Optimal Transportation of Multivariate Distributions using Copulas

N. Ghaffari¹, S. G. Walker¹,*

Department of Statistics and Data Sciences, University of Austin at Texas

Abstract
Optimal transports between multivariate distributions are difficult to find. Explicit results exist for multivariate Gaussian and special cases of elliptical distributions. Using results on the optimality of radial transformations and positive semidefinite operators, we find optimal transports between any two members of elliptical and simplicial distributions. We demonstrate an application for a Wasserstein barycenter problem, allowing us to obtain the barycenter among elliptical distributions with a common correlation structure.

Keywords: Wasserstein metric, optimal transport, simplicial distributions, elliptical distributions, Wasserstein barycenters

1. Introduction

The Wasserstein distances are metrics on the space of probability measures, and arise from the Kantorovich optimal transport problem. The Kantorovich transport problem is set up as follows: let \( \mathcal{P}^1(\mathbb{R}^d) \) denote the set of integrable probability measures on \( \mathbb{R}^d \), let \( P, Q \in \mathcal{P}^1(\mathbb{R}^d) \) be two such probability measures, and denote by \( \Pi(P, Q) \) the set of distributions on \( \mathbb{R}^{2d} \) with marginals \( P \) and \( Q \). Under an appropriate cost function \( c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+ \) the Kantorovich functional for a probability measure \( \pi \in \Pi(P, Q) \) is given by

\[
K_\pi(P, Q) \triangleq \int_{\mathbb{R}^{2d}} c(x, y)d\pi(x, y).
\]

The Kantorovich transport problem is to find the optimal measure \( \pi^* \in \Pi(P, Q) \) minimizing the Kantorovich functional

\[
\pi^* \triangleq \arg\min_{\pi \in \Pi} K_\pi(P, Q).
\]
The minimum of the Kantorovich functional $K^*$, attained under $\pi^*$, is a metric between $\mathbb{P}$ and $\mathbb{Q}$. For random variables $X \sim \mathbb{P}$ and $Y \sim \mathbb{Q}$, if $(X,Y) \sim \pi^*$, we say $X$ and $Y$ are optimally coupled. For more further details and more background on optimal transportation, see [14] and [15].

When the cost function is an $L^p$ distance, $c(x,y) = |x-y|^p$, the resulting metric is called the $p$-Wasserstein metric. Here we focus on the quadratic or 2-Wasserstein distance. Few explicit solutions exist. The major exception is the case of multivariate Gaussian measures; for two zero mean Gaussian measures with scale matrices $\Sigma_1 = A^T A$ and $\Sigma_2 = B^T B$, the optimal transport from the first measure to the second is given by

$$\phi^*(x) = B(ABAAB)^{-1/2}B^T x.$$  \hfill (3)

The optimal distance is also known,

$$W_2^2(\mathbb{P}, \mathbb{Q}) = \text{tr}(A^T A) + \text{tr}(B^T B) - 2\text{tr}\left(\sqrt{ABA}\right).$$  \hfill (4)

Some papers differ in formulations, since $\text{tr}\left(\sqrt{ABA}\right) = \text{tr}(AB)$. For further information on the Gaussian case, see [6], [12], [8], [9], and [13]. Optimal transports within certain classes of elliptically contoured distributions are also known, as in the Gaussian case, see [7].

In this paper we develop a general characterization of the optimal transport, under quadratic costs, between any two multivariate elliptically-contoured distributions, as well as between any two simplicially-contoured distributions. We present an application to the 2-Wasserstein barycenter problem. The remainder of the papers proceeds as follows. Section 2 presents relevant background information on elliptical and simplicial distributions. Section 3 presents established optimal coupling results under radial transforms and positive semidefinite matrices for the 2-Wasserstein. Section 4 applies these results to different classes of elliptical distributions and to simplicial distributions. Section 5 concludes with an application to the barycenter problem.

2. $L^1$ and $L^2$ Symmetric Distributions

In this section we provide background for simplicially and elliptically contoured distributions, starting with the former.

2.1. Simplicially Contoured Distributions

Let $U^{[d]}$ denote a uniform r.v. on the $d$-dimensional $L^1$ unit simplex or, equivalently, the $L^1$ unit sphere restricted to the positive orthant. When $d = 2$ this is simply a uniform random
variable (r.v.) on the straight line between $(1,0)$ and $(0,1)$. Let $R$ denote a nonnegative univariate random variable. Then $X \in \mathbb{R}_+^d$ has an $L^1$-symmetric, simplicially-contoured, simplicially-symmetric or simply simplicial distribution if $X \overset{d}{=} RU^d$, where $R$ is independent of $U^d$.

