

Multisensor Estimation Fusion on Statistical Manifold

Xiangbing Chen ¹  and Jie Zhou ^{2,*} ¹ Division of Mathematics, Sichuan University Jinjiang College, Meishan 620860, China² College of Mathematics, Sichuan University, Chengdu 610064, China

* Correspondence: jzhou@scu.edu.cn

Abstract: In the paper, we characterize local estimates from multiple distributed sensors as posterior probability densities, which are assumed to belong to a common parametric family. Adopting the information-geometric viewpoint, we consider such family as a Riemannian manifold endowed with the Fisher metric, and then formulate the fused density as an informative barycenter through minimizing the sum of its geodesic distances to all local posterior densities. Under the assumption of multivariate elliptical distribution (MED), two fusion methods are developed by using the minimal Manhattan distance instead of the geodesic distance on the manifold of MEDs, which both have the same mean estimation fusion, but different covariance estimation fusions. One obtains the fused covariance estimate by a robust fixed point iterative algorithm with theoretical convergence, and the other provides an explicit expression for the fused covariance estimate. At different heavy-tailed levels, the fusion results of two local estimates for a static target display that the two methods achieve a better approximate of the informative barycenter than some existing fusion methods. An application to distributed estimation fusion for dynamic systems with heavy-tailed process and observation noises is provided to demonstrate the performance of the two proposed fusion algorithms.

Keywords: distributed estimation fusion; elliptical distribution; information geometry; Manhattan distance; lie algebra

**Citation:** Chen, X.; Zhou, J.Multisensor Estimation Fusion on Statistical Manifold. *Entropy* **2022**, *24*, 1802. <https://doi.org/10.3390/e24121802>

Academic Editor: Geert Verdoolaege

Received: 22 October 2022

Accepted: 6 December 2022

Published: 9 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multisensor data fusion has been studied and widely applied to many important areas, such as image processing [1,2], wireless sensor networks [3,4] and remote sensing [5,6]. The research mainly centers on two basic architectures, i.e., centralized fusion and distributed fusion. The latter has less communication burden, higher flexibility and reliability, but challenges in dealing with the cross-correlation among local estimation errors caused by common process noises, related measurement noises and past data exchanges [7–9].

Considering the limitations of traditional Bayesian estimation methods such as some Kalman filtering variants, many estimation fusion methods only utilize the first two moments (i.e., mean and covariance) of each local posterior density for data fusion. For instance, the covariance intersection (CI) fusion method [10,11] propagates each local information pair consisting of the information matrix (i.e., the inverse covariance matrix) and the information vector (i.e., the state estimate left multiplied by the information matrix), and then provides a convex combination of all local information pair with specified normalized weights to yield a conservative fused estimate. Specifically, the determinant-minimization CI (DCI) method [12] selects weights by minimizing the trace or the determinant of fused covariance matrix, which also provided a set-theoretic interpretation [13]. Applying the set-theoretic criterion, some novel set-theoretic methods such as the relaxed Chebyshev center CI (RCC-CI) [14] and the analytic center CI (AC-CI) [15] take both the local state estimate and covariance matrix of estimation error into account to obtain the optimal weights, but the DCI only considers the latter.

The probability density function (PDF) constitutes a complete probabilistic description of the state estimate (i.e., the point estimate of the quantity of concern) from each

sensor, which includes a lot of important information, such as support, tail decay and multimodality, so it is more suitable for fusion than the first two moments. We note that there are already many fusion methods which combine the local estimates together in the form of PDFs. As examples, under Gaussian assumptions, ref. [16] minimizes the Shannon entropy of a log-linear combination of all local PDFs, and provides the fused estimates with the same form as the DCI, but in general with different weights; the fast CI (FCI) algorithm [17] finds the “halfway” point between two local PDFs through minimizing Chernoff information, which has been generalized to the case of multiple sensors; The Kullback–Leibler averaging (KLA) algorithm [18] seeks an average PDF to minimize the sum of KL divergences between such average and local posterior densities.

Although many PDF fusion rules have been established and found to be useful in specific applications, there is no consistent theoretical basis for the principles of these rules and their possible alternatives. Moreover, ref. [19] illustrates that the space of PDFs is generally not Euclidean. For these reasons, information geometry regards the parametric space of PDFs as a statistical manifold with a Riemannian structure, and then studies the intrinsic properties by the tools of modern differential geometry. The theoretical framework can be exploited for the PDF fusion by assuming that all local posterior PDFs as points belong to a common Riemannian manifold. For example, ref. [20] formulates the fusion result as the Wasserstein barycenter by minimizing the sum of its squared Wasserstein distances to Gaussian inputs. The Wasserstein distance as a geodesic distance on the Riemannian manifold equipped with the Wasserstein metric is usually used for optimal transportation [21,22], while the Fisher metric applies primarily to information science. Therefore, by endowing such space with a natural Riemannian structure (i.e., the Fisher metric and Levi–Civita connection), the resulting geodesic distance, also called the Rao distance, has been taken as an intrinsic measure for the dissimilarity between two PDFs and then applied in wide fields such as neural networks [23,24], signal processing [25,26], and statistical inference [27,28].

In addition, due to the intrinsic mechanism of applications, uncertain modeling errors and the existence of outliers, the class of multivariate elliptical distributions (MEDs) including multivariate Gaussian distribution (MGD), multivariate generalized Gaussian distribution (MGGD) [29], multivariate t -distribution (MTD) [30], symmetric multivariate Laplace distribution [31], contaminated Gaussian distribution, Gaussian–Student’s t mixture distribution and so on, has enjoyed a wide range of applications [32–41]. To our knowledge, the application of information geometry in multisensor estimation fusion has not been studied in depth except two previous papers [42,43]. The geodesic projection (GP) method [42] for distributed estimation fusion, minimizing the sum of geodesic projection distances onto the Gaussian submanifold with a fixed mean vector, formulates the fused PDF as an informative barycenter of the Gaussian manifold. And the QMMHD fusion algorithm [43] extends the PDF fusion on the Gaussian manifold to the MED manifold by minimizing the sum of Manhattan distances (MHDs) between the fused density and each local posterior density on the Riemannian manifold of MEDs; however, it suffers from two major drawbacks. One is that the covariance estimate of the fused PDF is not specially designed, just using the same form as the CI estimate, and the other is that the convergence of the QMMHD algorithm can not be guaranteed in theory. At present, an in-depth research on the PDF fusion of non-Gaussian manifolds is quite lacking. Our goal is to develop efficient distributed estimation fusion algorithms for various scenarios and applications under a unified MED fusion framework, rather than considering a single fusion rule or method for a particular MED.

In the work, using the minimal MHD instead of the geodesic distance as the loss function of geometric fusion criterion, we propose two distributed fusion methods under the MED assumptions. The main contributions are summarized as follows:

- (i) To exploit the non-Euclidean characteristics of probabilistic space and the decouple feature of Manhattan distance, we formulate a novel information-geometric criterion for fusion and discuss its inherent advantages.

- (ii) We derive the explicit expression for the MHD-based projection from a point onto the MED submanifold with a fixed mean, and then develop two fusion methods MHDP-I and MHDP-E by relaxing the fusion criterion, which both have the same mean estimation fusion, but differ in the form of the fused covariance. A fixed point iteration is given to obtain this fused mean.
- (iii) The MHDP-I obtains the fused covariance using a robust fixed point iterative algorithm with theoretical convergence, while the MHDP-E provides an explicit expression for the fused covariance by introducing a Lie algebraic structure. We also provide a crucial theorem presenting the exponential mapping and logarithmic mapping on the MED submanifold with a fixed mean.
- (iv) Simulations indicate that the two proposed information-geometric methods outperform some existing fusion methods under non-Gaussian distributions.

The outline of this paper is organized as follows. In Section 2, the basic facts for information geometry and some useful results concerning the MED manifold are introduced. Section 3 formulates an information-geometric fusion criterion based on the minimal MHD, and then proposes the MHDP-I and MHDP-E fusion methods in Section 4. Numerical examples in Section 5 are provided to demonstrate the superiority of the two methods. Section 6 gives a conclusion. All proofs are given in the Appendices A–F.

Notations

Throughout this paper, we use lightface letters to represent scalars and scalar-valued mappings, boldface lowercase letters to represent vectors, and boldface capital letters to represent matrices and matrix-valued mappings. All vectors are column vectors. The notation \mathbb{S}_+^m is the set of $m \times m$ real symmetric positive-definite matrices, \mathbb{R}_+ is the space of all positive real numbers, and \mathbb{R}^m denotes the set of all m -dimensional real column vectors. The symbol \mathbf{I}_m stands for the identity matrix of order m . For a matrix \mathbf{A} , \mathbf{A}^T and $\text{tr}(\mathbf{A})$ denote its transpose and trace, respectively. Moreover, $\text{Exp}(\cdot)$, $\text{Log}(\cdot)$ and $\text{arccosh}(\cdot)$ are matrix exponential, matrix logarithm and inverse hyperbolic functions, respectively.

2. Preliminaries

In this section, we review basic notions in the fields of information geometry and present some important results for later use.

2.1. Statistical Manifold and Fisher Metric

Consider a statistical model

$$\mathcal{S} = \left\{ p(\mathbf{x}; \boldsymbol{\theta}) : \boldsymbol{\theta} = (\theta^1, \dots, \theta^m) \in \Theta \right\} \quad (1)$$

of probability densities with the global coordinate system $\boldsymbol{\theta} = (\theta^1, \dots, \theta^m)$, where Θ is an open set of \mathbb{R}^m . Using the Fisher metric g induced by the $m \times m$ Fisher information matrix with the (i, j) -th entry

$$g_{ij}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \theta^i} \log p(\mathbf{x}; \boldsymbol{\theta}) \frac{\partial}{\partial \theta^j} \log p(\mathbf{x}; \boldsymbol{\theta}) \right], \quad (2)$$

where $\mathbb{E}_{\boldsymbol{\theta}}[\cdot]$ represents the expectation operator with respect to $p(\mathbf{x}; \boldsymbol{\theta})$, the statistical manifold \mathcal{S} can be endowed with a natural Riemannian differentiable structure [44]. For brevity, we shall abbreviate the point $p(\mathbf{x}; \boldsymbol{\theta})$ of \mathcal{S} as its coordinate $\boldsymbol{\theta}$.

Let $T_{\boldsymbol{\theta}}\mathcal{S}$ be the tangent vector space at the point $\boldsymbol{\theta}$ in \mathcal{S} , and an inner product on $T_{\boldsymbol{\theta}}\mathcal{S}$ is then defined by the Fisher metric g , written as

$$\langle \cdot, \cdot \rangle_{\boldsymbol{\theta}} : T_{\boldsymbol{\theta}}\mathcal{S} \times T_{\boldsymbol{\theta}}\mathcal{S} \rightarrow \mathbb{R}. \quad (3)$$

Thus, the geodesic distance between two points θ_0 and θ_1 in \mathcal{S} , also called the Rao distance or Fisher information distance, can be obtained through solving the following functional minimization problem

$$\ell_F(\theta_0, \theta_1) := \min_{\theta \in \Gamma} \int_0^1 \sqrt{\langle \dot{\theta}(t), \dot{\theta}(t) \rangle_{\theta(t)}} dt, \tag{4}$$

where the dot symbol over a variable signifies its derivative with respect to $t \in \mathbb{R}$, and the set Γ consists of all piecewise smooth curves linking θ_0 to θ_1 .

For any $\nu \in T_{\theta_0}\mathcal{S}$, there exists a unique geodesic segment $\gamma(t; \theta_0, \nu), t \in [0, 1]$, satisfying $\gamma(0) = \theta_0$ and $\dot{\gamma}(0) = \nu$, and then the exponential mapping is defined as

$$\begin{aligned} \exp_{\theta_0} : T_{\theta_0}\mathcal{S} &\rightarrow \mathcal{S} \\ \nu &\mapsto \exp_{\theta_0} \nu = \gamma(1; \theta_0, \nu). \end{aligned} \tag{5}$$

Conversely, the logarithmic mapping $\log_{\theta_0}(\theta_1)$ maps θ_1 into the tangent vector ν at θ_0 . As a generalization, the manifold retraction $R(\cdot)$ is a smooth mapping from the tangent bundle $T\mathcal{S} = \{T_{\theta}\mathcal{S} : \theta \in \Theta\}$ into \mathcal{S} (see, e.g., [45]), and its restriction $R_{\theta}(\cdot)$ on an open ball $B(\mathbf{0}, r)$ with radius r in $T_{\theta}\mathcal{S}$ satisfies $R_{\theta}(\mathbf{0}) = \theta$ and $dR_{\theta}|_0 = \text{id}_{T_{\theta}\mathcal{S}}$. Moreover, its inverse mapping $R_{\theta}^{-1}(\cdot)$, called the lifting mapping, exists in a neighborhood of θ . In particular, the exponential mapping and logarithmic mapping are the commonly used retraction mapping and lifting mapping, respectively.

