Multisensor Estimation Fusion Based on Kernel Mean Embedding

Mingjie Luo, Jie Zhou, Qingke Zou
College of Mathematics
Sichuan University
Chengdu, Sichuan 610064, China
lmj960407@163.com, jzhou@scu.edu.cn, m18208258544@163.com

***Abstract*—This work deals with the estimation fusion for distributed multisensor systems under the framework of local estimates being taken as probability density functions. The estimation fusion is formulated to an optimization problem that minimizes the sum of squared distances between the fused probability density and each local probability density. The maximum mean discrepancy, which is a distance between two probability density functions, is considered. It is defined by the kernel mean embeddings from the probability density function space to a reproducing kernel Hilbert space. For the quadratic, cubic and Gaussian kernels, either the analytical solutions are derived or the numerical methods are developed for solving the aforementioned optimization problem. Numerical experiments are provided to illustrate the performance of the proposed estimation fusion methods.**

***Keywords*—Distributed estimation fusion, reproducing kernel Hilbert space, kernel mean embedding, maximum mean discrepancy**

I. Introduction

Multisensor data fusion is widely applied in many fields such as target tracking, signal processing, machine learning and artificial intelligence over the past decades [1], [2], [3], [4], [5]. The purpose of multisensor data fusion is to fuse data from local sensors into a result which has a more precise and reliable description for the target. In existing literature, there are two common fusion architectures which are namely the centralized fusion and the distributed fusion. The latter has a less communication burden, higher survivability and faster computing speed so that it is more worthy of research on the distributed fusion [6], [7], [8], [9].

For a distributed tracking system, the cross-correlation of local estimation errors between two sensors does exist due to the common process noise and the exchange of transmitted data [10], [11]. In an ideal situation which the cross-correlation is known, the optimal fused result can be directly obtained according to the linear minimum mean-squared error (LMMSE) rule [12]. However, in the vast majority of practical cases, it is difficult to obtain the cross-correlation. In order to deal with the unknown cross-correlation, several approaches have been proposed. The most widely used method is called the covariance intersection (CI) algorithm [13], [14]. Based on different optimization criteria, a great deal of methods has been proposed. Two commonly used methods are the following. Minimizing the determinant of the covariance matrix of the fused result is adopted by the determinant-minimization CI (DCI) algorithm [15]. And the fast CI (FCI) considers a new criterion inspired by the Chernoff information criterion which finds the "halfway" point under the Kullback–Leibler (KL) divergence from two local probability density functions (PDFs) [16].

Although it is somewhat effective and convenient, the CI algorithm is very conservative. In order to make up for the deficiency, handling with PDFs which contain full distribution information becomes potentially more useful [17], [18], [19], [20]. In [19], several fusion approaches dealing with PDFs are divided into three categories. For the axiomatic approach, the fusion rule is defined indirectly according to a set of axioms [21], [22]. The optimization approach can derive the fusion rule by minimizing the weighted average of a discrepancy measure or distance between the PDFs. In [23] and [24], the minimization of the sum of KL divergences from the given local PDFs to the fused PDF is addressed. In [25] and [26], the estimation fusion is formulated based on the geodesic distance between the PDFs embedded on a Riemannian manifold by taking advantage of the information geometric approach. In [27], with regard to the supra-Bayesian approach, the fusion center is considered as a global posterior PDF given the local PDFs modeled as random observations.

This work provides an estimation fusion framework for the distributed systems in which the local estimates are characterized as the PDFs. The optimal fusion is formulated by embedding all local PDFs into a reproducing kernel Hilbert space (RKHS) [28], [29]. From most existing approaches for estimation fusion, we can find out that the distance or dissimilarity such as Shannon entropy [15], KL divergence [23] and geodesic distance [25] between two distributions plays an important role when constructing objective functions in the optimization problems. The inspiration is drawn from studies of the RKHS. In [30], [31], a class of distances on the probability space called the integral probability metric (IPM) is defined to describe the distance between two probability distributions. Under different function spaces, the integral

This work was supported in part by the Sichuan Science and Technology Program under Grant 2024NSFSC0444, the Key Laboratory of Data Protection and Intelligent Management (Sichuan University), Ministry of Education, China, and also the Fundamental Research Funds for the Central Universities under Grant SCU2023D008.

J. Zhou is the corresponding author.

probability metric can stand for various of distances such as Kolmogorov distance, total variation distance and Wasserstein distance [32]. When the chosen function space is a unit ball in the RKHS, the particular instance of the IPM is known as the maximum mean discrepancy (MMD) [32]. By reason of the reproducing property of functions in the RKHS, we can theoretically prove that the MMD between two distributions is equal to the norm of the difference between corresponding kernel mean embeddings. Moreover, the analytic form of the MMD can be derived for some special kernel functions and distributions. For three specific kernel functions which are respectively the quadratic polynomial kernel, cubic polynomial kernel and Gaussian kernel, under Gaussian distribution assumption, we formulate the corresponding optimization problems in which the common criterion is to minimize the sum of the squared MMDs between the given local PDFs in pairs, and derive the fusion formulas.

The structure of this paper is as follows. In Section II, some mathematical notations as well as several basic concepts and useful results of kernel mean embedding are introduced. The formulation of the optimal estimation fusion based on the kernel mean embedding is involved in Section III. Section IV presents to solve optimization problems and derive fused results under three different MMDs induced by corresponding kernel mean embeddings. Section V provides numerical experiments to verify the performance of the proposed methods. Conclusions are given in Section VI.

II. Preliminaries

This section gives a brief review of some mathematical notations as well as basic definitions of the RKHS and the kernel mean embedding. In addition, several useful results are also supplemented following the definitions.

A. Notations

Throughout this work, all vector-valued random variables are italicized. Scalars, vectors and matrices are respectively denoted by lightface, boldface lowercase and boldface uppercase letters.

Let $\mathbb{R}_+$ be the set of all positive real numbers and $\mathbb{N}$ denote the set of all natural numbers. $\mathbb{R}^n$ is the set of n-dimensional real vectors, and $\mathbb{S}_+^n$ represents the set of all $n \times n$ real symmetric and positive definite matrices. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The transpose of a vector or matrix is denoted by the symbol $(\cdot)^T$. The trace, norm, determinant and (i,j)-th entry of a matrix are respectively denoted by $\operatorname{tr}(\cdot)$, $\|\cdot\|$, $|\cdot|$ and $(\cdot)_{ij}$.

The Gaussian density function with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $N(\boldsymbol{m}, \boldsymbol{\Sigma})$, and by $N(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{\Sigma})$ to specify the variable $\boldsymbol{x}$.

