Sliced Wasserstein Barycenter of Multiple Densities

Sliced Wasserstein Barycenter of Multiple Densities

Nicolas Bonneel
and
Hanspeter Pfister

TR-02-13

Computer Science Group
Harvard University
Cambridge, Massachusetts

Sliced Wasserstein Barycenter of Multiple Densities

Nicolas Bonneel Hanspeter Pfister
Harvard University

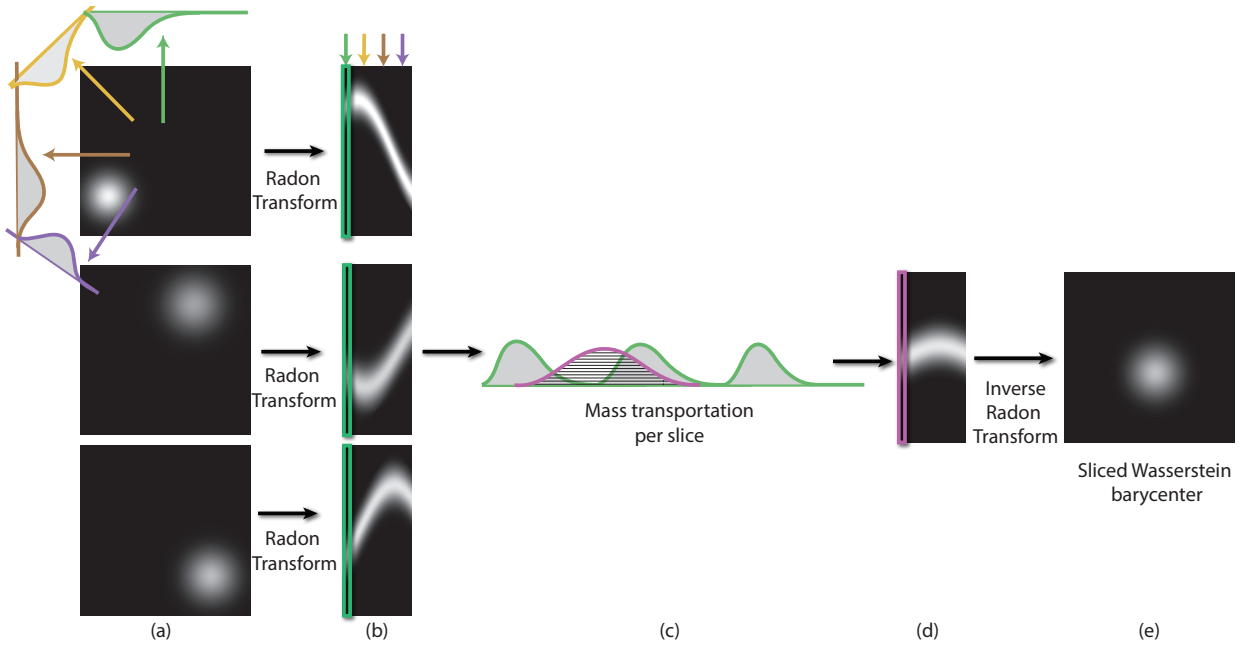

Figure 1: *From a set of discretized functions (a), we compute their Radon transform (b) and perform 1D mass transportation problems (c). We perform an inverse Radon transform on the result (d) to obtain the desired interpolated function (e).*

Abstract

The optimal mass transportation problem provides a framework for interpolating between different probability density functions (pdfs), warping one function toward another. Interpolating between two arbitrary pdfs can be challenging, but interpolating between more than two pdfs just remains untractable. We propose an approximation of such interpolations based on 1D projections, that is efficiently solved via Radon transforms. We observe the expected translational behavior of this interpolation on smooth 2D functions, and prove that it corresponds to the exact interpolant in a few particular cases.