2.2. Information Geometry of Elliptical Distributions

An m -dimensional random vector x has multivariate elliptical distribution $\text{EL}_h^m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its probability density function is of the form

$$p_h(x; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} h\left((x - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x - \boldsymbol{\mu})\right) \tag{6}$$

with some generating function $h(\cdot)$, where $\boldsymbol{\mu} \in \mathbb{R}^m$ is mean vector and $\boldsymbol{\Sigma} \in \mathbb{S}_+^m$ is positive-definite scatter matrix. From [46], \mathbf{x} has covariance matrix $\kappa_h \boldsymbol{\Sigma}$ with scale parameter κ_h . As two examples, the generating functions of the MGGD with shape parameter β and the MTD with ν degrees of freedom, respectively, are

$$h(u) = \frac{\beta \Gamma(m/2)}{\pi^{m/2} \Gamma(m/(2\beta)) 2^{m/(2\beta)}} \exp\left(-\frac{1}{2} u^\beta\right) \tag{7}$$

and

$$h(u) = \frac{\Gamma((\nu + m)/2)}{(\pi \nu)^{m/2} \Gamma(\nu/2)} \left(1 + \frac{u}{\nu}\right)^{-(\nu+m)/2}. \tag{8}$$

Suppose that $\boldsymbol{\Omega}$ is the lower-triangle Cholesky factor of $\boldsymbol{\Sigma}$, $\mathbf{z} = \boldsymbol{\Omega}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ and $\xi = d \log(h(\|\mathbf{z}\|^2)) / d(\|\mathbf{z}\|^2)$. Denote

$$a_h = \frac{1}{m} \mathbb{E}[\xi^2 \|\mathbf{z}\|^2], \quad b_h = \frac{1}{m(m+2)} \mathbb{E}[\xi^2 \|\mathbf{z}\|^4]. \tag{9}$$

It is easy to deduce that

- (i) for the MGGD, $a_h = \beta^2 \Gamma(2 + m/(2\beta) - 1/\beta) / (2^{1/\beta} m \Gamma(m/(2\beta)))$, $b_h = (m + 2\beta) / (4m + 8)$;
- (ii) and for the MTD, $a_h = b_h = (\nu + m) / (4(\nu + m + 2))$.

Remark 1. For other classical elliptical distributions, such as the Pearson-type VII class of distributions, the multivariate Cauchy distribution and a special subfamily of MEDs, the reference [47] has derived the analytic forms for a_h and b_h . However, in practical applications, some special MEDs

(e.g., the Gaussian mixture distribution and the Gaussian–Student’s t mixture distribution [41]) used to model the non-Gaussian distributions may not have their analytical forms, so we can use (9) to numerically calculate the two parameters, or approximate these MEDs using some specified classical elliptical distributions.

By introducing a Riemannian structure associated with the Fisher metric, the MED manifold

$$\mathcal{M} = \{EL_h^m(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^m, \boldsymbol{\Sigma} \in \mathbb{S}_+^m\} \tag{10}$$

with the same fixed function h can be regarded as a Riemannian manifold with coordinate system $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Note that the dimension of \mathcal{M} is $m + m(m + 1)/2$. From [47], the squared line element at a point $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in \mathcal{M} is

$$ds^2 = 4a_h d\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} d\boldsymbol{\mu} + 2b_h \text{tr}(\boldsymbol{\Sigma}^{-1} d\boldsymbol{\Sigma})^2 + \frac{4b_h - 1}{4} \text{tr}^2(\boldsymbol{\Sigma}^{-1} d\boldsymbol{\Sigma}), \tag{11}$$

and the geodesic equations are given by

$$\ddot{\boldsymbol{\mu}} - \dot{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{\mu}} = 0, \tag{12}$$

$$\ddot{\boldsymbol{\Sigma}} + a \dot{\boldsymbol{\mu}} \dot{\boldsymbol{\mu}}^T - b \dot{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{\mu}} \boldsymbol{\Sigma} - \dot{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{\Sigma}} = 0, \tag{13}$$

where $a = a_h/b_h$ and $b = a_h(4b_h - 1)/((8 + 4m)b_h^2 - mb_h)$.

Next, we derive the explicit expressions of the geodesic distance and geodesic curve between two given endpoints on the two-dimensional MED manifold \mathcal{M} (i.e., $m = 1$) with coordinate system (μ, σ^2) , which will be used in Section 5.1.

Theorem 1. Consider the two-dimensional MED manifold \mathcal{M} and two points $p_1 = EL_h^1(\mu_1, \sigma_1^2)$ and $p_2 = EL_h^1(\mu_2, \sigma_2^2)$ in \mathcal{M} . Let $c_h = 8a_h/(12b_h - 1)$ and $d_h = \sqrt{12b_h - 1}$, then the geodesic distance between p_1 and p_2 is

$$\ell_F = d_h \cdot \text{arccosh} \left(\frac{c_h(\mu_2 - \mu_1)^2 + 2(\sigma_1^2 + \sigma_2^2)}{4\sigma_1\sigma_2} \right). \tag{14}$$

Additionally, the geodesic curve between p_1 and p_2 is given by

(i) If $\mu_1 = \mu_2$, then

$$\mu(s) = \mu_1 = \mu_2, \tag{15}$$

$$\sigma^2(s) = \sigma_1^2 \cdot \exp(2s/d_h); \tag{16}$$

(ii) If $\mu_1 \neq \mu_2$, then

$$\mu(s) = \frac{1}{\sqrt{c_h}} (\delta_1 + 2\delta_2 \tanh(s/d_h + \varepsilon)), \tag{17}$$

$$\sigma^2(s) = 2\delta_2^2 \cosh^{-2}(s/d_h + \varepsilon), \tag{18}$$

where,

$$\delta_1 = \frac{2\sigma_2^2 + c_h\mu_2^2 - 2\sigma_1^2 - c_h\mu_1^2}{2\sqrt{c_h}(\mu_2 - \mu_1)}, \tag{19}$$

$$\delta_2 = \frac{1}{2} \sqrt{2\sigma_1^2 + (\sqrt{c_h}\mu_1 - \delta_1)^2}, \tag{20}$$

$$\varepsilon = \text{arctanh} \left(\frac{\sqrt{c_h}\mu_1 - \delta_1}{2\delta_2} \right). \tag{21}$$

Proof. See Appendix A. \square

However, in the general case for $m > 1$, how to obtain the closed form of Rao distance on the MED manifold \mathcal{M} remains an unsolved problem, except for some submanifolds with a constant mean vector or constant scatter matrix (see, e.g., [48]).

2.3. Some Submanifolds of Elliptical Distributions

One of the commonly used submanifolds of \mathcal{M} is

$$\mathcal{M}_{[\mu_0, \cdot]} = \{EL_h^m(\mu_0, \Sigma) : \Sigma \in \mathbb{S}_+^m\} \tag{22}$$

having a fixed mean μ_0 and the coordinate system (μ_0, Σ) or Σ for short. It is a totally geodesic submanifold, meaning that each geodesic of $\mathcal{M}_{[\mu_0, \cdot]}$ is also a geodesic of \mathcal{M} . As shown in [48], for any two tangent vectors $v_1, v_2 \in T_\theta(\mathcal{M}_{[\mu_0, \cdot]})$ at a point $\theta = (\mu_0, \Sigma)$, the Fisher inner product is given by

$$\langle v_1, v_2 \rangle_\theta = 2b_h \text{tr}(\Sigma^{-1} v_1 \Sigma^{-1} v_2) + \frac{4b_h - 1}{4} \text{tr}(\Sigma^{-1} v_1) \text{tr}(\Sigma^{-1} v_2), \tag{23}$$

and the Rao distance between two points $\theta_0 = (\mu_0, \Sigma_0)$ and $\theta_1 = (\mu_0, \Sigma_1)$ in $\mathcal{M}_{[\mu_0, \cdot]}$ equals

$$\ell_\Sigma(\theta_0, \theta_1) = \sqrt{2b_h \text{tr}(\text{Log}^2(\Sigma_0^{-1} \Sigma_1)) + \frac{4b_h - 1}{4} \text{tr}^2(\text{Log}(\Sigma_0^{-1} \Sigma_1))}. \tag{24}$$

Another special submanifold is defined as

$$\mathcal{M}_{[\cdot, \alpha \Sigma_0]} = \{EL_h^m(\mu, \alpha \Sigma_0) : \mu \in \mathbb{R}^m, \alpha \in \mathbb{R}_+\} \tag{25}$$

with a fixed scatter matrix $\Sigma_0 \in \mathbb{S}_+^m$ and the coordinate system $(\mu, \alpha \Sigma_0)$ or (μ, α) for short. As [43], an explicit expression for the Rao distance on $\mathcal{M}_{[\cdot, \alpha \Sigma_0]}$ is given as follows.

Lemma 1. Given two points $\theta_0 = (\mu_0, \Sigma_0)$ and $\theta_\alpha = (\mu, \alpha \Sigma_0)$ in $\mathcal{M}_{[\cdot, \alpha \Sigma_0]}$, let $c = 8a_h / (4m(m + 2)b_h - m^2)$ and $d_M = (\mu_0 - \mu)^T \Sigma_0^{-1} (\mu_0 - \mu)$, then on $\mathcal{M}_{[\cdot, \alpha \Sigma_0]}$, the Rao distance between θ_0 and θ_α is

$$\ell_\mu(\theta_0, \theta_\alpha) = \sqrt{\frac{2a_h}{c} \text{arccosh}\left(\frac{(cd_M + 2)^2}{8\alpha} + \frac{\alpha}{2} + \frac{cd_M}{2}\right)}. \tag{26}$$

2.4. Manhattan Distances

At present, there is no explicit expression for the Rao distance on the MED manifold \mathcal{M} . Intuitively motivated from the commonly adopted distance to measure the length of path along the coordinate curves on \mathcal{M} as the ℓ_1 distance in Euclidean space, we introduce an intermediate point $\theta_\alpha = (\mu_1, \alpha \Sigma_0) \in \mathcal{M}$ to link $\theta_0 = (\mu_0, \Sigma_0)$ and $\theta_1 = (\mu_1, \Sigma_1)$ for $\alpha \in \mathbb{R}^+$, and then construct a one-parametric class $\mathcal{A} = \{\ell_\alpha(\theta_0, \theta_1) : \alpha \in \mathbb{R}_+\}$ of MHDs from two Rao distances, where each member in \mathcal{A} satisfies

$$\ell_\alpha(\theta_0, \theta_1) := \ell_\mu(\theta_0, \theta_\alpha) + \ell_\Sigma(\theta_\alpha, \theta_1) \geq \ell_F(\theta_0, \theta_1). \tag{27}$$

Moreover, for providing tighter upper bounds for the Rao distance $\ell_F(\theta_0, \theta_1)$ on \mathcal{M} , the minimal MHD

$$\ell_{\hat{\alpha}}(\theta_0, \theta_1) = \min_{\alpha \in \mathbb{R}_+} \ell_\alpha(\theta_0, \theta_1) \tag{28}$$

as the minimum in the class \mathcal{A} can be obtained by optimally seeking the parameter $\hat{\alpha}$.

3. Fusion Criterion

Consider a distributed system with n sensors observing a common state $\mathbf{x} \in \mathbb{R}^m$. As many practical scenarios for dynamic target tracking in [49,50], the dynamic system is assumed having the heavy-tailed process and observation noises. We then adopt the MED assumption to better exploit heavy-tailed features inherent in noises, and denote by $p_k = \text{EL}_h^m(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ the local posterior density from the k -th sensor. The goal is to fuse all p_k into a single fused density with unavailable correlations among local sensors.