B. Kernel Function and Reproducing Kernel Hilbert Space

Given a vector space $\mathcal{X}$, a binary function $k(\cdot,\cdot): \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a *kernel function* or a *positive definite kernel* on $\mathcal{X}$ if $k(\cdot,\cdot)$ satisfies the following two conditions [28]:

1) $k(\boldsymbol{x}, \boldsymbol{y}) = k(\boldsymbol{y}, \boldsymbol{x})$ for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$;
2) The Gram matrix $\mathbf{K} = (k(\boldsymbol{x}_i, \boldsymbol{x}_j))_{i,j=1}^n$ is positive definite, i.e., $\sum_{i,j=1}^n c_i c_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0$, for any $n \in \mathbb{N}$, any choice $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$ and any $c_1, \ldots, c_n \in \mathbb{R}$.

Consider a Hilbert space $\mathcal{H}$ of functions. $\mathcal{H}$ is regarded as a RKHS when the evaluation functionals $\mathcal{F}_{\boldsymbol{x}}[f]$ defined as $\mathcal{F}_{\boldsymbol{x}}[f] = f(\boldsymbol{x})$ is bounded, i.e., there exists some $C > 0$ for all $\boldsymbol{x} \in \mathcal{X}$ such that

$$|\mathcal{F}_{\boldsymbol{x}}[f]| = |f(\boldsymbol{x})| \leq C\|f\|_{\mathcal{H}}, \forall f \in \mathcal{H}.$$

C. Kernel Mean Embedding

For a measurable space $\mathcal{X}$, $\mathcal{M}(\mathcal{X})$ denotes the space of probability measures over $\mathcal{X}$. When given a kernel function $k(\cdot,\cdot)$, in order to find images in $\mathcal{H}$ of elements over $\mathcal{M}(\mathcal{X})$, the mapping from $\mathcal{M}(\mathcal{X})$ to $\mathcal{H}$ can be defined as

$$\mu : \mathcal{M}(\mathcal{X}) \to \mathcal{H}, \quad \mathbb{P} \longmapsto \int k(\boldsymbol{x}, \cdot) \mathrm{d}\mathbb{P}(\boldsymbol{x}).$$

The above defined mapping from the probability measure space to the RKHS is called the *kernel mean embedding* of $\mathcal{M}(\mathcal{X})$ into $\mathcal{H}$ endowed with a kernel $k(\cdot,\cdot)$ [29], [33].

D. Integral Probability Metric and Maximum Mean Discrepancy

For arbitrary two probability distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ in $\mathcal{M}(\mathcal{X})$, a metric called the *integral probability metric* between them is defined as

$$\gamma[\mathcal{S}, \mathbb{P}_1, \mathbb{P}_2] = \sup_{f \in \mathcal{S}} \{g_1(\boldsymbol{x}) - g_2(\boldsymbol{y})\}, \tag{1}$$

where $g_i(\cdot) = \int f(\cdot)\mathrm{d}\mathbb{P}_i(\cdot)$ for $i = 1, 2$, $\mathcal{S}$ is a space of real-valued bounded measurable functions on $\mathcal{X}$. From the definition given by (1), we can see that the metric heavily depends on the choice of $\mathcal{S}$. Since it is hard to deal with the whole class of functions, some more restrictive classes are often considered. When the supremum in (1) is taken over functions in $\mathcal{S}_1 = \{f \| \|f\|_{\mathcal{H}} \leq 1\}$, the corresponding metric is known as the *maximum mean discrepancy* which is denoted by $\mathrm{MMD}[\mathcal{H}, \mathbb{P}_1, \mathbb{P}_2]$.

E. Several Useful Results in Kernel Mean Embedding

Here, we provide some results which will be used later. The detailed proofs can be found in the related references (e.g., [34]).

Lemma 2.1: The MMD of two probability distributions can be expressed as the norm of the difference between corresponding kernel mean embeddings, that is,

$$\begin{aligned}\mathrm{MMD}[\mathcal{H}, \mathbb{P}_1, \mathbb{P}_2] &= \sup_{\|f\|_{\mathcal{H} \leq 1}} \{g_1(\boldsymbol{x}) - g_2(\boldsymbol{y})\} \\ &= \sup_{\|f\|_{\mathcal{H} \leq 1}} \{\langle f, \mu_{\mathbb{P}_1} - \mu_{\mathbb{P}_2} \rangle_{\mathcal{H}}\} \\ &= \|\mu_{\mathbb{P}_1} - \mu_{\mathbb{P}_2}\|_{\mathcal{H}},\end{aligned}$$

where $\mu_{\mathbb{P}_i}$ is the kernel mean embedding of $\mathbb{P}_i$ for $i = 1, 2$. And it also can be expressed by the associated kernel function $k(\cdot,\cdot)$ as

$$\begin{aligned}\mathrm{MMD}^2[\mathcal{H}, \mathbb{P}_1, \mathbb{P}_2] = &\, \mathbb{E}_{\boldsymbol{x},\boldsymbol{x}'}[k(\boldsymbol{x}, \boldsymbol{x}')] + \mathbb{E}_{\boldsymbol{y},\boldsymbol{y}'}[k(\boldsymbol{y}, \boldsymbol{y}')] \\ &- 2\mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}[k(\boldsymbol{x}, \boldsymbol{y})],\end{aligned}$$

where $\boldsymbol{x}, \boldsymbol{x}' \sim \mathbb{P}_1$ and $\boldsymbol{y}, \boldsymbol{y}' \sim \mathbb{P}_2$ are respectively independent samples.

For some special kernel functions, the inner product of two Gaussian distributions can be expressed analytically [29], [35]. In fact, we have the following results.

