



**HAL**  
open science

# Marginal-preserving modified Wasserstein barycenters for Gaussian distributions and Gaussian mixtures

Maxime Dalery, Geneviève Dusson, Virginie Ehrlacher

► **To cite this version:**

Maxime Dalery, Geneviève Dusson, Virginie Ehrlacher. Marginal-preserving modified Wasserstein barycenters for Gaussian distributions and Gaussian mixtures. 2024. hal-04696783

**HAL Id: hal-04696783**

**<https://hal.science/hal-04696783v1>**

Preprint submitted on 13 Sep 2024

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

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

# MARGINAL-PRESERVING MODIFIED WASSERSTEIN BARYCENTERS FOR GAUSSIAN DISTRIBUTIONS AND GAUSSIAN MIXTURES <sup>\*</sup>

MAXIME DALERY <sup>\*</sup>, GENEVIÈVE DUSSON <sup>\*</sup>, AND VIRGINIE EHRLACHER <sup>†</sup>

**Abstract.** Wasserstein barycenters do not preserve marginals in general. In this work, we first characterize sufficient and necessary conditions for the Wasserstein barycenter between two Gaussian distributions to preserve marginals, and provide necessary conditions in the case of more than two Gaussians. We then propose modified Wasserstein barycenters that preserve the marginals of the distributions, both for Gaussian distributions and for mixtures of Gaussian distributions. In the case of Gaussian distributions, the marginal-preserving modified Wasserstein barycenters can be analytically computed, while for Gaussian mixtures, computing the marginal-preserving barycenter consists in a postprocessing of the Gaussian mixture Wasserstein barycenter. In both cases, we provide numerical simulations illustrating the difference between Wasserstein barycenters and modified marginal-preserving Wasserstein barycenters.

**1. Introduction.** Barycenters of a family of probability measures with finite second-order moments with respect to the Wasserstein-2 metric have been introduced and studied in the seminal paper [2]. Since then, these Wasserstein barycenters have been successfully used in a large variety of practical applications, ranging from imaging [17, 13], data science [18, 9], model order reduction techniques [12], tomographic reconstruction [1], and quantum chemistry [10]. In all these works, Wasserstein barycenters are used to build meaningful interpolations between various probability measures that are relevant to the considered application.

The motivation of the present work stems from the following well-known observation: given  $n_x$  and  $n_y$  two positive integers,  $\rho_0$  and  $\rho_1$  two probability measures defined on the cartesian product  $\mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$  with finite second-order moments, the marginals of the Wasserstein-2 barycenter between  $\rho_0$  and  $\rho_1$ , respectively in  $\mathbb{R}^{n_x}$  and  $\mathbb{R}^{n_y}$  are not equal in general to the Wasserstein-2 barycenters of the marginals of  $\rho_0$  and  $\rho_1$  respectively.

This is an issue in applicative contexts where interpolations between two or more probability measures defined on a cartesian product set have to be considered in a Wasserstein sense but where the preservation of marginals by the interpolation procedure is necessary to obtain relevant quantities. Indeed, in the study of statistical properties of large particle or agent systems, two-body correlations enable to obtain very precious insight on the statistical distribution of the particles or agents in the underlying system. The latter are probability measures defined on a cartesian product domain, that is the product of the set of possible particle or agent positions in the system with itself, the marginals of which are equal to the one-body particle density of the system. Interpolating between two-body correlations or one-body densities in a Wasserstein sense is meaningful in many situations, however, due to the issue highlighted above, the marginal of the interpolated two-body correlation is not equal to the interpolation of the one-body density in general, which is not satisfactory from a modeling point of view. This is the case for instance in quantum chemistry, see [10]. Other applications include population dynamics, where the populations are observed through the spectrum of several parameters, compartmental models in epidemiology, medical imaging with e.g. tumor growth study, as well as crowd modeling or kinetic equations in physics.

---

<sup>2</sup>Ecole Nationale des Ponts et Chaussées, CERMICS, and INRIA, MATHERIALS team-project, 6 & 8 Av. Pascal, 77455 Marne-la-Vallée, France, virginie.ehrlacher@enpc.fr.

<sup>1</sup>Université de Franche-Comté, CNRS, LmB, 25000 Besançon, France, maxime.dalery@univ-fcomte.fr, genevieve.dusson@math.cnrs.fr.

<sup>\*</sup> Funded/Co-funded by the European Union (ERC, HighLEAP, 101077204). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. GD was supported by the French ‘Investissements d’Avenir’ program, project Agence Nationale de la Recherche (ISITE-BFC) (contract ANR-15-IDEX-0003). GD was also supported by the Ecole des Ponts-ParisTech and region Bourgogne Franche-Comté.

It is thus of practical interest to define interpolations between measures defined on a cartesian product domain that are close to Wasserstein barycenters, with marginals that are guaranteed to be the Wasserstein barycenters of their marginals. The only attempt we are aware of in this direction is [1] where the authors propose a modified version of Wasserstein barycenters through an optimization problem where the marginal constraint was introduced in a penalized form. In the present work, the proposed modified Wasserstein barycenters exactly satisfy the marginal constraints. For simplicity, we restrict ourselves to Gaussian measures and Gaussian mixtures, which is already of significant practical interest in many applications including those related to quantum chemistry. Our main contributions are the following: (i) we give an explicit necessary and sufficient condition on the covariance matrices of two Gaussian measures for the marginals of their Wasserstein barycenter to be the Wasserstein barycenter of their marginals, whatever the value of the barycentric weights. We also show that, if this condition is not satisfied, for any non-trivial barycentric weights, the marginals of their Wasserstein barycenter is not equal to the Wasserstein barycenter of their marginals; (ii) we propose two definitions of modified marginal-preserving Wasserstein barycenters, prove their well-posedness; one of these definitions leads to an analytically computable barycenter (iii) we use this modified barycenter to define marginal-preserving barycenters for Gaussian mixtures with respect to the mixture Wasserstein distance (see for instance [11, 8]), for which the computation is done as a post-processing of the mixture Wasserstein barycenter; (iv) we illustrate the alternative definitions we propose here through various numerical tests to highlight the differences between the proposed marginal-preserving modified Wasserstein barycenters and the exact Wasserstein barycenters.

The outline of the article is the following. In Section 2, we recall the main classical notions related to optimal transport theory between Gaussian measures that will be useful in our analysis. In Section 3, we prove several results on the geometric mean between covariance matrices that will be useful in our analysis. In Section 4, we state our first main result concerning the necessary and sufficient condition on covariance matrices of Gaussian measures for their Wasserstein barycenter to be marginal-preserving. In Section 5, we introduce the modified marginal-preserving barycenters for Gaussian distributions. In Section 6, we extend the definition by proposing modified marginal-preserving barycenters for mixtures of Gaussian distributions. In Section 7, we provide numerical results illustrating the findings. Finally, in Section 8, we give some concluding remarks.

**2. Optimal transport for Gaussian distributions.** The aim of this section is to summarize some basic facts about optimal transport theory and Gaussian measures, that can e.g. be found in [16]. We first recall some fundamentals about the Wasserstein metric in Section 2.1. We then present some well-known properties of the Wasserstein metric for Gaussian measures in Section 2.2.

**2.1. Wasserstein-2 metric.** In the whole document we denote by  $\mathbb{N}^*$  the set  $\mathbb{N} \setminus \{0\}$ . Let  $n \in \mathbb{N}^*$  and let  $\mathcal{P}_2(\mathbb{R}^n)$  denote the set of probability measures on  $\mathbb{R}^n$  with finite second-order moments. The 2-Wasserstein distance over  $\mathcal{P}_2(\mathbb{R}^n)$  is defined for  $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^n)$  as

$$W_2(\rho_0, \rho_1) := \inf_{\pi \in \Pi(\rho_0, \rho_1)} \left( \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^2 d\pi(x, y) \right)^{1/2},$$

where  $\|\cdot\|$  denotes the euclidean norm of  $\mathbb{R}^n$  and  $\Pi(\rho_0, \rho_1)$  the set of probability measures over  $\mathbb{R}^n \times \mathbb{R}^n$  with marginals  $\rho_0$  and  $\rho_1$ , which is called the set of transport plans between  $\rho_0$  and  $\rho_1$ . More precisely,

$$\Pi(\rho_0, \rho_1) = \left\{ \pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^n) : \int_{y \in \mathbb{R}^n} d\pi(x, y) = d\rho_0(x), \int_{x \in \mathbb{R}^n} d\pi(x, y) = d\rho_1(y) \right\}.$$

For  $M \in \mathbb{N}^*$ , let us denote by

$$\Lambda_M = \left\{ (\lambda_1, \dots, \lambda_M) \in (\mathbb{R}_+)^M, \sum_{m=1}^M \lambda_m = 1 \right\}$$

the set of barycentric weights of cardinality  $M$ . The Wasserstein barycenter of a collection of  $M$  probability measures  $\boldsymbol{\rho} := (\rho^1, \dots, \rho^M) \in \mathcal{P}_2(\mathbb{R}^n)^M$  associated to a set of barycentric weights  $\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_M) \in \Lambda_M$  is then defined (see [2]) as the unique solution to the problem

$$(2.1) \quad \inf_{\rho \in \mathcal{P}_2(\mathbb{R}^n)} \sum_{m=1}^M \lambda_m W_2(\rho, \rho^m)^2.$$

The unique minimizer of (2.1) is denoted by  $\text{Bar}_{W_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho})$ . This barycenter is also related to the so-called multi-marginal optimal transport problem [2, 14, 15], defined, given  $M$  elements  $\boldsymbol{\rho} = (\rho^1, \dots, \rho^M)$  in  $\mathcal{P}_2(\mathbb{R}^n)^M$ , as

$$(2.2) \quad mmW_2(\boldsymbol{\rho}; \boldsymbol{\lambda}) := \inf_{\pi \in \Pi(\rho^1, \dots, \rho^M)} \left( \int_{(\mathbb{R}^n)^M} \frac{1}{2} \sum_{m_1, m_2=1}^M \lambda_{m_1} \lambda_{m_2} \|x_{m_1} - x_{m_2}\|^2 d\pi(x_1, \dots, x_M) \right)^{1/2},$$

where  $\Pi(\rho^1, \dots, \rho^M)$  is the set of probability measures over  $(\mathbb{R}^n)^M$  with marginals  $\rho_1, \dots, \rho_M$ , and there holds

$$mmW_2(\boldsymbol{\rho}; \boldsymbol{\lambda}) = \left[ \sum_{m=1}^M \lambda_m W_2(\text{Bar}_{W_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}), \rho^m)^2 \right]^{1/2}.$$

In all the sequel, we will assume that  $n = n_x + n_y$  for some  $n_x, n_y \in \mathbb{N}^*$ , so that  $\mathbb{R}^n = \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ . For a given probability measure  $\rho \in \mathcal{P}_2(\mathbb{R}^{n_x} \times \mathbb{R}^{n_y})$ , we will denote by  $\text{marg}_x(\rho) \in \mathcal{P}_2(\mathbb{R}^{n_x})$  the first marginal of  $\rho$  and by  $\text{marg}_y(\rho) \in \mathcal{P}_2(\mathbb{R}^{n_y})$  its second marginal. More precisely,

$$d\text{marg}_x(\rho)(x) = \int_{y \in \mathbb{R}^{n_y}} d\rho(x, y) \quad \text{and} \quad d\text{marg}_y(\rho)(y) = \int_{x \in \mathbb{R}^{n_x}} d\rho(x, y).$$

**2.2. Gaussian measures.** For any  $n \in \mathbb{N}^*$ , let us denote by  $\mathcal{S}_{+,*}^n$  (respectively  $\mathcal{S}_+^n, \mathcal{S}^n$ ) the set of symmetric definite positive (respectively symmetric semi-definite positive, symmetric) matrices of  $\mathbb{R}^{n \times n}$ . For any  $S \in \mathcal{S}_n$ , we denote by  $\|S\|_2$  its operator norm. For any  $S \in \mathcal{S}_+^n$ , we denote by  $S^{1/2}$  or by  $\sqrt{S}$  its unique symmetric semi-definite positive square root. For any  $S \in \mathcal{S}_{+,*}^n$ , we also denote by  $S^{-1/2}$  the inverse of its square root. For any  $\mu \in \mathbb{R}^n$  and any  $S \in \mathcal{S}_{+,*}^n$ , we denote by  $\mathcal{N}(\mu, S)$  the Gaussian probability measure on  $\mathbb{R}^n$  with mean  $\mu$  and covariance matrix  $S$ .

