



**HAL**  
open science

## Multivariate expectile-based distribution: properties, Bayesian inference, and applications

Julyan Arbel, Stéphane Girard, Hien Duy Nguyen, Antoine Usseglio-Carleve

### ► To cite this version:

Julyan Arbel, Stéphane Girard, Hien Duy Nguyen, Antoine Usseglio-Carleve. Multivariate expectile-based distribution: properties, Bayesian inference, and applications. *Journal of Statistical Planning and Inference*, 2023, 225, pp.146-170. 10.1016/j.jspi.2022.12.001 . hal-03428827v2

**HAL Id: hal-03428827**

**<https://inria.hal.science/hal-03428827v2>**

Submitted on 2 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Multivariate expectile-based distribution: properties, Bayesian inference, and applications

Julyan Arbel, Stéphane Girard

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK  
38000 Grenoble, France

and

Hien Duy Nguyen

School of Mathematics and Physics, Univ. of Queensland  
St. Lucia 4067, Queensland Australia

and

Antoine Usseglio-Carleve

Avignon Univ., Laboratoire de Mathématiques d'Avignon EA 2151  
84000 Avignon, France

## Abstract

Expectiles form a family of risk measures that have recently gained interest over the more common value-at-risk or return levels, primarily due to their capability to be determined by the probabilities of tail values and magnitudes of realisations at once. However, a prevalent and ongoing challenge of expectile inference is the problem of uncertainty quantification, which is especially critical in sensitive applications, such as in medical, environmental, or engineering tasks. We address this issue by developing a novel distribution, termed the multivariate expectile-based distribution (MED), that possesses an expectile as a closed-form parameter. Desirable properties of the distribution, such as log-concavity, make it an excellent fitting distribution in multivariate applications. Maximum likelihood estimation and Bayesian inference algorithms are described. Simulated examples and applications to expectile and mode estimation illustrate the usefulness of the MED for uncertainty quantification.

*Keywords:* Multivariate expectiles; Bayesian inference; Maximum likelihood estimation; Multivariate distribution; Risk measure.

## 1 Introduction

Risk assessment mainly relies on drawing inference from quantiles, often referred to as value-at-risk in finance or return levels in environmental science. A shortcoming of quantiles is that they only use the information on the frequency of events and not on the actual magnitudes. This

is an issue since taking account of the magnitudes of extreme events may be crucial in many fields of application. One way to tackle this problem is to work with expectiles, introduced in [Newey and Powell \(1987\)](#) since, unlike quantiles, expectiles depend on both the probability of tail values and the magnitude of realisations ([Kuan et al., 2009](#)).

In univariate settings, expectiles are defined as the minimizers of an asymmetric quadratic loss, and as such, they benefit from a number of desirable properties. They are well-defined and unique whenever the underlying distribution has a finite first moment; see for instance [Abdous and Remillard \(1995\)](#). Furthermore, they induce the only coherent, law-invariant, and elicitable risk measure (see [Ziegel, 2016](#)), and therefore benefit from a natural backtesting methodology. Expectiles are therefore a sensible risk management tool, as a complement or an alternative to quantiles.

In the last five years, there has been an increasing interest in estimating expectiles, both at fixed probability levels (see *e.g.* [Holzmann and Klar, 2016](#); [Krättschmer and Zähle, 2017](#)) as well as at extreme levels ([Daouia et al., 2018](#); [Girard et al., 2021](#)). Another recent line of work garnering increasing interest in mathematical finance consists of extending the expectile risk measure to the multivariate setting: see, for example, [Maume-Deschamps et al. \(2017, 2018\)](#); [Herrmann et al. \(2018\)](#); [Daouia and Painsaveine \(2019\)](#); [Beck et al. \(2021\)](#), among others.

Here, we focus on the notion of multivariate geometric expectiles, as introduced by [Herrmann et al. \(2018\)](#), which is the analog of the multivariate geometric notion of quantiles proposed by [Chaudhuri \(1996\)](#).

## 1.1 Our contribution

We construct a multivariate distribution (Section 2), which is based on the loss function used for defining multivariate geometric expectiles ([Herrmann et al., 2018](#)). This distribution, termed the multivariate expectile-based distribution (MED), is parameterized by a location vector, scale matrix, asymmetry level, and direction. We argue that inference employing the MED offers a practical way to handle uncertainty quantification in expectile estimation. Our specific contributions are summarized below.

We first prove a number of appealing properties of the MED: it is log-concave, unimodal, and the unique mode is an expectile whose parameters (asymmetry level and direction) are specified as closed-form parameters of the distribution. In doing so, we propose an elementary proof of the strict convexity of the above-mentioned loss function. We establish that the MED family is closed under linear transformation, leading to a simple standardization of the distribution.

We provide bounds on the MED density (Section 3), showing that it is sub-Gaussian. Limiting cases of the asymmetry level are studied, proving that the MED comprises the multivariate Gaussian and some asymmetric Gaussian distributions as special cases. Various closed-form moments are provided, along with marginal and conditional distributions (Section 4).

The main benefit of our proposal lies in inference. We propose a maximum likelihood approach (Section 5.1) which relies on optimization algorithms in Riemannian geometries. Consistency of the maximum likelihood estimator is established under compactness assumptions on the parameter space. We devise a Bayesian algorithm based on Hamiltonian Monte Carlo (Section 5.2). We also describe our accompanying R package ([R Core Team, 2019](#)) containing statis-

tical functions related to the multivariate expectile-based distribution<sup>1</sup>, and provide a simulation study (Section 6).

We finally showcase three multivariate applications of the MED (Section 7): Bayesian estimation of expectiles, of a distribution mode, and illustration of discriminant analysis. We show that a Bayesian approach relying on the MED yields well-calibrated credible intervals, thus addressing the uncertainty quantification shortcoming in expectile estimation.

All proofs and some additional implementation and application results are deferred to the Appendix.

## 1.2 Notations

We denote asymptotic equivalence between positive real functions  $f$  and  $g$  by  $f(y) \underset{y \rightarrow a}{\approx} g(y)$ , meaning that  $f(y)/g(y) \rightarrow 1$  when  $y \rightarrow a$  for  $a \in [-\infty, \infty]$ . Vectors and matrices are denoted with boldface, for instance  $\mathbf{Z} = (Z_1, \dots, Z_d)$  denotes a vector in  $\mathbb{R}^d$ ,  $\mathbf{0}$  denotes the vector of zeros, and  $(\mathbf{u}_1, \dots, \mathbf{u}_d)$  denotes the canonical basis of  $\mathbb{R}^d$ . We write the last  $d - 1$  entries of  $\mathbf{Z}$  as  $\mathbf{Z}_{-1}$ , in the sense that  $\mathbf{Z} = (Z_1, \mathbf{Z}_{-1})$ . The sphere in dimension  $d$  is denoted by  $\mathcal{S}^{d-1}$ . The set of  $d \times d$  symmetric positive definite matrices is denoted by  $\text{SPD}(d)$ , the identity matrix is denoted by  $\mathbf{I}_d$ . Let  $\langle \cdot, \cdot \rangle$  denote the usual dot product and let  $\|\cdot\|$  be the associated Euclidean norm. For any matrix  $\mathbf{M} \in \text{SPD}(d)$ , we set  $\langle \cdot, \cdot \rangle_{\mathbf{M}} := \langle \cdot, \mathbf{M} \cdot \rangle$  and denote by  $\|\cdot\|_{\mathbf{M}}$  the associated norm. The determinant of some matrix  $\mathbf{M}$  is denoted by  $|\mathbf{M}|$ . We denote by  $\Gamma$  the gamma function and by  ${}_2F_1$  the hypergeometric function. The  $d$ -dimensional Gaussian density with mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$  is denoted by  $\phi_d(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ . The Gaussian, uniform, Wishart, and inverse-Wishart distributions are respectively denoted by  $\mathcal{N}$ ,  $\mathcal{U}$ ,  $\mathcal{W}$ , and  $\mathcal{IW}$ .

## 2 Multivariate expectile-based distribution

For any  $\rho \in [0, 1)$ , for any  $\boldsymbol{\Sigma} \in \text{SPD}(d)$  and  $\boldsymbol{\nu} \in \mathbb{R}^d$ , such that  $\|\boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}} = 1$ , consider the loss function defined for any  $\mathbf{y} \in \mathbb{R}^d$  by

$$\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{y}) = \|\mathbf{y}\|_{\boldsymbol{\Sigma}^{-1}} (\|\mathbf{y}\|_{\boldsymbol{\Sigma}^{-1}} + \rho \langle \mathbf{y}, \boldsymbol{\nu} \rangle_{\boldsymbol{\Sigma}^{-1}}).$$

In the particular case where  $\boldsymbol{\Sigma} = \mathbf{I}_d$ , this loss function has been used to define the multivariate geometric expectile (Herrmann et al., 2018) of a random vector  $\mathbf{X}$  as

$$\arg \min_{\mathbf{x} \in \mathbb{R}^d} E(\Lambda_{\rho, \boldsymbol{\nu}, \mathbf{I}_d}(\mathbf{X} - \mathbf{x})).$$

It has been shown that this optimisation problem is well-defined and benefits from a unique solution provided that the margins of  $\mathbf{X}$  have finite second order moments (Proposition 4.6 of Herrmann et al., 2018). We show in Proposition A.1 in Appendix that this condition can be relaxed to assuming only a *finite first-order moment*. This notion can be extended without difficulty to any  $\boldsymbol{\Sigma} \in \text{SPD}(d)$  by setting:

$$\mathbf{e}_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} E(\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{X} - \mathbf{x})). \quad (1)$$

<sup>1</sup>Available at <https://github.com/AntoineUC/MEDdist>.

