On Wasserstein Barycenters and MMOSPA Estimation

Marcus Baum, Peter K. Willett, *Fellow, IEEE*, and Uwe D. Hanebeck

***Abstract*—The two title concepts have been evolving rather rapidly, but independent of each other. The Wasserstein barycenter, on one hand, has mostly made its appearance in image processing as it can describe a measure of similarity between images. Its minimization might, for example, suggest the best match in image alignment. On the other hand, MMOSPA estimation has been applied largely to multi-target tracking. The Optimal Sub-Pattern Assignment (OSPA) measures the distance between two sets and the Mean OSPA (MOSPA) can be minimized to give the Minimum MOPSA (MMOSPA), which improves MMSE estimation of the target locations when the labeling of the targets in the set is not important. Approximate and exact algorithms have evolved for both Wasserstein barycenters and MMOSPA estimation. Here, we draw connections between the two perspectives and elaborate how they can benefit from each other.**

***Index Terms*—Barycenter, earth mover's distance, image fusion, multi-target tracking, OSPA distance, Wasserstein distance.**

I. Introduction

THE mean (also called centroid or average) of a collection of objects is a fundamental mathematical concept, e.g., the arithmetic mean of the real numbers 2, 5, and 7 is given by $(2+5+7)/3$. In this letter, we are interested in determining the mean of objects in non-Euclidian space, where the addition operator is undefined. For example, the question is how to define the mean of sets such as $\{1,2\}$ and $\{5,6,10\}$. In this context, two mathematically-related concepts recently independently emerged–Minimum Mean OSPA (MMOSPA) estimation [25] and Wasserstein barycenters [1]. The contribution of this letter is to elaborate the relationship between these two concepts and put them into perspective. A main conclusion is that the Wasserstein barycenter for point clouds is equivalent to the empirical MMOSPA estimate. Finally, we point out by means of a simulation how the so-called sliced Wasserstein distance [33] can be used to facilitate MMOSPA estimation.

M. Baum and U. D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), D-76131 Karlsruhe, Germany (e-mail: marcus.baum@kit.edu; uwe.hanebeck@ieee.org).

P. Willett is with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: willett@engr.uconn.edu).

Both MMOSPA estimation and Wasserstein barycenters find their roots in the so-called Fréchet mean [10], which generalizes the concept of a mean to metric spaces. In analogy to the arithmetic mean, the Fréchet mean is defined as the object that minimizes the sum of squared distances to all objects in the collection. In the following, we introduce the general definition of the Fréchet mean for a random variable.

Definition 1: (Fréchet Mean). Let (S, D) be a metric space, where S is a set and D a metric. The Fréchet mean [10] $\hat{\mathcal{Y}}$ of a random variable $\boldsymbol{\chi}$ that takes values in S, i.e., $\boldsymbol{\chi} \sim \mu$ for a probability measure μ on S, is defined as

$$\hat{\mathcal{Y}} := \arg \inf_{\mathcal{Y} \in S} \mathrm{E}\{D(\boldsymbol{\chi}, \mathcal{Y})^2\}. \tag{1}$$

The so-called empirical Fréchet mean results from (1) by considering an empirical distribution, i.e.,

$$\hat{\mathcal{Y}} := \arg \inf_{\mathcal{Y} \in S} \frac{1}{N} \sum_{l=1}^{N} D(\chi_l, \mathcal{Y})^2, \tag{2}$$

where $\{\boldsymbol{\chi}_l\}_{l=1}^N$ are samples from $\boldsymbol{\chi}$.

Of course, one can use any exponent $p \geq 1$ instead of 2 in (1) and (2), e.g., for $p = 1$ one would get a median. In general, the Fréchet mean is not always unique so that there might be a set of means.

In the remainder of this letter, S consists of probability measures on d-dimensional Euclidean space [11]. Then, χ in (1) can be seen as a "meta" random variable whose realizations are probability measures. Probability measures allow us to describe several important real-world objects, e.g., an intensity image can be seen as a probability density function, and a set of targets such as airplanes can be represented as an empirical distribution (composed of point-masses).