If for  $i \in \{0, 1\}$ ,  $\rho_i = \mathcal{N}(\mu_i, S_i)$  is the Gaussian measure with mean  $m_i$  and covariance matrix  $S_i$ , with  $\mu_i \in \mathbb{R}^n$  and  $S_i \in \mathcal{S}_{+,*}^n$ , the 2-Wasserstein distance between  $\rho_0$  and  $\rho_1$  has a closed form expression which can be written as

$$W_2(\rho_0, \rho_1)^2 = \|\mu_1 - \mu_0\|^2 + \text{Tr} \left( S_0 + S_1 - 2 \left( \sqrt{S_0} S_1 \sqrt{S_0} \right)^{1/2} \right).$$

In the following, we denote by  $\mathcal{W}_2 : \mathcal{S}_+^n \times \mathcal{S}_+^n \rightarrow \mathbb{R}_+$  the Bures–Wasserstein metric [7]

$$\forall S_0, S_1 \in \mathcal{S}_+^n, \quad \mathcal{W}_2(S_0, S_1)^2 := \text{Tr} \left( S_0 + S_1 - 2 \left( \sqrt{S_0} S_1 \sqrt{S_0} \right)^{1/2} \right),$$

so that

$$\forall \mu_0, \mu_1 \in \mathbb{R}^n, \quad \forall S_0, S_1 \in \mathcal{S}_{+,*}^n, \quad W_2(\mathcal{N}(\mu_0, S_0), \mathcal{N}(\mu_1, S_1))^2 = \|\mu_0 - \mu_1\|^2 + \mathcal{W}_2(S_0, S_1)^2.$$

For any  $M \in \mathbb{N}^*$ ,  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_M) \in \Lambda_M$  and  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_M) \in \mathcal{P}_2(\mathbb{R}^n)^M$  so that for all  $i \in \{1, \dots, M\}$ ,  $\rho_i = \mathcal{N}(\mu_i, S_i)$  for some  $\mu_i \in \mathbb{R}^n$  and  $S_i \in \mathcal{S}_{+,*}^n$ , it holds that  $\text{Bar}_{\mathcal{W}_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}) = \mathcal{N}(\mu_*, S_*)$  where

$$\mu_* = \sum_{m=1}^M \lambda_m \mu_m,$$

and  $S_* \in \mathcal{S}_{+,*}^n$  is the unique symmetric positive definite matrix solution to the following equation

$$\sum_{m=1}^M \lambda_m \left( \sqrt{S_*} S_m \sqrt{S_*} \right)^{1/2} = S_*.$$

In the sequel, we will denote  $S_*$  by  $\text{Bar}_{\mathcal{W}_2}^{\boldsymbol{\lambda}}(\boldsymbol{S})$  where  $\boldsymbol{S} := (S_1, \dots, S_M)$ .

Let us note that, in the case where  $n = n_x + n_y$  for some  $n_x, n_y \in \mathbb{N}^*$ , if  $\rho = \mathcal{N}(\mu, S)$  with  $\mu = (\mu_x, \mu_y)$  for some  $\mu_x \in \mathbb{R}^{n_x}$  and  $\mu_y \in \mathbb{R}^{n_y}$ , and  $S \in \mathcal{S}_{+,*}^{n_x+n_y}$  written as a block matrix as  $S = \begin{pmatrix} S_x & S_{xy} \\ S_{xy}^\top & S_y \end{pmatrix}$  with  $S_x \in \mathcal{S}_{+,*}^{n_x}$ ,  $S_y \in \mathcal{S}_{+,*}^{n_y}$  and  $S_{xy} \in \mathbb{R}^{n_x \times n_y}$ , it holds that

$$\text{marg}_x(\rho) = \mathcal{N}(\mu_x, S_x) \quad \text{and} \quad \text{marg}_y(\rho) = \mathcal{N}(\mu_y, S_y).$$

This motivates us to introduce the following applications, which we call hereafter with a slight abuse of language, the marginals of covariance matrices:

$$\mathcal{M}_x : \left\{ \begin{array}{l} \mathcal{S}_+^{n_x+n_y} \rightarrow \mathcal{S}_+^{n_x} \\ S = \begin{pmatrix} S_x & S_{xy} \\ S_{xy}^\top & S_y \end{pmatrix} \mapsto S_x \end{array} \right\} \text{ and } \mathcal{M}_y : \left\{ \begin{array}{l} \mathcal{S}_+^{n_x+n_y} \rightarrow \mathcal{S}_+^{n_y} \\ S = \begin{pmatrix} S_x & S_{xy} \\ S_{xy}^\top & S_y \end{pmatrix} \mapsto S_y \end{array} \right\}.$$

The aim of Sections 3, 4 and 5 is to propose some definitions of marginal-preserving Wasserstein-like barycenters for a family of Gaussian distributions. The aim of Section 6 is to extend the definition of such preserving Wasserstein-like barycenters to families of mixtures of Gaussian distributions [11].

The definitions of modified marginal-preserving barycenters of a family of Gaussian distributions, which will be the object of Section 5.3, will be very closely linked to properties of their covariance matrices, and in particular their geometric mean. Hence, in Section 3 and 4, we first focus on the mathematical properties of these covariance matrices.

**3. Properties of the geometric mean of covariance matrices.** The aim of this section is to recall some well-known results about the geometric mean of positive definite matrices and prove some new results which will be useful in the rest of our analysis. Indeed this geometric mean studied by Bhatia and coauthors [5, Chapter 4] and [6] will play a key role in our subsequent analysis.

We begin by recalling the definition of the geometric mean of two symmetric positive definite matrices and its main properties.

DEFINITION 3.1. *Let  $n \in \mathbb{N}^*$  and  $S, T \in \mathcal{S}_{+,*}^n$ . We call geometric mean of  $S$  and  $T$  the matrix*

$$(3.1) \quad S\#T := S^{1/2} \left( S^{1/2} T^{-1} S^{1/2} \right)^{-1/2} S^{1/2}.$$

A few properties of this geometric mean will be useful.

LEMMA 3.2. [4, Theorem 4.1.3] *Let  $n \in \mathbb{N}^*$  and  $S, T \in \mathcal{S}_{+,*}^n$ . It then holds that*

- (i)  $S\#T$  is the unique matrix  $C \in \mathcal{S}_{+,*}^n$  solution to the equation  $CS^{-1}C = T$ ;
- (ii)  $S\#T = T\#S$ ;
- (iii)  $(S\#T)^{-1} = S^{-1}\#T^{-1}$ .

The following result is a preliminary lemma that will be very useful in the subsequent analysis.

LEMMA 3.3. *Let  $n_x, n_y \in \mathbb{N}^*$  and let  $S \in \mathcal{S}_{+,*}^{n_x+n_y}$  with block decomposition*

$$(3.2) \quad S = \begin{pmatrix} S_x & S_{xy} \\ S_{xy}^\top & S_y \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} T_x & T_{xy} \\ T_{xy}^\top & T_y \end{pmatrix},$$

with  $S_x, T_x \in \mathcal{S}_{+,*}^{n_x}$ ,  $S_y, T_y \in \mathcal{S}_{+,*}^{n_y}$  and  $S_{xy}, T_{xy} \in \mathbb{R}^{n_x \times n_y}$ . Let

$$(3.3) \quad \forall Z \in \mathcal{C}_{T_x, T_y} := \left\{ Z \in \mathbb{R}^{n_x \times n_y}, \|T_x^{-1/2} Z T_y^{-1/2}\|_2 < 1 \right\}, \quad T(Z) := \begin{pmatrix} T_x & Z \\ Z^\top & T_y \end{pmatrix}.$$

The function  $F$  defined as

$$(3.4) \quad F : \mathcal{C}_{T_x, T_y} \ni Z \mapsto \mathcal{W}_2(S, T(Z))^2 = \text{Tr} \left( S + T(Z) - 2 \left( \sqrt{S} T(Z) \sqrt{S} \right)^{1/2} \right)$$

is strictly convex. Moreover, the minimization problem

$$(3.5) \quad Z_{S,T}^* \in \underset{Z \in \mathcal{C}_{T_x, T_y}}{\text{argmin}} \mathcal{W}_2^2(S, T(Z)),$$

has a unique minimizer which is given by

$$(3.6) \quad Z_{S,T}^* = (S_x^{-1} \# T_x) S_{xy} (S_y^{-1} \# T_y).$$

REMARK 3.4. *Note that from the definition of  $\mathcal{C}_{T_x, T_y}$ , which is a convex cone, it is obvious to see that, for all  $Z \in \mathcal{C}_{T_x, T_y}$ ,  $T(Z) \in \mathcal{S}_{+,*}^{n_x+n_y}$  (see also [7, (14)]) with  $\mathcal{M}_x(T(Z)) = T_x$  and  $\mathcal{M}_y(T(Z)) = T_y$ .*

*Proof.* First, for  $n \in \mathbb{N}^*$ , the application  $\mathcal{S}_+^n \ni M \mapsto \text{Tr} M^{1/2}$  is a strictly operator concave function (see [5, Chapter V] and [7, Theorem 7]). Noting that  $\mathcal{C}_{T_x, T_y} \ni Z \mapsto \sqrt{S} T(Z) \sqrt{S}$  is an affine mapping, and that  $\mathcal{C}_{T_x, T_y} \ni Z \mapsto \text{Tr}(S + T(Z))$  is a constant function, we easily deduce that  $F$  is strictly (operator) convex.

We will now compute the first-order derivatives of  $F$  and show that there exists a unique stationary point belonging to  $\mathcal{C}_{T_x, T_y}$ . Note that, since  $\mathcal{C}_{T_x, T_y}$  is **not** a closed set, the existence of a minimizer is not obvious a priori.

For all  $1 \leq i, j \leq n$ , let us denote by  $J_{ij}$  the matrix

$$J_{ij} := \begin{pmatrix} 0_{n_x} & E_{ij} \\ E_{ij}^\top & 0_{n_y} \end{pmatrix}$$

where  $E_{ij} \in \mathbb{R}^{n_x \times n_y}$  is the matrix with all components equal to 0, except the  $(i, j)$ <sup>th</sup> component which is equal to 1. It is then easy to remark that for all  $1 \leq i \leq n_x, 1 \leq j \leq n_y$ , for all  $Z = (Z_{ij})_{1 \leq i \leq n_x, 1 \leq j \leq n_y} \in \mathcal{C}_{T_x, T_y}$ , it holds that  $\frac{\partial T}{\partial Z_{ij}}(Z) = J_{ij}$ . From [7, p. 182] the matrix square root defined on symmetric definite positive matrices  $\sqrt{\cdot} : M \in \mathcal{S}_+^n \mapsto \sqrt{M}$  has for derivative

$$\forall H \in \mathcal{S}^n, \quad d_{\sqrt{\cdot}}(M)(H) = \int_0^{+\infty} e^{-t\sqrt{M}} H e^{-t\sqrt{M}} dt,$$

from which we deduce that the application  $G : Z \in \mathcal{C}_{T_x, T_y} \mapsto \left( \sqrt{ST}(Z) \sqrt{S} \right)^{1/2}$  has partial derivatives equal to

$$\begin{aligned} \frac{\partial G}{\partial Z_{ij}}(Z) &= d_{\sqrt{\cdot}} \left( \sqrt{ST}(Z) \sqrt{S} \right) \left( \sqrt{S} \frac{\partial T}{\partial Z_{ij}}(Z) \sqrt{S} \right) \\ &= d_{\sqrt{\cdot}} \left( \sqrt{ST}(Z) \sqrt{S} \right) \left( \sqrt{S} J_{ij} \sqrt{S} \right) \\ (3.7) \quad &= \int_0^{+\infty} e^{-t(\sqrt{ST}(Z) \sqrt{S})^{1/2}} \left( \sqrt{S} J_{ij} \sqrt{S} \right) e^{-t(\sqrt{ST}(Z) \sqrt{S})^{1/2}} dt. \end{aligned}$$