Note that $U^d$ has a Dir$(1, \ldots, 1)$ distribution. Using this we have expectation given by

$$E[X] = E[R]E[U^d] = E[R] \left( \frac{1}{d}, \ldots, \frac{1}{d} \right), \quad (5)$$

and when $\Sigma$ is the covariance matrix of a Dir$(1, \ldots, 1)$ r.v. and $W$ is a square matrix with all entries 1, we have the covariance matrix of $X$ as

$$E[XX^T] - E[X]E[X]^T = E[R^2] \Sigma - \frac{E[R^2]}{d^2} W. \quad (7)$$

2.2. Spherical & Elliptically Contoured Distributions

Let $O(d)$ denote the set of $d \times d$ orthonormal matrices. A random vector $X \in \mathbb{R}^d$ is said to have a spherical distribution if $QX \overset{d}{=} X$, for all $Q \in O(d)$. Let the r.v. $U^d$ have a uniform distribution on the unit sphere of $\mathbb{R}^d$, and let $R$ denote a nonnegative r.v., independent of $U^d$, then any $X$ with a spherical distribution has the representation

$$X \overset{d}{=} RU^d. \quad (6)$$

Such zero mean spherical distributions have covariance matrix of the form

$$E[XX^T] = E[R^2]E[U^d(U^d)^T] = d^{-1}E[R^2]I. \quad (7)$$

Let $\|\cdot\|_2$ denote the Euclidean norm; spherical random variables $X$ have the following properties, $\|X\|_2 \overset{d}{=} R$, and $X/\|X\|_2 \overset{d}{=} U^d$. Let $\Sigma$ be a $d \times d$ matrix of rank $k \leq d$ and $A$ a $d \times k$ matrix such that $AA^T = \Sigma$. Then for any $X \in \mathbb{R}^k$ having a $k$-dimensional spherical distribution, we have that $Y \overset{d}{=} AX$ is an $L^2$-symmetric, or elliptical distribution. Here we focus on the case that $\Sigma$ is of full rank, hence $k = d$. Elliptical random variables have a stochastic representation related to that of spherical distributions,

$$Y \overset{d}{=} RAU^d. \quad (8)$$

The action of the matrix $A$ is to convert the uniform on the sphere $U^d$ to a uniform on the ellipse $AU^d$. Zero mean elliptical distributions have covariance matrices, related to those spherical distributions, of form

$$E[YY^T] = d^{-1}AE[R^2]A^T = d^{-1}E[R^2] \Sigma. \quad (9)$$

Note then that $\Sigma$ is a scale matrix and, in general (the Gaussian case being an exception), is not the covariance. For more background on elliptical distributions and for a derivation of their stochastic representation, see [4], and [3].
3. Optimal Coupling: Background

In this section we recall results in the literature on the general optimality of radial and positive semidefinite transformations. For more details and the derivation of these results, see [5].

3.1. Positive Semidefinite Linear Transformations

Optimal couplings on $\mathbb{R}^d$ admit the following characterizations,

Lemma 3.1. If $\phi^*$ is continuously differentiable, then $(X, \phi^*(X))$ is $2$-W optimal coupling if and only if

1. $\phi^*$ is monotone, i.e. $\langle x - y, \phi^*(x) - \phi^*(y) \rangle \geq 0, \forall x, y$

2. $D\phi^* = \left( \frac{\partial \phi^*}{\partial x_j} \right)$ is symmetric; i.e., $\frac{\partial \phi^*}{\partial x_j} = \frac{\partial \phi^*}{\partial x_i}$

For more on this, see [5] and [9]. An immediate consequence of this is that for any positive semidefinite (PSD) matrix $M$, $(X, MX)$ is an optimal coupling with respect to the $2$-Wasserstein, regardless of the distribution of $X$.

3.2. Radial Transformations

A function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a radial transform if it is of the form

$$
\varphi(x) = \begin{cases} 
\frac{\alpha(||x||)}{||x||} x, & x \neq 0 \\
0, & x = 0 
\end{cases}
$$

(10)

with $\alpha(\cdot)$ nondecreasing.

Proposition 3.1. For $X \sim P$ and $Y = \varphi(X) \sim Q$ for some radial transform $\varphi$, it is that $(X, \varphi(X))$ is an optimal coupling with respect to the $2$-Wasserstein distance.

When $||X||$ and $||Y||$ are continuous, we have $F_{||X||}(||x||) = F_{||Y||}(\alpha(||x||))$, implying that $\alpha(||x||) = F_{||Y||}^{-1}(F_{||X||}(||x||))$. Thus, such radial transforms are of the form $y = \phi^*(x) = ||y|| / ||x|| x$. In fact radial transforms have much broader optimality properties; radial transforms are optimal couplings with respect to any cost that is a convex function $h(||x - y||)$ of the distance between $x$ and $y$. For this result and the previous proposition, see [5].