Lemma 2.2: For $i = 1, 2$, if $\mathbb{P}_i$ is an n-dimensional Gaussian distribution with mean vector $\boldsymbol{m}_{\mathbb{P}_i}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbb{P}_i}$, then

1) When $k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\eta}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2\right)$,

$$\langle \mu_{\mathbb{P}_1}, \mu_{\mathbb{P}_2} \rangle_{\mathcal{H}} = \frac{\exp\left(-\frac{1}{2}\boldsymbol{m}_*^T \boldsymbol{\Sigma}_*^{-1} \boldsymbol{m}_*\right)}{|\eta \boldsymbol{\Sigma}_*|^{\frac{1}{2}}};$$

2) When $k(\boldsymbol{x}, \boldsymbol{y}) = (\langle \boldsymbol{x}, \boldsymbol{y} \rangle + 1)^2$,

$$\begin{aligned}\langle \mu_{\mathbb{P}_1}, \mu_{\mathbb{P}_2} \rangle_{\mathcal{H}} &= (\langle \boldsymbol{m}_{\mathbb{P}_1}, \boldsymbol{m}_{\mathbb{P}_2} \rangle + 1)^2 + \operatorname{tr}(\boldsymbol{\Sigma}_{\mathbb{P}_1} \boldsymbol{\Sigma}_{\mathbb{P}_2}) \\ &\quad + \boldsymbol{m}_{\mathbb{P}_1}^T \boldsymbol{\Sigma}_{\mathbb{P}_2} \boldsymbol{m}_{\mathbb{P}_1} + \boldsymbol{m}_{\mathbb{P}_2}^T \boldsymbol{\Sigma}_{\mathbb{P}_1} \boldsymbol{m}_{\mathbb{P}_2};\end{aligned}$$

3) When $k(\boldsymbol{x}, \boldsymbol{y}) = (\langle \boldsymbol{x}, \boldsymbol{y} \rangle + 1)^3$,

$$\begin{aligned}\langle \mu_{\mathbb{P}_1}, \mu_{\mathbb{P}_2} \rangle_{\mathcal{H}} &= (\langle \boldsymbol{m}_{\mathbb{P}_1}, \boldsymbol{m}_{\mathbb{P}_2} \rangle + 1)^3 + 6\boldsymbol{m}_{\mathbb{P}_1}^T \boldsymbol{\Sigma}_{\mathbb{P}_1} \boldsymbol{\Sigma}_{\mathbb{P}_2} \boldsymbol{m}_{\mathbb{P}_2} \\ &\quad + 3(\langle \boldsymbol{m}_{\mathbb{P}_1}, \boldsymbol{m}_{\mathbb{P}_2} \rangle + 1) \cdot b,\end{aligned}$$

where $\boldsymbol{m}_* = \boldsymbol{m}_{\mathbb{P}_1} - \boldsymbol{m}_{\mathbb{P}_2}$, $\boldsymbol{\Sigma}_* = \left(\boldsymbol{\Sigma}_{\mathbb{P}_1} + \boldsymbol{\Sigma}_{\mathbb{P}_2} + \eta^{-1}\mathbf{I}_n\right)$, $b = \left(\operatorname{tr}(\boldsymbol{\Sigma}_{\mathbb{P}_1} \boldsymbol{\Sigma}_{\mathbb{P}_2}) + \boldsymbol{m}_{\mathbb{P}_1}^T \boldsymbol{\Sigma}_{\mathbb{P}_2} \boldsymbol{m}_{\mathbb{P}_1} + \boldsymbol{m}_{\mathbb{P}_2}^T \boldsymbol{\Sigma}_{\mathbb{P}_1} \boldsymbol{m}_{\mathbb{P}_2}\right)$, η is a positive real number.

III. Formulation of Optimal Estimation Fusion

Consider a distributed system with ℓ sensors, and the observed target is $\boldsymbol{x} \in \mathbb{R}^n$. For the i-th sensor, it has the local posterior density $p_i(\boldsymbol{x}|\boldsymbol{z}_i)$ with the measurement $\boldsymbol{z}_i \in \mathbb{R}^n$. The purpose of distributed estimate fusion is to obtain a fused posterior density denoted as $\hat{p}(\boldsymbol{x}|\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell)$ by integrating the information of local estimates. For the convenience of writing, we abbreviate $p_i(\boldsymbol{x}|\boldsymbol{z}_i)$ and $\hat{p}(\boldsymbol{x}|\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell)$ as p_i and $\hat{p}$ in the following text, respectively.

Let $\mathcal{P}$ be a parametric probability density family and p_i belong to $\mathcal{P}$ for $i = 1, \ldots, \ell$. We aim to utilize a formulated optimization criterion to find the optimal objective as $\hat{p}$ in $\mathcal{P}$. Frequently-used criteria for fusing the PDFs stem from minimizing the sum of the dissimilarities or distances such as the Chernoff divergence, the KL divergence and the geodesic distance. From this perspective, the optimal fusion is formulated as

$$\hat{p} = \underset{p \in \mathcal{P}}{\arg\min} \sum_{i=1}^{\ell} d^2(p_i, p), \tag{2}$$

where $d(\cdot, \cdot)$ is a well-defined distance or dissimilarity between two PDFs.

With regard to (2), not all distributions can derive an explicit form of the objective function, so we conservatively take into account the family of Gaussian distributions which is the most familiar among various families. Under the framework of the family of Gaussian distributions, the original local and fused posterior PDFs can be respectively denoted as $N(\boldsymbol{x}; \hat{\boldsymbol{x}}_i, \mathbf{P}_i)$ and $N(\boldsymbol{x}; \hat{\boldsymbol{x}}, \mathbf{P})$, where $\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}} \in \mathbb{R}^n$ and $\mathbf{P}_i, \mathbf{P} \in \mathbb{S}_+^n$. The optimization problem (2) becomes

$$N(\boldsymbol{x}; \hat{\boldsymbol{x}}, \mathbf{P}) = \underset{\substack{N(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{\Sigma}) \in \mathcal{P} \\ (\boldsymbol{m}, \boldsymbol{\Sigma}) \in \mathbb{R}^n \times \mathbb{S}_+^n}}{\arg\min} \sum_{i=1}^{\ell} d^2(N_i, N(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{\Sigma})),$$

where N_i denotes $N(\boldsymbol{x}; \hat{\boldsymbol{x}}_i, \mathbf{P}_i)$.

IV. Solutions using Kernel Mean Embedding

According to Lemma 2.1, the squared MMD between two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ can be written as

$$\begin{aligned}\mathrm{MMD}^2[\mathcal{H}, \mathbb{P}_1, \mathbb{P}_2] &= \|\mu_{\mathbb{P}_1} - \mu_{\mathbb{P}_2}\|_{\mathcal{H}}^2 \\ &= \langle \mu_{\mathbb{P}_1}, \mu_{\mathbb{P}_1} \rangle_{\mathcal{H}} + \langle \mu_{\mathbb{P}_2}, \mu_{\mathbb{P}_2} \rangle_{\mathcal{H}} \\ &\quad - 2\langle \mu_{\mathbb{P}_1}, \mu_{\mathbb{P}_2} \rangle_{\mathcal{H}},\end{aligned}$$

where $\mathbb{P}_i$ is the corresponding distribution of the PDF p_i for $i = 1, 2$. When the distance in (2) is replaced by $\mathrm{MMD}[\mathcal{H}, \mathbb{P}, \mathbb{P}_i]$, the optimization problem (2) can be written as

