Information Fusion via the Wasserstein Barycenter in the Space of Probability Measures: Direct Fusion of Empirical Measures and Gaussian Fusion with Unknown Correlation

Adrian N. Bishop
Australian National University and NICTA
Email: adrian.bishop@anu.edu.au

***Abstract*—In this work, a general information fusion problem is formulated as an optimisation protocol in the space of probability measures (i.e. the so-called Wasserstein metric space). The high-level idea is to consider the data fusion result as the probability measure that is closest to a given collection of input measures in the sense that it will minimise the (weighted) Wasserstein distance between itself and the inputs. After formulating the general information fusion protocol, we consider the explicit computation of the fusion result for two special scenarios that occur frequently in practical applications. Firstly, we show how one can compute the general outcome explicitly with two Gaussian input measures (ignoring any correlation). We then examine the consistency of this result for the scenario in which the two Gaussian inputs have an unknown (but possibly non-zero) correlation. Secondly, we show how one can compute the general fusion result explicitly given two randomly sampled (discrete) empirical measures which typically have no common underlying support. Data fusion with empirical measures as input has wide applicability in applications involving Monte Carlo estimation etc.**

I. Introduction

In this work, a general information fusion problem is formulated as an optimisation protocol in the space of probability measures (i.e. the so-called Wasserstein metric space [1]). The high-level idea is to consider the data fusion outcome as the probability measure that is closest to a given collection of input measures in the sense that it will minimise the (weighted) Wasserstein distance between itself and the inputs.

The classical way to combine continuous conditional measures is to use Bayes rule, which (roughly) involves multiplying the measures together and then normalizing via an integral operation. Alternatively, one may find the weighted average of all the individual probability measures (i.e. sum the weighted measures (with the weights summing to 1) and take this as the combined belief); i.e. this is just a probability mixture (often referred to a linear opinion pool). Separate again, one may take the weighted average of the logarithm of the individual measures and then take the common belief to be the exponentiation of this weighted average (often referred to as a log-linear opinion pool). See [2]–[5] for a survey of these three well-studied solutions and variations. Other variations are discussed in [6], [7].

Other non-probabilistic (or generalised) approaches to information fusion and inference such as interval calculus, fuzzy logic, Dempster-Shafer etc. [6], [7] are not discussed here. Probability measures as arising in topics like random-set theory etc. [6] are also not discussed but presumably can be accommodated via the proposed framework for information fusion.

One novelty of the proposed general information fusion protocol is that it draws on a very general formulation in which input measures of a general type can be naturally accommodated; e.g. we will show how one can deal straightforwardly with empirical (discrete, randomly sampled) measures. Moreover, the formulation introduced appears to be a mathematically intuitive approach to information fusion drawing on the rigorous foundation of the Wasserstein metric [1]. The general algorithm introduced here for information fusion also lends itself neatly to distributed computation as outlined in separate work [8].

A. Direct Information Fusion of Empirical Measures

Note that one downside of existing probabilistic methods [2]–[5] for combining probability measures is that they do not naturally allow one to consider the direct combination of randomly sampled (discrete) measures [9] and/or non-standard probability measures. This limitation follows because these methods inherently act on measures as if they are 'functions' with common support. This assumption of a common support is extremely limiting. For example, the multiplication required by Bayes rule is simply impossible to do (directly) when dealing with empirical measures.

Traditionally, fusion of empirical measures typically involves temporarily transforming the empirical measure back to a continuous measure using a so-called Kernel method [10] at which point classical fusion results [2], [6], [7], [11] apply. The Kernel-based approaches typically scale poorly (without some form of clustering of the components) and may also perform poorly when the underlying sampled measures do not 'overlap' sufficiently well. Its worth noting the computationally efficient

A.N. Bishop is supported by NICTA and the ARC via a Discovery Early Career Researcher Award (DE-120102873). He is also supported by the US Air Force via AOARD and contract USAF-AOARD-144042.

approximations for sampling the output of the product of input measures, e.g. typically two Gaussian mixture inputs, which do not require an explicit computation of the product itself [12], [13]. Such methods naturally approximate the Kernel method approach to fusion with empirical inputs.

An advantage of the proposed method for information fusion is that it completely avoids the multiplication or summation of measures, a task that is impossible to do when dealing with randomly sampled measures. Instead, we specialise the general information fusion protocol proposed in this work and provide a directly computable protocol for the direct information fusion of two empirical (discrete) measures. This protocol works by finding a discrete measure that is closest in the sense of the Wasserstein metric [14] to the two empirical input measures. The real novelty of this presentation lies in the potential applications of such fusion in fields like Monte Carlo estimation [9] etc.

B. Gaussian Information Fusion in the Presence of Unknown Correlation

In classical Gaussian information fusion the optimal solution (in terms of minimising the variance of the fusion result) is straightforwardly computed when the correlation between the input measures is known [7]. In the case in which the correlation is unknown one may employ an algorithm such as covariance intersection (CI) [15], [16] which ensures the output is consistent in the sense that its variance estimate is never less than the actual variance that would arise if one knew the correlation (i.e. the fusion result is never over-confident but rather typically conservative with respect to the variance). The log-linear opinion pool [5] described previously is a generalisation of covariance intersection to arbitrary input measures. Other approaches to Guassian information fusion exist that consider the case in which the correlation between the input measures is unknown [17]–[21].

We specialise the general information fusion protocol proposed in this work and provide a directly computable protocol for the information fusion of two Gaussian input measures. This result produces a Gaussian fusion output that is the closest Gaussian measure in the (weighted) Wasserstein metric to the two input measures. The computation of this output does not take into account any dependence between the inputs and any such dependence, if it exists, does not appear in the computation. We then study the consistency of this fusion result, in the spirit of covariance intersection [15], when the inputs may or may not be independent.

C. Organisation

In the next section a general information fusion problem is formulated as an optimisation problem in the space of probability measures (i.e. the so-called Wasserstein metric space [1]). The high-level idea is to consider the data fusion outcome as the probability measure that is closest to the given collection of input measures in the sense that it will minimise the (weighted) Wasserstein distance between itself and the collection of inputs. In the subsequent two sections we consider the explicit computation of the data fusion result for two special scenarios that occur frequently in practical applications. Firstly, in Section III the general information fusion solution is computed explicitly for two Gaussian input measures (ignoring any correlation) and then the consistency of this result is examined for the scenario in which the two Gaussian inputs have an unknown (but possibly non-zero) correlation. Then in Section IV the general fusion problem is reduced to one involving two randomly sampled (discrete) empirical measures which typically have no common underlying support. A conclusion is given in Section V.

