An EM Algorithm for Mixtures of Hyperspheres

Julien Lesouple
Université de Toulouse
ENAC
Toulouse, France
julien.lesouple@enac.fr

Philippe Burger
UE APC 802
INRAE-centre Occitanie-Toulouse
Castanet Tolosan, France
philippe.burger@inrae.fr

Jean-Yves Tourneret
Université de Toulouse
IRIT/ENSEEIHT/TéSA
Toulouse, France
jean-yves.tourneret@toulouse-inp.fr

***Abstract*—This paper studies a new expectation maximization (EM) algorithm to estimate the centers and radii of multiple hyperspheres. The proposed method introduces latent variables indicating to which hypersphere each vector from the dataset belongs to, in addition to random latent vectors having an a priori von Mises-Fisher distribution characterizing the location of each vector on the different hyperspheres. This statistical model allows a complete data likelihood to be derived, whose expected value conditioned on the observed data has a known distribution. This property leads to a simple and efficient EM algorithm whose performance is evaluated for the estimation of hypersphere mixtures yielding promising results.**

***Index Terms*—Mixture distribution, hypersphere fitting, expectation-maximization algorithm.**

I. Introduction

Hypersphere estimation can be found in many applications including object tracking [1]–[3], robotics [4]–[6] or image processing and pattern recognition [7]–[9]. This problem was recently investigated in [10] for a single hypersphere by introducing latent variables defined as affine transformations of random vectors distributed according to von Mises-Fisher distributions. The von Mises-Fisher distribution is a probability distribution defined on the unit hypersphere, which is parameterized by a mean vector and a concentration parameter. This distribution reduces to the uniform distribution on the hypersphere when the concentration parameter equals zero, or to more informative distributions for other values of this concentration parameter.

An expectation-maximization (EM) algorithm [11] was investigated in [10] using variables with a von Mises-Fisher prior distribution, allowing the parameters of a single hypersphere (radius and center), and possibly the hyperparameters of the von Mises-Fisher distribution to be estimated. This paper generalizes this algorithm to estimate the parameters of a mixture of hyperspheres. In order to build this generalization, each observation is assigned a latent variable indicating the hypersphere it belongs to. The parameters of the different hyperspheres, the hyperparameters of the different von Mises-Fisher priors and the mixture proportions are then estimated conditionally to these latent variables, via a new EM algorithm generalizing the strategy introduced in [10].

The authors would like to thank the Agrophen platform (https://doi.org/10.15454/1.5483266728434124E12) from Phenome-Emphasis supported by the project Phenome-ANR-11-INBS-0012 for useful feedback about this work.

The paper is organized as follows. Section II extends the maximum likelihood (ML) formulation of the hypersphere fitting problem to a mixture of hyperspheres. A specific attention is devoted to the estimation of the model hyperparameters that can be estimated jointly with the hypersphere centers and radii and the corresponding noise variances. Section III evaluates the performance of the resulting EM algorithm for fitting a mixture of hyperspheres, using multiple experiments conducted on synthetic data. Conclusions and future works are reported in Section IV.

II. A New EM Algorithm for Mixtures of Hyperspheres

A. Problem Formulation

Consider n noisy measurements $\boldsymbol{y}_i \in \mathbb{R}^d, i = 1, ..., n$ located around K hyperspheres with radii $r_k > 0$ and centers $\boldsymbol{c}_k \in \mathbb{R}^d, k = 1, ..., K$. We assume that the noise realizations corrupting the observations are mutually independent and distributed according to the same isotropic multivariate Gaussian distribution. If an observation belongs to the kth hypersphere, the hypersphere fitting problem can then be formulated as an ML estimation problem by introducing hidden vectors $\boldsymbol{x}_i \in \mathcal{S}^{d-1}, i = 1, ..., n$, where $\mathcal{S}^{d-1}$ is the centered unit hypersphere in $\mathbb{R}^d$ [10]. These hidden vectors are unknown unit vectors located on the hypersphere such that

$$\boldsymbol{y}_i = \boldsymbol{c}_k + r_k \boldsymbol{x}_i + \boldsymbol{e}_i, \tag{1}$$

where $\boldsymbol{e}_i \sim \mathcal{N}(\boldsymbol{0}_d, \sigma_k^2 \boldsymbol{I}_d)$ is the ith model error, $\boldsymbol{0}_d$ is the zero vector of $\mathbb{R}^d$, $\sigma_k^2 > 0$ is the unknown noise variance and $\boldsymbol{I}_d$ is the $d \times d$ identity matrix. The vectors $\boldsymbol{x}_i$ are assigned independent von Mises-Fisher distributions denoted as $\boldsymbol{x}_i \sim \text{vMF}_d(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \kappa_k)$ with density

