Robust Hypersphere Fitting from Noisy Data Using an EM Algorithm

Robust Hypersphere Fitting from Noisy Data Using an EM Algorithm

Julien Lesouple
TéSA
Toulouse, France
julien.lesouple@tesa.prd.fr

Barbara Pilastre
TéSA
Toulouse, France
barbara.pilastre@tesa.prd.fr

Yoann Altmann
Heriot-Watt University
Edinburgh, Scotland
y.altmann@hw.ac.uk

Jean-Yves Tourneret
University of Toulouse
INP-ENSEEIHT/IRIT/TéSA
jean-yves.tourneret@toulouse-inp.fr

***Abstract*—This article studies a robust expectation maximization (EM) algorithm to solve the problem of hypersphere fitting. This algorithm relies on the introduction of random latent vectors having independent von Mises-Fisher distributions defined on the hypersphere and random latent vectors indicating the presence of potential outliers. This model leads to an inference problem that can be solved with a simple EM algorithm. The performance of the resulting robust hypersphere fitting algorithm is evaluated for circle and sphere fitting with promising results in terms of both estimation performance and computation time.**

***Index Terms*—Robust estimation, hypersphere fitting, expectation-maximization algorithm.**

I. Introduction

Fitting a circle, a sphere or more generally a hypersphere to a noisy point cloud is a recurrent problem 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] 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 hypersphere and parameterized by a mean vector and a concentration. This distribution reduces to the uniform distribution on the hypersphere when the concentration is zero, or to more informative distributions for other values of the concentration parameter. An expectation-maximization (EM) algorithm [11] was investigated in [10] using latent variables with von Mises-Fisher prior distributions, allowing the hypersphere parameters (radius and center), and possibly the von Mises-Fisher distribution hyperparameters to be estimated.

This paper studies a robust EM algorithm for hypersphere fitting allowing the hypersphere parameters to be estimated while being robust to the presence of potential outliers. The main contribution with respect to [10] is to introduce a mixture model for this estimation problem with one component corresponding to the inliers (located close to the hypersphere) and a second component allowing the presence of outliers. The proposed method robustifies the strategy introduced in [10], while allowing the model hyperparameters to be estimated. The paper is organized as follows. Section II recalls the maximum likelihood (ML) formulation of the hypersphere fitting problem and extends this formulation to the presence of outliers. A specific attention is devoted to the estimation of the model hyperparameters that can be estimated jointly with the hypersphere center and radius and the noise variance. Section III evaluates the performance of the resulting robust EM algorithm for hypersphere fitting through various experiments. Conclusions and future works are reported in Section IV.

Part of this work was supported by the Royal Academy of Engineering under the Research Fellowship scheme RF201617/16/31

II. A Robust EM Algorithm for Hypersphere Fitting

A. Problem Formulation

Consider n noisy measurements $\boldsymbol{y}_i \in \mathbb{R}^d, i = 1, ..., n$ located around a hypersphere with radius r and center $\boldsymbol{c} \in \mathbb{R}^d$. We assume that the noise realizations corrupting the observations are mutually independent and distributed according to the same isotropic multivariate Gaussian distribution. The hypersphere fitting problem can then be formulated as an ML estimation problem by introducing latent 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 latent vectors are the unknown unit vectors located on the hypersphere such that

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

where $\boldsymbol{e}_i \sim \mathcal{N}(\boldsymbol{0}_d, \sigma^2 \boldsymbol{I}_d)$ is the ith model error, $\boldsymbol{0}_d$ is the zero vector of $\mathbb{R}^d$, σ^2 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 $\boldsymbol{x}_i \sim \text{vMF}_d(\boldsymbol{x}_i; \boldsymbol{\mu}, \kappa)$ with density

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

where $\boldsymbol{\mu} \in \mathbb{R}^d$ is the mean direction with $\|\boldsymbol{\mu}\|_2 = 1$, $\kappa \geq 0$ is the concentration parameter, $1_{\mathcal{S}^{d-1}}(.)$ is the indicator function of $\mathcal{S}^{d-1}$, and $C_d(\kappa)$ is a normalization constant (recalled in [10]). Note that this distribution reduces to the uniform distribution on the hypersphere for $\kappa = 0$. 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 corresponds to a von Mises-Fisher distribution.