1 Introduction

Mass transportation consists in moving a probability density function (pdf) f_0 toward a pdf f_1 in a way that minimizes the effort needed to perform this motion [Villani 2003]. The effort is expressed as the cost it would require to move a pile of sand representing f_0 toward a hole made of f_1, by summing the cost (typically a squared distance) for each particle of sand to reach its destination in the hole (see Fig. 2). The mass transportation problem leads to an interpolation, called *displacement interpolation*, that consists in moving the function f_0 partway toward f_1, leading to an interpolated pdf f_t ($t \in [0,1]$). This has shown to have a wide range of applications, from computer graphics [Bonneel et al. 2011, Rabin et al. 2011, Matusik et al. 2005] to image processing [Pitié et al. 2005, Haker et al. 2004], due to the characteristic advection produced during the interpolation. While this interpolation can be very efficiently performed in one dimension, extending it to multiple dimensions is much more challenging and leads to computationally expensive solutions [Bonneel et al. 2011]. In addition, further extending the approach to an interpolation between more than two pdfs remains untractable.

We propose an approximation of the displacement interpolation based on 1D projections, efficiently performed with a Radon transform. A benefit of this approach is the handling of a displacement interpolation between any number of pdfs. This interpolant is called a Wasserstein barycenter, and the above approximation, the sliced Wasserstein barycenter. We evaluate our algorithm on smooth pdfs and images, and show that the expected translational behavior is obtained for smooth pdfs. While we show results in two dimensions, generalizing the method to higher dimensions is

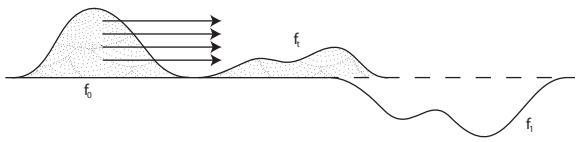

Figure 2: *The mass transportation problem consists in optimally moving a pile of sand represented by the function f_0 toward a hole made by the function f_1. This problem is also known as the Monge problem. During the motion, at an intermediate time t, an interpolated function f_t is obtained. This is called the* displacement interpolation*, and is generalized as a* Wasserstein barycenter *when considering more than $N = 2$ functions as input.*

straightforward. We demonstrate that our procedure also produces the exact Wasserstein barycenter in the case of translated and scaled functions, as well as isotropic Gaussians. We typically obtain the sliced Wasserstein barycenter of three functions more than four orders of magnitude faster than state of the art approaches for the exact Wasserstein barycenter of two functions [Papadakis et al. 2013]. We finally show how an existing formula used in computer graphics [Matusik et al. 2005] relates to the Wasserstein barycenter.

2 Previous work

2.1 Discrete Mass Transportation

Given a pdf $f_0(x_i)$ whose values are known at locations $x_i \in \mathbb{R}^n$, and $f_1(x_j)$, $x_j \in \mathbb{R}^n$, the mass transportation problem with quadratic cost consists in solving for a transport plan $t_{i,j}$ describing the amount of mass at location x_i in $f_0(x_i)$ that should move toward x_j so as to minimize a transportation cost. This is performed by solving:

$$\text{argmin}_{t_{i,j}} ||x_i - x_j||^2 \, t_{i,j} \tag{1}$$

$$\text{subject to} \quad t_{i,j} \geq 0 \tag{2}$$

$$\sum_i t_{i,j} = f_1(x_j) \tag{3}$$

$$\sum_j t_{i,j} = f_0(x_i) \tag{4}$$

where Eq. 1 is the cost of moving all the mass, Eq. 2 enforces the positivity of the transported mass $t_{i,j}$, and Eqs. 3 and 4 indicate that the entire mass going to x_j (resp. coming from x_i) is equal to $f_1(x_j)$ (resp. $f_0(x_i)$). Solving for $t_{i,j}$ and moving the mass only partway between x_i and x_j defines the displacement interpolation formulated in a discrete way. This problem can directly be solved via linear programming [Bonneel et al. 2011], or rather formulated in a continuous fashion by accounting for the particular structure of the quadratic cost problem. For instance, a fluid dynamic interpretation has been exploited by Benamou and Brenier [2000] which requires solving for a space-time Laplace equation. Haker et al. uses Brenier's polar factorization theorem to build a curl-free velocity field solving the transportation problem [Haker et al. 2004]. Proximal splitting has shown to improve performances [Papadakis et al. 2013] as well as geometric methods [Mérigot 2011]. These approaches however remain complex, and limited to the interpolation between two pdfs.