$$f_d(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \kappa_k) = C_d(\kappa_k) \exp\left(\kappa_k \boldsymbol{\mu}_k^T \boldsymbol{x}_i\right) 1_{\mathcal{S}^{d-1}}(\boldsymbol{x}_i), \tag{2}$$

where $\boldsymbol{\mu}_k \in \mathbb{R}^d$ is the mean direction with $\|\boldsymbol{\mu}_k\|_2 = 1$, $\kappa_k \geq 0$ is the concentration parameter, $1_{\mathcal{S}^{d-1}}(.)$ is the indicator function of $\mathcal{S}^{d-1}$, and $C_d(\kappa_k)$ is a normalization constant (recalled in [10]). Note that this distribution reduces to the uniform distribution on the hypersphere for $\kappa_k = 0$ and is more informative for $\kappa_k > 0$. It is well-suited for LiDAR applications whose calibration can be achieved using sphere imaging [12]. Indeed, in this case, the LiDAR beam only hits a part of a sphere, resulting in points located in this area, concentrated around a mean direction with a certain deviation around this direction, which is well modelled by a von Mises-Fisher distribution.

If each hypersphere to which an observation belongs would be known, the hypersphere fitting problem would simply consists in estimating the hypersphere radii $\boldsymbol{r} = \{r_1, \ldots, r_K\}$ and centres $\boldsymbol{c} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_K\}$ (and possibly the noise variances $\boldsymbol{\sigma}^2 = \{\sigma_1^2, \ldots, \sigma_K^2\}$) from the measurements $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$, given that the latent vectors $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ are missing. However, the hyperspheres associated with the different observations are generally unknown, which complicates the estimation problem significantly.

B. Likelihood and complete likelihood

Denoting as $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_K\}$ the vector of hypersphere proportions and using (1), the conditional distribution of $\boldsymbol{y}_i$ given $\boldsymbol{x}_i$ and the unknown proportions π_k is a mixture of Gaussian distributions, i.e.,

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{\theta}) = \sum_{k=1}^{K} \frac{\pi_k}{(2\pi\sigma_k^2)^{d/2}} \exp\left\{-\frac{\|\boldsymbol{y}_i - \boldsymbol{c}_k - r_k\boldsymbol{x}_i\|_2^2}{2\sigma_k^2}\right\}, \tag{3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_K^T)^T$ contains the unknown parameters of the proposed statistical model and $\boldsymbol{\theta}_k = (\pi_k, r_k, \boldsymbol{c}_k^T, \sigma_k^2)^T$ contains the unknown parameters of a single hypersphere. Note that the hypersphere proportions sum to one, i.e.,

$$\sum_{k=1}^{K} \pi_k = 1. \tag{4}$$

We propose to introduce latent vectors $\boldsymbol{z}_i \in \mathbb{R}^K, i = 1, \ldots, n$ such that $z_{ik} = 1$ if $\boldsymbol{y}_i$ belongs to the k-th sphere and $z_{ik} = 0$ otherwise. The set of vectors $\boldsymbol{z}_i$ (vectors of $\mathbb{R}^K$ with a single non-zero element equal to one) is denoted as $\mathcal{O}^K$. Based on (3), the following conditional likelihood can be obtained:

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}) = \prod_{k=1}^{K} \left[\frac{1}{(2\pi\sigma_k^2)^{d/2}} e^{-\frac{\|\boldsymbol{y}_i - \boldsymbol{c}_k - r_k\boldsymbol{x}_i\|_2^2}{2\sigma_k^2}}\right]^{z_{ik}}. \tag{5}$$

The latent vector $\boldsymbol{z}_i = (z_{i1}, ..., z_{iK})^T$ is naturally assigned a categorical distribution, i.e.,

$$p(\boldsymbol{z}_i|\boldsymbol{\theta}) = \prod_{k=1}^{K} \pi_k^{z_{ik}} 1_{\mathcal{O}^K}(\boldsymbol{z}_i), \tag{6}$$

where $1_{\mathcal{O}^K}(.)$ is the indicator function of $\mathcal{O}^K$. The density of $\boldsymbol{x}_i$ for a given sphere is obtained from (2), leading to

$$p(\boldsymbol{x}_i|\boldsymbol{z}_i, \boldsymbol{\theta}) = \prod_{k=1}^{K} \left[C_d(\kappa_k) \exp\left(\kappa_k \boldsymbol{\mu}_k^T \boldsymbol{x}_i\right)\right]^{z_{ik}} 1_{\mathcal{S}^{d-1}}(\boldsymbol{x}_i). \tag{7}$$

