# Score matching based assumed density filtering with the von Mises-Fisher distribution

Mario Bukal
University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Applied Mathematics
Unska 3, HR-10000 Zagreb, Croatia
Email: mario.bukal@fer.hr

Ivan Marković and Ivan Petrović
University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Control and Computer Engineering
Unska 3, HR-10000 Zagreb, Croatia
Email: ivan.markovic@fer.hr, ivan.petrovic@fer.hr

***Abstract*—Bayesian filters are often used in statistical inference and consist of recursively alternating between two steps: prediction and correction. Most commonly the Gaussian distribution is used within the Bayes filtering framework, but other distributions, which could model better the nature of the estimated phenomenon like the von Mises-Fisher distribution on the unit sphere, have also been subject of research interest. However, the von Mises-Fisher filter requires approximations since the prediction step does not yield an another von Mises-Fisher distribution. Furthermore, other advanced filtering methods require approximating a mixture of distributions with just a single component. In this paper we propose to use the score matching within the context of Bayesian assumed density filtering inlieu of the more common moment matching. Moment matching functions by assuming the type of the resulting distribution and then matching its moments with the prior distribution, which in the end minimizes the Kullback–Leibler divergence. Score matching also assumes the resulting distribution type, but finds optimal parameters by minimizing the relative Fisher information. In the paper we show that the score matching procedure results with identical performance, but with simpler equations that, unlike moment matching, do not require tedious numerical methods. In the end, we corroborate theoretical results by running the moment and score matching based filters for single and multiple object tracking on a large number of randomly generated trajectories on the unit sphere.**

## I. INTRODUCTION

Bayesian estimation is a widely used and powerful tool in statistical inference. For estimating the state of dynamic systems the procedure, called Bayesian filtering, evolves around formulating a prior distribution and iterating recursively two key steps: the prediction and the correction step. The prediction step entails computing an integral, i.e., convolution, of two probability density functions (pdfs), while the update step entails using the Bayes rule in order to obtain the posterior distribution. Prior to the Bayes filter design, the most appropriate models that describe well the state evolution and measurement generation are selected and, if they are nonlinear, how to best deal with these nonlinearities. Further challenges lie in modeling the noise in the system, dealing with false alarms/clutter, and if multiple object tracking is in question, also on how to deal with the measurement-to-object associations.

Another important aspect in Bayes filter design, which has attracted attention in recent years, is the choice of the distribution that best describes the underlying geometry of the phenomenon to be estimated. Most commonly for the Euclidean space the Gaussian distribution is used and the Bayesian filter, in that case, yields the linear Kalman filter. However, if the underlying geometry is not Euclidean, but rather, e.g., a unit sphere, other distributions such as the von Mises [1]–[5], von Mises-Fisher [6]–[8], Bingham [9]–[11], Fisher-Bingham and Kent distributions [12] can be used. Furthermore, for the case of bearing-only tracking the shifted Rayleigh filter was presented in [13] that exploits the essential structure of the nonlinearities present is such a tracking problem. Solving the Bayes filter equations can be challenging, and, moreover, directional distributions often do not stay closed in the same distribution space. For example, in the prediction step the von Mises-Fisher distribution will not yield a von Mises-Fisher distribution and approximation techniques need to be applied.

An example of such an approximation technique is moment matching, where the moments of the resulting distribution are matched to the moments of the prior distribution. This procedure is also the optimal choice in the sense of the Kullback-Leibler divergence [14]. Moment matching was analyzed in [5] for circular recursive non-linear estiamation. Therein, the authors presented a general framework for estimation of a circular state based on the wrapped normal distribution and the von Mises distribution. In [7] we have shown how moment matching can be used for the prediction step of the $(d-1)$–dimensional von Mises-Fisher filter, and how a mixture of von Mises-Fisher distributions can be approximated by a single von Mises-Fisher component [15]. This step was essential to obtain formulae for multiple object tracking techniques such as the joint probabilistic data association (JPDA) filter [16] and random finite sets (RFS) based filters [17] implemented with Gaussian mixtures, like the Gaussian mixture probability hypothesis density (PHD) filter [18], [19]. In [12] a method of *score matching estimation* was presented for compact oriented Riemannian manifolds with the main advantage being that for exponential family models, the need to work with awkward normalizing constants is avoided. A minor limitation of the method is that it requires the underlying distributions to have sufficiently smooth probability densities. This procedure was tailored to minimize the, therein dubbed, *Hyvärinen divergence*, which is in fact the *relative Fisher information* [20], [21]. The score matching estimator was originally developed for densities on the Euclidean space in [22] for the purpose of estimating statistical models where pdf is known only up to a multiplicative constant.