The hypersphere fitting problem consists of estimating the radius r and center $\boldsymbol{c}$ of the hypersphere (and possibly the

noise variance σ^2) from the measurements $\boldsymbol{Y} = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_n\}$, given that the latent vectors $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ are also unknown. To allow the presence of outliers in the observations, we borrow the idea of the so-called Maximum Likelihood Estimation SAmple Consensus (MLESAC) [13], which is a generalization of the RANdom SAmple Consensus (RANSAC), by introducing an outlier uniform distribution defined on a volume $A \subset \mathbb{R}^d$. This uniform distribution is defined as

$$p(\boldsymbol{y}_i) = \frac{1}{a} 1_A(\boldsymbol{y}_i), \tag{3}$$

where a is the volume of A, and 1_A is the indicator function on the set A. When there is no prior information about the location of the outliers, A can be chosen as the whole observation domain, i.e., the smallest hypercube containing all the observations, whose volume is a. In this case, the indicator can be omitted without loss of generality.

B. Likelihood and complete likelihood

The conditional distribution of $\boldsymbol{y}_i$ given $\boldsymbol{x}_i$ is a mixture between the uniform distribution (3) and the Gaussian distribution (1), i.e.,

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\gamma}{a} + \frac{1-\gamma}{(2\pi\sigma^2)^{d/2}} \exp\left\{-\frac{\|\boldsymbol{y}_i - \boldsymbol{c} - r\boldsymbol{x}_i\|_2^2}{2\sigma^2}\right\}, \tag{4}$$

where γ is the unknown proportion of outliers in the observations, and $\boldsymbol{\theta} = (r, \boldsymbol{c}^T, \sigma^2, \gamma)^T$ contains the unknown parameters of the proposed statistical model. We propose to include binary latent variables $z_i, i = 1, \ldots, n$ such that $z_i = 1$ if $\boldsymbol{y}_i$ is an outlier and $z_i = 0$ otherwise. The likelihood (4) can then be rewritten as

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i, z_i, \boldsymbol{\theta}) = \frac{1}{a^{z_i}} \left[\frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{\|\boldsymbol{y}_i - \boldsymbol{c} - r\boldsymbol{x}_i\|_2^2}{2\sigma^2}}\right]^{1-z_i}. \tag{5}$$

The latent variable z_i is naturally assigned a Bernoulli distribution with parameter γ, i.e.,

$$p(z_i|\boldsymbol{\theta}) = \gamma^{z_i}(1-\gamma)^{1-z_i}. \tag{6}$$

We also introduce the following notation

$$p(\boldsymbol{x}_i|z_i, \boldsymbol{\theta}) = p_1(\boldsymbol{x}_i)^{z_i} \left[C_d(\kappa) \exp\left(\kappa \boldsymbol{\mu}^T \boldsymbol{x}_i\right)\right]^{1-z_i}, \tag{7}$$

where $p_1(\boldsymbol{x}_i) = p(\boldsymbol{x}_i|z_i = 1, \boldsymbol{\theta})$ is the distribution assigned to the latent variable $\boldsymbol{x}_i$ when it corresponds to an outlier, and $p(\boldsymbol{x}_i|z_i = 0, \boldsymbol{\theta})$ is a von Mises-Fisher distribution with parameters κ and $\boldsymbol{\mu}$, which corresponds to the inlier distribution. We assume that $p_1(\boldsymbol{x}_i)$ does not depend on $\boldsymbol{\theta}$, which makes sense as outliers do not provide information about the hypersphere. In this case $p_1(\boldsymbol{x}_i)$ does not appear in the derivation of the algorithm.