Noting that  $\mathcal{C}_{T_x, T_y} \ni Z \mapsto \text{Tr}(S + T(Z))$  is a constant function, we see that

$$\frac{\partial F}{\partial Z_{ij}}(Z) = -2 \text{Tr} \left( \frac{\partial G}{\partial Z_{ij}}(Z) \right).$$

From (3.7) and using the cyclicity of the trace, we obtain

$$\frac{\partial F}{\partial Z_{ij}}(Z) = -2 \text{Tr} \left( \sqrt{S} \int_0^{+\infty} e^{-2t(\sqrt{ST}(Z) \sqrt{S})^{1/2}} dt \sqrt{S} J_{ij} \right).$$

Moreover, using that for any  $A \in \mathcal{S}_{+,*}^n$

$$\int_0^{+\infty} e^{-tA} dt = A^{-1},$$

with  $A = 2 \left( \sqrt{ST}(Z) \sqrt{S} \right)^{1/2}$  and (3.1) we obtain

$$\begin{aligned} \frac{\partial F}{\partial Z_{ij}}(Z) &= - \text{Tr} \left( \sqrt{S} \left( \sqrt{ST}(Z) \sqrt{S} \right)^{-1/2} \sqrt{S} J_{ij} \right) \\ &= - \text{Tr} \left( \left( T(Z)^{-1} \# S \right) J_{ij} \right) \\ &= -2 \left( T(Z)^{-1} \# S \right)_{xy, ij}, \end{aligned}$$

where  $\left(T(Z)^{-1}\#S\right)_{xy}$  is the right upper off-diagonal term of the matrix  $T(Z)^{-1}\#S$ . Hence the derivative of  $F$  is equal to 0 if and only if for all  $i = 1, \dots, n_x$  and all  $j = 1, \dots, n_y$ ,  $\left(T(Z)^{-1}\#S\right)_{xy,ij} = 0$ , that is if and only if  $T(Z)^{-1}\#S$  is block diagonal. This implies that the matrix  $S^{-1}\#T(Z)$  is also block diagonal. Since this matrix satisfies  $T(Z) = (S^{-1}\#T(Z))S(S^{-1}\#T(Z))$ , we obtain by computing the off-diagonal blocks that if there exists  $Z \in \mathcal{C}_{T_x, T_y}$  such that  $\frac{\partial F}{\partial Z_{ij}}(Z) = 0$  for all  $i = 1, \dots, n_x$  and all  $j = 1, \dots, n_y$ , then necessarily  $Z = Z_{S,T}^*$  where

$$(3.8) \quad Z_{S,T}^* := (S_x^{-1}\#T_x)S_{xy}(S_y^{-1}\#T_y).$$

Let us now prove that  $Z_{S,T}^*$  defined by (3.8) necessarily belongs to  $\mathcal{C}_{T_x, T_y}$ , which will yield the desired result. Since  $S \in \mathcal{S}_{+,*}^{n_x+n_y}$ , the Schur complement condition gives

$$S_x - S_{xy}S_y^{-1}S_{yx} \in \mathcal{S}_{+,*}^{n_x}.$$

Multiplying on both sides by  $(S_x^{-1}\#T_x) \in \mathcal{S}_{+,*}^{n_x}$ , we obtain

$$T_x - (S_x^{-1}\#T_x)S_{xy}S_y^{-1}S_{yx}(S_x^{-1}\#T_x) \in \mathcal{S}_{+,*}^{n_x}.$$

Noting that  $S_y^{-1} = (S_y^{-1}\#T_y)T_y^{-1}(S_y^{-1}\#T_y)$  leads to

$$I - T_x^{-1/2}(S_x^{-1}\#T_x)S_{xy}(S_y^{-1}\#T_y)T_y^{-1}(S_y^{-1}\#T_y)S_{yx}(S_x^{-1}\#T_x)T_x^{-1/2} \in \mathcal{S}_{+,*}^{n_x},$$

so that

$$\|T_x^{-1/2}(S_x^{-1}\#T_x)S_{xy}(S_y^{-1}\#T_y)T_y^{-1/2}\|_2 < 1,$$

that is  $\|T_x^{-1/2}Z_{S,T}^*T_y^{-1/2}\|_2 < 1$  and so  $Z_{S,T}^* \in \mathcal{C}_{T_x, T_y}$ . This concludes the proof.  $\square$

The following result links the blocks of covariance matrices and the structure of the geometric mean and will later play a role to characterize the covariance matrices with Wasserstein barycenters conserving marginals.

**LEMMA 3.5.** *Let  $n_x, n_y \in \mathbb{N}^*$  and let  $S, T \in \mathcal{S}_{+,*}^{n_x+n_y}$  with block decomposition (3.2), with  $S_x, T_x \in \mathcal{S}_{+,*}^{n_x}$ ,  $S_y, T_y \in \mathcal{S}_{+,*}^{n_y}$  and  $S_{xy}, T_{xy} \in \mathbb{R}^{n_x \times n_y}$ . Then, the four following assertions are equivalent.*

- (i) *the matrix  $S^{-1}\#T$  is block diagonal with block of dimensions of  $n_x \times n_x$  and  $n_y \times n_y$ ,*
- (ii) *the matrix  $S^{-1}\#T$  is equal to*

$$S^{-1}\#T = \begin{pmatrix} S_x^{-1}\#T_x & 0 \\ 0 & S_y^{-1}\#T_y \end{pmatrix},$$

- (iii) *the off-diagonal blocks verify*

$$T_{xy} = (S_x^{-1}\#T_x)S_{xy}(S_y^{-1}\#T_y);$$

- (iv) *we have the Pythagorean-type formula*

$$\mathcal{W}_2^2(S, T) = \mathcal{W}_2^2(S_x, T_x) + \mathcal{W}_2^2(S_y, T_y).$$

*In the following, if a pair  $(S, T) \in (\mathcal{S}_{+,*}^{n_x+n_y})^2$  meets one of these four assertions, we will write  $S \bowtie T$ .*

REMARK 3.6 (Not an equivalence relationship). *The relationship  $\bowtie$  is reflexive (i.e. for any  $S, T \in \mathcal{S}_{+,*}^{n_x+n_y}$ , if  $S \bowtie T$  then necessarily  $T \bowtie S$ ) but is not transitive. Indeed, for any  $R, S, T \in \mathcal{S}_{+,*}^{n_x+n_y}$ ,  $R \bowtie S$  and  $S \bowtie T$  do not necessarily imply that  $R \bowtie T$ . Hence,  $\bowtie$  is **not** an equivalence relationship.*

REMARK 3.7. *The solution to problem (3.5) satisfies  $T (Z_{S,T}^*) \bowtie S$ .*

*Proof.* Proof of (i)  $\Rightarrow$  (ii): If the matrix  $S^{-1}\#T$  is block diagonal with blocks  $C_x$  of dimension  $n_x \times n_x$  and  $C_y$  of dimension  $n_y \times n_y$ , then both  $C_x$  and  $C_y$  are symmetric positive definite matrices since  $S^{-1}\#T$  is positive definite. Moreover, noting that  $S^{-1}\#T$  is the only positive definite matrix such that

$$(S^{-1}\#T)S(S^{-1}\#T) = T,$$

and looking at the diagonal blocks identities, there holds

$$\begin{cases} C_x S_x C_x = T_x, \\ C_y S_y C_y = T_y. \end{cases}$$

Since  $S_x^{-1}\#T_x$  is the unique symmetric positive definite solution to the matrix equation  $C_x S_x C_x = T_x$ , there holds  $C_x = S_x^{-1}\#T_x$ . Similarly, we obtain  $C_y = S_y^{-1}\#T_y$ , so that

$$S^{-1}\#T = \begin{pmatrix} S_x^{-1}\#T_x & 0 \\ 0 & S_y^{-1}\#T_y \end{pmatrix}.$$

Proof of (ii)  $\Rightarrow$  (i): The result is straightforward.

Proof of (ii)  $\Rightarrow$  (iii): The matrix  $S^{-1}\#T$  satisfies  $T = (S^{-1}\#T)S(S^{-1}\#T)$ . Computing the off-diagonal blocks, there holds that  $T_{xy} = (S_x^{-1}\#T_x)S_{xy}(S_y^{-1}\#T_y)$ .

Proof of (iii)  $\Rightarrow$  (ii): There is a unique symmetric positive definite matrix  $C \in \mathcal{S}_{+,*}^{n_x+n_y}$  solution to the matrix equation  $CSC = T$ , which is  $S^{-1}\#T$ . By remarking that our guess  $\begin{pmatrix} S_x^{-1}\#T_x & 0 \\ 0 & S_y^{-1}\#T_y \end{pmatrix}$  is also a positive definite solution to the same equation, we then have that

$$S^{-1}\#T = \begin{pmatrix} S_x^{-1}\#T_x & 0 \\ 0 & S_y^{-1}\#T_y \end{pmatrix}.$$

Hence, at that point we have proved that (i)  $\Leftrightarrow$  (ii)  $\Leftrightarrow$  (iii).

Proof of (ii)  $\Rightarrow$  (iv): First, let us note that

$$\left(\sqrt{S}T\sqrt{S}\right)^{1/2} = \sqrt{S}(S^{-1}\#T)\sqrt{S},$$

and similarly

$$\left(\sqrt{S_x}T_x\sqrt{S_x}\right)^{1/2} = \sqrt{S_x}(S_x^{-1}\#T_x)\sqrt{S_x},$$

$$\left(\sqrt{S_y}T_y\sqrt{S_y}\right)^{1/2} = \sqrt{S_y}(S_y^{-1}\#T_y)\sqrt{S_y}.$$

Using the cyclicity of the trace, noting that  $(S^{-1}\#T)$  is block diagonal, and using the cyclicity of the trace again

$$\begin{aligned}\mathrm{Tr}\left(\left(\sqrt{ST}\sqrt{S}\right)^{1/2}\right) &= \mathrm{Tr}(S(S^{-1}\#T)) \\ &= \mathrm{Tr}(S_x(S_x^{-1}\#T_x)) + \mathrm{Tr}(S_y(S_y^{-1}\#T_y)) \\ &= \mathrm{Tr}\left(\sqrt{S_x}(S_x^{-1}\#T_x)\sqrt{S_x}\right) + \mathrm{Tr}\left(\sqrt{S_y}(S_y^{-1}\#T_y)\sqrt{S_y}\right).\end{aligned}$$

We then deduce that

$$\begin{aligned}\mathcal{W}_2^2(S, T) &= \mathrm{Tr}\left(T + S - 2\left(\left(\sqrt{ST}\sqrt{S}\right)^{1/2}\right)\right) \\ &= \mathrm{Tr}\left(T_x + S_x - 2\left(\left(\sqrt{S_x}T_x\sqrt{S_x}\right)^{1/2}\right)\right) \\ &\quad + \mathrm{Tr}\left(T_y + S_y - 2\left(\left(\sqrt{S_y}T_y\sqrt{S_y}\right)^{1/2}\right)\right) \\ &= \mathcal{W}_2^2(T_x, S_x) + \mathcal{W}_2^2(T_y, S_y),\end{aligned}$$

hence the desired result.