3.1. Information-Geometric Criterion

In practice, the Kullback–Leibler divergence (KLD) and Rao distance are widely accepted as basic information theoretic quantities that can capture higher-order statistical information, but the former is not a true distance in mathematics owing to the lack of triangle inequality and symmetry. Instead, as a natural intrinsic measure, the Rao distance not only enables us to deal with statistical issues while respecting the nonlinear geometry structure of MEDs, but is also invariant under coordinate transformations. From the viewpoint of information geometry, the geometrical illustration develops an intrinsic understanding of statistical models, and also provides a better avenue for estimation fusion. Then, taking the sum of geodesic distances $\ell_F(p_k, p)$ between $p = \text{EL}_h^m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and all p_k in the MED manifold \mathcal{M} as the cost function, we formulate a fusion criterion as

$$(\hat{\mathbf{x}}, \mathbf{P}) = \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^m, \boldsymbol{\Sigma} \in \mathbb{S}_+^m} \sum_{k=1}^n \ell_F(p_k, p) \tag{29}$$

to fuse all local available information p_k into a single posterior density $\hat{p} = \text{EL}_h^m(\hat{\mathbf{x}}, \mathbf{P})$.

The aforementioned inherent feature of the Rao distance ensures a unique fusion result in the multisensor fusion network involving different measurement frameworks. However, due to the absence of the closed form of Rao distance on the manifold of MEDs, the minimal MHD (28) as its tight upper bound is applied in the fusion criterion (29), that is,

$$\begin{aligned} (\hat{\mathbf{x}}, \mathbf{P}) &= \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^m, \boldsymbol{\Sigma} \in \mathbb{S}_+^m} \sum_{k=1}^n \min_{\alpha_k \in \mathbb{R}_+} \ell_{\alpha_k}(p_k, p) \\ &= \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^m, \boldsymbol{\Sigma} \in \mathbb{S}_+^m} \sum_{k=1}^n \min_{\alpha_k \in \mathbb{R}_+} (\ell_{\boldsymbol{\mu}}(p_k, p_{\alpha_k}) + \ell_{\boldsymbol{\Sigma}}(p_{\alpha_k}, p)), \end{aligned} \tag{30}$$

where the intermediate points $p_{\alpha_k} = \text{EL}_h^m(\boldsymbol{\mu}, \alpha_k \mathbf{P}_k)$.

The advantages of determining the minimal MHD to measure the dissimilarity between two MEDs are obvious:

- (i) The minimal MHD (28) is constructed from two Rao distances on two submanifolds (22) and (25) by inserting an intermediate point, thus having the decouple feature and inheriting some good properties from the Rao distance.
- (ii) As a tight upper bound for the Rao distance on the MED manifold, the minimal MHD (28) can be efficiently computed numerically. Moreover, its superior approximation performance has been fully verified by comparison with the Rao distances on the manifolds of MTDs and MGGDs (see [43] for more details).

3.2. Decoupling Fusion Criterion

Define the MHD-based projection distance (MHDPD) from a point $p_0 = \text{EL}_h^m(\hat{\mathbf{x}}_0, \mathbf{P}_0)$ in \mathcal{M} onto the submanifold $\mathcal{M}_{[\boldsymbol{\mu}, \cdot]}$ along the geodesic in $\mathcal{M}_{[\cdot, \alpha \mathbf{P}_0]}$ as

$$\ell_M(p_0, \mathcal{M}_{[\boldsymbol{\mu}, \cdot]}) := \min_{\alpha \in \mathbb{R}_+} \ell_{\boldsymbol{\mu}}(p_0, p_{\alpha}), \tag{31}$$

where $p_{\alpha} = \text{EL}_h^m(\boldsymbol{\mu}, \alpha \mathbf{P}_0)$.

Before further solving the optimization problem (30), we firstly derive the explicit expressions of the MHD-based projection (MHDP) from a point onto the submanifold $\mathcal{M}_{[\mu, \cdot]}$ and the corresponding MHDPD.

Theorem 2. The MHDPD from $p_0 = \text{EL}_{h'}^m(\hat{\mathbf{x}}_0, \mathbf{P}_0)$ in \mathcal{M} onto the submanifold $\mathcal{M}_{[\mu, \cdot]}$ is given as

$$\ell_M(p_0, \mathcal{M}_{[\mu, \cdot]}) = \sqrt{\frac{2a_h}{c}} \operatorname{arccosh}(cd_0 + 1) \tag{32}$$

with $d_0 = (\hat{\mathbf{x}}_0 - \boldsymbol{\mu})^T \mathbf{P}_0^{-1} (\hat{\mathbf{x}}_0 - \boldsymbol{\mu})$, and the corresponding MHDP is $p_{\hat{\alpha}_0} = \text{EL}_{h'}^m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the scatter matrix

$$\boldsymbol{\Sigma} = \hat{\alpha}_0 \mathbf{P}_0 \tag{33}$$

and the optimal scale factor

$$\hat{\alpha}_0 = \frac{c}{2} d_0 + 1. \tag{34}$$

Proof. See Appendix B. \square

As an application to distributed dynamic system, each available local estimate $(\hat{\mathbf{x}}_i, \mathbf{P}_i) \in \mathbb{R}^m \times \mathbb{S}_+^m$ should be limited to a bounded local feasible set S_i around the true location $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with high reliability (see, e.g., [51]). Therefore by the property of continuous function on compact set, it is straightforward to decompose the optimization (30) into two steps—first over $\boldsymbol{\mu}$ and then over $\boldsymbol{\Sigma}$:

$$\min_{\boldsymbol{\mu} \in \mathbb{R}^m} \min_{\boldsymbol{\Sigma} \in \mathbb{S}_+^m} \sum_{k=1}^n \min_{\alpha_k \in \mathbb{R}_+} (\ell_{\boldsymbol{\mu}}(p_k, p_{\alpha_k}) + \ell_{\boldsymbol{\Sigma}}(p_{\alpha_k}, p)). \tag{35}$$

Denote two functions of the variable $\boldsymbol{\mu} \in \mathbb{R}^m$ as follows:

$$d_k(\boldsymbol{\mu}) = (\hat{\mathbf{x}}_k - \boldsymbol{\mu})^T \mathbf{P}_k^{-1} (\hat{\mathbf{x}}_k - \boldsymbol{\mu}), \tag{36}$$

$$\hat{\alpha}_k(\boldsymbol{\mu}) = \frac{c}{2} d_k(\boldsymbol{\mu}) + 1. \tag{37}$$

Similar to the relaxing strategy in [42], we solve the optimization (35) by successively optimizing $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$. Specifically, exchanging the summation and the minimization over $\boldsymbol{\Sigma}$ and simplifying the objective function in the summation, (35) is then reduced to

$$\hat{\mathbf{x}} = \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^m} \sum_{k=1}^n \min_{\alpha_k \in \mathbb{R}_+} \ell_{\boldsymbol{\mu}}(p_k, p_{\alpha_k}), \tag{38}$$

or equivalently from (31), (32) and (34),

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^m} \sum_{k=1}^n \ell_M(p_k, \mathcal{M}_{[\boldsymbol{\mu}, \cdot]}) \\ &= \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^m} \sum_{k=1}^n \operatorname{arccosh}(cd_k(\boldsymbol{\mu}) + 1), \end{aligned} \tag{39}$$

where $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_n)$ with

$$\hat{\alpha}_k = \hat{\alpha}_k(\hat{\mathbf{x}}) = \frac{c}{2} d_k(\hat{\mathbf{x}}) + 1, \quad k = 1, \dots, n. \tag{40}$$

In general, (39) is not equivalent to the original optimization problem (35) owing to its providing a lower bound for the minimum of the objective function in (35), but it makes good sense for the mean fusion based on the following insights:

- (i) As depicted in Figure 1, we replace the local posterior densities p_k with the MHDPs $p_{\hat{\alpha}_k} = \text{EL}_h^m(\boldsymbol{\mu}, \hat{\alpha}_k \mathbf{P}_k)$ on $\mathcal{M}_{[\boldsymbol{\mu}, \cdot]}$ to measure the dissimilarity between the sought-for fused posterior density \hat{p} and p_k in the fusion criterion (30).
- (ii) The inner minimizations in (39) project each local density p_k onto the MED submanifold $\mathcal{M}_{[\boldsymbol{\mu}, \cdot]}$ with some common fixed mean to obtain its substitute $p_{\hat{\alpha}_k}$. An appropriate candidate $\hat{\mathbf{x}}$ for the specified mean variable $\boldsymbol{\mu}$ is selected by the outer optimization, minimizing the sum of MHDPDs from the local posterior densities p_k onto $\mathcal{M}_{[\boldsymbol{\mu}, \cdot]}$.

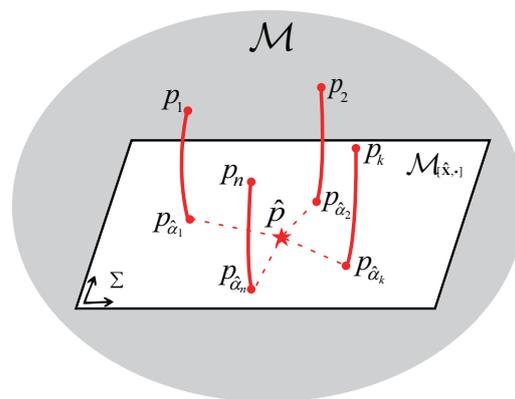


Figure 1. The fusion of n local posterior densities based on the MHDPs.

Furthermore, the mean estimation fusion (39) obtains the final MHDPs $p_{\hat{\alpha}_k} = \text{EL}_h^m(\hat{\mathbf{x}}, \hat{\alpha}_k \mathbf{P}_k)$ in $\mathcal{M}_{[\hat{\mathbf{x}}, \cdot]}$ with the proper scale factors $\hat{\alpha}_k$. Additionally, moving each local estimate $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ to its MHDP $(\hat{\mathbf{x}}, \hat{\alpha}_k \mathbf{P}_k)$ inevitably leads to an increase in the local estimation error. It is observed from (40) that $\hat{\alpha}_k(\hat{\mathbf{x}})$ will tend to the scalar 1 from the right hand side as the fused mean $\hat{\mathbf{x}}$ approaches the k -th local mean estimate $\hat{\mathbf{x}}_k$, so it is reasonable to replace the local scatter matrix \mathbf{P}_k with $\hat{\alpha}_k \mathbf{P}_k$. Then, we can fuse all MHDPs $p_{\hat{\alpha}_1}, \dots, p_{\hat{\alpha}_n}$ on the totally geodesic submanifold $\mathcal{M}_{[\hat{\mathbf{x}}, \cdot]}$ with the fixed $\hat{\mathbf{x}}$ to seek the fused scatter matrix \mathbf{P} , using the following fusion criterion:

$$\mathbf{P} = \arg \min_{\boldsymbol{\Sigma} \in \mathbb{S}_+^m} \sum_{k=1}^n \ell_{\boldsymbol{\Sigma}}^2(p_{\hat{\alpha}_k}, p). \tag{41}$$

It is worth noting that the squared geodesic distance, used as the cost function in (41), differs from the original fusion criterion (29). The main reason are as follows:

- (i) The covariance fusion (41) is indeed performed on the space \mathbb{S}_+^m , since the Rao distance $\ell_{\boldsymbol{\Sigma}}$ of $\mathcal{M}_{[\hat{\mathbf{x}}, \cdot]}$ only depends on the coordinate $\boldsymbol{\Sigma} \in \mathbb{S}_+^m$. Moreover, the Riemannian mean (also called Riemannian center of mass) of data points p_1, \dots, p_n in \mathbb{S}_+^m , which is the unique global minimizer of $\sum_{k=1}^n \ell^2(p_k, p)$ with a geodesic distance $\ell(\cdot, \cdot)$ on \mathbb{S}_+^m (see [52,53]), has been widely studied on the manifold of covariance matrices [54].
- (ii) The criterion $\sum_{k=1}^n \ell(p_k, p)$ gives a robust alternative for the center of mass, called the Riemannian median [55,56]. Both the Riemannian mean and the Riemannian median have been successfully applied to signal detection (e.g., [57,58]) and have shown their own advantages. However when only two sensors (i.e., $n = 2$) are considered, any point lying on the shortest geodesic segment between two points p_1 and p_2 can be regarded as the Riemannian median, which seems quite undesirable for fusion. In addition, the final mean fusion (39) does not contradict this claim owing to its adopting the MHDPD ℓ_M as the cost function.
- (iii) Due to the intractable root operation within the geodesic distance $\ell_{\boldsymbol{\Sigma}}$, the squared geodesic distance is considered for the convenience of solving the optimization prob-

lem (41). As illustrated in Section 4.2, we develop an efficient iterative strategy for the fused scatter estimate in the criterion (41) with at least linear convergence.

In summary, the final fused mean estimate $\hat{\mathbf{x}}$ and scatter matrix estimate \mathbf{P} can be obtained by solving (39) and (41), respectively.