In this paper we study the score matching estimator within the context of Bayesian estimation. Instead of using the moment matching principle for the prediction approximation and mixture component reduction, we propose to use the score matching. As a study example we use the von Mises Fisher distribution and derive the corresponding equations for the basic von Mises-Fisher filter, and probabilistic data association (PDA) and JPDA filters. The procedure actually amounts to minimizing the relative Fisher information, inlieu of Kullback-Leibler divergence. Unlike moment matching, which results in transcendental equations, the resulting score matching equations are linear. This results in computational advantages since numerical minimization techniques can be avoided. Performance wise, both moment and score matching produce the same results. This is surprising in some way, but it may be related to the fact that we are dealing with an exponential family. However, at the moment we cannot prove that moment and score matching methods should give the same results in the case of exponential families. In the end, we have conducted Monte Carlo simulations on a synthetic dataset in order to corroborate the theoretical results.

The paper is organized as follows. Section II introduces the theoretical background, formulates the problem, and also introduces the principles of moment and score matching. Section III derives the score matching formulae for the von Mises-Fisher filters, including the mixture component reduction required for the PDA and JPDA filters. Section IV presents the results of the synthetic data experiments, while in the end Section V concludes the paper.

## II. Problem statement

Bayesian filtering has the aim to estimate the posterior distribution of a state $\boldsymbol{x}^k$ based on the history of measurements $\boldsymbol{z}^{1:k}$, i.e., the pdf $p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k})$, by iterating two steps — prediction and update. In the prediction step, prior density $p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k-1})$ is obtained as a convolution of the state transition density $p(\boldsymbol{x}^k|\boldsymbol{x}^{1:k-1})$ with the posterior $p(\boldsymbol{x}^{k-1}|\boldsymbol{z}^{1:k-1})$ density from the previous time instance:

$$p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k-1}) = \int p(\boldsymbol{x}^k|\boldsymbol{x}^{1:k-1})p(\boldsymbol{x}^{k-1}|\boldsymbol{z}^{1:k-1})\mathrm{d}\boldsymbol{x}^{k-1}. \quad (1)$$

With having the predicted pdf computed, the Bayes rule is used to obtain the updated (posterior) density

$$p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k}) = \frac{p(\boldsymbol{z}^k|\boldsymbol{x}^k)p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k-1})}{p(\boldsymbol{z}^k|\boldsymbol{z}^{1:k-1})}\,, \quad (2)$$

where $p(\boldsymbol{z}^k|\boldsymbol{x}^k)$ denotes the measurement likelihood and the Bayes normalizer is given by

$$p(\boldsymbol{z}^k|\boldsymbol{z}^{1:k-1}) = \int p(\boldsymbol{z}^k|\boldsymbol{x}^k)p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k-1})\mathrm{d}\boldsymbol{x}^k\,. \quad (3)$$

Given the aforementioned formulae, the Bayes filter recursively estimates the posterior by alternating (1) and (2).

In moving object tracking other momenta appear which render the basic Bayes filtering insufficient. Namely, in the case of realistic single target tracking there are imperfect sensor detection and clutter, while in tracking multiple objects, there is also the problem of data association. To solve these problems seminal approaches can be applied, e.g., the PDAF for single object tracking in clutter, JPDAF [16], [23] and RFS based filters [17], like the PHD filter [18], for multiple object tracking in clutter. What is in common to data association filters and RFS filters, with implementation based on mixtures, is that at one point during the recursion the posterior is a mixure of densities

$$p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k}) = \sum_j w_j p_j(\boldsymbol{x}^k|\boldsymbol{z}^{1:k}), \quad (4)$$

which needs to be reduced to a single component density. In the PDAF and JPDAF case this is due to multiple data association hypotheses being generated thanks to clutter or multiple objects or both. Many of the RFS filters base their implementation on mixtures of distributions, like the Gaussian mixture PHD and CPHD [19], [24] filters and the recursion becomes intractable unless mixture component number reduction schemes are employed.

In this section we approach the problem formulation by assuming that the underlying distributions belong to an *exponential family*, which indeed includes many of well-known distributions like Gaussian, exponential, Poisson, von Mise-Fisher, etc. [25] A parametric set of probability distributions parametrized by the natural parameter $\boldsymbol{\theta} \in \mathbb{R}^d$ is called an exponential family if their pdf can be written in the form [25]

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \exp\left(T(\boldsymbol{x})\cdot\boldsymbol{\theta} - F(\boldsymbol{\theta}) + C(\boldsymbol{x})\right)\,, \quad (5)$$

where $T(\boldsymbol{x})$ is the sufficient statistics, $F(\boldsymbol{\theta})$ is the log-normalizing function and $C(\boldsymbol{x})$ denotes the carrier measure.