$$\begin{aligned}\hat{\mathbb{P}} &= \underset{\mathbb{P} \in \Omega}{\arg\min} \sum_{i=1}^{\ell} \mathrm{MMD}^2[\mathcal{H}, \mathbb{P}, \mathbb{P}_i] \\ &= \underset{\mathbb{P} \in \Omega}{\arg\min}\, \ell \langle \mu_{\mathbb{P}}, \mu_{\mathbb{P}} \rangle_{\mathcal{H}} + \sum_{i=1}^{\ell} \langle \mu_{\mathbb{P}_i}, \mu_{\mathbb{P}_i} \rangle_{\mathcal{H}} \\ &\quad - 2\sum_{i=1}^{\ell} \langle \mu_{\mathbb{P}}, \mu_{\mathbb{P}_i} \rangle_{\mathcal{H}},\end{aligned} \tag{3}$$

where Ω is a set of distributions. We can observe that the second term of the objective function in (3) is unrelated to $\mathbb{P}$ so that it can be regarded as a constant term which is neglected. The original optimization problem (3) can be additionally simplified as

$$\hat{\mathbb{P}} = \underset{\mathbb{P} \in \Omega}{\arg\min}\, \ell \langle \mu_{\mathbb{P}}, \mu_{\mathbb{P}} \rangle_{\mathcal{H}} - 2\sum_{i=1}^{\ell} \langle \mu_{\mathbb{P}}, \mu_{\mathbb{P}_i} \rangle_{\mathcal{H}}. \tag{4}$$

For the Gaussian case, due to Lemma 2.2, the inner product $\langle \mu_{\mathbb{P}_1}, \mu_{\mathbb{P}_2} \rangle_{\mathcal{H}}$ of two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ can be expressed analytically under three special kernel functions. We will discuss three cases separately and the concrete results are as follows. For each case, we first discuss the one-dimensional situation for convenience and then move on to multi-dimensional situation.

A. Quadratic Polynomial Kernel

When the kernel function is a quadratic polynomial kernel, in the one-dimensional situation, the inner product between $\mu_{N(m_i, \sigma_i^2)}$ and $\mu_{N(m_j, \sigma_j^2)}$ degenerates into

$$\begin{aligned}\left\langle \mu_{N(m_i, \sigma_i^2)}, \mu_{N(m_j, \sigma_j^2)} \right\rangle_{\mathcal{H}} &= m_i^2 m_j^2 + 2m_i m_j + 1 \\ &\quad + \sigma_i^2 \sigma_j^2 + m_i^2 \sigma_j^2 + m_j^2 \sigma_i^2.\end{aligned} \tag{5}$$

Substitute (5) into (4), the optimization problem for quadratic polynomial kernel in the one-dimensional situation can be written as

$$\begin{aligned} N\left(\hat{m},\hat{\sigma}^2\right) = \underset{N(m,\sigma^2)\in\Omega_1}{\arg\min}\ & \ell\left(m^4+2m^2+1+\sigma^4+2m^2\sigma^2\right) \\ & -2\sum_{i=1}^{\ell}\left(m_i^2m^2+2m_im+1\right. \\ & \left.+\sigma_i^2\sigma^2+m_i^2\sigma^2+\sigma_i^2m^2\right), \end{aligned} \tag{6}$$

where Ω_1 is the set of all one-dimensional Gaussian distributions. In order to derive the partial derivatives with respective to m and σ^2 conveniently, we transform the expression of (6) into

$$\begin{aligned} N\left(\hat{m},\hat{\sigma}^2\right) = \underset{N(m,\sigma^2)\in\Omega_1}{\arg\min}\ & \ell m^4 \\ & +\left(2\ell+2\ell\sigma^2-2\sum_{i=1}^{\ell}\left(m_i^2+\sigma_i^2\right)\right)m^2 \\ & -4\sum_{i=1}^{\ell}m_im+c(m), \end{aligned} \tag{7}$$

and

$$\begin{aligned} N\left(\hat{m},\hat{\sigma}^2\right) = \underset{N(m,\sigma^2)\in\Omega_1}{\arg\min}\ & \ell\sigma^4 \\ & +\left(2\ell m^2-2\sum_{i=1}^{\ell}\left(m_i^2+\sigma_i^2\right)\right)\sigma^2 \\ & +c(\sigma^2), \end{aligned} \tag{8}$$

where $c(m)$ and $c(\sigma^2)$ are the constants unrelated to m and σ^2 respectively.

Let the objective function in (6) denote as $f\left(m,\sigma^2\right)$. According to (7) and (8), taking partial derivatives of $f\left(m,\sigma^2\right)$ with respect to m and σ^2 respectively yields

$$\begin{aligned} \frac{\partial f\left(m,\sigma^2\right)}{\partial m} &\propto \ell m^3-\sum_{i=1}^{\ell}m_i \\ &\quad+\left(\ell+\ell\sigma^2-\sum_{i=1}^{\ell}\left(m_i^2+\sigma_i^2\right)\right)m, \\ \frac{\partial f\left(m,\sigma^2\right)}{\partial \sigma^2} &\propto \ell\sigma^2+\ell m^2-\sum_{i=1}^{\ell}\left(m_i^2+\sigma_i^2\right), \end{aligned}$$

where $A\propto B$ denotes that A is proportional to B.

The final fused results can be derived according to solving the simultaneous equations

$$\begin{cases} \dfrac{\partial f\left(m,\sigma^2\right)}{\partial m}=0, \\ \dfrac{\partial f\left(m,\sigma^2\right)}{\partial \sigma^2}=0, \end{cases} \tag{9}$$

and the specific form is shown in the following theorem.

Theorem 4.1: The analytic expressions for solutions of (9) are given by

$$\begin{aligned} m_{\text{quad}} &= \frac{1}{\ell}\sum_{i=1}^{\ell}m_i, \\ \sigma_{\text{quad}}^2 &= \frac{1}{\ell}\sum_{i=1}^{\ell}\left(m_i^2+\sigma_i^2\right)-\left(\frac{1}{\ell}\sum_{i=1}^{\ell}m_i\right)^2. \end{aligned}$$

Similar to the one-dimensional situation, the fused results for the n-dimensional situation are as follows:

$$\begin{aligned} \boldsymbol{m}_{\text{quad}} &= \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{m}_i, \\ \boldsymbol{\Sigma}_{\text{quad}} &= \frac{1}{\ell}\sum_{i=1}^{\ell}\left(\boldsymbol{m}_i\boldsymbol{m}_i^T+\boldsymbol{\Sigma}_i^2\right) \\ &\quad-\frac{1}{\ell^2}\left(\sum_{i=1}^{\ell}\boldsymbol{m}_i\right)\left(\sum_{i=1}^{\ell}\boldsymbol{m}_i\right)^T. \end{aligned}$$