So, as a distance measure on probability measures (and their corresponding probability density functions) in (1), we focus on the so-called Wasserstein distance [26], [38], which is widely used in many areas such as image processing. Intuitively, the Wasserstein distance measures the cost of transporting the mass from one density to the other.

Definition 2: (Wasserstein Distance). For $p > 0$, the p-Wasserstein distance[1] [26], [38] between two probability densities $f : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$W_p(g, f) := \left(\inf_{h \in H(g,f)} \int d(x,y)^p \cdot h(x,y)\, dx\, dy \right)^{1/p}, \tag{3}$$

[1]In order to keep the discussion short, we do not treat the case $p = \infty$ here. Essentially, all findings of this work are also valid for $p = \infty$.

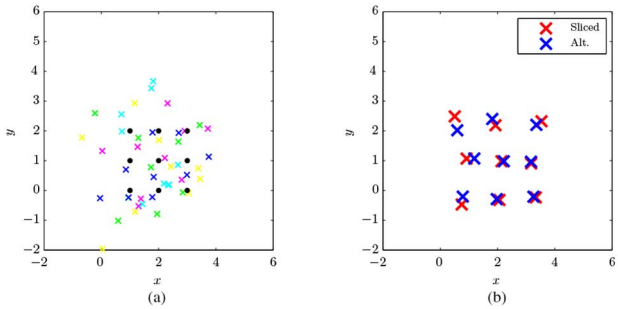

Fig. 1. Illustration of the considered scenario with $m = 9$ (a) Five example sets (in different colors) out of the 200 sets $\{X_l\}_{l=1}^{200}$. The dots represent the grid points used to generate the sets, see (15). (b) MMOSPA estimates resulting from the sliced Wasserstein distance optimization and the alternating optimization.

where $d(\cdot,\cdot)$ is a metric on $\mathbb{R}^d$ and $H(g,f)$ denotes the set of all joint densities $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ with marginals g and f, i.e., $\int h(x,y)\,dy = g(x)$ and $\int h(x,y)\,dx = f(y)$ for $x, y \in \mathbb{R}^d$.

Lemma 1: (Empirical Densities). For two d-dimensional empirical probability density functions $f(x) = \sum_{i=1}^m a_i \delta(x - x_i)$ and $g(y) = \sum_{j=1}^n b_j \delta(y - y_j)$ with weighted samples in which

- $\delta(\cdot)$ denotes the Dirac delta function,
- $x_i \in \mathbb{R}^d$ and $y_j \in \mathbb{R}^d$ are the sample locations, and
- $a_i \in \mathbb{R}$ and $b_j \in \mathbb{R}$ are the sample weights,

the Wasserstein distance becomes [26]

$$W_p(f,g) = \left(\inf_C \sum_{i=1}^m \sum_{j=1}^n C_{ij} d(x_i, y_j)^p \right)^{1/p}, \qquad (4)$$

where C is a matrix with positive entries such that $\sum_{j=1}^n C_{ij} = a_i$ and $\sum_{i=1}^m C_{ij} = b_j$.

In case of $m = n$ and equally weighted samples $f(x) = \frac{1}{m}\sum_{i=1}^m \delta(x - x_i)$ and $g(y) = \frac{1}{m}\sum_{j=1}^m \delta(y - y_j)$, we further obtain [26]

$$W_p(f,g) = \left(\frac{1}{m} \min_{\pi \in \Pi_m} \sum_{i=1}^m d(x_i, y_{\pi(i)})^p \right)^{1/p}, \qquad (5)$$

where Π_m denotes the set of all permutations of $\{1,\ldots,m\}$.

The Wasserstein distance (4) and (5) can be computed by solving a linear programming (LP) problem, for details see [26], [31], [34]. For example, with the Hungarian algorithm as described in [26], the runtime is cubic in the number of samples for solving (5).