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

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{Y}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\theta}) = \prod_{i=1}^{n} \int_{\mathcal{S}^{d-1}} \sum_{z_i \in \{0,1\}} p(\boldsymbol{y}_i, \boldsymbol{x}_i, z_i|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x}_i. \tag{8}$$

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 is

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

where $\boldsymbol{z} = \{z_1, ..., z_n\}$. Moreover, using (5), (6) and (7), the following result is obtained

$$\begin{aligned} p(\boldsymbol{y}_i, \boldsymbol{x}_i, z_i|\boldsymbol{\theta}) &= p(\boldsymbol{y}_i|\boldsymbol{x}_i, z_i, \boldsymbol{\theta}) p(\boldsymbol{x}_i|z_i, \boldsymbol{\theta}) p(z_i|\boldsymbol{\theta}), \\ &= \left[\frac{\gamma}{a} p_1(\boldsymbol{x}_i)\right]^{z_i} \left[\frac{1-\gamma}{(2\pi\sigma^2)^{d/2}} C_d(\kappa)\right]^{1-z_i} \\ &\times \left[\exp\left(-\frac{\|\boldsymbol{y}_i - \boldsymbol{c} - r\boldsymbol{x}_i\|_2^2}{2\sigma^2} + \kappa \boldsymbol{\mu}^T \boldsymbol{x}_i\right)\right]^{1-z_i}. \end{aligned} \tag{10}$$

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

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbb{E}_{\boldsymbol{X}, \boldsymbol{z}|\boldsymbol{Y}, \boldsymbol{\theta}^{(t)}} \left[\log \mathcal{L}_c(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{z})\right]. \tag{11}$$

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

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}). \tag{12}$$

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

$$\begin{aligned} \log \mathcal{L}_c(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{z}) &= K + \log(\gamma) \sum_{i=1}^{n} z_i \\ &+ \left[\log(1-\gamma) - \frac{d}{2}\log\left(\sigma^2\right) + \log C_d(\kappa)\right] \sum_{i=1}^{n} (1 - z_i) \\ &- \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\|\boldsymbol{y}_i - \boldsymbol{c}\|_2^2 + r^2\right)(1 - z_i) \\ &+ \sum_{i=1}^{n} \kappa_i \boldsymbol{\mu}_i^T \boldsymbol{x}_i (1 - z_i), \end{aligned} \tag{13}$$

where K is a term independent of $\boldsymbol{\theta}$ and

$$\kappa_i = \frac{\|r(\boldsymbol{y}_i - \boldsymbol{c}) + \sigma^2 \kappa \boldsymbol{\mu}\|_2}{\sigma^2}, \tag{14}$$

$$\boldsymbol{\mu}_i = \frac{r(\boldsymbol{y}_i - \boldsymbol{c}) + \sigma^2 \kappa \boldsymbol{\mu}}{\|r(\boldsymbol{y}_i - \boldsymbol{c}) + \sigma^2 \kappa \boldsymbol{\mu}\|_2}. \tag{15}$$

The distribution of $\boldsymbol{X}, \boldsymbol{z}|\boldsymbol{Y}, \boldsymbol{\theta}^{(t)}$ can be determined as

$$p(\boldsymbol{X}, \boldsymbol{z}|\boldsymbol{Y}, \boldsymbol{\theta}^{(t)}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i, z_i|\boldsymbol{y}_i, \boldsymbol{\theta}^{(t)}), \tag{16}$$

with

$$p(\boldsymbol{x}_i, z_i|\boldsymbol{y}_i, \boldsymbol{\theta}) \propto p(\boldsymbol{y}_i|\boldsymbol{x}_i, z_i, \boldsymbol{\theta})p(\boldsymbol{x}_i, z_i|\boldsymbol{\theta}), \quad (17)$$