B. Cubic Polynomial Kernel

When a cubic polynomial kernel is chosen, in the one-dimensional situation, perform same operations referring to the above subsection. The partial derivatives about m and σ^2 can be calculated and simplified as

$$\begin{aligned} \frac{\partial f\left(m,\sigma^2\right)}{\partial m} &\propto \ell m^5+\left(2\ell+4\ell\sigma^2\right)m^3 \\ &\quad-\alpha m^2+\left(\ell+2\ell\sigma^2+3\ell\sigma^4-2\beta\right)m \\ &\quad-\left(\delta+\alpha\sigma^2\right), & (10) \\ \frac{\partial f\left(m,\sigma^2\right)}{\partial \sigma^2} &\propto \left(3\ell m^2+\ell\right)\sigma^2 \\ &\quad+\left(\ell m^4+\ell m^2-\alpha m-\beta\right), & (11) \end{aligned}$$

where $\alpha=\sum_{i=1}^{\ell}m_i^3+3m_i\sigma_i^2$, $\beta=\sum_{i=1}^{\ell}m_i^2+\sigma_i^2$ and $\delta=\sum_{i=1}^{\ell}m_i$. Let the right-hand side of Equations (10) and (11) both equal to zero, the fused results can be sought by the theorem as below.

Theorem 4.2: For the case with cubic polynomial kernel in the one-dimensional situation, the fused mean m_{cubic} is the solution of

$$\begin{aligned} 8\ell^2m^7 &+ \left(13\ell^2-12\beta\ell\right)m^5+\left(2\alpha\ell-9\delta\ell\right)m^4 \\ &\cdot\left(6\ell^2-8\beta\ell\right)m^3+\left(2\alpha\ell+3\alpha\beta-6\delta\right)m^2 \\ &+\left(\ell^2+3\beta^2-\alpha^2\right)m-\delta\ell-\alpha\beta=0, \end{aligned} \tag{12}$$

and the corresponding fused variance σ_{cubic}^2 is

$$\sigma_{\text{cubic}}^2=-\frac{1}{3}m_{\text{cubic}}^2-\frac{2}{9}+\frac{9\alpha m_{\text{cubic}}+9\beta+2\ell}{9\ell\left(3m_{\text{cubic}}^2+1\right)}.$$

Similar to (10) and (11), the partial derivatives of the objective function with regard to $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ for the n-dimensional case after being simplified are given by

$$\begin{aligned}\frac{\partial f(\boldsymbol{m},\boldsymbol{\Sigma})}{\partial \boldsymbol{m}} \propto\, & \ell\left(\boldsymbol{m}^T\boldsymbol{m}\right)^2\boldsymbol{m}+2\ell\boldsymbol{m}^T\boldsymbol{m}\boldsymbol{m} \\ & +2\left(\ell\boldsymbol{m}^T\boldsymbol{\Sigma}\boldsymbol{m}\boldsymbol{m}+\ell\boldsymbol{m}^T\boldsymbol{m}\boldsymbol{\Sigma}\boldsymbol{m}\right) \\ & -\left(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2+\boldsymbol{\alpha}_3+\boldsymbol{\alpha}_4\right) \\ & +\Big(\ell\boldsymbol{m}+2\ell\boldsymbol{\Sigma}\boldsymbol{m}+2\ell\boldsymbol{\Sigma}^2\boldsymbol{m} \\ & +\frac{\ell}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)\boldsymbol{m}+\frac{\ell}{2}\operatorname{tr}\left(\left(\boldsymbol{\Sigma}^2\right)^T\right)\boldsymbol{m} \\ & -\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\Big)-\left(\boldsymbol{\delta}+\boldsymbol{\alpha}_5+\boldsymbol{\alpha}_6+\boldsymbol{\alpha}_7\right), \quad (13)\\ \frac{\partial f(\boldsymbol{m},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} \propto\, & \ell\left(\boldsymbol{m}^T\boldsymbol{m}\boldsymbol{\Sigma}+\boldsymbol{m}\boldsymbol{m}^T\boldsymbol{\Sigma}+\boldsymbol{\Sigma}\boldsymbol{m}\boldsymbol{m}^T+\boldsymbol{\Sigma}\right) \\ & +\Big(\ell\boldsymbol{m}^T\boldsymbol{m}\boldsymbol{m}\boldsymbol{m}^T+\ell\boldsymbol{m}\boldsymbol{m}^T \\ & -\mathbf{A}_1-\mathbf{A}_2-\mathbf{A}_3-\mathbf{B}_1-\mathbf{B}_2\Big), \quad (14)\end{aligned}$$

where

$$\begin{aligned}&\boldsymbol{\alpha}_1=\sum_{i=1}^{\ell}\left(\boldsymbol{m}^T\boldsymbol{m}_i\right)^2\boldsymbol{m}_i, \boldsymbol{\alpha}_2=\sum_{i=1}^{\ell}\boldsymbol{m}^T\boldsymbol{\Sigma}_i\boldsymbol{m}\boldsymbol{m}_i,\\ &\boldsymbol{\alpha}_3=\sum_{i=1}^{\ell}\boldsymbol{m}_i^T\boldsymbol{m}\boldsymbol{\Sigma}_i\boldsymbol{m}, \boldsymbol{\alpha}_4=\sum_{i=1}^{\ell}\boldsymbol{m}^T\boldsymbol{m}_i\boldsymbol{\Sigma}_i\boldsymbol{m},\\ &\boldsymbol{\alpha}_5=\sum_{i=1}^{\ell}\left(\operatorname{tr}\left(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}\right)^T\right)\boldsymbol{m}_i, \boldsymbol{\alpha}_6=\sum_{i=1}^{\ell}\boldsymbol{m}_i^T\boldsymbol{\Sigma}\boldsymbol{m}_i\boldsymbol{m}_i,\\ &\boldsymbol{\alpha}_7=2\sum_{i=1}^{\ell}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_i\boldsymbol{m}_i, \boldsymbol{\beta}_1=\ 2\sum_{i=1}^{\ell}\boldsymbol{m}^T\boldsymbol{m}_i\boldsymbol{m}_i,\\ &\boldsymbol{\beta}_2=2\sum_{i=1}^{\ell}\boldsymbol{\Sigma}_i\boldsymbol{m}, \boldsymbol{\delta}=\sum_{i=1}^{\ell}\boldsymbol{m}_i,\\ &\mathbf{A}_1=\sum_{i=1}^{\ell}\boldsymbol{m}_i^T\boldsymbol{m}\boldsymbol{\Sigma}_i, \mathbf{A}_2=\sum_{i=1}^{\ell}\boldsymbol{m}^T\boldsymbol{m}_i\boldsymbol{m}_i\boldsymbol{m}_i^T,\\ &\mathbf{A}_3=2\sum_{i=1}^{\ell}\boldsymbol{\Sigma}_i\boldsymbol{m}_i\boldsymbol{m}^T, \mathbf{B}_1=\sum_{i=1}^{\ell}\boldsymbol{\Sigma}_i,\\ &\mathbf{B}_2=\sum_{i=1}^{\ell}\boldsymbol{m}_i\boldsymbol{m}_i^T.\end{aligned}$$