Indeed, existence and unicity of the expectile defined in (1) stems from Proposition A.2 in Appendix on the strict convexity of  $\Lambda_{\rho, \nu, \Sigma}$ , which extends the result of Herrmann et al. (2018, Theorem 4.3) which deals with the particular case of  $\Lambda_{\rho, \nu, \mathbf{I}_d}$ . A proof relying on more elementary arguments than Herrmann et al. (2018)'s proof is provided in Appendix.

In view of (1), one clearly has  $e_{\rho, \nu, \Sigma}(\mathbf{X} + \mathbf{b}) = e_{\rho, \nu, \Sigma}(\mathbf{X}) + \mathbf{b}$ , for any  $\mathbf{b} \in \mathbb{R}^d$  and  $e_{\rho, \nu, \Sigma}(\sigma \mathbf{X}) = \sigma e_{\rho, \nu, \Sigma}(\mathbf{X})$  for any  $\sigma > 0$ , which are the equivariance properties of multivariate geometric expectiles with respect to translation and scale respectively. See Herrmann et al. (2018, Proposition 4.1 and Proposition 4.2) in the case where  $\Sigma = \mathbf{I}_d$ .

**Definition 1.** Let  $\rho \in [0, 1)$ ,  $\Sigma \in \text{SPD}(d)$  and  $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^d$  such that  $\|\boldsymbol{\nu}\|_{\Sigma^{-1}} = 1$ . The multivariate expectile-based distribution, denoted  $\text{MED}(\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho)$  or  $\text{MED}(\boldsymbol{\theta})$  where  $\boldsymbol{\theta}$  is a shorthand notation for  $\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}$ , and  $\rho$ , is defined on  $\mathbb{R}^d$  by its density

$$f_d(\mathbf{x}; \boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho) = \frac{c_d(\rho)}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \Lambda_{\rho, \nu, \Sigma}(\mathbf{x} - \boldsymbol{\mu})\right),$$

for any  $\mathbf{x} \in \mathbb{R}^d$ , where

$$c_d(\rho) = \sqrt{1 - \rho^2} \left( \frac{1 + \sqrt{1 - \rho^2}}{2} \right)^{(d-2)/2}. \quad (2)$$

See Lemma A.1 in the Appendix for a proof that  $f_d(\cdot; \boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho)$  integrates to one. A straightforward application of Proposition A.2 shows that the MED is log-concave, which implies in turn that it is unimodal, that its level sets are closed compact sets, and its marginal distributions remain log-concave (Brascamp and Lieb, 2002). Let us now show that the MED family is closed under linear transformations. Moreover, parameter  $\rho$  is invariant with respect to such transformation. The proof is straightforward.

**Proposition 1 (Linear transformation).** Let  $\mathbf{A}$  be a non-singular  $d \times d$  matrix and let  $\mathbf{b} \in \mathbb{R}^d$ . If  $\mathbf{X} \sim \text{MED}(\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho)$ , then  $\mathbf{AX} + \mathbf{b} \sim \text{MED}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^\top, \mathbf{A}\boldsymbol{\nu}, \rho)$ .

As an immediate consequence of Proposition 1, we have:

**Corollary 1 (Standardization).** Let  $\rho \in [0, 1)$ ,  $\Sigma \in \text{SPD}(d)$  and  $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^d$ , with  $\|\boldsymbol{\nu}\|_{\Sigma^{-1}} = 1$ . Consider a  $d \times d$  orthogonal matrix  $\mathbf{Q}$  such that  $\mathbf{Q}\Sigma^{-1/2}\boldsymbol{\nu} = \mathbf{u}_1$ . If  $\mathbf{Z} \sim \text{MED}(\mathbf{0}, \mathbf{I}_d, \mathbf{u}_1, \rho)$ , then  $\mathbf{X} := \Sigma^{1/2}\mathbf{Q}^\top\mathbf{Z} + \boldsymbol{\mu} \sim \text{MED}(\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho)$ .

In the sequel,  $\text{MED}(\mathbf{0}, \mathbf{I}_d, \mathbf{u}_1, \rho)$  is referred to as the standardized MED and, for the sake of presentation, it is denoted by  $\text{MED}(\rho)$ , and the associated density and objective function are denoted respectively by  $f_d(\cdot; \rho) := f_d(\cdot; \mathbf{0}, \mathbf{I}_d, \mathbf{u}_1, \rho)$  and  $\Lambda_\rho(\cdot) := \Lambda_{\mathbf{I}_d, \mathbf{u}_1, \rho}(\cdot)$ . Note that the standardization operation (or inverse standardization operation) requires obtaining a  $d \times d$  orthogonal matrix  $\mathbf{Q}$  such that  $\mathbf{Q}\Sigma^{-1/2}\boldsymbol{\nu} = \mathbf{u}_1$ , or such that  $\Sigma^{-1/2}\boldsymbol{\nu} = \mathbf{Q}^\top\mathbf{u}_1$ . The latter condition means that the first column of  $\mathbf{Q}^\top$  is equal to  $\Sigma^{-1/2}\boldsymbol{\nu}$ . Completing the last  $(d-1)$  columns of that matrix with the last  $(d-1)$  basis vectors  $\mathbf{u}_2, \dots, \mathbf{u}_d$  provides a  $d \times d$  matrix that can be transformed into a valid orthogonal matrix  $\mathbf{Q}^\top$  by using the Gram–Schmidt process.

**Proposition 2 (Particular cases).** For  $\mathbf{X} \sim \text{MED}(\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho)$ :

- (i) If  $\rho = 0$ , then  $\mathbf{X}$  is multivariate Gaussian  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .
- (ii) If  $\rho \rightarrow 1$ , let  $\boldsymbol{\Delta}_\rho = \text{diag}(\sqrt{1-\rho}, 1/\sqrt{2}, \dots, 1/\sqrt{2})$  and  $\mathbf{Z} = \mathbf{Q}\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$  the standardized MED; then  $\boldsymbol{\Delta}_\rho\mathbf{Z}$  converges in distribution to a half-standard Gaussian restricted to negative values of the first component.
- (iii) If  $d = 1$ , then  $X$  is the one-dimensional asymmetric Gaussian random variable of [Kato et al. \(2002\)](#).

We can interpret MED as a multivariate extension of the one-dimensional asymmetric Gaussian distribution (AGD) introduced in [Kato et al. \(2002\)](#). Tail properties of the one-dimensional AGD are established in [Chen and Huang \(2014\)](#), while this distribution is used in [Fu and Zhou \(2016\)](#) to build a model-based classifier. An alternative multivariate extension is applied in [He et al. \(2018\)](#), to the problem of image denoising. We also refer to [Majumdar and Paul \(2016\)](#) for an AGD-based construction of a stationary spatial noise process model, whose marginal distributions have certain (one-dimensional) expectiles equal to zero.

In the following,  $\rho \in [0, 1)$  is referred to as the asymmetry parameter, while  $\boldsymbol{\nu}$  is called the asymmetry direction. See [Figure 1](#) for an illustration of the possible shapes obtained by varying the parameter values of  $\rho$ ,  $\boldsymbol{\nu}$ , and  $\boldsymbol{\Sigma}$ . Parameter  $\boldsymbol{\mu}$  is simply a location parameter and its impact on the distribution is not illustrated (see [Proposition 3](#)). Scale matrix  $\boldsymbol{\Sigma}$  is a scale parameter that influences the shape of the distribution. The asymmetry direction  $\boldsymbol{\nu}$  clearly rotates the distribution, while the asymmetry parameter  $\rho$  stretches it in the direction opposite to  $\boldsymbol{\nu}$ . This counter-intuitive behaviour may be explained as follows. For a fixed (univariate) distribution, the expectile tends to  $+\infty$  as  $\rho \rightarrow 1$  (similarly to quantiles, see for instance [Daouia et al. \(2018\)](#)). Thus, to get a fixed expectile  $\boldsymbol{\mu}$ , the distribution should shift towards the negative real axis as  $\rho$  increases.

Characteristics of the location  $\boldsymbol{\mu}$  of the MED are investigated in the next result.

**Proposition 3** (Mode, expectile and expectation). *Let  $\mathbf{X} \sim \text{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$ . The distribution of  $\mathbf{X}$  is unimodal and satisfies:*

- (i)  $\text{mode}(\mathbf{X}) = \boldsymbol{\mu}$ ,
- (ii)  $\mathbf{e}_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{X}) = \boldsymbol{\mu}$ ,
- (iii)  $E(\mathbf{X}) = \boldsymbol{\mu} - \frac{\Gamma(\frac{d+3}{2})}{\sqrt{2}\Gamma(\frac{d+2}{2})} {}_2F_1\left(\frac{d+3}{4}, \frac{d+5}{4}; \frac{d+2}{2}; \rho^2\right) \rho\sqrt{1-\rho^2} \left(\frac{1+\sqrt{1-\rho^2}}{2}\right)^{\frac{d}{2}-1} \boldsymbol{\nu}$ .