II. Information Fusion via the Wasserstein Barrycenter

Consider a collection, $i \in \mathcal{V} = \{1, \ldots, n\}$ of Radon probability measures μ_i defined on the Borel sets of $(\mathbb{R}^m, d)$ with $0 < m < \infty$ where $d : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is the usual Euclidean distance. Define the space of all such measures on $(\mathbb{R}^m, d)$ by $\mathfrak{U}(\mathbb{R}^m)$ and the subset of all such measures with bounded, finite, p^{th} moment by $\mathfrak{U}_p(\mathbb{R}^m)$ for some suitably small $1 \leq p < \infty$. That is, $\mathfrak{U}_p$ is the collection of probability measures on the Borel sets of $(\mathbb{R}^m, d)$ such that

$$\int_{\mathbb{R}^m} d(\mathbf{x}, \mathbf{x}_0)^p \, \mathrm{d}\mu_i(\mathbf{x}) < \infty \tag{1}$$

for all bounded $\mathbf{x} \in \mathbb{R}^m$ and a given $\mathbf{x}_0 \in \mathbb{R}^m$.

One can then associate with the space $\mathfrak{U}_p$ a metric $\ell_p : \mathfrak{U}_p \times \mathfrak{U}_p \to \mathbb{R}$ defined by

$$\ell_p(\mu_i, \mu_j)^p = \inf_{\gamma \in \Gamma(\mu_i, \mu_j)} \int_{\mathbb{R}^m \times \mathbb{R}^m} d(\mathbf{x}_i, \mathbf{x}_j)^p \, \mathrm{d}\gamma(\mathbf{x}_i, \mathbf{x}_j) \tag{2}$$

where $\Gamma(\mu_i, \mu_j)$ denotes the collection of all measures on $\mathbb{R}^m \times \mathbb{R}^m$ with marginals μ_i and μ_j on the first and second factors respectively.

Suppose one wants to compute

$$\nu = \inf_{z \in \mathfrak{U}_p} \sum_{i \in \mathcal{V}} w_i \ell_p(z, \mu_i)^p \tag{3}$$

where $w_i \in (0, 1)$ and $\sum_{j \in \mathcal{V}} w_j = 1$. We neglect the trivial case in which $w_i(t) = 1$ for one i.

This operation is a form of information fusion in the sense that we are trying to find a measure that is the 'closest' measure to a collection of given input measures (in this case in the sense of the weighted Wasserstein distance). This formulation lends itself naturally to distributed implementation and the distributed fusion version of this algorithm along with convergence results are introduced in [8].

The Wasserstein metric captures the error in the expected value of a class of functions due to the approximation of one measure by another [22]. Thus, the fusion result can be viewed as a measure with an expected value (for a class of function) that is minimally different (simultaneously) from the same expectation of each input measure (in a weighted sense).

In the remainder, we consider the computation of (3) for two important (application heavy) special cases.

III. General Linear Information Fusion of Two Gaussian Estimates

We consider two (random variable) estimates $\mathbf{a} \sim \mathcal{N}(\mathbf{c}^*, \tilde{\mathbf{P}}_{aa})$ and $\mathbf{b} \sim \mathcal{N}(\mathbf{c}^*, \tilde{\mathbf{P}}_{bb})$ of some fixed parameter $\mathbf{c}^* \in \mathbb{R}^m$, $0 < m < \infty$. The estimation error of $\mathbf{a}$ and $\mathbf{b}$ are defined by the random variables

$$\tilde{\mathbf{a}} = \mathbf{a} - \mathbf{c}^* \ , \ \tilde{\mathbf{b}} = \mathbf{b} - \mathbf{c}^* \tag{4}$$

where, in this case,

$$\mathrm{E}[\tilde{\mathbf{a}}] = 0 \ , \ \tilde{\mathbf{P}}_{aa} = \mathrm{E}[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^\top] > \mathbf{0} \tag{5}$$

$$\mathrm{E}[\tilde{\mathbf{b}}] = 0 \ , \ \tilde{\mathbf{P}}_{bb} = \mathrm{E}[\tilde{\mathbf{b}}\tilde{\mathbf{b}}^\top] > \mathbf{0} \tag{6}$$

Although the true values $\tilde{\mathbf{P}}_{aa}$ and $\tilde{\mathbf{P}}_{bb}$ may not be known, consistent approximations $\mathbf{P}_{aa}$ and $\mathbf{P}_{bb}$ are assumed available where [1]

$$\mathbf{P}_{aa} \geq \tilde{\mathbf{P}}_{aa} \ , \ \mathbf{P}_{bb} \geq \tilde{\mathbf{P}}_{bb} \tag{7}$$

The cross-correlation matrix between the two estimates is denoted by $\tilde{\mathbf{P}}_{ab}$ and is defined by

$$\tilde{\mathbf{P}}_{ab} = \mathrm{E}[(\mathbf{a} - \mathbf{c}^*)(\mathbf{b} - \mathbf{c}^*)^\top] = \mathrm{E}[\tilde{\mathbf{a}}\tilde{\mathbf{b}}^\top] \tag{8}$$

This matrix may be known or unknown and may even be zero in some applications.

Let $\mathbf{c} \sim \mathcal{N}(\mathbf{c}^*, \tilde{\mathbf{P}}_{cc})$ denote a third estimate of $\mathbf{c}^*$ obtained via a linear combination of $\mathbf{a}$ and $\mathbf{b}$. That is

$$\mathbf{c} = \mathbf{K}_1\mathbf{a} + \mathbf{K}_2\mathbf{b} \tag{9}$$

where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and $\mathbf{K}_1, \mathbf{K}_2 \in \mathbb{R}^{n \times n}$. The error in this estimate is

$$\tilde{\mathbf{c}} = \mathbf{c} - \mathbf{c}^* \tag{10}$$

and obeys $\mathrm{E}[\tilde{\mathbf{c}}] = 0$ when $\mathbf{K}_1 + \mathbf{K}_2 = \mathbf{I}$.

The true covariance $\tilde{\mathbf{P}}_{cc} = \mathrm{E}[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^\top]$ is calculated by

$$\begin{aligned}\tilde{\mathbf{P}}_{cc} = &\mathbf{K}_1\tilde{\mathbf{P}}_{aa}\mathbf{K}_1^\top + \mathbf{K}_2\tilde{\mathbf{P}}_{bb}\mathbf{K}_2^\top \\ &+\mathbf{K}_1\tilde{\mathbf{P}}_{ab}\mathbf{K}_2^\top + \mathbf{K}_2\tilde{\mathbf{P}}_{ba}\mathbf{K}_1^\top\end{aligned} \tag{11}$$

and calculation of this term requires $\tilde{\mathbf{P}}_{ab} = \tilde{\mathbf{P}}_{ba}^\top$ be known (when it is non-zero).

