Sciences, SMAI 2026

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

them using Wasserstein barycenters developed by Carlier and Agueh [3], as shown in [4] for a toy model with one-dimensional particles. However due to the high-dimensionality of the considered measures for real systems (*e.g.* already 3-dimensional for the electronic density, and 6-dimensional for the electronic pair density), this does not seem feasible in practice using standard optimal transport algorithms, such as the now widely used Sinkhorn algorithm [5], see [6] for a monograph on the numerical aspects of optimal transport describing this algorithm.

Luckily, in [7], Delon and Desolneux proposed a Wasserstein-type distance and interpolation scheme based on the decomposition of probability measures as mixtures of Gaussian distributions. The modified optimal transport problem expressed in this context becomes independent of the underlying dimension of the distribution, or of its discrete spatial representation, so that the problem becomes extremely cheap to solve when the considered measures can be decomposed as convex combinations of a few Gaussian distributions. However, we encounter two limitations when trying to apply this strategy in the context of electronic structure calculations. First, for some models, the wave-functions are not as regular as Gaussians, and cusps are to be found around the nuclei positions in the molecular system. Therefore, the wave-function or its marginals would be better represented as a mixture of Slater-type functions, *i.e.* based on $\exp(-\alpha|x|)$, see [8], Figure 2 for an example on the H_2^+ molecule which is composed of two protons and one electron. Unfortunately these functions do not fit in the framework of [7]. Second, when considering fermions, the wave-function is anti-symmetric, so that its modulus square as well as its marginals is not only permutation symmetric, but also contains constraints coming from the anti-symmetry property. For example, the modulus square of the wave-function is zero when two variables are equal. This feature cannot be satisfied as such with convex combinations of strictly positive measures. Note that generic distributions as well as group symmetric mixtures appear in other contexts, such as portfolio theory in finance [9], or in image analysis where the described objects are defined up to rigid movements [10].

In this work, we therefore extend the theory presented in [7, 11] to generic mixture models. This allows us to consider both mixtures based on different distributions such as any elliptical distribution, but also any distribution transported with an affine map. These are the main cases where the Wasserstein barycenters between two different individual measures composing the mixtures (later called atoms) can be explicitely computed. We then provide various numerical results involving different types of distributions, as well as group-invariant measures for permutation with or without underlying antisymmetry, as well as rotation-invariant measures. Our main contributions in this article are listed below:

1. We prove in a general setting that for a set of atoms equipped with a metric that defines a geodesic space, the set of mixtures based on this set of atoms is also a geodesic space for the defined modified Wasserstein metric.
2. We provide numerical results for the modified Wasserstein barycenters in the case of location-scatter atoms and group-invariant measures.
3. We prove some sparsity and symmetry properties of optimal transport plans between measures that are invariant under some specific symmetries, namely the permutation of the variables, as well as a permutation-invariance arising from antisymmetry.

We leave the practical application of this theory to electronic structure calculations to a further work.

The rest of the article is organised as follows. In Section 2, we provide a few preliminaries on Wasserstein distances and barycenters that will be used in the subsequent sections. In Section 3, we give general conditions on a set of atoms for a modified Wasserstein distance defined on the associated set of mixtures to be computable by means of a sparse discrete optimal transport problem in the spirit of [7], and for the set of mixtures to be a geodesic space. We then focus our attention to two particular cases of interest: the case of location-scatter atoms in Section 4, and the case of symmetry group invariant measures in Section 5. We present an application of the theoretical concepts developped in this work related to electronic structure calculations in quantum chemistry in Section 6. Along the way, in Appendix A, we gather some properties satisfied by the exact optimal transport plan relative to the Wasserstein metric, as well as Wasserstein barycenters, between some particular group invariant measures.

2. Preliminaries on Wasserstein distances and barycenters

To start with, we introduce in this section a few objects related to Wasserstein metric and barycenters, that will be useful in the subsequent sections of this article, see for instance [6, 12, 13] for references. Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be a Borelian set. We denote by $\mathcal{P}(\Omega)$ the set of probability measures on Ω.

2.1. Wasserstein space and distance

Let $c : \Omega \times \Omega \to \mathbb{R}^+$ be a metric on Ω. For $p > 1$, we denote by $\mathcal{P}_p^c(\Omega)$ the set of probability measures on Ω with finite p^{th} moments, *i.e.*

$$\mathcal{P}_p^c(\Omega) := \left\{ \mu \in \mathcal{P}(\Omega), \quad \exists x_0 \in \Omega, \int_\Omega c(x, x_0)^p \mathrm{d}\mu(x) < +\infty \right\}.$$

The pth-Wasserstein distance between $\mu, \nu \in \mathcal{P}_p^c(\Omega)$ associated to c is defined as

$$W_p^c(\mu_0, \mu_1) := \left(\inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\Omega \times \Omega} c(x, y)^p \, \mathrm{d}\gamma(x, y) \right)^{1/p}, \tag{2.1}$$

where $\Pi(\mu_0, \mu_1)$ denotes the set of measures on $\Omega \times \Omega$ with marginals μ_0 and μ_1, also called the set of transport plans between μ_0 and μ_1.

An important particular example is the case where the metric c is the Euclidean distance on Ω, *i.e.* when

$$\forall x, y \in \Omega, \; c(x, y) := \|x - y\|. \tag{2.2}$$

In this case, we drop the superscript ^{c} in the notation W_p^c for the ease of readability. Then a Wasserstein distance of particular interest is the 2-Wasserstein distance with Euclidean distance on Ω defined by

$$W_2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \left(\int_{\Omega \times \Omega} \|x - y\|^2 \, \mathrm{d}\gamma(x, y) \right)^{1/2}. \tag{2.3}$$

The space $\mathcal{P}_2(\Omega)$ endowed with the distance W_2 is a metric space, usually called L^2-Wasserstein space (see [12] for more details). From [13],Theorem 1.17, there exists a unique optimal transport plan solution to the minimization problem (2.3) denoted by γ in the sequel, provided that μ_0 is absolutely continuous. Also, the optimal transport plan γ has the following form

$$\gamma = (\mathrm{Id}, T)\#\mu_0,$$

where $T : \Omega \to \Omega$ is an application called the optimal transport map and satisfying $T\#\mu_0 = \mu_1$. Here, we denote by $T\#\mu$ the push-forward measure of a measure μ on Ω by a map $T : \Omega \to \Omega$, that is the measure ν on Ω such that $\forall A \subset \Omega$, $T\#\mu(A) = \mu(T^{-1}(A))$. Similar results are available for more generic cost functions c, in particular when for all $x, y \in \Omega$, $c(x, y) = h(x - y)$ with h stricly convex and Ω compact [13].

The path $(\mu_t)_{t \in [0,1]}$ given by

$$\forall t \in [0, 1], \quad \mu_t = P_t\#\gamma = ((1 - t)\mathrm{Id} + tT)\#\mu_0, \quad \forall (x, y) \in \Omega^2, \; P_t(x, y) := (1 - t)x + ty,$$

then defines a constant speed geodesic in $\mathcal{P}_2(\Omega)$ between μ_0 and μ_1. The path $(\mu_t)_{t \in [0,1]}$ is called the McCann interpolation between μ_0 and μ_1 [14]. It can be shown that for all $t \in [0, 1]$, μ_t can be equivalently expressed as

the unique solution of the following minimization problem

$$\mu_t := \underset{\rho\in\mathcal{P}_2(\Omega)}{\operatorname{argmin}}\ (1-t)W_2^2(\rho,\mu_0) + tW_2^2(\rho,\mu_1). \tag{2.4}$$

2.2. Wasserstein barycenters

We next recall the notion of barycenters in the Wasserstein space which was introduced in [3] and can be seen as an extension of the McCann interpolation to a family of more than two measures. Let $n \in \mathbb{N}^*$ and let

$$\mathcal{L}_n := \left\{ \boldsymbol{t} := (t_1, \cdots, t_n) \in [0,1]^n, \quad \sum_{i=1}^n t_i = 1 \right\}$$

be the probability simplex of dimension $n-1$. For any family of probability measures $\boldsymbol{\mu} = (\mu_i)_{1\le i\le n} \in (\mathcal{P}_p^c(\Omega))^n$ and barycentric weights $\boldsymbol{t} = (t_i)_{1\le i\le n} \in \mathcal{L}_n$, if one of the measures μ_i has a density, there exists a unique minimizer to the problem

$$\inf_{\rho\in\mathcal{P}_p^c(\Omega)} \sum_{i=1}^n t_i W_p^c(\rho,\mu_i)^p, \tag{2.5}$$

which is the barycenter of the family of measures $\boldsymbol{\mu}$ with barycentric weights $\boldsymbol{t}$. The solutions of the barycenter problem are moreover related to the solutions of the following multi-marginal optimal transport problem [15]

$$W_p^c(\mu_1,\dots,\mu_n)^p := \inf_{\gamma\in\Pi(\mu_1,\dots,\mu_n)} \int_{\Omega^n} \frac{1}{2} \sum_{i,j=1}^n t_i t_j c(x_i,x_j)^p \,\mathrm{d}\gamma(x_1,\dots,x_n), \tag{2.6}$$

where $\Pi(\mu_1,\dots,\mu_n)$ denotes the set of probability measures on Ω^n having $\mu_1,\dots,\mu_n$ as marginals. In particular, at least in the case of the Euclidean distance (2.2) if Ω is a convex set and if (2.6) has a unique solution γ^*, there exists a unique solution ρ^* to (2.5) given by $\rho^* = B\#\gamma^*$, with $B : \Omega^n \to \Omega$ defined by $B(x_1,\dots,x_n) := \sum_{i=1}^n t_i x_i$ for all $(x_1,\dots,x_n) \in \Omega^n$ and the infima of (2.5) and (2.6) are equal.

2.3. Case of Gaussian distributions

To conclude this section, we recall some well-known results about Gaussian distributions for which many quantities introduced above can be made explicit. Here, $\Omega = \mathbb{R}^d$ and we denote by $\mathcal{S}_d$ the set of symmetric positive definite matrices of $\mathbb{R}^{d\times d}$. For any $m \in \mathbb{R}^d$ and $\Sigma \in \mathcal{S}_d$, the (normalized) Gaussian distribution $g_{m,\Sigma} \in \mathcal{P}_2(\mathbb{R}^d)$ is defined by

$$g_{m,\Sigma}(\,\mathrm{d}x) = G_{m,\Sigma}(x)\,\mathrm{d}x,$$

where

$$\forall x \in \mathbb{R}^d, \quad G_{m,\Sigma}(x) := \frac{1}{(2\pi)^{d/2}\sqrt{\det\Sigma}} \exp\left((x-m)^T\Sigma^{-1}(x-m)\right).$$

For any $m_0, m_1 \in \mathbb{R}^d$ and $\Sigma_0, \Sigma_1 \in \mathcal{S}_d$, denoting by $g_0 := g_{m_0,\Sigma_0}$ and $g_1 := g_{m_1,\Sigma_1}$, there holds

$$W_2^2(g_0,g_1) = \|m_0 - m_1\|^2 + \operatorname{Tr}\left(\Sigma_0 + \Sigma_1 - 2\left(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2}\right)^{1/2}\right).$$

Finally, the Wasserstein barycenter between two and more generally n Gaussian measures is also a Gaussian measure. Precisely, for $\boldsymbol{t} = (t_i)_{1\leq i\leq n} \in \mathcal{L}_n$, the unique minimizer of (2.5) is given by

$$g_{\boldsymbol{t}}(\,\mathrm{d}x) = G_{m_{\boldsymbol{t}},\Sigma_{\boldsymbol{t}}}(x)\,\mathrm{d}x, \tag{2.7}$$

where $m_{\boldsymbol{t}}$ and $\Sigma_{\boldsymbol{t}}$ satisfy

$$m_{\boldsymbol{t}} = \sum_{i=1}^{n} t_i m_i, \quad \Sigma_{\boldsymbol{t}} = \sum_{i=1}^{n} t_i \left(\Sigma_{\boldsymbol{t}}^{1/2}\Sigma_i\Sigma_{\boldsymbol{t}}^{1/2}\right)^{1/2}. \tag{2.8}$$

3. Wasserstein-type distance and barycenters for generic mixtures

In this section, we follow the works [7, 16] which proposed a modified Wasserstein distance for Gaussian mixtures, and we generalize the theory to a larger setting. We exhibit sufficient conditions on the densities constituting the mixtures and the definition of the modified Wasserstein distance on mixtures to define a metric and geodesic space. We then prove a similar result as in [7], which is the equivalence between a discrete optimisation problem and a continuous one in the case where the considered mixtures are identifiable.

3.1. Distance and barycenters between generic mixtures

We start by defining a set of particular densities, which we call atoms, and mixtures thereof.

Definition 3.1 ($\mathcal{A}$-mixture)**.** Let $\mathcal{A}$ be a subset of $\mathcal{P}(\Omega)$, called dictionnary of atoms hereafter. We denote by $\mathcal{M}(\mathcal{A})$ the set of finite mixtures of atoms $\mathcal{A}$, *i.e.* the set of probability measures μ of $\mathcal{P}(\Omega)$ such that there exists $K \in \mathbb{N}^*$, $\boldsymbol{a} := (a^1, \cdots, a^K) \in \mathcal{A}^K$ and $\boldsymbol{\lambda} := (\lambda^1, \cdots, \lambda^K) \in \mathcal{L}_K$ such that

$$\mu = \sum_{k=1}^{K} \lambda^k a^k.$$

Typical examples of dictionaries of atoms are

(i) the set of non-degenerate Gaussian measures *i.e.*

$$\mathcal{A}_g^d := \left\{g_{m,\Sigma},\ m \in \mathbb{R}^d,\ \Sigma \in \mathcal{S}_d\right\};$$

(ii) the set of possibly degenerate Gaussian measures *i.e.*

$$\overline{\mathcal{A}}_g^d := \left\{g_{m,\Sigma},\ m \in \mathbb{R}^d,\ \Sigma \in \overline{\mathcal{S}}_d\right\};$$

where $\overline{\mathcal{S}}_d$ is the set of symmetric positive (but not necessarily definite) matrices of $\mathbb{R}^{d\times d}$. The set $\mathcal{M}(\mathcal{A}_g^d)$ (respectively $\mathcal{M}\left(\overline{\mathcal{A}}_g^d\right)$) is then the set of finite Gaussian mixtures (respectively possibly degenerate Gaussian mixtures) of dimension d. Note that both $\mathcal{A}_g^d$ and $\overline{\mathcal{A}}_g^d$ are geodesic spaces when endowed with the metric W_2.

In order to define a distance on the set of mixtures as well as establish many properties on the space of mixtures which we study in the sequel, the mixtures need to be identifiable, that is they need to satisfy the following assumption.

Assumption 3.2 (Identifiability)**.** The set of mixtures $\mathcal{M}(\mathcal{A})$ is identifiable, that is given two mixtures $\mu_0 = \sum_{j=1}^{J} \lambda_0^j a_0^j$, $\mu_1 = \sum_{k=1}^{K} \lambda_1^k a_1^k$, where the atomic functions $(a_0^j)_{j=1,\ldots,J}$ are pairwise distinct, and similarly for

$(a_1^k)_{k=1,\dots,K}$, then $\mu_0 = \mu_1$ if and only if $J = K$, and the indices in the sums can be reordered such that all $\lambda_0^k = \lambda_1^k$ and $a_0^k = a_1^k$ for all $k = 1, \dots, K$.

A classical example of identifiable mixtures is the set of Gaussian mixtures $\mathcal{M}(\mathcal{A}_g^d)$, see *e.g.* [7], Appendix, as well as elliptical distributions defined below in (4.4). We then introduce a map $\delta : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_+$, which will need to satisfy the following assumption.

Assumption 3.3. The application $\delta : \mathcal{A} \times \mathcal{A} \to \mathbb{R}^+$ defines a metric on $\mathcal{A}$ and $(\mathcal{A}, \delta)$ is a geodesic space.