The vector  $\boldsymbol{\mu}$  is both the mode and an expectile of the MED. In contrast, the expectation of the MED also depends on the asymmetry direction: it is located in the two-dimensional subspace spanned by  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$ . Specifying the above expectation in the one-dimensional case  $\text{MED}(\mu, \sigma, 1, \rho)$ , we recover the expectation of the one-dimensional asymmetric Gaussian of [Kato et al. \(2002\)](#):  $E\left(\frac{X-\mu}{\sigma}\right) = -\frac{c_1(\rho)\sqrt{2}}{\sqrt{\pi}} \frac{\rho}{1-\rho^2}$ . Second moment and skewness centered with respect to  $\mu$  are given by

$$E\left(\frac{(X-\mu)^2}{\sigma^2}\right) = \frac{c_1(\rho)}{2} \left[ \frac{1}{(1-\rho)^{3/2}} + \frac{1}{(1+\rho)^{3/2}} \right] \quad \text{and} \quad E\left(\frac{(X-\mu)^3}{\sigma^3}\right) = -\frac{c_1(\rho)\sqrt{2}}{\sqrt{\pi}} \frac{4\rho}{(1-\rho^2)^2},$$

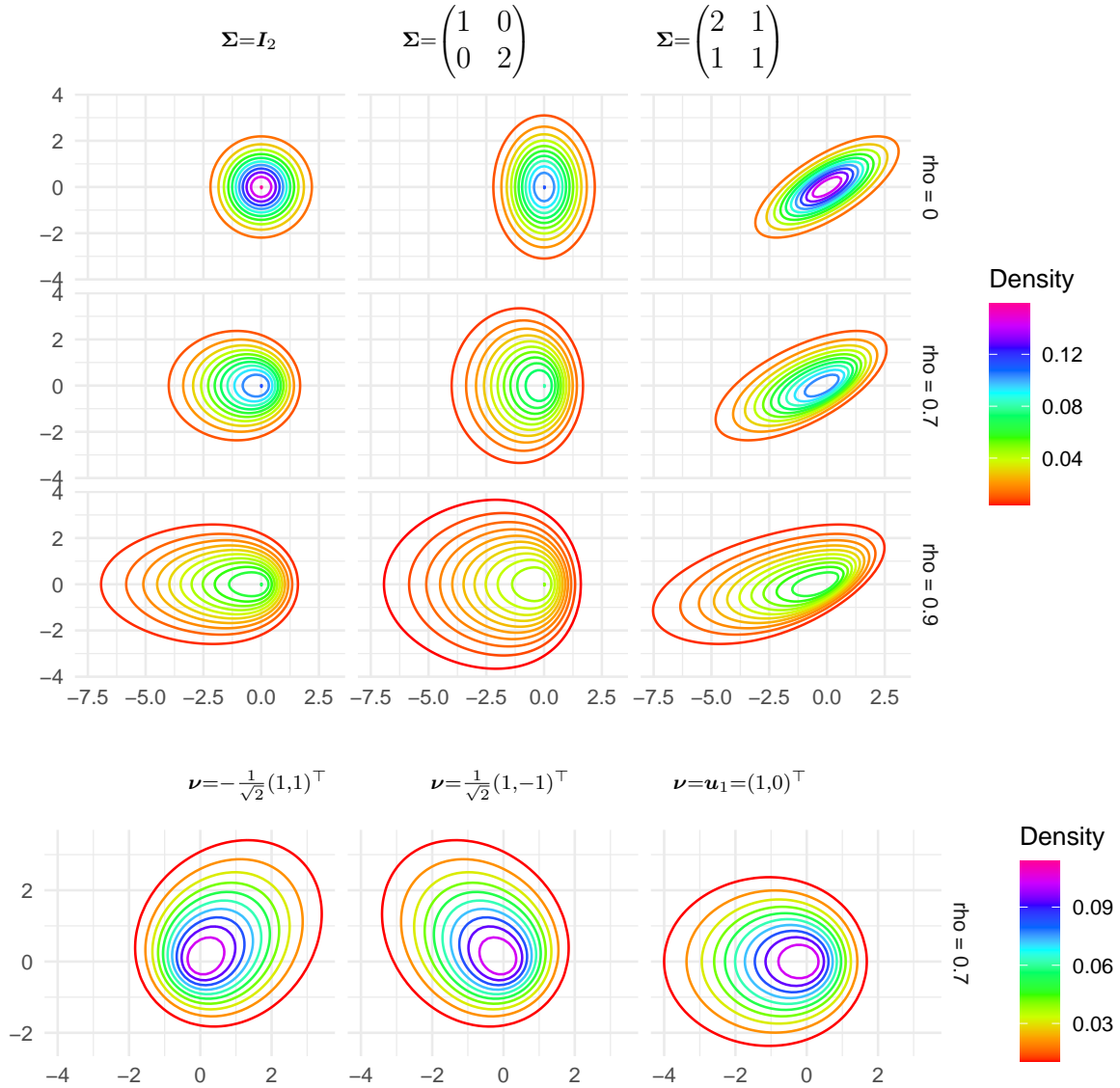


Figure 1: Contour plots of densities  $f_2(\cdot; \mathbf{0}, \Sigma, \boldsymbol{\nu}, \rho)$  in  $\mathbb{R}^2$ . Top panel: asymmetry direction  $\boldsymbol{\nu} = \mathbf{u}_1$  fixed, varying asymmetry parameter  $\rho \in \{0, 0.7, 0.9\}$  (row-wise) and matrices  $\Sigma$  (column-wise). Bottom panel: fixed  $\Sigma = \mathbf{I}_2$ ,  $\rho = 0.7$ , and varying direction  $\boldsymbol{\nu}$ .

respectively. It can be seen from Figure 7 in Appendix that asymmetry increases with  $\rho$ , corroborating the illustration of Figure 1. The role of the asymmetry parameter is further illustrated with the next proposition which studies the variations of expectiles with respect to variations of the asymmetry parameter for a given MED.

**Proposition 4** (Expectile with a different asymmetry parameter). *Let  $\rho_1, \rho_2 \in [0, 1)$ . The expectile of parameters  $(\rho_2, \mathbf{u}_1, \mathbf{I}_d)$  for the standardized  $\mathbf{Z} \sim \text{MED}(\rho_1)$  is supported on the first axis, and satisfies the following equivalences*

- $\langle \mathbf{e}_{\rho_2, \mathbf{u}_1, \mathbf{I}_d}(\mathbf{Z}), \mathbf{u}_1 \rangle > 0$  if and only if  $\rho_2 > \rho_1$ ;
- $\langle \mathbf{e}_{\rho_2, \mathbf{u}_1, \mathbf{I}_d}(\mathbf{Z}), \mathbf{u}_1 \rangle = 0$  (ie  $\mathbf{e}_{\rho_2, \mathbf{u}_1, \mathbf{I}_d}(\mathbf{Z}) = \mathbf{0}$ ) if and only if  $\rho_2 = \rho_1$ ;
- $\langle \mathbf{e}_{\rho_2, \mathbf{u}_1, \mathbf{I}_d}(\mathbf{Z}), \mathbf{u}_1 \rangle < 0$  if and only if  $\rho_2 < \rho_1$ .

### 3 Moments, sub-Gaussianity and bounds

The next Proposition provides both upper and lower bounds on the MED density based on the Gaussian density. This result will reveal itself to be useful to sample from the MED, using an acceptance-rejection method; see Section 5.3. To this end, let us consider, for any  $\rho \in [0, 1)$ ,

$$m_d(\rho) = c_d(\rho)(1 + \rho)^{-d/2} \text{ and } M_d(\rho) = c_d(\rho)(1 - \rho)^{-d/2}. \quad (3)$$

**Proposition 5** (Density bounds). *For any  $\mathbf{x} \in \mathbb{R}^d$ , one has*

$$m_d(\rho)\phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1 + \rho}\right) \leq f_d(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho) \leq M_d(\rho)\phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1 - \rho}\right).$$

As a consequence, one can easily derive bounds on moments of a projected MED random vector.

**Corollary 2** (Bounds on projected moments). *Let  $\mathbf{X} \sim \text{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$ . For any non-zero  $\boldsymbol{\beta} \in \mathbb{R}^d$  and  $p > 0$ , one has*

$$m_d(\rho) \left(\frac{2}{1 + \rho}\right)^{p/2} \leq \frac{\pi^{1/2} E(|\langle \boldsymbol{\beta}, \mathbf{X} - \boldsymbol{\mu} \rangle_{\boldsymbol{\Sigma}^{-1}}|^p)}{\Gamma((p + 1)/2)} \leq M_d(\rho) \left(\frac{2}{1 - \rho}\right)^{p/2}.$$

In the terminology of Vladimirova et al. (2020), MED random variables are thus sub-Weibull with optimal tail parameter 1/2, that is to say, they are sub-Gaussian, but no lighter than sub-Gaussian. The sub-Weibull property in a multivariate setting entails projections  $\langle \boldsymbol{\beta}, \mathbf{Z} \rangle_{\boldsymbol{\Sigma}^{-1}}$  for any non-zero vector  $\boldsymbol{\beta} \in \mathbb{R}^d$  to satisfy the (univariate) sub-Weibull property.

The next proposition provides the quadratic moment of the MED.

**Proposition 6** (Quadratic moment). *Let  $\mathbf{X} \sim \text{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$ . Then*

$$E[\|\mathbf{X} - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}}^2] = \frac{d}{\sqrt{1 - \rho^2}} + \frac{2\rho^2}{(1 - \rho^2)(1 + \sqrt{1 - \rho^2})}.$$



We always have  $E[\|\mathbf{X} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2] \geq d$ . If  $\rho = 0$ , we get  $E[\|\mathbf{X} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2] = d$ , the mean of a  $\chi^2$  distribution with  $d$  degrees of freedom. The next proposition provides the limiting behaviour of the expectation and quadratic moment when  $\rho \rightarrow 1$ .