We are mainly interested in the construction of an estimator $\mathbf{c}$ defined by some $\mathbf{K}_1$ and $\mathbf{K}_2$ and also an estimate $\mathbf{P}_{cc}$ of $\tilde{\mathbf{P}}_{cc}$ when the cross-correlation $\tilde{\mathbf{P}}_{ab}$ is non-zero but unknown. We are further interested in certain properties of the resulting $\mathbf{P}_{cc}$. In particular, we are interested in the property of consistency

$$\mathbf{P}_{cc} \geq \tilde{\mathbf{P}}_{cc} \tag{12}$$

where $\tilde{\mathbf{P}}_{cc}$ is given by (11) which depends on the cross-correlation $\tilde{\mathbf{P}}_{ab}$ or some estimation thereof which we assume unavailable.

Definition 1. *Suppose $\tilde{\mathbf{P}}_{aa}$ and $\tilde{\mathbf{P}}_{bb}$ are given along with $\tilde{\mathbf{P}}_{ab} = \tilde{\mathbf{P}}_{ba}^\top$. Suppose $\tilde{\mathbf{P}}_{ab} = \tilde{\mathbf{P}}_{ba}^\top$ is non-zero. Suppose an estimator $\mathbf{c}$ for $\mathbf{c}^*$ is given in the form (9). The true covariance of $\mathbf{c}$ is denoted by $\tilde{\mathbf{P}}_{cc}$. An estimate $\mathbf{P}_{\mathbf{cc}}$ of $\tilde{\mathbf{P}}_{cc}$ is said to be consistent if*

$$\mathbf{P}_{cc} \geq \tilde{\mathbf{P}}_{cc} \tag{13}$$

where $\tilde{\mathbf{P}}_{cc}$ is defined by (11).

It is often true that ignoring the correlation $\tilde{\mathbf{P}}_{ab}$ when fusing $\mathbf{a}$ and $\mathbf{b}$ can lead to overly confident results; i.e. the resulting estimate of $\mathbf{P}_{cc}$ will be inconsistent as per Definition 1. Some algorithms, such as covariance intersection (CI), see [15], [16], on the other hand are designed to generate consistent estimates when the cross-correlation is unknown. In many cases, the resulting estimators are considerably conservative.

A. Information Fusion in the Wasserstein Space of Gaussian Probability Measures

Consider two Gaussian probability measures μ_i, $i \in \{a, b\}$ defined on the Borel sets of $(\mathbb{R}^m, d)$ and suppose that μ_a and μ_b admit Gaussian densities of the form $\mathcal{N}(\mathbf{c}^*, \tilde{\mathbf{P}}_{aa})$ and $\mathcal{N}(\mathbf{c}^*, \tilde{\mathbf{P}}_{bb})$ given some fixed parameter $\mathbf{c}^*$. Define the space of all such measures on $(\mathbb{R}^m, d)$ with bounded second moment by $\mathfrak{G}_2(\mathbb{R}^m) \subset \mathfrak{U}_2(\mathbb{R}^m)$.

Suppose one wants to compute

$$\mu_c \quad = \quad \inf_{z \in \mathfrak{G}_p} w_1\ell_p(z, \mu_a)^p + w_2\ell_p(z, \mu_b)^p \tag{14}$$

where $w_i \in (0, 1)$ and $(w_1 + w_2) = 1$. This operation is clearly a special case of (3) where $p = 2$, $|\mathcal{V}| = 2$ and we restrict ourselves to the space $\mathfrak{G}_2(\mathbb{R}^m) \subset \mathfrak{U}_2(\mathbb{R}^m)$. Note that this operation does not consider any dependence between the inputs.

Theorem 1 ([14], [23])**.** *Suppose that μ_a and μ_b admit Gaussian probability densities of the form $\mathcal{N}(\mathbf{c}^*, \tilde{\mathbf{P}}_{aa})$ and $\mathcal{N}(\mathbf{c}^*, \tilde{\mathbf{P}}_{bb})$ given some fixed parameter $\mathbf{c}^*$. Then μ_c defined as the solution to*

$$\mu_c = \inf_{z \in \mathfrak{G}_2} (1 - w)\ell_2(z, \mu_a)^2 + w\ell_2(z, \mu_b)^2 \tag{15}$$

with $w \in (0, 1)$ exists and is unique. Moreover, μ_c is in $\mathfrak{G}_2$ and admits a Gaussian density of the form $\mathcal{N}(\widehat{\mathbf{c}}, \widehat{\mathbf{P}}_{cc})$ where

$$\widehat{\mathbf{c}} = (1 - w)\mathbf{c}^* + w\mathbf{c}^* = \mathbf{c}^* \tag{16}$$

$$\widehat{\mathbf{P}}_{cc} = \left((1 - w)\mathbf{I} + w\widehat{\mathbf{\Pi}}\right)\tilde{\mathbf{P}}_{aa}\left((1 - w)\mathbf{I} + w\widehat{\mathbf{\Pi}}\right) \tag{17}$$

where $\widehat{\mathbf{\Pi}} = \tilde{\mathbf{P}}_{bb}^{1/2}(\tilde{\mathbf{P}}_{bb}^{1/2}\tilde{\mathbf{P}}_{aa}\tilde{\mathbf{P}}_{bb}^{1/2})^{-\frac{1}{2}}\tilde{\mathbf{P}}_{bb}^{1/2}$.

We now have the following information fusion estimate (termed the **Wasserstein-Gaussian Information Fusion Algorithm**):

$$\boxed{\begin{aligned}\mathbf{c} &= (1 - w)\mathbf{a} + w\mathbf{b} \\ \mathbf{P}_{cc} &= ((1 - w)\mathbf{I} + w\mathbf{\Pi})\,\mathbf{P}_{aa}\,((1 - w)\mathbf{I} + w\mathbf{\Pi}) \\ \mathbf{\Pi} &= \mathbf{P}_{bb}^{1/2}(\mathbf{P}_{bb}^{1/2}\mathbf{P}_{aa}\mathbf{P}_{bb}^{1/2})^{-\frac{1}{2}}\mathbf{P}_{bb}^{1/2}, \quad w \in (0, 1)\end{aligned}}$$

This algorithm is motivated by the fact that it defines the best approximation to the measure that is closest in the weighted Wasserstein sense to two input measures defined exactly by the distributions of $\mathbf{a}$ and $\mathbf{b}$. The output measure is thus a kind of weighted average between two input measures

[1] This inequality is in the sense of matrix positive definiteness.

in the sense of a probabilistically justified and intuitive metric (i.e. the Wasserstein metric). It is appealing since finding a measure that is close to two input measures in this sense does not depend on the correlation $\tilde{\mathbf{P}}_{ab}$ which is often assumed unknown (but possibly non-zero).