And the fused results $\boldsymbol{m}_{\text{cubic}}$ and $\boldsymbol{\Sigma}_{\text{cubic}}$ satisfy $\partial f(\boldsymbol{m},\boldsymbol{\Sigma})/\partial\boldsymbol{m}=0$ and $\partial f(\boldsymbol{m},\boldsymbol{\Sigma})/\partial\boldsymbol{\Sigma}=0$.

From (12), (13) and (14), we can see that the analytic expressions of fused results in both one-dimensional and multi-dimensional cases are trickily derived. Some numerical methods can be employed to solve (12) and find the zero-points of (13) and (14). For (12), the bisection method and Newton's method are adopted to find an approximate solution of m_{cubic}. The gradient descent method can be utilized for obtain the zero points of (13) and (14) to seek the fused results $\boldsymbol{m}_{\text{cubic}}$ and $\boldsymbol{\Sigma}_{\text{cubic}}$.

C. Gaussian Kernel

When a Gaussian kernel is selected, the partial derivatives in the one-dimensional situation are given by

$$\frac{\partial f\left(m,\sigma^2\right)}{\partial m} \propto 2\eta\sum_{i=1}^{\ell}(m-m_i)\lambda_i^{-\frac{3}{2}}\phi_i, \quad (15)$$

$$\begin{aligned}\frac{\partial f\left(m,\sigma^2\right)}{\partial \sigma^2} \propto & -\eta\ell\left(2\eta\sigma^2+1\right)^{-\frac{3}{2}}-\sum_{i=1}^{\ell}(m-m_i)^2\eta^2\phi_i\lambda_i^{-\frac{5}{2}}\\ & +\sum_{i=1}^{\ell}\eta\phi_i\lambda_i^{-\frac{3}{2}}, \quad (16)\end{aligned}$$

where

$$\lambda_i=\eta\sigma^2+\eta\sigma_i^2+1, \phi_i=\exp\left(\frac{-\eta(m-m_i)^2}{2\lambda_i}\right).$$

As for the n-dimensional case, the partial derivatives of the objective function with regard to $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ are

$$\frac{\partial f(\boldsymbol{m},\boldsymbol{\Sigma})}{\partial \boldsymbol{m}} \propto 2\sum_{i=1}^{\ell}\left(\psi_i\left|\boldsymbol{\Lambda}_i\right|^{-\frac{1}{2}}\left(\frac{\boldsymbol{\Lambda}_i}{\eta}\right)^{-1}\boldsymbol{\varepsilon}_i(\boldsymbol{m})\right), \quad (17)$$

$$\begin{aligned}\frac{\partial f(\boldsymbol{m},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} \propto & -\ell\eta\left|\boldsymbol{\Lambda}'\right|^{-\frac{1}{2}}\left(\boldsymbol{\Lambda}'\right)^{-1}-\sum_{i=1}^{\ell}\Big(\psi_i\left|\boldsymbol{\Lambda}_i\right|^{-\frac{1}{2}}\\ & \cdot\left(\frac{\boldsymbol{\Lambda}_i}{\eta}\right)^{-1}\boldsymbol{\varepsilon}_i(\boldsymbol{m})\left(\boldsymbol{\varepsilon}_i(\boldsymbol{m})\right)^T\left(\frac{\boldsymbol{\Lambda}_i}{\eta}\right)^{-1}\\ & -\eta\psi_i\left|\boldsymbol{\Lambda}_i\right|^{-\frac{1}{2}}\boldsymbol{\Lambda}_i^{-1}\Big), \quad (18)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\Lambda}_i&=\eta\boldsymbol{\Sigma}+\eta\boldsymbol{\Sigma}_i+\mathbf{I}_n,\\ \boldsymbol{\Lambda}'&=2\eta\boldsymbol{\Sigma}+\mathbf{I}_n,\\ \boldsymbol{\varepsilon}_i(\boldsymbol{m})&=\boldsymbol{m}-\boldsymbol{m}_i,\\ \psi_i&=\exp\left(-\frac{\boldsymbol{\varepsilon}_i(\boldsymbol{m})^T\left(\frac{\boldsymbol{\Lambda}_i}{\eta}\right)^{-1}\boldsymbol{\varepsilon}_i(\boldsymbol{m})}{2}\right).\end{aligned}$$

Let the partial derivatives (15)–(18) be zero, the solutions are respectively denoted by m_{Gauss} and σ_{Gauss} for the one-dimensional case and by $\boldsymbol{m}_{\text{Gauss}}$ and $\boldsymbol{\Sigma}_{\text{Gauss}}$ for the multi-dimensional case. Although it is difficult to obtain the analytic solutions, one can use numerical algorithms to seek approximate solutions as the same as the cube polynomial kernel case. Specifically, the gradient descent method and Newton's method can be employed to find the zero points of (15)–(18).

V. Numerical Examples

In this section, we present the performance of the proposed method through numerical experiments considered in [25] and [36]. The PDF in the proposed method is available by embedding the first two order moments of the each local estimate into a Gaussian distribution which is reasonable due to maximum entropy principle. The DCI and FCI algorithms are adopted for comparison. The root mean squared error (RMSE) is adopted to evaluate the performance of the fused estimate. It is estimated by averaging 500 Monte Carlo runs.

The free parameter η in the Gauss kernel is considered to be fixed, and let $\eta = 1$.

A. One-Dimensional Static Case

Suppose two local sensors are set to observe a scalar static target with the state $x \sim N\left(\bar{x}, \bar{P}\right)$. For each sensor, the observation equation is

$$z_i = x + v_i, \ \ i = 1, 2,$$

where $[v_1, v_2]$ is zero-mean Gaussian noise with covariance matrix

$$\mathbf{R} = \begin{bmatrix} R_1 & \rho\sqrt{R_1 R_2} \\ \rho\sqrt{R_1 R_2} & R_2 \end{bmatrix},$$

and its components are independent of x. The local sensors can obtain LMMSE estimates $(\hat{x}_i, P_i)$ by given prior information of x and measurements z_i.

