Hypersphere Fitting from Noisy Data Using an EM Algorithm - Supplementary Material.

Julien LESOUPLE[1], Barbare PILASTRE[1], Yoann ALTMANN[3] and Jean-Yves TOURNERET[1,3]

[1]TéSA, 7 Boulevard de la Gare, 31500 Toulouse, France
[2]School of Engineering, Heriot-Watt University, EH14 4AS Edinburgh, U.K
[3]ENSEEIHT-IRIT, 2 Rue Camichel, 31071 Toulouse, France

January 6, 2021

Abstract

This document contains supplementary materials associated with the letter [1].

1 Introduction

The aim of the paper [1] is to estimate hyperspheres from noisy data clouds using the Expectation-Maximization (EM) with Von Mises-Fisher priors. This document details some derivations and shows additional experiments.

2 Derivations

2.1 Problem formulation

We consider n noisy measurements $\boldsymbol{z}_i \in \mathbb{R}^d, i = 1, ..., n$ 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. We introduce latent vectors $\boldsymbol{x}_i \in \mathbb{R}^d, i = 1, ..., n$ corresponding to the unknown true locations corrupted by an additive white Gaussian noise $\boldsymbol{n}_i$, i.e.,

$$\boldsymbol{z}_i = \boldsymbol{x}_i + \boldsymbol{n}_i, \tag{TR-1}$$

where $\boldsymbol{n}_i \sim \mathcal{N}(\boldsymbol{0}_d, \sigma^2 \boldsymbol{I}_d)$, $\boldsymbol{0}_d$ is the zero vector of $\mathbb{R}^d$, σ^2 is the unknown noise power and $\boldsymbol{I}_d$ is the $d \times d$ identity matrix. The vectors $\boldsymbol{x}_i$ are located on the hypersphere of radius r and center $\boldsymbol{c}$ and can thus be defined as affine transformations of unit random vectors $\boldsymbol{u}_i \in \mathbb{R}^d$ (with $\|\boldsymbol{u}_i\|_2 = 1$) such that

$$\boldsymbol{x}_i = \boldsymbol{c} + r\boldsymbol{u}_i. \tag{TR-2}$$

The vectors $\boldsymbol{u}_i$ denoted as latent vectors are located on the hypersphere $\mathcal{H}_d$ of $\mathbb{R}^d$ defined by $\|\boldsymbol{u}_i\|_2 = 1$. They are assigned an a priori von Mises-Fisher probability distribution denoted as $\boldsymbol{u}_i \sim \text{vMF}_d(\boldsymbol{u}_i; \boldsymbol{\mu}, \kappa)$ with the following density

$$f_d(\boldsymbol{u}_i; \boldsymbol{\mu}, \kappa) = C_d(\kappa) \exp\left(\kappa \boldsymbol{\mu}^T \boldsymbol{u}_i\right) 1_{\mathcal{H}_d}(\boldsymbol{u}_i), \tag{TR-3}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ is the mean direction (with $\|\boldsymbol{\mu}\|_2 = 1$), $\kappa \geq 0$ is the concentration parameter, and $C_d(\kappa)$ is the normalization constant given by

$$C_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)}, \tag{TR-4}$$

where $I_\nu(.)$ denotes the modified Bessel function of first kind of parameter ν.

2.2 Likelihood and complete likelihood

We store the unknown parameters of the statistical model in the vector $\boldsymbol{\theta} = (r, \boldsymbol{c}^T, \sigma)^T \in \mathbb{R}^{d+2}$ and the measurements in a matrix $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n) \in \mathbb{R}^{d\times n}$. The likelihood is then

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{Z}) &= \prod_{i=1}^{n} p(\boldsymbol{z}_i|\boldsymbol{\theta}) \\
&= \prod_{i=1}^{n} \int p(\boldsymbol{z}_i|\boldsymbol{u}_i, \boldsymbol{\theta}) p(\boldsymbol{u}_i) \mathrm{d}\boldsymbol{u}_i \\
&\propto \prod_{i=1}^{n} (\sigma^2)^{-\frac{d}{2}} \int_{\mathcal{H}} \exp\left(-\frac{1}{2\sigma^2}\left[(\boldsymbol{z}_i - \boldsymbol{c} - r\boldsymbol{u}_i)^T (\boldsymbol{z}_i - \boldsymbol{c} - r\boldsymbol{u}_i)\right]\right) \exp\left(\kappa \boldsymbol{\mu}^T \boldsymbol{u}_i\right) \mathrm{d}\boldsymbol{u}_i \\
&\propto \prod_{i=1}^{n} (\sigma^2)^{-\frac{d}{2}} \int_{\mathcal{H}} \exp\left(-\frac{1}{2\sigma^2}\left[\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2 - 2\left\{r(\boldsymbol{z}_i - \boldsymbol{c})^T + \sigma^2\kappa\boldsymbol{\mu}^T\right\}\boldsymbol{u}_i\right]\right) \mathrm{d}\boldsymbol{u}_i \\
&\propto \prod_{i=1}^{n} (\sigma^2)^{-\frac{d}{2}} \exp\left(-\frac{1}{2\sigma^2}\left[\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2\right]\right) \int_{\mathcal{H}} \exp\left(\frac{r(\boldsymbol{z}_i - \boldsymbol{c})^T\boldsymbol{u}_i + \sigma^2\kappa\boldsymbol{\mu}^T\boldsymbol{u}_i}{\sigma^2}\right) \mathrm{d}\boldsymbol{u}_i \\
&\propto \prod_{i=1}^{n} (\sigma^2)^{-\frac{d}{2}} \exp\left(-\frac{1}{2\sigma^2}\left[\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2\right]\right) \int_{\mathcal{H}} \exp\left(\frac{\|r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu}\|_2}{\sigma^2} \frac{r(\boldsymbol{z}_i - \boldsymbol{c})^T + \sigma^2\kappa\boldsymbol{\mu}^T}{\|r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu}\|_2}\boldsymbol{u}_i\right) \mathrm{d}\boldsymbol{u}_i.
\end{aligned} \tag{TR-5}$$

We denote

$$\kappa_i = \frac{\|r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu}\|_2}{\sigma^2} \tag{TR-6}$$

$$\boldsymbol{\mu}_i = \frac{r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu}}{\|r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu}\|_2}, \tag{TR-7}$$

to have

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{Z}) &\propto \prod_{i=1}^{n} (\sigma^2)^{-\frac{d}{2}} \exp\left(-\frac{1}{2\sigma^2}\left[\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2\right]\right) \int_{\mathcal{H}} \exp\left(\kappa_i\boldsymbol{\mu}_i^T\boldsymbol{u}_i\right) \mathrm{d}\boldsymbol{u}_i \\
&\propto \prod_{i=1}^{N} (\sigma^2)^{-\frac{d}{2}} \exp\left(-\frac{1}{2\sigma^2}\left[\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2\right]\right) C_d(\kappa_i)^{-1} \\
&\propto (\sigma^2)^{-\frac{nd}{2}} \exp\left(-\sum_{i=1}^{n} \frac{1}{2\sigma^2}\left[\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2\right]\right) \prod_{i=1}^{n} \frac{I_{d/2-1}(\kappa_i)}{\kappa_i^{d/2-1}}.
\end{aligned} \tag{TR-8}$$

Note that we have used the property that f_d in (TR-3) is a pdf whose integral is one. If one wants to compute the maximum likelihood estimator of $\boldsymbol{\theta}$, one has to differentiate the log-likelihood with respect to $\boldsymbol{\theta}$. This log-likelihood is defined by

$$\log \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{Z}) = K - \frac{nd}{2}\log\sigma^2 - \sum_{i=1}^{n} \frac{1}{2\sigma^2}\left[\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2\right] + \sum_{i=1}^{n} \log\left(\frac{I_{d/2-1}(\kappa_i)}{\kappa_i^{d/2-1}}\right). \tag{TR-9}$$