The (marginal) likelihood of this model, which does not involve the latent vectors $(\boldsymbol{x}_i, \boldsymbol{z}_i)$, is

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{Y}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\theta}) = \prod_{i=1}^{n} \int_{\mathcal{S}^{d-1}} \sum_{\boldsymbol{z}_i \in \mathcal{O}^K} p(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x}_i. \tag{8}$$

As explained in [10] for hypersphere fitting, a closed-form expression for the ML estimator (MLE) of $\boldsymbol{\theta}$ cannot be derived. Instead, we propose to use the EM algorithm [11] to estimate the unknown vector $\boldsymbol{\theta}$. The so-called complete likelihood associated with the previous mixture model for K hyperspheres is

$$\mathcal{L}_c(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}), \tag{9}$$

where $\boldsymbol{Z} = \{\boldsymbol{z}_1, ..., \boldsymbol{z}_n\}$. Moreover, using the relation $p(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}) = p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})p(\boldsymbol{x}_i|\boldsymbol{z}_i, \boldsymbol{\theta})p(\boldsymbol{z}_i|\boldsymbol{\theta})$ and (5), (6) and (7), the following result is obtained

$$p(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}) = \prod_{k=1}^{K} \left[\frac{\pi_k C_d(\kappa_k)}{(2\pi\sigma_k^2)^{d/2}} \exp\left(-\frac{\|\boldsymbol{y}_i - \boldsymbol{c}_k - r_k\boldsymbol{x}_i\|_2^2 - 2\sigma_k^2\kappa_k\boldsymbol{\mu}_k^T\boldsymbol{x}_i}{2\sigma_k^2}\right)\right]^{z_{ik}},$$

where the indicator function has been omitted for brevity and it is assumed that $\boldsymbol{x}_i \in \mathcal{S}^{d-1}$ and $\boldsymbol{z}_i \in \mathcal{O}^K$.

C. Proposed EM Algorithm

The EM algorithm alternates between two steps referred to as expectation (E) and maximization (M) steps that are recalled below for iteration $(t+1)$ [11]:

1- The E-step consists of computing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$, the expected value of the complete data log-likelihood given the observed data and the current parameter estimate $\boldsymbol{\theta}^{(t)}$, defined as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbb{E}_{\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{Y}, \boldsymbol{\theta}^{(t)}} \left[\log \mathcal{L}_c(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z})\right]. \tag{10}$$

2- The M-step consists of estimating $\boldsymbol{\theta}^{(t+1)}$ by solving

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}). \quad \text{s.t.} \sum_{k=1}^{K} \pi_k = 1. \tag{11}$$

The complete log-likelihood can be computed using (9) and (10). Straightforward computations lead to

$$\begin{aligned}\log \mathcal{L}_c(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}) = &\sum_{k=1}^{K} \left[\log \pi_k - \frac{d}{2}\log\sigma_k^2 + \log C_d(\kappa_k)\right] \sum_{i=1}^{n} z_{ik} \\ &- \sum_{k=1}^{K} \frac{1}{2\sigma_k^2} \sum_{i=1}^{n} \left(\|\boldsymbol{y}_i - \boldsymbol{c}_k\|_2^2 + r_k^2\right) z_{ik} \\ &+ \sum_{k=1}^{K}\sum_{i=1}^{n} \kappa_{ik}\boldsymbol{\mu}_{ik}^T\boldsymbol{x}_i z_{ik} + C,\end{aligned} \tag{12}$$

where C is an additive term independent of $\boldsymbol{\theta}$ and

$$\kappa_{ik} = \frac{\|r_k(\boldsymbol{y}_i - \boldsymbol{c}_k) + \sigma_k^2\kappa_k\boldsymbol{\mu}_k\|_2}{\sigma_k^2}, \tag{13}$$

$$\boldsymbol{\mu}_{ik} = \frac{r_k(\boldsymbol{y}_i - \boldsymbol{c}_k) + \sigma_k^2\kappa_k\boldsymbol{\mu}_k}{\|r(\boldsymbol{y}_i - \boldsymbol{c}_k) + \sigma_k^2\kappa_k\boldsymbol{\mu}_k\|_2}. \tag{14}$$