**Proposition 7** (Asymptotics of moments). *Let  $\mathbf{X} \sim \text{MED}(\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho)$ . Then, when  $\rho \rightarrow 1$ , we have the following asymptotic equivalents*

$$(i) \quad E[\mathbf{X} - \boldsymbol{\mu}] \underset{\rho \rightarrow 1}{\approx} -\frac{\Gamma(\frac{d+3}{2})}{2^{d/2}\Gamma(\frac{d+3}{4})\Gamma(\frac{d+5}{4})} \frac{1}{\sqrt{1-\rho}} \boldsymbol{\nu},$$

$$(ii) \quad E[\|\mathbf{X} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2] \underset{\rho \rightarrow 1}{\approx} \frac{1}{1-\rho}.$$

## 4 Marginal & conditional distributions

In this section, we provide some marginal and conditional distributions of the MED that illustrate the distribution's shape in the asymmetry direction and in the hyperplane orthogonal to it. Let us first focus on the margins of the standard MED. As alluded to in the Introduction, the log-concavity of the MED implies log-concavity of any of its marginal distributions. Two distributions are considered: the univariate margin in the asymmetry direction  $\mathbf{u}_1$  and the  $(d-1)$ -dimensional margin in the orthogonal direction, denoted by  $\mathbf{u}_1^\perp$ . In the second case, the marginal distribution is elliptically contoured (Cambanis et al., 1981) or, more precisely, spherical when  $\Sigma = \mathbf{I}_d$ .

**Proposition 8** (Marginals). *Let  $\mathbf{Z} \sim \text{MED}(\rho)$  with  $d \geq 2$ , and write  $\mathbf{Z} = (Z_1, \mathbf{Z}_{-1})$ .*

(i) *The marginal density of  $Z_1$  (in the direction  $\mathbf{u}_1$ ), denoted by  $f_{Z_1}$ , satisfies for any  $z_1 \in \mathbb{R}$ :*

$$\frac{1}{(1+\rho)^{(d-1)/2}} \exp\left(-\frac{z_1^2}{2}(1+\rho)\right) \leq \frac{(2\pi)^{1/2}}{c_d(\rho)} f_{Z_1}(z_1) \leq \frac{1}{(1-\rho)^{(d-1)/2}} \exp\left(-\frac{z_1^2}{2}(1-\rho)\right),$$

and

$$f_{Z_1}(z_1) \underset{z_1 \rightarrow +\infty}{\approx} \frac{c_d(\rho)}{(2\pi)^{1/2}} \left(\frac{1}{1+\rho/2}\right)^{(d-1)/2} \exp\left(-\frac{z_1^2}{2}(1+\rho)\right), \quad (4)$$

$$f_{Z_1}(z_1) \underset{z_1 \rightarrow -\infty}{\approx} \frac{c_d(\rho)}{(2\pi)^{1/2}} \left(\frac{1}{1-\rho/2}\right)^{(d-1)/2} \exp\left(-\frac{z_1^2}{2}(1-\rho)\right). \quad (5)$$

(ii) *The marginal density of  $\mathbf{Z}_{-1}$ , denoted by  $f_{\mathbf{Z}_{-1}}$ , is an elliptically contoured distribution in the hyperplane orthogonal to  $\mathbf{u}_1$ . That is, there exists a density  $f_\perp$  on  $\mathbb{R}_+$  such that for any  $\mathbf{z}_{-1} \in \mathbb{R}^{d-1}$ :*

$$f_{\mathbf{Z}_{-1}}(\mathbf{z}_{-1}) = \tilde{f}(\|\mathbf{z}_{-1}\|), \quad (6)$$

and

$$\tilde{f}(z) \underset{z \rightarrow \infty}{\approx} \frac{1}{(2\pi\sigma^2(\rho))^{(d-1)/2}} \exp\left(-\frac{z^2}{2\sigma^2(\rho)}\right), \quad (7)$$

where  $\sigma^2(\rho) = \frac{2}{1+\sqrt{1-\rho^2}}$ .

Let us note that the bounds and the asymptotic equivalents in Proposition 8 (i) all coincide in the particular case where  $\rho = 0$ , yielding back the standard Gaussian density. When  $\rho \in (0, 1)$ , the gap between asymptotic equivalences (4) and (5) shows that the univariate marginal distribution of (i) has asymmetric tails. The asymptotic equivalence of (6) and (7) shows that the  $(d - 1)$ -dimensional marginal of (ii) reduces to a centered Gaussian density on  $\mathbb{R}^{d-1}$ .

We conclude this section with some one-dimensional and  $(d - 1)$ -dimensional conditional distributions.

**Proposition 9** (Conditionals). *Let  $\mathbf{Z} \sim \text{MED}(\rho)$  with  $d \geq 2$ .*

(i) *Let  $\mathbf{w} \in \mathbb{R}^d$ , such that  $\|\mathbf{w}\| = 1$ , and denote by  $P_{\mathbf{w}^\perp}(\mathbf{Z})$  the projection of  $\mathbf{Z}$  on the hyperplane orthogonal to  $\mathbf{w}$ . The one-dimensional conditional distribution of  $\langle \mathbf{w}, \mathbf{Z} \rangle$  given that  $P_{\mathbf{w}^\perp}(\mathbf{Z}) = \mathbf{0}$  coincides with the one-dimensional MED with  $\mu = 0$ ,  $\Sigma = 1$ , asymmetry parameter equal to  $\rho |\langle \mathbf{w}, \mathbf{u}_1 \rangle|$ , and asymmetry vector  $\text{sign}(w_1)$ , where  $w_1 = \langle \mathbf{w}, \mathbf{u}_1 \rangle$ ,*

$$\langle \mathbf{w}, \mathbf{Z} \rangle \mid P_{\mathbf{w}^\perp}(\mathbf{Z}) = \mathbf{0} \sim \text{MED}(0, 1, \rho |\langle \mathbf{w}, \mathbf{u}_1 \rangle|, \text{sign}(\langle \mathbf{w}, \mathbf{u}_1 \rangle)).$$

(ii) *Write  $\mathbf{Z} = (Z_1, \mathbf{Z}_{-1})$ . Then, the  $(d - 1)$ -dimensional conditional of  $\mathbf{Z}_{-1}$ , given  $Z_1 = 0$ , is a standard multivariate Gaussian distribution:*

$$\mathbf{Z}_{-1} \mid Z_1 = 0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d-1}).$$

Note that in (i) above, the considered distribution is one-dimensional, thus the asymmetry vector is either 1 or  $-1$ , which can be written as the sign of a scalar product. According to Proposition 2 (iii), the conditional distribution given in (i) also corresponds to the one-dimensional asymmetric Gaussian density introduced in Kato et al. (2002). When projecting in this way, the asymmetry parameter is reduced by a factor of  $|\langle \mathbf{w}, \mathbf{u}_1 \rangle| \leq 1$ . In particular, it remains unchanged if the projection direction  $\mathbf{w}$  is identical to  $\mathbf{u}_1$ , while it vanishes when  $\mathbf{w} \perp \mathbf{u}_1$ ; in the latter case, the distribution obtained coincides with the standard Gaussian distribution.

## 5 Statistical inference

In this section, we start by describing inference approaches for fitting the MED, either based on maximum likelihood or on a Bayesian model. We conclude by explaining how the statistical functions for MLE and Bayesian inference can be used in R via the **MEDdist** package.

### 5.1 Maximum likelihood estimation

Let  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho, \boldsymbol{\nu})$  denote the parameter vector containing the elements  $\boldsymbol{\mu} \in \mathbb{R}^d$ ,  $\boldsymbol{\Sigma} \in \text{SPD}(d)$ ,  $\rho \in [0, 1)$ , and  $\boldsymbol{\nu}$ . To produce an algorithm for computing the maximum likelihood estimator (MLE), we shall use the more convenient parameterization  $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho, \tilde{\boldsymbol{\nu}})$ , where  $\tilde{\boldsymbol{\nu}} = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\nu} \in \mathcal{S}^{d-1}$ . Given data  $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{\text{iid}}{\sim} \text{MED}(\tilde{\boldsymbol{\theta}})$ , the negative log-likelihood can be

rewritten as

$$\begin{aligned} \ell_n(\tilde{\boldsymbol{\theta}}) = & -n \log(c_d(\rho)) + \frac{n}{2} \log |\boldsymbol{\Sigma}| + \frac{1}{2} \sum_{i=1}^n \|\mathbf{X}_i - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}}^2 \\ & + \frac{\rho}{2} \sum_{i=1}^n \langle \mathbf{X}_i - \boldsymbol{\mu}, \tilde{\boldsymbol{v}} \rangle_{\boldsymbol{\Sigma}^{-1/2}} \|\mathbf{X}_i - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}} + C, \end{aligned}$$

where  $C$  is an additive constant that does not depend on  $\tilde{\boldsymbol{\theta}}$ .