Derivating the log-likelihood with respect to r yields

$$\frac{\partial \log \mathcal{L}}{\partial r} = -n\frac{r}{\sigma^2} + \sum_{i=1}^{n} \frac{\partial}{\partial r} \log\left(\frac{I_{d/2-1}(\kappa_i)}{\kappa_i^{d/2-1}}\right),$$

with

$$\frac{\partial}{\partial r}\log\left(\frac{I_{d/2-1}\left(\kappa_i\right)}{\kappa_i^{d/2-1}}\right)=\frac{\partial\kappa_i}{\partial r}\frac{\partial}{\partial\kappa_i}\log\left(\frac{I_{d/2-1}\left(\kappa_i\right)}{\kappa_i^{d/2-1}}\right), \tag{TR-10}$$

and [2, Chap. 10.29.4]

$$\begin{aligned}&\frac{1}{x}\frac{\partial}{\partial x}x^{-\nu}I_\nu=x^{-\nu-1}I_{\nu+1}(x)\\ \Rightarrow&\frac{\partial}{\partial x}x^{-\nu}I_\nu=x^{-\nu}I_{\nu+1}(x)\\ \Rightarrow&\frac{\partial}{\partial\kappa_i}\frac{I_{d/2-1}}{\kappa_i^{d/2-1}}=\frac{I_{d/2}(\kappa_i)}{\kappa_i^{d/2-1}},\end{aligned} \tag{TR-11}$$

so

$$\begin{aligned}\frac{\partial}{\partial\kappa_i}\log\left(\frac{I_{d/2-1}\left(\kappa_i\right)}{\kappa_i^{d/2-1}}\right)&=\frac{\frac{I_{d/2}(\kappa_i)}{\kappa_i^{d/2-1}}}{\frac{I_{d/2-1}(\kappa_i)}{\kappa_i^{d/2-1}}}\\&=\frac{I_{d/2}(\kappa_i)}{I_{d/2-1}\left(\kappa_i\right)}.\end{aligned} \tag{TR-12}$$

On the other hand

$$\begin{aligned}\frac{\partial\kappa_i}{\partial r}&=\frac{\partial}{\partial r}\frac{\|r(\boldsymbol{z}_i-\boldsymbol{c})+\sigma^2\kappa\boldsymbol{\mu}\|_2}{\sigma^2}\\&=\frac{1}{\sigma^2}(\boldsymbol{z}_i-\boldsymbol{c})^T\frac{r(\boldsymbol{z}_i-\boldsymbol{c})+\sigma^2\kappa\boldsymbol{\mu}}{\|r(\boldsymbol{z}_i-\boldsymbol{c})+\sigma^2\kappa\boldsymbol{\mu}\|_2},\end{aligned} \tag{TR-13}$$

leading to

$$\frac{\partial\log\mathcal{L}}{\partial r}=-n\frac{r}{\sigma^2}+\sum_{i=1}^n\frac{1}{\sigma^2}(\boldsymbol{z}_i-\boldsymbol{c})^T\frac{r(\boldsymbol{z}_i-\boldsymbol{c})+\sigma^2\kappa\boldsymbol{\mu}}{\|r(\boldsymbol{z}_i-\boldsymbol{c})+\sigma^2\kappa\boldsymbol{\mu}\|_2}\frac{I_{d/2}\left(\frac{\|r(\boldsymbol{z}_i-\boldsymbol{c})+\sigma^2\kappa\boldsymbol{\mu}\|_2}{\sigma^2}\right)}{I_{d/2-1}\left(\frac{\|r(\boldsymbol{z}_i-\boldsymbol{c})+\sigma^2\kappa\boldsymbol{\mu}\|_2}{\sigma^2}\right)}. \tag{TR-14}$$

Solving for $\frac{\partial\log\mathcal{L}}{\partial r}=0$ seems pretty hard (and we have to do the same for $\boldsymbol{c}$ and σ) so we propose to study an EM algorithm to simplify the estimation problem. The EM algorithm relies on the so-called complete likelihood defined as

$$\begin{aligned}\mathcal{L}_c\left(\boldsymbol{\theta};\boldsymbol{Z},\boldsymbol{U}\right)&=\prod_{i=1}^n p(\boldsymbol{z}_i,\boldsymbol{u}_i|\boldsymbol{\theta})\\&=\prod_{i=1}^n p(\boldsymbol{z}_i|\boldsymbol{\theta},\boldsymbol{u}_i)p(\boldsymbol{u}_i)\\&\propto\prod_{i=1}^n(\sigma^2)^{-\frac{d}{2}}\exp\left(-\frac{1}{2\sigma^2}\left[(\boldsymbol{z}_i-\boldsymbol{c}-r\boldsymbol{u}_i)^T\left(\boldsymbol{z}_i-\boldsymbol{c}-r\boldsymbol{u}_i\right)\right]\right)\exp\left(\kappa\boldsymbol{\mu}^T\boldsymbol{u}_i\right)1_{\mathcal{H}}(\boldsymbol{u}_{\boldsymbol{i}})\\&\propto(\sigma^2)^{-\frac{nd}{2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n\left[(\boldsymbol{z}_i-\boldsymbol{c}-r\boldsymbol{u}_i)^T\left(\boldsymbol{z}_i-\boldsymbol{c}-r\boldsymbol{u}_i\right)-2\sigma^2\kappa\boldsymbol{\mu}^T\boldsymbol{u}_i\right]\right)1_{\mathcal{H}}(\boldsymbol{u}_{\boldsymbol{i}})\\&\propto(\sigma^2)^{-\frac{nd}{2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n\left[\|\boldsymbol{z}_i-\boldsymbol{c}\|_2^2+r^2-2(r(\boldsymbol{z}_i-\boldsymbol{c})+\sigma^2\kappa\boldsymbol{\mu})^T\boldsymbol{u}_i\right]\right)1_{\mathcal{H}}(\boldsymbol{u}_{\boldsymbol{i}}),\end{aligned} \tag{TR-15}$$

where $\boldsymbol{U}=(\boldsymbol{u}_1,...,\boldsymbol{u}_n)\in\mathbb{R}^{d\times n}$.

2.3 E step

The E step consists in computing the function

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbb{E}_{\boldsymbol{U}|\boldsymbol{Z},\boldsymbol{\theta}^{(t)}}\left[\log \mathcal{L}_c\left(\boldsymbol{\theta}; \boldsymbol{Z}, \boldsymbol{U}\right)\right]. \tag{TR-16}$$

Assuming $\boldsymbol{u}_i \in \mathcal{H}_d$ to alleviate the notations, we have

$$\log \mathcal{L}_c\left(\boldsymbol{\theta}; \boldsymbol{Z}, \boldsymbol{U}\right) = K - \frac{nd}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2 - 2(r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu})^T\boldsymbol{u}_i\right],$$

hence

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \mathbb{E}_{\boldsymbol{U}|\boldsymbol{Z},\boldsymbol{\theta}^{(t)}}\left[K - \frac{nd}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2 - 2(r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu})^T\boldsymbol{u}_i\right)\right]\\
&= K - \frac{nd}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2 - 2(r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu})^T\mathbb{E}_{\boldsymbol{U}|\boldsymbol{Z},\boldsymbol{\theta}^{(t)}}\left[\boldsymbol{u}_i\right]\right).
\end{aligned} \tag{TR-17}$$