Proof of (iv)  $\Rightarrow$  (ii): Let us denote by  $T^*$  the matrix

$$T^* = \begin{pmatrix} T_x & (S_x^{-1}\#T_x)S_{xy}(S_y^{-1}\#T_y) \\ (S_y^{-1}\#T_y)S_{yx}(S_x^{-1}\#T_x) & T_y \end{pmatrix}$$

which satisfies (iii). Using that (i)  $\Leftrightarrow$  (ii)  $\Leftrightarrow$  (iii) and (ii)  $\Rightarrow$  (iv),  $T^*$  satisfies (iv). Therefore, there holds

$$\mathcal{W}_2^2(S, T^*) = \mathcal{W}_2^2(S_x, T_x) + \mathcal{W}_2^2(S_y, T_y).$$

From Lemma 3.3,  $T^*$  is the solution to the optimization problem (3.5). Hence

$$\mathcal{W}_2^2(S, T^*) \leq \mathcal{W}_2^2(S, T) = \mathcal{W}_2^2(S_x, T_x) + \mathcal{W}_2^2(S_y, T_y),$$

using the assumption, from which we deduce that

$$\mathcal{W}_2^2(S, T^*) = \mathcal{W}_2^2(S, T),$$

and thus  $T^* = T$ , noting that problem (3.5) admits a unique solution. Hence  $T$  satisfies (iii) and the proof is complete.  $\square$

**4. Marginals of Bures–Wasserstein barycenters.** In this section, we present a first result linking the preservation of the marginals of the Bures–Wasserstein barycenter between two covariance matrices  $S$  and  $T$  to the condition  $S \bowtie T$ .

**THEOREM 4.1.** *Let  $S, T \in \mathcal{S}_{+,*}^{n_x+n_y}$  with block decomposition (3.2), with  $S_x, T_x \in \mathcal{S}_{+,*}^{n_x}$ ,  $S_y, T_y \in \mathcal{S}_{+,*}^{n_y}$  and  $S_{xy}, T_{xy} \in \mathbb{R}^{n_x \times n_y}$ . Then, the five following statements are equivalent:*

- (i)  $\exists \lambda \in ]0, 1[$ ,  $\mathcal{M}_x\left(\mathrm{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T)\right) = \mathrm{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S_x, T_x)$ ;
- (ii)  $\forall \lambda \in [0, 1]$ ,  $\mathcal{M}_x\left(\mathrm{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T)\right) = \mathrm{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S_x, T_x)$ ;
- (iii)  $\exists \lambda \in ]0, 1[$ ,  $\mathcal{M}_y\left(\mathrm{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T)\right) = \mathrm{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S_y, T_y)$ ;

- (iv)  $\forall \lambda \in [0, 1]$ ,  $\mathcal{M}_y \left( \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T) \right) = \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S_y, T_y)$ ;  
(v)  $S \bowtie T$ .

*Proof.* Proof of (i)  $\Leftrightarrow$  (ii): First there holds

$$(4.1) \quad \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T) = ((1-\lambda)I_{n_x+n_y} + \lambda(S^{-1}\#T)) S ((1-\lambda)I_{n_x+n_y} + (S^{-1}\#T)).$$

Denoting  $S^{-1}\#T = \begin{pmatrix} A_x & A_{xy} \\ A_{xy}^\top & A_y \end{pmatrix}$ , the marginal in  $x$  of the barycenter is

$$\begin{aligned} \mathcal{M}_x \left( \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T) \right) &= (1-\lambda)^2 S_x + \lambda(1-\lambda)(S_x A_x + A_x S_x + S_{xy} A_{xy}^\top + A_{xy} S_{xy}^\top) \\ &\quad + \lambda^2 (A_x S_x A_x + A_x S_{xy} A_{xy}^\top + A_{xy} S_{xy}^\top A_x + A_{xy} S_y A_{xy}^\top). \end{aligned}$$

Now, the barycenter of the marginals in  $x$  is

$$\text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S_x, T_x) = ((1-\lambda)I_{n_x} + \lambda(S_x^{-1}\#T_x)) S_x ((1-\lambda)I_{n_x} + (S_x^{-1}\#T_x)).$$

Both functions  $[0, 1] \ni \lambda \mapsto \mathcal{M}_x \left( \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T) \right)$  and  $[0, 1] \ni \lambda \mapsto \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S_x, T_x)$  are second order polynomials. They coincide at  $\lambda = 0$  and  $\lambda = 1$ , thus proving the equivalence between the first and second statements.

Proof of (ii)  $\Rightarrow$  (v): Assuming both polynomials  $[0, 1] \ni \lambda \mapsto \mathcal{M}_x \left( \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T) \right)$  and  $[0, 1] \ni \lambda \mapsto \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S_x, T_x)$  are equal, then each coefficient of these functions in the basis  $\{(1-\lambda)^2, \lambda(1-\lambda), \lambda^2\}$  are also equal, hence in particular

$$\begin{cases} S_x A_x + A_x S_x + S_{xy} A_{xy}^\top + A_{xy} S_{xy}^\top = S_x (S_x^{-1}\#T_x) + (S_x^{-1}\#T_x) S_x, \\ A_x S_x A_x + A_x S_{xy} A_{xy}^\top + A_{xy} S_{xy}^\top A_x + A_{xy} S_y A_{xy}^\top = T_x. \end{cases}$$

Denoting by  $B = \sqrt{S_x} (A_x - (S_x^{-1}\#T_x)) + \sqrt{S_x}^{-1} S_{xy} A_{xy}^\top$ , we rewrite the system as

$$\begin{cases} \sqrt{S_x} B + B^\top \sqrt{S_x} = 0, \\ B^\top B + (S_x^{-1}\#T_x) \sqrt{S_x} B + B^\top \sqrt{S_x} (S_x^{-1}\#T_x) + A_{xy} (S_y - S_{xy}^\top S_x^{-1} S_{xy}) A_{xy}^\top = 0. \end{cases}$$

Since  $(S_x^{-1}\#T_x)$  is a symmetric matrix, there exists an orthogonal matrix  $P$  such that  $P(S_x^{-1}\#T_x)P^\top$  is a diagonal matrix. Therefore noting  $\tilde{B} = PB P^\top$ ,  $\tilde{S}_x = P S_x P^\top$ , so that  $\sqrt{\tilde{S}_x} = P \sqrt{S_x} P^\top$ , we obtain

$$\begin{cases} \sqrt{\tilde{S}_x} \tilde{B} + \tilde{B}^\top \sqrt{\tilde{S}_x} = 0, \\ \tilde{B}^\top \tilde{B} + P(S_x^{-1}\#T_x)P^\top \sqrt{\tilde{S}_x} \tilde{B} + \tilde{B}^\top \sqrt{\tilde{S}_x} P(S_x^{-1}\#T_x)P^\top \\ \quad + (P A_{xy}) (S_y - S_{xy}^\top S_x^{-1} S_{xy}) (P A_{xy})^\top = 0. \end{cases}$$

From the first equation, we deduce that the matrix  $\sqrt{\tilde{S}_x} \tilde{B}$  is antisymmetric. In particular, as  $P(S_x^{-1}\#T_x)P^\top$  is diagonal, we have that the diagonals of the two matrices  $P(S_x^{-1}\#T_x)P^\top \sqrt{\tilde{S}_x} \tilde{B}$  and  $\tilde{B}^\top \sqrt{\tilde{S}_x} P(S_x^{-1}\#T_x)P^\top$  are filled with zeros. As both remaining matrices  $B^\top B$  and  $(P A_{xy}) (S_y - S_{xy}^\top S_x^{-1} S_{xy}) (P A_{xy})^\top$  are positive semi-definite matrices, we deduce they are both equal to  $0_{n_x}$ . Hence, as  $S_y - S_{xy}^\top S_x^{-1} S_{xy}$

is a positive definite matrix as a Schur complement of  $S \in \mathcal{S}_{+,*}^{n_x+n_y}$ , we deduce that  $PA_{xy}$ , and thus  $A_{xy}$  are equal to zero as well.

This finally shows that the matrix  $S^{-1}\#T$  is block diagonal, hence, by Lemma 3.5, that  $S \bowtie T$ .

Proof of (v)  $\Rightarrow$  (ii): Using Lemma 3.5, there holds

$$S^{-1}\#T = \begin{pmatrix} S_x^{-1}\#T_x & 0 \\ 0 & S_y^{-1}\#\tau_y \end{pmatrix}.$$

Computing the marginals from (4.1), we easily obtain that for any  $\lambda \in [0, 1]$ ,

$$\mathcal{M}_x \left( \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S, T) \right) = \text{Bar}_{\mathcal{W}_2}^{(1-\lambda, \lambda)}(S_x, T_x).$$

The proofs can be extended similarly to the marginals with respect to the  $y$  variable. This concludes the proof.  $\square$

The extension of this result to a family of Gaussians with given covariance matrices seems more involved and so far, we have only managed to prove a partial result. First we present a necessary condition for the marginals to be preserved by the Wasserstein barycenter.

**THEOREM 4.2.** *Let  $M \in \mathbb{N}^*$  and let  $S^1, \dots, S^M \in \mathcal{S}_{+,*}^{n_x+n_y}$  so that for all  $1 \leq m \leq M$ , the matrix  $S^m$  admits the block decomposition*

$$(4.2) \quad S^m = \begin{pmatrix} S_x^m & S_{xy}^m \\ S_{xy}^m \top & S_y^m \end{pmatrix}.$$

*If for all  $\lambda \in \Lambda_M$ ,  $\mathcal{M}_x \left( \text{Bar}_{\mathcal{W}_2}^\lambda(S^1, \dots, S^M) \right) = \text{Bar}_{\mathcal{W}_2}^\lambda(S_x^1, \dots, S_x^M)$  or for all  $\lambda \in \Lambda_M$ ,  $\mathcal{M}_y \left( \text{Bar}_{\mathcal{W}_2}^\lambda((S^1, \dots, S^M)) \right) = \text{Bar}_{\mathcal{W}_2}^\lambda((S_y^1, \dots, S_y^M))$ , then*

$$\forall i, j \in \{1, \dots, M\}, \quad S^i \bowtie S^j.$$

*Proof.* Let assume that for all  $\lambda \in \Lambda_M$ ,

$$\mathcal{M}_x \left( \text{Bar}_{\mathcal{W}_2}^\lambda(S^1, \dots, S^M) \right) = \text{Bar}_{\mathcal{W}_2}^\lambda(S_x^1, \dots, S_x^M).$$

Let  $i, j \in \{1, \dots, M\}$ , taking weights  $\lambda$  with zeros except for the  $i^{\text{th}}$  and  $j^{\text{th}}$  variable, we apply Theorem 4.1 which implies that  $S^i \bowtie S^j$ . The proof is similar if the assumption is satisfied for the  $y$  marginals instead of the  $x$  marginals.  $\square$

We now turn to the inverse statement. For this, we need to make an additional assumption.

**THEOREM 4.3.** *Let  $M \in \mathbb{N}^*$  and let  $S^1, \dots, S^M \in \mathcal{S}_{+,*}^{n_x+n_y}$  so that for all  $1 \leq m \leq M$ ,  $S^m$  admits the block decomposition (4.2). Assume that*

$$(4.3) \quad \forall i, j \in \{1, \dots, M\}, \quad S^i \bowtie S^j,$$

*and that one of the following two assumptions below is satisfied*

- (i) *The matrices  $(S^i)^{-1}\#S^j$  for all  $i, j = 1, \dots, M$  commute.*
- (ii) *The off-diagonal blocks satisfy for all  $m = 1, \dots, M$ ,  $S_{xy}^m = 0$ .*

Then, for all  $\lambda \in \Lambda_M$ ,

$$\mathcal{M}_x \left( \text{Bar}_{\mathcal{W}_2}^\lambda(S^1, \dots, S^M) \right) = \text{Bar}_{\mathcal{W}_2}^\lambda(S_x^1, \dots, S_x^M),$$

and

$$\mathcal{M}_y \left( \text{Bar}_{\mathcal{W}_2}^\lambda(S^1, \dots, S^M) \right) = \text{Bar}_{\mathcal{W}_2}^\lambda(S_y^1, \dots, S_y^M).$$

REMARK 4.4. Assumption (i) is always satisfied if  $n_x = n_y = 1$ , meaning that the condition (4.3) is in this case a sufficient and necessary condition for the marginals to be preserved by the Wasserstein barycenter.