4. Two MHDP-Based Fusion Methods

In this section, we first derive a fixed point iteration for the mean fusion (39), and then provide two different forms of covariance estimation fusion by iteratively solving the optimization problem (41) and introducing the Lie algebraic structure on $\mathcal{M}_{[\hat{\mathbf{x}}, \cdot]}$ to obtain an explicit form of the fused covariance, respectively.

4.1. Mean Estimation Fusion

Define two auxiliary functions

$$\phi(x) = \operatorname{arccosh}(x), \quad x \geq 1, \tag{42}$$

$$\check{d}_k(\mathbf{x}) = cd_k(\mathbf{x}) + 1, \quad \mathbf{x} \in \mathbb{R}^m, \tag{43}$$

and then rewrite (39) as

$$\hat{\mathbf{x}} = \arg \min_{\mu \in \mathbb{R}^m} \sum_{k=1}^n \phi(\check{d}_k(\mu)). \tag{44}$$

Theorem 3. *The solution of the optimization problem (44) can be expressed in the following implicit form*

$$\hat{\mathbf{x}} = \left(\sum_{k=1}^n \omega_k(\hat{\mathbf{x}}) \mathbf{P}_k^{-1} \right)^{-1} \sum_{k=1}^n \omega_k(\hat{\mathbf{x}}) \mathbf{P}_k^{-1} \hat{\mathbf{x}}_k, \tag{45}$$

where the normalized weights are given as

$$\omega_k(\hat{\mathbf{x}}) = \frac{\phi(\check{d}_k(\hat{\mathbf{x}}))}{\sum_{k=1}^n \phi(\check{d}_k(\hat{\mathbf{x}}))}, \quad k = 1, \dots, n. \tag{46}$$

Proof. See Appendix C. \square

In Theorem 3, the fused mean estimate $\hat{\mathbf{x}}$ given by (45) has the same form as the traditional CI estimate, but in general with different weights. To solve the implicit Equation (45) for $\hat{\mathbf{x}}$, we derive a fixed point iteration for seeking the final $\hat{\mathbf{x}}$ in the following theorem.

Theorem 4. *By adopting the fixed point iteration*

$$\mu_{t+1} = \left(\sum_{k=1}^n \hat{\omega}_k(\mu_t) \mathbf{P}_k^{-1} \right)^{-1} \sum_{k=1}^n \hat{\omega}_k(\mu_t) \mathbf{P}_k^{-1} \hat{\mathbf{x}}_k \tag{47}$$

with the normalized weights

$$\hat{\omega}_k(\mu_t) = \frac{\left(\check{d}_k^2(\mu_t) - 1 \right)^{-\frac{1}{2}}}{\sum_{k=1}^n \left(\check{d}_k^2(\mu_t) - 1 \right)^{-\frac{1}{2}}}, \quad k = 1, \dots, n, \tag{48}$$

the resulting sequence $\{\mu_t, t \in \mathbb{N}\}$ converges to a solution $\hat{\mathbf{x}}$ of (45) as t tends to infinity.

Proof. See Appendix D. \square

As shown in Algorithm 1, the fixed point iteration (47) is adopted to calculate the optimal estimate of mean $\hat{\mathbf{x}}$ in (44), where the convergence of iteration can be guaranteed by Theorem 4.

Algorithm 1: MHDP-based mean estimation fusion.

Input : $\{(\hat{\mathbf{x}}_k, \mathbf{P}_k)\}_{k=1,\dots,n}$ and tolerance ϵ
Output: $\hat{\mathbf{x}}$
 1 Set $t = 0, r = \epsilon$;
 2 Initiate $\boldsymbol{\mu}_0$ using the CI estimate with equal weights;
 3 **while** $r \geq \epsilon$ **do**
 4 Run the mean iteration using (47);
 5 Compute the Frobenius norm $r = \|\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t\|$;
 6 $t \leftarrow t + 1$;
 7 **end**
 8 **return** $\hat{\mathbf{x}} = \boldsymbol{\mu}_t$

Remark 2. Similar to the GP method in [42], the fused scatter matrix estimate seems available to be set as

$$\mathbf{P} = \left(\sum_{k=1}^n \omega_k \mathbf{P}_k^{-1} \right)^{-1}, \tag{49}$$

since the fused mean estimate $\hat{\mathbf{x}}$ has the CI form (45), and then the fused covariance estimate is naturally given as

$$\mathbf{Q} = \kappa_h \mathbf{P} = \left(\sum_{k=1}^n \omega_k \mathbf{Q}_k^{-1} \right)^{-1} \tag{50}$$

with each local covariance $\mathbf{Q}_k = \kappa_h \mathbf{P}_k$. Nevertheless, (50) is reasonable to be replaced by the following two MHDP-based covariance fusion forms owing to its subjectively adopting the covariance fusion.

4.2. Iterative Solution for Covariance Estimation Fusion

Define an auxiliary function

$$\psi_k(\boldsymbol{\Sigma}) = \ell_{\boldsymbol{\Sigma}}^2(p_{\hat{\mathbf{x}}_k}, p), \tag{51}$$

and then inserting the Rao distance (24) into (51) yields

$$\psi_k(\boldsymbol{\Sigma}) = 2b_h \text{tr}(\mathbf{T}_k^2) + \frac{4b_h - 1}{4} \text{tr}^2(\mathbf{T}_k), \tag{52}$$

where,

$$\check{\mathbf{P}}_k = \hat{\mathbf{x}}_k \mathbf{P}_k, \tag{53}$$

$$\mathbf{T}_k = \text{Log}(\check{\mathbf{P}}_k^{-1} \boldsymbol{\Sigma}). \tag{54}$$

In the theorem given below, we derive the Riemannian gradient of $\psi_k(\boldsymbol{\Sigma})$ on $\mathcal{M}_{[\hat{\mathbf{x}}, \cdot]}$.

Theorem 5. The Riemannian gradient of $\psi_k(\boldsymbol{\Sigma})$ on the submanifold $\mathcal{M}_{[\hat{\mathbf{x}}, \cdot]}$ is given as

$$\nabla \psi_k = 2\boldsymbol{\Sigma} \mathbf{T}_k. \tag{55}$$

Proof. See Appendix E. \square

Furthermore, denoting the objective function in (41) as

$$\psi(\Sigma) = \sum_{k=1}^n \psi_k(\Sigma), \tag{56}$$

and differentiating $\psi(\cdot)$ with respect to Σ , we have the Riemannian gradient vector

$$\nabla\psi = \sum_{k=1}^n \nabla\psi_k = 2\Sigma \sum_{k=1}^n \text{Log}(\check{\mathbf{P}}_k^{-1}\Sigma). \tag{57}$$

As a result, the minimizer \mathbf{P} of (41) satisfies $\nabla\psi = 0$, i.e.,

$$\frac{1}{n} \sum_{k=1}^n \text{Log}(\check{\mathbf{P}}_k^{-1}\Sigma) = \mathbf{0}. \tag{58}$$

Remark 3. The average Σ of n symmetric positive-definite matrices $\check{\mathbf{P}}_1, \dots, \check{\mathbf{P}}_n$, satisfying the barycentric equation (58), is also called the Riemannian barycenter [59].

In general, (58) cannot be solved explicitly due to the noncommutative nature on \mathbb{S}_+^m , but we formulate the fixed point iteration for the scatter estimation fusion as follows:

$$\Sigma^{(t+1)} = \Sigma^{(t)} \cdot \text{Exp}\left(\frac{1}{n} \sum_{k=1}^n \text{Log}(\check{\mathbf{P}}_k^{-1}\Sigma^{(t)})\right), \tag{59}$$

which is a very efficient iterative strategy for solving (58) with at least linear convergence (see [60] for a detailed proof). The fusion method above is called the MHDP-I method (i.e., Algorithm 2).

Algorithm 2: MHDP-I fusion algorithm.

Input : $\{(\hat{\mathbf{x}}_k, \mathbf{P}_k)\}_{k=1, \dots, n}$ and tolerance ϵ
Output: $\hat{\mathbf{x}}, \mathbf{Q}$

- 1 Compute $\hat{\mathbf{x}}$ using Algorithm 1;
- 2 Compute $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_n)$ using equation (40);
- 3 Compute $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$ using equation (46);
- 4 Set $t = 0, r_1 = r_2 = \epsilon$ and $\check{\mathbf{P}}_k = \hat{\alpha}_k \mathbf{P}_k, k = 1, \dots, n$;
- 5 Initiate $\Sigma^{(0)}$ using equation (49);
- 6 **while** $r \geq \epsilon$ **do**
- 7 Compute $\Sigma^{(t+1)}$ using equation (59);
- 8 Compute the Euclidean norm $r = \|\Sigma^{(t+1)} - \Sigma^{(t)}\|$;
- 9 $t \leftarrow t + 1$;
- 10 **end**
- 11 **return** $\hat{\mathbf{x}}, \mathbf{Q} = \kappa_h \Sigma^{(t)}$

4.3. Explicit Solution for Covariance Estimation Fusion

The submanifold $\mathcal{M}_{[\hat{\mathbf{x}}]}$ with the fixed mean $\hat{\mathbf{x}}$ can be identified with the symmetric positive-definite matrix space \mathbb{S}_+^m equipped with the same Riemannian metric induced by (23). Meanwhile, we can endow \mathbb{S}_+^m with a group structure depending on the following definition:

Definition 1. $\mathcal{S}_+^m = (\mathbb{S}_+^m, \odot, i, e)$ is a commutative group whose group operations are defined as follows:

- (i) Multiplication operation: $g_1 \odot g_2 := \text{Exp}(\text{Log}(g_1) + \text{Log}(g_2))$, for any $g_1, g_2 \in \mathcal{S}_+^m$;
- (ii) Inverse operation: $i(g) := g^{-1}$, for any $g \in \mathcal{S}_+^m$;
- (iii) Identity element: $e = \mathbf{I}_m$.

It is easy to verify that the two operations \odot and i are compatible with the Riemannian structure on \mathcal{S}_+^m (see, e.g., [61]), hence \mathcal{S}_+^m is a Lie group.

As is well known, the Lie group is locally equivalent to a linear space, in which the local neighborhood of any element can be described by its tangent space. The tangent space $T_e \mathcal{M}_{[\check{x}, \cdot]}$ at the identity element $e = \mathbf{I}_m$, also called Lie algebra \mathfrak{g} , is indeed isomorphic to the space \mathbb{S}^m of real symmetric matrices. Theoretically, the retraction mapping $R(\cdot)$ establishes a vector space isomorphism from $(\mathbb{S}^m, +)$ to $(\mathcal{M}_{[\check{x}, \cdot]}, \odot)$. Then, we endow $\mathcal{M}_{[\check{x}, \cdot]}$, namely \mathcal{S}_+^m , with a ‘‘Lie vector space structure’’.

It is important to stress that, based on the linear structure of Lie algebra, we can handle the covariance/scatter estimation fusion with the Euclidean operation, i.e., the arithmetic averaging in the Log-Euclidean domain, instead of the Riemannian one. It is explained as follows:

- (i) Move the MHDPs $\check{\mathbf{P}}_1, \dots, \check{\mathbf{P}}_n$ into some neighborhood of the identity element $e = \mathbf{I}_m \in \mathcal{M}_{[\check{x}, \cdot]}$ via using a left translation $L_{\mathbf{P}}(\cdot)$ by the sought-for scatter estimate \mathbf{P} to obtain $L_{\mathbf{P}}(\check{\mathbf{P}}_k) = \mathbf{P}^{-1} \odot \check{\mathbf{P}}_k$. Then, all $L_{\mathbf{P}}(\check{\mathbf{P}}_k)$ can be shifted to \mathfrak{g} by the lifting mapping $R_e^{-1}(\cdot)$ at e , obtaining the vectors $R_e^{-1}(L_{\mathbf{P}}(\check{\mathbf{P}}_k))$ for $k = 1, \dots, n$.
- (ii) The arithmetic average of $R_e^{-1}(L_{\mathbf{P}}(\check{\mathbf{P}}_k))$ is the best way to minimize the total displacements from each $\check{\mathbf{P}}_k$ to the sought-for \mathbf{P} on the Lie algebra \mathfrak{g} owing to the linear structure of \mathfrak{g} (see, e.g., [60]), hence $\frac{1}{n} \sum_{k=1}^n R_e^{-1}(\mathbf{P}^{-1} \odot \check{\mathbf{P}}_k) = \mathbf{0}$, or equivalently,

$$\mathbf{P} = \mathbf{P} \odot R_e \left(\frac{1}{n} \sum_{k=1}^n R_e^{-1}(\mathbf{P}^{-1} \odot \check{\mathbf{P}}_k) \right). \tag{60}$$

Let the retraction mapping $R_e(\cdot) = \exp_e(\cdot)$ and the lifting mapping $R_e^{-1}(\cdot) = \log_e(\cdot)$ on $\mathcal{M}_{[\check{x}, \cdot]}$, whose analytical expressions are given in the theorem below.