Definition 3.4 (Mixture distance)**.** Let $\mathcal{A} \subset \mathcal{P}(\Omega)$ be a dictionary of atoms, and let $\delta : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_+$ be a metric on $\mathcal{A}$. We assume that $\mathcal{M}(\mathcal{A})$ satisfies Assumption 3.2. For all $p > 1$, we define the application $\delta_{\mathcal{M},p} : \mathcal{M}(\mathcal{A}) \times \mathcal{M}(\mathcal{A}) \to \mathbb{R}_+$ as follows: for all $\mu_0 = \sum_{j=1}^J \lambda_0^j a_0^j \in \mathcal{M}(\mathcal{A})$ and $\mu_1 = \sum_{k=1}^K \lambda_1^k a_1^k \in \mathcal{M}(\mathcal{A})$,

$$\delta_{\mathcal{M},p}(\mu_0, \mu_1)^p := \min_{w:=(w_{jk})_{\substack{1\le j\le J,\\ 1\le k\le K}} \in \Pi(\Lambda_0,\Lambda_1)} \sum_{j=1}^J \sum_{k=1}^K w_{jk} \delta^p(a_0^j, a_1^k), \tag{3.1}$$

$$\text{with} \quad \Pi(\Lambda_0, \Lambda_1) := \Bigg\{ w := (w_{jk})_{\substack{1\le j\le J,\\ 1\le k\le K}} \in \mathbb{R}_+^{J\times K}, \\ \forall 1 \le j \le J, \sum_{k=1}^K w_{jk} = \lambda_0^j, \quad \forall 1 \le k \le K, \sum_{j=1}^J w_{jk} = \lambda_1^k \Bigg\}.$$

Remark 3.5. It is easily seen that there exists at least one minimizer $w^* := (w_{jk}^*)_{1\le j\le J, 1\le k\le K}$ to problem (3.1). However, let us point out here that uniqueness is not guaranteed.

Then, the following results can be easily proved.

Proposition 3.6. *For all $p > 1$, if $\delta : \mathcal{A} \times \mathcal{A} \to \mathbb{R}^+$ defines a metric, the application $\delta_{\mathcal{M},p}$ defines a metric on the set of $\mathcal{A}$-mixtures $\mathcal{M}(\mathcal{A})$.*

The proof of this result follows from [6], Proposition 2.2. It also follows from [16], Theorem 1 where this result was derived for the particular case of Gaussian mixtures. Therefore, we do not provide the proof here.

Proposition 3.7 (Geodesic space)**.** *Under Assumptions 3.2 and 3.3, the space $\mathcal{M}(\mathcal{A})$ equipped with the metric $\delta_{\mathcal{M},p}$ is a geodesic space.*

The proof of this result is similar to the proof given in [16], Section III therefore we do not detail it here. In the case where $(\mathcal{M}(\mathcal{A}), \delta_{\mathcal{M},p})$ is a geodesic space, we can express barycenters in terms of geodesics of atoms.

Corollary 3.8. *Let us assume that Assumptions 3.2 and 3.3 hold. Let $\mu_0 = \sum_{j=1}^J \lambda_0^j a_0^j$ and $\mu_1 = \sum_{k=1}^K \lambda_1^k a_1^k$ be two elements of $\mathcal{M}(\mathcal{A})$. The barycenters (for $\delta_{\mathcal{M},p}$) between μ_0 and μ_1 belong to $\mathcal{M}(\mathcal{A})$ and can be written as*

$$\forall t \in [0,1], \quad \mu_t = \sum_{j=1}^J \sum_{k=1}^K w_{jk}^* a_t^{j,k},$$

where $w^* = (w_{jk}^*)_{1\le j\le J, 1\le k\le K}$ is a solution to (3.1) and $(a_t^{j,k})_{t\in[0,1]}$ are constant speed geodesic between respectively a_0^j and a_1^k.

3.2. Particular case of Wasserstein metric

We focus here on the specific case where $\mathcal{A} \subset \mathcal{P}_p^c(\Omega)$ for some $p > 1$ and metric $c : \Omega \times \Omega \to \mathbb{R}^+$, and where the atom metric δ is defined by the Wasserstein metric W_p^c, *i.e.*

$$\forall a_0, a_1 \in \mathcal{A}, \quad \delta(a_0, a_1) = \left(\inf_{\gamma \in \Pi(a_0,a_1)} \int_{\Omega\times\Omega} (c(x,y))^p \mathrm{d}\gamma(x,y) \right)^{1/p}. \tag{3.2}$$

Let us point out that Assumption 3.3 is indeed satisfied in this setting. We first investigate two-marginal problems, and then show that the theory can easily be extended to multi-marginal problems.

3.2.1. Two-marginal problem

Let us show that the discrete problem (3.1) is in fact equivalent to a continuous problem, similar to the one presented in [7]. To obtain this equivalence, we need to make the following assumption, which is the existence of an optimal transport plan between any atoms, on top of the identifiability of the mixtures of atoms that is Assumption 3.2.

Assumption 3.9 (Existence of optimal transport plan)**.** The set of atoms $\mathcal{A}$ and the cost function c are such that for any $a_0, a_1 \in \mathcal{A}$, there exists at least one solution to the optimal transport problem

$$\inf_{\gamma \in \Pi(a_0,a_1)} \int_{\Omega\times\Omega} c(x,y)^p \mathrm{d}\gamma(x,y). \tag{3.3}$$

The set of minimizers of (3.3) is then denoted by $\Gamma^*_{a_0,a_1}$.

This assumption is for instance satisfied if Ω is compact and c is continuous, see *e.g.* [6], Remark 2.3. We then define the set of optimal transport plans of the atoms as well as mixtures thereof.

Definition 3.10 (Admissible set of atom transport plans)**.** A set of transport plans $\Gamma(\mathcal{A}) \subset \mathcal{P}(\Omega \times \Omega)$ is said admissible for the set of atoms $\mathcal{A}$ if and only if

$$\Gamma(\mathcal{A}) = \bigcup_{a_0,a_1 \in \mathcal{A}} \Gamma_{a_0,a_1},$$

where for all $a_0, a_1 \in \mathcal{A}$, Γ_{a_0,a_1} is a convex set such that

$$\Gamma^*_{a_0,a_1} \subset \Gamma_{a_0,a_1} \subset \Pi(a_0, a_1).$$

Definition 3.11 (Mixtures of admissible atom transport plans)**.** Let $\Gamma(\mathcal{A})$ be an admissible set of atom transport plans. We define $\Gamma_{\mathcal{M}}(\mathcal{A}) := \mathcal{M}(\Gamma(\mathcal{A}))$ as the set of finite mixtures of optimal transport plans of atoms $\Gamma(\mathcal{A})$, *i.e.* the set of probability measures γ of $\mathcal{P}(\Omega \times \Omega)$ such that there exists $K \in \mathbb{N}^*$, $\boldsymbol{\gamma} := (\gamma^1, \cdots, \gamma^K) \in \Gamma(\mathcal{A})^K$ and $\boldsymbol{\lambda} := (\lambda^1, \cdots, \lambda^K) \in \mathcal{L}_K$ such that

$$\gamma = \sum_{k=1}^{K} \lambda^k \gamma^k.$$

Then, we have the following result which extends the arguments presented in [7], Proposition 4. We present its proof in Appendix B for the sake of completeness.

Proposition 3.12. *Let $\Gamma(\mathcal{A})$ be an admissible set of atom transport plans. Let $\mu_0 = \sum_{j=1}^{J} \lambda_0^j a_0^j \in \mathcal{M}(\mathcal{A})$ and $\mu_1 = \sum_{k=1}^{K} \lambda_1^k a_1^k \in \mathcal{M}(\mathcal{A})$. Let us define*

$$\widetilde{\delta}_{\mathcal{M},p}(\mu_0, \mu_1) := \left(\inf_{\gamma \in \Pi(\mu_0,\mu_1) \cap \Gamma_{\mathcal{M}}(\mathcal{A})} \int_{\Omega \times \Omega} c(x,y)^p \, \mathrm{d}\gamma(x,y) \right)^{1/p}. \tag{3.4}$$

Then, under Assumptions 3.2 and 3.9, problem (3.4) is equivalent to (3.1), i.e. $\widetilde{\delta}_{\mathcal{M},p}(\mu_0, \mu_1) = \delta_{\mathcal{M},p}(\mu_0, \mu_1)$. Moreover, the sets of minimizers of (3.4) is equal to the set of measures γ which can be written as

$$\gamma = \sum_{j=1}^{J} \sum_{k=1}^{K} w_{jk}^* \gamma_{jk}, \tag{3.5}$$

with $w^* := (w_{jk}^*)_{1 \le j \le J, 1 \le k \le K}$ a minimizer of (3.1), and $\gamma_{jk} \in \Gamma^*_{a_0^j, a_1^k}$ is an optimal transport plan between a_0^j and a_1^k.

In the following, we will give practical examples where some admissible $\Gamma(\mathcal{A})$ can be made explicit. We would like to highlight the fact that the mixture distance defined in (3.4) does not depend on $\Gamma(\mathcal{A})$ provided that the conditions presented in Definition 3.10 hold, and that in practice, the knowledge of $\Gamma(\mathcal{A})$ is not needed to compute the mixture distance. Moreover, a direct consequence of Proposition 3.12 is that for any $\mu_0, \mu_1 \in \mathcal{M}(\mathcal{A})$, it holds that

$$\widetilde{\delta}_{\mathcal{M},p}(\mu_0, \mu_1) = \delta_{\mathcal{M},p}(\mu_0, \mu_1) \ge W_p(\mu_0, \mu_1).$$

Remark 3.13. In the case where $\mathcal{A}$ is the set of atomic Gaussian measures $\mathcal{A}_g^d$, the result of [7], Proposition 4 corresponds to Proposition 3.12 for a set of atomic transport plans chosen as

$$\Gamma(\mathcal{A}) := \bigcup_{a_0, a_1 \in \mathcal{A}_g^d} \left(\overline{\mathcal{A}}_g^{2d} \cap \Pi(a_0, a_1) \right) = \overline{\mathcal{A}}_g^{2d}.$$

Indeed for all $a_0, a_1 \in \mathcal{A}_g^d$, the set $\Gamma_{a_0,a_1} := \left(\overline{\mathcal{A}}_g^{2d} \cap \Pi(a_0, a_1) \right)$ is convex and such that

$$\{\gamma^*_{a_0,a_1}\} \subset \Gamma_{a_0,a_1} \subset \Pi(a_0, a_1).$$

Remark 3.14. In the work by Delon and Desolneux [7], it is mentioned that other distributions than Gaussians can be used, as long as they satisfy two conditions: the identifiability property (Asm. 3.2), and a marginal consistency, *i.e.* that transport plans are mixtures of 2-dimensional atoms. In fact, what we have shown in the previous section is that we do not need the marginal consistency, as we use for transport plans mixtures of transport plans between atoms. Interestingly, for Gaussians, the set of transport plans between atoms can be chosen as the set of (degenerate) Gaussians in dimension $2d$.

3.2.2. Multi-marginal problems

The theory naturally extends to multi-marginal problems. For this, we need to assume the existence of multi-marginal optimal transport plans. We then define the extensions of Definitions 3.10 and 3.11 to the multi-marginal case.

Assumption 3.15 (Existence of multi-marginal transport plan)**.** The set of atoms $\mathcal{A}$ and the metric c are such that for all $Q \geq 2$, for any $\boldsymbol{t} := (t_q)_{1\leq q\leq Q} \in \mathcal{L}_Q$ and for any $a_1, \cdots, a_Q \in \mathcal{A}$, there exists at least one solution to the multi-marginal optimal transport problem

$$\delta^{\boldsymbol{t}}_{Q,\mathcal{A},p}(a_1,\dots,a_Q) = \left(\inf_{\gamma\in\Pi(a_1,\dots,a_Q)} C^{\boldsymbol{t}}_Q(x_1,\dots,x_Q)\mathrm{d}\gamma(x_1,\dots,x_Q)\right)^{1/p}, \tag{3.6}$$

where for all $(x_1,\dots,x_Q) \in \Omega^Q$,

$$C^{\boldsymbol{t}}_Q(x_1,\dots,x_Q) := \frac{1}{2}\sum_{q=1}^{Q}\sum_{q'=1}^{Q} t_q t_{q'} c(x_q,x_{q'})^p. \tag{3.7}$$

The set of minimizers of (3.6) is then denoted by $\Gamma^{*,\boldsymbol{t}}_{a_1,\dots,a_Q}$.

Definition 3.16. We say that $\Gamma_Q(\mathcal{A}) \subset \mathcal{P}(\Omega^Q)$ is an admissible set of atom multi-marginal transport plans if

$$\Gamma_Q(\mathcal{A}) = \bigcup_{a_1,\cdots,a_Q\in\mathcal{A}} \Gamma_{a_1,\dots,a_Q},$$

where $\Gamma_{a_1,\dots,a_Q}$ is any convex set such that

$$\bigcup_{\boldsymbol{t}\in\mathcal{L}_Q} \Gamma^{*,\boldsymbol{t}}_{a_1,\dots,a_Q} \subset \Gamma_{a_1,\dots,a_Q} \subset \Pi(a_1,a_2,\dots,a_Q).$$

Definition 3.17. We define $\Gamma_{Q,\mathcal{M}}(\mathcal{A})$ as the set of finite mixtures of admissible atom multi-marginal optimal plans $\Gamma_Q(\mathcal{A})$, *i.e.* the set of probability measures of $\mathcal{P}(\Omega^Q)$ such that there exists $K \in \mathbb{N}^*$, $\boldsymbol{\gamma} := (\gamma^1,\cdots,\gamma^K) \in \Gamma_Q(\mathcal{A})^K$ and $\boldsymbol{\lambda} := (\lambda^1,\cdots,\lambda^K) \in \mathcal{L}_K$ such that

$$\gamma = \sum_{k=1}^{K} \lambda^k \gamma^k.$$

Now, let $Q \geq 2$ and for all $1 \leq q \leq Q$, let $K_q \in \mathbb{N}^*$, $\boldsymbol{\lambda}_q := (\lambda_q^{k_q})_{1\leq k_q\leq K_q} \in \mathcal{L}_{K_q}$ and $\mu_q := \sum_{k_q=1}^{K_q} \lambda_q^{k_q} a_q^{k_q} \in \mathcal{M}(\mathcal{A})$. For all $\boldsymbol{t} := (t_q)_{1\leq q\leq Q} \in \mathcal{L}_Q$, we define the multi-marginal transport problem as

$$\widetilde{\delta}^{\boldsymbol{t}}_{Q,\mathcal{M},p}(\mu_1,\dots,\mu_Q) = \left(\inf_{\gamma\in\Pi(\mu_1,\dots,\mu_Q)\cap\Gamma_{Q,\mathcal{M}}(\mathcal{A})} C^{\boldsymbol{t}}_Q(x_1,\dots,x_Q)\mathrm{d}\gamma(x_1,\dots,x_Q)\right)^{1/p}, \tag{3.8}$$

with $C^{\boldsymbol{t}}_Q$ defined by (3.7).

Using a similar approach as in the previous section, we prove the following result.