Using Bayes' theorem and the derivations that led to (TR-5) and (TR-15) allows the following result to be obtained

$$\begin{aligned}
p(\boldsymbol{U}|\boldsymbol{Z}, \boldsymbol{\theta}^{(t)}) &= \prod_{i=1}^{n} p(\boldsymbol{u}_i|\boldsymbol{z}_i, \boldsymbol{\theta}^{(t)})\\
&\propto \prod_{i=1}^{n} p(\boldsymbol{z}_i|\boldsymbol{u}_i, \boldsymbol{\theta}^{(t)})p(\boldsymbol{u}_i)\\
&\propto ((\sigma^{(t)})^2)^{-\frac{nd}{2}}\exp\left(-\frac{1}{2(\sigma^{(t)})^2}\sum_{i=1}^{n}\left[\|\boldsymbol{z}_i - \boldsymbol{c}^{(t)}\|_2^2 + (r^{(t)})^2 - 2(r^{(t)}(\boldsymbol{z}_i - \boldsymbol{c}^{(t)}) + (\sigma^{(t)})^2\kappa\boldsymbol{\mu})^T\boldsymbol{u}_i\right]\right)1_{\mathcal{H}}(\boldsymbol{u_i})\\
&\propto \prod_{i=1}^{n}\exp\left(\kappa_i^{(t)}(\boldsymbol{\mu}_i^{(t)})^T\boldsymbol{u}_i\right)1_{\mathcal{H}}(\boldsymbol{u_i}),
\end{aligned} \tag{TR-18}$$

where

$$\kappa_i^{(t)} = \frac{\|r^{(t)}(\boldsymbol{z}_i - \boldsymbol{c}^{(t)}) + (\sigma^{(t)})^2\kappa\boldsymbol{\mu}\|_2}{(\sigma^{(t)})^2} \tag{TR-19}$$

$$\boldsymbol{\mu}_i^{(t)} = \frac{r^{(t)}(\boldsymbol{z}_i - \boldsymbol{c}^{(t)}) + (\sigma^{(t)})^2\kappa\boldsymbol{\mu}}{\|r^{(t)}(\boldsymbol{z}_i - \boldsymbol{c}^{(t)}) + (\sigma^{(t)})^2\kappa\boldsymbol{\mu}\|_2}. \tag{TR-20}$$

Thus, the conditional distribution of $\boldsymbol{u}_i$ given $\boldsymbol{Z}$ and $\boldsymbol{\theta}^{(t)}$ is a von Mises-Fisher distribution with parameters $\kappa_i^{(t)}$ and $\boldsymbol{\mu}_i^{(t)}$, whose expectation is

$$\mathbb{E}_{\boldsymbol{U}|\boldsymbol{Z},\boldsymbol{\theta}^{(t)}}\left[\boldsymbol{u}_i\right] = \frac{I_{d/2}(\kappa_i^{(t)})}{I_{d/2-1}(\kappa_i^{(t)})}\boldsymbol{\mu}_i^{(t)}, \tag{TR-21}$$

that can be plugged into (TR-17) to obtain

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = K - \frac{nd}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\|\boldsymbol{z}_i - \boldsymbol{c}\|_2^2 + r^2 - 2(r(\boldsymbol{z}_i - \boldsymbol{c}) + \sigma^2\kappa\boldsymbol{\mu})^T\frac{I_{d/2}(\kappa_i^{(t)})}{I_{d/2-1}(\kappa_i^{(t)})}\boldsymbol{\mu}_i^{(t)}\right).$$

2.4 M step

In the M step, the function Q is optimized with respect to $\boldsymbol{\theta}$. To simplify the computations, we define

$$\boldsymbol{\alpha}_i^{(t)} = \frac{I_{d/2}(\kappa_i^{(t)})}{I_{d/2-1}(\kappa_i^{(t)})}\boldsymbol{\mu}_i^{(t)}, \tag{TR-22}$$

to obtain

$$\begin{aligned}\arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \arg\min_{r,\boldsymbol{c},\sigma} \frac{nd}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\boldsymbol{c}^T\boldsymbol{c} + \boldsymbol{z}_i^T\boldsymbol{z}_i + r^2 - 2\boldsymbol{c}^T\boldsymbol{z}_i + 2r\boldsymbol{c}^T\boldsymbol{\alpha}_i^{(t)} - 2r\boldsymbol{z}_i^T\boldsymbol{\alpha}_i^{(t)} - 2\sigma^2\kappa\boldsymbol{\mu}^T\boldsymbol{\alpha}_i^{(t)}\right)\\ &= \arg\min_{r,\boldsymbol{c},\sigma} \frac{nd}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\left(n\boldsymbol{c}^T\boldsymbol{c} + \sum_{i=1}^{n}\boldsymbol{z}_i^T\boldsymbol{z}_i + nr^2 - 2\boldsymbol{c}^T\sum_{i=1}^{n}\boldsymbol{z}_i + 2r\boldsymbol{c}^T\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{(t)} - 2r\sum_{i=1}^{n}\boldsymbol{z}_i^T\boldsymbol{\alpha}_i^{(t)}\right).\end{aligned} \tag{TR-23}$$

The problem can be decoupled with $(r,\boldsymbol{c})$ on one side and σ^2 on the other side because

$$\begin{aligned}&\arg\min_{r,\boldsymbol{c}} \frac{nd}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\left(n\boldsymbol{c}^T\boldsymbol{c} + \sum_{i=1}^{n}\boldsymbol{z}_i^T\boldsymbol{z}_i + nr^2 - 2\boldsymbol{c}^T\sum_{i=1}^{n}\boldsymbol{z}_i + 2r\boldsymbol{c}^T\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{(t)} - 2r\sum_{i=1}^{n}\boldsymbol{z}_i^T\boldsymbol{\alpha}_i^{(t)}\right)\\ &= \arg\min_{r,\boldsymbol{c}}\left(n\boldsymbol{c}^T\boldsymbol{c} + nr^2 - 2\boldsymbol{c}^T\sum_{i=1}^{n}\boldsymbol{z}_i + 2r\boldsymbol{c}^T\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{(t)} - 2r\sum_{i=1}^{n}\boldsymbol{z}_i^T\boldsymbol{\alpha}_i^{(t)}\right).\end{aligned} \tag{TR-24}$$

Denoting

$$\boldsymbol{H}^{(t)} = \begin{bmatrix} n & \sum_{i=1}^{n}(\boldsymbol{\alpha}_i^{(t)})^T \\ \sum_{i=1}^{n}\boldsymbol{\alpha}_i^{(t)} & nI_d\end{bmatrix} \tag{TR-25}$$

$$\boldsymbol{f}^{(t)} = \begin{bmatrix}\sum_{i=1}^{n}(\boldsymbol{\alpha}_i^{(t)})^T\boldsymbol{z}_i \\ \sum_{i=1}^{n}\boldsymbol{z}_i\end{bmatrix}, \tag{TR-26}$$

Problem (TR-24) can be written as

$$(r^{(t+1)}, \boldsymbol{c}^{(t+1)}) = \arg\min_{\boldsymbol{\theta}_0=(r,\boldsymbol{c}^T)^T\in\mathbb{R}^{d+1}} \frac{1}{2}\boldsymbol{\theta}_0^T\boldsymbol{H}^{(t)}\boldsymbol{\theta}_0 - (\boldsymbol{f}^{(t)})^T\boldsymbol{\theta}_0. \tag{TR-27}$$

2.4.1 Determinant of the matrix $\boldsymbol{H}^{(t)}$

We propose to show by induction on the dimension d that the determinant of the matrix $\boldsymbol{H}^{(t)}$ is

$$\det(\boldsymbol{H}^{(t)}) = n^{d-1}\left(n^2 - \left\|\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{(t)}\right\|_2^2\right). \tag{TR-28}$$

To simplify the notations, we have removed the superscript $^{(t)}$ from $\boldsymbol{H}$ and have included a subscript associated with the dimension of the hypersphere d. Moreover, we introduce the notation $\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{(t)} = \boldsymbol{a}_d = (a_1,\ldots,a_d)^T \in \mathbb{R}^d$.