Note that (13) and (14) ensure that $\boldsymbol{\mu}_{ik}$ has a unit norm, as required for the von Mises-Fisher distributions. The distribution of $\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{Y}, \boldsymbol{\theta}^{(t)}$ can be determined as

$$p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{Y}, \boldsymbol{\theta}^{(t)}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{y}_i, \boldsymbol{\theta}^{(t)}), \quad (15)$$

where each term can be determined using (5), (6) and (7), i.e.,

$$p(\boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{y}_i, \boldsymbol{\theta}) \propto p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}) p(\boldsymbol{x}_i|\boldsymbol{z}_i, \boldsymbol{\theta}) p(\boldsymbol{z}_i|\boldsymbol{\theta}), \quad (16)$$

$$\propto \prod_{k=1}^{K} \left[\tilde{\gamma}_{ik} f_d(\boldsymbol{x}_i; \boldsymbol{\mu}_{ik}, \kappa_{ik})\right]^{z_{ik}}, \quad (17)$$

where $\propto$ means "proportional to" and

$$\tilde{\gamma}_{ik} = \frac{\pi_k}{(\sigma_k^2)^{d/2}} \frac{C_d(\kappa_k)}{C_d(\kappa_{ik})} \exp\left(-\frac{\|\boldsymbol{y}_i - \boldsymbol{c}_k\|_2^2 + r_k^2}{2\sigma_k^2}\right). \quad (18)$$

Defining

$$\gamma_{ik} = \frac{\tilde{\gamma}_{ik}}{\sum_{k=1}^{K} \tilde{\gamma}_{ik}}, \quad k = 1, ..., K \quad (19)$$

the following results are obtained:

$$\mathbb{E}_{\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{\theta}^{(t)}}[z_{ik}] = \gamma_{ik}^{(t)}, \quad (20)$$

$$\mathbb{E}_{\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{\theta}^{(t)}}[\boldsymbol{x}_i z_{ik}] = \gamma_{ik}^{(t)} A_d(\kappa_{ik}^{(t)}) \boldsymbol{\mu}_{ik}^{(t)}, \quad (21)$$

where $\kappa_{ik}^{(t)}, \boldsymbol{\mu}_{ik}^{(t)}$ and $\gamma_{ik}^{(t)}$ are computed from (13), (14), (18), and (19) using the current values of $r_k, \boldsymbol{c}_k, \sigma_k^2, \pi_k, \kappa_k$ and $\boldsymbol{\mu}_k$. Note that (21) has been obtained using the mean of a von Mises-Fisher distribution, where

$$A_d(\kappa) = \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}, \quad (22)$$

and where $I_\nu(.)$ denotes the modified Bessel function of the first kind of parameter ν [13, Chap. 10.25].

After substituting these expectations into (12), the constrained maximization of the function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\theta}$ leads to the following updates for $r_k, \boldsymbol{c}_k, \sigma_k^2$, and π_k

$$r_k^{(t+1)} = \frac{1}{1 - \overline{\boldsymbol{u}}_{kt}^T \overline{\boldsymbol{u}}_{kt}} (\overline{\boldsymbol{u}^T \boldsymbol{y}}_{kt} - \overline{\boldsymbol{u}}_{kt}^T \overline{\boldsymbol{y}}_{kt}), \quad (23)$$

$$\boldsymbol{c}_k^{(t+1)} = \overline{\boldsymbol{y}}_{kt} - r_k^{(t+1)} \overline{\boldsymbol{u}}_{kt}, \quad (24)$$

$$d\sigma_k^{2(t+1)} = \overline{\|\boldsymbol{y}\|_2^2}_{kt} + \|c_k^{(t+1)}\|_2^2 + {r_k^{(t+1)}}^2 - 2\left\{ {\boldsymbol{c}_k^{(t+1)}}^T \overline{\boldsymbol{y}}_{kt} + r_k^{(t+1)} \left[\overline{\boldsymbol{u}^T \boldsymbol{y}}_{kt} - \overline{\boldsymbol{u}}_{kt}^T \boldsymbol{c}_k^{(t+1)}\right] \right\}, \quad (25)$$

$$\pi_k^{(t+1)} = \frac{\gamma_k^{(t)}}{n}, \quad (26)$$

with

$$\gamma_k^{(t)} = \sum_{i=1}^{n} \gamma_{ik}^{(t)}, \quad \overline{\boldsymbol{u}}_{kt} = \frac{1}{\gamma_k^{(t)}} \sum_{i=1}^{n} \gamma_{ik}^{(t)} \boldsymbol{\alpha}_{ik}^{(t)}, \quad (27)$$

$$\boldsymbol{\alpha}_{ik}^{(t)} = A_d(\kappa_{ik}^{(t)}) \boldsymbol{\mu}_{ik}^{(t)}, \quad \overline{\boldsymbol{y}}_{kt} = \frac{1}{\gamma_k^{(t)}} \sum_{i=1}^{n} \gamma_{ik}^{(t)} \boldsymbol{y}_i, \quad (28)$$

$$\overline{\boldsymbol{u}^T \boldsymbol{y}}_{kt} = \frac{1}{\gamma_{ik}^{(t)}} \sum_{i=1}^{n} \gamma_{ik}^{(t)} \boldsymbol{y}_i^T \boldsymbol{\alpha}_{ik}^{(t)}, \quad \overline{\|\boldsymbol{y}\|_2^2}_{kt} = \frac{1}{\gamma_{ik}^{(t)}} \sum_{i=1}^{n} \gamma_{ik}^{(t)} \|\boldsymbol{y}_i\|_2^2. \quad (29)$$