Proposition 3.18. *Under Assumption 3.2 and Assumption 3.15, it holds that problem* (3.8) *is equivalent to the following discrete problem*

$$\left(\delta^{\boldsymbol{t}}_{Q,\mathcal{M},p}(\mu_1,\dots,\mu_Q)\right)^p = \inf_{w\in\Pi(\boldsymbol{\lambda}_1,\dots,\boldsymbol{\lambda}_Q)} \sum_{k_1,k_2,\dots,k_Q=1}^{K_1,K_2,\dots,K_Q} w_{k_1,k_2,\dots,k_Q}\left(\delta^{\boldsymbol{t}}_{Q,\mathcal{A},p}(a_1^{k_1},a_2^{k_2},\dots,a_Q^{k_Q})\right)^p, \tag{3.9}$$

$$\textit{with } \Pi(\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_Q) := \Bigg\{ w \in \mathbb{R}_+^{K_1\times\ldots\times K_Q}, \\ \forall 1 \le q \le Q,\ \forall 1 \le k_q \le K_q, \sum_{k_1,\ldots,k_{q-1},k_{q+1},\ldots,k_Q=1}^{K_1,\ldots,K_{q-1},K_{q+1},\ldots,K_Q} w_{k_1,\ldots,k_Q} = \lambda_q^{k_q} \Bigg\}.$$

Moreover, denoting by $\mathcal{K} := \{1,\ldots,K_1\}\times\{1,\ldots,K_2\}\times\ldots\times\{1,\ldots,K_Q\}$*, any solution to* (3.8) *can be written as*

$$\gamma^* = \sum_{\mathbf{k}\in\mathcal{K}} w^*_{\mathbf{k}}\gamma^*_{\mathbf{k}}, \tag{3.10}$$

where $(w^*_{\mathbf{k}})_{\mathbf{k}\in\mathcal{K}}$ *is a solution to* (3.9) *and* $\gamma^*_{\mathbf{k}} \in \Gamma^{*,\boldsymbol{t}}_{a_1^{k_1},\ldots,a_Q^{k_Q}}$ *for all* $\mathbf{k} = (k_1,\ldots,k_Q) \in \mathcal{K}$*. As a consequence, any barycenter of* $(\mu_1,\cdots,\mu_Q)$ *with barycentric weights* $\boldsymbol{t}$ *can be written as*

$$\mathrm{bar}_{\boldsymbol{t}}(\mu_1,\ldots,\mu_Q) = \sum_{\mathbf{k}\in\mathcal{K}} w^*_{\mathbf{k}}\,\mathrm{bar}_{\boldsymbol{t}}(a_1^{k_1},\ldots,a_Q^{k_Q}), \tag{3.11}$$

where $\mathrm{bar}_{\boldsymbol{t}}(a_1^{k_1},\ldots,a_Q^{k_Q})$ *is the barycenter of* $(a_1^{k_1},\ldots,a_Q^{k_Q})$ *with barycentric weights* $\boldsymbol{t}$*. Finally, any solution* $(w^*_{\mathbf{k}})_{\mathbf{k}\in\mathcal{K}}$ *to* (3.9) *contains at most* $K_1+K_2+\ldots+K_Q-Q+1$ *nonzeros components.*

Proof. The proof for the equivalence between the discrete and continuous optimization problem is similar to the proof of Proposition 3.12. The structure of the barycenter is a direct consequence of equation (3.10). The maximum number of nonzeros components comes from the form of the discrete problem, which is a linear program with $K_1+K_2+\ldots+K_Q-Q+1$ affine constraints, therefore there exists a solution with at most $K_1+K_2+\ldots+K_Q-Q+1$ nonzero components, see [17], Theorem 2 or [18], Appendix A. □

3.3. General comments on mixture Wasserstein distances

In the end, as soon as Assumption 3.2 and Assumption 3.3 are satisfied for the atoms, we can define a geodesic space on mixtures of these atoms. Moreover, if the metric on the atoms corresponds to a Wasserstein distance for which there is existence of optimal transport plans (Asm. 3.9), then there is an equivalence between the discrete problem (3.1) and the continuous one (3.4). Let us highlight here that the obtained mixture distance then does not depend on the choice of admissible set of atom transport plans provided that it satisfies the conditions of Definition 3.10.

The above mentioned sufficient conditions are pretty simple. However, all calculations on mixtures include distance calculations on the set of atoms, as well as multi-marginal calculations for the computation of barycenters. Therefore, the efficiency of the mixture calculations will highly depend on the cost of computing distances and barycenters between atoms. In the best case scenario, these should be explicit. This is for example the case of Gaussian measures, which motivated the two contributions [7, 16].

For the sake of simplicity, in the rest of the paper, we will focus on the case of the 2-Wasserstein distance with quadratic cost (*i.e.* when $p=2$, $c(x,y)=\|x-y\|$) and $\Omega=\mathbb{R}^d$ for some $d\in\mathbb{N}^*$. The aim of the two next sections is to highlight how the general framework introduced in the previous section can be used for other sets of atoms than Gaussians, for which computing distances and barycenters is explicit or at least computationally efficient. We will focus here on two such families of atoms:

- location-scatter measures (which includes elliptic distributions) in Section 4;
- group invariant measures in Section 5.

In each case, we discuss about satisfying Assumptions 3.2, 3.3 and 3.9 and comment about the practical computation of the mixture distance for such atoms.

For several families of atoms fitting into the scope of Section 4 and Section 5, we also provide some illustrative numerical examples. All the numerical tests presented in this article have been implemented with the Julia language [19] and the Wasserstein barycenters for the W_2 metric have been computed with the Python optimal transport library POT [20]. All Wasserstein barycenters for the W_2 metric are computed using the log-sinkhorn algorithm to avoid numerical errors, with a regularization parameter of 10^{-4} and a maximum number of iterations of 10000. The code used for generating all the figures can be downloaded at https://github.com/dussong/W2_mixtures.jl/.

For one-dimensional distributions, we could also compute W_2 barycenters using the inverse cumulative distribution functions (icdf) of the considered probability measures, using the fact that the icdf of the W_2 barycenter is the convex combination of the icdfs of the probability measures. In practice, the numerical results using the icdfs were highly similar to the ones obtained with the Sinkhorn algorithm. The latter are thus not reported here. Moreover, the computation of the barycenter with this technique also depends on a grid parameter used to discretize the distributions.

The one-dimensional cases presented in the following sections are computed on a grid containing 200 points and the two-dimensional cases on a grid with 50 points per dimension.

In terms of computational cost, it is difficult to provide proper timings to compare the cost of computing the W_2 barycenters and the mixture baycenters based on the mixture metric denoted by $\delta_{\mathcal{M},p}$ as it highly depends on the choice of the grid, the smoothing parameter, and number of iterations for the Sinkhorn algorithm for the W_2 calculations. Note however that the cost of computing the mixture Wasserstein barycenters does not depend on any spatial grid, as the size of the problem is only related to the number of atoms in the considered mixtures.

Therefore, with the provided parameters, the order of magnitude for computing one mixture barycenter is less than 1ms while the computation of the W_2 barycenters is of the order of 10s for the one-dimensional cases and about 8 minutes for the two-dimensional cases, hence several order of magnitude more expensive than the mixture-based calculations.

4. Mixtures of location-scatter atoms

In this section, we start by stating results on location-scatter measures, which correspond to families of probability measures generated from affine transformations. In some specific cases described more precisely below, transport maps and Wasserstein barycenters are explicitely computable. We then turn to a few practical examples, including elliptical distributions and affine-generated measures. For the sake of simplicity, we focus in this section on the 2-Wasserstein distance with quadratic cost (*i.e.* when $p = 2$ and $c(x, y) = \|x - y\|$) and $\Omega = \mathbb{R}^d$.

4.1. Properties of location-scatter measures

Location-scatter measures are generated from affine transformations of a given probability measure. We define the corresponding set of atoms as follows.

Definition 4.1 (Location-scatter atoms)**.** We define the set of location-scatter atoms generated from $a \in \mathcal{P}_2(\mathbb{R}^d)$ as

$$\mathcal{A} := \left\{ T\#a, \quad T : x \in \mathbb{R}^d \mapsto Ax + b,\ A \in \mathcal{S}_d,\ b \in \mathbb{R}^d \right\}. \tag{4.1}$$

We then formulate a generic classical result on location-scatter measures, which includes an explicit expression for the Wasserstein distance. This result is a rewriting in our context of [21], Theorem 2.3, itself based on [22], Theorem 2.1. We refer the reader to [23], Theorem 7.2.1 for its proof.

Theorem 4.2. *Let $\mathcal{A}$ be a set of location-scatter atoms generated from $a \in \mathcal{P}_2(\mathbb{R}^d)$ in the sense of Definition 4.1. Let us assume that the measure a has mean $m_a \in \mathbb{R}^d$ and a covariance matrix $\Sigma_a \in \mathcal{S}_d$. Let $a_0, a_1 \in \mathcal{A}$ be such that there exist $T_0, T_1 : \mathbb{R}^d \to \mathbb{R}^d$ with $T_0(x) = A_0 x + b_0$, $T_1(x) = A_1 x + b_1$, $A_0, A_1 \in \mathcal{S}_d$ and $b_0, b_1 \in \mathbb{R}^d$ such that $a_0 = T_0 \# a$, $a_1 = T_1 \# a$. Then, a_0 and a_1 have respectively means m_i and covariance Σ_i for $i = 0, 1$ defined by*

$$m_i = A_i m_a + b_i, \qquad \Sigma_i = A_i \Sigma_a A_i^T.$$

Moreover, the Wasserstein distance squared between a_0 and a_1 satisfies

$$W_2^2(a_0, a_1) = \|m_0 - m_1\|^2 + \mathrm{Tr}\left(\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}\right), \tag{4.2}$$

and the optimal transport map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T \# a_0 = a_1$ is given by

$$Tx = Ax + (m_0 - m_1), \quad \text{with } A = \Sigma_0^{-1/2}\left(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2}\right)^{1/2}\Sigma_0^{-1/2}. \tag{4.3}$$

A nice example of location-scatter atoms is the case of a generative atom a that is an absolutely continuous radially symmetric with finite second-order moments probability measure on $\mathbb{R}^d$, *i.e.* with an associated density μ satisfying $\mu(x) = \xi(|x|)$ for some function $\xi \in L^1(\mathbb{R}^d)$ [23], Theorem 7.2.1. This encompasses the case of elliptical distributions defined by

$$\forall x \in \mathbb{R}^d, \quad f_{m,\Sigma}(x) = \frac{1}{Z_{h,\Sigma,m}} h((x-m)^T\Sigma^{-1}(x-m)), \tag{4.4}$$

for $m \in \mathbb{R}^d$, $\Sigma \in \mathcal{S}_d$, where the generative atom a can be taken as f_{0,I_d}, denoting by I_d the identity matrix of size d.

We can express a similar theorem for the multi-marginal optimal transport problem.

Theorem 4.3. *Let $\mathcal{A}$ be a set of location-scatter atoms generated from $a \in \mathcal{P}_2(\mathbb{R}^d)$ in the sense of Definition 4.1. Let us assume that the measure a has a mean $m_a \in \mathbb{R}^d$ and a covariance matrix $\Sigma_a \in \mathcal{S}_d$. Let $a_1, a_2, \ldots, a_Q \in \mathcal{A}$ be such that there exist $T_1, \ldots, T_Q : \mathbb{R}^d \to \mathbb{R}^d$ with $T_q(x) = A_q x + m_q$, $A_q \in \mathcal{S}_d$ and $m_q \in \mathbb{R}^d$ such that for $q = 1, \ldots, Q$, $a_q = T_q \# a$, and a_q has mean m_q and covariance matrix Σ_q.*

The multi-marginal optimal transport problem with parameters $\boldsymbol{t} = (t_1, \ldots, t_Q)$ has for minimal cost

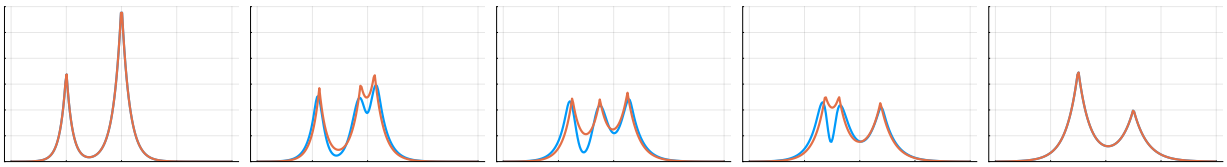
$$\delta^{\boldsymbol{t}}_{Q,\mathcal{A},2}(a_1, \ldots, a_Q) = \sum_{q=1}^{Q} t_q W_2^2(a_q, \bar{a}_{\boldsymbol{t}}),$$

where $\bar{a}_{\boldsymbol{t}} = T_{\boldsymbol{t}} \# a$ is the Wasserstein barycenter with $T_{\boldsymbol{t}} = Sx + m$, where S is the only positive definite matrix satisfying

$$S = \sum_{q=1}^{Q} t_q (S^{1/2}\Sigma_q S^{1/2})^{1/2}, \quad \text{and} \quad m = \sum_{q=1}^{Q} t_q m_q. \tag{4.5}$$

Proof. This theorem is a reformulation of [21], Corollary 4.5. □

With this result, we immediately obtain that the set of location-scatter atoms is a geodesic space, using the geodesics defined by the barycenters, which guarantees the applicability of the results of Section 2 for the location-scatter atoms.



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

FIGURE 1. Wasserstein barycenters between two mixtures of Slater-type elliptical distributions for the W_2 metric (blue) and the $W_{2,\mathcal{M}}$ metric (red).

4.2. Numerical tests

In this section, we provide a few practical examples for three types of mixtures. Two are based on elliptical distributions, namely Slater functions and the Wigner semicircle distribution. For elliptical distributions we need explicit expressions of the means and covariance matrices of the distributions in order to use formulas (4.2) and (4.5) to compute the Wasserstein distance and barycenters. We provide them in Appendix C, Lemma C.1. The third one is based on the gamma distribution, and illustrates that the framework developed in this article is not limited to elliptical distributions.

4.2.1. Slater-type elliptical distributions

We consider here Slater-type elliptical distributions, that is we consider $h(x) = \exp(-\alpha_d |x|^{1/2})$ where α_d is adapted with respect to d in order to satisfy equation (C.2), *i.e.*

$$\alpha_d = \sqrt{d+1}, \quad \text{and} \quad Z_\Sigma = \frac{2\pi^{d/2}}{(d+1)^{d/2}} \frac{\Gamma(d)}{\Gamma(d/2)} |\det \Sigma|^{1/2}.$$

The identifiability of the mixtures of such atoms can be proved with similar arguments as in [7], Proposition 2. Below we provide one-dimensional and two-dimensional examples for Slater-type elliptical distributions.

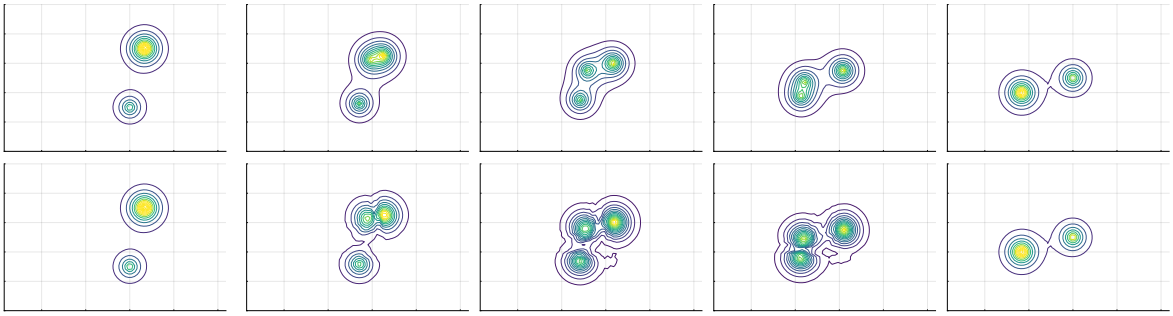
One-dimensional case. Using the previous argument, we consider the function $h(x) = \exp(-\sqrt{2}|x|^{1/2})$. In Figure 1, we present the Wasserstein barycenters with respect to the W_2 and $W_{2,\mathcal{M}}$ metrics between two mixtures of two atoms each. We observe that the two barycenters look quite different. In particular the W_2 barycenter is smoother than the $W_{2,\mathcal{M}}$ barycenter, in the sense that the $W_{2,\mathcal{M}}$ barycenter has sharper cusps W_2 barycenter.