In this simulation, the correlation coefficient ρ varies in the interval $(-1, 1)$ and $\bar{x} = 1, \bar{P} = 25, R_1 = 3, R_2 = 6$.

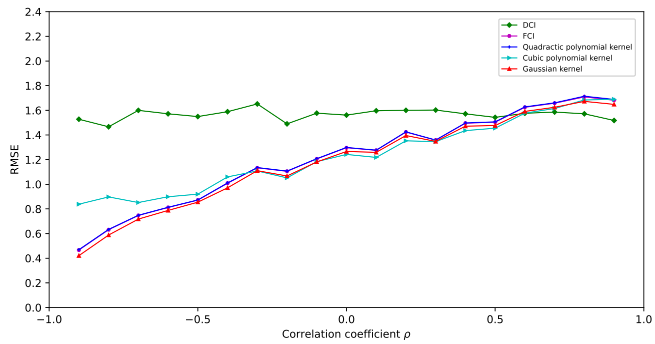

Fig. 1. The RMSEs of fused mean estimates.

The RMSEs of fused estimates by the considered methods are illustrated in Fig. 1. We can see from it that, in contrast to DCI algorithm, all three proposed methods through kernel mean embeddings perform well for negative ρ, and worse performance for large positive ρ. Compared to the FCI algorithm, the proposed method is nearly equal performance for quadratic polynomial kernel, more effective except $\rho \in (-1, -0.3)$ for cubic polynomial kernel, and consistently better for Gaussian kernel.

B. Two-Dimensional Dynamic Case

Consider a dynamic system with two sensors to track some target. The specific system is modeled by

$$\boldsymbol{x}_{k+1} = \begin{bmatrix} 1 & \Delta T \\ 0 & 1 \end{bmatrix} \boldsymbol{x}_k + \begin{bmatrix} (\Delta T)^2/2 \\ \Delta T \end{bmatrix} w_k,$$
$$z_k^i = \mathbf{H}_i \boldsymbol{x}_k^i + v_k^i, \ \ i = 1, 2$$

where k is the time index, the sampling interval $\Delta T = 1$, the components of the state $\boldsymbol{x}_k$ are respectively the position and velocity, $\mathbf{H}_i$ is the measurement matrix of the i-th sensor, w_k is zero-mean Gaussian process noise with variance Q, $\left[v_k^1, v_k^2\right]$ is zero-mean Gaussian measurement noise with covariance matrix $\mathbf{R}_k$, and all process and measurement noises are mutually independent and independent of $\boldsymbol{x}_0$.

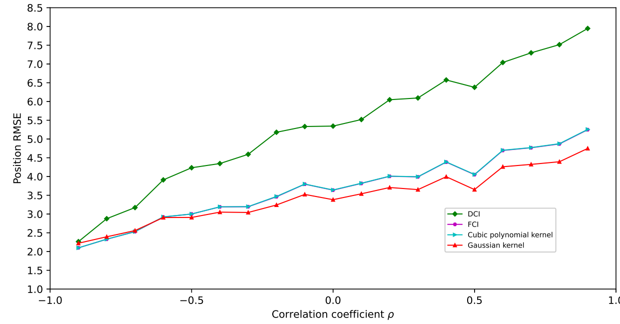

Fig. 2. The RMSEs of the fused position estimates.

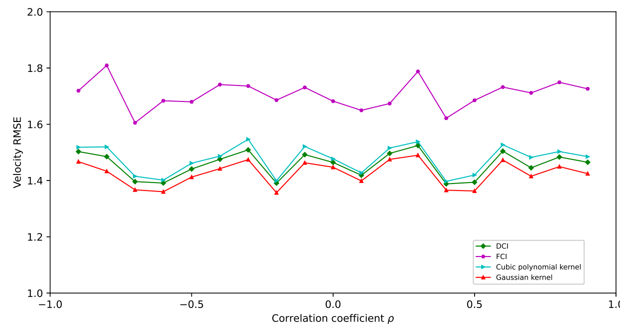

Fig. 3. The RMSEs of the fused velocity estimates.

For the i-th sensor, the local estimate $\boldsymbol{x}_k^i$ and its estimation error matrix $\mathbf{P}_k^i$ at $k\Delta T$ instant are obtained by Kalman filtering, and then the local PDF is $N\left(\boldsymbol{x}_k^i, \mathbf{P}_k^i\right)$ by embedding to the space of Gaussian distributions.

In this simulation, the covariance matrix $\mathbf{R}_k$ is

$$\mathbf{R}_k = \begin{bmatrix} R_k^1 & R_k^{12} \\ R_k^{21} & R_k^2 \end{bmatrix},$$

where $R_k^{12} = R_k^{21} \triangleq \rho\sqrt{R_k^1 R_k^2}$. The initial state $\boldsymbol{x}_0$ is generated from a Gaussian distribution with mean $\bar{\boldsymbol{x}}$ and covariance matrix $\bar{\mathbf{P}}$. Similar to the above state case, we evaluate the RMSEs for different methods at $k = 10$. The concrete parameters are set as

$$\mathbf{H}_1 = [1, 0], \mathbf{H}_2 = [0, 1], Q = 5, R_k^1 = 10, R_k^2 = 2,$$
$$\bar{\boldsymbol{x}} = \begin{bmatrix} 7 \\ 4 \end{bmatrix}, \bar{\mathbf{P}} = \begin{bmatrix} 20 & 0 \\ 0 & 10 \end{bmatrix}.$$

Figs. 2 and 3 show the RMSEs of the fused position and velocity respectively. From Fig. 2, we can observe that, for the position estimation, the proposed methods with cubic polynomial and Gaussian kernels outperform the DCI algorithm for $\rho \in (-1, 1)$. Note that the case for cubic polynomial kernel behaves nearly the same with the FCI algorithm while the performance is totally better than the FCI algorithm for Gaussian kernel except for ρ being negatively large. For the velocity estimation shown in Fig. 3, the proposed method with cubic polynomial kernel is slightly worse than the DCI algorithm but better than the FCI algorithm. The case for Gaussian kernel outperforms the DCI and FCI algorithms consistently. Note that the performance of the proposed method with quadratic polynomial kernel is worse than two classical CI methods which is not presented in figures. One possible reason for this phenomenon is that the simple kernel results in an inaccurate estimate of covariance matrix.