$$\propto [\tilde{\pi}_{i,1}p_1(\boldsymbol{x}_i)]^{z_i} [\tilde{\pi}_{i,2}f_d(\boldsymbol{x}_i; \boldsymbol{\mu}_i, \kappa_i)]^{1\text{-}z_i}, \quad (18)$$

where $\propto$ means "proportional to" and

$$\tilde{\pi}_{i,1} = \frac{\gamma}{a}, \quad (19)$$

$$\tilde{\pi}_{i,2} = \frac{1-\gamma}{(2\pi\sigma^2)^{d/2}} \frac{C_d(\kappa)}{C_d(\kappa_i)} \exp\left(-\frac{\|\boldsymbol{y}_i - \boldsymbol{c}\|_2^2 + r^2}{2\sigma^2}\right). \quad (20)$$

Defining

$$\pi_{i,1} = \frac{\tilde{\pi}_{i,1}}{\tilde{\pi}_{i,1} + \tilde{\pi}_{i,2}}, \quad \pi_{i,2} = 1 - \pi_{i,1}, \quad (21)$$

the following results are obtained:

$$\mathbb{E}_{\boldsymbol{X},\boldsymbol{z}|\boldsymbol{Y},\boldsymbol{\theta}^{(t)}}[z_i] = \pi_{i,1}^{(t)}, \quad (22)$$

$$\mathbb{E}_{\boldsymbol{X},\boldsymbol{z}|\boldsymbol{Y},\boldsymbol{\theta}^{(t)}}[1 - z_i] = \pi_{i,2}^{(t)}, \quad (23)$$

$$\mathbb{E}_{\boldsymbol{X},\boldsymbol{z}|\boldsymbol{Y},\boldsymbol{\theta}^{(t)}}[\boldsymbol{x}_i(1 - z_i)] = \pi_{i,2}^{(t)} A_d(\kappa_i^{(t)})\boldsymbol{\mu}_i^{(t)}, \quad (24)$$

where $\kappa_i^{(t)}, \boldsymbol{\mu}_i^{(t)}, \pi_{i,1}^{(t)}$ and $\pi_{i,2}^{(t)}$ are computed from (14), (15), (19), (20), and (21) using the current values of $r, \boldsymbol{c}, \sigma^2, \gamma, \kappa$ and $\boldsymbol{\mu}$. Note that (24) has been obtained using the mean of a von Mises-Fisher distribution, where

$$A_d(\kappa) = \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}, \quad (25)$$

where $I_\nu(.)$ denotes the modified Bessel function of first kind of parameter ν [14, Chap. 10.25].

After substituting these expectations into (13), the maximization of the function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\theta}$ leads to the following updates for r, $\boldsymbol{c}, \sigma^2$, and γ

$$r^{(t+1)} = \frac{1}{1 - \overline{\boldsymbol{u}}_t^T \overline{\boldsymbol{u}}_t}(\overline{\boldsymbol{u}^T \boldsymbol{y}}_t - \overline{\boldsymbol{u}}_t^T \overline{\boldsymbol{y}}_t), \quad (26)$$

$$\boldsymbol{c}^{(t+1)} = \overline{\boldsymbol{z}}_t - r^{(t+1)}\overline{\boldsymbol{u}}_t, \quad (27)$$

$$d\sigma^{2(t+1)} = \overline{\|\boldsymbol{y}\|_2^2}_t + \|c^{(t+1)}\|_2^2 + r^{(t+1)^2} - 2\left\{\boldsymbol{c}^{(t+1)^T}\overline{\boldsymbol{y}}_t + r^{(t+1)}\left[\overline{\boldsymbol{u}^T\boldsymbol{y}}_t - \overline{\boldsymbol{u}}_t^T \boldsymbol{c}^{(t+1)}\right]\right\}, \quad (28)$$