Remark 1: The 1-Wasserstein distance is also known as the Earth Mover's Distance (EMD) [31], [34], [39] and the Mallows distance [28].

An empirical probability density with equally weighted samples $f(x) = \frac{1}{m}\sum_{i=1}^m \delta(x - x_i)$ naturally specifies a finite point set $X = \{x_1,\ldots,x_m\} \subset \mathbb{R}^d$. Hence, the Wasserstein distance can be seen as a distance measure for sets and we write $W_p(X,Y)$ to denote the Wasserstein distance between the two sets X and Y.

In the context of multi-target tracking, the Wasserstein distance for finite point sets has been introduced in [26] in order to assess the quality of multi-target trackers. However, it turned out that counterintuitive results for sets with different cardinalities [35] may occur. As a consequence, a new distance measure–the Optimal Sub-Pattern Assignment Metric (OSPA) [35]–has been constructed on the basis of the Wasserstein distance. For sets with the same cardinality, i.e., $m = n$, the OSPA distance coincides with the Wasserstein distance.

Definition 3 (OSPA Distance.). The OSPA [35] distance between two finite point sets $X = \{x_1,\ldots,x_m\} \subset \mathbb{R}^d$ and $Y = \{y_1,\ldots,y_n\} \subset \mathbb{R}^d$ is defined as

$$\mathrm{OSPA}_{p,c}(X,Y) := \left(\frac{1}{n} \min_{\pi \in \Pi_n} \sum_{i=1}^m d_c(x_i, y_{\pi(i)})^p + c^p \cdot (n - m) \right)^{1/p} \qquad (6)$$

if $m \le n$, and $\mathrm{OSPA}_{p,c}(X,Y) := \mathrm{OSPA}_{p,c}(Y,X)$ if $m > n$, where $d_c(x,y) := \min\{c, d(x,y)\}$ with cut-off parameter $c > 0$ and distance metric $d(x,y)$ for $x, y \in \mathbb{R}^d$.

II. MMOSPA Estimation

Multi-target tracking [4] deals with the problem of estimating the number and locations of multiple moving targets such as airplanes based on sensor measurements, e.g., from radar or sonar. From a mathematical point of view, the target states can be captured with a finite point set, where each element represents the state of a target. In order to estimate the target states, i.e., the finite point set, a rigorous theory called Finite Set Statistics (FISST) [29] has been developed, which is built upon an engineer-friendly version of point process theory.

Multi-target tracking is inherently intractable: Due to unknown measurement-to-target associations one has to deal with an exponential number of possible associations. Hence, it becomes necessary to perform sophisticated approximations in order to get efficient algorithms. The key idea of Minimum Mean OSPA (MMOSPA) estimation [25], [36] is to perform these approximations with an eye to the OSPA distance: The final objective is to obtain target estimates that minimize the mean OSPA distance. The development of MMOSPA estimation in [25], [36] has led to a variety of new methods and applications in multi-target tracking, see [2], [3], [7], [8], [15]–[19], [21], [22], [25], [36].

Mathematically, the MMOSPA estimate is the Fréchet mean (1) for the (posterior) random finite set with respect to the OSPA distance. Hence, S consists of finite sets $X = \{x_1,\ldots,x_m\} \subset \mathbb{R}^d$ with $m \ge 0$ and $D(X,Y)$ (with $X, Y \in S$) is the OSPA distance, so that [25]

$$\hat{Y}^{\mathrm{OSPA}} = \arg \inf_{Y \in S} \mathrm{E}\{\mathrm{OSPA}_{p,c}(\boldsymbol{X}, Y)^p\}, \qquad (7)$$

where the set-valued random variable $\boldsymbol{X}$ is distributed according to a random finite set (RFS) density [29]

$$\boldsymbol{X} \sim p(X). \qquad (8)$$

Remark 2: Replacing the OSPA distance in (7) with the Wasserstein distance leads to a Minimum Mean Wasserstein estimator, which is discussed in [19], [26] in a multi-target tracking (not barycenter) context. For the case $m = n$, the MMOSPA estimator coincides with the Minimum Mean Wasserstein estimator as the OSPA and Wasserstein distance are the same.