Despite this complexity in the general case, a few particular cases remain tractable, even for interpolating between multiple pdfs. A first instance is the interpolation between multiple n-dimensional Gaussian densities that can be performed with a fixed-point method (or with a closed-form solution in the case of two n-dimensional Gaussians only) [Ferradans et al. 2013]. A second instance is the interpolation between arbitrary one-dimensional pdfs, as explained next.

2.2 Sliced approximations

Computing an optimal Wasserstein barycenter on the real line – i.e., a displacement interpolation between multiple 1D pdfs f_i weighted by ω_i with $\sum_{i=1}^{N} \omega_i = 1$ – is a straightforward process [Agueh and Carlier 2011]. Benefitting from the monotonous transport plan, Rabin et al. [Rabin et al. 2011] hence perform a stochastic gradient descent to compute an approximate Wasserstein barycenter of point distributions that sample several pdfs. This approximation is called the *Sliced Wasserstein Barycenter*, and consists in projecting and ordering these points on random 1D lines, ultimately solving for a set of 1D mass transportation problems. The main difference with our method is that we are directly interested in the interpolated pdf rather than a set of points that samples it. Pitié et al. [2005] use a similar projection-based approach to transfer colors in images.

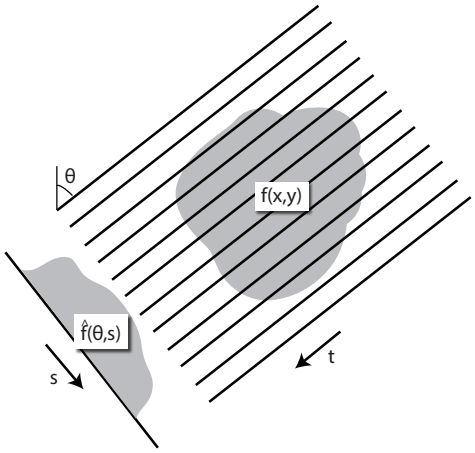

Figure 3: *The 2D Radon transform projects a function $f(x, y)$ onto a 1D slice perpendicular to θ. Using different values of θ, we obtain the transformed function $\hat{f}(\theta, s)$.*

3 Sliced Wasserstein barycenter

Relying on 1D projections, we first describe how we simplify the exact Wasserstein barycenter on the real line. We then describe how these projections are used to build an approximate n-dimensional interpolation.

3.1 Multi-marginal interpolation on the real line

Agueh and Carlier [2011] introduced the notion of barycenter in the Wasserstein space as a way to characterize a weighted average of probability distributions, and provided a practical method to compute it on the real line. We further simplify the expression of this Wasserstein barycenter, and show that it can simply be computed as the derivative of the inverse of a weighted sum of inverse cumulative density functions, as has been intuitively used by Matusik et al. [2005] in the context of texture interpolation.

Assuming that f_i (for $i = 1 \ldots N$) has no atom and has finite second moments, the barycenter $\bar{f} = \text{bar}(f_i, \omega_i)$ can be computed as follows. Let F_i be the cumulative density function (cdf) of f_i:

$$F_i(t) = \int_0^t f_i(x) \, \mathrm{dx} \tag{5}$$

Let F_i^{-1} denote the generalized inverse:

$$F_i^{-1}(t) = \inf\{x \in \mathbb{R}; F_i(x) > t\} \tag{6}$$

And let $\sharp$ denote the push-forward operator for densities, that is, in our case and for a nondecreasing function T,

$$\bar{f} = T\sharp f_1 \iff f_1(x) = \bar{f}(T(x))\,T'(x)$$

Finally, let $T_i = F_i^{-1} \circ F_1$. Then, the barycenter can be computed as [Agueh and Carlier 2011]:

$$\bar{f} = \mathrm{bar}(f_i, \omega_i) = \left(\sum_{i=1}^{N} \omega_i T_i\right) \sharp f_1$$

We now show how this can be further simplified. We have,

$$f_1(x) = \bar{f}(T(x))\,T'(x) \iff F_1 = \bar{F} \circ T$$

with $\bar{F}$ the cdf of $\bar{f}$. It follows that

$$\bar{F}^{-1} = \left(\sum_{i=1}^{N} \omega_i T_i\right) \circ F_1^{-1}$$

Using the expression of T_i, we obtain the inverse cumulative density function of the barycenter as:

$$\bar{F}^{-1} = \sum_{i=1}^{N} \omega_i F_i^{-1} \tag{7}$$

The resulting 1D function can finally be inverted using Eq.6 and differentiated to obtain the barycenter $\bar{f}$. This corresponds to the formula used by Matusik et al. [2005].