$$\gamma^{(t+1)} = 1 - \frac{\pi_2^{(t)}}{n}, \quad (29)$$

with

$$\pi_2^{(t)} = \sum_{i=1}^n \pi_{i,2}^{(t)}, \quad \overline{\boldsymbol{u}}_t = \frac{1}{\pi_2^{(t)}} \sum_{i=1}^n \pi_{i,2}^{(t)} \boldsymbol{\alpha}_i^{(t)}, \quad (30)$$

$$\boldsymbol{\alpha}_i^{(t)} = A_d(\kappa_i^{(t)})\boldsymbol{\mu}_i^{(t)}, \quad \overline{\boldsymbol{y}}_t = \frac{1}{\pi_2^{(t)}} \sum_{i=1}^n \pi_{i,2}^{(t)} \boldsymbol{y}_i, \quad (31)$$

$$\overline{\boldsymbol{u}^T\boldsymbol{y}}_t = \frac{1}{\pi_2^{(t)}} \sum_{i=1}^n \pi_{i,2}^{(t)} \boldsymbol{y}_i^T \boldsymbol{\alpha}_i^{(t)}, \quad \overline{\|\boldsymbol{y}\|_2^2}_t = \frac{1}{\pi_2^{(t)}} \sum_{i=1}^n \pi_{i,2}^{(t)} \|\boldsymbol{y}_i\|_2^2. \quad (32)$$

Note that the quantities with bars and subscript t are the weighted sum of these quantities over the weights corresponding to the inlier class at iteration t, and that $\boldsymbol{\alpha}_i^{(t)}$ are means of von Mises-Fisher distributions with parameters $\kappa_i^{(t)}$ and $\boldsymbol{\mu}_i^{(t)}$.

Finally, $\pi_{i,2}^{(t)}$ is the a posteriori probability that vector $\#i$ belongs to the inlier class, which is a useful piece of information. Indeed, it can be used to define an outlier detector, e.g., declaring that $\boldsymbol{y}_i$ is an outlier if $\pi_{i,2}^{(t)} < 0.5$ (maximum a posteriori detector).

D. Hyperparameter Estimation

The method presented before assumes that the hyperparameters κ and $\boldsymbol{\mu}$ 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 and $\boldsymbol{\mu}$ in (13)). 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.,

$$\kappa^{(t+1)} = A_d^{-1}(\|\overline{\boldsymbol{u}}_t\|_2), \quad \boldsymbol{\mu}^{(t+1)} = \frac{\overline{\boldsymbol{u}}_t}{\|\overline{\boldsymbol{u}}_t\|_2}. \quad (33)$$

Note that these equations have been obtained by using the expressions of the ML estimators for the von Mises-Fisher distribution parameters [15, 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 [16].

III. Experiments

Several experiments have been performed to illustrate the robustness of the proposed approach to outliers. Before presenting the results, note that the proposed algorithm only requires two parameters to be set: the stopping criterion of the algorithm (we chose to set a fixed number of iterations in our experiments) and the volume occupied by the outliers a, which can be fixed as the volume of the observation window. It is the volume of the smallest hypercube englobing all the observations, obtained by multiplying the differences between the maximum and minimum observed values of the different features. Note that the curse of dimensionality might occur as d increases. However, the experiments presented in this paper are restricted to 2D and 3D datasets.

In all the experiments, the center has been initialized by the mean of the noisy measurements denoted as $\boldsymbol{c}_0$, the initial radius has been fixed to its MLE given $\boldsymbol{c}_0$, i.e., $r_0 = \frac{1}{n}\sum_{i=1}^n \|\boldsymbol{z}_i - \boldsymbol{c}_0\|$, and the noise variance by its MLE given $(\boldsymbol{c}_0, r_0)$, i.e., $\sigma_0^2 = \frac{1}{nd}\sum_{i=1}^n \|\boldsymbol{z}_i - \boldsymbol{c}_0\|^2 - \frac{1}{d}r_0^2$. Moreover, the concentration parameter was initialized to $\kappa_0 = 1$, the mean direction to $\boldsymbol{c}_0/\|\boldsymbol{c}_0\|_2$ and the ratio of outliers to $\gamma_0 = 0.1$. Note that in all scenarios considered in this paper, the hyperparameters $\kappa, \boldsymbol{\mu}$ are unknown and therefore estimated jointly with the parameters $r, \boldsymbol{c}, \sigma^2$ and γ.