Theorem 6. *The following properties hold:*

- (i) Let \mathbf{H} be a non-singular matrix. The metric on $\mathcal{M}_{[\check{x}, \cdot]}$ originated from (23) is invariant under the congruent transformation: $\Sigma \rightarrow \mathbf{H}^T \Sigma \mathbf{H}$ for any $\Sigma \in \mathbb{S}_+^m$.
- (ii) The exponential mapping and logarithmic mapping on $\mathcal{M}_{[\check{x}, \cdot]}$ are, respectively, given by

$$\exp_{\Sigma_1}(\mathbf{v}) = \Sigma_1 \text{Exp}(\Sigma_1^{-1} \mathbf{v}), \tag{61}$$

$$\log_{\Sigma_1}(\Sigma_2) = \Sigma_1 \text{Log}(\Sigma_1^{-1} \Sigma_2), \tag{62}$$

for any $\Sigma_1, \Sigma_2 \in \mathcal{M}_{[\check{x}, \cdot]}$ and $\mathbf{v} \in T_{\Sigma_1} \mathcal{M}_{[\check{x}, \cdot]}$.

Proof. See Appendix F. \square

Substituting (61) and (62) into (60) yields

$$\mathbf{P} = \mathbf{P} \odot \text{Exp} \left(\frac{1}{n} \sum_{k=1}^n \text{Log}(\mathbf{P}^{-1} \odot \check{\mathbf{P}}_k) \right). \tag{63}$$

Then, by $\mathbf{P}^{-1} \odot \check{\mathbf{P}}_k = \text{Exp}(\text{Log} \check{\mathbf{P}}_k - \text{Log} \mathbf{P})$ and the definition of \odot , we have the fused scatter matrix estimate

$$\mathbf{P} = \text{Exp} \left(\frac{1}{n} \sum_{k=1}^n \text{Log}(\check{\mathbf{P}}_k) \right), \tag{64}$$

and the fused covariance matrix estimate

$$\mathbf{Q} = \kappa_h \text{Exp} \left(\frac{1}{n} \sum_{k=1}^n \text{Log}(\check{\mathbf{P}}_k) \right). \tag{65}$$

Remark 4. As shown in [60], the explicit scatter matrix estimate (64) coincides with the arithmetic mean in the log-domain, i.e., the Log-Euclidean Fréchet mean.

In summary, the fusion method above is called the MHDP-E method (i.e., Algorithm 3). For comparison, the MHDP-I and MHDP-E fusion algorithms have the same mean fusion (see Algorithm 1), but the latter possesses an explicit form for the fused covariance matrix estimate (65). In addition, many existing fusion methods have identical forms as the CI, but differ from each other in the weights. For example, the KLA has equal weights $1/N$, while many others, including DCI, FCI, GP and QMMHD, have their own weights, depending on local estimates. Taking two sensors tracking a one-dimensional target for instance, the fused estimates for the KLA, DCI, FCI, GP and QMMHD theoretically lie on the (straight) Euclidean line segment between two local estimates from sensors, as shown in Figure 2 of Section 5.1. However, the MED manifold \mathcal{M} with the Riemannian structure is generally not flat [47], so the proposed MHDP-I and MHDP-E are more reasonable owing to their improving covariance estimation fusion based on this geometric structure.

Algorithm 3: MHDP-E fusion algorithm.

Input : $\{(\hat{x}_k, \mathbf{P}_k)\}_{k=1,\dots,n}$

Output: \hat{x}, \mathbf{Q}

- 1 Compute \hat{x} using Algorithm 1;
- 2 Compute $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_n)$ using Equation (40);
- 3 Set $\check{\mathbf{P}}_k = \hat{\alpha}_k \mathbf{P}_k, k = 1, \dots, n$;
- 4 Compute \mathbf{Q} using Equation (65);
- 5 return \hat{x}, \mathbf{Q}

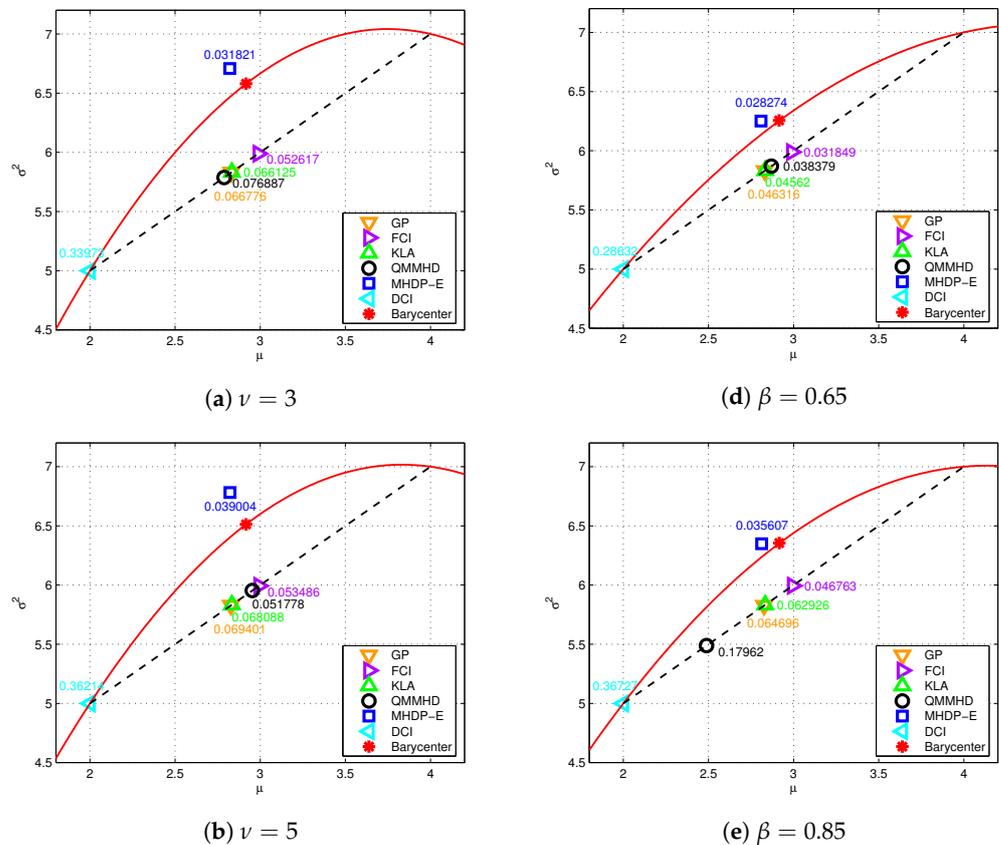


Figure 2. Cont.

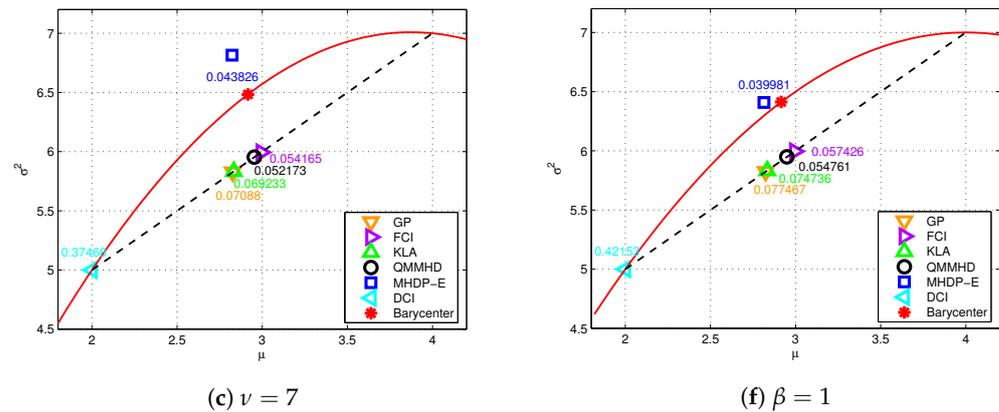


Figure 2. Six subfigures are depicted for the geodesic distance between the fused density and the informative barycenter. The various fused results are labeled by the specified marks. Subfigures (a–c) correspond to MTDs with DOF parameters $\nu = 3, 5$ and 7 , respectively. Subfigures (d–f) correspond to MGGDs with shape parameters $\beta = 0.65, 0.85$ and 1 , respectively. In each subfigure, the geodesic distances between the informative barycenter and the fused densities are displayed, and the geodesic distance between the two local densities is represented by the solid red curve.

5. Numerical Examples

In this section, we provide two numerical examples to show the performance of two fusion methods MHDP-I and MHDP-E, including static and dynamic target tracking. The traditional fusion algorithms DCI [16], FCI [15], and KLA [18], and two information-geometric methods GP [42] and QMMHD [43] are taken for comparison.

5.1. Static Target Tracking

To intuitively evaluate the performance of these distributed fusion methods, we consider two local estimates $(\hat{x}_1, P_1) = (2, 5)$ and $(\hat{x}_2, P_2) = (4, 7)$ of a static target from two local sensors, similar to the settings in [42]. Note that the MHDP-I and MHDP-E methods have the same covariance fusion due to the commutability of two first-order matrices. In Figure 2, six subfigures are drawn to display the geodesic distance between the fused density and the informative barycenter under six different MED assumptions, respectively. Here, the informative barycenter as the optimal solution of (29) is an arbitrary point on the geodesic segment linking (\hat{x}_1, P_1) and (\hat{x}_2, P_2) , and we uniquely determine it by minimizing the sum of its squared distances to the two local estimates. Since the MHDP-I and MHDP-E are developed based on the relaxation of (29), learning about the informative barycenter can improve our understanding of their fusion performance as the heavy-tailed level varies.

- (i) The MTD has the degree of freedom (DOF) parameter $\nu > 2$ (the higher ν , the lower the heavy-tailed level), and is almost equivalent to the Gaussian distribution as ν tends to infinity. As depicted in subfigures (a), (b) and (c), the MHDP-E is consistently superior to other methods regardless of the DOF, since the fused density of the MHDP-E is closest to the informative barycenter.
- (ii) The MGGD has the shape parameter $\beta > 0$ (the higher β , the lower the heavy-tailed level) and corresponds to the Gaussian distribution when $\beta = 1$. Similar to the MTD, subfigures (d), (e) and (f) show the MHDP-E outperforms other fusion methods for different shape parameters.

Moreover, the aforementioned fusion methods DCI, FCI, KLA, GP and QMMHD have the same form as (45), but with different weights in general. Thus in Figure 2 their fused estimates lie on the (straight) Euclidean segment between (\hat{x}_1, P_1) and (\hat{x}_2, P_2) , while the MHDP-E puts its estimate fairly close to the (curved) geodesic segment. Therefore, the MHDP-I and MHDP-E are considered more reasonable because the MED manifold \mathcal{M} is indeed non-Euclidean.

5.2. Dynamic Target Tracking

Consider a distributed dynamic system with n sensors and one fusion center for tracking a two-dimensional moving target:

$$\mathbf{x}_{k+1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} w_k, \tag{66}$$

$$z_k^i = \mathbf{h}_k^i \mathbf{x}_k + v_k^i, \quad i = 1, \dots, n_0, \tag{67}$$

$$\mathbf{z}_k^i = \mathbf{H}_k^i \mathbf{x}_k + \mathbf{v}_k^i, \quad i = n_0 + 1, \dots, n, \tag{68}$$

where there are n_0 sensors for one-dimensional measurements and $n - n_0$ sensors for multi-dimensional measurements, the state vector \mathbf{x}_k consists of the position and velocity, and the sampling interval $T = 1$ s. As [49,50] for more practical scenarios, the outlier contaminated state and measurement noises are modeled as

$$w_k \sim \begin{cases} N(0, Q), & \text{w.p. } 0.9, \\ N(0, 200Q), & \text{w.p. } 0.1, \end{cases} \tag{69}$$

$$v_k^i \sim \begin{cases} N(0, R_k^i), & \text{w.p. } 0.9, \\ N(0, uR_k^i), & \text{w.p. } 0.1, \end{cases} \quad i = 1, \dots, n_0, \tag{70}$$

$$\mathbf{v}_k^i \sim \begin{cases} N(0, \mathbf{R}_k^i), & \text{w.p. } 0.9, \\ N(0, u\mathbf{R}_k^i), & \text{w.p. } 0.1, \end{cases} \quad i = n_0 + 1, \dots, n, \tag{71}$$

where ‘‘w.p.’’ denotes ‘‘with probability’’.