Two-dimensional case. In two dimensions, we consider $h(x) = \exp(-\sqrt{3}|x|^{1/2})$, for which we check that equation (C.2) is satisfied. In Figures 2 and 3, we present the $W_{2,\mathcal{M}}$ and W_2 barycenters for mixtures of two-dimensional Slater-type elliptical distributions. As for the one-dimensional case, the W_2 barycenter is smoother than the $W_{2,\mathcal{M}}$ barycenter.

4.2.2. Wigner-semicircle-type elliptical distributions

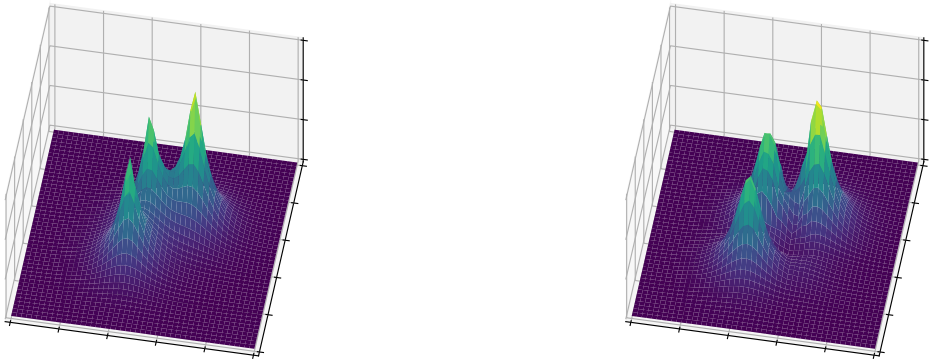
We now consider elliptical distributions based on the Wigner-semicircle distribution, *i.e.* we consider

$$h(x) = \begin{cases} \sqrt{1-\alpha_d x}, & x \leq \frac{1}{\alpha_d} \\ 0, & x > \frac{1}{\alpha_d}, \end{cases}$$



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

FIGURE 2. Contour plots of $W_{2,\mathcal{M}}$ (top) and W_2 (bottom) barycenters between two mixtures of Slater-type elliptic distributions.



(a) $W_{2,\mathcal{M}}$-barycenter (b) W_2-barycenter

FIGURE 3. Comparison between W_2 and $W_{2,\mathcal{M}}$ barycenters for $t = 0.5$ and the mixtures presented on Figure 2.

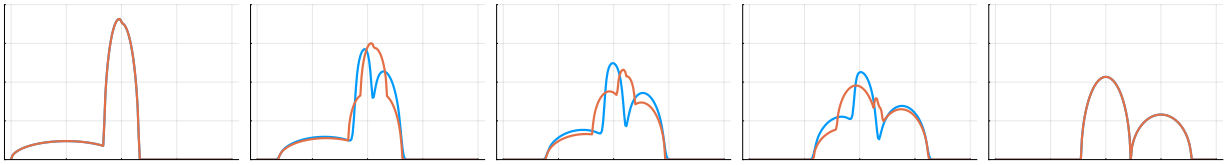
For $\alpha_d = \frac{1}{d+3}$, one easily checks that equation (C.2) is satisfied. Therefore, the densities of the probability distributions take the form

$$\forall x \in \mathbb{R}^d, \quad g_{m,\Sigma}(x) = \frac{1}{Z_\Sigma}\sqrt{1 - \frac{1}{d+3}(x-m)^T\Sigma^{-1}(x-m)},$$

with

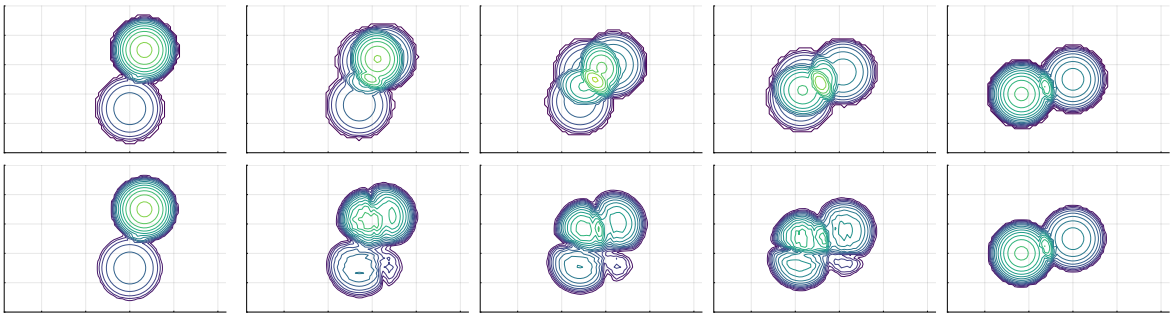
$$Z_\Sigma = \frac{\pi^{\frac{d+1}{2}}(d+3)^{d/2}}{2\Gamma\left(\frac{d+3}{2}\right)}|\det\Sigma|^{1/2}.$$

The identifiability of the mixtures based on these Wigner-semicircle type elliptical distribution distributions can be easily proved noting that the atoms can be iteratively uniquely characterized by their support.



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

FIGURE 4. Wasserstein barycenters between two mixtures of Wigner-semicircle elliptical distributions for the W_2 metric (blue) and the $W_{2,\mathcal{M}}$ metric (red).



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

FIGURE 5. Contour plots of $W_{2,\mathcal{M}}$ (top) and W_2 (bottom) barycenters between two mixtures of Wigner semicircle elliptic distributions.

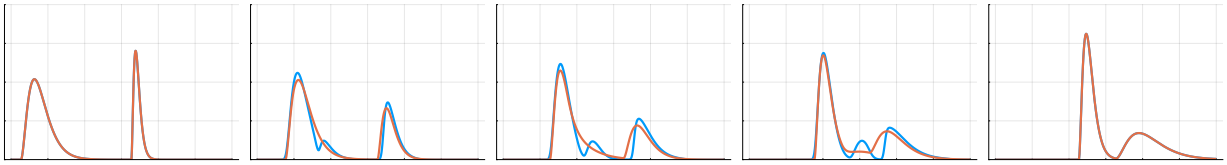
One-dimensional case. We take $h(x) = \begin{cases} \sqrt{1 - x/4}, & x \leq 4 \\ 0, & x > 4 \end{cases}$. In Figure 4, we plot the W_2 and $W_{2,\mathcal{M}}$ barycenters between two mixtures of two atoms each. We observe that the two types of barycenters behave quite differently, as the zones with largest densities as pretty different.

Two-dimensional case. We take $h(x) = \begin{cases} \sqrt{1 - x/5}, & x \leq 5 \\ 0, & x > 5 \end{cases}$. In Figure 5 we plot the W_2 and $W_{2,\mathcal{M}}$ barycenters between two mixtures of two atoms each for the two-dimensional case. The observations are similar to the one-dimensional case.

4.2.3. Gamma-distribution-based atoms

Finally, we consider atoms based on a Gamma distribution in $\mathbb{R}$. Since Gamma distributions are not elliptical distributions, we choose one particular gamma distribution and generate atoms from location-scatters of this particular distribution. For $\alpha, \beta > 0$, the probability density function of the gamma distribution with these parameters is

$$\forall x \geq 0, \quad f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x},$$



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

FIGURE 6. Wasserstein barycenters between two mixtures of Gamma distributions for the W_2 metric (blue) and the $W_{2,\mathcal{M}}$ metric (red).

and has mean α/β and variance α/β^2, so that we choose α, β such that $\alpha/\beta^2 = 1$, and we then define the atoms characterized by their mean m and covariance $\Sigma > 0$ by

$$\forall x \in \mathbb{R}, \quad g_{m,\Sigma}(x) = \begin{cases} \frac{1}{\sqrt{\Sigma}} f_{\alpha,\beta}(\Sigma^{-1/2}(x - m + \frac{\alpha}{\beta})), & x \geq m - \frac{\alpha}{\beta} \\ 0, & x < m - \frac{\alpha}{\beta}. \end{cases}$$

As for the Slater-type case, it is easy to prove identifiability of the mixtures by following the arguments presented in [7], Proposition 2.

In Figure 6, we present the W_2 and $W_{2,\mathcal{M}}$ barycenters between two mixtures of two atoms each, where we have taken $\alpha = 3$, $\beta = 9$ to define the atom having a variance of one. We observe that the $W_{2,\mathcal{M}}$ barycenter seems smoother and to have a lower middle mass movement than the W_2 barycenter.

5. Mixture distance for group invariant measures

In this section, we consider dictionaries of atoms that are defined as sets of symmetric probability measures, *i.e.* invariant with respect to some transformation, as stated in the following definition. Here, Ω is a convex open subset (or the closure of a convex open subset) of $\mathbb{R}^d$, $p > 1$ and $c : \Omega \times \Omega \to \mathbb{R}_+$ a given metric on Ω. We first define a distance well-suited for these symmetric probability measure, and we then define a mixture distance that is similar to the distance defined in (3.1) based on these symmetric atoms.

Definition 5.1 (Invariant measure)**.** Let $S : \Omega \to \Omega$ be a measurable function from Ω to itself. A measure $\mu \in \mathcal{P}(\Omega)$ is said to be invariant with respect to S if $S\#\mu = \mu$, that is if for every measurable subset $A \subset \Omega$, $\mu\left(S^{-1}(A)\right) = \mu(A)$.

Typical examples we have in mind are measures that are invariant with respect to permutations of the variable ordering, as in the case of marginals of the modulus square of the many-body wavefunction in quantum chemistry. Another example is given by the set of measures that are invariant with more generic isometries, such as rotations, rigid-body motions, or combination thereof.

5.1. General case

We now introduce dictionaries of symmetric atoms in order to define a mixture distance on symmetric measures. Let G be a finite or compact group acting on $\mathcal{P}_2(\Omega)$ through a group action denoted by $\cdot$. We denote by H the normalised Haar measure on G. Note that for finite groups, this measure corresponds to Dirac masses on the elements of the group with equal weight that is the inverse of the cardinal of the group.

Let $\mathcal{A}$ be a set of atoms such that for all $a \in \mathcal{A}$ and all $g \in G$, $g \cdot a \in \mathcal{A}$. We define $\mathcal{A}_{\text{sym}}$ as the set of symmetric measures defined from $\mathcal{A}$ as follows:

$$\mathcal{A}_{\text{sym}} = \left\{ \int_G g \cdot a \, H(dg), \quad a \in \mathcal{A} \right\}. \tag{5.1}$$

For any $a \in \mathcal{A}$, we denote by

$$S(a) := \int_G g \cdot a \, H(dg).$$

Note that, for any $g \in G$ and any $a \in \mathcal{A}$, $S(g \cdot a) = S(a)$. This leads us to introduce a metric on $\mathcal{A}_{\text{sym}}$. As we will see in Section 6, $S(a)$ could in principle be defined differently, provided that it satisfies the required symmetry. In order to impose uniqueness up to the group action, we make the following assumption.

Assumption 5.2. For any $a_1, a_2 \in \mathcal{A}$, there holds $S(a_1) = S(a_2)$ if and only if there exists $g \in G$ such that $a_1 = g \cdot a_2$.

This assumption is well-adapted to atoms generating identifiable mixtures, but not satisfied in general. Let us give a toy example, taking $\Omega = \left(-\frac{1}{2}, \frac{1}{2}\right)$ and the two-element group $(e, \tilde{e})$. Let us consider the following group action: for all $a \in \mathcal{P}_2(\Omega)$,

$$e \cdot a = a, \quad \tilde{e} \cdot a = (-\text{Id})\#a. \tag{5.2}$$

If the dictionary of atoms $\mathcal{A}$ is chosen so that (i) for all $a \in \mathcal{A}$, $(-\text{Id})\#a \in \mathcal{A}$ and (ii) the set mixtures $\mathcal{M}(\mathcal{A})$ is identifiable, it is easy to show that Assumption 5.2 is satisfied. However, taking $a_0(\,\mathrm{d}x) = 2\mathbb{1}_{[-1/2,0]}(x)\,\mathrm{d}x$ and $a_1(\,\mathrm{d}x) = \frac{2}{1+\mathrm{e}^{-x}}\,\mathrm{d}x$, there holds $\frac{1}{2}a_0 + \frac{1}{2}(-\text{Id})\#a_0 = \frac{1}{2}a_1 + \frac{1}{2}(-\text{Id})\#a_1$, while it does not hold that $a_1 = a_0$ or $a_1 = (-\text{Id})\#a_0$.

A metric on the set of symmetric measures $\mathcal{A}_{\text{sym}}$ is defined as follows.

Proposition 5.3. *Let $d : \mathcal{A} \times \mathcal{A} \to \mathbb{R}^+$ be a metric such that for all $a_0, a_1 \in \mathcal{A}$ and all $g \in G$,*

$$d(g \cdot a_0, g \cdot a_1) = d(a_0, a_1). \tag{5.3}$$

Then, the map $\bar{d} : \mathcal{A}_{\text{sym}} \times \mathcal{A}_{\text{sym}} \to \mathbb{R}^+$ defined by

$$\forall a_0, a_1 \in \mathcal{A}, \quad \bar{d}(\bar{a}_0, \bar{a}_1) = \inf_{g \in G} d(a_0, g \cdot a_1), \qquad \textit{where } \bar{a}_0 = S(a_0),\ \bar{a}_1 = S(a_1), \tag{5.4}$$

is a metric on $\mathcal{A}_{\text{sym}}$.

Remark 5.4. The definition given in Proposition 5.3 is equivalent to

$$\forall \bar{a}_0, \bar{a}_1 \in \mathcal{A}_{\text{sym}}, \quad \bar{d}(\bar{a}_0, \bar{a}_1) = \inf_{g_0, g_1 \in G} d(g_0 \cdot a_0, g_1 \cdot a_1), \quad \text{where } \bar{a}_0 = S(a_0),\ \bar{a}_1 = S(a_1),$$

due to (5.3).

Remark 5.5. In some cases, it is clear which g realizes the minimum in (5.4) which means that the distance $\bar{d}(a_0, a_1)$ can be easily computed by first computing $g \cdot a_1$ and then computing the distance $d(a_0, g \cdot a_1)$. For example for the parity group acting on probability measures, one can restrict the problem to probability measures with positive mean. However nothing guarantees for a generic group that a unique transformation can be applied to all elements such that the minimization problem (5.4) can be avoided.