A. Illustrations on a synthetic 2D Dataset

The proposed method referred to as REM (for robust EM), is compared to the original EM proposed in [10]. The estimation performance is first evaluated using a simple scenario with $n = 200$, $d = 2$, $r = 6$, $\boldsymbol{c} = (-5, 5)^T$, $\sigma^2 = 0.25$ and $\boldsymbol{\mu} = (1, 1)^T/\sqrt{2}$. The estimation results obtained for representative

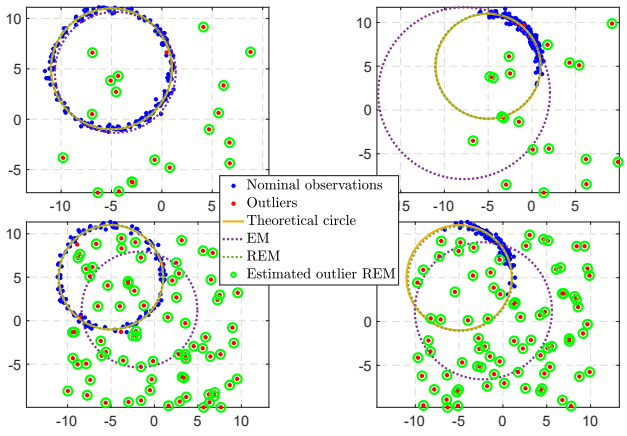


Fig. 1: Comparison of EM and robust EM. The first and second columns correspond to $\kappa = 0$ and $\kappa = 6$. The first and second rows correspond to $\gamma = 0.1$ and $\gamma = 0.4$. The blue points are the normal vectors (inliers) whereas the red points are the outliers. The yellow circle is the theoretical one used to generate the different observations, the dotted purple circle is obtained with the EM algorithm and the dotted green circle is obtained using the proposed robust EM algorithm. Finally, data circled in green are the data detected as outliers by REM.

values of γ and κ are depicted in Fig. 1[1]. The outliers were generated using a uniform distribution on $[-10, 10] \times [-10, 10]$ and the number of iterations of any EM algorithm was fixed to 100. As one can see, the EM algorithm is strongly affected by the presence of outliers, whereas REM provides better results, even with a proportion $\gamma = 40\%$ of outliers and a high value of the concentration parameter.

B. Monte-Carlo simulations

This section first analyzes the robustness of REM to the level of outliers γ. All the results presented here have been averaged using 500 Monte-Carlo runs. For $d = 2$, the configuration is the same as in Section III-A, whereas for $d = 3$, the parameters were fixed to $\boldsymbol{c} = (-5, 5, 3)^T$, $\boldsymbol{\mu} = (1, 1, 1)^T/\sqrt{3}$, and the outliers are sampled uniformly in the cube $[-10, 10]^3$. Once again, the algorithms are challenged in two configurations $\kappa = 0$ (uniform distribution) and $\kappa = 6$ (informative distribution), with the outlier ratio varying between 0 and 1. Figs. 2a and 2b show the averaged mean square errors (MSEs) of the vector containing the parameters of interest, namely $(r, \boldsymbol{c}^T, \sigma^2)^T$, and the vector of hyperparameters $(\kappa, \boldsymbol{\mu}^T)^T$. The proposed REM method was also compared to a robust version of the EM algorithm using the RANSAC [17] paradigm. Note that the RANSAC algorithm is also an iterative method. Therefore the combination of RANSAC and EM has a higher execution time compared to the proposed REM. RANSAC samples k_0 sets of size n_0 from the observations. From the k-th subset, it computes the corresponding solution $\boldsymbol{c}^{(k)}, r^{(k)}$ using this subset and evaluates the data that are in good agreement

[1]All the codes are available on the first author webpage http://perso.tesa.prd.fr/jlesouple/codes.html

TABLE I: Comparison of computation times (in seconds) of EM, REM and EM+RANSAC for $\gamma = 0.2$.