Our study example in the present paper is the von Mises-Fisher distribution, which is a parametric probability distribution on the unit $(d-1)$–dimensional sphere $\mathbb{S}^{d-1}$

$$p(\boldsymbol{x};\boldsymbol{\mu},\kappa) = C_d(\kappa)\exp(\kappa\boldsymbol{\mu}\cdot\boldsymbol{x}), \quad \boldsymbol{x}\in\mathbb{S}^{d-1}\,, \quad (6)$$

where $\kappa \geq 0$ and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$ denote the concentration and the mean direction, respectively, while expression

$$C_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2}I_{d/2-1}(\kappa)} \quad (7)$$

is the normalization constant with respect to the uniform surface measure and $I_p$ denotes the modified Bessel function of the first kind and order $p$ [26]. The von Mises-Fisher distributions constitute an exponential family parametrized by the natural parameter $\boldsymbol{\theta} = \kappa\boldsymbol{\mu} \in \mathbb{R}^d$, with the log-normalizer given by

$$F_d(\boldsymbol{\theta}) = -\log C_d(|\boldsymbol{\theta}|)\,, \tag{8}$$

$C(\boldsymbol{x}) = 0$, and the minimal sufficient statistics $T(\boldsymbol{x}) = \boldsymbol{x}$.

In [7] we discussed on models and how von Mises-Fisher distribution can be used to describe all three ingredients of the Bayesian filter: posterior, state transition and measurement likelihood distribution. With the underlying distributions defined, all that remains is to evaluate (1)–(3) to form the basic Bayes filter, and (4) in order to obtain a tool so that the resulting filter can be applied in more advanced techniques like the data association filters or mixture implementations of RFS based filters. Taking the starting distribution in the Bayesian filter to be a member of an exponential family, there is no guarantee that distributions in (1)–(2) will remain in the same exponential family. Typically they will not, like for instance von Mises-Fisher distributions. Hence, the main challenge therein is to find distributions of the given type, which are in some sense close to the obtained distributions in (1)–(2). A common approch is to use statistical divergences as measures of distance. In the sequel, we analyze and compare two methods: (i) moment matching and (ii) score matching, which find the closest to (1)–(2) exponential family distributions by minimizing the Kullback-Leibler divergence and the relative Fisher information, respectively. And we present explicit results for the von Mises-Fisher distributions.

### A. Moment matching

Kullback-Leibler divergence (or relative entropy) between two distributions with densities $p$ and $q$ is defined by the following expression [20], [21]

$$\mathcal{H}(p,q) = \int p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x}\,. \tag{9}$$

In applications, $p$ is typically the true distribution (density) and $q$ its approximation, and (9) measures the information loss after substituting $p$ by $q$. Hence, in our case, distributions (1)–(2) will play the role of $p$, and we will look for a von Mises-Fisher distribution $q$, which minimizes (9). On the other hand, it is easy to see that the above minimization of the Kullback-Leibler divergence is equivalent to the maximization of the Shannon entropy $-\int q(\boldsymbol{x})\log q(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ under the moment constraint, i.e., $\int \boldsymbol{x}p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \int \boldsymbol{x}q(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$, hence the name moment matching. Since the moment of the $(d-1)$–dimensional von Mises-Fisher distribution $q$ calculates as [7]

$$\mathbb{E}[\boldsymbol{x}] = \int_{\mathbb{S}^{d-1}} \boldsymbol{x} q(\boldsymbol{x};\boldsymbol{\mu},\kappa)\mathrm{d}\boldsymbol{x} = \nabla F_d(\boldsymbol{\theta}) = A_d(\kappa)\boldsymbol{\mu}\,, \tag{10}$$

yielding the *directional mean*, where $A_d(\kappa)$ is the ratio of the following Bessel functions

$$A_d(\kappa) = \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}\,, \tag{11}$$

the optimal von Mises-Fisher distribution of the same moment $\boldsymbol{m}$ as $p$ is given by the natural parameter $\boldsymbol{\theta}^*$, which solves the following equation

$$\nabla F_d(\boldsymbol{\theta}) = \boldsymbol{m}\,. \tag{12}$$

For the prediction step (1), the directional mean of the predicted von Mises-Fisher evaluates to

$$\mathbb{E}[\boldsymbol{x}^k|\boldsymbol{z}^{1:k-1}] = A_d(\kappa_\tau)A_d(\kappa^{k-1})\boldsymbol{\mu}^{k-1} \tag{13}$$

where $\kappa_\tau$ is, e.g., process noise concentration parameter. The next step is to find moments of the assumed von Mises-Fisher distribution, $q(\boldsymbol{x};\boldsymbol{\mu}^k,\boldsymbol{\mu}^k)$, that match (13). These are as follows [7]

$$\boldsymbol{\mu}^k = \boldsymbol{\mu}^{k-1}\,, \qquad \kappa^k = A_d^{-1}(A_d(\kappa_\tau)A_d(\kappa^{k-1}))\,. \tag{14}$$