Proof. First, $\bar{d}$ is clearly symmetric. Second, if $\bar{d}(\bar{a}_0, \bar{a}_1) = 0$, there exists $g \in G$ such that $d(a_0, g \cdot a_1) = 0$ and so $a_0 = g \cdot a_1$. Therefore, $S(a_0) = S(g \cdot a_1) = S(a_1)$, *i.e.* $\bar{a}_0 = \bar{a}_1$. Third, we prove the triangle inequality. Let $a_0, a_1, a_2 \in \mathcal{A}$ so that $\bar{a}_0, \bar{a}_1, \bar{a}_2 \in \mathcal{A}_{\text{sym}}$. It then holds that

$$\begin{aligned}\bar{d}(\bar{a}_0, \bar{a}_2) &= \inf_{g_2 \in G} d(a_0, g_2 \cdot a_2) \\ &\leq \inf_{g_1, g_2 \in G} [d(a_0, g_1 \cdot a_1) + d(g_1 \cdot a_1, g_2 \cdot a_2)] \\ &\leq \bar{d}(\bar{a}_0, \bar{a}_1) + \bar{d}(\bar{a}_1, \bar{a}_2).\end{aligned}$$

Hence $\bar{d}$ is a metric. □

Proposition 5.6. *$\mathcal{A}_{\text{sym}}$ equipped with the metric $\bar{d}$ is a geodesic space.*

Proof. The proof is similar to that of Proposition 3.7. We consider paths $(\rho_t)_{t\in[0,1]}$ with $\rho_t \in \mathcal{A}_{\text{sym}}$ for all $t \in [0,1]$, we define the length of the path relative to the $\bar{d}$ metric as

$$\text{Len}_{\bar{d}}(\rho) = \sup_{\substack{N\in\mathbb{N}^*,\ (t_i)_{0\leq i\leq N}\in[0,1]^{N+1} \\ 0=t_0<t_1<\ldots<t_N=1}} \sum_{i=1}^{N} \bar{d}(\rho_{t_{i-1}}, \rho_{t_i}) \tag{5.5}$$

and we show that given any two points $\bar{a}_0, \bar{a}_1 \in \mathcal{A}_{\text{sym}}$, there exists a path between them the length of which equals the distance $\bar{d}(\bar{a}_0, \bar{a}_1)$. We first show using the triangle inequality that $\text{Len}_{\bar{d}}(\rho) \geq \bar{d}(\bar{a}_0, \bar{a}_1)$. The equality is shown by defining $\bar{g} = \text{argmin}_{g\in G} d(a_0, g \cdot a_1)$, where $\bar{a}_0 = S(a_0), a_0 \in \mathcal{A}$ and $\bar{a}_1 = S(a_1), a_1 \in \mathcal{A}$, and taking $(\mu_t)_{t\in[0,1]}$ be a constant speed geodesic between a_0 and $\bar{g} \cdot a_1$. Noting that for all $0 \leq s, t \leq 1$,

$$d(\mu_t, \mu_s) = |t - s| d(a_0, \bar{g} \cdot a_1).$$

we define $\bar{\mu}_t = S(\mu_t)$ to obtain that for any $0 \leq s, t \leq 1$,

$$\bar{d}(\bar{\mu}_t, \bar{\mu}_s) = \inf_{\widetilde{g}\in G} d(\mu_t, \widetilde{g} \cdot \mu_s) \leq d(\mu_t, \mu_s) \leq |t - s| d(a_0, \bar{g} \cdot a_1) = |t - s| \bar{d}(\bar{a}_0, \bar{a}_1),$$

from which we deduce that $\text{Len}_{\bar{d}}((\bar{\mu}_t)_{t\in[0,1]}) = \bar{d}(\bar{a}_0, \bar{a}_1)$. We conclude that $\mathcal{A}_{\text{sym}}$ equipped with $\bar{d}$ is a geodesic space. □

Having now proved that $(\mathcal{A}_{\text{sym}}, \bar{d})$ is a geodesic space, we can now express barycenters in terms of geodesics of atoms.

Corollary 5.7. *Let $\bar{a}_0 = S(a_0)$ and $\bar{a}_1 = S(a_1)$ be two elements of $\mathcal{A}_{\text{sym}}$. Let $\bar{g} = \text{argmin}_{g\in G} d(a_0, g \cdot a_1)$. The barycenters between $\bar{a}_0$ and $\bar{a}_1$ belong to $\mathcal{A}_{\text{sym}}$ and can be written as*

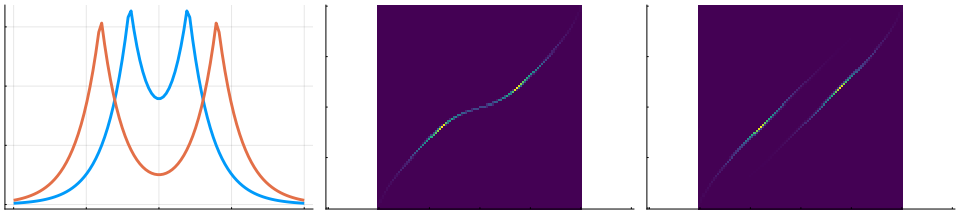
$$\forall t \in [0,1], \quad \bar{a}_t = S(a_t),$$

where $(a_t)_{t\in[0,1]}$ is a constant speed geodesic between a_0 and $\bar{g} \cdot a_1$.

Note that if Assumption 5.2 is satisfied, it is easy to show that the barycenter between $\bar{a}_0$ and $\bar{a}_1$ does not depend on the choice of a_0 and a_1.

5.2. Wasserstein case

In this section, we focus on the particular case where d is the Wasserstein distance defined in (2.1), and where the group action of G on $\mathcal{P}_p^c(\Omega)$ is inherited from a group action of G onto Ω as follows: for all $g \in G$, $g \cdot a = a \# T_g$ where $T_g : \Omega \to \Omega$ is defined by $T_g(x) = g \cdot x$ for all $x \in \Omega$. For all $k \in \mathbb{N}^*$, we define $T_g^k : \Omega^k \to \Omega^k$ as the map such that $T_g^k(x_1, \cdots, x_k) = (g \cdot x_1, \cdots, g \cdot x_k)$ for all $x_1, \cdots, x_k \in \Omega$. Using a slight abuse of notation, for all



(a) two even densities (b) W_2 optimal transport plan (c) $W_{2,\mathcal{M}}$ optimal transport plan

FIGURE 7. Comparison between W_2 and $W_{2,\mathcal{M}}$ transport plans between two even densities.

$k \in \mathbb{N}^*$, and for all $\gamma \in \mathcal{P}_p(\Omega^k)$, we denote by $g \cdot \gamma = \gamma \# T_g^k$. Then upon the assumption that $c(g \cdot x, g \cdot y) = c(x, y)$, for all $x, y \in \Omega$ and $g \in G$ so that Proposition 5.3 holds, there holds

$$\bar{d}(\bar{a}_0, \bar{a}_1) = \left[\inf_{g_0, g_1 \in G} \inf_{\gamma \in \Pi(g_0 \cdot a_0, g_1 \cdot a_1)} \int c(x, y)^p \mathrm{d}\gamma(x, y) \right]^{1/p}. \tag{5.6}$$

By analogy, one can define the following symmetrized optimal transport plan as

$$\bar{\gamma} = \int_{g \in G} g \cdot \gamma \, H(dg).$$

Note that this is not equivalent to the true Wasserstein transport plan. Indeed, the transport plans are in general different, as shown in Figure 7 for two densities being even hence symmetric with respect to the group action defined in (5.2). This can be easily understood. For the true Wasserstein distance, the optimal transport plan satisfies the monotone rearrangment property, while in the case of the symmetrized distance, the atom is first transported to the closest one, and then the transport plan is symmetrized. However, this allows for a simple generalization to the multi-marginal problem.

Definition 5.8 (Symmetric multi-marginal problem)**.** Let $Q \in \mathbb{N}^*$. Let $\boldsymbol{t} := (t_q)_{1 \leq q \leq Q} \in \mathcal{L}_Q$. Let $a_1, \ldots, a_Q \in \mathcal{A}$ and $\bar{a}_1 = S(a_1), \ldots, \bar{a}_Q = S(a_Q)$ for $\mathcal{A}$ a set of atoms. We define the symmetric multi-marginal transport problem by

$$\bar{d}_Q^{\boldsymbol{t}}(\bar{a}_1, \ldots, \bar{a}_Q) = \left(\inf_{g_1, \ldots, g_Q \in G} \inf_{\gamma \in \Pi(g_1 \cdot a_1, \ldots, g_Q \cdot a_Q)} s(x_1, \ldots, x_Q) \mathrm{d}\gamma(x_1, \ldots, x_Q) \right)^{1/p}, \tag{5.7}$$

with

$$s(x_1, \ldots, x_Q) = \frac{1}{2} \sum_{q=1}^{Q} \sum_{q'=1}^{Q} t_q t_{q'} c(x_q, x_{q'})^p.$$

Denoting by $\Gamma^{\boldsymbol{t}}_{\bar{a}_1, \ldots, \bar{a}_Q}$ the set of minimizers of (5.7), for all $\gamma \in \Gamma^{\boldsymbol{t}}_{\bar{a}_1, \ldots, \bar{a}_Q}$, we define the associated symmetrized optimal transport plan as

$$\bar{\gamma} = \int_{g \in G} g \cdot \gamma \, H(dg).$$

Finally, assuming that Wasserstein barycenters between atoms in $\mathcal{A}$ also belong to $\mathcal{A}$, the barycenters of $(\bar{a}_1, \cdots, \bar{a}_Q)$ with barycentric weights $\boldsymbol{t}$ can be written as

$$\overline{\mathrm{bar}}_{\boldsymbol{t}}(\bar{a}_1, \ldots, \bar{a}_Q) = S(\mathrm{bar}_{\boldsymbol{t}}(\bar{g}_1 \cdot a_1, \ldots, \bar{g}_Q \cdot a_Q)) \tag{5.8}$$

where $(\bar{g}_1, \ldots, \bar{g}_Q) \in G^Q$ is such that

$$(\bar{g}_1, \ldots \bar{g}_Q) \in \underset{g_1,\ldots,g_Q \in G}{\mathrm{arginf}} \inf_{\gamma \in \Pi(g_1 \cdot a_1, \ldots, g_Q \cdot a_Q)} s(x_1, \ldots, x_Q) \mathrm{d}\gamma(x_1, \ldots, x_Q).$$

With this definition, we show that the multi-marginal problem is equivalent to the symmetric barycentric problem.

Proposition 5.9. *The infimum in* (5.7) *matches the following barycentric minimization problem*

$$\bar{d}^{\boldsymbol{t}}_Q(\bar{a}_1, \ldots, \bar{a}_Q)^p = \inf_{\bar{a} \in \mathcal{A}_{\mathrm{sym}}} \sum_{q=1}^{Q} t_q \bar{d}(\bar{a}, \bar{a}_q)^p. \tag{5.9}$$

Proof. Starting from the right handside of (5.9), and using the distance definition (5.4), we obtain

$$\inf_{\bar{a} \in \mathcal{A}_{\mathrm{sym}}} \sum_{q=1}^{Q} t_q \bar{d}(\bar{a}, \bar{a}_q)^p = \inf_{a \in \mathcal{A}} \sum_{q=1}^{Q} t_q [\inf_{g_q \in G} d(a, g_q \cdot a_q)]^p = \inf_{g_1,\ldots,g_Q \in G} \inf_{a \in \mathcal{A}} \sum_{q=1}^{Q} t_q d(a, g_q \cdot a_q)^p.$$

Then using the equivalence between the multi-marginal problem and the barycentric problem on the set of atoms $\mathcal{A}$, which is well-posed since the atoms are stable under Wasserstein barycenter proves (5.9). □

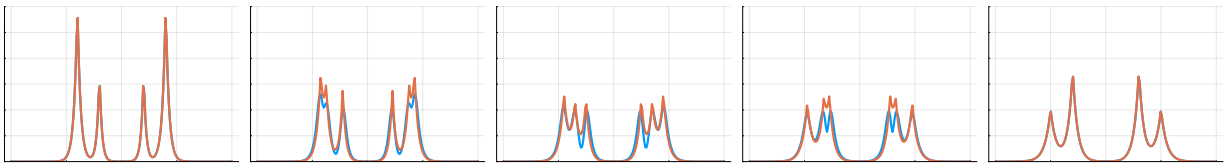
Using the defined distance $\bar{d}$ on the atoms and corresponding atomic barycenters, we define the corresponding mixture distance following Definition 3.4, and mixture barycenters following Proposition 3.18.

5.3. Examples

We now provide a few examples in dimension $d = 1, 2$ for group actions defined over permutation groups and the rotation group $SO(2)$. For simplicity, we consider the case $\Omega = \mathbb{R}^d$, $p = 2$ and $c(x, y) = \|x - y\|$. All the numerical results presented below are performed with the same setting as in Section 4.2 for the W_2 barycenters. The $W_{2,\mathcal{M}}$ barycenters are computed following Definition (5.8). In terms of computational cost, let us remark that the computational cost of the mixture barycenter highly depends on the cost of computing the barycenter between two atoms. This cost indeed depends on the cardinal of the group if finite, or the complexity of solving (5.4) for infinite groups. However, in the tests presented below, the cardinal of the considered finite groups, which are the permutation groups, is only two, therefore the computation of the mixture barycenters stays orders of magnitude cheaper than the computation of the W_2 barycenters.

5.3.1. Parity group

We first consider the case $d = 1$, $\Omega = \mathbb{R}^d$ and where G is defined as the two-element group $(e, \tilde{e})$ with the group action defined in (5.2). We compute the W_2 and $W_{2,\mathcal{M}}$ barycenters for the symmetric measure of a mixture of two Slater functions with respect to this group action and present them in Figure 8. We observe that the barycenters correspond to the symmetrized version of the barycenters presented on Figure 1 which is naturally expected.



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

FIGURE 8. Wasserstein barycenters between two mixtures of symmetric Slater distributions for the W_2 metric (blue) and the $W_{2,\mathcal{M}}$ metric (red).

5.3.2. Permutation group

We then consider the permutation group $\mathfrak{S}_n$ on the set $\{1, \dots, n\}$. We define $\Omega = (\mathbb{R}^d)^n$ and a group action $\cdot$ on elements of $\mathcal{P}_2(\Omega)$ by

$$\forall a \in \mathcal{P}_2(\Omega), \quad \forall \sigma \in \mathfrak{S}_n, \quad \sigma \cdot a = a \# T_\sigma,$$

where

$$T_\sigma : \left\{ \begin{array}{ccc} \Omega & \to & \Omega \\ (x_1, \dots, x_n) & \mapsto & (x_{\sigma(1)}, \dots, x_{\sigma(n)}), \end{array} \right. \tag{5.10}$$

which corresponds to a permutation of the variables. It is easy to check that this indeed defines a group action. For a given set of atomic measures $\mathcal{A}$, the corresponding set of symmetric atomic measures $\mathcal{A}_{\text{sym}}$ is then given as the set of elements $\bar{a}$ defined for all $a \in \mathcal{A}$ by

$$\bar{a} = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sigma \cdot a.$$

The computation of the symmetric barycenters can be then easily done, provided that the underlying set of atoms allows for an efficient - or even explicit - computation of distances d. Choosing for the atoms $\mathcal{A}$ any set of location-scatter atoms, the computation of the symmetric Wasserstein distance $\bar{d}$ only requires the computation of $n!$ explicit Wasserstein distances, which stays cheap for moderate values of n. In Figure 9 we plot the W_2 barycenters between symmetric mixtures of Gaussian measures for the group action defined above and compare it to the symmetric barycenter based on Gaussian mixtures. We observe that the mixture barycenter is smoother than the W_2 barycenter, and obtained at a fraction of the cost.