- Initialization $d=1$

When $d=1$, we have

$$\boldsymbol{H}_1 = \begin{bmatrix} n & a_1 \\ a_1 & n\end{bmatrix}, \tag{TR-29}$$

and

$$\begin{aligned}\det(\boldsymbol{H}_1) &= n^2 - a_1^2\\ &= n^{1-1}(n^2 - \|a_1\|_2^2).\end{aligned} \tag{TR-30}$$

Therefore the property is true for $d=1$.

- Induction

We assume that for a given $d \geq 1$ we have

$$\det(\boldsymbol{H}_d) = n^{d-1}\left(n^2 - \|\boldsymbol{a}_d\|_2^2\right). \tag{TR-31}$$

and we show that

$$\det(\boldsymbol{H}_{d+1}) = n^{d}\left(n^2 - \|\boldsymbol{a}_{d+1}\|_2^2\right), \tag{TR-32}$$

Using a block decomposition of $\boldsymbol{H}_{d+1}$ versus $\boldsymbol{H}_d$, we obtain

$$\begin{aligned}
\det(\boldsymbol{H}_{d+1}) &= \det\left(\begin{bmatrix} & & & a_{d+1} \\ & & & 0 \\ & \boldsymbol{H}_d & & \vdots \\ & & & 0 \\ a_{d+1} & 0 & \dots\ 0 & n \end{bmatrix}\right) \\
&= n \det(\boldsymbol{H}_d) + (-1)^{d+1} a_{d+1} \det\left(\begin{bmatrix} \boldsymbol{a}_d & & nI_d & \\ a_{d+1} & 0 & \dots & 0 \end{bmatrix}\right) \\
&= n \det(\boldsymbol{H}_d) + (-1)^{d+1}(-1)^d a_{d+1}^2 n^d \\
&= n n^{d-1}\left(n^2 - \|\boldsymbol{a}_d\|_2^2\right) - a_{d+1}^2 n^d \\
&= n^d\left(n^2 - \|\boldsymbol{a}_d\|_2^2 - a_{d+1}^2\right) \\
&= n^d\left(n^2 - \|\boldsymbol{a}_{d+1}\|_2^2\right),
\end{aligned}$$

which proves the induction and thus (TR-28).

2.4.2 Invertibility of the matrix $\boldsymbol{H}_d$

We would like to show that for any $d \geq 2$, the matrix $\boldsymbol{H}_d$ is invertible. We have from (TR-28)

$$\det(\boldsymbol{H}_d) = n^{d-1}\left(n^2 - \left\|\sum_{i=1}^{n} \boldsymbol{\alpha}_i^{(t)}\right\|_2^2\right). \tag{TR-33}$$

Applying the triangular inequality leads to

$$\left\|\sum_{i=1}^{n} \boldsymbol{\alpha}_i^{(t)}\right\|_2 < \sum_{i=1}^{n} \left\|\boldsymbol{\alpha}_i^{(t)}\right\|_2, \tag{TR-34}$$

where

$$\begin{aligned}
\left\|\boldsymbol{\alpha}_i^{(t)}\right\|_2 &= \left\|\frac{I_{d/2}(\kappa_i^{(t)})}{I_{d/2-1}(\kappa_i^{(t)})}\boldsymbol{\mu}_i^{(t)}\right\|_2 \\
&= \frac{I_{d/2}(\kappa_i^{(t)})}{I_{d/2-1}(\kappa_i^{(t)})}\left\|\boldsymbol{\mu}_i^{(t)}\right\|_2 \\
&\leq 1,
\end{aligned} \tag{TR-35}$$

because $\left\|\boldsymbol{\mu}_i^{(t)}\right\|_2 = 1$ and I_ν is a decreasing function of ν [2, Chap. 10.37]. Using (TR-34) yields

$$\left\|\sum_{i=1}^{n} \boldsymbol{\alpha}_i^{(t)}\right\|_2 < n. \tag{TR-36}$$

The strict inequality in (TR-34) is justified since all the $\boldsymbol{\mu}_i^{(t)}$, and hence all the $\boldsymbol{\alpha}_i^{(t)}$, are not colinear. Finally, we have

$$\left\| \sum_{i=1}^{n} \boldsymbol{\alpha}_i^{(t)} \right\|_2^2 < n^2 \Leftrightarrow \det(\boldsymbol{H}_d) > 0 \quad \text{for any } d \geq 2, \tag{TR-37}$$

which proves the invertibility of the matrix $\boldsymbol{H}^{(t)}$ for any dimension d. Moreover, $\boldsymbol{H}^{(t)}$ is a so-called arrowhead matrix whose inverse has a closed-form expression [3]

$$\boldsymbol{H}_d^{-1} = \frac{1}{n^2 - \boldsymbol{a}_d^T \boldsymbol{a}_d} \begin{bmatrix} n & -\boldsymbol{a}_d^T \\ -\boldsymbol{a}_d & \frac{n^2 - \boldsymbol{a}_d^T \boldsymbol{a}_d}{n} I_d + \frac{1}{n} \boldsymbol{a}_d \boldsymbol{a}_d^T \end{bmatrix}. \tag{TR-38}$$

2.4.3 Solution of the minimization problem

The problem (TR-27) has a global minimum if $\boldsymbol{H}^{(t)}$ is symmetric positive definite. Here, $\boldsymbol{H}^{(t)}$ is clearly symmetric. Using the Sylvester criterion, a matrix $\boldsymbol{A} = (a_{i,j})_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$ is positive definite if and only if all the n submatrices $\boldsymbol{A}_p = (a_{i,j})_{1 \leq i,j \leq p}, p = 1, \ldots, n$ have a strictly positive determinant. For $p = 2, \ldots, d+1$, the submatrices of $\boldsymbol{H}_d$ are the matrices $\boldsymbol{H}_1, ..., \boldsymbol{H}_d$ defined before. We have seen that the determinants of these matrices are strictly positive. The missing submatrix corresponds to $p = 1$, i.e., to the first coefficient of the matrix $\boldsymbol{H}^{(t)}$, which is n and thus is strictly positive. Therefore the Sylvester criterion allows us to claim that the matrix $\boldsymbol{H}^{(t)}$ is positive definite, and therefore the problem (TR-27) has a unique solution, given by

$$\begin{bmatrix} r^{(t+1)} \\ \boldsymbol{c}^{(t+1)} \end{bmatrix} = (\boldsymbol{H}^{(t)})^{-1} \boldsymbol{f}^{(t)}. \tag{TR-39}$$

We can deduce the solution for $\sigma^{(t)}$ solving

$$\arg\min_{\sigma} \frac{nd}{2} \log \sigma^2 + \frac{1}{2\sigma^2} M^{(t+1)}, \tag{TR-40}$$

where

$$\begin{aligned} M^{(t+1)} &= \sum_{i=1}^{n} \left((\boldsymbol{c}^{(t+1)})^T \boldsymbol{c}^{(t+1)} + (r^{(t+1)})^2 - 2(\boldsymbol{c}^{(t+1)})^T \boldsymbol{z}_i + 2r^{(t+1)} (\boldsymbol{c}^{(t+1)})^T \boldsymbol{\alpha}_i^{(t)} - 2r^{(t+1)} \boldsymbol{z}_i^T \boldsymbol{\alpha}_i^{(t)} + \boldsymbol{z}_i^T \boldsymbol{z}_i \right) \\ &= 2 \left(\frac{1}{2} \boldsymbol{\theta}_0^T \boldsymbol{H}^{(t)} \boldsymbol{\theta}_0 - (\boldsymbol{f}^{(t)})^T \boldsymbol{\theta}_0 \right) + \sum_{i=1}^{n} \|\boldsymbol{z}_i\|_2^2 \\ &= \sum_{i=1}^{n} \|\boldsymbol{z}_i\|_2^2 - (\boldsymbol{f}^{(t)})^T (\boldsymbol{H}^{(t)})^{-1} \boldsymbol{f}^{(t)} \\ &= \sum_{i=1}^{n} \|\boldsymbol{z}_i\|_2^2 - (\boldsymbol{f}^{(t)})^T \begin{bmatrix} r^{(t+1)} \\ \boldsymbol{c}^{(t+1)} \end{bmatrix}. \end{aligned} \tag{TR-41}$$