*Proof.* First, we assume that (i) holds. From [3, Theorem 4.2], the following sequence converges to the Wasserstein barycenter with weights  $\lambda \in \Lambda_M$ :  $B_0 \in \mathcal{S}_{+,*}^{n_x+n_y}$ , and for  $k \geq 1$ ,

$$B_{k+1} = B_k^{-1/2} \left( \sum_{m=1}^M \lambda_m \left( B_k^{1/2} S^m B_k^{1/2} \right)^{1/2} \right)^2 B_k^{-1/2},$$

which can be rewritten as

$$B_k^{-1} \# B_{k+1} = \sum_{m=1}^M \lambda_m (B_k^{-1} \# S^m).$$

Choosing  $B_0 = S^1$ , there holds for all  $m = 1, \dots, M$ ,  $B_0 \bowtie S^m$ . Then  $B_1$  is the only matrix satisfying

$$(S^1)^{-1} \# B_1 = \sum_{m=1}^M \lambda_m \left( (S^1)^{-1} \# S^m \right).$$

Using (i), the matrix  $(S^1)^{-1} \# B_1$  commutes with  $(S^1)^{-1} \# S^m$  for  $m = 1, \dots, M$  hence with  $((S^1)^{-1} \# S^m)^{-1}$ . Therefore for  $m = 1, \dots, M$  the matrices  $B_1^{-1} \# S^m$  are respectively equal to

$$B_1^{-1} \# S^m = \left( (S^1)^{-1} \# B_1 \right)^{-1} \left( (S^1)^{-1} \# S^m \right),$$

and block diagonal since each term on the righthandside is a block diagonal matrix. Moreover, combining the two previous equations leads to

$$\sum_{m=1}^M \lambda_m (B_1^{-1} \# S^m) = I_{n_x+n_y},$$

so that  $B_1$  is the Wasserstein barycenter  $\text{Bar}_{\mathcal{W}_2}^\lambda(S^1, \dots, S^M)$ . As each of the matrices  $B_1^{-1} \# S^m$  is block diagonal, and denoting by  $B_{1x}$  and  $B_{1y}$  the diagonal blocks of  $B_1$  we can now write the two equations

$$\begin{cases} \sum_{m=1}^M \lambda_m (B_{1x}^{-1} \# S_x^m) = I_{n_x}, \\ \sum_{m=1}^M \lambda_m (B_{1y}^{-1} \# S_y^m) = I_{n_y}, \end{cases}$$

which show that  $B_{1x} = \text{Bar}_{\mathcal{W}_2}^\lambda(S_x^1, \dots, S_x^M)$  (resp.  $B_{1y} = \text{Bar}_{\mathcal{W}_2}^\lambda(S_y^1, \dots, S_y^M)$ ). Noting that  $\mathcal{M}_x(B_1) = B_{1x}$  and  $\mathcal{M}_y(B_1) = B_{1y}$  concludes the proof.

Second we assume that (ii) holds. If the matrices  $S^m$  for  $m = 1, \dots, M$  are block diagonal, noting that  $B_0 = S^1$  is also block diagonal, are chosen to be block diagonal, then so are all iterates  $(B_k)_k$  of the iterative algorithm. Therefore at the limit, the property is kept and the barycenter  $\text{Bar}_{\mathcal{W}_2}^\lambda(S^1, \dots, S^M)$  is also block diagonal. We easily deduce that the marginals match.  $\square$

**5. Marginal-preserving modified Wasserstein barycenters between several Gaussian distributions.** The aim of this section is to propose, based on the theoretical results presented in Section 3 the definition of *modified* Wasserstein barycenters between Gaussian distributions that are marginal-preserving. We propose here two different ways to define such modified barycenters, which yield to two different objects. The first approach consists in finding the marginal-preserving Gaussian distribution which is the closest in the Wasserstein sense to the exact Wasserstein barycenter (which is not marginal-preserving in general) and is presented in Section 5.1. The second one consists in solving a barycentric optimization problem and is presented in Section 5.2.

In all this section, we denote by  $M \in \mathbb{N}^*$  a positive integer, by  $\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_M) \in \Lambda_M$  and by  $\mathbf{S} := (S^1, \dots, S^M) \in (\mathcal{S}_{+,*}^n)^M$ . We also assume that for all  $m = 1, \dots, M$ , the matrix  $S^m$  admits the block-decomposition (4.2), with  $S_x^m \in \mathcal{S}_{+,*}^{n_x}$ ,  $S_y^m \in \mathcal{S}_{+,*}^{n_y}$  and  $S_{xy}^m \in \mathbb{R}^{n_x \times n_y}$ . We also denote by  $\mathbf{S}_x := (S_x^1, \dots, S_x^M)$  and by  $\mathbf{S}_y := (S_y^1, \dots, S_y^M)$ . Let us also denote by  $\mathfrak{b}_x^\lambda := \text{Bar}_{\mathcal{W}_2}^\lambda(\mathbf{S}_x)$  and by  $\mathfrak{b}_y^\lambda := \text{Bar}_{\mathcal{W}_2}^\lambda(\mathbf{S}_y)$ .

**5.1. Modified barycenter based on minimal distance between Bures–Wasserstein barycenter and marginal-preserving barycenter.** Our first definition of modified marginal-preserving Wasserstein barycenter is based on the following theorem.

**THEOREM 5.1.** *There exists a unique solution  $Z^*$  to the minimization problem*

$$(5.1) \quad Z^* = \underset{Z \in \mathcal{C}_{\mathfrak{b}_x^\lambda, \mathfrak{b}_y^\lambda}}{\text{argmin}} \mathcal{W}_2^2 \left( \text{Bar}_{\mathcal{W}_2}^\lambda(\mathbf{S}), \mathfrak{B}^\lambda(Z) \right),$$

where

$$\mathcal{C}_{\mathfrak{b}_x^\lambda, \mathfrak{b}_y^\lambda} := \left\{ Z \in \mathbb{R}^{n_x \times n_y}, \left\| \sqrt{\mathfrak{b}_x^\lambda}^{-1} Z \sqrt{\mathfrak{b}_y^\lambda}^{-1} \right\|_2 < 1 \right\},$$

and for all  $Z \in \mathcal{C}_{\mathfrak{b}_x^\lambda, \mathfrak{b}_y^\lambda}$ ,

$$\mathfrak{B}^\lambda(Z) := \begin{pmatrix} \mathfrak{b}_x^\lambda & Z \\ Z^\top & \mathfrak{b}_y^\lambda \end{pmatrix}.$$

Assuming that  $\text{Bar}_{\mathcal{W}_2}^\lambda(\mathbf{S})$  admits the following block-decomposition

$$\text{Bar}_{\mathcal{W}_2}^\lambda(\mathbf{S}) = \begin{pmatrix} b_x^\lambda & b_{xy}^\lambda \\ b_{xy}^{\lambda \top} & b_y^\lambda \end{pmatrix},$$

with  $b_x^\lambda \in \mathcal{S}_{+,*}^{n_x}$ ,  $b_y^\lambda \in \mathcal{S}_{+,*}^{n_y}$  and  $b_{xy}^\lambda \in \mathbb{R}^{n_x \times n_y}$ , it holds that

$$Z^* = \left( b_x^{\lambda^{-1}} \# b_x^\lambda \right) b_{xy}^\lambda \left( b_y^{\lambda^{-1}} \# b_y^\lambda \right).$$

*Proof.* It is a direct consequence of Lemma 3.3 applied with  $S = \text{Bar}_{\mathcal{W}_2}^\lambda(\mathbf{S})$ ,  $T_x = b_x^\lambda$  and  $T_y = b_y^\lambda$ .  $\square$

In the following, we will denote this modified marginal-preserving barycenter by  $\mathfrak{B}^\lambda(\mathbf{S}) := \mathfrak{B}^\lambda(Z^*)$  where  $Z^*$  is the unique solution of the minimization problem (5.1). The main advantage of this first definition is that the solution of the problem is explicit, provided one has computed the exact Wasserstein barycenter  $\text{Bar}_{\mathcal{W}_2}^\lambda(\mathbf{S})$ .

**5.2. Modified barycenter based on Bures–Wasserstein barycenter optimization problem with marginal constraints.** We now present another natural definition of modified marginal-preserving Wasserstein barycenter, considering the following optimization problem, which is an alternative to (5.1): we know that there exists a unique solution  $Z^* \in \overline{\mathcal{C}_{\mathbf{b}_x^\lambda, \mathbf{b}_y^\lambda}}$  to

$$(5.2) \quad Z^* = \underset{Z \in \overline{\mathcal{C}_{\mathbf{b}_x^\lambda, \mathbf{b}_y^\lambda}}}{\text{argmin}} \sum_{m=1}^M \lambda_m \mathcal{W}_2^2(S_m, \mathcal{B}^\lambda(Z)),$$

where

$$\overline{\mathcal{C}_{\mathbf{b}_x^\lambda, \mathbf{b}_y^\lambda}} = \left\{ Z \in \mathbb{R}^{n_x \times n_y}, \left\| \sqrt{\mathbf{b}_x^\lambda}^{-1} Z \sqrt{\mathbf{b}_y^\lambda}^{-1} \right\|_2 \leq 1 \right\},$$

and for all  $Z \in \overline{\mathcal{C}_{\mathbf{b}_x^\lambda, \mathbf{b}_y^\lambda}}$ ,

$$\mathcal{B}^\lambda(Z) = \begin{pmatrix} \mathbf{b}_x^\lambda & Z \\ Z^\top & \mathbf{b}_y^\lambda \end{pmatrix}.$$

Indeed, (5.2) reads as a minimization problem of a strictly convex function over a convex compact domain. If such a minimizer  $Z^*$  is an element of  $\mathcal{C}_{\mathbf{b}_x^\lambda, \mathbf{b}_y^\lambda}$ , which we conjecture to be true, but were unable to prove, then it is necessarily solution to the equation

$$\left[ \sum_{m=1}^M \lambda_m \mathcal{B}^\lambda(Z^*)^{-1} \# S^m \right]_{xy} = 0.$$

In the following, we will denote by  $\mathcal{B}^\lambda(\mathbf{S}) := \mathcal{B}^\lambda(Z^*)$  where  $Z^*$  is the unique solution of the convex minimization problem (5.2), that is found using standard optimization packages.

Let us emphasize here that, in general,  $\mathcal{B}^\lambda(\mathbf{S})$  and  $\mathfrak{B}^\lambda(\mathbf{S})$  are different as will be shown in the numerical tests presented in Section 7.

**5.3. Extension to Gaussian distributions.** In the previous sections, we presented two ways of defining modified marginal-preserving Bures–Wasserstein barycenters associated to a family of covariance matrices. These two definitions can be easily and naturally extended to define marginal-preserving Wasserstein barycenters of a family of Gaussian probability distributions as follows. Let  $M \in \mathbb{N}^*$ , let  $\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_M) \in \Lambda_M$ ,  $\mathbf{S} := (S^1, \dots, S^M) \in (\mathcal{S}_{+,*}^n)^M$  and  $\boldsymbol{\mu} := (\mu^1, \dots, \mu^M) \in (\mathbb{R}^n)^M$ . For all  $1 \leq m \leq M$ , we denote by  $\rho^m = \mathcal{N}(\mu^m, S^m)$  and by  $\boldsymbol{\rho} := (\rho^1, \dots, \rho^M)$ .

Then, using a slight abuse of notation, we define two marginal-preserving barycenters of the family  $\boldsymbol{\rho}$  with weights  $\boldsymbol{\lambda}$  as follows:

$$\mathcal{B}^\lambda(\boldsymbol{\rho}) := \mathcal{N}(\boldsymbol{\mu}, \mathcal{B}^\lambda(\mathbf{S})) \quad \text{and} \quad \mathfrak{B}^\lambda(\boldsymbol{\rho}) := \mathcal{N}(\boldsymbol{\mu}, \mathfrak{B}^\lambda(\mathbf{S})),$$

where

$$\boldsymbol{\mu} := \sum_{m=1}^M \lambda_m \mu^m.$$

This amounts to remark that the centers of the Gaussian distributions do not play a role in the marginal-preservation of the barycenters.

## 6. Modified marginal-preserving barycenters for Gaussian mixtures.

The aim of this section is to propose a method to define a marginal-preserving Wasserstein-like barycenter for a family of mixtures of Gaussian distributions. In Section 6.1, we recall the definition and main properties of the Gaussian mixture Wasserstein metric proposed in [11, 8]. In Section 6.2, we introduce the proposed marginal-preserving mixture barycenter based on a marginal-constrained barycenter formulation for Gaussian mixtures.