Spherical equivalence of distributions is defined as follows. Let $X \sim P$ and $Y \sim Q$ be distributions on $\mathbb{R}^d$ and $||X|| \sim P^{||\cdot||}$ and $||Y|| \sim Q^{||\cdot||}$ be the distributions of their respective
norms. Denote by $p_t$ and $q_t$ the $t$-th quantile of $\|\cdot\|$ and $\|\cdot\|$, respectively, for $t \in (0,1)$. Then $P$ and $Q$ are spherically equivalent if

$$P(x \in D | \|x\| = p_t) = P\left(y \in \frac{q_t}{p_t}D | \|y\| = q_t\right) \quad \forall t, \forall D \subseteq \mathbb{R}^n. \quad (11)$$

For spherically equivalent distributions, we cite the following two propositions; for the proofs see [5].

**Proposition 3.2.** Let $P$ and $Q$ be spherically equivalent. If $X \sim P$ and $Y \sim Q$ then $y \|y\| \overset{d}{=} x \|x\|$.  

**Proposition 3.3.** Let $P$ and $Q$ be two probability distributions and let $X \sim P$.

1. If $P$ and $Q$ are spherically equivalent and $\|X\|$ is continuous, then there exists a radial transformation $\varphi$ such that $\varphi(X) \sim Q$.

2. If $\varphi$ is a radial transformation such that $\varphi(X) \sim Q$, and $\alpha$ is strictly increasing, then $P$ and $Q$ are spherically equivalent.

3.3. Radial and PSD Transforms

After establishing the optimality of PSD matrices and radial transforms, it is natural to consider whether the composite transform is also optimal. However, [3] show that for $M$ PSD and $\alpha$ increasing ($X, (\alpha(\|X\|)/\|X\|)MX)$ is not necessarily an optimal transport. Nevertheless, the following does always result in an optimal transport,

**Proposition 3.4.** For $M$ a PSD matrix and $\alpha$ a nondecreasing function on $\mathbb{R}^+$ we have that

$$\left(X, \frac{\alpha(\|X\|_M)}{\|X\|_M} MX\right)$$

where $\|x\|_M = (x^T M x)^{1/2}$, is an optimal coupling with respect to the 2-Wasserstein. For the proof, see [3].

4. Optimal Coupling of Simplicial & Elliptical Distributions

Here we use radial and PSD transforms to establish optimal couplings and explicit Wasserstein distances for simplicial and elliptical distributions. Optimal couplings between simplicial distributions can be characterized through radial transforms. For optimal couplings between elliptical distributions we present three distinct cases. The first case, that of PSD transformations for location-scatter subfamilies of elliptical distributions, is known in the literature. The case of radial transformations and radial plus PSD transformations, to our knowledge, has yet to be demonstrated.
4.1. Simplicial Distributions

Let $X$ and $Y$ be simplicially contoured distributions in $\mathbb{R}^d$ with stochastic representations $X \overset{d}{=} RU^d$ and $Y \overset{d}{=} SU^d$. As before, $R$ and $S$ represent nonnegative one–dimensional random variables independent of $U^d$, a uniform random variable on the unit simplex. Then a simple radial transformation defines the optimal transport

**Theorem 4.1** (Simplicial Distributions). For $X$ and $Y$ defined as above, we have optimal transport,

$$
\phi^*(x) = \frac{\|y\|_2}{\|x\|_2} x = \frac{F^{-1}_S(F_R(\|x\|_1))}{\|x\|_1} x
$$

(12)

With 2-Wasserstein distance, formulated in terms of their stochastic representation,

$$
W_2^2(P, Q) = (E[R^2] + E[S^2] - 2E[RS]) \frac{d - 1}{d(d + 1)}
$$

(13)

**Proof.** First, note that $\phi^*(\cdot)$ is a radial transform, hence optimal by construction. To see that we obtain the correct target distribution,

$$
\frac{\|Y\|_2}{\|X\|_2} X = \frac{S \|U^d\|_2}{R \|U^d\|_2} X = \frac{S}{R} RU^d = SU^d \overset{d}{=} Y
$$