3.2 Interpolation in $\mathbb{R}^2$

We build upon the Sliced Wasserstein Barycenter but directly perform our computations with pdfs rather than a sampling of the pdfs. We rely on the Radon transform [Deans 2007], typically used in medical imaging for the tomography problem, to represent a pdf in $\mathbb{R}^n$ in the space of its projections on $(n-1)$-D hyperplanes. To make the implementation simpler, we will focus on the case $n = 2$ for which the Radon transform consists in projecting a 2D image on 1D lines.

In 2D, the Radon transform is expressed as:

$$\hat{f}(\theta, s) = R(f(x,y))(\theta, s) \tag{8}$$

$$= \int_{-\infty}^{+\infty} f(t\,\sin\theta + s\,\cos\theta, -t\cos\theta + s\,\sin\theta)\,\mathrm{dt} \tag{9}$$

and intuitively represents the sum of the function $f(x,y)$ accumulated in the direction $(\sin\theta, \cos\theta)$ (see Fig.3). The inverse Radon transform is expressed as:

$$f(x,y) = \int_0^{2\pi} \hat{f}(\theta, x\,\cos\theta + y\,\sin\theta)\,\mathrm{d}\theta$$

and can be regarded as the inverse operation, bringing back the transformed function $\hat{f}$ back into its original domain. Fortunately, both the Radon transform and its inverse can be efficiently computed via fast Fourier transforms, thanks to the Fourier Slice theorem [Deans 2007].

To compute the Sliced Wasserstein Barycenter of $f_1, f_2, \ldots, f_N$ with weights $\omega_1, \omega_2, \ldots, \omega_N$, we first Radon transform each $f_i(x,y)$ to obtain $\hat{f}_i(\theta, s)$. We then consider each $\hat{f}_i(\theta, \cdot)$ as a set of 1-D functions (or *slices*) parameterized by θ. For each slice θ, we compute a barycenter $\hat{\bar{f}}(\theta, \cdot) = \mathrm{bar}(\hat{f}_i(\theta, \cdot), \omega_i)$ using Eq. 7. We finally compute the inverse Radon transform of $\hat{\bar{f}}(\theta, s)$ to obtain the sliced Wasserstein barycenter $\bar{f}(x,y)$ (see Alg. 1 and Fig.1).

Algorithm 1 Compute $\bar{f} = \mathrm{bar}(f_i, \omega_i)$

$\hat{f}_i(\theta, s) \leftarrow \mathrm{radon}(f_i(x,y))$
for all angle θ_j, $j = 1 \ldots K$ **do**
  $\hat{F}_i \leftarrow \mathrm{cdf}(\hat{f}_i(\theta_j, .))$, $\forall i$, using Eq. 5
  $\hat{F}_i^{-1}$ is obtained from $\hat{F}_i$, $\forall i$, with Eq. 6
  $\hat{\bar{F}}^{-1}(\theta_j, \cdot) \leftarrow \sum_{i=1}^{N} \omega_i \hat{F}_i^{-1}$
  $\hat{\bar{F}}(\theta_j, \cdot)$ is obtained from $\hat{\bar{F}}^{-1}(\theta_j, \cdot)$ using Eq. 6
  $\hat{\bar{f}}(\theta_j, \cdot)$ is obtained from $\hat{\bar{F}}(\theta_j, \cdot)$ by numerical differentiation
end for
return $\bar{f}(x,y) \leftarrow \mathrm{inverse\_radon}(\hat{\bar{f}}(\theta, s))$

3.3 Properties

From properties of the Radon transform and displacement interpolation on the real line, we show through the following three theorems that the sliced Wasserstein barycenter recovers certain desirable properties of the exact Wasserstein barycenter.