Based on (7), it is obvious to define an empirical MMOSPA estimate [15], [17], [21], [25] according to

$$\hat{Y}^{\mathrm{OSPA}} := \arg \inf_{Y \in S} \frac{1}{N} \sum_{l=1}^{N} \mathrm{OSPA}_{p,c}(X_l, Y)^p, \qquad (9)$$

where $\{X_l\}_{l=1}^N$ are the sampled sets.

Up to now, work on MMOSPA estimation focused on the case $m = n$ and a variety of computationally tractable methods is available:

- An alternating optimization method is introduced in [25], which can be used for both the empirical MMOSPA (9) and the general MMOSPA (7).
- Exact algorithms for the empirical MMOSPA estimate (9) have been developed in [6], [9]. In general, the time complexity is polynomial in the number of point sets N but exponential in the number of elements m.
- A greedy approximate algorithm for empirical MMOSPA estimates (9) is discussed in [17], [19].
- An approximation for Gaussian mixture RFS (8) based on the Chernov bound is derived in [22].

III. Wasserstein Barycenter

The concept of a Wassertein barycenter [1] recently gained significant interest in image processing, see for example [1], [23], [27], [33], [39]. A Wasserstein barycenter is essentially an empirical Fréchet mean (2) with respect to the Wasserstein distance for probability densities. For the probability densities $\{f_l\}_{l=1}^N$, the Wasserstein barycenter [1], [23], [33], [39] is given by

$$\hat{g} := \arg \inf_{g \in S} \frac{1}{N} \sum_{l=1}^{N} W_p(f_l, g)^p. \qquad (10)$$

Depending on the specific form of $\{f_l\}_{l=1}^N$ and g, one can distinguish the following special cases (see also the discussion in [23]):

A. Empirical Densities $\{f_l\}_{l=1}^N$ and g: Sample Weights of g are Optimized; Sample Locations of g are Fixed.

This special case is interesting for image processing as an intensity image can be interpreted as an empirical distribution for which the sample locations represent the pixel locations and the weights specify the pixel intensities [23]. In this manner, the Wasserstein barycenter calculates a mean image. In [23], it is shown that the cost function in (10) is convex and a projected subgradient method for optimization is derived.

B. Empirical Densities $\{f_l\}_{l=1}^N$ and g: Sample Weights and Sample Locations of g are Optimized (Maximum Number of Samples is Restricted to be Less Than k).

For the general case of empirical probability densities, [23] derived an optimization method, which alternates between optimizing the weights and the locations. For $N = 1$, this case is equivalent to the well-known k-means clustering problem. If $N = 1$ and uniform sample weights of g are enforced additionally, a constrained k-means problem is obtained as discussed in [23].

C. Empirical Densities $\{f_l\}_{l=1}^N$ and g with Equally Weighted Samples: Sample Locations of g are Optimized.

An empirical density with equally weighted samples represents a point set (also called point cloud) as discussed earlier. Hence, in this case, the Wasserstein barycenter of a collection of point sets $\{X_l\}_{l=1}^N$ is the point set $\hat{Y}^{\mathrm{W}}$ that minimizes the sum of Wasserstein distances to all point sets in the collection, i.e.,

$$\hat{Y}^{\mathrm{W}} := \arg \inf_{Y \in S} \frac{1}{N} \sum_{l=1}^{N} \mathrm{W}_p(X_l, Y)^p. \qquad (11)$$

In order to directly minimize (11) the alternating optimization methods developed in [23] can be used.