Practically, this estimation algorithm is taking two Gaussian (random variable) inputs $\mathbf{a}$ and $\mathbf{b}$ and producing a third (random variable) output $\mathbf{c} = (1-w)\mathbf{a} + w\mathbf{b}$ which has mean $\mathbf{c}^*$ and true covariance

$$\begin{aligned}\tilde{\mathbf{P}}_{cc} = (1-w)^2\tilde{\mathbf{P}}_{aa} + w^2\tilde{\mathbf{P}}_{bb} \\ +w(1-w)\tilde{\mathbf{P}}_{ab} + w(1-w)\tilde{\mathbf{P}}_{ba}\end{aligned} \tag{18}$$

from (11). Of course, given $\mathbf{c} = (1-w)\mathbf{a} + w\mathbf{b}$ (or any other estimator of the form (9)) it is not possible to compute the true covariance (or use (11) at all; say with estimates $\mathbf{P}_{aa} \geq \tilde{\mathbf{P}}_{aa}$ etc.) when $\tilde{\mathbf{P}}_{ab}$ is unknown and non-zero. Hence the construction of $\mathbf{P}_{cc}$ as a sole function of $\mathbf{P}_{aa}$ and $\mathbf{P}_{bb}$. Of course, we want $\mathbf{P}_{cc}$ to be a good representation of $\tilde{\mathbf{P}}_{cc}$ but further it is often accepted that one also wants consistency $\mathbf{P}_{cc} \geq \tilde{\mathbf{P}}_{cc}$ so that the estimator is not over-confident.

We now have the following main result.

Proposition 1. *Let $w \in (0,1)$ and suppose an estimator $\mathbf{c}$ for $\mathbf{c}^*$ is given by the Wasserstein-Gaussian Information Fusion Algorithm along with the associated covariance matrix $\mathbf{P}_{cc}$. Computation of $\mathbf{P}_{cc}$ uses $\mathbf{P}_{aa} \geq \tilde{\mathbf{P}}_{aa}$ and $\mathbf{P}_{bb} \geq \tilde{\mathbf{P}}_{bb}$ which are guaranteed consistent. Then*

$$\mathbf{P}_{cc} \geq \tilde{\mathbf{P}}_{cc} \tag{19}$$

is guaranteed if and only if

$$(\tilde{\mathbf{P}}_{aa}\tilde{\mathbf{P}}_{bb})^{1/2} + (\tilde{\mathbf{P}}_{bb}\tilde{\mathbf{P}}_{aa})^{1/2} \geq \tilde{\mathbf{P}}_{ab} + \tilde{\mathbf{P}}_{ba} \tag{20}$$

where $\tilde{\mathbf{P}}_{cc}$ is the true covariance of $\mathbf{c}$ given by (18).

Before proceeding to the proof we need the following lemmas.

Lemma 1. *Let $\mathbf{\Pi} = \mathbf{P}_{bb}^{1/2}(\mathbf{P}_{bb}^{1/2}\mathbf{P}_{aa}\mathbf{P}_{bb}^{1/2})^{-\frac{1}{2}}\mathbf{P}_{bb}^{1/2}$. The following statements hold:*

1) $\mathbf{\Pi} = \mathbf{P}_{aa}^{-1/2}(\mathbf{P}_{aa}^{1/2}\mathbf{P}_{bb}\mathbf{P}_{aa}^{1/2})^{\frac{1}{2}}\mathbf{P}_{aa}^{-1/2}$
2) $\mathbf{\Pi}\mathbf{P}_{aa}\mathbf{\Pi} = \mathbf{P}_{bb}$

Proof: Matrices in the form of $\mathbf{\Pi}$ are well studied and actually correspond to the matrix geometric mean of $\mathbf{P}_{bb}$ and $\mathbf{P}_{aa}^{-1}$. The identities are found in [24]. ∎

Lemma 2. *Let $\mathbf{\Pi} = \mathbf{P}_{aa}^{-1/2}(\mathbf{P}_{aa}^{1/2}\mathbf{P}_{bb}\mathbf{P}_{aa}^{1/2})^{\frac{1}{2}}\mathbf{P}_{aa}^{-1/2}$. The following statements hold:*

1) $\mathbf{\Pi}\mathbf{P}_{aa} = (\mathbf{P}_{bb}\mathbf{P}_{aa})^{1/2}$
2) $\mathbf{P}_{aa}\mathbf{\Pi} = (\mathbf{P}_{aa}\mathbf{P}_{bb})^{1/2}$

Proof: With $\mathbf{\Pi}$ as defined we use the following identities $\mathbf{\Pi} = (\mathbf{P}_{bb}\mathbf{P}_{aa})^{1/2}\mathbf{P}_{aa}^{-1} = \mathbf{P}_{aa}^{-1}(\mathbf{P}_{aa}\mathbf{P}_{bb})^{1/2}$ as found in [24]. ∎

We now proceed to the proof of the main proposition.

of Proposition 1: For the remainder of the proof let $\mathbf{\Pi} = \mathbf{P}_{bb}^{1/2}(\mathbf{P}_{bb}^{1/2}\mathbf{P}_{aa}\mathbf{P}_{bb}^{1/2})^{-\frac{1}{2}}\mathbf{P}_{bb}^{1/2}$. Now by expanding the quadratic form of $\mathbf{P}_{cc}$ we get

$$\begin{aligned}\mathbf{P}_{cc} - \tilde{\mathbf{P}}_{cc} &= w^2\mathbf{\Pi}\mathbf{P}_{aa}\mathbf{\Pi} + (1-w)^2\mathbf{P}_{aa} \\ &\quad +w(1-w)\left[\mathbf{\Pi}\mathbf{P}_{aa} + \mathbf{P}_{aa}\mathbf{\Pi}\right] \\ &\quad -w(1-w)\left[\tilde{\mathbf{P}}_{ab} + \tilde{\mathbf{P}}_{ba}\right] \\ &\quad -w^2\tilde{\mathbf{P}}_{bb} - (1-w)^2\tilde{\mathbf{P}}_{aa}\end{aligned} \tag{21}$$

From Lemma 1 we have $\mathbf{\Pi}\mathbf{P}_{aa}\mathbf{\Pi} = \mathbf{P}_{bb}$ and thus

$$\begin{aligned}\mathbf{P}_{cc} - \tilde{\mathbf{P}}_{cc} &\geq w(1-w)\left[\mathbf{\Pi}\mathbf{P}_{aa} + \mathbf{P}_{aa}\mathbf{\Pi}\right] \\ &\quad -w(1-w)\left[\tilde{\mathbf{P}}_{ab} + \tilde{\mathbf{P}}_{ba}\right] \\ &= w(1-w)\left[\mathbf{\Pi}\mathbf{P}_{aa} + \mathbf{P}_{aa}\mathbf{\Pi} - \tilde{\mathbf{P}}_{ab} - \tilde{\mathbf{P}}_{ba}\right]\end{aligned} \tag{22}$$

using the relations $\mathbf{P}_{aa} \geq \tilde{\mathbf{P}}_{aa}$ and $\mathbf{P}_{bb} \geq \tilde{\mathbf{P}}_{bb}$.