Theorem 1. *The sliced Wasserstein barycenter of multiple translated copies of a probability distribution with density h is another translated copy of h, and is equal to the exact Wasserstein barycenter.*

Proof. Let $f_i(x,y) = h(x - x_i, y - y_i)$. The Radon transform of f_i is given by $\hat{f}_i(\theta, s - x_i\cos\theta - y_i\sin\theta)$. For a given slice, the 1D Wasserstein barycenter $\hat{\bar{f}}(\theta, \cdot)$ is given by $\hat{\bar{f}}(\theta, s) = \hat{h}(\theta, s - \sum_i \omega_i(x_i\cos\theta + y_i\sin\theta))$. This corresponds to the Radon transform of $\bar{f}(x,y) = h(x - \sum \omega_i x_i, y - \sum \omega_i y_i)$, as can be seen by the change of variable $t' \to t + \sum \omega_i(x_i\cos\theta - y_i\sin\theta)$ in Eq. 9, similarly to the exact Wasserstein barycenter [McCann 1997]. □

Theorem 2. *The sliced Wasserstein barycenter of multiple uniformly scaled copies of a probability distribution with density h is another uniformly scaled copy of h, and is equal to the exact Wasserstein barycenter.*

Proof. Let $f_i(x,y) = h(\lambda_i x, \lambda_i y)$. The Radon transform of f_i is given by $\frac{1}{\lambda_i}\hat{f}_i(\theta, \lambda_i s)$. For a given slice, the 1D Wasserstein barycenter $\hat{\bar{f}}(\theta, \cdot)$ is given by $\hat{\bar{f}}(\theta, s) = \hat{h}(\theta, s\sum_i \omega_i\lambda_i)$, up to a constant scaling. This corresponds to the Radon transform of $\bar{f}(x,y) = h(x\sum \omega_i\lambda_i, y\sum \omega_i\lambda_i)$, similarly to the exact Wasserstein barycenter [McCann 1997]. □

Theorem 3. *The sliced Wasserstein barycenter of isotropic Gaussians is equal to the exact Wasserstein barycenter.*

Proof. The Radon transform of an isotropic 2D Gaussian distribution $\mathcal{N}((\mu_x, \mu_y), \sigma)$ is given by (formula adapted from [Deans 2007] to non-centered Gaussians) :

$$\hat{f}(\theta, s) = \frac{1}{\sqrt{2\,\pi}} \exp(-(s - \mu_x\cos\theta - \mu_y\sin\theta)^2/(2\sigma^2))$$

For a fixed θ, and up to a normalization constant, this is a 1D Gaussian $\mathcal{N}(\mu_x\cos\theta + \mu_y\sin\theta, \sigma)$. The exact Wasserstein barycenter of a set of 1D Gaussians $\mathrm{bar}(\mathcal{N}(\mu_i, \sigma_i), \omega_i)$ is a 1D Gaussian $\mathcal{N}(\sum \omega_i\mu_i, \sum \omega_i\sigma_i)$. It follows that the 1D Wasserstein barycenter of each slice in the Radon domain is a Gaussian with averaged

mean and standard deviation. We hence recognize the Radon transform of a 2D Gaussian with averaged mean and standard deviation. This coincides with the exact 2D Wasserstein barycenter of isotropic Gaussians, and can also be seen as a corollary of theorems 1 and 2. □

However, the sliced barycenter of anisotropic Gaussians might not remain Gaussian as in the exact case. This can be easily seen when these Gaussians degenerate and become rank deficient[1] (see Fig. 7, top row). The sliced approximation is in general not equal to the original problem. We evaluate the differences in Sec. 4.2.

4 Results and discussion

In this section, we provide implementation details and discuss our results.