The number $M^{(t+1)}$ can be expressed as

$$\begin{aligned} M^{(t+1)} &= \sum_{i=1}^{n} \left[(\boldsymbol{c}^{(t+1)})^T \boldsymbol{c}^{(t+1)} + (r^{(t+1)})^2 - 2(\boldsymbol{c}^{(t+1)})^T \boldsymbol{z}_i + 2r^{(t+1)} (\boldsymbol{c}^{(t+1)})^T \boldsymbol{\alpha}_i^{(t)} - 2r^{(t+1)} \boldsymbol{z}_i^T \boldsymbol{\alpha}_i^{(t)} + \boldsymbol{z}_i^T \boldsymbol{z}_i \right] \\ &= \sum_{i=1}^{n} \left[(\boldsymbol{c}^{(t+1)} - \boldsymbol{z}_i + r^{(t+1)} \boldsymbol{\alpha}_i^{(t)})^T (\boldsymbol{c}^{(t+1)} - \boldsymbol{z}_i + r^{(t+1)} \boldsymbol{\alpha}_i^{(t)}) + (r^{(t+1)})^2 (1 - (\boldsymbol{\alpha}_i^{(t)})^T \boldsymbol{\alpha}_i^{(t)}) \right] \\ &= \sum_{i=1}^{n} \left[\left\| \boldsymbol{c}^{(t+1)} - \boldsymbol{z}_i + r^{(t+1)} \boldsymbol{\alpha}_i^{(t)} \right\|_2^2 + (r^{(t+1)})^2 \left(1 - \left\| \boldsymbol{\alpha}_i^{(t)} \right\|_2^2 \right) \right]. \end{aligned} \tag{TR-42}$$

Using (TR-35), we obtain $\left(1 - \left\|\boldsymbol{\alpha}_i^{(t)}\right\|_2^2\right) \geq 0$ for any $i = 1, \ldots, n$, which guarantees

$$M^{(t+1)} \geq 0. \tag{TR-43}$$

Denoting $f_\sigma : \sigma \mapsto \frac{nd}{2} \log \sigma^2 + \frac{1}{2\sigma^2} M^{(t+1)}$ the function to minimize with respect to σ, we have

$$\frac{\mathrm{d}f_\sigma}{\mathrm{d}\sigma} = nd\frac{1}{\sigma} - \frac{1}{\sigma^3} M^{(t+1)}, \tag{TR-44}$$

and

$$\frac{\mathrm{d}^2 f_\sigma}{\mathrm{d}\sigma^2} = -nd\frac{1}{\sigma^2} + \frac{3}{\sigma^4} M^{(t+1)}. \tag{TR-45}$$

Our problem has a unique global optimum given by

$$\frac{\mathrm{d}f_\sigma}{\mathrm{d}\sigma} = 0 \Leftrightarrow \sigma^{(t+1)} = \sqrt{\frac{M^{(t+1)}}{nd}}. \tag{TR-46}$$

The seconde derivative of this function at this point is

$$\begin{aligned} \frac{\mathrm{d}^2 f_\sigma}{\mathrm{d}\sigma^2}(\sigma^{(t+1)}) &= -\frac{(nd)^2}{M^{(t+1)}} + \frac{3(nd)^2}{(M^{(t+1)})^2} M^{(t+1)} \\ &= \frac{2(nd)^2}{M^{(t+1)}}, \end{aligned} \tag{TR-47}$$

which is positive ensuring that the optimum of f_σ is a minimum.

3 Additional experiments

This section presents several additional results compared to those available in the signal processing letter [1].

3.1 Uniform prior in 2D

This section shows more results in 2 dimensions, obtained using a uniform prior over the unit circle for the latent variable. The scenario is explained in Fig. 1, where $\boldsymbol{c} = (2, 3), r = 5$ and $\sigma = 0.1$. For the simulations, 500 Monte-Carlo runs were ran for each value of $\sigma \in [0.01, 10]$. For each run, the center has been sampled uniformly on the grid $\{-5, -4, \ldots, 4, 5\} \times \{-5, -4, \ldots, 4, 5\}$ and r is an integer uniformly sampled between 1 and 10. Corresponding MSEs for $\boldsymbol{\theta} = (r, \boldsymbol{c}^T)^T$ are displayed in Fig. 2 and for σ in Fig. 3. In addition here, we show the MSEs on the radius r and the center $\boldsymbol{c}$ in Fig. 4a and Fig. 4b.

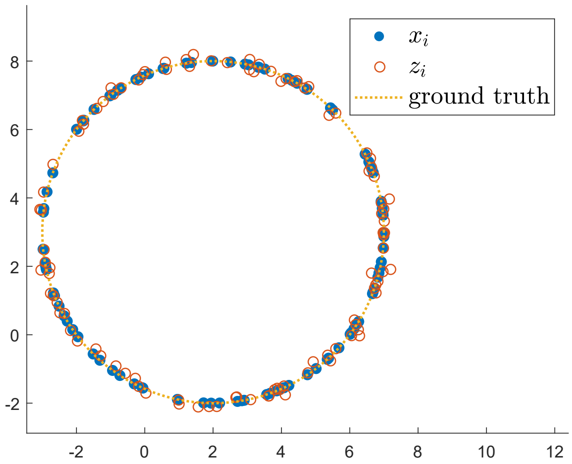


Figure 1: Scenario with a uniform prior for the latent variables, with $\boldsymbol{c} = (2, 3), r = 5$ and $\sigma = 0.1$.

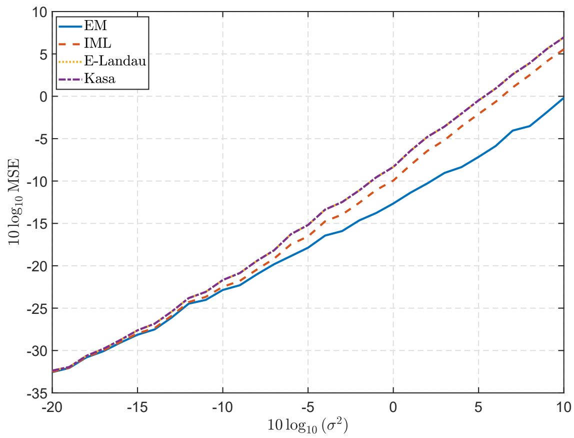


Figure 2: MSEs for $\hat{\boldsymbol{\theta}} = (\hat{r}, \hat{\boldsymbol{c}}^T)^T$ for E-Landau, Kasa, IML and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when latent variables have a uniform prior ($\kappa = 0$) in 2D.