Note that the quantities with bars and subscript kt are the weighted sums of these quantities over the weights for the kth hypersphere at iteration t, and that $\boldsymbol{\alpha}_{ik}^{(t)}$ is the mean of a von Mises-Fisher distributions with parameters $\kappa_{ik}^{(t)}$ and $\boldsymbol{\mu}_{ik}^{(t)}$.

Finally, note that $\gamma_{ik}^{(t)}$ is the a posteriori probability that observation $\#i$ belongs to the kth hypersphere, which is a useful piece of information. Indeed, it can be used to assign each data to one of the hyperspheres according to the maximum a posteriori (MAP) rule: $\boldsymbol{y}_i$ belongs to the kth hypersphere if and only if $k = \arg\max_j \gamma_{ij}^{(t)}$.

D. Hyperparameter Estimation

The method presented before assumes that the hyperparameters κ_k and $\boldsymbol{\mu}_k$ of the hidden variables $\boldsymbol{x}_i$ are known. When these parameters are unknown, they can be estimated using different methods presented in [11], such as empirical or hierarchical Bayesian inference. In this paper, we propose to include these parameters in the vector $\boldsymbol{\theta}$ (which explains why some terms depend on κ_k and $\boldsymbol{\mu}_k$ in (12)). This strategy results in additional updates for their estimates in the M-step using their MLE given the current estimation of the hidden variables, i.e.,

$$\kappa_k^{(t+1)} = A_d^{-1}(\|\overline{\boldsymbol{u}}_{kt}\|_2), \quad \boldsymbol{\mu}_k^{(t+1)} = \frac{\overline{\boldsymbol{u}}_{kt}}{\|\overline{\boldsymbol{u}}_{kt}\|_2}. \quad (30)$$

Note that these equations have been obtained by using the expressions of the MLEs of the parameters of a von Mises-Fisher distribution [14, Chap. 10.3.1]. Note also that the inverse function A_d^{-1} has no closed-form expression but can be computed using a two-steps iterative method [15].

III. SIMULATION RESULTS

In all the experiments presented in this paper, the EM algorithm was initialized as follows: the dataset was partitioned into K clusters (corresponding to the number of hyperspheres) using the k-means algorithm [16]. Then, the iterative maximum likelihood (IML) algorithm of [17] was run on each cluster to estimate the hypersphere radii and centers denoted as $\hat{\boldsymbol{c}}_k$ and $\hat{r}_k$ for $k = 1, ..., K$. The other parameters (noise variances, mixture proportions and hyperparameters of the von Mises-Fisher distributions) were finally initialized using their ML estimators [18, Chap. 7], i.e., for hypersphere k, $\hat{\sigma}_k^2 = \frac{1}{d}\left(\|\bar{\boldsymbol{y}}_k - \hat{\boldsymbol{c}}_k\|_2^2 - \hat{r}_k^2\right), \hat{\pi}_k = \frac{n_k}{n}, \hat{\mu}_k = \frac{\bar{\boldsymbol{y}}_k}{\|\bar{\boldsymbol{y}}_k\|_2}$ and $\hat{\kappa}_k = A_d^{-1}(\|\bar{\boldsymbol{y}}_k\|_2)$, where $\bar{\boldsymbol{y}}_k$ is the mean of the n_k observations assigned to hypersphere k, and A_d^{-1} has been defined previously. Note that in all experiments, the hyperparameters $\kappa_k, \boldsymbol{\mu}_k$ are unknown and are therefore estimated jointly with the other parameters $r_k, \boldsymbol{c}_k, \sigma_k^2$ and π_k.

A. Simulated 2D LiDAR

The first experiments were conducted with $d = 2$, $K = 8$ and $r_k = 3, \kappa_k = 5, \pi_k = \frac{1}{K}, \sigma_k^2 = 0.1$ for $k = 1, \ldots, K$. The geometry configuration is a simplified version of the LiDAR calibration problem. A LiDAR is assumed to be in the center of the scene. Thus, the values of the mean vectors $\boldsymbol{\mu}_k$ are the unitary vectors pointing from each circle center to the LiDAR.