**6.1. Some fundamental results on Gaussian mixtures.** In this section, we recall some classical results from [11] on mixture Wasserstein distance and barycenters for Gaussian mixtures. We say that a probability measure  $\rho \in \mathcal{P}_2(\mathbb{R}^n)$  is a finite Gaussian mixture if and only if there exists  $K \in \mathbb{N}^*$ ,  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K) \in \Lambda_K$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K) \in (\mathbb{R}^n)^K$  and  $\boldsymbol{S} = (S_1, \dots, S_K) \in (\mathcal{S}_{+,*}^n)^K$  such that

$$\rho = \sum_{k=1}^K \pi_k \mathcal{N}(\mu_k, S_k).$$

In [11, 8], the authors introduced a notion of Wasserstein-like barycenter, called mixture Wasserstein barycenter, of a family of Gaussian mixtures which is defined as follows. Let  $M \geq 2$  and for all  $1 \leq m \leq M$ , let  $\rho^m$  be some finite Gaussian mixture such that

$$\rho^m = \sum_{k^m=1}^{K^m} \pi_{k^m}^m \mathcal{N}(\mu_{k^m}^m, S_{k^m}^m),$$

with  $K^m \in \mathbb{N}^*$ ,  $\boldsymbol{\pi}^m = (\pi_1^m, \dots, \pi_{K^m}^m) \in \Lambda_{K^m}$ ,  $\boldsymbol{\mu}^m = (\mu_1^m, \dots, \mu_{K^m}^m) \in (\mathbb{R}^n)^{K^m}$  and  $\boldsymbol{S}^m = (S_1^m, \dots, S_{K^m}^m) \in (\mathcal{S}_{+,*}^n)^{K^m}$ . For all  $1 \leq m \leq M$ , we make the assumption that all pairs  $(\mu_{k^m}^m, S_{k^m}^m)$  are distinct from one another and denote by  $g_{k^m}^m := \mathcal{N}(\mu_{k^m}^m, S_{k^m}^m)$  for all  $1 \leq k^m \leq K^m$ .

We first recall the definition of the mixture Wasserstein distance between two Gaussian mixtures  $\rho^1$  and  $\rho^2$ . The latter is denoted by  $MW_2(\rho^1, \rho^2)$  and is computed as follows:

$$(6.1) \quad MW_2^2(\rho^1, \rho^2) := \min_{\boldsymbol{v} \in \Gamma(\boldsymbol{\pi}^1, \boldsymbol{\pi}^2)} \sum_{k^1=1}^{K^1} \sum_{k^2=1}^{K^2} v_{k^1 k^2} W_2^2(g_{k^1}^1, g_{k^2}^2),$$

where

$$\Gamma(\boldsymbol{\pi}^1, \boldsymbol{\pi}^2) := \left\{ \boldsymbol{v} := (v_{k^1 k^2})_{1 \leq k^1 \leq K^1, 1 \leq k^2 \leq K^2} \in \mathbb{R}_+^{K^1 \times K^2} : \begin{array}{l} \forall 1 \leq k^1 \leq K^1, \sum_{k^2=1}^{K^2} v_{k^1 k^2} = \pi_{k^1}^1, \\ \forall 1 \leq k^2 \leq K^2, \sum_{k^1=1}^{K^1} v_{k^1 k^2} = \pi_{k^2}^2 \end{array} \right\}.$$

Now, the mixture Wasserstein barycenter of the family  $\boldsymbol{\rho} = (\rho^1, \dots, \rho^M)$  with weights  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_M) \in \Lambda_M$  is given as follows. Let

$$\boldsymbol{K} := \{\boldsymbol{k} := (k^1, \dots, k^M) \in \{1, \dots, K^1\} \times \dots \times \{1, \dots, K^M\}\}.$$

Consider the discrete optimization problem: find  $\boldsymbol{w}^*(\boldsymbol{\lambda}) := (w_{\boldsymbol{k}}^*(\boldsymbol{\lambda}))_{\boldsymbol{k} \in \boldsymbol{K}} \in \mathbb{R}_+^{\boldsymbol{K}}$  solution to

$$(6.2) \quad \min_{\mathbf{w} := (w_{\mathbf{k}})_{\mathbf{k} \in \mathbf{K}} \in \Gamma(\boldsymbol{\pi}^1, \dots, \boldsymbol{\pi}^M)} \sum_{\mathbf{k} := (k^1, \dots, k^M) \in \mathbf{K}} w_{\mathbf{k}} \text{mm}W_2^2(g_{k^1}^1, \dots, g_{k^M}^M; \boldsymbol{\lambda}),$$

where  $\text{mm}W_2$  is defined in (2.2), and

$$\Gamma(\boldsymbol{\pi}^1, \dots, \boldsymbol{\pi}^M) := \left\{ \mathbf{w} = (w_{\mathbf{k}})_{\mathbf{k} \in \mathbf{K}} \in (\mathbb{R}_+)^{\mathbf{K}} : \forall 1 \leq m \leq M, \forall 1 \leq \ell^m \leq K^m, \sum_{\mathbf{k} \in \mathbf{K} : k^m = \ell^m} w_{\mathbf{k}} = \pi_{\ell^m}^m \right\}.$$

The mixture Wasserstein barycenter of the family  $\boldsymbol{\rho}$  with weights  $\boldsymbol{\lambda}$  is then defined as

$$(6.3) \quad \text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}) := \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}}^*(\boldsymbol{\lambda}) \text{Bar}_{W_2}^{\boldsymbol{\lambda}}(g_{k^1}^1, \dots, g_{k^M}^M).$$

It is then proved in [11] that  $\text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho})$  is indeed the barycenter of the family of finite Gaussian mixtures  $\boldsymbol{\rho}$  with weights  $\boldsymbol{\lambda}$  with respect to the mixture Wasserstein distance. In particular,  $\text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho})$  is itself a Gaussian mixture.

**6.2. Modified marginal-preserving Wasserstein barycenters for Gaussian mixtures.** The aim of this section is to detail the proposed definition of modified marginal-preserving Gaussian mixture Wasserstein barycenters, using the notation introduced in the previous section. First, since the exact mixture Wasserstein barycenter  $\text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho})$  is a Gaussian mixture, it can be written as follows

$$(6.4) \quad \text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}) = \sum_{i=1}^I \pi_i \mathcal{N}(\nu^i, T^i),$$

for some  $I \in \mathbb{N}^*$ ,  $\boldsymbol{\pi} = (\pi_i)_{1 \leq i \leq I} \in \Lambda_I$ , and for all  $1 \leq i \leq I$ ,  $\nu^i \in \mathbb{R}^{n_x + n_y}$  and

$$T^i := \begin{pmatrix} T_x^i & T_{xy}^i \\ (T_{xy}^i)^\top & T_y^i \end{pmatrix} \in S_{+,*}^{n_x + n_y},$$

with  $T_x^i \in \mathcal{S}_{+,*}^{n_x}$ ,  $T_y^i \in \mathcal{S}_{+,*}^{n_y}$  and  $T_{xy}^i \in \mathbb{R}^{n_x \times n_y}$ . For  $z = x, y$ , we denote by  $\boldsymbol{\rho}_z := (\rho_z^1, \dots, \rho_z^M)$  where for all  $1 \leq m \leq M$ ,  $\rho_z^m = \text{marg}_z(\rho^m)$ . Similarly, we can write  $\text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}_x)$  and  $\text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}_y)$  under the form

$$\text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}_x) = \sum_{k=1}^K \alpha_x^k \mathcal{N}(\mu_x^k, S_x^k), \quad \text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}_y) = \sum_{\ell=1}^L \beta_y^\ell \mathcal{N}(\mu_y^\ell, S_y^\ell),$$

for some  $K, L \in \mathbb{N}^*$ ,  $\boldsymbol{\alpha}_x := (\alpha_x^k)_{1 \leq k \leq K} \in \Lambda_K$ ,  $\boldsymbol{\beta}_y := (\beta_y^\ell)_{1 \leq \ell \leq L} \in \Lambda_L$ . In addition, for all  $1 \leq k \leq K$ ,  $\mu_x^k \in \mathbb{R}^{n_x}$  and  $S_x^k \in \mathcal{S}_{+,*}^{n_x}$ , and for all  $1 \leq \ell \leq L$ ,  $\mu_y^\ell \in \mathbb{R}^{n_y}$  and  $S_y^\ell \in \mathcal{S}_{+,*}^{n_y}$ .

Further it is easy to see that the set of Gaussian mixtures  $\rho$  satisfying the following marginal constraints

$$\text{marg}_x(\rho) = \text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}_x), \quad \text{marg}_y(\rho) = \text{Bar}_{MW_2}^{\boldsymbol{\lambda}}(\boldsymbol{\rho}_y),$$

is equal to the set  $\mathcal{B}$  defined as

$$\mathcal{B} := \left\{ \rho := \sum_{k=1}^K \sum_{\ell=1}^L \sum_{j=1}^{J(k,\ell)} \gamma_{k\ell j} \mathcal{N} \left( \begin{pmatrix} \mu_x^k \\ \mu_y^\ell \end{pmatrix}, \begin{pmatrix} S_x^k & Z_{k\ell j} \\ (Z_{k\ell j})^\top & S_y^\ell \end{pmatrix} \right) : \right. \\ \forall k = 1, \dots, K, \forall \ell = 1, \dots, L, J(k, \ell) \in \mathbb{N}^*, \\ \forall j = 1, \dots, J(k, \ell), Z_{k\ell j} \in \mathcal{C}_{S_x^k, S_y^\ell}, \quad \gamma_{k\ell j} \geq 0, \\ \left. \sum_{\ell=1}^L \sum_{j=1}^{J(k,\ell)} \gamma_{k\ell j} = \alpha_x^k, \quad \sum_{k=1}^K \sum_{j=1}^{J(k,\ell)} \gamma_{k\ell j} = \beta_y^\ell \right\}.$$

where for any  $T_x \in \mathcal{S}_{+,*}^{n_x}$  and  $T_y \in \mathcal{S}_{+,*}^{n_y}$ , the set  $\mathcal{C}_{T_x, T_y}$  is defined in (3.3). Now, for all  $J \in \mathbb{N}^*$ , we introduce the set  $\mathcal{B}_J \subset \mathcal{B}$  as

$$\mathcal{B}_J := \left\{ \rho := \sum_{k=1}^K \sum_{\ell=1}^L \sum_{j=1}^J \gamma_{k\ell j} \mathcal{N} \left( \begin{pmatrix} \mu_x^k \\ \mu_y^\ell \end{pmatrix}, \begin{pmatrix} S_x^k & Z_{k\ell j} \\ (Z_{k\ell j})^\top & S_y^\ell \end{pmatrix} \right) : \right. \\ \forall k = 1, \dots, K, \forall \ell = 1, \dots, L, \forall j = 1, \dots, J, \\ Z_{k\ell j} \in \mathcal{C}_{S_x^k, S_y^\ell}, \quad \gamma_{k\ell j} \geq 0, \\ \left. \sum_{\ell=1}^L \sum_{j=1}^J \gamma_{k\ell j} = \alpha_x^k, \quad \sum_{k=1}^K \sum_{j=1}^J \gamma_{k\ell j} = \beta_y^\ell \right\}.$$

Our approach is motivated as follows: it would be natural to look for a modified marginal-preserving Wasserstein barycenter by considering the following constrained optimization problem, which amounts to minimizing the mixture Wasserstein distance between the exact barycenter and the marginal-preserving barycenter

$$\inf_{\rho \in \mathcal{B}} MW_2^2(\text{Bar}_{MW_2}^\lambda(\rho), \rho).$$

Unfortunately this yields to a very complex optimization problem, even restricting it to the smaller set  $\mathcal{B}_J$  for some  $J \in \mathbb{N}^*$ . Indeed, in both cases, solving this problem requires to optimize upon matrices  $Z_{k\ell j} \in \mathcal{C}_{S_x^k, S_y^\ell}$  and weights  $\gamma_{k\ell j} \geq 0$  for  $1 \leq k \leq K$ ,  $1 \leq \ell \leq L$  and  $1 \leq j \leq J$ , which is a very costly and high-dimensional nonlinear optimization problem.