As before $S/R$ denotes the function that maps the $\omega$-th quantile of $R$ to the $\omega$-th quantile of $S$. This is explicitly formulated through $F_R$ and $F_S^{-1}$, the distribution and quantile function of $R$ and $S$, respectively. Then note that $\|X\|_1 = R$ by construction, so

$$
Y \overset{d}{=} \frac{F^{-1}_S(F_R(\|X\|_1))}{\|X\|_1} X
$$

The 2-Wasserstein distance is then easily obtained through the stochastic representation components:

$$
W_2^2(P, Q) = E[X^TX] + E[Y^TY] - 2E[X^TY] = E[R^2] \text{dvar}(U_i^d) + E[S^2] \text{dvar}(U_i^d) - 2E[RS] \text{dvar}(U_i^d) = (E[R^2] + E[S^2] - 2E[RS]) \frac{d - 1}{d(d + 1)}
$$

\[\square\]
4.2. Elliptical Distributions of Related Class

Let $X \sim P$ and $Y \sim Q$ be two elliptically distributed random variables admitting stochastic representations $X \overset{d}{=} RAU^{(d)}$ and $Y \overset{d}{=} RBU^{(d)}$. Thus $X$ and $Y$ are of the same family of elliptical distributions, having the same nonnegative random variable $R$, but generally having different correlations through differing scale matrices. Adding a location parameter makes these families a location-scatter subfamily of elliptical distributions. These are the “classes of related elliptically symmetric probability measures” that Gelbrich mentions; see [7].

Theorem 4.2 (Gelbrich, 1990). Let elliptical $X$ and $Y$ be distributed as above. Then the optimal transport is a PSD matrix given by,

$$
\phi^*(x) = M = B(BAAB)^{-1/2}Bx.
$$

and the corresponding distance may be obtained,

$$
W_2^2(P, Q) = \frac{E[R^2]}{d} (\text{tr}(AA) + \text{tr}(BB) - 2\text{tr}(AB)).
$$

Proof. By construction, $M$ is PSD, so $(X, MX)$ is an optimal coupling. That $M$ transports $X$ to $Y$ is readily observable as $MX = RMAU^{(d)} = RBU^{(d)} = Y$. The formulation for the Wasserstein distance is easily verified using the stochastic representation of elliptical distributions:

$$
W_2^2(P, Q) = \frac{E[R^2]}{d} \left( \text{tr}(AA) + \frac{E[R^2]}{d} \text{tr}(BB) - 2\frac{E[R^2]}{d} \text{tr}(AB) \right)
= \frac{E[R^2]}{d} \left( \text{tr}(AA) + \text{tr}(BB) - 2\text{tr}(AB) \right).
$$

We note here that $E[R^2] = d$ in the Gaussian case, hence the scale matrix is the covariance matrix; in the Student $t$ case, $E[R^2] = (\nu/(\nu - 2))d$; by contrast Gelbrich represents this distance with the covariance matrices, rather than separately displaying scale matrices and the nonnegative random variable component.

This covers the case of multivariate Gaussian distributions, appearing in [6], [12], [9], [8], and [13]. It also extends to elliptical distributions having the same nonnegative, univariate random variable in their stochastic representation, e.g. multivariate $t$ distributions of equal degrees of freedom, but possibly differing correlations.
4.3. Spherically Equivalent Elliptical Distributions

Let \( X \sim P \) and \( Y \sim Q \) be elliptically distributed random variables with stochastic representation \( X \overset{d}{=} RAU(d) \) and \( Y \overset{d}{=} SAU(d) \), respectively. Then \( X \) and \( Y \) have the same scale matrix and hence correlation structure, but different nonnegative random variables. It is easy to see that these distributions are spherically equivalent.

**Lemma 4.1.** For \( X \) and \( Y \) distributed as above, \( X \) and \( Y \) are spherically equivalent.

**Proof.** Let \( D \subseteq \mathbb{R}^d \) be any Borel subset and \( p_t \) and \( q_t \) the \( t \)-th quantile for the distributions of \( X \) and \( Y \), respectively. Note \( P \left( X \in D \mid \|X\|_2 = p_t \right) = P \left( RAU(d) \in D \mid \|AU(d)\|_2 = p_t \right) = P \left( UA \in D/p_t \right), \) where \( U_A = AU(d) / \|AU(d)\|_2 \). Also

\[
P \left( Y \in \frac{q_t}{p_t} D \mid \|Y\|_2 = q_t \right) = P \left( SAU(d) \in \frac{q_t}{p_t} D \mid \|AU(d)\|_2 = q_t \right) = P \left( U_A \in D/p_t \right),
\]

which completes the proof. \( \square \)

**Theorem 4.3** (Spherically Equivalent Elliptical Distributions). For \( X \) and \( Y \) distributed as above, the optimal transport taking \( X \) to \( Y \) is given by the radial transform,

\[
\phi^*(x) = \frac{\|y\|_2}{\|x\|_2} = \frac{F^{−1}_S(F_R(\|x\|_2))}{\|x\|_2}x.
\] (16)

With 2-Wasserstein distance,

\[
W^2_2(P, Q) = \frac{\text{tr}(AA)}{d} \left( E[R^2] + E[S^2] - 2E[RS] \right).
\] (17)

**Proof.** Since \( \phi^*(\cdot) \) is a radial transform, it is optimal. Then it is easily verified that we obtain the desired target distributions, for

\[
\frac{\|Y\|_2}{\|X\|_2} X = \frac{S}{R} \frac{\|AU(d)\|_2}{\|AU(d)\|_2} X = \frac{S}{R} RAU(d) = SAU(d) \overset{d}{=} Y
\]