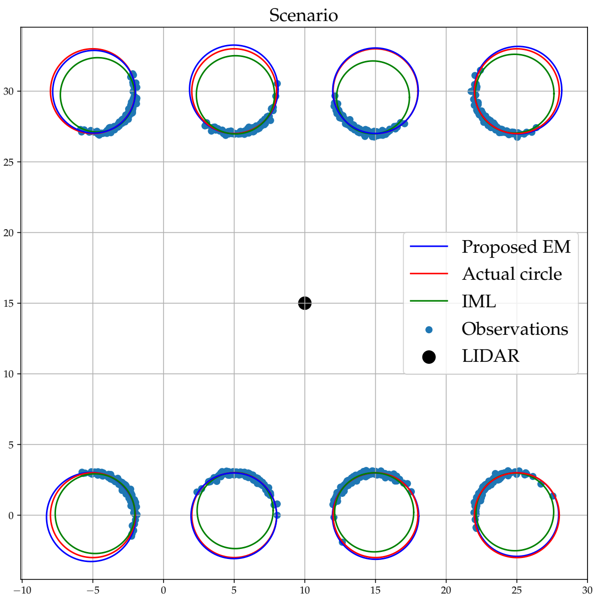

Fig. 1: Observations, ground truth and estimated circles.

After defining the problem geometry, $n = 1000$ observations were generated according to (3). The LiDAR configuration, the observations and the estimated circles are displayed in Fig. 1. The proposed algorithm provides better results than the IML reference, due to the use of the von Mises-Fisher prior distributions. Examples of mean square errors (MSEs) of $\boldsymbol{\theta}_k$ for the different circles are displayed in Fig. 2 as a function of the iteration number, illustrating the algorithm convergence. Note that the final values of these MSEs depend on the circle positions with respect to the LiDAR and the values of the model hyperparameters. Finally, as the most important parameters for LiDAR calibration are the circle centers, the MSEs of $\boldsymbol{c}$ are displayed in Fig. 3 versus the number of iterations. Having a global MSE for $\boldsymbol{c}$ lower than -10 dB is sufficiently accurate for LiDAR calibration, which is an interesting result.

B. Monte-Carlo simulations

To better understand the behaviour of the proposed method as a function of the noise level, $N_{\text{MC}} = 500$ Monte-Carlo simulations were performed for different values of the noise variance σ^2. The resulting averaged MSEs of $\boldsymbol{\theta}$ are shown in Fig. 4 for both the proposed EM approach and the benchmark IML method. Figs. 5 and 6 display the average MSEs of the vector of circle centres $\boldsymbol{c}$ and the MSEs of the hyperparameter vector $\left(\kappa_k, \boldsymbol{\mu}_{\boldsymbol{k}}^T\right)^T$. All these results show that the proposed EM algorithm has a good performance, when compared to the IML method.

IV. Conclusion

This paper studied a new EM algorithm for estimating the centers and radii of multiple hyperspheres for LiDAR calibration. A mixture of von Mises-Fisher distributions was assigned to hidden variables located on the unit hypersphere

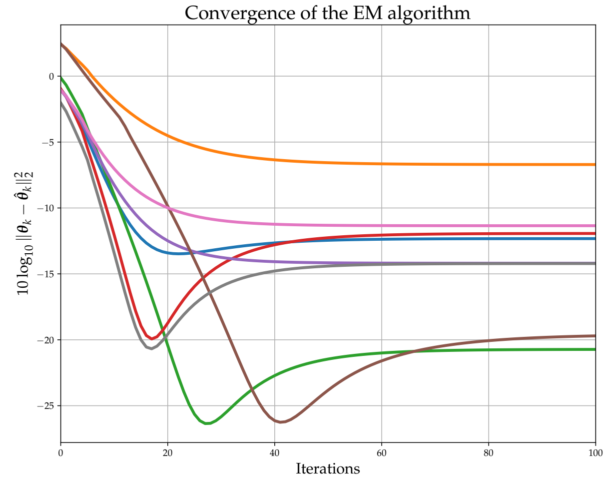

Fig. 2: MSEs of $\boldsymbol{\theta}_k$ versus the number of iterations of the EM algorithm (initial values obtained using the k-means algorithm and IML initialization).