We therefore simplify the optimization problem one step further, building on the results obtained in the previous section. We choose  $J = I$  and, for all  $1 \leq k \leq K$ ,  $1 \leq \ell \leq L$  and  $1 \leq j \leq I$ , inspired by the result of Lemma 3.3, we fix the values of the matrices  $Z_{k\ell j}$  to be equal to

$$Z_{k\ell j} := ((T_x^j)^{-1} \# S_x^k) T_{xy}^j ((T_y^j)^{-1} \# S_y^\ell).$$

This amounts to choosing the modified optimization set

$$(6.5) \quad \tilde{\mathcal{B}}_I := \left\{ \rho := \sum_{k=1}^K \sum_{\ell=1}^L \sum_{j=1}^I \gamma_{k\ell j} \mathcal{N} \left( \begin{pmatrix} \mu_x^k \\ \mu_y^\ell \end{pmatrix}, \begin{pmatrix} S_x^k & Z_{k\ell j} \\ (Z_{k\ell j})^\top & S_y^\ell \end{pmatrix} \right), \right. \\ \forall k = 1, \dots, K, \forall \ell = 1, \dots, L, \forall j = 1, \dots, I, \\ Z_{k\ell j} = ((T_x^j)^{-1} \# S_x^k) T_{xy}^j ((T_y^j)^{-1} \# S_y^\ell), \quad \gamma_{k\ell j} \geq 0, \\ \left. \sum_{\ell=1}^L \sum_{j=1}^I \gamma_{k\ell j} = \alpha_x^k, \quad \sum_{k=1}^K \sum_{j=1}^I \gamma_{k\ell j} = \beta_y^\ell \right\}.$$

We thus consider the following now linear optimization problem

$$\min_{\rho \in \tilde{\mathcal{B}}_I} MW_2^2(\text{Bar}_{MW_2}^\lambda(\rho), \rho).$$

Using (6.4) and the definition of the mixture distance (6.1), this problem can be written as

$$\min_{\substack{\gamma \in \mathbb{R}_+^{K \times L \times I} \\ \sum_{\ell=1}^L \sum_{j=1}^J \gamma_{k\ell j} = \alpha_x^k, \\ \sum_{k=1}^K \sum_{j=1}^J \gamma_{k\ell j} = \beta_y^\ell}} \min_{\substack{\mathbf{w} \in \mathbb{R}_+^{K \times L \times I \times I} \\ \sum_{k\ell j} w_{k\ell j} = \pi_i \\ \sum_i w_{k\ell j} = \gamma_{k\ell j}}} \sum_{k\ell j} w_{k\ell j} W_2^2 \left( \mathcal{N} \left( \begin{pmatrix} \mu_x^k \\ \mu_y^\ell \end{pmatrix}, \begin{pmatrix} S_x^k & Z_{k\ell j} \\ Z_{k\ell j}^\top & S_y^\ell \end{pmatrix} \right), \mathcal{N}(\nu^i, T^i) \right),$$

that we recast as

$$(6.6) \quad \min_{\substack{\mathbf{w} \in \mathbb{R}_+^{K \times L \times I \times I} \\ \sum_{k\ell j} w_{k\ell j} = \pi_i \\ \sum_{i\ell j} w_{k\ell j} = \alpha_x^k \\ \sum_{ikj} w_{k\ell j} = \beta_y^\ell}} \sum_{k\ell j} w_{k\ell j} W_2^2 \left( \mathcal{N} \left( \begin{pmatrix} \mu_x^k \\ \mu_y^\ell \end{pmatrix}, \begin{pmatrix} S_x^k & Z_{k\ell j} \\ (Z_{k\ell j})^\top & S_y^\ell \end{pmatrix} \right), \mathcal{N}(\nu^i, T^i) \right),$$

and set

$$\gamma_{k\ell j} := \sum_{i=1}^I w_{k\ell j}.$$

The proposed marginal-preserving modified Wasserstein barycenter  $\mathfrak{B}^\lambda(\rho)$  is finally defined for Gaussian mixtures as

$$(6.7) \quad \mathfrak{B}^\lambda(\rho) := \sum_{k=1}^K \sum_{\ell=1}^L \sum_{j=1}^I \gamma_{k\ell j} \mathcal{N} \left( \begin{pmatrix} \mu_x^k \\ \mu_y^\ell \end{pmatrix}, \begin{pmatrix} S_x^k & Z_{k\ell j} \\ (Z_{k\ell j})^\top & S_y^\ell \end{pmatrix} \right).$$

This barycenter is therefore computed as a post-processing of the true mixture Wasserstein barycenter of a family of Gaussian mixtures. It requires the resolution of an additional linear programming problem of size  $KLI^2$ . This procedure gives satisfying numerical results, as we will present in Section 7.

**7. Numerical results.** In this section, we provide a few numerical results illustrating the findings of the previous sections. The code generating the figures of this articles can be found on the Github repository <https://github.com/mdalery/PreservingMarginals>.

**7.1. Marginals of Wasserstein barycenters.** We start by providing in Figure 1 an example of discrepancy between the marginals of the Wasserstein barycenters and the Wasserstein barycenters of the marginals. More precisely, we plot on the top line the Wasserstein barycenters between two two-dimensional Gaussian distributions  $g \sim \mathcal{N}(0, S)$  and  $h \sim \mathcal{N}(0, T)$  with

$$S = \begin{pmatrix} 0.3 & 0.15 \\ 0.15 & 0.15 \end{pmatrix}, \quad T = \begin{pmatrix} 0.1 & -0.1 \\ -0.1 & 0.2 \end{pmatrix}.$$

On the second line (respectively on the bottom line), we plot the marginals with respect to the  $x$  variable (resp.  $y$ ) of the barycenter together with the barycenters of the marginals which correspond to  $g_x \sim \mathcal{N}(0, S_x)$  with  $S_x = (0.3)$  and  $h_x \sim \mathcal{N}(0, T_x)$

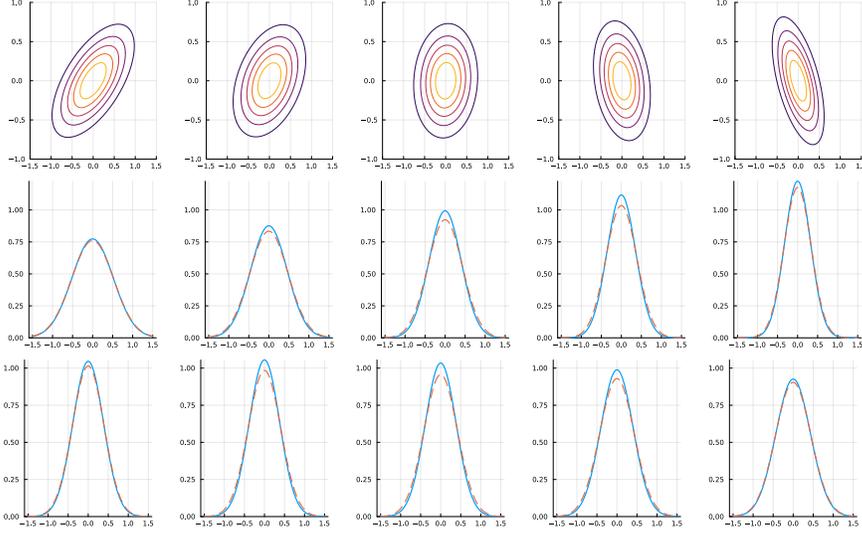


FIG. 1. *Up:* contour plot of the Wasserstein barycenter  $\text{Bar}_{W_2}^{(1-\lambda, \lambda)}(g, h)$  for  $\lambda \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$  from left to right. *Middle:* plot of marginals with respect to  $x$  variable,  $\text{marg}_x(\text{Bar}_{W_2}^{(1-\lambda, \lambda)}(g, h))$  in solid blue and  $\text{Bar}_{W_2}^{(1-\lambda, \lambda)}(g_x, h_x)$  in dash orange. *Bottom:* plot of marginals with respect to  $y$  variable,  $\text{marg}_y(\text{Bar}_{W_2}^{(1-\lambda, \lambda)}(g, h))$  in solid blue and  $\text{Bar}_{W_2}^{(1-\lambda, \lambda)}(g_y, h_y)$  in dash orange.

with  $T_x = (0.1)$  (resp.  $g_y \sim \mathcal{N}(0, S_y)$  with  $S_y = (0.15)$  and  $h_y \sim \mathcal{N}(0, T_y)$  with  $T_y = (0.2)$ ). We indeed observe that the marginals do not match in general.

We emphasize the difference by plotting in Figure 2 the values of the covariance matrices corresponding to the marginals of the Wasserstein barycenter between  $S$  and  $T$  together with the Wasserstein between the marginals of  $S$  and  $T$ .

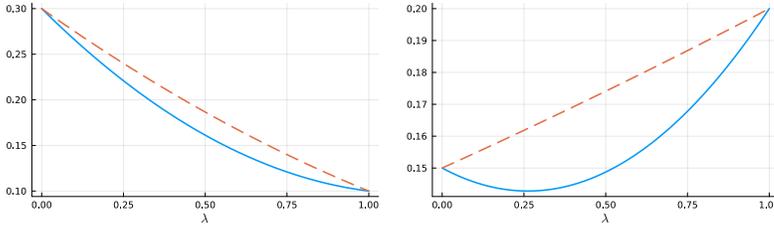


FIG. 2. *Left:* plot of  $\lambda \mapsto \mathcal{M}_x(\text{Bar}_{W_2}^{(1-\lambda, \lambda)}(S, T))$  in solid blue and  $\lambda \mapsto \text{Bar}_{W_2}^{(1-\lambda, \lambda)}(S_x, T_x)$  in dash orange. *Right:* plot of  $\lambda \mapsto \mathcal{M}_y(\text{Bar}_{W_2}^{(1-\lambda, \lambda)}(S, T))$  in solid blue and  $\lambda \mapsto \text{Bar}_{W_2}^{(1-\lambda, \lambda)}(S_y, T_y)$  in dash orange.

**7.2. Comparison of modified Wasserstein barycenters.** We now compare the true Wasserstein barycenter with the two different modified Wasserstein barycenters introduced respectively in Sections 5.1 and 5.2. More precisely, we compare these three barycenters in two examples, first in Figure 3 between two Gaussian distribu-

tions  $g_1 \sim \mathcal{N}(0, S_1)$  and  $g_2 \sim \mathcal{N}(0, S_2)$  with

$$S_1 = \begin{pmatrix} 1 & 0.999 \\ 0.999 & 1 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 1 & -0.999\sqrt{2} \\ -0.999\sqrt{2} & 2 \end{pmatrix},$$

and second on Figure 4 between  $g_3 \sim \mathcal{N}(0, S_3)$  and  $g_4 \sim \mathcal{N}(0, S_4)$  with

$$S_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad S_4 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

On Figure 3 (resp. Figure 4), we present a contour plot of the true Wasserstein barycenter  $\text{Bar}_{W_2}^\lambda(g^1, g^2)$  (resp.  $\text{Bar}_{W_2}^\lambda(g^3, g^4)$ ) on the top line, a contour plot of the modified Wasserstein barycenter  $\mathfrak{B}^\lambda(g^1, g^2)$  (resp.  $\mathfrak{B}^\lambda(g^3, g^4)$ ) defined in Section 5.1 in the middle line, a contour plot of the modified Wasserstein barycenter  $\mathcal{B}^\lambda(g^1, g^2)$  (resp.  $\mathcal{B}^\lambda(g^3, g^4)$ ) defined in Section 5.2 on the bottom line.