We note that the elements of  $\tilde{\boldsymbol{\theta}}$  are constrained to four distinct types of Euclidean sets, namely,  $\boldsymbol{\mu} \in \mathbb{R}^d$  is unrestricted,  $\rho \in [0, 1)$  is restricted to a bounded interval,  $\boldsymbol{\Sigma} \in \text{SPD}(d)$  is restricted to the positive definite matrix manifold, and  $\tilde{\boldsymbol{v}} \in \mathcal{S}^{d-1}$  is restricted to the spherical manifold. Due to the non-separable dependence of the elements of  $\tilde{\boldsymbol{\theta}}$  and the nonlinear manner in which  $\ell_n$  is dependent on  $\tilde{\boldsymbol{\theta}}$  making a closed form solution unavailable, we have found it most convenient to compute the MLE

$$\tilde{\boldsymbol{\theta}}_n = \arg \min_{\tilde{\boldsymbol{\theta}}} \ell_n(\tilde{\boldsymbol{\theta}})$$

via a coordinate descent algorithm, in order to simplify the optimization process by only considering one of the four domain constraints at a time. Let  $(\tilde{\boldsymbol{\theta}}^{(t)})_{t=1}^{\infty}$  be a sequence of coordinate descent algorithm iterates, where  $\tilde{\boldsymbol{\theta}}^{(t)} = (\boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)}, \rho^{(t)}, \tilde{\boldsymbol{v}}^{(t)})$  denotes the  $t$ th iteration. With  $\text{rem}(a, b)$  denoting the remainder of  $a \in \mathbb{N}$  divided by  $b \in \mathbb{N}$ , and initializing at some  $\tilde{\boldsymbol{\theta}}^{(0)}$ , the cyclic coordinate descent algorithm proceeds via the following rules, for each  $t \in \mathbb{N}$ :

- If  $\text{rem}(t, 4) = 0$ , set  $\tilde{\boldsymbol{\theta}}^{(t)} = (\boldsymbol{\mu}_*, \boldsymbol{\Sigma}^{(t-1)}, \rho^{(t-1)}, \tilde{\boldsymbol{v}}^{(t-1)})$ , where

$$\boldsymbol{\mu}_* = \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^d} \ell_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{(t-1)}, \rho^{(t-1)}, \tilde{\boldsymbol{v}}^{(t-1)}), \quad (8)$$

- If  $\text{rem}(t, 4) = 1$ , set  $\tilde{\boldsymbol{\theta}}^{(t)} = (\boldsymbol{\mu}^{(t-1)}, \boldsymbol{\Sigma}_*, \rho^{(t-1)}, \tilde{\boldsymbol{v}}^{(t-1)})$ , where

$$\boldsymbol{\Sigma}_* = \arg \min_{\boldsymbol{\Sigma} \in \text{SPD}(d)} \ell_n(\boldsymbol{\mu}^{(t-1)}, \boldsymbol{\Sigma}, \rho^{(t-1)}, \tilde{\boldsymbol{v}}^{(t-1)}), \quad (9)$$

- If  $\text{rem}(t, 4) = 2$ , set  $\tilde{\boldsymbol{\theta}}^{(t)} = (\boldsymbol{\mu}^{(t-1)}, \boldsymbol{\Sigma}^{(t-1)}, \rho_*, \tilde{\boldsymbol{v}}^{(t-1)})$ , where

$$\rho_* = \arg \min_{\rho \in [0, \infty)} \ell_n(\boldsymbol{\mu}^{(t-1)}, \boldsymbol{\Sigma}^{(t-1)}, \rho, \tilde{\boldsymbol{v}}^{(t-1)}), \text{ and} \quad (10)$$

- If  $\text{rem}(t, 4) = 3$ , set  $\tilde{\boldsymbol{\theta}}^{(t)} = (\boldsymbol{\mu}^{(t-1)}, \boldsymbol{\Sigma}^{(t-1)}, \rho^{(t-1)}, \tilde{\boldsymbol{v}}_*)$ , where

$$\tilde{\boldsymbol{v}}_* = \arg \min_{\tilde{\boldsymbol{v}} \in \mathcal{S}^{d-1}} \ell_n(\boldsymbol{\mu}^{(t-1)}, \boldsymbol{\Sigma}^{(t-1)}, \rho^{(t-1)}, \tilde{\boldsymbol{v}}). \quad (11)$$

### 5.1.1 Numerical considerations

In practice, the coordinate descent algorithm is run for a sufficiently large number of iterations  $T \in \mathbb{N}$ , at which point the algorithm is considered to be convergent. We then use the value  $\tilde{\boldsymbol{\theta}}^{(T)}$  as the MLE.

The computation of (8) and (10) can be performed via typical optimization routines due to their mundane domains. However, the computation of (9) and (11) require the enforcement of matrix manifold restrictions and thus specialized optimization routines. Here, we make use of the Riemannian optimization framework of [Absil et al. \(2009\)](#), which is implemented via the libraries of [Huang et al. \(2018\)](#) and [Martin et al. \(2020\)](#). The availability of such Riemannian manifold optimization tools is also a motivating factor for our use of a coordinate descent algorithm for MLE.

By definition, the coordinate descent algorithm yields a decreasing sequence of objective values  $(\ell_n(\tilde{\boldsymbol{\theta}}^{(t)}))_{t=1}^{\infty}$ . However, without further assumptions, it is not possible to make stronger statements regarding the behavior of the sequences  $(\tilde{\boldsymbol{\theta}}^{(t)})_{t=1}^{\infty}$  and  $(\ell_n(\tilde{\boldsymbol{\theta}}^{(t)}))_{t=1}^{\infty}$ . Furthermore, due to the spherical matrix manifold restriction on  $\tilde{\boldsymbol{\nu}}$ , the typical directional derivative-based arguments for convergence of [Tseng \(2001\)](#) cannot be applied and thus we cannot guarantee convergence via application of generic optimization routines. However, if each of the sub-problems (8)–(11) are solved using trust region-based algorithms, then the recent theory of [Boumal et al. \(2018\)](#) and [Tian et al. \(2019\)](#) can be applied to obtain global convergence guarantees if a random sub-problem is solved at iteration  $t$  instead of cycling through the sub-problems (8)–(11), as we have proposed. However, empirically, we have found convergence to be robust to the choice of algorithms used to compute (8)–(11) and to the use of cyclic coordinate descent.

Even for simple choices of algorithms to use for solving the sub-problems (8)–(11), the combinatorial calculation of the computational complexity is prohibitively difficult, as can be seen via the example calculations of [Boumal et al. \(2018\)](#) and [Huang and Gallivan \(2022\)](#). However, at the very least, when implementing the coordinate descent algorithm using trust region-based solvers, we have the result that each of the sub-problems converge superlinearly [Huang et al. \(2015\)](#) and that the norm of the tangent vector of the sub-problem iterate after  $O(1/\epsilon_{\text{tan}}^2)$  iterations can be bounded from above by a constant  $\epsilon_{\text{tan}} > 0$ . Here the tangent vector in Riemannian geometry is analogous to the usual gradient in Euclidean geometry, and a tangent vector value of size zero implies a stationary point, like the typical Fermat condition.

To make more quantitative statements, we conduct a simulation study where we implement coordinate descent with sub-problems (8) and (10) solved using the `optim` function in R, where we use the Broyden–Fletcher–Goldfarb–Shanno algorithm in the case of (8) and the Dekker–Brent algorithm in the case of (10) (see [Quarteroni et al., 2010](#), Chs. 6 and 7). Sub-problems (9) and (11) are solved using the `manifold.optim` function from the `ManifoldOptim` package ([Martin et al. 2020](#)) in R, implementing the Riemannian symmetric rank one trust region method of [Huang et al. \(2015\)](#).

In our study, we simulate  $n \in \{50, 100, 200, 500, 1000\}$  observations  $\mathbf{X}_1, \dots, \mathbf{X}_n$  from a  $d \in \{2, 4, 6, 8, 10\}$  dimensional normal distribution with randomized mean vectors and covariance matrices from normal and Wishart distributions, respectively. Here, we do not simulate from the broader class of MED distributions due to the efficiency characteristics of our current sampling algorithm for larger values of  $d$ ; see [Figure 3](#). For each pair  $(n, d)$  we simulate 20

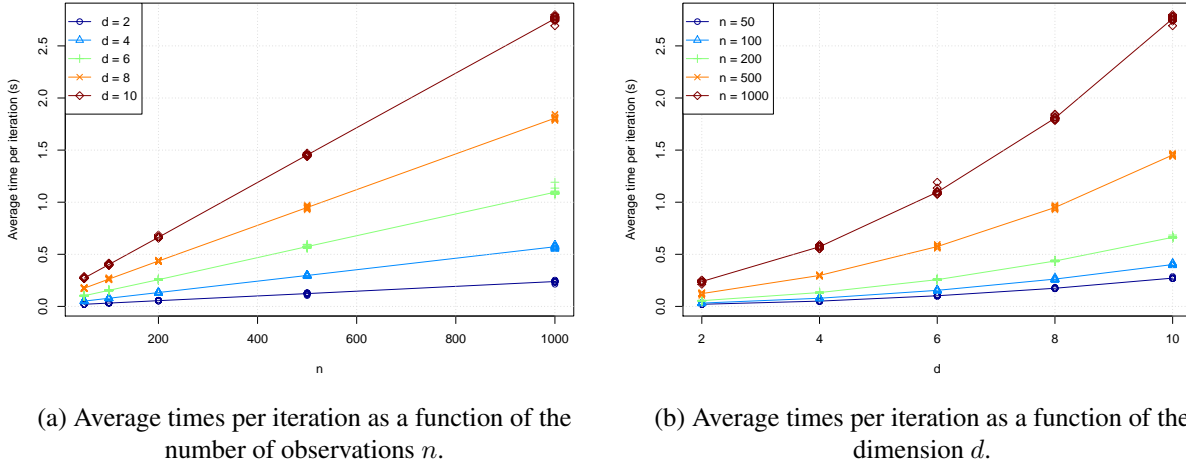


Figure 2: Computational times per iteration of the coordinate descent algorithm. Points represent results from individual replicates of each case pair  $(n, d)$ .

replicate data sets, from which we then run  $T = 80$  iterations of our specified coordinate descent algorithm (so that each of the sub-problems (8)–(11) is solved 20 times). The average times per iteration depending on  $n$  and  $d$  on a PC (Intel i7-9700 3.00 GHz CPU, 32 GB of RAM, running Windows 10 and R 4.1.0) are plotted in Figure 2. The computational cost of the coordinate descent algorithm appears to be linear in  $n$  and quadratic in  $d$ . Although the linear scaling in  $n$  is expected, we note that the quadratic scaling in  $d$  is surprising since the computation of  $\ell_n$  requires matrix multiplications and inversions, which are typically operations of complexity order  $O(d^3)$ . As such, we advise that our empirical results regarding the complexity with respect to  $d$  should be used cautiously and should not be extrapolated beyond the bounds of the simulation experiment.