It then follows that $\mathbf{P}_{cc} - \tilde{\mathbf{P}}_{cc} \geq \mathbf{0}$ if and only if

$$\mathbf{\Pi}\mathbf{P}_{aa} + \mathbf{P}_{aa}\mathbf{\Pi} - \tilde{\mathbf{P}}_{ab} - \tilde{\mathbf{P}}_{ba} \geq \mathbf{0} \tag{23}$$

Applying Lemma 2 it follows that $\mathbf{P}_{cc} - \tilde{\mathbf{P}}_{cc} \geq \mathbf{0}$ if and only if

$$(\mathbf{P}_{aa}\mathbf{P}_{bb})^{1/2} + (\mathbf{P}_{bb}\mathbf{P}_{aa})^{1/2} - \tilde{\mathbf{P}}_{ab} - \tilde{\mathbf{P}}_{ba} \geq \mathbf{0} \tag{24}$$

and to guarantee consistency for all possible $\mathbf{P}_{aa} \geq \tilde{\mathbf{P}}_{aa}$ and $\mathbf{P}_{bb} \geq \tilde{\mathbf{P}}_{bb}$ we set $\mathbf{P}_{aa} = \tilde{\mathbf{P}}_{aa}$ and $\mathbf{P}_{bb} = \tilde{\mathbf{P}}_{bb}$ and the proof is complete. ∎

One may question whether the inequality

$$(\tilde{\mathbf{P}}_{aa}\tilde{\mathbf{P}}_{bb})^{1/2} + (\tilde{\mathbf{P}}_{bb}\tilde{\mathbf{P}}_{aa})^{1/2} - \tilde{\mathbf{P}}_{ab} - \tilde{\mathbf{P}}_{ba} \geq \mathbf{0} \tag{25}$$

required by the proposition is automatically satisfied. Indeed, the authors initially suspected this may be the case. We explore this idea further now. We note the following lemma.

Lemma 3. *Let $\mathbf{P}_{aa} > \mathbf{0}$, $\mathbf{P}_{bb} > \mathbf{0}$ and $\mathbf{P}_{ab}$ be defined as before and given and then define*

$$\mathbf{M} = \begin{bmatrix} \mathbf{P}_{aa} & \mathbf{P}_{ab} \\ \mathbf{P}_{ba} & \mathbf{P}_{bb} \end{bmatrix} \tag{26}$$

noting $\mathbf{M} > \mathbf{0}$. Then there exists a contractive matrix $\mathbf{S}$ defined to obey $\mathbf{I} - \mathbf{S}\mathbf{S}^\top \geq \mathbf{0}$ such that $\mathbf{P}_{ab} + \mathbf{P}_{ba} = \mathbf{P}_{aa}^{1/2}\mathbf{S}\mathbf{P}_{bb}^{1/2} + \mathbf{P}_{bb}^{1/2}\mathbf{S}^\top\mathbf{P}_{aa}^{1/2}$.

Proof: See [24]. ∎

This lemma may be used to generalise the notion of a correlation coefficient found in the scalar case.

Corollary 1. *Let $w \in (0,1)$ and suppose a scalar estimator c for $c^* \in \mathbb{R}$ is given by the Wasserstein-Gaussian Information Fusion Algorithm along with the associated variance p_{cc}. Computation of p_{cc} uses the scalars $p_{aa} \geq \tilde{p}_{aa}$ and $p_{bb} \geq \tilde{p}_{bb}$ which are guaranteed consistent. Then it always holds that*

$$p_{cc} \geq \tilde{p}_{cc} \tag{27}$$

where $\tilde{p}_{cc}$ is the true covariance of c given by (18).

Proof: We need

$$(\tilde{p}_{aa}\tilde{p}_{bb})^{1/2} + (\tilde{p}_{bb}\tilde{p}_{aa})^{1/2} - \tilde{p}_{ab} - \tilde{p}_{ba} \geq 0 \tag{28}$$

and from Lemma 3 note $\tilde{p}_{ab} = \tilde{p}_{ba} = s(\tilde{p}_{bb}\tilde{p}_{aa})^{1/2}$ for some $s \in [-1, 1]$. ■

In the scalar case, it is interesting to note that $p_{cc} = \tilde{p}_{cc}$ when $p_{aa} = \tilde{p}_{aa}$ and $p_{bb} = \tilde{p}_{bb}$ and the correlation coefficient is $s = 1$. Indeed, the variance estimate becomes more conservative compared to the underlying true variance as s decreases from 1.

In any finite dimension the required condition for consistency can be rewritten using Lemma 3 as

$$(\tilde{\mathbf{P}}_{aa}\tilde{\mathbf{P}}_{bb})^{1/2} + (\tilde{\mathbf{P}}_{bb}\tilde{\mathbf{P}}_{aa})^{1/2} - \tilde{\mathbf{P}}_{aa}^{1/2}\mathbf{S}\tilde{\mathbf{P}}_{bb}^{1/2} - \tilde{\mathbf{P}}_{bb}^{1/2}\mathbf{S}^{\top}\tilde{\mathbf{P}}_{aa}^{1/2} \geq \mathbf{0} \quad (29)$$

where $\mathbf{S}$ is a contractive matrix obeying $\mathbf{I} - \mathbf{S}\mathbf{S}^{\top} \geq \mathbf{0}$. Unfortunately, in higher dimensional cases (beyond scalars) the inequality required for consistency may fail to hold and consistency is dependent on the relationship of the correlation in the individual estimates to the underlying individual estimation covariance estimates.

One further important note is that the relative difference $\mathbf{P}_{cc} - \tilde{\mathbf{P}}_{cc}$ is independent of $w \in (0, 1)$. This is not to say that $\mathbf{P}_{cc}$ may not improve with w but rather that $(\mathbf{P}_{cc} - \tilde{\mathbf{P}}_{cc})$ does not vary.