The local state estimate $\hat{\mathbf{x}}_i(k|k)$ and scatter matrix $\mathbf{P}_i(k|k)$ of estimation error are computed by using the specified filter, when the i -th sensor obtains its own measurement at the instant k . The fusion center fuses all local estimates from the sensors to obtain the final estimate $(\hat{\mathbf{x}}(k|k), \mathbf{P}(k|k))$ by applying some specified fusion methods, and then transmits it back to each local sensor. To demonstrate the performance of various fusion methods, i.e., the DCI, FCI, KLA, GP, QMMHD, MHDP-I and MHDP-E, we use the Student’s t based filter (STF) [49] and the Gaussian–Student’s t mixture distribution based Kalman filter (GSTMDKF) [41] to compare the root mean squared errors (RMSEs) of the fused results in the following two cases:

- (i) Case I (two sensors, i.e., $n_0 = n = 2$): The variances $Q = 1$ and $R_k^1 = R_k^2 = 16$, the measurement matrices

$$\mathbf{h}_k^1 = [1, 0], \quad \mathbf{h}_k^2 = [0, 1], \tag{72}$$

and the two filters are initialized as

$$\hat{\mathbf{x}}(0|0) = \begin{bmatrix} 0 \\ 5 \end{bmatrix}, \quad \mathbf{P}(0|0) = \begin{bmatrix} 10 & \\ & 25 \end{bmatrix}. \tag{73}$$

- (ii) Case II (three sensors, i.e., $n_0 = 2, n = 3$): The system parameters $Q, R_k^1, R_k^2, \mathbf{h}_k^1, \mathbf{h}_k^2, \hat{\mathbf{x}}(0|0), \mathbf{P}(0|0)$ have the same settings as Case I, and also

$$\mathbf{H}_k^3 = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}, \quad \mathbf{R}_k^3 = \begin{bmatrix} 16 & 4 \\ 4 & 6 \end{bmatrix}. \tag{74}$$

It is seen from (69)–(71) that the process and observation noises follow heavy-tailed distributions. The STF models the initial state vector and the noise distributions with $u = 200$ as follows:

$$\mathbf{x}_0 \sim \text{MTD}(\hat{\mathbf{x}}(0|0), \mathbf{P}(0|0), \nu_1), \tag{75}$$

$$w_k \sim \text{MTD}(0, Q, \nu_2), \tag{76}$$

$$v_k^i \sim \text{MTD}(0, R_k^i, \nu_3), \quad i = 1, \dots, n_0, \tag{77}$$

$$\mathbf{v}_k^i \sim \text{MTD}(0, \mathbf{R}_k^i, \nu_4), \quad i = n_0 + 1, \dots, n. \tag{78}$$

By (75)–(78), each local posterior density can be represented by the MTD with the DOF parameter $\nu = 3$, i.e., $\nu_i = 3, i = 1, \dots, 4$. In the GSTMDKF setup, a Gaussian–Student’s t mixture distribution (see [41] for more details) with the fixed DOF 3 is used to model the heavy-tailed noises (69)–(71) with $u = 100$, the initial state vector follows a Gaussian distribution, i.e., $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}(0|0), \mathbf{P}(0|0))$, and other parameters are the same as those in [41]. As explained in [41], the GSTM distribution has the same efficiency as the Student’s t distribution in the presence of outliers, and thus our proposed MHDP-E and MHDP-I methods fuse all posterior densities by modeling them as the MTDs with the common DOF parameter 3. Note that the DCI, FCI, KLA, and GP only utilize the first two moments (i.e., $\hat{\mathbf{x}}(k|k)$ and $\kappa\mathbf{P}(k|k)$ for the i -th local posterior density), and the FCI, KLA and GP fusion methods are developed under Gaussian assumption so that the posterior density for the i -th sensor can be modeled as $N(\hat{\mathbf{x}}_i(k|k), \kappa\mathbf{P}_i(k|k))$.

By adopting two different filters STF and GSTMDKF, Figure 3 shows the RMSEs of the position over 100 time steps and 500 Monte Carlo runs in two different cases using the aforementioned fusion methods, while the DCI does not appear in Figure 3 owing to it behaving worst. We can see that the proposed MHDP-I and MHDP-E are consistently better than other fusion methods when different filtering methods are adopted. Moreover, Table 1 reports the averaged root mean squared errors (ARMSEs) of the various methods in both cases. It is evident that the performance of all fusion methods in three-sensor example has a significant improvement over that in the two-sensor case, and also the MHDP-I performs slightly better than the MHDP-E when using the GSTMDKF. However in contrast to the MHDP-I, the MHDP-E has lower computational complexity due to the explicit expression for the fused covariance estimate.

Table 1. The Position ARMSEs of seven fusion methods in both cases by the STF and the GSTMDKF.

Method	STF		GSTMDKF	
	Case I	Case II	Case I	Case II
DCI	28.99	26.76	46.01	35.84
GP	8.54	5.93	10.35	6.47
FCI	8.23	6.15	7.82	5.36
KLA	7.92	5.96	9.47	6.49
QMMHD	7.06	5.45	7.78	5.65
MHDP-E	5.72	3.35	5.91	3.64
MHDP-I	5.72	3.36	5.90	3.63

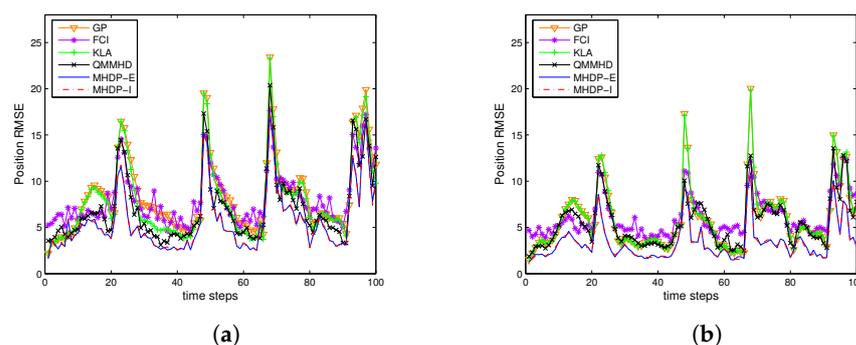


Figure 3. Cont.

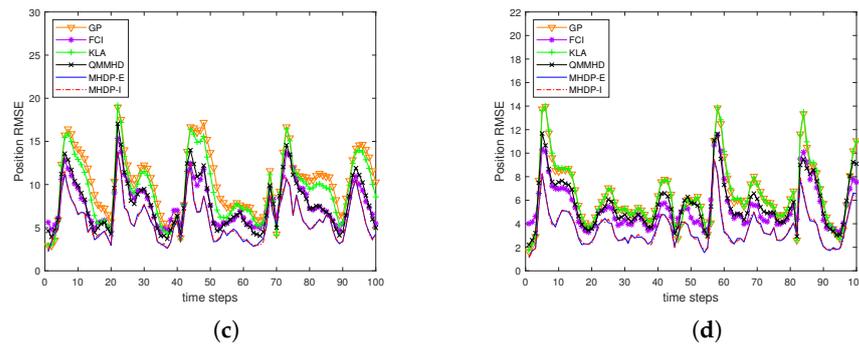


Figure 3. The position RMSEs of fused results vs. the instant k from 1 to 100 by (a) the STF in Case I; (b) the STF in Case II; (c) the GSTMDKF in Case I; and (d) the GSTMDKF in Case II.

6. Conclusions

In this work, we formulate an information-geometric fusion criterion, taking the geodesic distance as the loss function, and implement the fusion of local posterior densities on the MED manifold endowed with the Fisher metric. Two distributed estimation fusion algorithms MHD-E and MHD-I are proposed by using the minimal Manhattan distance instead of the geodesic distance on the manifold of MEDs, which both have the same mean estimation fusion, but differ in the form of covariance estimation fusion. On the MED submanifold with a fixed mean, the MHD-I fuses all MHD-based projections of local posterior densities by minimizing the squared geodesic distances from a sought-for fused density. We have developed a robust fixed point iterative algorithm with theoretical convergence to compute the covariance estimate of such fused density. By introducing a specified Lie algebraic structure on the aforementioned submanifold and deriving its exponential and logarithmic mappings, the MHD-E has provided an explicit expression for the fused covariance estimate. Numerical examples have shown better performance of the two proposed information-geometric methods than five other fusion methods.

Author Contributions: X.C. put forward the original ideas and conducted the research; J.Z. raised the research question, reviewed this paper and provided improvement suggestions. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Sichuan Science and Technology Program under Grant 2022JDRC0068, and the National Natural Science Foundation of China under Grants 61374027, 11871357 and 12271376.

Acknowledgments: We thank the anonymous reviewers for their suggestions on improving the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

Let $c_h = 8a_h / (12b_h - 1)$ and the variable transformation

$$\check{\mu} = \sqrt{c_h} \mu. \tag{A1}$$

By using (13), the geodesic equations on the MED manifold with the coordinate system (μ, Σ) can be rewritten as

$$\check{\ddot{\mu}} - \check{\Sigma} \Sigma^{-1} \dot{\check{\mu}} = 0, \tag{A2}$$

$$\check{\Sigma} + \dot{\check{\mu}}^2 - \Sigma^{-1} \dot{\Sigma}^2 = 0. \tag{A3}$$

Then, the squared line element along the geodesic becomes

$$\begin{aligned} d\check{s}^2 &= (d\check{\mu})^2 \Sigma^{-1} + \frac{1}{2}(\Sigma^{-1} d\Sigma)^2 \\ &= \frac{2}{12b_h - 1} \left(4a_h (d\mu)^2 \Sigma^{-1} + (3b_h - 1/4)(\Sigma^{-1} d\Sigma)^2 \right). \end{aligned} \tag{A4}$$

Compared with (11), we obtain $\check{s} = \sqrt{2/(12b_h - 1)}s$, where s is the arc length parameter before the transformation (A1). Therefore, by Theorem 6.5 in [62], the geodesic distance between p_1 and p_2 is given as

$$\ell_F = d_h \cdot \operatorname{arccosh} \left(\frac{c_h(\mu_2 - \mu_1)^2 + 2(\sigma_1^2 + \sigma_2^2)}{4\sigma_1\sigma_2} \right) \tag{A5}$$

with $d_h = \sqrt{12b_h - 1}$. To that end, using Theorem 6.2 in [62], we obtain the geodesic curve on the MED manifold as

$$\mu(s) = (\delta_1 + 2\delta_2 \tanh(s/d_h + \varepsilon)) / \sqrt{c_h}, \tag{A6}$$

$$\sigma^2(s) = 2\delta_2^2 \cosh^{-2}(s/d_h + \varepsilon), \tag{A7}$$

where δ_1, δ_2 and ε can be calculated by boundary conditions. Note that for the case of $\mu_1 = \mu_2$, the geodesic given by (15) and (16) are easily obtained by Equation (6.4) in [62].

Appendix B. Proof of Theorem 2

Define an auxiliary function

$$f(\alpha) = \frac{(cd_M + 2)^2}{8\alpha} + \frac{\alpha}{2} + \frac{cd_M}{2}, \quad \alpha \in \mathbb{R}_+. \tag{A8}$$

It is easy to see from (26) and (A8) that the minimization for (31) is equivalent to minimizing $f(\alpha)$ with respect to $\alpha \in \mathbb{R}_+$. We have thus $\ell_M(p_0, \mathcal{M}_{[\mu, \cdot]}) = \min_{\alpha \in \mathbb{R}_+} \ell_\mu(p_0, p_\alpha)$, the optimal scale factor

$$\begin{aligned} \hat{\alpha}_0 &:= \arg \min_{\alpha \in \mathbb{R}_+} \ell_\mu(p_0, p_\alpha) \\ &= \arg \min_{\alpha \in \mathbb{R}_+} \left(\frac{(cd_0 + 2)^2}{8\alpha} + \frac{\alpha}{2} + \frac{cd_0}{2} \right) \\ &= \frac{cd_0}{2} + 1 \end{aligned} \tag{A9}$$

with $d_0 = (\hat{\mathbf{x}}_0 - \boldsymbol{\mu})^T \mathbf{P}_0^{-1} (\hat{\mathbf{x}}_0 - \boldsymbol{\mu})$ and the unique MHDP $p_{\hat{\alpha}_0} = \text{EL}_h^m(\boldsymbol{\mu}, \hat{\alpha}_0 \mathbf{P}_0)$. Therefore, substituting (A9) into (26) yields the MHDPD (32).