A similar procedure can be carried out for a mixture of von Mises-Fisher distributions, $\sum_j w_j p(\boldsymbol{x};\mu_j^k,\kappa_j^k)$, and the corresponding directional mean is

$$\mathbb{E}[\boldsymbol{x}^k|z^{1:k}] = \sum_j w_j A_d(\kappa_j^k)\boldsymbol{\mu}_j^k\,, \tag{15}$$

while the single moment matched component, $p(\boldsymbol{x};\mu^k,\kappa^k)$ , is found by solving the following equations [7]

$$\kappa^k = A_d^{-1}\left(\Big|\sum_j w_j A_d(\kappa_j^k)\boldsymbol{\mu}_j^k\Big|\right), \tag{16}$$

$$\boldsymbol{\mu}^k = \Big(\sum_j w_j A_d(\kappa_j^k)\boldsymbol{\mu}_j^k\Big)\Big/A_d(\kappa^k)\,. \tag{17}$$

From previous relations we can notice that for the von Mises-Fisher distribution the resulting moment matching equations are transcendental and one needs to resort to numerical methods in order to compute the inverse $A_d^{-1}(\cdot)$. In the sequel we turn our attention to the explicit case of the von Mises-Fisher distribution on the unit 2–sphere.

### B. Score matching

Another important statistical divergence measure is the relative Fisher information defined by [20], [21]

$$\mathcal{I}(p,q) = \int_{\mathbb{S}^2} p(\xi)\left|\nabla_\xi \log\frac{p(\xi)}{q(\xi)}\right|^2 \mathrm{d}\sigma(\xi)\,, \tag{18}$$

where $\xi$ denote local coordinates on $\mathbb{S}^2$. Let $p$ be given distribution on $\mathbb{S}^2$, we want to find von Mises-Fisher distribution $q$ with natural parameter $\boldsymbol{\theta} = \kappa\boldsymbol{\theta}$, i.e.,

$$q(\boldsymbol{x}) = C_3(\kappa)e^{\kappa\boldsymbol{\mu}\cdot\boldsymbol{x}}\,,$$

which minimizes the relative Fisher information (18). Note that distributions in (18) are given in local coordinates on $\mathbb{S}^2$, while we now write $q$ in Euclidean coordinates. First, using the Stokes theorem, the relative Fisher information can be written as the following integral [12]

$$\begin{aligned}\mathcal{I}(p,q) = \int_{\mathbb{S}^2} p(\xi)\,\big(&|\nabla_\xi \log q(\xi)|^2 \\ &+2\Delta_\xi \log q(\xi)\big)\,\mathrm{d}\sigma(\xi)\,,\end{aligned} \tag{19}$$

where $\nabla_\xi$ and $\Delta_\xi$ denote the gradient and Laplace (Laplace-Beltrami) operators in local coordinates on the 2–sphere.

Relations between local and Euclidean coordinates are the following

$$\begin{aligned} |\nabla_\xi \log q(\xi)|^2 &= |P(\boldsymbol{x})\nabla_{\boldsymbol{x}} \log q(\boldsymbol{x})|^2 \,, \\ \Delta_\xi \log q(\xi) &= \mathrm{Tr}(P(\boldsymbol{x})\nabla_{\boldsymbol{x}}^T P(\boldsymbol{x})\nabla_{\boldsymbol{x}} \log q(\boldsymbol{x}))\,, \end{aligned}$$

where $\mathrm{Tr}(\cdot)$ designates matrix trace and $P(\boldsymbol{x}) = I_3 - \boldsymbol{x}\boldsymbol{x}^T$ denotes the projection operator. Straightforward calculations then yield for the first term in (19):

$$\begin{aligned} \nabla_\xi \log q(\xi) &= P(\boldsymbol{x})\nabla_{\boldsymbol{x}}(\boldsymbol{\theta}\cdot\boldsymbol{x} - F(\boldsymbol{\theta})) \\ &= P(\boldsymbol{x})\boldsymbol{\theta} = \boldsymbol{\theta} - (\theta\cdot\boldsymbol{x})\boldsymbol{x}\,, \end{aligned}$$

from which we can easily compute the squared norm

$$|P(\boldsymbol{x})\boldsymbol{\theta}|^2 = |\boldsymbol{\theta}|^2 - (\boldsymbol{\theta}\cdot\boldsymbol{x})^2.$$

Next, the second term in (19) evalutes to

$$\begin{aligned} \Delta_\xi \log q(\xi) &= \mathrm{Tr}(P(\boldsymbol{x})\nabla_{\boldsymbol{x}}^T P(\boldsymbol{x})\boldsymbol{\theta}) \\ &= \mathrm{Tr}(-(\boldsymbol{\theta}\cdot\boldsymbol{x})I_3 - \boldsymbol{\theta}\boldsymbol{x}^T + 2(\boldsymbol{\theta}\cdot\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^T) \\ &= -2\boldsymbol{\theta}\cdot\boldsymbol{x}\,. \end{aligned}$$