4.1 Implementation

We implemented our approach in Matlab. The cdf is computed with the trapeze method, and the function inversion is done using a linear search and a piecewise linear approximation of the function. The inverse Radon transform is computed by Matlab using a linear interpolation and a Hann filter, and a discretization every one degree (we did not notice further improvement by refining the discretization). Differentiation is performed using second order centered finite differences. Radon transforms are precomputed so that multiple displacement interpolations can be performed on the same functions with different parameters more efficiently.

4.2 Results

We analyze the method for both speed and accuracy. Although we do not aim at accurately representing the exact Wasserstein barycenter, we aim at loosely mimicking its behavior at a significant speed improvement.

Speed: A typical interpolation of three two-dimensional pdfs discretized on a 1024×1024 pixel grid requires 5 seconds for the initial Radon transforms, and 0.4 seconds to compute the barycenter and perform the inverse Radon transform. After storing the initial Radon transforms, we can typically compute 200 barycenters for the same three functions, in approximately 88 seconds, on a single core with unoptimized Matlab code. For comparison, the interpolation between two two-dimensions pdfs discretized on a 128×128 pixel grid requires 1721 seconds using [Papadakis et al. 2013], resuting in the exact interpolation, up to numerical precision (see Fig. 4). Using a 128×128 discretization, our method requires 0.1 second to produce the sliced Wasserstein barycenter of three functions, corresponding to a speedup of four orders of magnitude.

Qualitive results: In addition to the theoretical guaranties provided by this interpolation, we evaluate the sliced Wasserstein interpolation on several cases, including the interpolation between sums of Gaussians (Fig. 4 and 5 left), smooth images (Fig. 5 right) and photographs (Fig. 6). We show that interpolating between smooth probability distributions often produce a desired motion, while it may not be suited for image interpolation. We verify that the interpolation between isotropic Gaussians remains Gaussian (Fig. 1). Interpolation between four distributions behaves similarly to the case of three distributions (Fig. 8). In some highly anisotropic cases (Fig. 7) the resulting distribution might not be appropriate.

[1] We thank Marie-Paule Cani for this example.

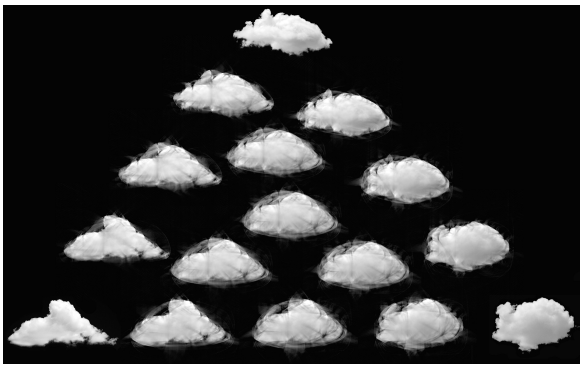

Figure 6: *The sliced Wasserstein barycenter does not properly handle non-smooth image interpolation as shown on this cloud example, and exhibits slicing artifacts.*

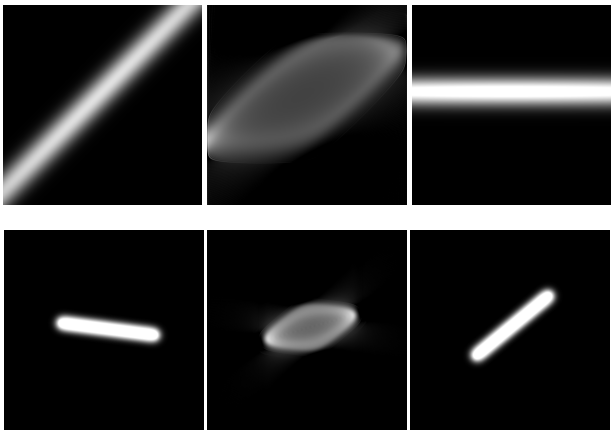

Figure 7: *The sliced Wasserstein barycenter between two anisotropic Gaussians does remain Gaussian, as shown with the degenerate truncated Gaussian on the top row. A line segment is not interpolated as a line segment (bottom row).*