### 5.1.2 Consistency of the MLE

Given that MLE  $\tilde{\theta}_n$  can be obtained, we require some additional assumptions to guarantee its consistency. A minimal assumption is that the set of minimizers

$$\tilde{\Theta}_0 = \left\{ \tilde{\theta}_0 : E \log f_d(\mathbf{X}; \tilde{\theta}_0) \leq \inf_{\tilde{\theta}} E \log f_d(\mathbf{X}; \tilde{\theta}) \right\}$$

is contained in a compact set  $\bar{\Theta}$ . Given our domain specifications of  $\tilde{\theta}$ , and following the approach Ritter (2014, Sec. B.6.2), we can always choose such a compact set to be of the form

$$\bar{\Theta} = \left\{ \tilde{\theta} : \rho \in [0, 1 - \gamma], \boldsymbol{\mu} \in [-m, m]^d, v_1 \leq \lambda_{(1)}(\boldsymbol{\Sigma}) \leq \lambda_{(d)}(\boldsymbol{\Sigma}) \leq v_d, \tilde{\mathbf{v}} \in \mathcal{S}^{d-1} \right\}, \quad (12)$$

where  $\gamma \in (0, 1)$  and  $v_1 > 0$  are sufficiently small and  $m, v_d > 0$  are sufficiently large. Here,  $\lambda_{(1)}(\boldsymbol{\Sigma})$  and  $\lambda_{(d)}(\boldsymbol{\Sigma})$  are the smallest and largest eigenvalues of  $\boldsymbol{\Sigma}$ , respectively. The constraint does not need to be enforced explicitly, but for numerical stability, we found it useful to solve subproblem (10) using the domain  $[0, 1 - \gamma]$ , with  $\gamma = 10^{-6}$ . The next result follows by an application of Theorem 5.3 in Shapiro et al. (2021). See Appendix A.1 regarding the proof.

**Proposition 10.** *If  $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{\text{iid}}{\sim} \text{MED}(\tilde{\boldsymbol{\theta}})$  and  $\tilde{\Theta}_0$  is contained in a compact set  $\bar{\Theta}$ , then  $\tilde{\boldsymbol{\theta}}_n$  is consistent in the sense that, as  $n \rightarrow \infty$ ,*

$$\inf_{\tilde{\boldsymbol{\theta}} \in \tilde{\Theta}_0} \|\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}\| \xrightarrow{a.s.} 0.$$

## 5.2 Bayesian inference

An interesting feature of Bayesian inference is that some expert knowledge can be incorporated in the prior distribution and that the posterior distribution can be used to provide uncertainty estimates with credible bands. This is complementing the previous section as well as the work of [Herrmann et al. \(2018\)](#).

We adopt the same parameterization as that of the previous section, which replaces  $\boldsymbol{\nu}$  in  $\boldsymbol{\theta}$ , by  $\tilde{\boldsymbol{\nu}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu}$ . Instead of the covariance matrix, it is more efficient from an implementation viewpoint to work with the precision matrix  $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} \in \text{SPD}(d)$ . A simple strategy consists in assigning independent priors on  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Omega}$ ,  $\rho$  and  $\tilde{\boldsymbol{\nu}}$ . First, a normal-inverse-Wishart prior on  $(\boldsymbol{\mu}, \mathbf{K})$  parameters, where  $\mathbf{K}$  is the covariance matrix parameter of the normal prior for  $\boldsymbol{\mu}$ :

$$\begin{aligned} \boldsymbol{\mu} \mid \mathbf{K} &\sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{K}), \\ \mathbf{K} &\sim \mathcal{IW}(\nu_K, \mathbf{S}_K), \end{aligned}$$

where  $\boldsymbol{\mu}_0 \in \mathbb{R}^d$ ,  $\nu_K \in (d-1, \infty)$ , and  $\mathbf{S}_K \in \text{SPD}(d)$  can be fixed or be given hyperpriors. Second, a Wishart prior is placed on  $\boldsymbol{\Omega}$ :

$$\boldsymbol{\Omega} \sim \mathcal{W}(\nu_\Omega, \mathbf{S}_\Omega),$$

where  $\nu_\Omega \in (d-1, \infty)$  and  $\mathbf{S}_\Omega$  is symmetric and positive definite, and can be fixed or can be given hyperpriors. And finally, uniform priors are placed on  $\rho$  and  $\tilde{\boldsymbol{\nu}}$ :

$$\rho \sim \mathcal{U}(0, 1), \quad \tilde{\boldsymbol{\nu}} \sim \mathcal{U}(\mathcal{S}^{d-1}).$$

Note that in the presence of prior information, more informative priors than the ones presented here could be considered in order to encode such prior knowledge.

Finally, the data generating process for observations  $\mathbf{X}_1, \dots, \mathbf{X}_n$  is the MED:

$$\mathbf{X}_1, \dots, \mathbf{X}_n \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \rho, \tilde{\boldsymbol{\nu}}, \boldsymbol{\varphi} \stackrel{\text{iid}}{\sim} \text{MED}(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}, \boldsymbol{\Omega}^{-1/2}\tilde{\boldsymbol{\nu}}, \rho),$$

where  $\boldsymbol{\varphi}$  represents potential hyperparameters described above. We have implemented the model in the Stan probabilistic programming language ([Carpenter et al., 2017](#)), which uses Hamiltonian Monte Carlo. See [Appendix B.2](#) for implementation details.

## 5.3 The MEDdist R package

We describe an R package termed **MEDdist**<sup>2</sup> containing the MLE and Bayesian procedures introduced above. Density, distribution function, quantile function and random generation for the MED are also documented.

<sup>2</sup>Available at <https://github.com/AntoineUC/MEDdist>.

- `fitmed`: estimates the parameters  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$ ,  $\rho$  and  $\tilde{\boldsymbol{\nu}}$  from a sample (input) using the MLE procedure of Section 5.1. The starting parameters  $\boldsymbol{\mu}^{(0)}$ ,  $\boldsymbol{\Sigma}^{(0)}$ ,  $\rho^{(0)}$  and  $\tilde{\boldsymbol{\nu}}^{(0)}$  may be chosen as inputs, as well as the maximum number of iterations `maxiter`.
- `dmed`: returns the value of the p.d.f. of the MED with parameters  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$ ,  $\rho$  and  $\boldsymbol{\nu}$  (inputs) at point  $\boldsymbol{x} \in \mathbb{R}^d$  (input) using Equation (2).
- `pmed`: returns the value of the c.d.f. (only in dimension  $d = 1$ ) of the MED with parameters  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$ ,  $\rho$  and  $\boldsymbol{\nu}$  (inputs) at point  $x \in \mathbb{R}$  (input).
- `qmed`: returns the quantiles (only in dimension  $d = 1$ ) of the MED with parameters  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$ ,  $\rho$  and  $\boldsymbol{\nu}$  (inputs) of level  $\tau \in (0, 1)$  (input).
- `rmed`: returns a sample of size  $n$  (input) from the MED with parameters  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$ ,  $\rho$  and  $\boldsymbol{\nu}$  (inputs), see below.
- `bayexpectile` and `baymode`: return posterior mean estimators for multivariate expectile and mode, respectively, under a sample from the MED. These functions, containing Stan codes including prior specification, are provided in Appendix B.2, and detailed in Section 7.

Sampling from the distribution is done by using the bounds of Section 3. Indeed, the upper bound of Equation 3 and Proposition 5 leads to a simple rejection algorithm, where the proposal distribution is multivariate Gaussian. The acceptance probability,  $M_d(\rho)^{-1}$ , deteriorates with  $\rho$  and  $d$  as follows (see also Figure 3):

- polynomially in  $\rho$  as  $\rho \rightarrow 1$  (for fixed  $d$ ):  $(M_d(\rho))^{-1} \underset{\rho \rightarrow 1}{\approx} a(1 - \rho)^{\frac{d-1}{2}}$ , for some constant  $a$ ;
- exponentially in  $d$  as  $d \rightarrow \infty$  (for fixed  $\rho$ ):  $(M_d(\rho))^{-1} \underset{d \rightarrow \infty}{\approx} A(\rho) (B(\rho))^{d/2}$ , for some constants (in  $\rho$ )  $A(\rho)$  and  $B(\rho)$ , where  $B(\rho) < 1$ .

Some potential refinements for the proposal distribution taking the form of mixtures of truncated multivariate Gaussian distributions are not investigated here.

## 6 MLE simulation study

To assess the performance of our procedure for computing the MLE, as well as the quality of the MLEs themselves, we have conducted a set of simulation studies in R. Within these studies, we have implemented the coordinate descent algorithm that is described in Section 5.1 using the same choices of sub-problem solvers as in Section 5.1.1.

For our simulations we set  $d = 2$  and simulate  $n \in \{50, 100, 200, 500, 1000\}$  observations  $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$  according to a MED( $\tilde{\boldsymbol{\theta}}$ ) distribution with  $\boldsymbol{\mu}^\top = (0, 0)$ ,  $\boldsymbol{\Sigma} \in \{\boldsymbol{I}_2, (\boldsymbol{I}_2 + \boldsymbol{J}_2)/2\}$ ,  $\rho \in \{0.2, 0.5, 0.8\}$ , and  $\tilde{\boldsymbol{\nu}}^\top = (1, 1)/\sqrt{2}$ . for each choice of  $n$  and  $\tilde{\boldsymbol{\theta}}$ , we perform 100 simulations where we compute the MLE  $\tilde{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}^{(T)}$ , for  $T = 400$  iterations. Here,  $\boldsymbol{J}_2$  is a  $2 \times 2$  matrix of 1s. To aggregate the performance over the 100 simulations, we first transform  $\tilde{\boldsymbol{\nu}}_n$  and  $\tilde{\boldsymbol{\nu}}$  to the more interpretable polar-angle forms  $\phi_n = \arctan(\tilde{\nu}_2/\tilde{\nu}_1)$  and  $\phi = \arctan(\nu_2/\nu_1) = \pi/4$ , respectively.