Given the above relations, (19) now simplifies to the following integral

$$\mathcal{I}(p,q) = \int_{\mathbb{S}^2} (|\boldsymbol{\theta}|^2 - (\boldsymbol{\theta}\cdot\boldsymbol{x})^2 - 4\boldsymbol{\theta}\cdot\boldsymbol{x}) p(\boldsymbol{x})\mathrm{d}\sigma(\boldsymbol{x})\,,$$

which remains to be minimized with respect to $\boldsymbol{\theta}$. Taking the gradient with respect to $\boldsymbol{\theta}$, we obtain the equation

$$\boldsymbol{\theta} - \int_{\mathbb{S}^2} (\boldsymbol{\theta}\cdot\boldsymbol{x})\boldsymbol{x} p(\boldsymbol{x})\mathrm{d}\sigma(\boldsymbol{x}) = 2\boldsymbol{m}\,,$$

where $\boldsymbol{m} = \mathbb{E}(\boldsymbol{x})$. The latter yields the linear equation for $\boldsymbol{\theta}$

$$(\mathbb{I}_3 - \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^T))\boldsymbol{\theta} = 2\boldsymbol{m}\,, \tag{20}$$

with $\mathbb{I}_3$ the indentity matrix, and which is far more convenient to deal with, than the moment matching counterpart (12).

## III. Assumed density filtering with the von Mises-Fisher distribution

In this section we calculate the optimal von Mises-Fisher distribution, in the sense of the score matching method, for the predicted distribution (1) of the Bayes filter, and for the mixture of von Mises-Fisher distributions.

According to [7], $p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k-1})$ is given by

$$\begin{aligned} &p(\boldsymbol{x}^k|\boldsymbol{z}^{1:k-1}) \\ &= C_3(\kappa_\tau)C_3(\kappa^{k-1})\int_{\mathbb{S}^2} e^{(\kappa_\tau \boldsymbol{x}^k + \kappa^{k-1}\boldsymbol{\mu}^{k-1})\cdot\boldsymbol{x}^{k-1}} \mathrm{d}\sigma(\boldsymbol{x}^{k-1})\,. \end{aligned} \tag{21}$$

In order to compute the optimal $\boldsymbol{\theta}$ in (20), we have to first compute the moment vector $\boldsymbol{m}$ and matrix $M = \mathbb{I}_3 - \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^T)$. We have already seen in (13) that

$$\boldsymbol{m} = A_3(\kappa_\tau)A_3(\kappa^{k-1})\boldsymbol{\mu}^{k-1}\,. \tag{22}$$

It remains to evaluate matrix $M$, i.e., the matrix of second moments $\mathbb{E}(\boldsymbol{x}^k(\boldsymbol{x}^k)^T)$. Using the calculations provided in the Appendix, for $i,j = 1,2,3$, we obtain:

$$\begin{aligned} \mathbb{E}(x_i^k x_j^k) &= \int_{\mathbb{S}^2} x_i^k x_j^k p(\boldsymbol{x}^k)\mathrm{d}\sigma(\boldsymbol{x}^k) \\ &= C_3(\kappa_\tau)C_3(\kappa^{k-1})\int_{\mathbb{S}^2} e^{\kappa^{k-1}\boldsymbol{\mu}^{k-1}\cdot\boldsymbol{x}^{k-1}} \\ &\qquad \left(\int_{\mathbb{S}^2} x_i^k x_j^k e^{\kappa_\tau \boldsymbol{x}^k\cdot\boldsymbol{x}^{k-1}}\mathrm{d}\sigma(\boldsymbol{x}^k)\right)\mathrm{d}\sigma(\boldsymbol{x}^{k-1}) \\ &= C_3(\kappa^{k-1})\left(-B_3(\kappa_\tau) + 2A_3^2(\kappa_\tau) - \frac{A_3(\kappa_\tau)}{\kappa_\tau}\right) \\ &\qquad \int_{\mathbb{S}^2} x_i^{k-1}x_j^{k-1} e^{\kappa^{k-1}\boldsymbol{\mu}^{k-1}\cdot\boldsymbol{x}^{k-1}}\mathrm{d}\sigma(\boldsymbol{x}^{k-1}) \\ &\quad + C_3(\kappa^{k-1})\frac{A_3(\kappa_\tau)}{\kappa_\tau}\int_{\mathbb{S}^2}\delta_{ij} e^{\kappa^{k-1}\boldsymbol{\mu}^{k-1}\cdot\boldsymbol{x}^{k-1}}\mathrm{d}\sigma(\boldsymbol{x}^{k-1}) \\ &= \left(-B_3(\kappa_\tau) + 2A_3^2(\kappa_\tau) - \frac{A_3(\kappa_\tau)}{\kappa_\tau}\right) \\ &\qquad \times\Bigg(\bigg(-B_3(\kappa^{k-1}) + 2A_3^2(\kappa^{k-1}) \\ &\qquad - \frac{A_3(\kappa^{k-1})}{\kappa^{k-1}}\bigg)\mu_i^{k-1}\mu_j^{k-1} \\ &\qquad + \frac{A_3(\kappa^{k-1})}{\kappa^{k-1}}\delta_{ij}\Bigg) + \frac{A_3(\kappa_\tau)}{\kappa_\tau}\delta_{ij}\,, \end{aligned}$$