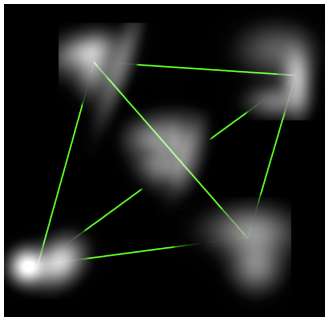

Figure 8: *The sliced Wasserstein barycenter of four 2D distributions with equal weights.*

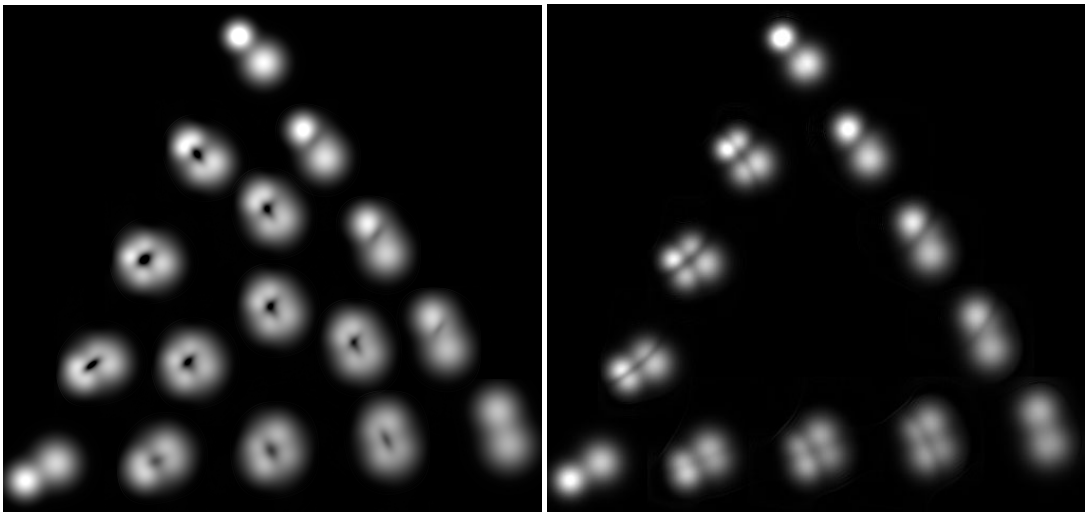

Figure 4: *The sliced Wasserstein barycenter (left) does not, in general, produce the same result as the exact Wasserstein barycenter (right, [Papadakis et al. 2013]). However, it is much faster (more than four orders of magnitude), works equally well for the barycenter of $N > 2$ probability distributions and shares similar properties.*

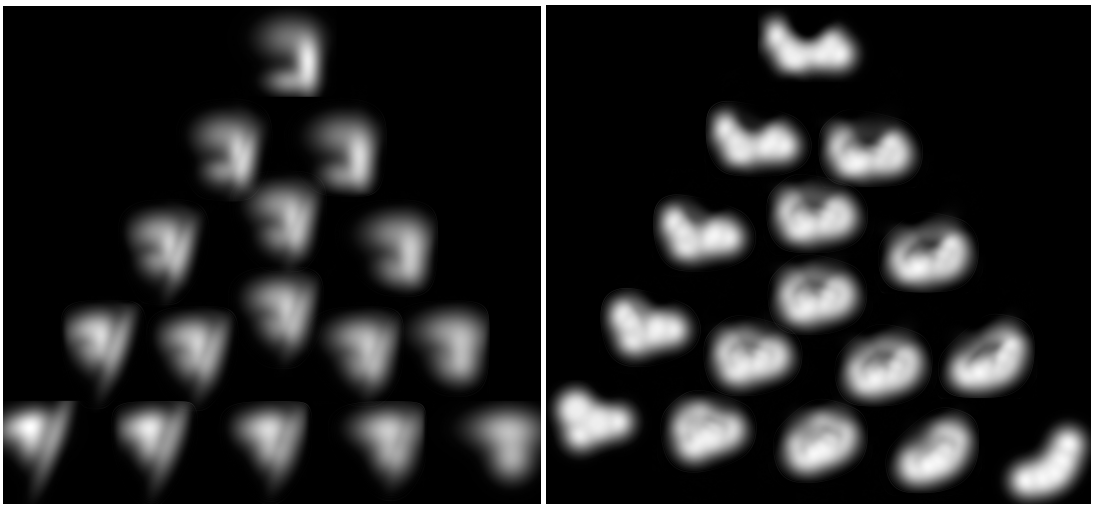

Figure 5: *The sliced Wasserstein barycenter properly advects smooth functions, as shown on this sum of 2D Gaussians (left) and hand drawn non-centered blobs (right).*