Again, the 2-Wasserstein distance can be readily verified through the stochastic representation components:

\[
W^2_2(P, Q) = \frac{E[R^2]}{d} \text{tr} (AA) + \frac{E[S^2]}{d} \text{tr} (AA) - 2\frac{E[RS]}{d} \text{tr} (AA)
\]

\[
= \frac{\text{tr}(AA)}{d} \left( E[R^2] + E[S^2] - 2E[RS] \right).
\]
That is we map the $t$-th quantile of $R$ to the $t$-th quantile of $S$. Such a map implicitly defines a nondecreasing map on $\|X\|_2$ because of the cancellation of $\|AU^{(d)}\|_2$. If we know the scale matrix and know the distribution of $R$ and $S$, this yields a simple method for defining the optimal transport, since $R = \|A^{-1}X\|_2$. Thus $F_S^{-1}(F_R(\|A^{-1}X\|_2)) / \|A^{-1}X\|_2$ is the optimal transport for such distributions.

This case addresses the example of multivariate $t$ distributions with the same scale matrix, but different degrees of freedom. The resulting 2-Wasserstein distance can formulated via the elements of the stochastic representation, similar to the previous case.

4.4. General Elliptical Distributions

Let $X \overset{d}{=} RAU^{(d)}$ and $Y \overset{d}{=} SBU^{(d)}$ be general elliptical distributions with different nonnegative random variables and different scale matrices.

**Theorem 4.4 (General Elliptical Distributions).** For $X$ and $Y$ distributed as above, the optimal transport is a combination of the previous two cases,

$$\phi^*(x) = F_S^{-1}(F_R(\|A^{-1}X\|_2)) \frac{Mx}{\|A^{-1}X\|_2} \quad (18)$$

As before, using the stochastic representation, the 2-Wasserstein distance may be formulated,

$$W_2^2(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} (E[R^2] \text{tr}(AA) + E[S^2] \text{tr}(BB) - 2E[RS] \text{tr}(AB)). \quad (19)$$

**Proof.** To see that this is optimal note that because $R$ is scalar-valued, $\|X\|_M = (X^TMX)^{1/2}$ can be written as $R((AU^{(d)})^TM(AU^{(d)}))$. Then the map $\alpha$ mapping $R((AU^{(d)})^TM(AU^{(d)})) \mapsto S((AU^{(d)})^TM(AU^{(d)}))$ such that $t$-th quantile of $R$ maps to the $t$-th quantile of $S$ is nondecreasing and

$$\frac{\alpha(\|X\|_M)}{\|X\|_M} = \frac{S((AU^{(n)})^TM(AU^{(n)}))^{1/2}}{R((AU^{(n)})^TM(AU^{(n)}))^{1/2}} = \frac{S}{R}.$$

Thus, as in the case of radial transforms, for elliptical distributions, mapping $R$ to $S$ by quantiles, implicitly defines a nondecreasing function $\alpha$ on $\|\cdot\|_M$ and the other components of the mapping cancel in the radial transform ratio. Hence $\phi^*(x)$ as defined is of the form

$$\frac{\alpha(\|x\|_M)}{\|x\|_M} Mx.$$
hence an optimal transport and delivers the desired target distribution. The distance formulation is again easily verified by using the stochastic representation: As before, using the stochastic representation, the 2-Wasserstein distance may be formulated,

$$W_2^2(P, Q) = \frac{E[R^2]}{d} \text{tr} (AA) + \frac{E[S^2]}{d} \text{tr} (AA) - 2 \frac{E[RS]}{d} \text{tr} (AB)$$

$$= \frac{1}{d} \left( E[R^2] \text{tr} (AA) + E[S^2] \text{tr} (BB) - 2 E[RS] \text{tr} (AB) \right).$$

As an illustration, we can formulate an optimal transport for general multivariate $t$-distributions. Consider $T \in \mathbb{R}^d$ a zero mean $t$-distribution with degrees of freedom $\nu$ and scale matrix $\Sigma = A^T A$, (hence covariance matrix $\Lambda = \frac{\nu}{\nu - 2} \Sigma$). Then we have the standard representation $T \overset{d}{=} \sqrt{\nu/\chi^2_\nu} AZ$ for $\chi^2_\nu$ a $\chi^2$ distribution with $\nu$ degrees of freedom and $Z$ a $d$-dimensional standard normal. Then $Z$ has a standard spherical representation $\|Z\|_2 \overset{d}{=} U^{(d)}$ where $\|Z\|_2$ is a $\chi$ distribution with $d$ degrees of freedom. Then for $Q = \|Z\|_2 \sqrt{\nu/\chi^2_\nu}$ we have that $Q^2/d$ is an $F$ distribution with $(\nu, d)$ degrees of freedom, and we have elliptical representation $T \overset{d}{=} QAU^{(d)}$ and $\|A^{-1}T\|_2 = Q$.