where $B_3(\kappa) = C_3''(\kappa)/C_3(\kappa)$ (cf. (31)) and $\delta_{ij}$ denotes the Kronecker symbol. Thus, if $M$ is regular we obtain

$$\boldsymbol{\theta} = 2M^{-1}\boldsymbol{m}\,, \tag{23}$$

and the optimal parameters of the score matched distribution are determined as

$$\begin{aligned} \kappa^k &= |2A_3(\kappa_\tau)A_3(\kappa^{k-1})M^{-1}\boldsymbol{\mu}^{k-1}| \\ \boldsymbol{\mu}^k &= \frac{2A_3(\kappa_\tau)A_3(\kappa^{k-1})}{\kappa^k}M^{-1}\boldsymbol{\mu}^{k-1}. \end{aligned} \tag{24}$$

By inspecting the abover relations we can notice that parameters can be computed in a straightforward manner, as opposed to the transcendental form of (14).

If we now assume that $p$ is a mixture of von Mises-Fisher distributions:

$$p(\boldsymbol{x}) = \sum_{l=1}^{L}\omega_l C_3(\kappa_l)e^{\kappa_l\boldsymbol{\mu}_l\cdot\boldsymbol{x}}\,,$$

then the matrix of second moment evaluates according to

$$\mathbb{E}(x_i x_j) = \sum_{l=1}^{L}\omega_l C_3(\kappa_l)\int_{\mathbb{S}^2} x_i x_j e^{\kappa_l\boldsymbol{\mu}_l\cdot\boldsymbol{x}}\mathrm{d}\sigma(\boldsymbol{x}) \tag{25}$$

$$= \sum_{l=1}^{L}\omega_l\left(\left(-B_3(\kappa_l) + 2A_3^2(\kappa_l) - \frac{A_3(\kappa_l)}{\kappa_l}\right)\mu_i^l\mu_j^l + \frac{A_3(\kappa_l)}{\kappa_l}\delta_{ij}\right)\,, \tag{26}$$

and the corresponding parameters of the approximating single component can be computed from (23) as in (24).

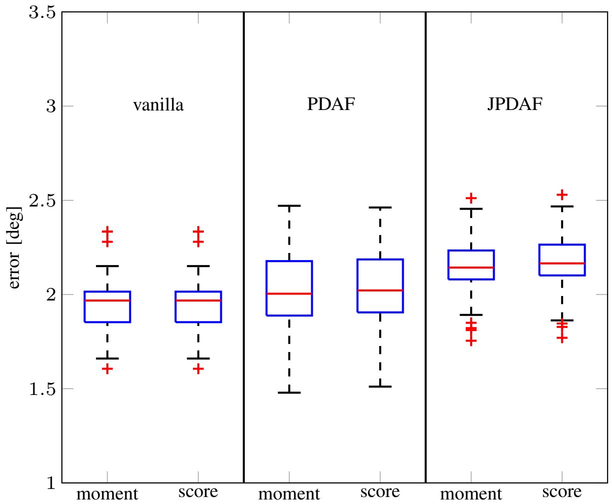


Fig. 1. Matlab's boxplot of the von Mises-Fisher basic, PDA and JPDA filters on 100 randomly generated trajectories on the unit 2–sphere.

## IV. Experimental results

In order to validate the filter performance based on score matching and compare it to the moment matching approach we generated a series of synthetic experiments. We performed tests for three different filter tracking examples: (i) basic von Mises-Fisher filter, (ii) von Mises-Fisher PDA filter, and (iii) von Mises-Fisher JPDA filter. For all the three examples we had moving object dynamics with the mean value of angular velocities 0.5 deg/step, and the standard deviation was set to $10^{-3}$ deg/step. For the first example, i.e., the basic von Mises-Fisher filter, we generated 100 Monte Carlo runs where the probability of detection was $p_D = 1$, no clutter was present, and von Mises-Fisher sensor noise concentration parameter was set to corresponding 5 deg. For the von Mises-Fisher PDA filter, we generated also 100 Monte Carlo runs where the probability of detection was $p_D = 0.95$, clutter rate was set to $\lambda = 1.25{\cdot}4\pi$, and von Mises-Fisher sensor noise concentration parameter was set to corresponding 5 deg. Finally, for the von Mises-Fisher JPDA filter, we generated 100 Monte Carlo runs with the same parameters as for the PDA filter case, except that the number of objects in the scene was an integer value randomly drawn from the interval $[1, 5]$. The estimation error was computed as the great circle distance between the ground truth and the estimated state:

$$\Delta_{\text{err}} = \arccos(\boldsymbol{\mu}_{\text{gt}} \cdot \boldsymbol{\mu}^k). \tag{27}$$

Note that for the von Mises-Fisher JPDA filter the Hungarian algorithm [27], [28] was ran in order to optimally associate the filters to the ground truth trajectories. The filters had the same parameters set, the process noise was set to $\kappa_\tau = 750$, while the measurement noise was set so as to have the concentration parameter corresponding to the standard deviation of 5 deg.