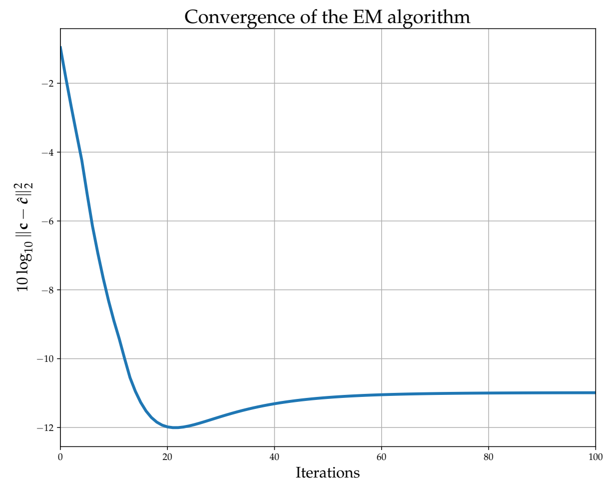

Fig. 3: MSE of $\boldsymbol{c}$ versus the number of iterations for the proposed EM algorithm.

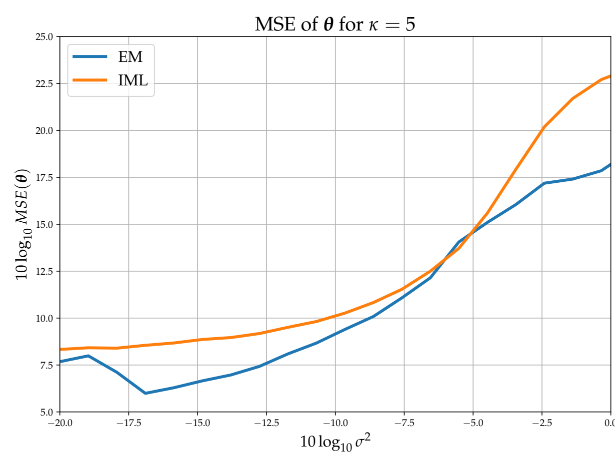

Fig. 4: Averaged MSE of $\boldsymbol{\theta}$ versus σ^2 - 500 Monte-Carlo iterations.

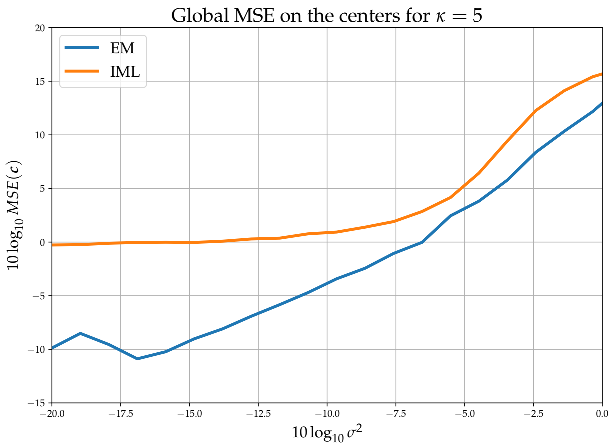

Fig. 5: MSE of $\boldsymbol{c}$ versus the noise variance σ^2 - 500 Monte-Carlo iterations.

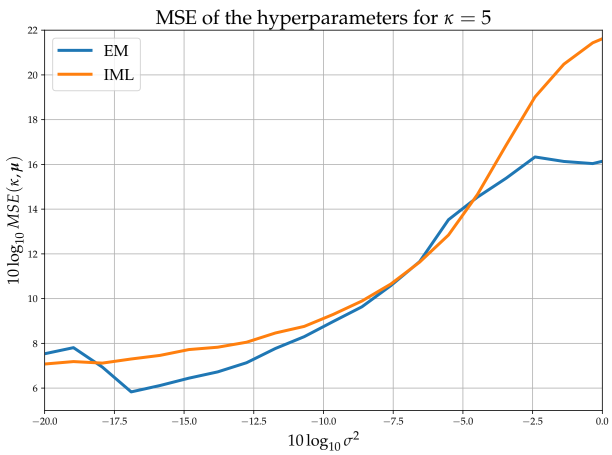

Fig. 6: MSE of $\left(\kappa, \boldsymbol{\mu}^T\right)^T$ versus the noise variance σ^2 - 500 Monte-Carlo iterations.

indicating that the observations are close to some parts of the hyperspheres, depending on the LiDAR location with respect to these hyperspheres. Categorical distributions were also assigned to hypershere indicators indicating to which hypersphere the observations correspond. The proposed algorithm only requires two parameters to be adjusted: the stopping criterion for the EM algorithm and the number of hyperspheres. Note that the number of hyperspheres will be known for practical applications related to LiDAR calibration. However, it could be estimated using criteria such as the Bayesian Information Criterion (BIC) [19], or using a split and merge approach as in [20]. Finally, it is interesting to note that the hyperparameters of the von Mises-Fisher distributions are also estimated by the algorithm. The proposed algorithm was evaluated for circle fitting in a realistic scenario. The results obtained on simulated data are very encouraging showing the competitiveness of the proposed approach with respect to the IML reference method. Future work include the generalization of the proposed work to the robust estimation of hypersphere mixtures, when outliers are contaminating the LiDAR measurements. It would be also interesting to analyse the sensitivity of the algorithm to its initialization, to the number of iterations for IML, to hyperparameter estimation... Finally, the proposed algorithm could be generalized to other noise distributions, e.g., to mitigate the impact of outliers or to take into account the geometry of the problem.