Algorithm	$\kappa = 0$ $d = 2$	$\kappa = 6$ $d = 2$	$\kappa = 0$ $d = 3$	$\kappa = 6$ $d = 3$
EM	0.014	0.014	0.015	0.017
REM	0.010	0.013	0.009	0.018
RANSAC+EM	0.530	0.529	2.171	3.537

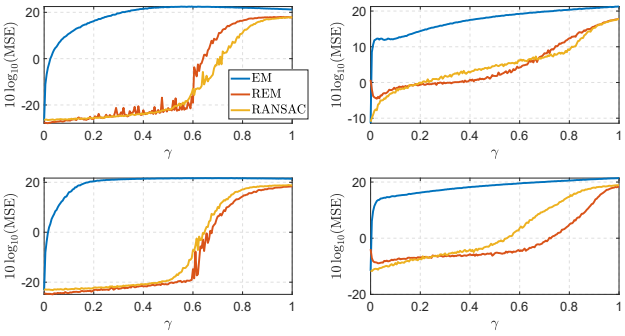
with these parameters. The parameters are finally estimated using all the conform data obtained with the subsets that lead to the maximum number of conform data. The RANSAC parameters were set as advised in [17] : $n_0 = 2d + 3$ for the minimum subset size, $k_0 = \log(1-p)/\log(1-w^{n_0})$ for the number of subset to sample, where $p = 0.9$ is the desired probability of having at least one subset containing only inliers, and $w = 0.5$ is the a priori proportion of outliers. Finally, a data $\boldsymbol{x}_i$ is declared as conform with the model when $|\|\boldsymbol{x_i} - \boldsymbol{c}^{(k)}\|_2^2 - {r^{(k)}}^2| \leq S$ with $S = 8$ fixed by cross-validation. As one can see, the EM solution is not robust to the presence of outliers, contrary to REM (until a breakpoint close to 60%) and RANSAC. The advantage of the proposed REM is its reduced execution time with respect to RANSAC.

The next experiments study the convergence speed of the algorithm (versus the number of iterations) and its robustness with respect to the noise variance. For those experiments, all the parameters have been set as explained before, and the outlier ratio is $\gamma = 0.1$. The influence of the number of iterations is depicted in Fig. 3a whereas that of σ^2 can be observed in Fig. 3b. As one can see, the REM algorithm takes more time to converge when $\kappa > 0$, i.e., when the distribution of the latent variables is not uniform. Regarding the noise variance, there seems to be a breakdown around $\sigma^2 = -10$dB beyond which the algorithm performance collapses. Note that in the extreme cases where $\gamma > 60\%$ and/or the noise variance is large, it becomes extremely difficult to discriminate the noisy observations from the outliers, especially if their proportion is unknown. This could be addressed by assigning informative hyperpriors to γ and σ^2. Finally, Table I provides quantitative values of the algorithm execution times confirming the interest of the proposed approach.

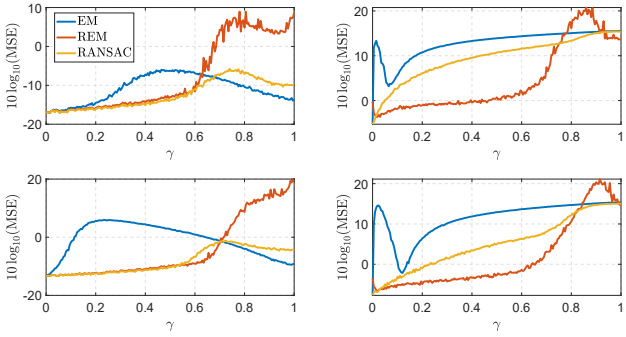
IV. Conclusion

This paper proposed a robust EM algorithm for hypersphere fitting. The algorithm was derived assuming a uniform distribution for the outliers and von Mises-Fisher distributions for latent variables associated with the observations. The resulting algorithm only requires two parameters to be adjusted: the stopping criterion for the EM iterations and the volume of the outlier distribution. Note that the hyperparameters of the von Mises-fisher distributions assigned to the latent variables can be also estimated by the algorithm.