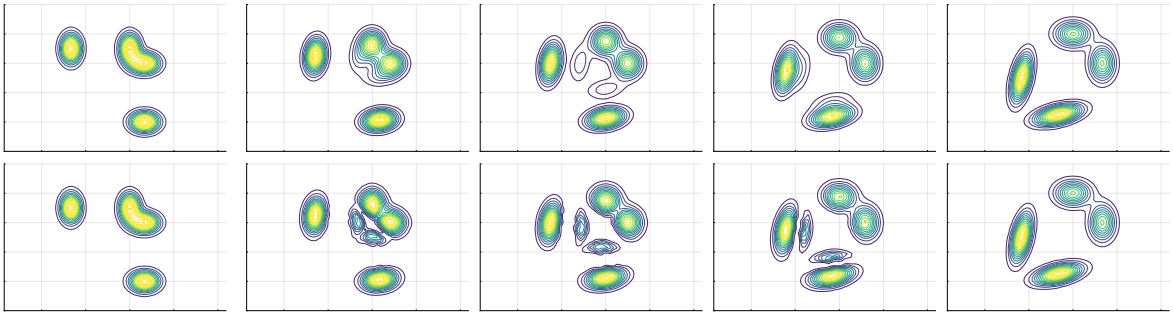
5.3.3. Rotation group $SO(2)$

Finally, we consider the case where $\Omega = \mathbb{R}^2$ and the two-dimensional rotation group $SO(2)$, for which we can, as for the permutation group, define a natural group action $\cdot$ *via*

$$\forall a \in \mathcal{A}, \quad \forall Q \in SO(2), \quad (Q \cdot a) = a \# T_Q,$$

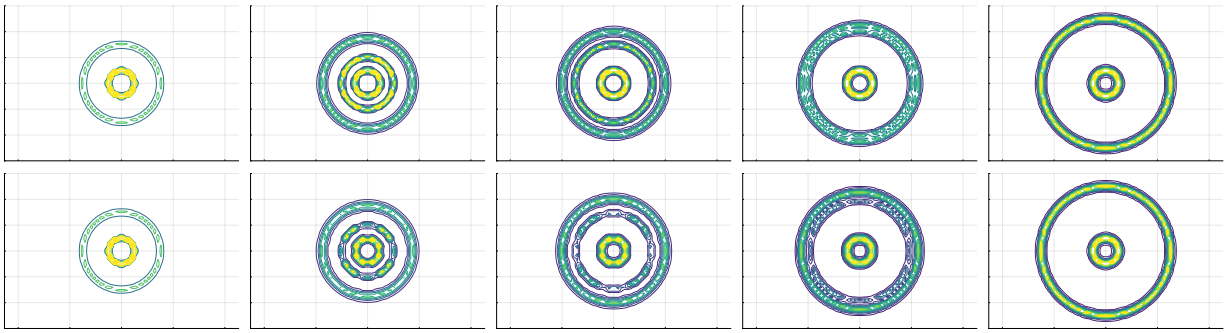
with $T_Q : \mathbb{R}^2 \ni \mathbf{x} \mapsto Q\mathbf{x}$. Note that such a group action can similarly be defined for $SO(n)$. From definition (5.1) the symmetric atoms $\bar{a}$ of $\mathcal{A}_{\text{sym}}$ write

$$\bar{a} = \int Q \cdot a \, H(dQ).$$



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

FIGURE 9. Contour plots of $W_{2,\mathcal{M}}$ (top) and W_2 (bottom) barycenters between two mixtures of symmetric Gaussian distributions.



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

FIGURE 10. Contour plots of $W_{2,\mathcal{M}}$ (top) and W_2 (bottom) barycenters between two mixtures of $SO(2)$-symmetric measures based on Slater-type distributions.

We now provide a numerical example, taking $\mathcal{A}$ as the set of two-dimensional Slater-type distributions. The main difficulty and difference with the previous examples is that the group is not finite. Therefore, problem (5.4) which has to be solved to compute the distance between elements in $\mathcal{A}_{\text{sym}}$ as well as barycenters is now a continuous optimization problem. In this two-dimensional case, there is in fact only one parameter to optimize which is the angle of the rotation, and we chose to perform this optimization numerically. This increases the computational cost of the distance and barycenters, however the computational cost stays way below the cost of computing the W_2 Wasserstein barycenters. In Figure 10 we provide an example of barycenters between two mixtures of symmetric distributions, themselves based on one Slater function each. We observe that the W_2 Wasserstein barycenter computed with the Sinkhorn algorithm is way less regular than the modified Wasserstein barycenter.

6. Application in quantum chemistry

The aim of this section is to present some applications of generic mixture Wasserstein distances and barycenters to quantum chemistry applications, which require the use of atoms different from classical Gaussian distributions. More precisely, we consider typical solutions to quantum chemistry problems which exhibits additional symmetries arising from the antisymmetry of the wavefunction with respect to exchange of particles. Therefore we present barycenters between symmetry group invariant mixtures based on Gaussian distributions. Other examples beyond Gaussian mixtures using Slater-type distributions, which naturally arise in the study of a Schrödinger-type eigenvalue problem are presented in [24].

6.1. Preliminaries on electronic structure theory

In quantum chemistry, the state of a system of N electrons is fully characterized by its electronic wavefunction, which is a function $\psi \in L^2(\mathbb{R}^{dN};\mathbb{C})$ with d the dimension of the underlying physical space (typically $d = 1, 2, 3$). This wavefunction is normalized in the sense that

$$\int_{\mathbb{R}^{dN}} |\psi|^2 = 1.$$

Note that we omit here the spin variables for simplicity of the presentation. Since electrons are fermions, the function ψ is antisymmetric with respect to permutations of the ordering of the variables: for all $\sigma \in \mathfrak{S}_N$ that is a permutation of the set $\{1, \cdots, N\}$ and all $(x_1, \ldots, x_N) \in (\mathbb{R}^d)^N$, there holds

$$\psi(x_{\sigma(1)}, \ldots, x_{\sigma(N)}) = \epsilon(\sigma)\psi(x_1, \ldots, x_N),$$

where $\epsilon(\sigma)$ denotes the parity of the permutation σ.

Since the wavefunction is defined on a high-dimensional space when the number of electrons in the system is large, it is more convenient to handle normalized n-body densities for some $1 \leq n \leq N$ defined as follows:

$$\forall \boldsymbol{x} = (x_1, x_2, \ldots x_n) \in (\mathbb{R}^d)^n, \quad \rho_n(\boldsymbol{x}) := \int_{(x_{n+1},\ldots,x_N)\in(\mathbb{R}^d)^{N-n}} |\psi(\boldsymbol{x})|^2 \, \mathrm{d}x_{n+1} \ldots \mathrm{d}x_N. \tag{6.1}$$

Thus for all $1 \leq n \leq N$, ρ_n can be seen as the density associated to a probability measure on $\mathbb{R}^{dn}$.

For simplicity we focus now on the 2-body density $\rho_2 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ which belongs to $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. From its definition, it is clear that (i) ρ_2 is symmetric (*i.e.* $\rho_2(x_1, x_2) = \rho_2(x_2, x_1)$; (ii), it vanishes on the set $\{(x, x) : x \in \mathbb{R}^d\}$.

Mixtures of well-chosen could be adapted to approximate 2-body (or more generally n-body) densities if they were satisfying these two properties. Moreover, mixture Wasserstein-like barycenters between such approximations would provide meaningful interpolations between these objects, in particular preserving the symmetry properties of 2-body densities highlighted above.

6.2. Description of the method relying on mixtures of group-invariant measures

We consider the permutation group $G = \mathfrak{S}_2$ acting on $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ defined in Section 5.3.2. The group G only contains two elements, *i.e.* $G = \{\mathrm{Id}, \sigma\}$ where Id is the identity map of the set $\{1, 2\}$ and σ is the permutation of the elements 1 and 2. In addition, we introduce a subset $\mathcal{A} \subset \mathcal{P}(\mathbb{R}^{d\times d})$ of atomic probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ absolutely continuous with respect to the Lebesgue measure (*e.g.* Gaussian or Slater distributions) satisfying Assumption 3.2, Assumption 3.3 with $\delta = W_2$ and Assumption 3.9. We also denote by $\mathcal{A}_{\mathrm{sym}}$ the set of symmetric atomic measures under the action of G defined in Section 5.

We define $\mathcal{A}_0 = \{a \in \mathcal{A} : a \neq \sigma \cdot a\}$ which is the set of atomic measures which are not symmetric with respect to the map T_σ defined in (5.10). For any $a \in \mathcal{A}_0$, we denote by S_a the probability measure on $\mathbb{R}^{d\times d}$ defined by

$$S_a := \frac{1}{Z}\left(a - \sigma \cdot a\right)^2 \quad \text{with } Z = \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(a - \sigma \cdot a\right)^2,$$

that is

$$S_a(x_1, x_2) = \frac{1}{Z}\left(a(x_1, x_2) - a(x_2, x_1)\right)^2,$$

and we consider the set of atomic 2-body probability densities as

$$\mathcal{A}_{2b} := \{S_a : \ a \in \mathcal{A}_0\}.$$

We then consider 2-body electronic densities that can be written under the form

$$\rho_2(x_1, x_2) = \sum_{k=1}^{K} \pi_k S_{a_k}(x_1, x_2),$$

where $K \in \mathbb{N}^*$, $(\pi_k)_{1\leq k\leq K} \in \mathcal{L}_K$ a collection of barycentric weights, $a_k \in \mathcal{A}_0$, and $Z_k := \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(a_k(x_1, x_2) - a_k(x_2, x_1)\right)^2 \, \mathrm{d}x_1 \, \mathrm{d}x_2$ for $k = 1, \ldots, K$. Note that for all $1 \leq k \leq K$, since $a_k \in \mathcal{A}_0$, $Z_k > 0$.

Remark 6.1. In this section, we present the method in the case of 2-body densities for the sake of simplicity of the presentation. However, note that it is straightforward to extend the proposed method to general n-body densities with arbitrary $n \geq 2$. To this aim, one may consider a set of atomic measures $\mathcal{A}$ on $(\mathbb{R}^d)^n$ and the corresponding set of atomic measures $\mathcal{A}_{nb}$ may be defined as the set of probability measures of the form

$$\frac{1}{Z}\left(\sum_{\sigma\in\mathfrak{S}_n} \epsilon(\sigma)\sigma\cdot a\right)^2,$$

where a belongs to a set of atomic probability measures on $(\mathbb{R}^d)^n$ such that $\sigma\cdot a \neq \widetilde{\sigma}\cdot a$ for all $\sigma, \widetilde{\sigma} \in \mathfrak{S}_n$ with $\sigma \neq \widetilde{\sigma}$, and

$$Z = \int_{(\mathbb{R}^d)^n} \left(\sum_{\sigma\in\mathfrak{S}_n} \epsilon(\sigma)\sigma\cdot a\right)^2.$$

We assume in the following that the set $\mathcal{A}_{2b}$ is identifiable. Note that this is indeed the case when the set $\mathcal{A}$ is chosen to be the set of Gaussian measures. Then, there is a unique injective linear map $\mathcal{I} : \mathcal{M}(\mathcal{A}_{2b}) \to \mathcal{M}(\mathcal{A}_{\mathrm{sym}})$ such that, for all $a \in \mathcal{A}_0$, $\mathcal{I}(S_a) = \overline{a}$ where $\overline{a} = \frac{1}{2}(a + \sigma\cdot a)$.

We then define a mixture distance δ on $\mathcal{M}(\mathcal{A}_{2b})$ which is defined as follows: for all $S, S' \in \mathcal{M}(\mathcal{A}_{2b})$,

$$\delta(S, S') = \overline{W_2}_{,\mathcal{M}}(\mathcal{I}(S), \mathcal{I}(S')),$$

where $\overline{W_2}$ is the symmetrized Wasserstein metric on $\mathcal{A}_{\mathrm{sym}}$ as defined in Section 5, and $\overline{W_2}_{,\mathcal{M}}$ is the mixture symmetrized Wasserstein metric defined on $\mathcal{M}(\mathcal{A}_{\mathrm{sym}})$.

Let us now explain how we compute Wasserstein-like barycenters between elements of $\mathcal{M}(\mathcal{A}_{2b})$. For any $Q \in \mathbb{N}^*$, and any collection $\boldsymbol{S} = (S_1, \ldots, S_Q)$ of elements of $\mathcal{M}(\mathcal{A}_{2b})$ and any collection $\boldsymbol{t} = (t_1, \ldots, t_Q) \in \mathcal{L}_Q$ of barycentric weights, using the results of Section 5, one can construct a mixture Wasserstein barycenter of the family $(\mathcal{I}(S_1), \ldots, \mathcal{I}(S_Q)) \in \mathcal{M}(\mathcal{A}_{\mathrm{sym}})$ with barycentric weights $\boldsymbol{t}$ for the mixture symmetrized Wasserstein metric $\overline{W_2}_{,\mathcal{M}}$. This barycenter, denoted by τ, is then an element of $\mathcal{M}(\mathcal{A}_{\mathrm{sym}})$ which reads as

$$\tau = \sum_{k=1}^{K} \pi_k \frac{1}{2}(a_k + \sigma\cdot a_k) \tag{6.2}$$

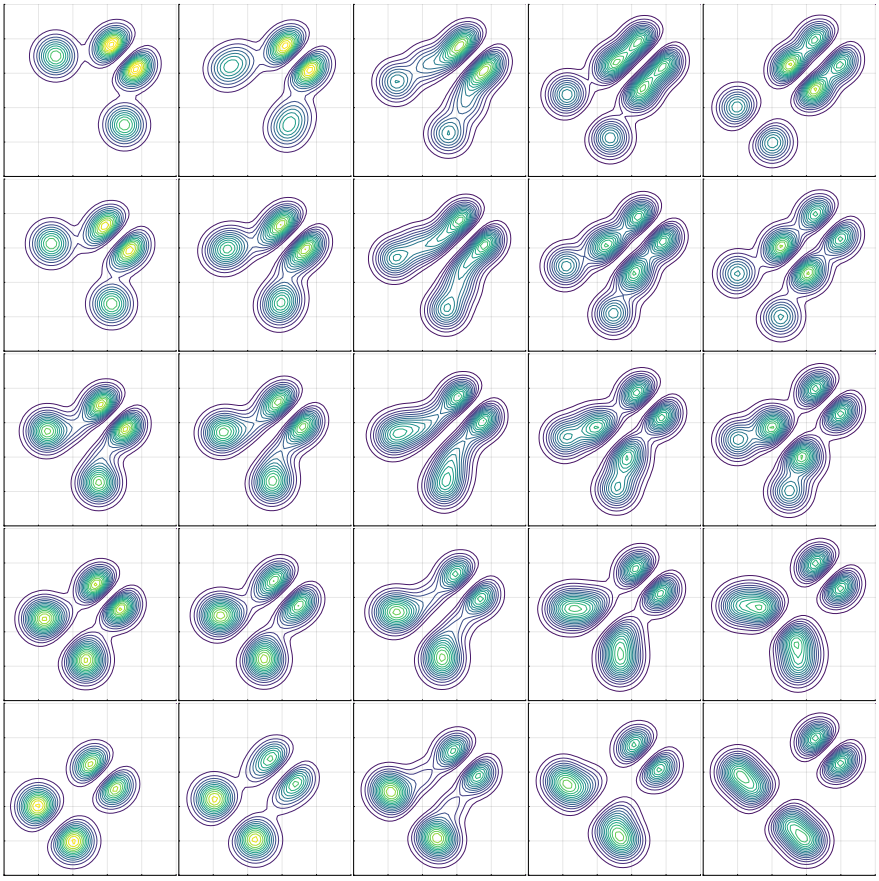


FIGURE 11. Contour plots of δ barycenters between four mixtures of squared Slater determinants based on Gaussian functions.

for some $K \in \mathbb{N}^*$, $(\pi_k)_{1\leq k\leq K} \in \mathcal{L}_K$ and $a_k \in \mathcal{A}$ for $k = 1, \dots, K$. In the case when $a_k \in \mathcal{A}_0$ for all $k = 1, \dots, K$, we can then define a barycenter for the family $\boldsymbol{S}$ with barycentric weights $\boldsymbol{t}$ as

$$\overline{\mathrm{bar}}_{\boldsymbol{t}}(\boldsymbol{S}) = \sum_{k=1}^{K} \pi_k S_{a_k}. \tag{6.3}$$

Note that the definition (6.3) is well-defined for any choice of $(a_k)_{1\leq k\leq K}$ such that (6.2) holds.

However, if there exists some $1 \leq k_0 \leq K$ such that $a_{k_0} \in \mathcal{A} \setminus \mathcal{A}_0$, *i.e.* a_{k_0} is permutation-invariant, we are not able to define a relevant barycenter since the probability measure $S_{a_{k_0}}$ is not defined in this case. This is also the reason why the metric space $(\mathcal{M}(\mathcal{A}_{2b}), \delta)$ is *a priori* not a geodesic space. We emphasize on the fact that we never encountered this situation in practice in the numerical test cases reported in the next section.