Conjecture 1. *Let* $\mathbf{S}$ *obey* $\mathbf{I} - \mathbf{S}\mathbf{S}^{\top} \geq \mathbf{0}$ *then*

$$(\tilde{\mathbf{P}}_{aa}\tilde{\mathbf{P}}_{bb})^{1/2} + (\tilde{\mathbf{P}}_{bb}\tilde{\mathbf{P}}_{aa})^{1/2} - \tilde{\mathbf{P}}_{aa}^{1/2}\mathbf{S}\tilde{\mathbf{P}}_{bb}^{1/2} - \tilde{\mathbf{P}}_{bb}^{1/2}\mathbf{S}^{\top}\tilde{\mathbf{P}}_{aa}^{1/2} \not\geq \mathbf{0} \quad (30)$$

for all $\tilde{\mathbf{P}}_{aa} \geq \mathbf{0}$ *and* $\tilde{\mathbf{P}}_{bb} \geq \mathbf{0}$.

This conjecture has been tested via numerous random examples. The significance of this conjecture is that it immediately implies that either the proposed estimator is consistent $(\mathbf{P}_{cc} - \tilde{\mathbf{P}}_{cc}) \geq \mathbf{0}$ or that $\mathbf{P}_{cc} \not\geq \tilde{\mathbf{P}}_{cc}$ & $\tilde{\mathbf{P}}_{cc} \not\geq \mathbf{P}_{cc}$. In other words, the proposed estimated covariance may be consistent but it will never be less (in the positive-semidefinite sense) than the true variance of the estimate. That is, the conjecture implies $\tilde{\mathbf{P}}_{cc} \not\geq \mathbf{P}_{cc}$. This is a desirable property which implies that the proposed covariance estimate is never overly confident.

Finally, we note that in practice one is not typically interested in the estimator for all values $w \in (0, 1)$ but rather in the estimator for a particular value of w. We now have the following information fusion estimate (termed the **Optimal Wasserstein-Gaussian Information Fusion Algorithm**):

$$\begin{array}{rcl} \mathbf{c} & = & (1-\overline{w})\mathbf{a} + \overline{w}\mathbf{b} \\ \mathbf{P}_{cc} & = & ((1-\overline{w})\mathbf{I} + \overline{w}\mathbf{\Pi})\,\mathbf{P}_{aa}\,((1-\overline{w})\mathbf{I} + \overline{w}\mathbf{\Pi}) \\ \mathbf{\Pi} & = & \mathbf{P}_{bb}^{1/2}(\mathbf{P}_{bb}^{1/2}\mathbf{P}_{aa}\mathbf{P}_{bb}^{1/2})^{-\frac{1}{2}}\mathbf{P}_{bb}^{1/2} \\ \overline{w} & = & \operatorname{argmin}_{w\in(0,1)} \operatorname{tr}\mathbf{P}_{cc}|_{\overline{w}=w} \end{array}$$

We do not explore the details of computing an optimal $\overline{w}$ via a particular optimisation method here but note simply that numerous protocols are applicable including a simple line search algorithm. Other criterion beyond the trace could also be substituted without difficulty. As noted previously the particular value of w has no effect on the relative value of $(\mathbf{P}_{cc} - \tilde{\mathbf{P}}_{cc})$.

IV. Direct Information Fusion of Empirical Measures

Again let $|\mathcal{V}| = 2$ and suppose that an arbitrary continuous $\mu_i \in \mathfrak{U}_p$ supported on $\mathbb{R}^m$ is approximated by some discrete empirical measure $\tilde{\mu}_i^N$ such that $N \to \infty$ implies $\tilde{\mu}_i^N \to \mu_i$ implies $\ell_p(\tilde{\mu}_i^N, \mu_i) \to 0$ for all $1 \leq p < \infty$ [2]. Here

$$\tilde{\mu}_i^N = \frac{1}{N}\sum_{j=1}^N \delta_{x_j^i} \quad (31)$$

where, for example, x_j^i can be considered a realisation of a random variable drawn independently from μ_i. Let $\mathscr{U}_p \subset \mathfrak{U}_p$ denote the space of all such measures. We drop the superscript N but note that each $\tilde{\mu}_i$, $\forall i \in \{1, 2\}$ is defined by the same N.

Let $p = 2$ going forward. Consider the finite set $\mathcal{U}_i = \{x_1^i, \ \ldots, x_N^i\}$ of sample points (indexed from 1 to N). One can write $\ell_2(\tilde{\mu}_i, \tilde{\mu}_j)^2$ as

$$\ell_2(\tilde{\mu}_i, \tilde{\mu}_j)^2 = \min_{\sigma_k^j, \ k\in\{1,\ldots,N\}} \ \frac{1}{N}\sum_{k=1}^N \|x_k^i - x_{\sigma_k^j}^j\|^2 \quad (32)$$

where the minimisation is taken over all permutations $\sigma^j : \{1, \ldots, N\} \to \{1, \ldots, N\}$.

It now follows that

$$\tilde{\nu} \ = \ \inf_{z\in\mathscr{U}_2} \sum_{i\in\{1,2\}} \omega_i \ell_2(z, \tilde{\mu}_i)^2 = \frac{1}{N}\sum_{i=1}^N \delta_{z_i^*} \quad (33)$$

where $\mathcal{Z}^* = \{z_1^*, \ \ldots, \ z_N^*\}$ is a finite set.

For a given $\mathcal{Z}^*$ we can define a vector $\mathbf{z}^* = \left[z_1^{*\top} \ \ldots \ z_N^{*\top}\right]^{\top}$ and

$$\mathbf{z}^* \ = \ \operatorname*{argmin}_{z_i\in\mathbb{R}^m} \left[\omega_1 \min_{\sigma^1} \frac{1}{N}\sum_{k=1}^N \|z_k - x_{\sigma_k^1}^1\|^2 + \omega_2 \min_{\sigma^2} \frac{1}{N}\sum_{k=1}^N \|z_k - x_{\sigma_k^2}^2\|^2\right] \quad (34)$$

For some fixed permutation σ^j, $\forall j$, consider

$$\mathbf{z}_* = \operatorname*{argmin}_{z_i\in\mathbb{R}^m} \ \left[\tfrac{1}{N}\sum_{k=1}^N \left[\omega_1\|z_k - x_{\sigma_k^1}^1\|^2 + \omega_2\|z_k - x_{\sigma_k^2}^2\|^2\right]\right] \quad (35)$$

The solution to this optimisation problem is easily given by

$$z_{k*} = \omega_1 x_{\sigma_k^1}^1 + \omega_2 x_{\sigma_k^2}^2, \quad \forall k \in \{1, \ldots, N\} \quad (36)$$

with $\omega_i \in (0, 1)$ and $\omega_1 + \omega_2 = 1$.