In the context of image processing, Wasserstein barycenters for point clouds were introduced in [33] for texture mixing. As the cubic time-complexity for evaluating the Wasserstein distance might be too high for image processing applications, [33] introduced the so-called sliced Wasserstein distance, which exploits that the Wasserstein distance can be computed efficiently in one-dimensional space, i.e.,

$$W_p^{\mathrm{sliced}}(X, Y) := \left(\int_{\theta \in \Theta^d} W_p(X_\theta, Y_\theta)^p d\theta \right)^{1/p}, \qquad (12)$$

where Θ^d is the d-dimensional unit sphere and $X_\theta = \{x_i^T \theta\}_{i=1}^n$ results from projecting all elements of X onto the unit vector θ. In order to compute barycenters based on (12), [33] introduced a stochastic gradient descent algorithm.

Actually, this special case reveals a relationship that has not been mentioned in literature yet: According to the discussion in the previous section about MMOSPA estimation, the Wasserstein barycenter for point clouds (11) is equivalent to the empirical MMOSPA estimate (9). As a consequence, approximate and exact MMOSPA algorithms can be used for Wasserstein barycenters of point clouds (and vice versa). Actually, the alternating algorithm for the empirical MMOSPA estimate [25] can bee seen as a special case of the methods presented in [23].

Recently, further interesting applications of the point cloud case have evolved: In [12], Wasserstein barycenters for point clouds are used for fusing empirical probability densities (in this case $N = 2$). In applied topology, the statistical analysis of persistence diagrams can be performed based on Wasserstein barycenters for point clouds (plus copies of diagonal elements) [30], [37]. In this context, [37] investigates an alternating optimization method similar to [25]. Another similar alternating approach has been proposed in [24] in order to compute median point sets under affine transformations.

D. General Case: Arbitrary $\{f_l\}_{l=1}^N$ and g.

Approximate algorithms for the general case that are based on discretization are discussed in [13]. Note that if $\{f_l\}_{l=1}^N$ are (zero-mean) Gaussian densities, it can be shown that the Wasserstein barycenter is also Gaussian and given by the root of a matrix equation [1].

IV. Alternative Barycenters for Point Sets

The evaluation of the Wasserstein distance has a rather high runtime complexity [26]. Due to the optimization over permutations in (6), it is also difficult to find closed-form solutions for the expectation (7). In this context, we discussed a modification called sliced Wasserstein distance [33] in Section III-C, which can be evaluated more efficiently. Next, we mention two further computationally attractive distance criteria that have been used to define barycenters for point sets, i.e., the kernel distance [32] and a probabilistic similarity [14].

The *kernel distance* [32] between two point sets $X = \{x_1, \dots, x_m\}$ and $Y = \{y_1, \dots, y_n\}$ is defined as

$$\begin{aligned}\mathrm{KD}(X,Y) :&= \sum_{i_1=1}^{m}\sum_{i_2=1}^{m} k(x_{i_1}, x_{i_2}) \\ &+ \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} k(y_{j_1}, y_{j_2}) - 2\sum_{i=1}^{m}\sum_{j=1}^{n} k(x_i, y_j),\end{aligned} \tag{13}$$

where $k(\cdot,\cdot)$ is a suitable kernel function, e.g., Gaussian. Fréchet means and barycenters based on the kernel distance were investigated in a multi-target tracking context in [5]. An advantage is that closed-form expressions for (7) are available for Gaussian mixture RFS densities (8).

In the context of point set registration, [14] defines a similarity measure for two sets X and Y by considering the probability that Y is a noisy observation of X. For this purpose, a soft data association model is used, which is also known in multi-target tracking [20], in order to define

$$\mathrm{probsim}(X,Y) := \prod_{i=1}^{m}\sum_{j=1}^{n} \mathcal{N}(x_i - y_j, \Sigma), \tag{14}$$

where Σ is the covariance of the noise. Barycenters for this *probabilistic similarity* are defined in [14] and also an alternating optimization procedure based on the Expectation-Maximization (EM) algorithm is derived.