6.3. Numerical results

To illustrate the numerical behaviour of the proposed method, we plot on Figures 11 and 12 the barycenters between four mixtures of squares of Slater determinants respectively for the δ and W_2 distances. We use bilinear weights with respect to the four corners of the square to compute the barycenters. First we observe that both types of barycenters satisfy that they are zero on the diagonal, as is constrained by the form of the densities

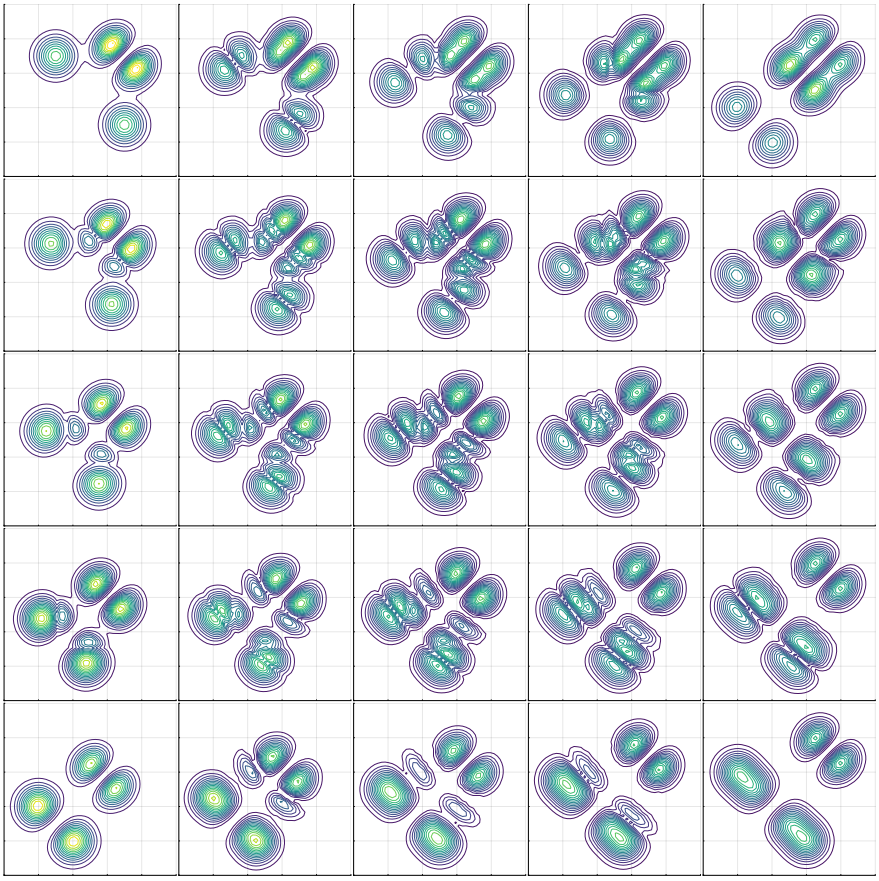


FIGURE 12. Contour plots of W_2 barycenters between four mixtures of squared Slater determinants based on Gaussian functions.

for the δ distance, and naturally expected from Proposition A.2 for the 2-Wasserstein barycenters. Second, we observe that the mixture Wasserstein barycenters are smoother than the W_2-barycenters and therefore seem to be better suited for interpolations of electronic densities arising from antisymetry. The application of the Wasserstein-type distance and Wasserstein-type barycenters introduced here to accelerate computations in quantum chemistry will be addressed in a future work.

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreements EMC2 No 810367 and HighLEAP No. 101077204). This work was supported by the French 'Investissements d'Avenir' program, project Agence Nationale de la Recherche (ISITE-BFC) (contract ANR-15-IDEX-0003), and the ANR JCJC NUMERIQ (contract ANR-24-CE46-2255). G.D. was also supported by the Ecole des Ponts-ParisTech. VE acknowledges support from the ANR project COMODO (ANR-19-CE46-0002).

DATA AVAILABILITY STATEMENT

The research code associated with this article are available at https://github.com/dussong/W2_mixtures.jl/.

References

[1] E. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris and Y. Maday, Computational quantum chemistry: a primer, in *Handbook of Numerical Analysis*, vol. 10. Elsevier (2003) 3–270.

[2] T. Helgaker, P. Jorgensen and J. Olsen, Molecular Electronic-Structure Theory. John Wiley & Sons (2014).

[3] M. Agueh and G. Carlier, Barycenters in the wasserstein space. *SIAM J. Math. Anal.* **43** (2011) 904–924.

[4] G. Dusson, G. Friesecke and C. Klüppelberg and G. Friesecke, Copula methods for modeling pair densities in density functional theory. *J. Chem. Phys.* **162** (2025) 144109.

[5] R. Sinkhorn, Diagonal equivalence to matrices with prescribed row and column sums. *Am. Math. Mon.* **74** (1967) 402–405.

[6] G. Peyré and M. Cuturi, Computational optimal transport: with applications to data science. *Found. Trends® Mach. Learn.* **11** (2019) 355–607.

[7] J. Delon and A. Desolneux, A Wasserstein-type distance in the space of gaussian mixture models. *SIAM J. Imaging Sci.* **13** (2020) 936–970.

[8] D.H. Pham, *Galerkin method using optimized wavelet-Gaussian mixed bases for electronic structure calculations in quantum chemistry.* PhD thesis, Université Grenoble Alpes (2017).

[9] A.K Gupta, T. Varga and T. Bodnar, Elliptically Contoured Models in Statistics and Portfolio Theory. Springer, New York (2013).

[10] E. Mehr, A. Lieutier, F.S. Bermudez, V. Guitteny, N. Thome and M. Cord, Manifold learning in quotient spaces, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* IEEE (2018) 9165–9174.

[11] H. Chen and G. Friesecke, Pair densities in density functional theory. *Multiscale Model. Simul.* **13** (2015) 1259–1289.

[12] C. Villani, Optimal Transport. Springer, Berlin, Heidelberg (2009).

[13] F. Santambrogio, Optimal Transport for Applied Mathematicians. Springer International Publishing (2015).

[14] R.J. McCann, A convexity principle for interacting gases. *Adv. Math.* **128** (1997) 153–179.

[15] W. Gangbo and A. Świech, Optimal maps for the multidimensional Monge–Kantorovich problem. *Commun. Pure Appl. Math.* **51** (1998) 23–45.

[16] Y. Chen, T.T. Georgiou and A. Tannenbaum, Optimal transport for gaussian mixture models. *IEEE Access* **7** (2018) 6269–6278.

[17] E. Anderes, S. Borgwardt and J. Miller, Discrete Wasserstein barycenters: optimal transport for discrete data. *Math. Methods Oper. Res.* **84** (2016) 389–409.

[18] G. Friesecke and M. Penka, The GenCol algorithm for high-dimensional optimal transport: general formulation and application to barycenters and Wasserstein splines. *SIAM J. Math. Data Sci.* **5** (2023) 899–919.

[19] J. Bezanson, A. Edelman, S. Karpinski and V.B. Shah, Julia: a fresh approach to numerical computing. *SIAM Rev.* **59** (2017) 65–98.

[20] R. Flamary, N. Courty, A. Gramfort, M.Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N.T.H. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D.J. Sutherland, R. Tavenard, A. Tong and T. Vayer. Pot: Python optimal transport. *J. Mach. Learn. Res.* **22** (2021) 1–8.

[21] P.C. Álvarez-Esteban, E. del Barrio, J.A. Cuesta-Albertos and C. Matrán, A fixed-point approach to barycenters in Wasserstein space. *J. Math. Anal. Appl.* **441** (2016) 744–762.

[22] J.A. Cuesta-Albertos, C. Matrán-Bea and A. Tuero-Diaz, On lower bounds for the 2-Wasserstein metric in a Hilbert space. *J. Theoret. Probab.* **9** (1996) 263–283.

[23] G. Friesecke, Optimal Transport: A Comprehensive Introduction to Modeling, Analysis, Simulation, Applications. Society for Industrial and Applied Mathematics, Philadelphia, PA (2024).

[24] Dalery, Maxime, et al. Nonlinear reduced basis using mixture Wasserstein barycenters: application to an eigenvalue problem inspired from quantum chemistry. in press in *Constructive Approximation* (2023).

Please help to maintain this journal in open access!

This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting subscribers@edpsciences.org.

More information, including a list of supporters and financial transparency reports, is available at https://edpsciences.org/en/subscribe-to-open-s2o.

Appendix A. Properties of optimal transport plans and Wasserstein barycenters for group invariant measures

In this Appendix, we prove some interesting results on the symmetry properties of optimal transport plans and Wasserstein barycenters between symmetric measures. The results of this section are not directly linked to the work on mixtures we presented in the core sections of this article, but they provide interesting insight on the symmetry properties of Wasserstein barycenters in particular for some of the group-invariant measures considered in Section 5. First, we prove that the optimal transport plan between symmetric measures in a multi-marginal context is symmetric. Second, we show that the Wasserstein barycenter between several symmetric measures is symmetric as well. Third, we prove some sparsity properties on the support of the optimal transport plan in the case where the symmetry is a reflection, as is the case in the quantum chemistry applications presented in Section 6.

We first study the symmetry of the transport plan between symmetric measures. Let $\mu_1, \dots, \mu_n$ be probability measures in $\mathcal{P}_p^c(\Omega)$ and $C : \Omega^n \to \mathbb{R}_+ \cup \{+\infty\}$ be a multi-dimensional cost function. We consider here the following multi-marginal optimal transport problem: find $\gamma \in \Pi(\mu_1, \dots, \mu_n)$ solution to

$$\inf_{\gamma \in \Pi(\mu_1,\dots,\mu_n)} \int_{\Omega^n} C \, \mathrm{d}\gamma. \tag{A.1}$$

We have the following proposition.

Proposition A.1. *Let $S : \Omega \to \Omega$ be a $\mathcal{C}^1$-diffeomorphism such that $|\det \nabla S(x)| = 1$ for all $x \in \Omega$. We introduce*

$$S_n : \left\{ \begin{array}{ccc} \Omega^n & \to & \Omega^n \\ (x_1, \cdots, x_n) & \mapsto & (S(x_1), \cdots, S(x_n)). \end{array} \right. \tag{A.2}$$

In addition, let $\mu_1, \dots, \mu_n$ be probability measures in $\mathcal{P}(\Omega)$ that are invariant under S. Let us assume that C is invariant with respect to S_n, that is

$$C(S_n(x_1, \dots, x_n)) = C(S(x_1), \dots, S(x_n)) = C(x_1, \dots, x_n), \quad \forall (x_1, \dots, x_n) \in \Omega^n, \tag{A.3}$$

and that there exists a unique solution γ to (A.1). Then γ is invariant with respect to S_n.

Proof. Let γ be the optimal transport plan solution to (A.1). Let $\widetilde{\gamma} := S_n \# \gamma$. Let us prove that $\widetilde{\gamma} = \gamma$. First, it is easy to see that for all $1 \le i \le n$, $P_i \# \widetilde{\gamma} = S \# \mu_i$. Since μ_i is invariant under S for all $1 \le i \le n$, it holds that $\widetilde{\gamma} \in \Pi(\mu_1, \cdots, \mu_n)$.

Moreover using a change of variables, we obtain from (A.3) that

$$\int_{\Omega^n} C(x_1, \dots, x_n) \, \mathrm{d}\widetilde{\gamma} = \int_{\Omega^n} C(S(x_1), \dots, S(x_n)) \, \mathrm{d}\gamma = \int_{\Omega^n} C(x_1, \dots, x_n) \, \mathrm{d}\gamma.$$

Thus, $\widetilde{\gamma}$ is also an optimal transport plan between $\mu_1, \dots, \mu_n$ which implies that $\widetilde{\gamma} = \gamma$, hence the desired result. □

We now prove that the 2-Wasserstein barycenter associated to the Euclidean distance between symmetric measures is symmetric.

Proposition A.2. *Let us assume that Ω is convex and that for all $x, y \in \Omega$, $c(x,y) = \|x-y\|$. Let $S : \Omega \to \Omega$ be a measurable map such that for all $x \in \Omega$, $S(x) = Ax + b$ with $b \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d\times d}$ a unitary matrix. Let $\mu_1, \dots, \mu_n$ be probability measures in $\mathcal{P}_p^c(\Omega)$ that are invariant under S. Let $\boldsymbol{t} = (t_1, \dots, t_n) \in \mathcal{L}_n$ and let us assume that there exists a unique optimal transport plan solution to (A.1) with $C = C_n^{\boldsymbol{t}}$ where for all $(x_1, \dots, x_n) \in \Omega^n$,*

$$C_n^{\boldsymbol{t}}(x_1, \dots, x_n) := \frac{1}{2} \sum_{1\le i,j\le n} t_i t_j c(x_i, x_j)^p. \tag{A.4}$$

Let ν be the p-Wasserstein barycenter between the measures $\mu_1, \dots, \mu_n$ with weigths $\boldsymbol{t}$. Then ν is also invariant under S.

Proof. It suffices to show that for all real-valued measurable functions $f : \Omega \to \mathbb{R}$, there holds

$$\int_\Omega f(x)\,\mathrm{d}\nu(x) = \int_\Omega f(S(x))\,\mathrm{d}\nu(x). \tag{A.5}$$

Let f be a measurable real-valued function defined on Ω. From [3], there holds $\nu = T\#\gamma$ with

$$T(x_1, \dots, x_n) = \sum_{i=1}^n t_i x_i, \quad \forall (x_1, \cdots, x_n) \in \Omega^n,$$

and γ the unique solution to (A.1) associated to the cost function $C = C_n^{\boldsymbol{t}}$ defined in (A.4).

Thus

$$\int_\Omega f(x)\,\mathrm{d}\nu(x) = \int_{\Omega^n} f(T(x_1, \dots, x_n))\,\mathrm{d}\gamma(x_1, \dots, x_n), \tag{A.6}$$

and

$$\int_\Omega f(S(x))\,\mathrm{d}\nu(x) = \int_{\Omega^n} f(S(T(x_1, \dots, x_n)))\,\mathrm{d}\gamma(x_1, \dots, x_n).$$

Using the assumption on S and the fact that $\sum_{i=1}^n t_i = 1$, we obtain

$$S(T(x_1, \dots, x_n)) = T(S(x_1), \cdots, S(x_n)), \quad \forall (x_1, \cdots, x_n) \in \Omega^n.$$

As a consequence, noting that the assumptions on S and Proposition A.1 show that γ is invariant under S_n where S_n is the map defined by (A.2), and using the assumption on the cost function (A.3),

$$\begin{aligned}\int_\Omega f(S(x))\,\mathrm{d}\nu(x) &= \int_{\Omega^n} f(T(S(x_1), \dots, S(x_n)))\,\mathrm{d}\gamma(x_1, \dots, x_n)\\ &= \int_{\Omega^n} f(T(S_n(x_1, \dots, x_n)))\,\mathrm{d}\gamma(x_1, \dots, x_n)\\ &= \int_{\Omega^n} f(T(x_1, \dots, x_n))\,\mathrm{d}\gamma(x_1, \dots, x_n).\end{aligned}$$

Using (A.6) finishes the proof. □

Finally, we prove that when the considered symmetry is a reflection and the domain can be split into two parts such that $S(\Omega_1) = \Omega_2$ and $S(\Omega_2) = \Omega_1$, the optimal transport plan between symmetric measures is zero on many parts of the domain Ω^n. Indeed, the only parts of the domain where the optimal transport plan is nonzero are Ω_1^n and Ω_2^n.