Now it then follows that

$$\sigma^* \ = \ \operatorname*{argmin}_{\sigma^j, \ j\in\{1,2\}} \frac{1}{N}\sum_{k=1}^N \left[\|z_{k*} - x_{\sigma_k^1}^1\|^2 + \|z_{k*} - x_{\sigma_k^2}^2\|^2\right] \quad (37)$$

$$= \ \operatorname*{argmin}_{\sigma^j, \ j\in\{1,2\}} \frac{1}{N}\sum_{k=1}^N \left[\|\omega_1(x_{\sigma_k^1}^1 - x_{\sigma_k^2}^2)\|^2 + \|\omega_2(x_{\sigma_k^1}^1 - x_{\sigma_k^2}^2)\|^2\right] \quad (38)$$

[2] More generally, $\tilde{\mu}_i^N$ as used here could be any discrete normalised probability measure.

and then

$$\mathbf{z}^* = \mathbf{z}_* \big|_{\sigma^j=\sigma^{j*},\ j\in\{1,2\}} \tag{39}$$

where the arguments of the minimisation are the permutations $\sigma^j : \{1,\ldots,N\} \to \{1,\ldots,N\}$.

We then have the following main result.

Theorem 2. *Let $i \in \{1,2\}$ and consider two measures*

$$\tilde{\mu}_i^N = \tfrac{1}{N}\sum_{j=1}^N \delta_{x_j^i} \tag{40}$$

along with the finite sets $\mathcal{U}_i = \{x_1^i, \ldots, x_N^i\}$ of sample points. Then the globally optimal solution to

$$\tilde{\nu} \quad = \quad \inf_{z\in\mathscr{U}_2} \sum_{i\in\{1,2\}} \omega_i \ell_2(z,\tilde{\mu}_i)^2 \tag{41}$$

is given by

$$\tilde{\nu} \quad = \quad \tfrac{1}{N}\sum_{k=1}^N \delta_{z_k^*} \tag{42}$$

where

$$z_k^* = \omega_1 x^1_{\sigma_k^{1*}} + \omega_2 x^2_{\sigma_k^{2*}} \tag{43}$$

and where σ_k^{i*} is given as the output of Algorithm 1.

Let $\mathrm{ind}(x_j^i) = j$. Then Algorithm 1 is a solution to the optimisation problem in (38).

Algorithm 1 Solution to the Optimal Information Fusion Problem with Randomly Sampled Empirical Measures

$\widehat{\mathcal{U}}_i = \mathcal{U}_i,\ \forall i \in \{1,2\}$
for $k = 1$ to N **do**
  nearest $= \infty$
  for each element x_i^1 in $\widehat{\mathcal{U}}_1$ **do**
    for each element x_j^2 in $\widehat{\mathcal{U}}_2$ **do**
      if $\|x_i^1 - x_j^2\|^2 <$ nearest **then**
        $\sigma_k^{1*} = \mathrm{ind}(x_i^1),\ \sigma_k^{2*} = \mathrm{ind}(x_j^2)$
        nearest $= \|x_i^1 - x_j^2\|^2$
      end if
    end for
  end for
  $\widehat{\mathcal{U}}_i = \widehat{\mathcal{U}}_i \setminus \{x^i_{\sigma_k^i}\},\ \forall i \in \{1,2\}$
end for

Proof: Given the derivation to (39) it remains only to show that Algorithm 1 solves the optimisation problem given in (38). Algorithm 1 pairs points in $\mathcal{U}_1$ with points in $\mathcal{U}_2$ based on the (squared) Euclidean distance between them, i.e. the outcome implies $x^1_{\sigma_1^{1*}}$ and $x^2_{\sigma_1^{2*}}$ are the two closest points to each other when picking disjointly from both $\mathcal{U}_1$ and $\mathcal{U}_2$. Then $x^1_{\sigma_2^{1*}}$ and $x^2_{\sigma_2^{2*}}$ are the next two closest points and so on.

Consider the problem in (38) and note that the $x^1_{\sigma_1^{1*}}$ and $x^2_{\sigma_1^{2*}}$ minimises the first $k = 1$ summation term over any other possible pairing of points. Fixing these two points it then follows that the pair $x^1_{\sigma_2^{1*}}$ and $x^2_{\sigma_2^{2*}}$ minimises the second $k = 2$ summation term over any other possible pairing of points and so on for all k. Because σ_k^{i*} only appears in one summation term in (38) and of course the ordering in k is unimportant it follows that this minimises the total summation. This completes the proof. ∎

A more general discussion on the existence and uniqueness of solutions to the general optimisation problem for computing a measure in $\mathfrak{U}_p$ that is closest (in a weighted Wasserstein sense) to a given set of (possibly more than 2) input measures is found in [14]. For discrete input measures in $\mathscr{U}_2$ the optimal solution is not necessarily unique (though it seems it is generically unique). It is worth noting [25] where a similar mathematical formalism is given for computing a discrete measure that is closest (in the Wasserstein sense) to two discrete input measures. No reference to empirical measures is given in [25] and so the inputs may have common support. The application considered in [25] is one of texture mixing in computer vision and no relation to information fusion is discussed. More efficient approximations (or relaxations) to the underlying optimisation problem may also be considered [26] to reduce the complexity.

Moreover, we note that the algorithm presented here is just a direct approach for computing a specific case of the well-known McCann interpolant measure (i.e. a measure that lies on the geodesic connecting two input measures) when the input measures are discrete [23].

The real novelty of this presentation lies in the connection to information fusion of sampled probability measures and the potential applications of such in fields like Monte-Carlo estimation [9] etc.

A. Illustrative Examples

Two arbitrary Gaussian mixture densities are considered each with 8 components. These mixtures are randomly sampled at $N = 1000$ to generate $\tilde{\mu}_i$, $\forall i \in \{1,2\}$. The individual μ_i, $\forall i \in \{1,2\}$ along with the sample points $\mathcal{U}_i$, $\forall i \in \{1,2\}$ are shown in Figure 1.

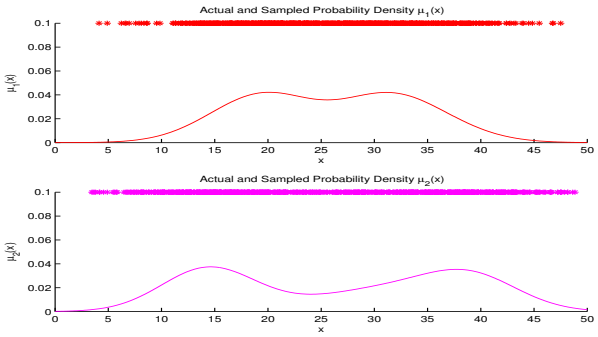

Fig. 1. The two continuous inputs μ_i, $\forall i \in \{1,2\}$ along with the sample points $\mathcal{U}_i$, $\forall i \in \{1,2\}$. The samples are shown at a height of 0.1 for clarity.