Appendix C. Proof of Theorem 3

By the Lagrange multiplier method, the minimizer $\hat{\mathbf{x}}$ of (44) must satisfy the following equality:

$$\sum_{k=1}^n \phi(\check{d}_k(\boldsymbol{\mu})) \mathbf{P}_k^{-1} (\boldsymbol{\mu} - \hat{\mathbf{x}}_k) = 0. \tag{A10}$$

Taking further derivation, we immediately have (45) and (46). Additionally, it is obvious that $\phi(x) \geq 0$ for $x \geq 1$, and thus each $\omega_k \in [0, 1]$.

Appendix D. Proof of Theorem 4

Define an auxiliary function

$$\varphi(x) = \frac{x_1 + x}{\sqrt{x_1^2 - 1} + \sqrt{x^2 - 1}}, \quad x \in (1, +\infty) \tag{A11}$$

with $x_1 > 1$. It is easy to know $\dot{\varphi}(x) \leq 0$, and thus we have the inequality

$$\frac{x_1 + x_2}{\sqrt{x_1^2 - 1} + \sqrt{x_2^2 - 1}} \leq \frac{x_1}{\sqrt{x_1^2 - 1}} \tag{A12}$$

for $x_2 \geq x_1 > 1$.

Furthermore, define two functions of $\boldsymbol{\mu} \in \mathbb{R}^m$ as follows:

$$\check{\phi}(\boldsymbol{\mu}) = \sum_{k=1}^n \phi(\check{d}_k(\boldsymbol{\mu})) = \sum_{k=1}^n \operatorname{arccosh}(\check{d}_k(\boldsymbol{\mu})), \tag{A13}$$

$$\check{\phi}_{\boldsymbol{\mu}_t}(\boldsymbol{\mu}) = \check{\phi}(\boldsymbol{\mu}_t) + \sum_{k=1}^n \frac{\check{d}_k(\boldsymbol{\mu}) - \check{d}_k(\boldsymbol{\mu}_t)}{\sqrt{\check{d}_k^2(\boldsymbol{\mu}_t) - 1}}. \tag{A14}$$

Utilizing two inequalities $\log(x) \leq x - 1$ and (A12), we obtain

$$\begin{aligned} & \operatorname{arccosh}(\check{d}_k(\boldsymbol{\mu})) - \operatorname{arccosh}(\check{d}_k(\boldsymbol{\mu}_t)) \\ & \leq \frac{\check{d}_k(\boldsymbol{\mu}) + \sqrt{\check{d}_k^2(\boldsymbol{\mu}) - 1}}{\check{d}_k(\boldsymbol{\mu}_t) + \sqrt{\check{d}_k^2(\boldsymbol{\mu}_t) - 1}} - 1 \\ & = (\check{d}_k(\boldsymbol{\mu}) - \check{d}_k(\boldsymbol{\mu}_t)) \frac{1 + \frac{\check{d}_k(\boldsymbol{\mu}) + \check{d}_k(\boldsymbol{\mu}_t)}{\sqrt{\check{d}_k^2(\boldsymbol{\mu}) - 1} + \sqrt{\check{d}_k^2(\boldsymbol{\mu}_t) - 1}}}{\check{d}_k(\boldsymbol{\mu}_t) + \sqrt{\check{d}_k^2(\boldsymbol{\mu}_t) - 1}} \\ & \leq \frac{\check{d}_k(\boldsymbol{\mu}) - \check{d}_k(\boldsymbol{\mu}_t)}{\sqrt{\check{d}_k^2(\boldsymbol{\mu}_t) - 1}}. \end{aligned} \tag{A15}$$

By (A13), (A14) and (A15), it follows that $\check{\phi}(\boldsymbol{\mu}) \leq \check{\phi}_{\boldsymbol{\mu}_t}(\boldsymbol{\mu})$ and $\check{\phi}(\boldsymbol{\mu}_t) = \check{\phi}_{\boldsymbol{\mu}_t}(\boldsymbol{\mu}_t)$, and thus $\check{\phi}(\boldsymbol{\mu})$ is majorized by $\check{\phi}_{\boldsymbol{\mu}_t}(\boldsymbol{\mu})$ at point $\boldsymbol{\mu}_t$ under the majorization–minimization framework [63].

Consequently, setting the gradient of $\check{\phi}_{\boldsymbol{\mu}_t}(\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$ as zero, we conclude that

$$\sum_{k=1}^n \frac{\mathbf{P}_k^{-1}(\boldsymbol{\mu} - \hat{\mathbf{x}}_k)}{\sqrt{\check{d}_k^2(\boldsymbol{\mu}_t) - 1}} = 0, \tag{A16}$$

and then the fixed point iteration (47) for the mean estimate $\hat{\mathbf{x}}$ follows from (A16).

Appendix E. Proof of Theorem 5

Lemma A1 (See [52]). *If $\mathbf{X}(t)$ is an invertible real matrix-valued function of real variable t and its eigenvalues lie on the positive real line, then*

$$\frac{d}{dt} \operatorname{tr}(\operatorname{Log}^2 \mathbf{X}(t)) = 2 \operatorname{tr} \left(\operatorname{Log} \mathbf{X}(t) \mathbf{X}^{-1}(t) \frac{d}{dt} \mathbf{X}(t) \right). \tag{A17}$$

Let $\gamma(t; \theta, \nu)$ denote the geodesic emanating from the point $\theta = \Sigma$ at $t = 0$ in an arbitrary direction $\nu \in T_{\theta} \mathcal{M}_{[\hat{x}, \cdot]}$. Inspired by the method in the proof of Lemma A1 in [52], we can easily know

$$\begin{aligned} \frac{d}{dt} \text{tr}(\text{Log } \gamma(t)) &= \text{tr} \left(\int_0^1 ((\gamma(t) - \mathbf{I})s + \mathbf{I})^{-2} ds \cdot \frac{d\gamma(t)}{dt} \right) \\ &= \text{tr} \left(\gamma^{-1}(t) \frac{d}{dt} \gamma(t) \right). \end{aligned} \tag{A18}$$

Applying (A17) and (A18), and differentiating $\psi_k(\gamma(t))$ with respect to t , we have

$$\begin{aligned} \frac{d}{dt} \psi_k(\gamma(t)) &= 4b_h \text{tr} \left(\mathbf{T}_k(t) \gamma^{-1}(t) \frac{d\gamma(t)}{dt} \right) \\ &\quad + \left(2b_h - \frac{1}{2} \right) \text{tr}(\mathbf{T}_k(t)) \text{tr} \left(\gamma^{-1}(t) \frac{d\gamma(t)}{dt} \right), \end{aligned} \tag{A19}$$

where $\mathbf{T}_k(t) := \text{Log}(\check{\mathbf{P}}_k^{-1} \gamma(t))$. As a result,

$$\frac{d}{dt} \psi_k(\gamma(t)) \Big|_{t=0} = 2b_h \text{tr}(2\mathbf{T}_k \Sigma^{-1} \nu) + \frac{4b_h - 1}{4} \text{tr}(2\mathbf{T}_k) \text{tr}(\Sigma^{-1} \nu). \tag{A20}$$

According to the arbitrariness of vector ν and the Riemannian inner product (23), we obtain the Riemannian gradient

$$\nabla \psi_k = 2\Sigma \mathbf{T}_k. \tag{A21}$$

Appendix F. Proof of Theorem 6

(i) From the inner product (23), we have the squared line element

$$ds^2 = 2b_h \text{tr}(\Sigma^{-1} d\Sigma)^2 + \frac{4b_h - 1}{4} \text{tr}^2(\Sigma^{-1} d\Sigma). \tag{A22}$$

By taking the congruent transformation $\Sigma \rightarrow \mathbf{H}^T \Sigma \mathbf{H}$, it is not difficult to see that the squared infinitesimal distance (A22) is invariant.

(ii) Suppose

$$r_k = \log \lambda_k, \quad k = 1, \dots, m, \tag{A23}$$

where λ_k are the eigenvalues of $\Sigma_1^{-1} \Sigma_2$. There exists a nonsingular matrix \mathbf{H} such that

$$\mathbf{H}^T \Sigma_1 \mathbf{H} = \mathbf{I}_m, \quad \mathbf{H}^T \Sigma_2 \mathbf{H} = \Lambda, \tag{A24}$$

where $\Lambda = \text{diag}([\lambda_1, \dots, \lambda_m])$. According to the geodesic from \mathbf{I}_m to Λ on $\mathcal{M}_{[\hat{x}, \cdot]}$

$$r_k(t) = a_k t + b_k, \quad k = 1, \dots, m \tag{A25}$$

with initial conditions $r_k(0) = 0$ and $r_k(1) = r_k$ (see [48]), we have $a_k = \log \lambda_k$, $b_k = 0$ and then the geodesic direction

$$\dot{\Lambda}(0) = \text{Log}(\Lambda). \tag{A26}$$

Let $\gamma(t), t \in [0, 1]$ be the shortest geodesic curve connecting two points Σ_1 and Σ_2 on $\mathcal{M}_{[\hat{x}, \cdot]}$. Due to the invariance of geodesic distance under the congruent transformation, we have

$$\dot{\Lambda}(0) = \mathbf{H}^T \dot{\gamma}(0) \mathbf{H}, \tag{A27}$$

and thus substituting (A26) into (A27) yields

$$\hat{\gamma}(0) = (\mathbf{H}^T)^{-1} \text{Log}(\mathbf{\Lambda}) \mathbf{H}^{-1}. \quad (\text{A28})$$

Moreover, by the first equality in (A24) it follows that $\mathbf{\Sigma}_1 = (\mathbf{H}^T)^{-1} \mathbf{H}^{-1}$, and inserting it into (A28), we obtain

$$\begin{aligned} \hat{\gamma}(0) &= (\mathbf{H}^T)^{-1} \cdot \text{Log}(\mathbf{H}^T \mathbf{\Sigma}_2 \mathbf{H}) \cdot \mathbf{H}^{-1} \\ &= (\mathbf{H}^T)^{-1} \cdot \text{Log}(\mathbf{H}^{-1} \mathbf{\Sigma}_1^{-1} \mathbf{\Sigma}_2 \mathbf{H}) \cdot \mathbf{H}^{-1} \\ &= (\mathbf{H} \mathbf{H}^T)^{-1} \cdot \text{Log}(\mathbf{\Sigma}_1^{-1} \mathbf{\Sigma}_2) \\ &= \mathbf{\Sigma}_1 \cdot \text{Log}(\mathbf{\Sigma}_1^{-1} \mathbf{\Sigma}_2), \end{aligned} \quad (\text{A29})$$

i.e., the logarithmic mapping

$$\log_{\mathbf{\Sigma}_1}(\mathbf{\Sigma}_2) = \mathbf{\Sigma}_1 \text{Log}(\mathbf{\Sigma}_1^{-1} \mathbf{\Sigma}_2). \quad (\text{A30})$$

As a result, replacing the left hand side of (A30) with $\nu \in T_{\mathbf{\Sigma}_1} \mathcal{M}_{[\hat{x}, \cdot]}$, we obtain the exponential mapping

$$\exp_{\mathbf{\Sigma}_1}(\nu) = \mathbf{\Sigma}_1 \text{Exp}(\mathbf{\Sigma}_1^{-1} \nu). \quad (\text{A31})$$