VI. Conclusions

This work provides a framework based on kernel mean embedding for distributed estimation fusion. Specifically, we formulate the optimal fusion into an optimization problem to minimize the sum of the squared distances between distributions in pairs, and adopt the MMD as the distance of two embedded probability densities in an RKHS. Under Gaussian distribution assumption, by analytically deriving the MMDs for three kernels, the objective functions in optimal fusion problems are specified. For the quadratic polynomial kernel, the fused estimate can be expressed analytically, while for the other two kernels, the numerical method can be employed. Simulations show that the proposed fusion methods with three kernels are generally effective, and specially the case of Gaussian kernel consistently outperforms than the DCI and FCI fusion methods.

Further research for more general local PDFs such as ellipsoidal distributions and for heterogeneous distributions under the framework of kernel mean embedding will be done in the future. Specially, more efficient algorithms for solving the related optimization problem will also be considered.

References

[1] Y. Bar-Shalom and X. R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT, USA: YBS Publishing, 1995.

[2] Y. Zhu, *Multisensor Decision and Estimation Fusion*. Boston, MA, USA: Kluwer Academic Publishers, 2002.

[3] M. E. Liggins, D. L. Hall, and J. Llinas, Eds., *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2009.

[4] S. Yu, L.-C. Tranchevent, B. D. Moor, and Y. Moreau, *Kernel-based Data Fusion for Machine Learning: Methods and Applications in Bioinformatics and Text Mining*. Berlin, Germany: Springer, 2011.

[5] P. Siarry, M. Jabbar, R. Aluvalu, A. Abraham, and A. Madureira, Eds., *The Fusion of Internet of Things, Artificial Intelligence, and Cloud Computing in Health Care*. Cham, Switzerland: Springer, 2021.

[6] H. M. La and W. Sheng, "Distributed sensor fusion for scalar field mapping using mobile sensor networks," *IEEE Transactions on Cybernetics*, vol. 43, no. 2, pp. 766–778, Apr. 2013.

[7] B. Chen, W.-A. Zhang, and L. Yu, "Distributed fusion estimation with missing measurements, random transmission delays and packet dropouts," *IEEE Transactions on Automatic Control*, vol. 59, no. 7, pp. 1961–1967, Jul. 2014.

[8] M. A. Bakr and S. Lee, "Distributed multisensor data fusion under unknown correlation and data inconsistency," *Sensors*, vol. 17, no. 11, p. 2472, Oct. 2017.

[9] X. Su, K. Fan, and W. Shi, "Privacy-preserving distributed data fusion based on attribute protection," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5765–5777, Oct. 2019.

[10] Y. Bar-Shalom, "On the track-to-track correlation problem," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 571–572, Apr. 1981.

[11] Y. Bar-Shalom and L. Campo, "The effect of the common process noise on the two-sensor fused-track covariance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 6, pp. 803–805, Nov. 1986.

[12] X. R. Li, Y. Zhu, J. Wang, and C. Han, "Optimal linear estimation fusion. I. unified fusion rules," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2192–2208, Sep. 2003.

[13] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proceedings of the American Control Conference*, Albuquerque, NM, USA, Jun. 1997, pp. 2369–2373.

[14] L. Chen, P. Arambel, and R. Mehra, "Estimation under unknown correlation: covariance intersection revisited," *IEEE Transactions on Automatic Control*, vol. 47, no. 11, pp. 1879–1882, Nov. 2002.

[15] M. B. Hurley, "An information theoretic justification for covariance intersection and its generalization," in *Proceedings of the Fifth International Conference on Information Fusion*, Annapolis, MD, USA, Jul. 2002, pp. 505–511.

[16] Y. Wang and X. R. Li, "A fast and fault-tolerant convex combination fusion algorithm under unknown cross-correlation," in *Proceedings of the 12th International Conference on Information Fusion*, Seattle, WA, USA, Jul. 2009, pp. 571–578.

[17] M. Üney, D. E. Clark, and S. J. Julier, "Distributed fusion of PHD filters via exponential mixture densities," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 521–531, Jun. 2013.

[18] Y. Cai and L.-H. Lim, "Distances between probability distributions of different dimensions," *IEEE Transactions on Information Theory*, vol. 68, no. 6, pp. 4020–4031, Jun. 2022.

[19] G. Koliander, Y. El-Laham, P. M. Djuric, and F. Hlawatsch, "Fusion of probability density functions," *Proceedings of the IEEE*, vol. 110, no. 4, pp. 404–453, Apr. 2022.

[20] P. Wu, T. Imbiriba, V. Elvira, and P. Closas, "Bayesian data fusion with shared priors," *IEEE Transactions on Signal Processing*, vol. 72, pp. 275–288, 2024.

[21] C. Genest and J. V. Zidek, "Combining probability distributions: A critique and an annotated bibliography," *Statistical Science*, vol. 1, no. 1, pp. 114–135, Feb. 1986.

[22] R. T. Stewart and I. O. Quintana, "Probabilistic opinion pooling with imprecise probabilities," *Journal of Philosophical Logic*, vol. 47, pp. 17–45, Jan. 2017.

[23] G. Battistelli and L. Chisci, "Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, Mar. 2014.

[24] W. Li and F. Yang, "Information fusion over network dynamics with unknown correlations: An overview," *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 2, Jun. 2023, Art. no. 100003.

[25] M. Tang, Y. Rong, J. Zhou, and X. R. Li, "Information geometric approach to multisensor estimation fusion," *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 279–292, Jan. 2019.

[26] X. Chen and J. Zhou, "Multisensor estimation fusion on statistical manifold," *Entropy*, vol. 24, Dec. 2022, Art. no. 1802.

[27] P. A. Morris, "Combining expert judgments: A Bayesian approach," *Management Science*, vol. 23, no. 7, pp. 679–693, Mar. 1977.

[28] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.

[29] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. New York, NY, USA: Springer, 2004.

[30] A. Müller, "Integral probability metrics and their generating classes of functions," *Advances in Applied Probability*, vol. 29, no. 2, pp. 429–443, Jun. 1997.

[31] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. R. G. Lanckriet, "On the empirical estimation of integral probability metrics," *Electronic Journal of Statistics*, vol. 6, pp. 1550–1599, Jan. 2012.

[32] K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf, "Kernel mean embedding of distributions: A review and beyond," *Foundations and Trends in Machine Learning*, vol. 10, no. 1–2, pp. 1–141, Jun. 2017.

[33] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. G. Lanckriet, "Hilbert space embeddings and metrics on probability measures," *The Journal of Machine Learning Research*, vol. 11, pp. 1517–1561, Aug. 2010.

[34] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006.

[35] L. Song, X. Zhang, A. Smola, A. Gretton, and B. Schölkopf, "Tailoring density estimation via reproducing kernel moment matching," in *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, Finland, Jul. 2008, pp. 992–999.

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