We observe on Figure 3 that for  $g_1$  and  $g_2$ , the modified marginal-preserving barycenters are very different from the true Wasserstein barycenters. Moreover, we also observe a clear difference between the two modified Wasserstein barycenters, illustrating that the two alternative definitions in Sections 5.1 and 5.2 lead to different results. We emphasize this difference on Figure 5, where we plot the off-diagonal coefficient of the covariance matrices for the two marginal-preserving barycenters between the two Gaussians of Figure 3. We also observe on Figure 4 that for some Gaussian distributions, the modified Wasserstein barycenters are very close to the true Wasserstein barycenters.

Numerically speaking, computing the solution to problem (5.1) is easy as the solution to this problem is explicit and given in Theorem 5.1 while the solution to problem (5.2) is not explicit and requires in practice to solve an optimization problem. In our numerical simulations, we computed the solution to (5.2) using nonlinear convex optimization algorithms from the Julia packages JuMP and Ipopt.

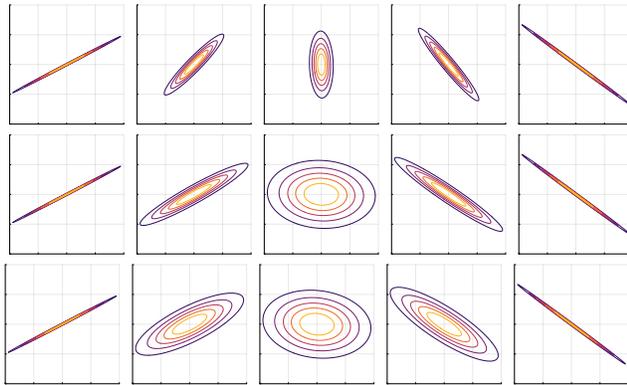


FIG. 3. Top: contour plots of true Wasserstein barycenters  $\text{Bar}_{W_2}^\lambda(g_1, g_2)$ , middle: contour plots of modified Wasserstein barycenters  $\mathfrak{B}^\lambda(g_1, g_2)$ , bottom: contour plots of modified Wasserstein barycenters  $\mathcal{B}^\lambda(g_1, g_2)$ . Left to right: weights  $(1 - \lambda, \lambda)$  for  $\lambda = 0, 0.25, 0.5, 0.75, 1$ .

**7.3. Modified mixture barycenters.** We finally provide an example of a marginal-preserving barycenter between four shapes fitted as mixtures of Gaussian distributions. The fit with respect to the shapes is performed using Python module

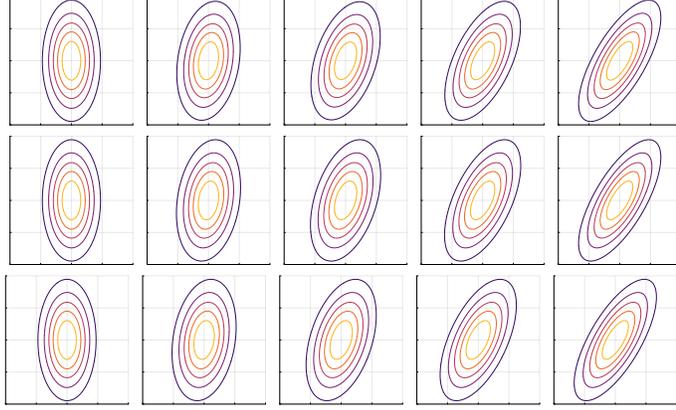


FIG. 4. *Top: contour plots of true Wasserstein barycenters  $\text{Bar}_{W_2}^\lambda(g_3, g_4)$ , middle: contour plots of modified Wasserstein barycenters  $\mathfrak{B}^\lambda(g_3, g_4)$ , bottom: contour plots of modified Wasserstein barycenters  $\mathcal{B}^\lambda(g_3, g_4)$ . Left to right: weights  $(1 - \lambda, \lambda)$  for  $\lambda = 0, 0.25, 0.5, 0.75, 1$ .*

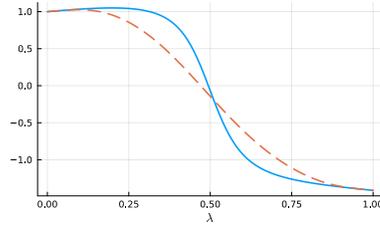


FIG. 5. *Comparison of the off-diagonal terms of the two marginal-preserving barycenters in an extreme case. Plot of  $\lambda \mapsto (\mathfrak{B}^\lambda(S_1, S_2))_{xy}$  in solid blue and  $\lambda \mapsto (\mathcal{B}^\lambda(S_1, S_2))_{xy}$  in dash orange.*

`scikit-learn`. Twenty Gaussians are used for fitting each image. We use bilinear weights with respect to the four corners of the square to compute the barycenters. We plot on Figure 6 the modified marginal-preserving barycenters defined in (6.7) while we plot on Figure 7 the mixture Wasserstein barycenters defined in (6.3). We observe a visible difference between the two barycenters but which does not seem to affect the quality of the barycentric images. Therefore, the definition of the marginal-preserving barycenters for Gaussian mixtures seems very satisfying.

**8. Conclusion.** The aim of this paper is to propose marginal-preserving Wasserstein-like barycenters. After presenting some theoretical results on the necessary conditions to preserve marginal consistency in Wasserstein barycenters for Gaussian distributions, we introduce modified Wasserstein barycenter that indeed preserve the marginals for Gaussian distributions as well as for mixture of Gaussian distributions. This opens the way to very interesting applications, e.g. in quantum chemistry for the fitting of pair densities preserving the electronic density, which is the marginal of the pair density. Other applications include population dynamics, crowd modeling, as well as fitting of distributions arising from kinetic partial differential equations. A natural follow-up question is whether these modified barycenters arise from a modified Wasserstein distance, which would lead to a natural extension of this work to generic probability measures. This is subject to future work.

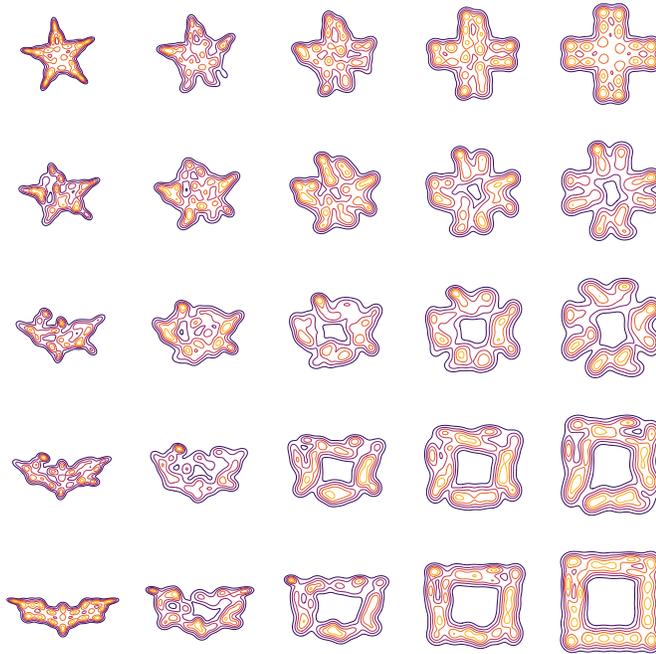


FIG. 6. Marginal-preserving modified Wasserstein barycenters between 4 mixtures of 20 Gaussian components.

#### REFERENCES

- [1] I. ABRAHAM, R. ABRAHAM, M. BERGOUNIOUX, AND G. CARLIER, *Tomographic reconstruction from a few views: a multi-marginal optimal transport approach*, Applied mathematics & optimization, 75 (2017), pp. 55–73.
- [2] M. AGUEH AND G. CARLIER, *Barycenters in the Wasserstein space*, SIAM Journal on Mathematical Analysis, 43 (2011), pp. 904–924.
- [3] 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), pp. 744–762.
- [4] R. BHATIA, *Positive definite matrices*, in Positive Definite Matrices, Princeton university press, 2009.
- [5] ———, *Matrix Analysis*, Springer Science & Business Media, Dec. 2013.
- [6] ———, *The riemannian mean of positive matrices*, in Matrix Information Geometry, F. Nielsen and R. Bhatia, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 35–51.
- [7] R. BHATIA, T. JAIN, AND Y. LIM, *On the Bures–Wasserstein distance between positive definite matrices*, Expo. Math., 37 (2019), pp. 165–191.
- [8] Y. CHEN, T. T. GEORGIU, AND A. TANNENBAUM, *Optimal transport for gaussian mixture models*, IEEE Access, 7 (2018), pp. 6269–6278.
- [9] K. CHENG, S. AERON, M. C. HUGHES, AND E. L. MILLER, *Dynamical Wasserstein barycenters for time-series modeling*, Advances in Neural Information Processing Systems, 34 (2021), pp. 27991–28003.
- [10] M. DALERY, G. DUSSON, V. EHRLACHER, AND A. LOZINSKI, *Nonlinear reduced basis using mixture Wasserstein barycenters: application to an eigenvalue problem inspired from quantum chemistry*, arXiv preprint arXiv:2307.15423, (2023).
- [11] J. DELON AND A. DESOLNEUX, *A Wasserstein-type distance in the space of gaussian mixture models*, SIAM Journal on Imaging Sciences, 13 (2020), pp. 936–970.
- [12] V. EHRLACHER, D. LOMBARDI, O. MULA, AND F.-X. VIALARD, *Nonlinear model reduction on metric spaces. application to one-dimensional conservative pdes in Wasserstein spaces*, ESAIM: Mathematical Modelling and Numerical Analysis, 54 (2020), pp. 2159–2197.

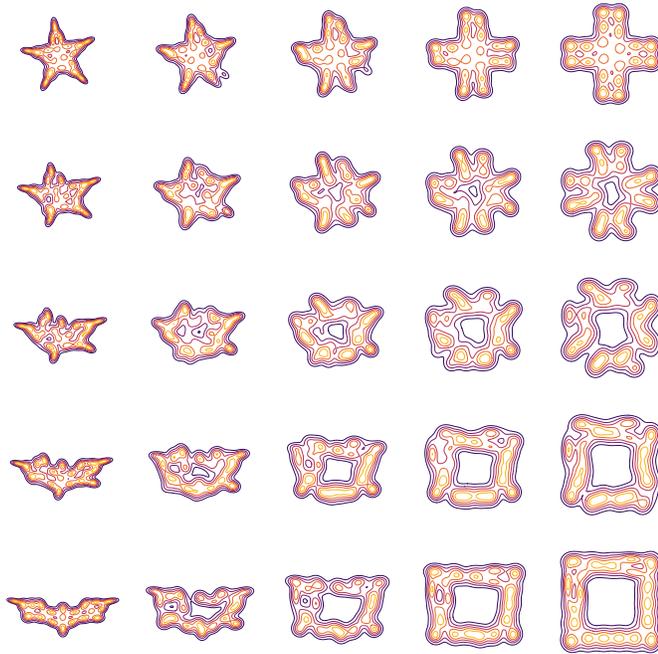


FIG. 7. *Exact Wasserstein barycenters between 4 mixtures of 20 Gaussian components for the metric  $MW_2$ .*

- [13] G. FRIESECKE AND M. PENKA, *The gencol algorithm for high-dimensional optimal transport: general formulation and application to barycenters and Wasserstein splines*, SIAM Journal on Mathematics of Data Science, 5 (2023), pp. 899–919.
- [14] W. GANGBO AND A. ŚWIĘCH, *Optimal maps for the multidimensional Monge-Kantorovich problem*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 51 (1998), pp. 23–45.
- [15] B. PASS, *Multi-marginal optimal transport: theory and applications*, ESAIM: Mathematical Modelling and Numerical Analysis, 49 (2015), pp. 1771–1790.
- [16] G. PEYRÉ AND M. CUTURI, *Computational optimal transport: With applications to data science*, Foundations and Trends® in Machine Learning, 11 (2019), pp. 355–607.
- [17] J. RABIN, G. PEYRÉ, J. DELON, AND M. BERNOT, *Wasserstein barycenter and its application to texture mixing*, in Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3, Springer, 2012, pp. 435–446.
- [18] A. J. STROMME, *Wasserstein barycenters: statistics and optimization*, PhD thesis, Massachusetts Institute of Technology, 2020.