Now, consider $X, Y \in \mathbb{R}^d$ $t$-distributions with degrees of freedom $\nu_x$ and $\nu_y$ and scale matrices $\Sigma_x = A^T A$, $\Sigma_y = B^T B$, respectively. Define $Q_x = \|A^{-1}X\|_2$ and $Q_y = \|B^{-1}Y\|_2$ as above, then the optimal transport is

$$\phi^*(x) = \sqrt{\frac{dF_{Q_y}^{-1} \left(F_{Q_x} \left( \frac{\|A^{-1}x\|_2}{d} \right) \right)}{\|A^{-1}x\|_2}} B(BAAB)^{-1/2} B x.$$  \hspace{1cm} (20)

5. The Wasserstein Barycenter Problem

In this section we describe the Wasserstein barycenter problem and show an application of the result on spherically equivalent elliptical distributions to obtaining the 2-Wasserstein barycenter among distributions within these classes of elliptical distributions.

The $W_2$-barycenter problem is as follows: Given distributions $\nu_1, \ldots, \nu_n \in \mathcal{P}_2(\mathbb{R}^d)$ in the set of square integrable $d$-dimensional probability distributions and weights $\{\lambda_i\}_{i=1}^n$ such that $\sum_{i=1}^n \lambda_i = 1$, find a distribution $\bar{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ such that

$$\bar{\mu} \triangleq \arg\min_{\mu \in \mathcal{P}_2} \sum_{i=1}^n \lambda_i W_2^2(\nu_i, \mu)$$  \hspace{1cm} (21)
For existence, consistency, and uniqueness, see [1] and [10]. Then $\bar{\mu}$ is the barycenter (w.r.t. the $W_2$ distance) of $\nu_1, \ldots, \nu_n$. It is a minimizer for the barycentric cost functional,

$$V(\mu) \triangleq \sum_{i=1}^{n} \lambda_i W_2^2(\nu_i, \mu)$$

(22)

5.1. Fixed-Point Approach

In the following $\mathcal{P}_{2,ac}(\mathbb{R}^d)$ denotes the set of absolutely continuous, square-integrable probability measures on $\mathbb{R}$. The authors in [2] develop a fixed point approach to finding the barycenters; a functional $G : \mathcal{P}_{2,ac}(\mathbb{R}^d) \to \mathcal{P}_{2,ac}(\mathbb{R}^d)$ is constructed whose fixed points, under basic assumptions, are the barycenters of $\nu_1, \ldots, \nu_n \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$ with weights $\lambda_1, \ldots, \lambda_n$. Let $\mu \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$, and define, for $j = 1, \ldots, n$, $T_j$ such that $T_j(X) \sim \nu_j$ for $X \sim \mu$. So $T_j$ is an optimal transport, known to exist from optimal transportation theory, and $W_2^2(\mu, \nu_j) = E[\|X - T_j(X)\|^2]$. Then

$$G(\mu) \triangleq \mathcal{L} \left( \sum_{i=1}^{n} \lambda_i T_j(X) \right),$$

(23)

where $\mathcal{L}$ denotes the distribution, or “law,” of the underlying random variable.

Theorem 5.1. If $\nu_j$ has a density, $j = 1, \ldots, n$ then $G$ maps $\mathcal{P}_{2,ac}(\mathbb{R}^d)$ into $\mathcal{P}_{2,ac}(\mathbb{R}^d)$ and it is continuous for the $W_2$ metric.

Proposition 5.1. If $\mu \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$ then

$$V(\mu) \geq V(G(\mu)) + W_2^2(\mu, G(\mu)).$$

As a consequence, $V(\mu) \geq V(G(\mu))$, with strict inequality if $\mu \neq G(\mu)$. For a barycenter $\bar{\mu}$, $G(\bar{\mu}) = \bar{\mu}$

For a full derivation of this method, see [2].

5.2. Elliptical Distributions with Common Correlation

In [2] the authors use their fixed point method to obtain the barycenter between elliptical distributions of a common family, i.e. elliptical distributions with a common nonnegative random variable $R$ but possibly different scale matrices $\Sigma$. At each iteration, the output is another elliptical distribution in the same family, i.e. with the same $R$. Thus only the scale matrix varies with iteration until the barycenter has been obtained.
Here we demonstrate how to obtain the barycenter between elliptic distributions with common correlation structure but potentially different nonnegative random variables $R$. In particular, the fixed point operator of [2] reduces to a one dimensional problem and converges in one iteration. The barycenter is thus a quantile mixture of the nonnegative $R$ with the same correlation structure.