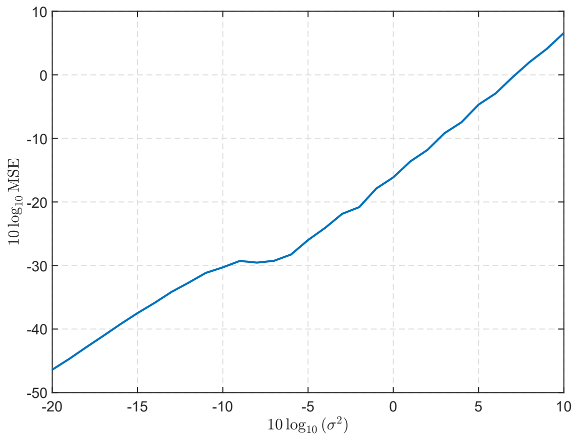
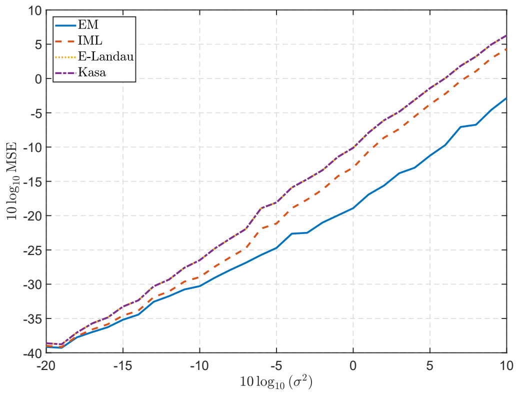
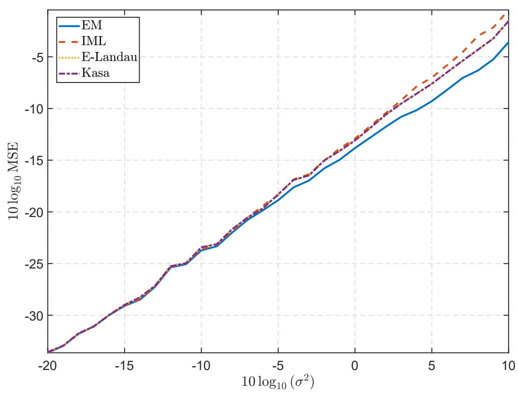


Figure 3: MSEs for σ^2 EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when latent variables have a uniform prior ($\kappa = 0$) in 2D.



(a) MSE on the radius.



(b) MSE on the center.

Figure 4: MSEs for the radius (left) and the center (right) for E-Landau, Kasa, IML and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when latent variables have a uniform prior ($\kappa = 0$) in 2D.

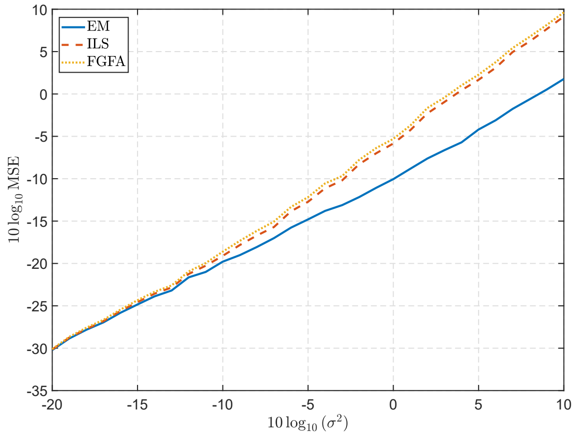


Figure 5: MSEs for $\hat{\boldsymbol{\theta}} = (\hat{r}, \hat{\boldsymbol{c}}^T)^T$ for E-Landau, Kasa, IML and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when latent variables have a uniform prior ($\kappa = 0$) in 3D.

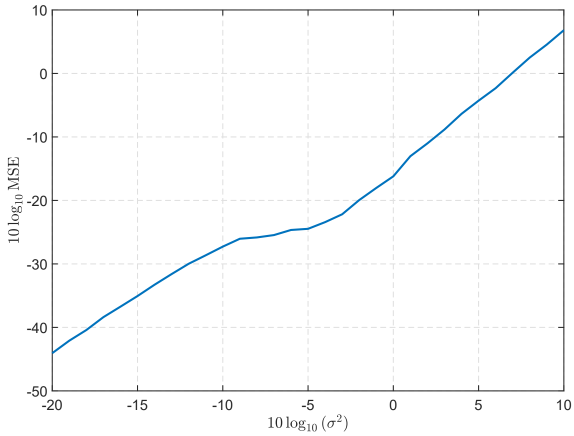
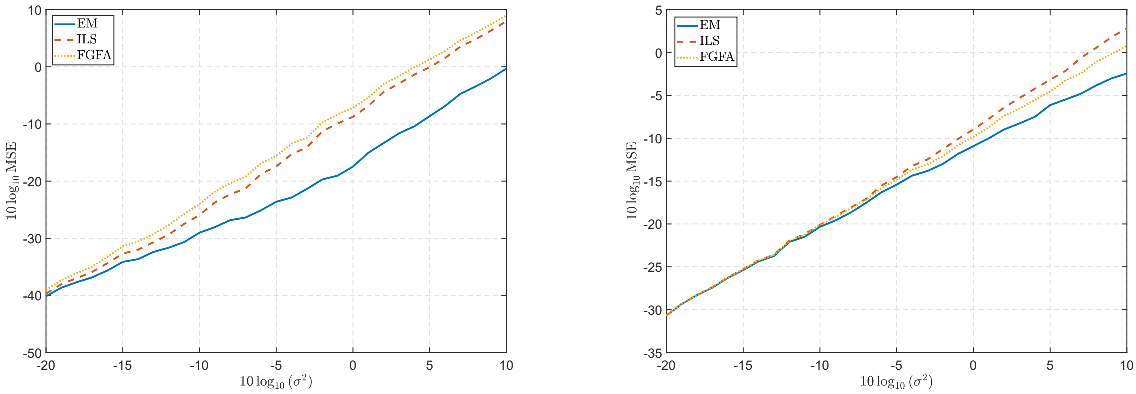


Figure 6: MSEs for σ^2 EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when latent variables have a uniform prior ($\kappa = 0$) in 3D.

3.2 Uniform prior in 3D

This section shows more results in 3 dimensions, obtained using a uniform prior over the unit circle for the latent variable. For the simulations, 500 Monte-Carlo runs were ran for each value of $\sigma \in [0.01, 10]$. For each run, the center has been sampled uniformly on the grid $\{-5, -4, \ldots, 4, 5\} \times \{-5, -4, \ldots, 4, 5\}$ and r is an integer uniformly sampled between 1 and 10. Corresponding MSEs for $\boldsymbol{\theta} = (r, \boldsymbol{c})$ are displayed in Fig. 5 and for σ in Fig. 6. In addition here, we show the MSEs on the radius r and the center $\boldsymbol{c}$ in Fig. 7a and Fig. 7b.



(a) MSE on the radius.

(b) MSE on the center.

Figure 7: MSEs for the radius (left) and the center (right) for FGFA, ILS and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when latent variables have a uniform prior ($\kappa = 0$) in 3D.

3.3 Informative prior in 2D

This section presents some simulation results obtained with a von Mises-Fisher prior with parameters $\kappa = 2$ and $\boldsymbol{\mu} = \left[\cos(\frac{\pi}{4}) \quad \sin(\frac{\pi}{4})\right]^T$. The scenario is illustrated in Fig. 8 for $\boldsymbol{c} = (1, -2), r = 5$ and $\sigma = 0.1$. The same previous strategy was applied to obtain the corresponding MSEs for $\hat{\boldsymbol{\theta}}$, which are displayed in Fig. 9 and in Fig. 10 for σ^2 . We also display the MSEs for both the radius r and the center $\boldsymbol{c}$ in Figs. 11a and 11b. The results confirm the interest of the EM algorithm for hypersphere fitting using an informative prior defined as a von Mises-Fisher distribution.