Figure 1 shows the results of moment and score matching von Mises-Fisher variants for the basic, PDA and JPDA filter trajectories. We can see from the figure that the results are practically identical; the basic von Mises-Fisher filter had the same median value of the estimation error. The PDA filter had the median of 2.00 deg and 2.02 deg, while JPDA filter had the median of 2.14 deg and 2.16 deg for the moment and score matching, respectively. If we look the moment and score matching results for the von Mises-Fisher filter prediction, i.e., (14) and (23), respectively, we can see that the formulae are quite different, but the simulations corroborate that the obtained result is the same. Indeed, the prediction yields the same mean value and concentration parameter for both moment and score matching. Minor differences in the PDA and JPDA filter can be attributed to to update step, which essentially entails reducing a mixture of von Mises-Fisher distributions to a single von Mises-Fisher component. In this step moment matching results with a transcendental equation for which numerical optimization needs to be employed, while score matching has a straightforward calculation. Optimization can yield slightly different values for the concentration parameter which propagates in the future steps, thus accounting for slight differences in the final result of Monte Carlo simulations. The experimental results come as somewhat of a surprise. We hypothesize that the minimizer of both the Kullback-Leibler divergence and relative Fisher information is the same; however, at the moment we do not have a formal proof for this hypothesis. Nevertheless, advantages from the computational complexity perspective are still valid.

Furthermore, when working with the von Mises-Fisher distribution, care should be taken when it comes to evaluating the distribution, since for larger concentration parameters one can encounter numerical difficulties. Namely, the $\exp$ and $\sinh$ function can quickly exceed the range of double precision floating point representation. Evaluation of the von Mises-Fisher distribution is a necessary step in the PDA and JPDA filters; first in validation gating and then in computing the association probabilities. Numerical issues can be avoided by using the method described in [29], [30], while discussion for the case of the von Mises distribution can be found in [31]. These approaches basically entail lifting the distribution normalizer to the exponent, and approximating the resulting logarithm of the Bessel function by a series expansion.

## V. Conclusion

In this paper we have studied the application of the score matching within the context of Bayesian filtering. For some classes of distributions, like the von Mises-Fisher distribution, filter prediction does not yield an another von Mises-Fisher distribution, hence approximation techniques need to applied. Furthermore, filtering methods that can deal with false alarms/clutter and multiple moving objects, require approximation of a distribution mixture by a single component. For example, the PDA and JPDA filters, due to multiple hypothesis generation, the posterior density is a mixture and the filter principle requires the posterior mixture to be reduced to a single distribution. In this paper, as a study example, we chose the von Mises-Fisher distribution for which we derived the prediction equation approximation and mixture component reduction formulae for score matching. As opposed to moment matching, which yields transcendental equations, score matching resulted with equations that do not require numerical procedures. The experimental Monte Carlo simulations showed performance of both filters to be identical. This was a surprising result, and we hypothesize that it is related to the fact that we are dealing with an exponential family. However, for the moment we cannot prove that these two different methods should yield the same results. The procedure is not limited to the von Mises-Fisher distribution, and the presented framework can be applied to other directional distributions of interest that constitute an exponential family of distributions.


## ACKNOWLEDGMENT

This research has been supported by the Ministry of Science and Education of the Republic of Croatia under the grant "Centre of Research Excellence for Data Science and Cooperative Systems".


## APPENDIX

Here we provide some necessary auxilliary identities obtained by straightforward calculations. By first calculating the directional mean of the von Mises-Fisher distribution

$$\mathbb{E}(\boldsymbol{x}) = C(\kappa) \int_{\mathbb{S}^2} \boldsymbol{x} e^{\kappa \boldsymbol{\mu} \cdot \boldsymbol{x}} \mathrm{d}\sigma(\boldsymbol{x}) = A_3(\kappa)\boldsymbol{\mu} \,. \tag{28}$$

we can derive the matrix of second moments which equals

$$\begin{aligned} \mathbb{E}(x_i x_j) &= C(\kappa) \int_{\mathbb{S}^2} x_i x_j e^{\kappa \boldsymbol{\mu} \cdot \boldsymbol{x}} \mathrm{d}\sigma(\boldsymbol{x}) \\ &= \frac{\partial^2 F(\kappa \boldsymbol{\mu})}{\partial \theta_i \partial \theta_j} + A_3^2(\kappa) \mu_i \mu_j \,, \end{aligned} \tag{29}$$

where

$$\begin{aligned} \frac{\partial^2 F(\kappa \boldsymbol{\mu})}{\partial \theta_i \partial \theta_j} = &\left( -B_3(\kappa) + A_3^2(\kappa) - \frac{A_3(\kappa)}{\kappa} \right) \mu_i \mu_j \\ &+ \frac{A_3(\kappa)}{\kappa} \delta_{ij} \,, \end{aligned} \tag{30}$$

and

$$B_3(\kappa) = \frac{C_3''(\kappa)}{C_3(\kappa)} = \frac{1}{\tanh^2 \kappa} - \frac{2}{\kappa \tanh \kappa} + \frac{1}{\sinh^2 \kappa} \,. \tag{31}$$

## REFERENCES


[1] M. Azmani, S. Reboul, J.-B. Choquel, and M. Benjelloun, "A recursive fusion filter for angular data," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2009, pp. 882–887.