Let $\nu_1, \ldots, \nu_n$ denote elliptical distributions with common scale matrices $\Sigma = A^T A$ and nonnegative random variables $S_1, \ldots, S_n$, respectively, in their stochastic representation. Let $\mu$ be the initial proposal, an elliptical distribution represented with scale matrix $\Sigma$ and nonnegative random variable $R$. Let $T_j$, for $j = 1, \ldots, n$, represent the optimal transport between $\mu$ and $\nu_j$.

**Theorem 5.2.** Given elliptical distributions $\nu_j$, $j = 1, \ldots, n$, and $\mu$ as defined above, the fixed point operator $G(\mu)$ converges in one iteration to the barycenter, an elliptical distribution with scale matrix $A$ and nonnegative random variable $R^\dagger$ such that

$$F_{R^\dagger}^{-1}(\omega) = \sum_{j=1}^{n} \lambda_j F_{S_j}^{-1}(\omega),$$

i.e. $R^\dagger$ is a weighted quantile mixture of the nonnegative $S_1, \ldots, S_n$.

**Proof.** Denote $X \sim \mu$ and $Y_j \sim \nu_j$. Recall $G(\mu) = \mathcal{L}\left(\sum_{j=1}^{n} \lambda_j T_j(X)\right)$. Since the $\nu_j$ and $\mu$ are elliptical with common scale matrices and $T_j$ is the optimal transport we may write

$$\sum_{j=1}^{n} \lambda_j T_j(x) = \sum_{j=1}^{n} \lambda_j \frac{y_j}{\|x\|_2} x \frac{\|y_j\|_2}{\|x\|_2} = \sum_{j=1}^{n} \lambda_j \frac{F_{S_j}^{-1}(F_R(\|A^{-1}x\|_2))}{\|A^{-1}x\|_2} R AU^{(d)}.$$  

Pulling the common $AU^{(d)}$ out of the summand and noting that $\|A^{-1}x\|_2 \overset{d}{=} R$

$$\sum_{j=1}^{n} \lambda_j T_j(x) = AU^{(d)} \sum_{j=1}^{n} \lambda_j F_{S_j}^{-1}(F_R(\|A^{-1}x\|_2))$$

$$= AU^{(d)} \sum_{j=1}^{n} \lambda_j F_{S_j}^{-1}(\omega)$$

The last equality through noting that $F_R(\|A^{-1}x\|_2) = \omega$ is common to each term of the summand, so that the final form is a weighted mixture of quantiles. Since each $S_j$ is a nonnegative scalar random variable it is easily observed that the quantile mixture of $S_j$ terms is
also a nonnegative scalar random variable. As the $AU^{(d)}$ is unchanged the resulting distribution is an elliptical distribution with the same correlation (hence spherically equivalent) and a "quantile averaged" radial component. To see that it converges in one iteration let $\mu^\dagger$ denote the output, i.e.

$$
\mu^\dagger = \mathcal{L} \left( AU^{(d)} \sum_{j=1}^{n} \lambda_j F_{S_j}^{-1} \left( FR\left(\|A^{-1}x\|_2\right)\right) \right).
$$

Then for $Z \sim \mu^\dagger$ we have that $Z \overset{d}{=} R^\dagger A U^{(d)}$ where $F_{R^\dagger}^{-1}(\omega) = \sum_{j=1}^{n} \lambda_j F_{S_j}^{-1}(\omega)$. Let $T^\dagger_j(Z)$ denote the optimal map taking $Z$ to $Y_j$. Then it is easily verified that

$$
\sum_{j=1}^{n} \lambda_j T^\dagger_j(z) = AU^{(d)} \sum_{j=1}^{n} \lambda_j F_{S_j}^{-1}(\omega),
$$

where $\omega$ denotes $F_{R^\dagger}(\|A^{-1}z\|_2)$. Hence $G(\mu^\dagger) = \mu^\dagger$.

This is the case of elliptical distributions with common correlation but different radial components. For example, this case covers the barycenter of $t$ distributions with different degrees of freedom but the same correlations. We note that this also covers the case of finding the $W_2$ barycenter among $d$-dimensional simplicial distributions. The intuition is much the same and the resulting barycenter is also a quantile average of nonnegative radial components of each constituent measure in the barycenter.

5.3. Illustration

As a numerical illustration we find the barycenter for $t$ distributions with common covariance matrices but different degrees of freedom. The barycenter, as previously demonstrated, is elliptical with the nonnegative radial component a quantile average of the constituents. In general, the resulting elliptical barycenter is not a $t$ distribution, but in practice it is close to a $t$ distribution; using the 2-Wasserstein distance, we can obtain degrees of freedom for the $t$ distribution closest to this barycenter.