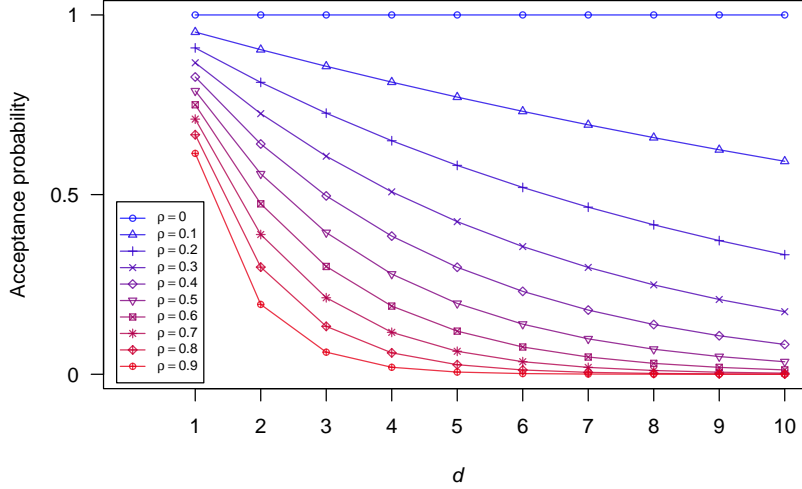


Figure 3: Acceptance probability of the rejection sampler `rmed` from the MED distribution, for varying values of asymmetry parameter  $\rho \in \{0, 0.1, \dots, 0.9\}$  and dimension  $1 \leq d \leq 10$ .

For each  $\mu_{1n}, \mu_{2n}, \Sigma_{11,n}, \Sigma_{12,n}, \Sigma_{22,n}, \rho_n$ , we report the absolute mean differences (absolute bias) between these estimators and their respective generative parameters  $\mu_1, \mu_2, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}, \rho$ , as well as the standard deviations, over the 100 simulations of  $\tilde{\theta}_n$ . To compare the quality of the sample of estimates of  $\phi_n$  for  $\phi$ , we report the absolute circular mean differences and circular standard deviations, as per [Mardia and Jupp \(2000, Ch. 2\)](#). The results for the  $\Sigma = \mathbf{I}_2$  and  $\Sigma = (\mathbf{I}_2 + \mathbf{J}_2)/2$  simulation scenarios are presented in [Tables 1 and 2](#), respectively.

From [Tables 1 and 2](#), we observe that in every case, other than when  $\rho = 0.2$  and  $n \in \{50, 100, 200\}$ , the biases of the MLEs are all within a standard deviation of the samples. Furthermore, we observe that in all cases of  $\rho$  and  $\Sigma$ , the sample standard deviations of the estimate samples are decreasing in  $n$ . These two initial observations provide empirical verification of the statistical consistency and numerical accuracy of the MLE estimator and the coordinate descent algorithm for its computation.

However, the estimates of  $\rho$  appear to exhibit a noticeable bias when  $\rho = 0.2$ . By inspecting the results in the  $\rho = 0.5$  and  $\rho = 0.8$  cases, it appears that the biases and standard deviations of the estimates of  $\rho$  are decreasing as  $\rho$  increases. Furthermore, we see that although  $\rho$  does not seem to affect the bias of other parameters, it does appear that the standard deviations of the  $\phi$  estimates are larger when  $\rho$  is small, and decreasing as  $\rho$  increases. Notably, there does not appear to be any other effects due to the generative value of  $\rho$  with respect to the biases and standard deviations of the samples of estimates of  $\mu$  and  $\Sigma$ .



Table 1: Absolute biases (standard font) and standard deviations (italic) of the MLE estimates across 100 simulations for different values of  $n$  and  $\rho$ , and  $\Sigma = \mathbf{I}_2$ .

$\Sigma$	$\rho$	$n$	$\mu_{1n}$	$\mu_{2n}$	$\Sigma_{11,n}$	$\Sigma_{12,n}$	$\Sigma_{22,n}$	$\rho_n$	$\phi_n$	
$\mathbf{I}_2$	0.2	50	0.021	0.096	0.236	0.002	0.183	0.322	0.410	
			<i>0.480</i>	<i>0.410</i>	<i>0.222</i>	<i>0.145</i>	<i>0.224</i>	<i>0.217</i>	<i>1.841</i>	
		100	0.034	0.048	0.156	0.012	0.087	0.209	0.317	
			<i>0.313</i>	<i>0.295</i>	<i>0.142</i>	<i>0.109</i>	<i>0.160</i>	<i>0.17</i>	<i>1.315</i>	
		200	0.022	0.011	0.071	0.012	0.047	0.152	0.069	
			<i>0.238</i>	<i>0.248</i>	<i>0.119</i>	<i>0.076</i>	<i>0.112</i>	<i>0.152</i>	<i>1.207</i>	
	500	0.003	0.008	0.016	0.002	0.021	0.044	0.050		
		<i>0.123</i>	<i>0.123</i>	<i>0.073</i>	<i>0.048</i>	<i>0.074</i>	<i>0.101</i>	<i>0.702</i>		
	1000	0.002	0.003	0.007	0.003	0.005	0.016	0.053		
		<i>0.104</i>	<i>0.108</i>	<i>0.043</i>	<i>0.033</i>	<i>0.049</i>	<i>0.096</i>	<i>0.659</i>		
	$\mathbf{I}_2$	0.5	50	0.016	0.006	0.131	0.010	0.168	0.121	0.004
				<i>0.363</i>	<i>0.440</i>	<i>0.292</i>	<i>0.186</i>	<i>0.270</i>	<i>0.212</i>	<i>0.889</i>
100			0.029	0.044	0.090	0.019	0.078	0.055	0.055	
			<i>0.273</i>	<i>0.350</i>	<i>0.201</i>	<i>0.116</i>	<i>0.225</i>	<i>0.166</i>	<i>0.683</i>	
200			0.006	0.037	0.032	0.006	0.027	0.026	0.033	
			<i>0.226</i>	<i>0.220</i>	<i>0.156</i>	<i>0.096</i>	<i>0.148</i>	<i>0.137</i>	<i>0.450</i>	
500		0.020	0.002	0.003	0.002	0.015	0.003	0.011		
		<i>0.139</i>	<i>0.133</i>	<i>0.091</i>	<i>0.054</i>	<i>0.101</i>	<i>0.094</i>	<i>0.255</i>		
1000		0.007	0.011	0.003	0.002	0.001	0.003	0.000		
		<i>0.095</i>	<i>0.087</i>	<i>0.077</i>	<i>0.042</i>	<i>0.069</i>	<i>0.068</i>	<i>0.175</i>		
$\mathbf{I}_2$		0.8	50	0.045	0.080	0.051	0.018	0.056	0.012	0.023
				<i>0.345</i>	<i>0.490</i>	<i>0.292</i>	<i>0.212</i>	<i>0.356</i>	<i>0.115</i>	<i>0.431</i>
	100		0.039	0.015	0.003	0.009	0.039	0.006	0.007	
			<i>0.262</i>	<i>0.299</i>	<i>0.250</i>	<i>0.179</i>	<i>0.253</i>	<i>0.109</i>	<i>0.263</i>	
	200		0.051	0.002	0.043	0.012	0.030	0.007	0.041	
			<i>0.199</i>	<i>0.234</i>	<i>0.201</i>	<i>0.146</i>	<i>0.218</i>	<i>0.087</i>	<i>0.147</i>	
	500	0.026	0.001	0.017	0.008	0.030	0.008	0.007		
		<i>0.116</i>	<i>0.139</i>	<i>0.114</i>	<i>0.080</i>	<i>0.133</i>	<i>0.047</i>	<i>0.083</i>		
	1000	0.001	0.002	0.005	0.004	0.002	0.001	0.001		
		<i>0.080</i>	<i>0.107</i>	<i>0.074</i>	<i>0.056</i>	<i>0.092</i>	<i>0.032</i>	<i>0.076</i>		

Table 2: Absolute biases (standard font) and standard deviations (italic) of the MLE estimates across 100 simulations for different values of  $n$  and  $\rho$ , and  $\Sigma = (\mathbf{I}_2 + \mathbf{J}_2) / 2$ .