4.3 Conclusion

We have shown that a sliced Wasserstein barycenter can be appropriate to interpolate between two smooth probability distributions, at a fraction of the cost of the exact computation, using the Radon transform. It also provides the flexibility to interpolate between more than two probability distributions. Samples drawn from sliced Wasserstein barycenters have already found applications in computer graphics [Rabin et al. 2011], showing the practical importance of this approximation. We have demonstrated that the sliced Wasserstein barycenter of isotropic Gaussians and of translated and scaled distributions are equal to the exact Wasserstein barycenter. Further theoretical guarantees would provide increased robustness in methods that make use the sliced approximation [Pitié et al. 2005, Rabin et al. 2011]. Although we focused on the 2D case, the same principles apply in higher dimension. For instance, the 3D case would amount to projecting onto planes, which would then reduce to the 2D case. Such implementation could directly be handled with projections onto lines, as has been proposed in [Rabin et al. 2011]. Finally, the Radon transform approach can make it difficult to recover the transport plan. We hope that this work will spur further research in efficient approximations of mass transportation and Wasserstein barycenters in high dimension.

References

[Agueh and Carlier 2011] AGUEH, M., AND CARLIER, G. 2011. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis 43*, 2, 904–924.

[Benamou and Brenier 2000] BENAMOU, J.-D., AND BRENIER, Y. 2000. A computational fluid mechanics solution to the monge-kantorovich mass

transfer problem. *Numerische Mathematik 84*, 3, 375–393.

[Bonneel et al. 2011] BONNEEL, N., VAN DE PANNE, M., PARIS, S., AND HEIDRICH, W. 2011. Displacement interpolation using lagrangian mass transport. *ACM Transactions on Graphics (Proceedings of SIGGRAPH ASIA 2011) 30*, 6.

[Deans 2007] DEANS, S. R. 2007. *The Radon transform and some of its applications*. Courier Dover Publications.

[Ferradans et al. 2013] FERRADANS, S., XIA, G.-S., PEYRÉ, G., AND AUJOL, J.-F. 2013. Optimal transport mixing of gaussian texture models. In *Proc. SSVM'13*.

[Haker et al. 2004] HAKER, S., ZHU, L., TANNENBAUM, A., AND ANGENENT, S. 2004. Optimal mass transport for registration and warping. *International Journal of Computer Vision 60*, 3, 225–240.

[Matusik et al. 2005] MATUSIK, W., ZWICKER, M., AND DURAND, F. 2005. Texture design using a simplicial complex of morphable textures. In *ACM Transactions on Graphics (TOG)*, vol. 24, ACM, 787–794.

[McCann 1997] MCCANN, R. J. 1997. A convexity principle for interacting gases. *advances in mathematics 128*, 1, 153–179.

[Mérigot 2011] MÉRIGOT, Q. 2011. A multiscale approach to optimal transport. *Comput. Graph. Forum 30*, 5, 1584–1592.

[Papadakis et al. 2013] PAPADAKIS, N., PEYRÉ, G., AND OUDET, E. 2013. Optimal transport with proximal splitting. Rapport de recherche hal-00816211, HAL.

[Pitié et al. 2005] PITIÉ, F., KOKARAM, A. C., AND DAHYOT, R. 2005. N-dimensional probability density function transfer and its application to color transfer. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2, IEEE, 1434–1439.

[Rabin et al. 2011] RABIN, J., PEYRÉ, G., DELON, J., AND BERNOT, M. 2011. Wassertein barycenter and its applications to texture mixing. In *LNCS, Proc. SSVM'11*, Springer, vol. 6667, 435–446.

[Villani 2003] VILLANI, C. 2003. *Topics in optimal transportation*, vol. 58. American Mathematical Soc.