The solution to the direct sampled fusion problem

$$\tilde{\nu} \quad = \quad \inf_{z\in\mathscr{U}_2} \sum_{i\in\{1,2\}} \tfrac{1}{2}\ell_2(z,\tilde{\mu}_i)^2 \tag{44}$$

is then computed according to Theorem 2. For visualization, a Kernel method is applied to the fused sample points defining $\tilde{\nu}$ in order to obtain

$$\nu \quad = \quad \tfrac{1}{Nh\sqrt{2\pi}}\sum_{i=1}^N \exp\left[-\tfrac{(x-z_i^*)}{2h^2}\right] \tag{45}$$

with bandwidth $h = (\frac{4\gamma^5}{3N})^{1/5}$ where γ is the sample standard deviation [10]. Both ν and $\tilde{\nu}$ are displayed in Figure 2.

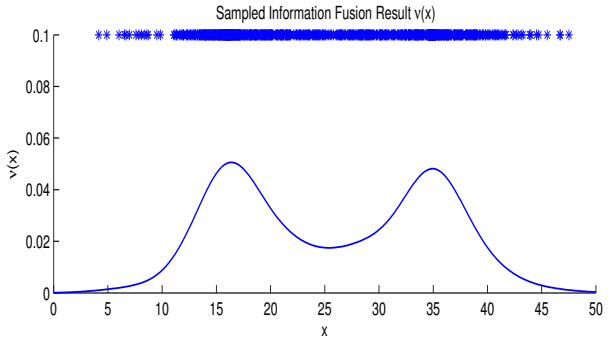

Fig. 2. The sampled information fusion solution $\tilde{\nu}$ and a corresponding Kernel-based density estimate ν. The sample points are plotted at a height of 0.1 for clarity. The sampled solution $\tilde{\nu}$ (and thus the corresponding density estimation ν) is computed using only the sets $\mathcal{U}_i$, $\forall i \in \{1, 2\}$ as input.

V. Concluding Remarks

This work introduced an information fusion protocol that delivers a probability measure that is the 'closest' measure to a collection of given input measures (in the sense of the weighted Wasserstein metric on the space of probability measures).

We considered the explicit computation of this information fusion result for two important (application driven) special cases. Firstly, we examined the case of two Gaussian input measures which may or may not be independent and we detailed the computation of the fusion result and explored the consistency of this result. Secondly, we examined the case of two empirical (randomly sampled) input measures and we provided an information fusion computation that works directly on the two discrete empirical samples.

In the Gaussian case, no comparison, either through simulation or analysis, with covariance intersection or log-linear opinion pools [5], [15], [16] has been considered and such work would be necessary before Wasserstein information fusion was considered applicable. Comparison with other related work [17]–[21] is also important and a potential topic for future work.

Similarly, in the case of empirical measures, further study and analysis is required. Also, comparison with those computationally efficient approximations for sampling the output of the product of input measures [12], [13] among other approaches is needed.

References

[1] L. Ambrosio, N. Gigli, and G. Savaraé, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Basel, Switzerland: Birkhauser Verlag, 2005.

[2] R. Clemen and R. Winkler, "Combining probability distributions from experts in risk analysis," *Risk Analysis*, vol. 19, no. 2, pp. 187–203, 1999.

[3] S. Julier, T. Bailey, and J. Uhlmann, "Using exponential mixture models for suboptimal distributed data fusion," in *Proc. of the 2006 Nonlinear Statistical Signal Processing Workshop*, 2006, pp. 160–163.

[4] A. Abbas, "A Kullback-Leibler view of linear and log-linear pools," *Decision Analysis*, vol. 6, no. 1, pp. 25–37, 2009.

[5] T. Bailey, S. Julier, and G. Agamennoni, "On conservative fusion of information with unknown non-gaussian dependence," in *Proc. of the 15th Int. Conference on Information Fusion*, 2012, pp. 1876–1883.

[6] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House, 2007.

[7] H. Durrant-Whyte and T. Henderson, "Multisensor data fusion," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Berlin, Germany: Springer-Verlag, 2008.

[8] A. Bishop and A. Doucet, "Distributed nonlinear consensus in the space of probability measures," in *Proc. of the 19th IFAC World Congress*, Cape Town, South Africa, August 2014.

[9] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[10] B. Silverman, *Density Estimation for Statistics and Data Analysis*. CRC Press, 1986.

[11] M. Liggins, C.-Y. Chong, I. Kadar, M. Alford, V. Vannicola, and S. Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 95–107, 1997.

[12] A. Ihler, E. Sudderth, W. Freeman, and A. Willsky, "Efficient multiscale sampling from products of Gaussian mixtures." in *Proc. of 17th Annual Conf. on Neural Information Processing Systems*, 2003.

[13] E. Sudderth, A. Ihler, M. Isard, W. Freeman, and A. Willsky, "Nonparametric belief propagation," *Communications of the ACM*, vol. 53, no. 10, pp. 95–103, 2010.

[14] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," *SIAM Journal on Mathematical Analysis*, vol. 43, no. 2, pp. 904–924, 2011.

[15] S. Julier and J. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. of the 1997 American Control Conference*, 1997, pp. 2369–2373.

[16] L. Chen, P. Arambel, and R. Mehra, "Estimation under unknown correlation: Covariance intersection revisited," *IEEE Trans. on Auto. Control*, vol. 47, no. 11, pp. 1879–1882, 2002.

[17] S. Julier, "An Empirical Study into the Use of Chernoff Information for Robust, Distributed Fusion of Gaussian Mixture Models," in *Proceedings 9th International Conference on Information Fusion*, 2006.

[18] M. Reinhardt, B. Noack, M. Baum, and U. Hanebeck, "Analysis of set-theoretic and stochastic models for fusion under unknown correlations," in *Proceedings 14th International Conference on Information Fusion*, 2011.

[19] C.-Y. Chong and S. Mori, "Convex combination and covariance intersection algorithms in distributed fusion," in *Proc. of the 4th International Conf. on Information Fusion*, 2001.

[20] Y. Wang and X. Li, "Distributed estimation fusion with unavailable cross-correlation," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 259–278, January 2012.

[21] J. Sijs and M. Lazar, "State fusion with unknown correlation: Ellipsoidal intersection," *Automatica*, vol. 48, no. 8, 2012.

[22] A. Gibbs and F. Su, "On choosing and bounding probability metrics," *International Statistical Review*, vol. 70, no. 3, pp. 419–435, 2002.

[23] R. McCann, "A convexity principle for interacting gases," *Advances in Mathematics*, vol. 128, no. 1, pp. 153–179, 1997.

[24] D. Bernstein, *Matrix mathematics: theory, facts, and formulas*. Princeton University Press, 2009.

[25] 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*. Springer, 2012, pp. 435–446.

[26] M. Cuturi and A. Doucet, "Fast computation of Wasserstein barycenters," *arXiv preprint arXiv:1310.4375*, 2013.