V. Experiment

In this section, we demonstrate how a multi-target estimation scenario can benefit from algorithms for Wasserstein barycenters. For this purpose, we consider the MMOSPA estimate for a random finite set (RFS) in (7) in two different settings. In the first setting, there are six targets in two-dimensional space, i.e., $m = n = 6$ and $d = 2$. The RFS (8) is of the form

$$p(X = \{x_1, \dots, x_6\}) := \sum_{\pi\in\Pi_6}\prod_{i=1}^{6} \mathcal{N}(x_i - \mu_{\pi(i)}, \Sigma), \tag{15}$$

and $p(X) := 0$ otherwise. The means of the Gaussians $\{\mu_i\}_{i=1}^{m}$ in (15) are arranged in a regular 2×3 grid with spacing 1, and $\Sigma = diag([1,1])$ is given. We randomly draw $N = 200$ samples from the Gaussian mixture RFS (15). Each sample of (15) is a set containing $m = 6$ elements. Based on these samples, the objective is to calculate the empirical MMOSPA (9). The second setting coincides exactly with the first setting except that there are $m = 9$ targets and the $\{\mu_i\}_{i=1}^{m}$ are arranged in a regular 3×3 grid. The above scenario might occur in a group tracking scenario, to track a large number of closely-spaced targets.

TABLE I
Evaluation Results: Optimizing the Sliced Wasserstein Distance vs. Alternating Optimization of the Wasserstein/OSPA Distance. Comparison is Done with Respect to the Runtime (in Seconds) and the Mean OSPA (MOSPA) Distance of the Result to all Sets $\{X_l\}_{l=1}^{200}$. The Experiments have been Performed on a Standard Desktop Computer with Non-Optimized MATLAB Implementations. (a) $m = 6$ (b) $m = 9$.

	Sliced	**Alt.**
Time (s)	3.0224	5.6313
MOSPA	1.2371	1.1813

(a)

	Sliced	**Alt.**
Time (s)	4.5314	53.9180
MOSPA	0.9952	0.9600

(b)

We use two different methods for minimizing (9): The gradient descent algorithm for the sliced Wasserstein distance [33] (with 20 directions, see (12)) and the alternating optimization algorithm from [25]. For the sake of intuition, Fig. 1 shows example sets and the resulting MMOSPA estimates, i.e., barycenters, for the second setting. Both algorithms have been initialized with the grid $\{\mu_i\}_{i=1}^{m}$ and iterations were stopped when the change in the (sliced) Wasserstein distance was below 0.001. Table I depicts a comparison of both algorithms with respect to runtime and quality. While both algorithms seem to have the same performance for the six target scenario, the situation changes for nine targets: the sliced Wasserstein distance algorithm is significantly faster. The reason is that the alternating optimization algorithm computes, at each iteration, the optimal permutations for all sets $\{X_l\}_{l=1}^{200}$, which has a cubic runtime in m. The sliced Wasserstein distance avoids this computational complexity, but comes with a loss of optimality. All told, the sliced Wasserstein distance might be suitable for large-scale tracking problems that are not yet tractable with current MMOSPA estimation techniques.

VI. Discussion and Conclusions

This letter is about two research areas that have recently gained increasing interest–Wasserstein barycenters and MMOSPA estimators. To the authors' knowledge, this is the first work that clarifies the relationship between these two areas. We have derived both Wasserstein barycenters and MMOSPA estimators from fundamental mathematical concepts, i.e., the Fréchet mean and the Wasserstein distance, and pointed out recent algorithms, variations, applications, and further developments. In particular, we revealed that the empirical MMOSPA estimate is equivalent to the Wasserstein barycenter for point clouds, which both have been recently independently discussed in [15], [17], [18], [21], [25] and [33]. The two research areas might benefit from the discovered equivalence, e.g., algorithms and insights for solving the underlying optimization problem can be shared. An example of the use of an algorithm from the barycenter community (sliced Wasserstein distance) in the MMOSPA context is in Section V. Ideas from the MMOSPA literature that might be of use to the barycenter community include the exact algorithms for the empirical MMOSPA estimate [6], [9] and the iterative algorithm [25], although the significance of the latter is perhaps already well appreciated [23].