Possible application areas include financial and econometric data, where heavy tails are frequently encountered; $t$ distributions exhibit heavy tails with the heaviness increasing as the degrees of freedom, $\nu$, decrease. As such, $t$ distributions often more accurately model the risk of extreme events in financial and economic markets when compared to Gaussian distributions. Different techniques exist for modeling degrees of freedom for estimating tail risk; these include tail index techniques and excess kurtosis. For more, see for example [11].
For illustration, we consider a scenario where the covariance matrix and the degrees of freedom for a multivariate model are estimated separately. If we have a single estimate for the covariance and several different estimates for the degrees of freedom, we may use a Wasserstein barycenter to find an “average” distribution and hence degrees of freedom. For simplicity, we assume all distributions are centered and have unit variance scale matrices with a common correlation structure. These scale matrices are multiplied by \((\nu - 2)/\nu\), since the covariance is \(\nu/(\nu - 2)\) multiplied by the scale matrix; this insures all distributions have the same covariance matrix. For \(\mu_1, \ldots, \mu_n\) in this setup, having nonnegative radial components \(Q_1, \ldots, Q_n\) and degrees of freedom \(\nu_1, \ldots, \nu_n\) respectively the nonnegative radial component of the barycenter \(\bar{Q}\) is in terms of its distribution function

\[
F^{-1}_{\bar{Q}}(\cdot) = \sum_{j=1}^{n} \frac{\lambda_j(\nu_j - 2)}{\nu_j} F^{-1}_{Q_j}(\cdot).
\]

We then find the \(Q^*\) for a \(t\) distribution with \(\nu^*\) degrees of freedom, that minimizes the 2-Wasserstein distance to \(\bar{Q}\).

We first consider a barycenter between two 2-dimensional \(t\) distributions with degrees of freedom 3 and 27 each with weight 1/2; the former has fairly heavy tails while the latter is fairly close to a Gaussian. The resulting barycenter was found to be closest to a \(t\) distributions with 5.5 degrees of freedom. This is in contrast to the arithmetic average of 15 the degrees of freedom and indicates that Wasserstein barycenter places greater weight on the heavy tails and greater uncertainty associated with 3 degrees of freedom.

Next we examined the case of the barycenter twenty eight different 2-dimensional \(t\) distributions with common covariance and degrees of freedom 3, 4, \ldots, 30, each with weight 1/28. This barycenter was closest to a \(t\) distribution with 11.3 degrees of freedom whereas the arithmetic mean is 16 degrees of freedom. The arithmetic average weights each degree of freedom equally while the Wasserstein barycenter takes into account the underlying measures and hence, as opposed to the arithmetic mean, the barycenter weight lower degrees of freedom more heavily emphasizing the greater uncertainty represented by these distributions.

References

References

[1] M. Agueh and G Carlier. Barycenters in the wasserstein space. *SIAM Journal of Mathematical Analysis*, 43:904–924, 2011.
[2] P.C. Álvarez Esteban, E. del Barrio, J.A. Cuesta-Albertos, and C. Matrán. A fixed-point approach to barycenters in wasserstein space. *Journal of Multivariate Analysis*, 441:744–762, 2016.

[3] T. Anderson and K. Fang. Theory and applications of elliptically contoured and related distributions. Technical Report 24, Department of Statistics, Stanford University, 1990.

[4] S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11:368–385, 1981.

[5] J.A. Cuesta-Albertos, L. Rüschendorf, and A. Tuero-Diaz. Optimal coupling of multivariate distributions and stochastic processes. *Journal of Multivariate Analysis*, 46:335–361, 1993.

[6] D.C. Dowson and B.V. Landau. The fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis*, 12:450–455, 1982.

[7] M. Gelbrich. On a formula for the $L^2$ wasserstein distance between measures on euclidean and hilbert spaces. *Mathematische Nachrichten*, 147:185–203, 1990.

[8] Clark R. Givens and Rae Michael Shortt. A class of wasserstein metrics for probability distributions. *Michigan Math Journal*, 31:231–240, 1984.

[9] M. Knott and C.S. Smith. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43:39–49, 1984.

[10] T. Le Gouic and J. Loubes. Existence and consistency of wasserstein barycenters. *Probability Theory and Related Fields*, 168:901–917, 2017.

[11] Chu-Hsiung Lin and Shan-Shan Shen. Can the student-$t$ distribution provide accurate at risk value? *Journal of Financial Risk*, 7:292–301, 2006.

[12] I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263, 1982.

[13] A. Takatsu. Wasserstein geometry of gaussian measures. *Osaka Journal of Mathematics*, 48:1005–1026, 2011.

[14] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics. American Mathematical Society, 2003.

[15] C. Villani. *Optimal Transport, Old and New*. Springer, 2006.