[2] I. Marković, I. Petrović, I. Markovic, and I. Petrovic, "Bearing-only tracking with a mixture of von Mises distributions," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 707–712.

[3] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive monlinear filtering for angular data based on circular distributions," in *American Control Conference (ACC)*, 2013, pp. 5439–5445.

[4] G. Stienne, S. Reboul, M. Azmani, J. Choquel, and M. Benjelloun, "A multi-temporal multi-sensor circular fusion filter," *Information Fusion*, vol. 18, pp. 86–100, jul 2014.

[5] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive Bayesian filtering in circular state spaces," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 3, pp. 70–87, 2016.

[6] I. Marković, F. Chaumette, and I. Petrović, "Moving object detection, tracking and following using an omnidirectional camera on a mobile robot," in *International Conference on Robotics and Automation (ICRA)*, 2014.

[7] I. Marković, M. Bukal, J. Ćesić, and I. Petrović, "Multitarget tracking with the von Mises-Fisher filter and probabilistic data association," *Journal of Advances in Information Fusion*, 2016.

[8] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Unscented von Mises-Fisher filtering," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 463–467, 2016.

[9] J. Glover and L. P. Kaelbling, "Tracking 3-D rotations with the quaternion Bingham filter," Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology (MIT), Tech. Rep., 2013.

[10] G. Kurz, I. Gilitschenski, S. Julier, and U. D. Hanebeck, "Recursive estimation of orientation based on the Bingham distribution," in *International Conference on Information Fusion (FUSION)*, 2013, pp. 1–16.

[11] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Unscented Orientation Estimation Based on the Bingham Distribution," *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 172–177, 2016.

[12] J. T. Kent, I. Hussein, and M. K. Jah, "Directional Distributions in Tracking of Space Debris," in *International Confernce on Information Fusion (FUSION)*, 2016, p. 6.

[13] J. M. C. Clark, R. B. Vinter, and M. M. Yaqoob, "The shifted Rayleigh filter for bearings only tracking," in *International Conference on Information Fusion (FUSION)*, 2005, pp. 93–100.

[14] O. Schwander and F. Nielsen, "Learning mixtures by simplifying kernel density estimators," in *Matrix Information Geometry*, F. Nielsen and R. Bhatia, Eds. Springer Berlin Heidelberg, 2013, pp. 403–426.

[15] I. Marković, J. Ćesić, and I. Petrović, "Von Mises mixture PHD filter," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2229–2233, 2015.

[16] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association filter," *IEEE Journal of Oceanic Engineering*, vol. 8, no. 3, pp. 173–184, 1983.

[17] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Arthech House, 2007.

[18] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, oct 2003.

[19] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.

[20] S. Kullback, *Information Theory and Statistics*. New York: Dover Publications, 1997.

[21] C. Villani, *Topics in Optimal Transportation*. American Mathematical Society, 2003.

[22] A. Hyvärinen, "Estimation of non-normalized statistical models by score matching," *Journal of Machine Learning Research*, vol. 6, pp. 695–708, 2006.

[23] Y. Bar-Shalom and E. Tse, "Sonar tracking of multiple targets using joint probabilistic data association filter," *Automatica*, vol. 11, pp. 451–460, 1975.

[24] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3553–3567, jul 2007.

[25] F. Nielsen and V. Garcia, "Statistical exponential families: A digest with flash cards," *arXiv:0911.4863*, 2009.

[26] K. V. Mardia and P. E. Jupp, *Directional Statistics*. New York: Wiley, 1999.

[27] H. W. Kuhn, "The Hungarian method for the assignement problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.

[28] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.

[29] S. Sra, "A short note on parameter approximation for von Mises-Fisher distributions: and a fast implementation of Is(x)," *Computational Statistics*, vol. 27, no. 1, pp. 177–190, feb 2011.

[30] A. Tanabe, K. Fukumizu, S. Oba, T. Takenouchi, and S. Ishii, "Parameter estimation for von Mises–Fisher distributions," *Computational Statistics*, vol. 22, no. 1, pp. 145–157, 2007.

[31] I. Marković, A. Portello, P. Danès, and I. Petrović, "Active speaker localization with circular likelihoods and bootstrap filtering," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 2914–2920.