REFERENCES

[1] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," *SIAM J. Math. Anal.*, vol. 43, no. 2, pp. 904–924, 2011.

[2] E. H. Aoki, Y. Boers, L. Svensson, P. Mandal, and A. Bagchi, "A Bayesian look at the optimal track labelling problem," in *Data Fusion Target Tracking Conf. (DF TT 2012): Algorithms Applications, 9th IET*, 2012, pp. 1–6.

[3] E. Aoki, Y. Boers, L. Svensson, P. Mandal, and A. Bagchi, "The Rao-Blackwellized marginal m-smc filter for Bayesian multi-target tracking and labelling," in *15th Int. Conf. Information Fusion (FUSION)*, 2012, pp. 90–97.

[4] Y. Bar-Shalom, P. K. Willett, and X. Tian*, Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT, USA: YBS, 2011.

[5] M. Baum, P. Ruoff, D. Itte, and U. D. Hanebeck, "Optimal point estimates for multi-target states based on kernel distances," in *Proc. 51st IEEE Conf. Decision and Control (CDC 2012)*, Maui, HI, USA, Dec. 2012.

[6] M. Baum, P. Willett, and U. D. Hanebeck, "Calculating some exact MMOSPA estimates for particle distributions," in *Proc. 15th Int. Conf. Information Fusion (Fusion 2012)*, Singapore, Jul. 2012.

[7] M. Baum, P. Willett, and U. D. Hanebeck, "MMOSPA-based direction-of-arrival estimation for planar antenna arrays," in *Proc. Eighth IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2014)*, A Coruña, Spain, Jun. 2014.

[8] M. Baum, P. Willett, and U. D. Hanebeck, "MMOSPA-based track extraction in the PHD filter–a justification for k-means clustering," in *Proc. 53rd IEEE Conf. Decision and Control (CDC 2014)*, Los Angeles, CA, USA, Dec. 2014.

[9] M. Baum, P. Willett, and U. D. Hanebeck, "Polynomial-time algorithms for the exact MMOSPA estimate of a multi-object probability density represented by particles," *IEEE Trans. Signal Process.*, 2015.

[10] R. Bhattacharya and V. Patrangenaru, "Large sample theory of intrinsic and extrinsic sample means on manifolds," *Ann. Statist.*, vol. 31, no. 1, pp. 1–29, Feb. 2003.

[11] J. Bigot and T. Klein, "Consistent estimation of a population barycenter in the Wasserstein space," in *arXiv:1212.2562* , Dec. 2012.

[12] A. Bishop, "Information fusion via the Wasserstein barycenter in the space of probability measures: Direct fusion of empirical measures and gaussian fusion with unknown correlation," in *17th Int. Conf. Information Fusion*, Salamanca, Spain, 2014.

[13] G. Carlier, A. Oberman, and E. Oudet, "Numerical methods for matching for teams and Wasserstein barycenters," McGill Univ., Montreal, QC, Canada, Tech. Rep., May 2014 [Online]. Available: arXiv:1411.3602

[14] B. Combes, M. Fournier, D. Kennedy, J. Braga, N. Roberts, and S. Prima, "EM-ICP strategies for joint mean shape and correspondences estimation: Applications to statistical analysis of shape and of asymmetry," in *2011 IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, Mar. 2011, pp. 1257–1263.

[15] D. F. Crouse, P. Willett, and Y. Bar-Shalom, "Developing a real-time track display that operators do not hate," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3441–3447, Jul. 2011.

[16] D. F. Crouse, P. Willett, and Y. Bar-Shalom, "Generalizations of Blom and Bloem's PDF decomposition for permutation-invariant estimation," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 3840–3843.

[17] D. F. Crouse, P. Willett, Y. Bar-Shalom, and L. Svensson, "Aspects of MMOSPA estimation," in *Proc. the 50th IEEE Conf. Decision and Control and Eur. Control Conf.*, Orlando, FL, Dec. 2011.