References

- Du, J.; Fang, M.; Yu, Y.; Lu, G. An adaptive two-scale biomedical image fusion method with statistical comparisons. *Comput. Methods Programs Biomed.* **2020**, *196*, 105603. [[CrossRef](#)] [[PubMed](#)]
- Zhao, F.; Zhao, W.; Yao, L.; Liu, Y. Self-supervised feature adaption for infrared and visible image fusion. *Inf. Fusion* **2021**, *76*, 189–203. [[CrossRef](#)]
- Shafran-Nathan, R.; Etzion, Y.; Broday, D. Fusion of land use regression modeling output and wireless distributed sensor network measurements into a high spatiotemporally-resolved NO₂ product. *Environ. Pollut.* **2021**, *271*, 116334. [[CrossRef](#)]
- Xiao, K.; Wang, R.; Deng, H.; Zhang, L.; Yang, C. Energy-aware scheduling for information fusion in wireless sensor network surveillance. *Inf. Fusion* **2019**, *48*, 95–106. [[CrossRef](#)]
- Salcedo-Sanz, S.; Ghamisi, P.; Piles, M.; Werner, M.; Cuadra, L.; Moreno-Martínez, A.; Izquierdo-Verdiguier, E.; Muñoz-Marí, J.; Mosavi, A.; Camps-Valls, G. Machine learning information fusion in earth observation: A comprehensive review of methods, applications and data sources. *Inf. Fusion* **2020**, *63*, 256–272. [[CrossRef](#)]
- Liggins, M.E.; Hall, D.L.; Llinas, J. (Eds.) *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2009.
- Li, G.; Battistelli, G.; Yi, W.; Kong, L. Distributed multi-sensor multi-view fusion based on generalized covariance intersection. *Signal Process.* **2020**, *166*, 107246. [[CrossRef](#)]
- Bar-Shalom, Y.; Campo, L. The effect of the common process noise on the two-sensor fused-track covariance. *IEEE Trans. Aerosp. Electron. Syst.* **1986**, *22*, 803–805. [[CrossRef](#)]
- Li, X.R.; Zhu, Y.; Wang, J.; Han, C. Optimal linear estimation fusion—Part I: Unified fusion rules. *IEEE Trans. Inf. Theory* **2003**, *49*, 2192–2208. [[CrossRef](#)]
- Julier, S.J.; Uhlmann, J.K. A non-divergent estimation algorithm in the presence of unknown correlations. In Proceedings of the 1997 American Control Conference, Albuquerque, NM, USA, 6 June 1997; pp. 2369–2373.
- Reinhardt, M.; Noack, B.; Arambel, P.O.; Hanebeck, U.D. Minimum covariance bounds for the fusion under unknown correlations. *IEEE Signal Process. Lett.* **2015**, *22*, 1210–1214. [[CrossRef](#)]
- Julier, S.J.; Uhlmann, J.K. General decentralized data fusion with covariance intersection (CI). In *Handbook of Multisensor Data Fusion: Theory and Practice*; CRC Press: Boca Raton, FL, USA, 2001; pp. 12.1–12.5.
- Chong, C.Y.; Mori, S. Convex combination and covariance intersection algorithms in distributed fusion. In Proceedings of the 4th International Conference of Information Fusion, Sun City, South Africa, 1–4 November 2001; Volume 1, pp. 11–18.
- Eldar, Y.C.; Beck, A.; Teboulle, M. A minimax Chebyshev estimator for bounded error estimation. *IEEE Trans. Signal Process.* **2008**, *56*, 1388–1397. [[CrossRef](#)]
- Wang, Y.; Li, X.R. Distributed estimation fusion under unknown cross-correlation: An analytic center approach. In Proceedings of the 13th International Conference on Information Fusion, Edinburgh, UK, 26–29 July 2010; pp. 1–8.
- Hurley, M.B. An information theoretic justification for covariance intersection and its generalization. In Proceedings of the 5th International Conference on Information Fusion, Annapolis, MA, USA, 8–11 July 2002; pp. 505–511.
- Wang, Y.; Li, X.R. 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, 6–9 July 2009; pp. 571–578.

18. Battistelli, G.; Chisci, L. Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica* **2014**, *50*, 707–718. [[CrossRef](#)]
19. Costa, S.I.R.; Santos, S.A.; Strapasson, J.E. Fisher information distance: A geometrical reading. *Discret. Appl. Math.* **2015**, *197*, 59–69. [[CrossRef](#)]
20. Bishop, A.N. Information fusion via the Wasserstein barycenter in the space of probability measures: Direct fusion of empirical measures and Gaussian fusion with unknown correlation. In Proceedings of the 17th International Conference on Information Fusion, Salamanca, Spain, 7–10 July 2014; pp. 1–7.
21. Takatsu, A. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **2011**, *48*, 1005–1026.
22. Puccetti, G.; Rüschendorf, L.; Vanduffel, S. On the computation of Wasserstein barycenters. *J. Multivar. Anal.* **2020**, *176*, 104581. [[CrossRef](#)]
23. Amari, S.I.; Kurata, K.; Nagaoka, H. Information geometry of Boltzmann machines. *IEEE Trans. Neural Netw.* **1992**, *3*, 260–271. [[CrossRef](#)]
24. Amari, S.I. Information geometry of the EM and em algorithms for neural networks. *Neural Netw.* **1995**, *8*, 1379–1408. [[CrossRef](#)]
25. Cheng, Y.; Wang, X.; Morelande, M.; Moran, B. Information geometry of target tracking sensor networks. *Inf. Fusion* **2013**, *14*, 311–326. [[CrossRef](#)]
26. Eilders, M.J. Decentralized Riemannian Particle Filtering with Applications to Multi-Agent Localization. Ph.D. Thesis, Air Force Institute of Technology, Dayton, OH, USA, 2012.
27. Amari, S.I.; Kawanabe, M. Information geometry of estimating functions in semi-parametric statistical models. *Bernoulli* **1997**, *3*, 29–54. [[CrossRef](#)]
28. Rong, Y.; Tang, M.; Zhou, J. Intrinsic losses based on information geometry and their applications. *Entropy* **2017**, *19*, 405. [[CrossRef](#)]
29. Gómez, E.; Gomez-Viilegas, M.A.; Marín, J.M. A multivariate generalization of the power exponential family of distributions. *Commun. Stat.—Theory Methods* **1998**, *27*, 589–600. [[CrossRef](#)]
30. Samuel, K. *Multivariate t-Distributions and Their Applications*; Cambridge University Press: Cambridge, UK, 2004.
31. Kotz, S.; Kozubowski, T.J.; Podgórski, K. (Eds.) *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*; Birkhäuser: Boston, MA, USA, 2001.
32. Wen, C.; Cheng, X.; Xu, D. Filter design based on characteristic functions for one class of multi-dimensional nonlinear non-Gaussian systems. *Automatica* **2017**, *82*, 171–180. [[CrossRef](#)]
33. Chen, B.; Liu, X.; Zhao, H.; Príncipe, J.C. Maximum correntropy Kalman filter. *Automatica* **2017**, *76*, 70–77. [[CrossRef](#)]
34. Agamennoni, G.; Nieto, J.I.; Nebot, E.M. Approximate inference in state-space models with heavy-tailed noise. *IEEE Trans. Signal Process.* **2012**, *60*, 5024–5037. [[CrossRef](#)]
35. Do, M.N.; Vetterli, M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback–Leibler distance. *IEEE Trans. Image Process.* **2002**, *11*, 146–158. [[CrossRef](#)]
36. Tzagkarakis, G.; Beferull-Lozano, B.; Tsakalides, P. Rotation-invariant texture retrieval with Gaussianized steerable pyramids. *IEEE Trans. Image Process.* **2006**, *15*, 2702–2718. [[CrossRef](#)] [[PubMed](#)]
37. Verdoolaege, G.; Scheunders, P. Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination. *Int. J. Comput. Vis.* **2011**, *95*, 265–286. [[CrossRef](#)]
38. Xue, C.; Huang, Y.; Zhu, F.; Zhang, Y.; Chambers, J.A. An outlier-robust Kalman filter with adaptive selection of elliptically contoured distributions. *IEEE Trans. Signal Process.* **2022**, *70*, 994–1009. [[CrossRef](#)]
39. Huang, Y.; Zhang, Y.; Zhao, Y.; Chambers, J.A. A novel outlier-robust Kalman filtering framework based on statistical similarity measure. *IEEE Trans. Signal Process.* **2021**, *66*, 2677–2692. [[CrossRef](#)]
40. Zhu, F.; Huang, Y.; Xue, C.; Mihaylova, L.; Chambers, J.A. A sliding window variational outlier-robust Kalman filter based on Student’s *t* noise modelling. *IEEE Trans. Signal Process.* **2022**, *58*, 4835–4849.
41. Huang, Y.; Zhang, Y.; Zhao, Y.; Chambers, J.A. A novel robust Gaussian–Student’s *t* mixture distribution based Kalman filter. *IEEE Trans. Signal Process.* **2019**, *67*, 3606–3620. [[CrossRef](#)]
42. Tang, M.; Rong, Y.; Zhou, J.; Li, X.R. Information geometric approach to multisensor estimation fusion. *IEEE Trans. Signal Process.* **2019**, *67*, 279–292. [[CrossRef](#)]
43. Chen, X.; Zhou, J.; Hu, S. Upper bounds for Rao distance on the manifold of multivariate elliptical distributions. *Automatica* **2021**, *129*, 109604. [[CrossRef](#)]
44. Amari, S.I.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2000.
45. Absil, P.A.; Mahony, R.; Sepulchre, R. *Optimization Algorithms on Matrix Manifolds*; Princeton University Press: Princeton, NJ, USA, 2009.
46. Mitchell, A.F.S. The information matrix, skewness tensor and α -connections for the general multivariate elliptic distribution. *Ann. Inst. Stat. Math.* **1989**, *41*, 289–304. [[CrossRef](#)]
47. Calvo, M.; Oller, J.M. A distance between elliptical distributions based in an embedding into the Siegel group. *J. Comput. Appl. Math.* **2002**, *145*, 319–334. [[CrossRef](#)]
48. Berkane, M.; Oden, K.; Bentler, P.M. Geodesic estimation in elliptical distributions. *J. Multivar. Anal.* **1997**, *63*, 35–46. [[CrossRef](#)]
49. Huang, Y.; Zhang, Y.; Ning, L.; Chambers, J.A. Robust Student’s *t* based nonlinear filter and smoother. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 2586–2596. [[CrossRef](#)]

50. Huang, Y.; Zhang, Y.; Chambers, J.A. A novel Kullback–Leibler divergence minimization-based adaptive Student’s *t*-filter. *IEEE Trans. Signal Process.* **2019**, *67*, 5417–5432. [[CrossRef](#)]
51. Schewpe, F.C. *Uncertain Dynamic Systems*; Prentice Hall: Englewood Cliffs, NJ, USA, 1973.
52. Moakher, M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **2005**, *26*, 735–747. [[CrossRef](#)]
53. Said, S.; Bombrun, L.; Berthoumieu, Y.; Manton, J.H. Riemannian Gaussian distributions on the space of symmetric positive definite matrices. *IEEE Trans. Inf. Theory* **2017**, *63*, 2153–2170. [[CrossRef](#)]
54. Ilea, I.; Bombrun, L.; Terebes, R.; Borda, M.; Germain, C. An M-estimator for robust centroid estimation on the manifold of covariance matrices. *IEEE Signal Process. Lett.* **2016**, *23*, 1255–1259. [[CrossRef](#)]
55. Fletcher, P.T.; Venkatasubramanian, S.; Joshi, S. The geometric median on Riemannian manifolds with application to robust atlas estimation. *Neuroimage* **2009**, *45*, S143–S152. [[CrossRef](#)]
56. Hajri, H.; Ilea, I.; Said, S.; Bombrun, L.; Berthoumieu, Y. Riemannian Laplace distribution on the space of symmetric positive definite matrices. *Entropy* **2016**, *18*, 98. [[CrossRef](#)]
57. Arnaudon, M.; Barbaresco, F.; Yang, L. Riemannian medians and means with applications to radar signal processing. *IEEE J. Sel. Top. Signal Process.* **2013**, *7*, 595–604. [[CrossRef](#)]
58. Wong, K.M.; Zhang, J.K.; Liang, J.; Jiang, H. Mean and median of PSD matrices on a Riemannian manifold: Application to detection of narrow-band sonar signals. *IEEE Trans. Signal Process.* **2017**, *65*, 6536–6550. [[CrossRef](#)]
59. Grove, K.; Karcher, H.; Ruh, E. Jacobi fields and Finsler metrics on compact Lie groups with an application to differentiable pinching problems. *Math. Ann.* **1974**, *211*, 7–21. [[CrossRef](#)]
60. Arsigny, V.; Fillard, P.; Pennec, X.; Ayache, N. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **2007**, *29*, 328–347. [[CrossRef](#)]
61. Li, P.; Wang, Q.; Zeng, H.; Zhang, L. Local log-Euclidean multivariate Gaussian descriptor and its application to image classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 803–817. [[CrossRef](#)]
62. Skovgaard, L.T. A Riemannian geometry of the multivariate normal model. *Scand. J. Stat.* **1984**, *11*, 211–223.
63. Hunter, D.R.; Lange, K. A tutorial on MM algorithms. *Am. Stat.* **2004**, *58*, 30–37. [[CrossRef](#)]