References

[1] M. Baum and U. D. Hanebeck, “Random Hypersurface Models for Extended Object Tracking,” in *Proc. IEEE Int. Symp. Sign. Process. Inf. Tech. (ISSPIT 2009)*, Ajman, United Arab Emirates, Dec. 2009, pp. 178–183.

[2] N. Wahlström and E. Özkan, “Extended Target Tracking Using Gaussian Processes,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 4165–4178, May 2015.

[3] D. Epstein and D. Feldman, “Sphere Fitting with Applications to Machine Tracking,” *Algorithms*, vol. 13, no. 8, p. 177, July 2020.

[4] F. Sandoval, “An Algorithm for Fitting 2-D Data on the Circle: Applications to Mobile Robotics,” *IEEE Signal Process. Lett.*, vol. 15, pp. 127–130, Jan. 2008.

[5] F. Bonin-Font, A. Ortiz, and G. Oliver, “Visual Navigation for Mobile Robots: A Survey,” *J. Intell. Robot. Syst.*, vol. 53, p. 263–296, Nov. 2008.

[6] D. Epstein and D. Feldman, “Quadcopter Tracks Quadcopter via Real Time Shape Fitting,” *IEEE Robot. Autom. Lett.*, vol. 3, p. 544–550, Jan. 2018.

[7] D. Lin and C. Yang, “Real-Time Eye Detection Using Face-Circle Fitting and Dark-Pixel Filtering,” in *Proc. IEEE Int. Conf. on Multimedia ans Expo (ICME’2004)*, Taipei, Taiwan, June 2004, pp. 1167–1170.

[8] L. Pan, W. Chu, J. M. Saragih, F. De la Torre, and M. Xie, “Fast and Robust Circular Object Detection With Probabilistic Pairwise Voting,” *IEEE Sign. Process. Let.*, vol. 18, no. 11, pp. 639–642, Sept. 2011.

[9] A. Geiger, P. Lenz, and R. Urtasun, “Are We Ready for Autonomous Driving? The Kitti Vision Benchmark Suite,” in *Proc. IEEE Int. Conf. Conf. Comput. Vis. Pattern Recognit. (CVPR’2012)*, Providence, RI, USA, Jun. 2012, pp. 3354–3361.

[10] J. Lesouple, B. Pilastre, Y. Altmann, and J.-Y. Tourneret, “Hypersphere Fitting from Noisy Data Using an EM Algorithm,” *IEEE Signal Process. Lett.*, vol. 28, pp. 314–318, 1 2021.

[11] A. Dempster, N. Laird, and D. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *J. R. Stat. Soc. Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[12] M. Ruan and D. Huber, “Calibration of 3D Sensors Using a Spherical Target,” in *Proc. 2nd International Conference on 3D Vision*, Tokyo, Japan, Dec. 2014, pp. 187–193.

[13] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

[14] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, Inc., 1999.

[15] S. Sra, “A Short Note on Parameter Approximation for von Mises-Fisher Distributions: and a Fast Implementation of $I_s(x)$,” *Computational Statistics*, vol. 27, no. 1, pp. 177–190, March 2012.

[16] J. B. MacQueen, “Some Methods for classification and Analysis of Multivariate Observations,” in *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, Berkeley, California, 1965, pp. 281–297.

[17] W. Li, J. Zhong, T. A. Gulliver, B. Rong, R. Hu, and Y. Qian, “Fitting Noisy Data to a Circle: A Simple Iterative Maximum Likelihood Approach,” in *Proc. IEEE Int. Conf. on Communications*, Kyoto, Japan, July 2011, pp. 1 – 5.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. USA: Prentice-Hall, Inc., 1993.

[19] G. Schwarz, “Estimating the Dimension of a Model,” *Ann. of Statist.*, vol. 6, no. 2, pp. 461 – 464, March 1978.

[20] Z. Zhang, C. Chen, J. Sun, and K. Luk Chan, “EM Algorithms for Gaussian Mixtures with Split-and-Merge Operation,” *Pattern Recognition*, vol. 36, no. 9, pp. 1973–1983, Sept. 2003.