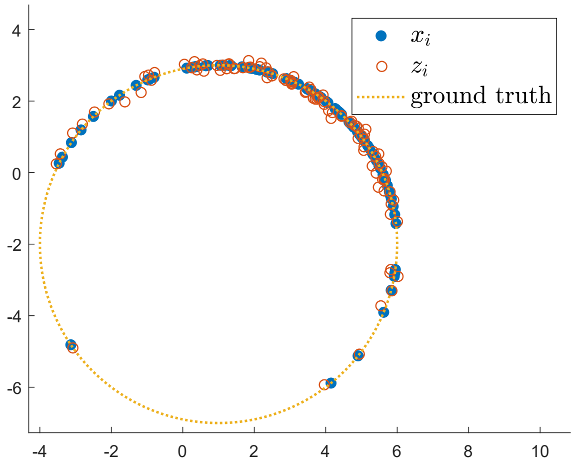


Figure 8: Scenario for a von Mises-Fisher prior with parameters $\kappa = 2$ and $\boldsymbol{\mu}$ corresponding to an angle $\pi/4$, for $\boldsymbol{c} = (1, -2), r = 5$ and $\sigma = 0.1$.

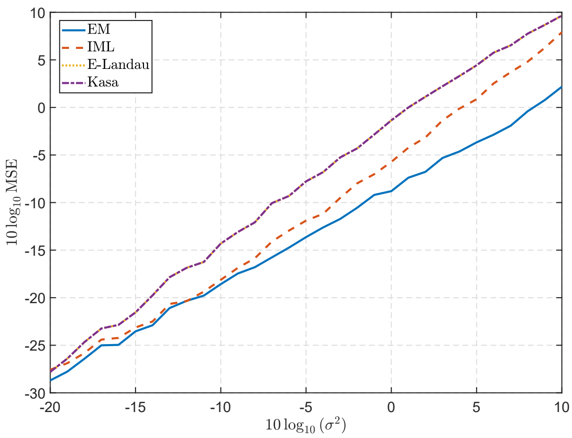


Figure 9: MSEs for $\hat{\boldsymbol{\theta}} = (r, \boldsymbol{c}^T)^T$ for E-Landau, Kasa, IML and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when data have a von Mises prior ($\kappa = 2$) in 2D.

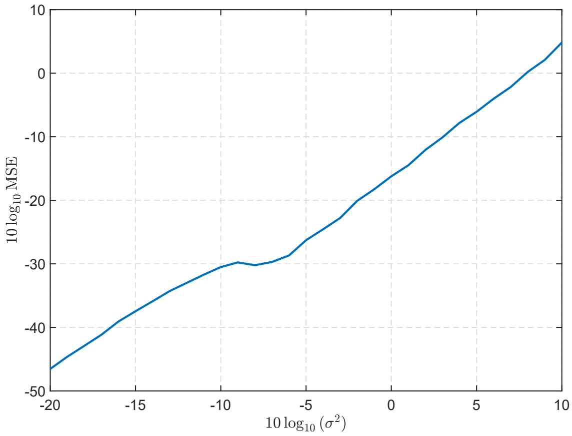
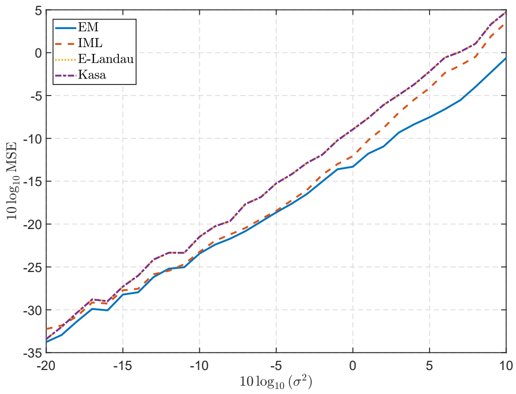
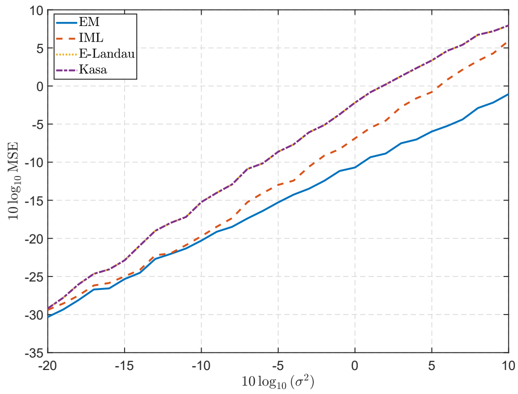


Figure 10: MSE for σ^2 EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when latent variables have an informative prior in 2D.



(a) MSE on the radius.



(b) MSE on the center.

Figure 11: MSEs for the radius (left) and the center (right) for E-Landau, Kasa, IML and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when data have a von Mises-Fisher prior ($\kappa = 2$) in 2D.

3.4 Informative prior in 3D

This section presents some simulation results obtained with a von Mises-Fisher prior with parameters $\kappa = 2$ and $\boldsymbol{\mu} = [\sin(\pi/4)\cos(\pi/3), \sin(\pi/4)\sin(\pi/3), \cos(\pi/4)]^T$. The same previous strategy was applied to obtain the corresponding MSEs for $\hat{\boldsymbol{\theta}}$, which are displayed in Fig. 12 and in Fig. 13 for σ^2 . We also display the MSEs for both the radius r and the center $\boldsymbol{c}$ in Figs. 14a and 14b. The results confirm the interest of the EM algorithm for hypersphere fitting using an informative prior defined as a von Mises-Fisher distribution.

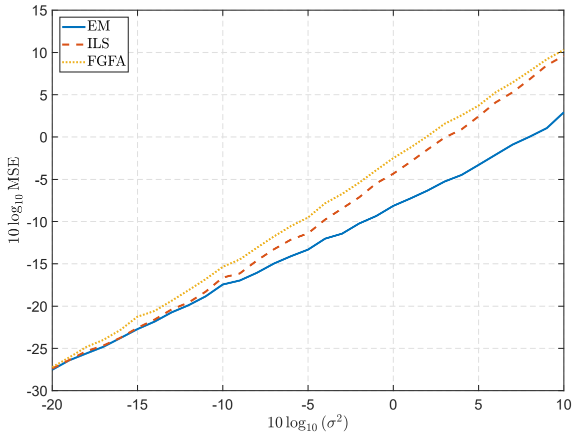


Figure 12: MSEs for $\hat{\boldsymbol{\theta}} = (r, \boldsymbol{c}^T)^T$ for FGFA, ILS and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when data have a von Mises prior ($\kappa = 2$) in 3D.

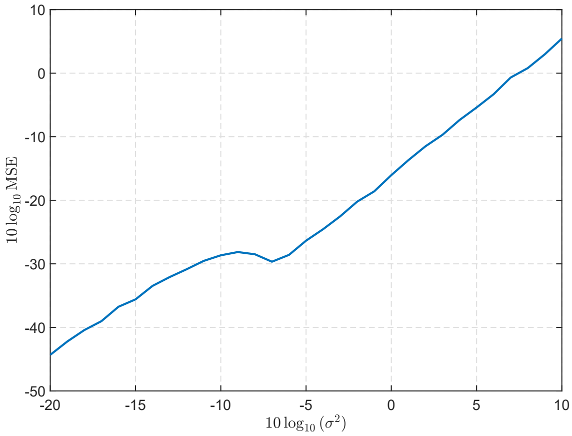
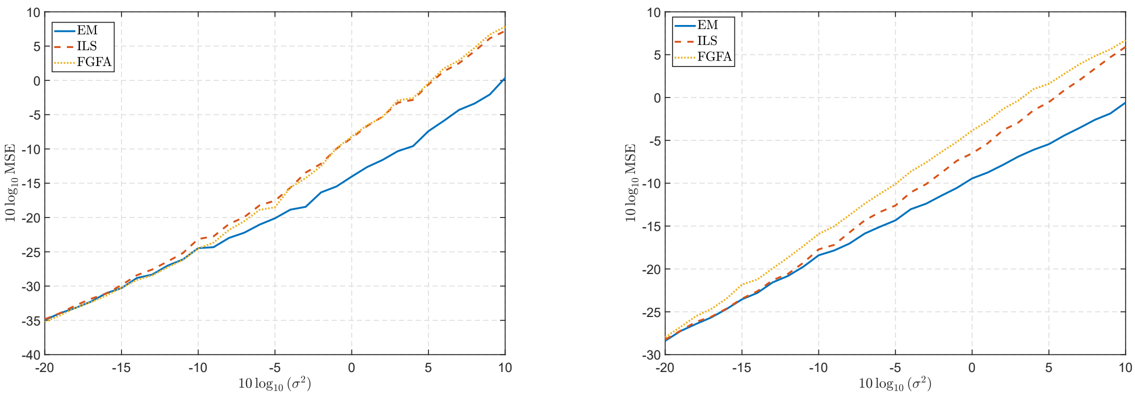


Figure 13: MSE for σ^2 EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when latent variables have an informative prior in 3D.