[18] D. F. Crouse, P. Willett, L. Svensson, and Y. Bar-Shalom, "Comparison of compressed sensing, ML, and MMOSPA estimation for radar superresolution," in *Proc. 45th Asilomar Conf. Signals, Systems and Computers*, 2011.

[19] D. F. Crouse, "Advances in displaying uncertain estimates of multiple targets," *SPIE–Signal Processing, Sensor Fusion, and Target Recognition XXII*, vol. 8745, pp. 874504–874504–31, 2013.

[20] D. F. Crouse, M. Guerriero, P. Willett, R. Streit, and D. Dunham, "A look at the PMHT," in *Proc. 12th Int. Conf. Information Fusion (Fusion 2009)*, Jul. 2009.

[21] D. F. Crouse, P. Willett, L. Svensson, D. Svensson, and M. Guerriero, "The set MHT," in *Proc. 14th Int. Conf. Information Fusion (Fusion 2011)*, Jul. 2011.

[22] D. F. Crouse, P. Willett, M. Guerriero, and L. Svensson, "An approximate minimum MOSPA estimator," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 3644–3647.

[23] M. Cuturi and A. Doucet, "Fast computation of Wasserstein barycenters," in *Proc. 31st Int. Conf. Machine Learning (ICML-14)*, Oct. 2014.

[24] H. Ding and J. Xu, "Finding median point-set using earth movers distance," in *Proc. 28th AAAI Conf. Artificial Intelligence (AAAI 2014)*, 2014.

[25] M. Guerriero, L. Svensson, D. Svensson, and P. Willett, "Shooting two birds with two bullets: How to find minimum mean OSPA estimates," in *Proc. 13th Int. Conf. Information Fusion (Fusion 2010)*, 2010.

[26] J. Hoffman and R. Mahler, "Multitarget miss distance and its applications," in *Proc. Fifth Int. Conf. Information Fusion*, Jul. 2002, vol. 1, pp. 149–155, vol.1.

[27] S. Justin, R. Rustamov, L. Guibas, and A. Butscher, "Earth mover's distances on discrete surfaces," *ACM Trans. Graphics (SIGGRAPH)*, 2014.

[28] E. Levina and P. Bickel, "The earth mover's distance is the Mallows distance: Some insights from statistics," in *Eighth IEEE Int. Conf. Computer Vision*, 2001, vol. 2, pp. 251–256, vol.2.

[29] R. Mahler, ""Statistics 101" for multisensor, multitarget data fusion," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 53–64, Jan. 2004.

[30] Y. Mileyko, S. Mukherjee, and J. Harer, "Probability measures on the space of persistence diagrams," *Inv. Probl.*, vol. 27, no. 12, p. 124007, 2011.

[31] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *IEEE 12th Int. Conf. Computer Vision (ICCV)*, 2009.

[32] J. M. Phillips and S. Venkatasubramanian, "A gentle introduction to the kernel distance," *CoRR*, vol. abs/1103.1625, 2011.

[33] J. Rabin, G. Peyr, J. Delon, and M. Bernot, "Wasserstein barycenter and its application to texture mixing," in *Scale Space and Variational Methods in Computer Vision*, ser. Lecture Notes in Computer Science. Heidelberg, Germany: Springer Berlin, 2012, vol. 6667, pp. 435–446.

[34] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000.

[35] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, 2008.

[36] L. Svensson, D. Svensson, M. Guerriero, and P. Willett, "Set JPDA filter for multi-target tracking," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4677–4691, Oct. 2011.

[37] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer, "Fréchet means for distributions of persistence diagrams," in *arXiv:1206.2790* , Jun. 2012.

[38] C. Villani*, Optimal Transport*. Berlin, Germany: Springer Verlag, 2009.

[39] G. Zen and E. Ricci, "Earth mover's prototypes: A convex learning approach for discovering activity patterns in dynamic scenes," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2011, pp. 3225–3232.