The proposed algorithm was evaluated for circle and sphere fitting in various scenarios. The results obtained on simulated data are encouraging and show the competitiveness of the proposed approach with respect to the classical RANSAC algorithm, requiring less hyperparameters to adjust and a

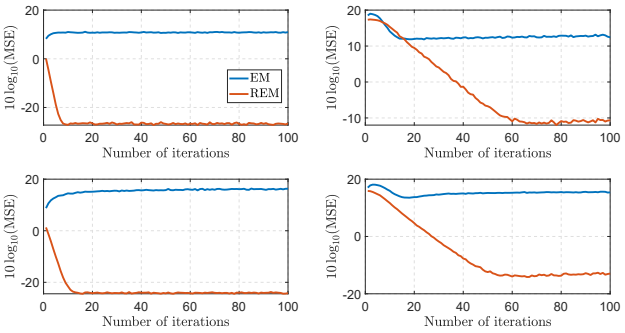


(a) Parameters of interest $(r, \boldsymbol{c}^T, \sigma^2)^T$.

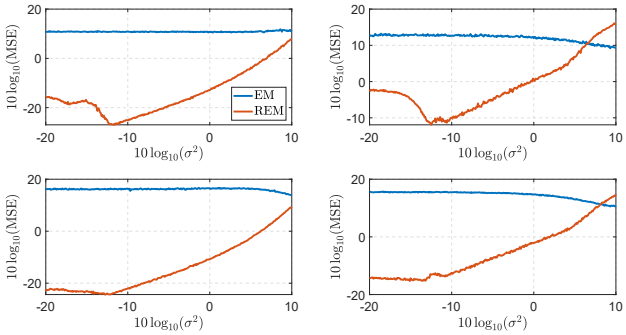


(b) Hyperparameters $(\kappa, \boldsymbol{\mu}^T)^T$.

Fig. 2: Comparison of MSEs (in dB) of EM, REM and EM+RANSAC for the parameters and hyperparameters versus the percentage of outliers. First row: $d = 2$, second row: $d = 3$, First column: $\kappa = 0$, second column: $\kappa = 6$. Note that when $\kappa = 0$, the estimation error of $\boldsymbol{\mu}$ is not taken into account since the von-Mises distribution does not depend on $\boldsymbol{\mu}$.



(a) Influence of the number of iterations.



(b) Influence of the noise variance σ^2.

Fig. 3: Comparison of MSEs (in dB) of EM and REM for the parameters of interest $(r, \boldsymbol{c}^T, \sigma^2)^T$. First row: $d = 2$, second row: $d = 3$, first column: $\kappa = 0$, second column: $\kappa = 6$.

significantly reduced execution time. Future work includes the generalization of the proposed work to the robust estimation of several hyperspheres with application to LIDAR calibration for real data. It would be also interesting to study the optimal performance of estimators for hypersphere fitting, e.g., through the derivation of Cramér-Rao bounds.

REFERENCES

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

[2] 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.

[3] 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.

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

[6] 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.

[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] 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.

[9] 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.

[10] J. Lesouple, B. Pilastre, Y. Altmann, and J.-Y. Tourneret, "Hypersphere Fitting from Noisy Data Using an EM Algorithm," *IEEE Signal Processing Letters*, to appear 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] P. H. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, April 2000.

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

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

[16] 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.

[17] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. Of the ACM*, vol. 24, no. 6, pp. 381–395, June. 1981.