Proposition A.3. *Let $\mu_1, \dots, \mu_n$ be probability measures in $\mathcal{P}_p^c(\Omega)$ invariant under a map $S : \Omega \to \Omega$. We assume that S is a reflection (i.e. S is an isometry in the sense that $c(S(x), S(y)) = c(x, y)$ for all $x, y \in \Omega$ and such that $S^2 = Id$) and that there exist Ω_1, Ω_2 such that $\Omega_1 \cap \Omega_2 = \emptyset$, $\Omega_1 \cup \Omega_2 = \Omega$, and that $S(\Omega_1) = \Omega_2$, $S(\Omega_2) = \Omega_1$. We also assume that for all $(x, y) \in (\Omega_1 \times \Omega_1) \cup (\Omega_2 \times \Omega_2)$,*

$$c(x, y) \leq c(S(x), y).$$

Let $\boldsymbol{t} = (t_1, \dots, t_n) \in \mathcal{L}_n$. Let us assume that there exists a unique optimal transport plan $\gamma \in \Pi(\mu_1, \dots, \mu_n)$ solution to (A.1) with $C = C_n^{\boldsymbol{t}}$ defined in (A.4). Then, the support of γ is included in $\Omega_1^n \cup \Omega_2^n$.

Proof. For any $\boldsymbol{i} = (i_1, i_2, \dots, i_n) \in \{0, 1\}^n$, denoting by $\boldsymbol{x} = (x_1, \dots, x_n) \in \Omega^n$, we define the map

$$S_{\boldsymbol{i}} : \left\{ \begin{array}{ccc} \Omega^n & \to & \Omega^n \\ \boldsymbol{x} & \mapsto & (S^{i_1}(x_1), S^{i_2}(x_2), \dots, S^{i_n}(x_n)), \end{array} \right. \tag{A.7}$$

so that $S^2_{i_1, i_2, \dots, i_n} = \mathrm{Id}$. We define a non-negative measure $\widetilde{\gamma}$ on Ω^n as follows: for any measurable subset $B \subset \Omega^n$, we define

$$\widetilde{\gamma}(B) = \frac{1}{2} \sum_{\boldsymbol{i} \in \{0,1\}^n} (S_{\boldsymbol{i}} \# \gamma) \left(B \cap (\Omega_1^n \cup \Omega_2^n)\right). \tag{A.8}$$

By construction, the support of $\tilde{\gamma}$ is included in $\Omega_1^n \cup \Omega_2^n$. Let us prove that $\tilde{\gamma}$ is an optimal Wasserstein transport plan between $\mu_1, \dots, \mu_n$. To this aim, we first show that $\tilde{\gamma}$ has for marginals $\mu_1, \dots, \mu_n$, starting with the first marginal, the others being dealt with similarly. Let $f : \Omega \to \mathbb{R}$ be a measurable map and define by $F : \Omega^n \to \mathbb{R}$ the function such that $F(\boldsymbol{x}) = f(x_1)$ for all $\boldsymbol{x} \in \Omega^n$. Using (A.8) we obtain

$$\int_{\Omega^n} f(x_1) \, \mathrm{d}\tilde{\gamma}(\boldsymbol{x}) = \int_{\Omega^n} F(\boldsymbol{x}) \, \mathrm{d}\tilde{\gamma}(\boldsymbol{x}) = \frac{1}{2} \sum_{\boldsymbol{i} \in \{0,1\}^n} \int_{\Omega_1^n \cup \Omega_2^n} F(\boldsymbol{x}) \, d(S_{\boldsymbol{i}} \# \gamma)(\boldsymbol{x}).$$

Introducing functions S^{i_k} in the arguments of the function F which do not change the values of F for k from 2 to n, and then using a change of variables and writing explicitly the sum on i_1, we obtain

$$\begin{aligned} \int_{\Omega^n} f(x_1) \, \mathrm{d}\tilde{\gamma}(\boldsymbol{x}) &= \frac{1}{2} \sum_{\boldsymbol{i} \in \{0,1\}^n} \int_{\Omega_1^n \cup \Omega_2^n} F(x_1, S^{i_2}(x_2), \dots, S^{i_n}(x_n)) \, d(S_{\boldsymbol{i}} \# \gamma)(\boldsymbol{x}) \\ &= \frac{1}{2} \sum_{i_2, \cdots, i_n = 0}^{1} \int_{S_{0, i_2, \dots, i_n}(\Omega_1^n \cup \Omega_2^n)} F(\boldsymbol{x}) \, \mathrm{d}\gamma(\boldsymbol{x}) \\ &+ \frac{1}{2} \sum_{i_2, \dots, i_n = 0}^{1} \int_{S_{1, i_2, \dots, i_n}\left(\Omega_1^n \cup \Omega_2^n\right)} F(S(x_1), x_2, \dots, x_n) \, \mathrm{d}\gamma(\boldsymbol{x}). \end{aligned}$$

Using properties of S leads to

$$\begin{aligned} \int_{\Omega^n} f(x_1) \, \mathrm{d}\tilde{\gamma}(\boldsymbol{x}) &= \frac{1}{2} \int_{(\Omega_1 \cup \Omega_2) \times \Omega^{n-1}} \left[F(\boldsymbol{x}) + F(S(x_1), x_2, \dots, x_n)\right] \, \mathrm{d}\gamma(\boldsymbol{x}) \\ &= \frac{1}{2} \int_{\Omega} f(x_1) \, \mathrm{d}\mu_1(x_1) + \frac{1}{2} \int_{\Omega} f(S(x_1)) \, \mathrm{d}\mu_1(x_1). \end{aligned}$$

Noting that f is symmetric under S, we obtain that the first marginal of $\widetilde{\gamma}$ is μ_1. The proof for the other marginals are similar. Hence $\widetilde{\gamma} \in \Pi(\mu_1, \dots, \mu_n)$. Now, let us prove that $\tilde{\gamma}$ is an optimal Wasserstein transport plan. Indeed, noting that

$$c(S(x), S(y)) = c(x, y) \ \text{ if } (x, y) \in (\Omega_1 \times \Omega_1) \cup (\Omega_2 \times \Omega_2)$$

and

$$c(S(x), y) \leq c(x, y) \text{ if } (x, y) \in (\Omega_1 \times \Omega_2) \cup (\Omega_2 \times \Omega_1),$$

there holds

$$\begin{aligned}
\int_{\Omega^n} \sum_{1\leq k,l\leq n} t_k t_l c(x_k, x_l)^p \,\mathrm{d}\tilde{\gamma}(\boldsymbol{x}) &= \frac{1}{2} \sum_{\boldsymbol{i}\in\{0,1\}^n} \int_{\Omega_1^n\cup\Omega_2^n} \sum_{1\leq k,l\leq n} t_k t_l c(x_k, x_l)^p d(S_{\boldsymbol{i}}\#\gamma)(\boldsymbol{x}) \\
&= \frac{1}{2} \sum_{\boldsymbol{i}\in\{0,1\}^n} \int_{S_{\boldsymbol{i}}(\Omega_1^n\cup\Omega_2^n)} \sum_{1\leq k,l\leq n} t_k t_l c(S^{i_k}(x_k), S^{i_l}(x_l))^p \,\mathrm{d}\gamma(\boldsymbol{x}) \\
&\leq \frac{1}{2} \sum_{\boldsymbol{i}\in\{0,1\}^n} \int_{S_{\boldsymbol{i}}(\Omega_1^n\cup\Omega_2^n)} \sum_{1\leq k,l\leq n} t_k t_l c(x_k, x_l)^p \,\mathrm{d}\gamma(\boldsymbol{x}) \\
&= \int_{\Omega^n} \sum_{1\leq k,l\leq n} t_k t_l c(x_k, x_l)^p \,\mathrm{d}\gamma(\boldsymbol{x}).
\end{aligned}$$

Therefore, γ being a Wasserstein optimal transport plan between $\mu_1, \ldots, \mu_n$, so is $\tilde{\gamma}$. The desired result then follows from the uniqueness of the optimal transport plan. □

Appendix B. Proof of Proposition 3.12

For completeness, we now present the proof of Proposition 3.12.

Proof. Let $\mu_0 = \sum_{j=1}^J \lambda_0^j a_0^j$, $\mu_1 = \sum_{k=1}^K \lambda_1^k a_1^k$ be two mixtures in $\mathcal{M}(\mathcal{A})$, with a_0^j (respectively a_1^k) all distinct. First, let $w^* \in \mathbb{R}^{J\times K}$ be a solution to (3.1), and let for $j = 1, \ldots, J$ and $k = 1, \ldots, K$, $\gamma_{jk} := \gamma^*_{a_0^j, a_1^k}$ so that $\gamma_{jk} \in \Gamma^*_{a_0^j, a_1^k} \subset \Gamma_{a_0^j, a_1^k}$. Then let us define $\gamma^* = \sum_{j=1}^J \sum_{k=1}^K w^*_{jk}\gamma_{jk}$. There holds $\gamma^* \in \Gamma_{\mathcal{M}}(\mathcal{A}) \cap \Pi(\mu_0, \mu_1)$. Moreover,

$$\begin{aligned}
\widetilde{\delta}_{\mathcal{M},p}(\mu_0, \mu_1)^p &\leq \sum_{j=1}^J \sum_{k=1}^K w^*_{jk} \int_{\Omega\times\Omega} c(x, y)^p \mathrm{d}\gamma_{jk}(x, y) = \sum_{j=1}^J \sum_{k=1}^K w^*_{jk} \delta^p(a_0^j, a_1^k) \\
&= \delta_{\mathcal{M},p}(\mu_0, \mu_1)^p.
\end{aligned}$$

Second, let $\gamma \in \Gamma_{\mathcal{M}}(\mathcal{A})$. Then, there exists $I \in \mathbb{N}^*$ such that $\gamma = \sum_{i=1}^I \lambda^i \gamma^i$ with $\gamma^i \in \Gamma(\mathcal{A})$. Using that $P_0\#\gamma = \mu_0$, we obtain

$$\sum_{i=1}^I \lambda^i P_0\#\gamma^i = \sum_{j=1}^J \lambda_0^j a_0^j.$$

Using the fact that for all $1 \leq i \leq I$, $P_0\#\gamma^i \in \mathcal{A}$ and using the identifiability Assumption 3.2, we obtain that for each $i = 1, \ldots, I$, there exists $j \in \{1, \ldots, J\}$ such that $P_0\#\gamma^i = a_0^j$. Similarly, for each $i = 1, \ldots, I$, there exists $k \in \{1, \ldots, K\}$ such that $P_1\#\gamma^i = a_1^k$. Hence for all $1 \leq i \leq I$, $\gamma^i \in \Pi(a_0^j, a_1^k)$ for some $j \in \{1, \ldots, J\}, k \in \{1, \ldots, J\}$. For all $1 \leq j \leq J$ and $1 \leq k \leq K$, let $I_{jk} := \left\{1 \leq i \leq I,\ \gamma^i \in \Pi(a_0^j, a_1^k)\right\}$. The collection of sets $(I_{jk})_{1\leq j\leq J, 1\leq k\leq K}$ then forms a partition of the set $\{1, \ldots, I\}$ and for all $1 \leq j \leq J$ and $1 \leq k \leq K$, we denote by $\Lambda_{jk} := \sum_{i\in I_{jk}} \lambda^i$. For all $1 \leq j \leq J$ and $1 \leq k \leq K$, we also denote by

$$\overline{\gamma}_{jk} := \begin{cases} \frac{1}{\Lambda_{jk}} \sum_{i\in I_{jk}} \lambda^i \gamma^i & \text{if } \Lambda_{jk} \neq 0, \\ \gamma^*_{a_0^j, a_1^k} & \text{if } \Lambda_{jk} = 0. \end{cases}$$

Then, it necessarily holds that for all $1 \leq j \leq J$, $1 \leq k \leq K$, $\overline{\gamma}_{jk} \in \Gamma_{a_0^j, a_1^k}$ since $\Gamma_{a_0^j, a_1^k}$ is a convex set and $\gamma^i \in \Gamma_{a_0^j, a_1^k}$ for all $i \in I_{jk}$. Furthermore, we have $\gamma = \sum_{j=1}^J \sum_{k=1}^K \Lambda_{jk} \overline{\gamma}_{jk}$. Thus,

$$\begin{aligned}\int_{\Omega\times\Omega} c(x,y)^p \mathrm{d}\gamma(x,y) &= \sum_{j=1}^J \sum_{k=1}^K w_{jk} \int_{\Omega\times\Omega} c(x,y)^p \mathrm{d}\overline{\gamma}_{jk}(x,y) \\ &\geq \sum_{j=1}^J \sum_{k=1}^K w_{jk} \delta^p(a_0^j, a_1^k) \geq \delta^p_{\mathcal{M},p}(\mu_0,\mu_1),\end{aligned}$$

where we have used (3.2). This inequality being valid for any $\gamma \in \Gamma_{\mathcal{M}}(\mathcal{A})$, we conclude that $\widetilde{\delta}_{\mathcal{M},p}(\mu_0,\mu_1) \geq \delta_{\mathcal{M},p}(\mu_0,\mu_1)$. Equation (3.5) is then a straightforward consequence of this proof. □

Appendix C. Explicit computations for elliptical distributions

In this section, we provide the explicit calculations of the means and covariance matrices of elliptical distributions.

Lemma C.1. *For a given $h : \mathbb{R}^+ \to \mathbb{R}^+$, then the elliptical distribution defined on $\mathbb{R}^d$ via*

$$\forall x \in \mathbb{R}^d, \quad g_{m,\Sigma}(x) = \frac{1}{Z_{m,\Sigma,h}} h((x-m)^T \Sigma^{-1}(x-m)),$$

with

$$Z_{m,\Sigma,h} = \frac{2\pi^{d/2}}{\Gamma(d/2)} |\det \Sigma|^{1/2} \int_{\mathbb{R}^+} r^{d-1} h(r^2) \mathrm{d}r, \tag{C.1}$$

has mean m. Moreover if

$$\frac{\int_{\mathbb{R}^+} r^{d+1} h(r^2)\mathrm{d}x}{\int_{\mathbb{R}^+} r^{d-1} h(r^2)\mathrm{d}r} = d, \tag{C.2}$$

then $g_{m,\Sigma}$ has covariance Σ.

Proof. First, let us check that $g_{m,\Sigma}$ indeed defines a probability distribution. Using a first change of variable $y = \Sigma^{-1/2}(x-m)$, a second change of variables with spherical coordinates, and using that the surface of the $d-$sphere of radius r is $\frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1}$ we obtain

$$\int_{\mathbb{R}^d} h((x-m)^T \Sigma^{-1}(x-m))\mathrm{d}x = \frac{2\pi^{d/2}}{\Gamma(d/2)} |\det \Sigma|^{1/2} \int_{\mathbb{R}^+} r^{d-1} h(r^2)\mathrm{d}r,$$

from which we easily deduce using (C.1) that $g_{m,\Sigma}$ is normalized to one. Second, the mean of $g_{m,\Sigma}$ satisfies

$$\int_{\mathbb{R}^d} x g_{m,\Sigma}(x)\mathrm{d}x = \int_{\mathbb{R}^d} (x+m) g_{0,\Sigma}(x)\mathrm{d}x = m,$$

where we have used a change of variable, and that $g_{0,\Sigma}(-x) = g_{0,\Sigma}(x)$ for all $x \in \mathbb{R}^d$. Third, the covariance matrix of $g_{m,\Sigma}$ can be computed as

$$\int_{\mathbb{R}^d} (x-m)(x-m)^T g_{m,\Sigma}(x)\mathrm{d}x = \Sigma^{1/2} \int_{\mathbb{R}^d} y y^T g_{0,I}(y)\mathrm{d}y [\Sigma^{1/2}]^T.$$

Denoting $y = (y_1, \ldots, y_d)$, noting that by symmetry, the off-diagonal terms $\int_{\mathbb{R}^d} y_i y_j g_{0,I}(y)\mathrm{d}y$ are zero and that the diagonal terms $\int_{\mathbb{R}^d} y_i^2 g_{0,I}(y)\mathrm{d}y$ are all equal to

$$\frac{1}{d}\int_{\mathbb{R}^d} \|y\|^2 g_{0,I}(y)\mathrm{d}y = \frac{2\pi^{d/2}}{d\Gamma(d/2)}\int_{\mathbb{R}^+} r^{d+1} h(r^2)\mathrm{d}r,$$

we obtain the result. □