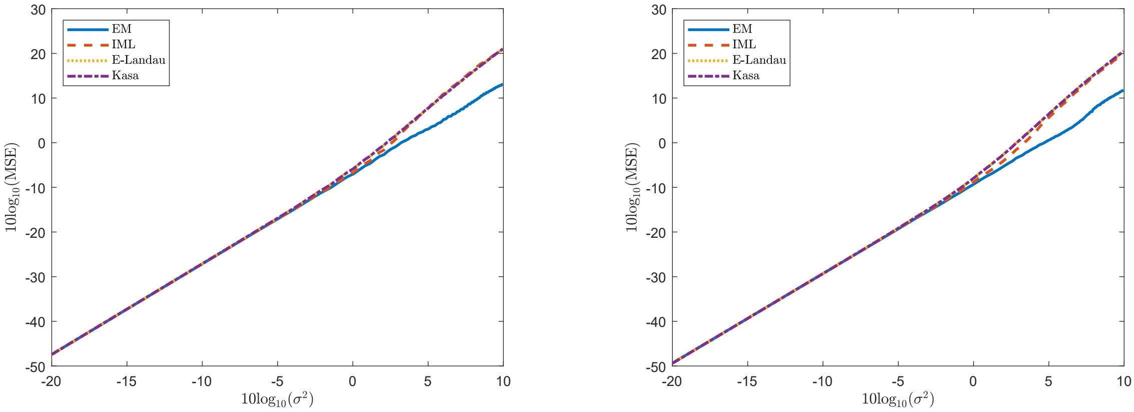
(a) MSE on the radius.

(b) MSE on the center.

Figure 14: MSEs for the radius (left) and the center (right) for FGFA, ILS and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when data have a von Mises-Fisher prior ($\kappa = 2$) in 3D.

3.5 Influence of n

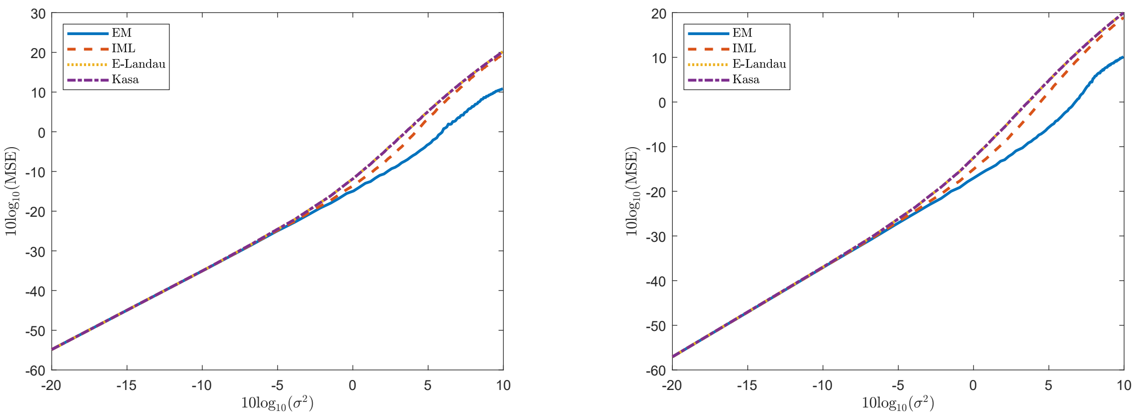
The previous results were obtained with $n = 100$ data. Here we will display the results on the first scenario (uniform prior) with $n = 50$ and $n = 30$. Fig. 15 shows the MSE on $\hat{\boldsymbol{\theta}}$, Fig. 16 shows the MSE on the radius, and Fig. 17 shows the MSE for the center. As one can see, the proposed method offers the best MSEs in any case.



(a) MSE on $\hat{\boldsymbol{\theta}}$ for $n = 30$ data.

(b) MSE on $\hat{\boldsymbol{\theta}}$ for $n = 50$ data.

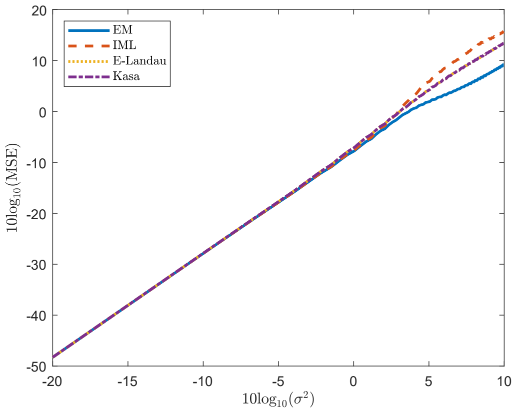
Figure 15: MSEs for $\hat{\boldsymbol{\theta}}$ for $n = 30$ (left) and $n = 50$ (right) for E-Landau, Kasa, IML and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when data have a uniform prior ($\kappa = 0$).



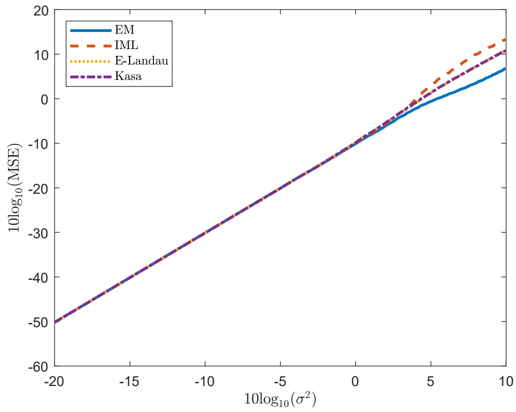
(a) MSE on the radius for $n = 30$ data.

(b) MSE on the radius for $n = 50$ data.

Figure 16: MSEs for the radius for $n = 30$ (left) and $n = 50$ (right) for E-Landau, Kasa, IML and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when data have a uniform prior ($\kappa = 0$).



(a) MSE on the center for $n = 30$ data.



(b) MSE on the center for $n = 50$ data.

Figure 17: MSEs for the center for $n = 30$ (left) and $n = 50$ (right) for E-Landau, Kasa, IML and EM (proposed method) versus noise power σ^2 (500 Monte Carlo runs) when data have a uniform prior ($\kappa = 0$).

References

[1] J. Lesouple, B. Pilastre, Y. Altmann, and J.-Y. Tourneret, "Hypersphere Fitting from Noisy Data Using an EM Algorithm," *Submitted to IEEE Signal Processing Letters*, 2021.

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

[3] D. K. Salkuyeh and F. P. A. Beik, "An Explicit Formula for the Inverse of Arrowhead and Doubly Arrow Matrices," *International Journal of Applied and Computational Mathematics*, vol. 4, no. 3, pp. 1–8, June 2018.