$\Sigma$	$\rho$	$n$	$\mu_{1n}$	$\mu_{2n}$	$\Sigma_{11,n}$	$\Sigma_{12,n}$	$\Sigma_{22,n}$	$\rho_n$	$\phi_n$	
$(\mathbf{I}_2 + \mathbf{J}_2) / 2$	0.2	50	0.007	0.009	0.134	0.082	0.180	0.308	0.049	
			<i>0.396</i>	<i>0.401</i>	<i>0.254</i>	<i>0.183</i>	<i>0.225</i>	<i>0.175</i>	<i>1.555</i>	
		100	0.029	0.054	0.107	0.062	0.081	0.185	0.014	
			<i>0.282</i>	<i>0.289</i>	<i>0.162</i>	<i>0.118</i>	<i>0.165</i>	<i>0.164</i>	<i>1.451</i>	
		200	0.010	0.027	0.046	0.019	0.050	0.120	0.118	
			<i>0.209</i>	<i>0.234</i>	<i>0.104</i>	<i>0.081</i>	<i>0.111</i>	<i>0.137</i>	<i>1.120</i>	
	500	0.011	0.015	0.015	0.001	0.010	0.031	0.103		
		<i>0.138</i>	<i>0.128</i>	<i>0.073</i>	<i>0.060</i>	<i>0.072</i>	<i>0.105</i>	<i>0.864</i>		
	1000	0.008	0.011	0.011	0.001	0.001	0.016	0.063		
		<i>0.082</i>	<i>0.093</i>	<i>0.055</i>	<i>0.039</i>	<i>0.054</i>	<i>0.074</i>	<i>0.505</i>		
	$(\mathbf{I}_2 + \mathbf{J}_2) / 2$	0.5	50	0.037	0.056	0.076	0.016	0.065	0.070	0.075
				<i>0.323</i>	<i>0.346</i>	<i>0.273</i>	<i>0.211</i>	<i>0.268</i>	<i>0.178</i>	<i>0.833</i>
100			0.014	0.007	0.052	0.026	0.060	0.044	0.039	
			<i>0.232</i>	<i>0.263</i>	<i>0.201</i>	<i>0.165</i>	<i>0.197</i>	<i>0.144</i>	<i>0.553</i>	
200			0.025	0.046	0.004	0.013	0.005	0.012	0.077	
			<i>0.187</i>	<i>0.195</i>	<i>0.129</i>	<i>0.099</i>	<i>0.139</i>	<i>0.123</i>	<i>0.482</i>	
500		0.028	0.011	0.022	0.011	0.011	0.016	0.055		
		<i>0.119</i>	<i>0.132</i>	<i>0.092</i>	<i>0.078</i>	<i>0.099</i>	<i>0.095</i>	<i>0.203</i>		
1000		0.009	0.015	0.001	0.002	0.004	0.005	0.017		
		<i>0.087</i>	<i>0.095</i>	<i>0.066</i>	<i>0.055</i>	<i>0.072</i>	<i>0.063</i>	<i>0.166</i>		
$(\mathbf{I}_2 + \mathbf{J}_2) / 2$		0.8	50	0.029	0.101	0.049	0.003	0.034	0.004	0.066
				<i>0.302</i>	<i>0.297</i>	<i>0.222</i>	<i>0.229</i>	<i>0.300</i>	<i>0.069</i>	<i>0.308</i>
	100		0.001	0.019	0.026	0.002	0.009	0.002	0.024	
			<i>0.177</i>	<i>0.224</i>	<i>0.177</i>	<i>0.164</i>	<i>0.235</i>	<i>0.045</i>	<i>0.206</i>	
	200		0.006	0.030	0.024	0.000	0.005	0.001	0.043	
			<i>0.140</i>	<i>0.179</i>	<i>0.136</i>	<i>0.127</i>	<i>0.162</i>	<i>0.044</i>	<i>0.153</i>	
	500	0.003	0.009	0.011	0.008	0.011	0.002	0.006		
		<i>0.079</i>	<i>0.089</i>	<i>0.068</i>	<i>0.061</i>	<i>0.079</i>	<i>0.021</i>	<i>0.073</i>		
	1000	0.000	0.008	0.007	0.012	0.007	0.001	0.002		
		<i>0.057</i>	<i>0.062</i>	<i>0.057</i>	<i>0.057</i>	<i>0.067</i>	<i>0.018</i>	<i>0.049</i>		

## 7 Multivariate applications

### 7.1 Bayesian estimation of expectiles

In this section, we estimate geometric expectiles in the Bayesian framework described in Section 5.2. We adopt the bivariate simulation experiments of Herrmann et al. (2018). More specifically, we consider four bivariate random variables  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(4)}$  (See Table 6 in the Appendix for a description of the data distributions).

We are interested in geometric expectiles  $e_{\rho, \nu(\phi), \mathbf{I}_d}(\mathbf{X}^{(i)})$ , where  $\rho = 0.98$  and the whole spectrum  $\nu(\phi) = (\cos(\phi), \sin(\phi))$  with  $\phi \in [0, 2\pi)$  is considered. According to the Bayesian model of Section 5.2, estimation of such expectiles is carried out by simply setting the parameters  $(\rho, \nu, \Sigma)$  to fixed desired values  $(0.98, \nu(\phi), \mathbf{I}_d)$  (i.e. no prior is set) and by estimating the expectile through parameter  $\mu$ .

We sampled  $n = 1,000$  observations for each of the four distribution settings. For each value of  $\phi$  in the grid  $\{\phi_\ell = 2\pi(\ell - 1)/100, \ell = 1, \dots, 100\}$ , we used Stan (code described in Appendix B.2) with 2,000 HMC iterations after a 1,000-iteration burn-in. Stan provides a  $\hat{R}$  value of 1 for all of the chains, thus not indicating convergence issues. Expectiles were estimated by the posterior mean. For each value of  $\phi$ , credible bands could be computed as those regions containing a prespecified proportion of iterations from the HMC chain (see, for instance, Section 7.2). We have followed a different approach in order to better visualize the credible regions as bands around the estimated expectiles. More specifically, we used multivariate quantiles (in the sense of Chaudhuri, 1996) of orders 2.5% and 97.5%, to get pointwise credible bounds for each value of  $\phi_\ell$ , and then connected them with a curve visible as blue dashed lines on Figure 4. Expectiles were also estimated using the maximum likelihood approach of Herrmann et al. (2018). For each of the four distribution settings, Figure 4 displays contour plots of the true distribution, and the Bayesian and MLE estimates in solid blue and red lines, respectively. Eight values for  $\phi$  defined as  $\varphi_j = j\pi/8, j = 0, \dots, 7$ , are marked along the expectile curves.

We can see from Figure 4 that Bayesian estimates of the expectiles are almost indistinguishable from MLE estimates. The main asset of our Bayesian approach is the availability of uncertainty quantification of these estimates.

### 7.2 Bayesian estimation of the mode of a distribution

We take advantage of the property that the MED mode coincides with its location parameter  $\mu$  (Proposition 3 (i)) to propose a procedure for estimating the mode of a distribution.

Consider the four distribution settings of the previous section. For each of them, we generated  $n$  observations,  $n \in \{100; 500; 1,000\}$ . We used the Bayesian model of Section 5.2 implemented in Stan (code described in Appendix B.2) with 2,000 HMC iterations after a 1,000-iteration burn-in. Stan provides a  $\hat{R}$  value of 1 for all of the chains, thus non indicating any convergence issue. As a competitor to the MED, we used a multivariate Gaussian distribution as a fitting distribution. The Gaussian distribution can be simply obtained from the MED by fixing the  $\rho$  parameter to 0, see Proposition 2 (i) (this automatically makes the  $\nu$  parameter deprecated). The Gaussian distribution also satisfies the property that the mode is available in closed-form (as its mean). For both cases, expectiles were estimated by the posterior mean. In contrast to Section 7.1, 95%

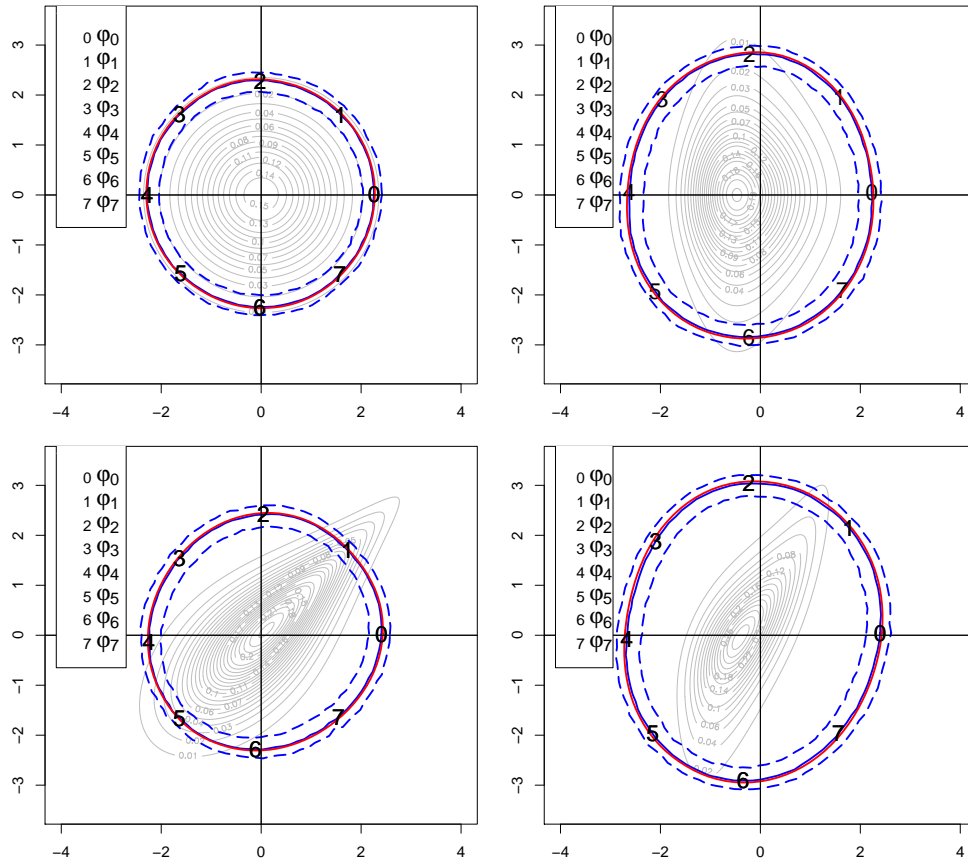


Figure 4: Red: Empirical estimates of expectiles. Blue: Bayesian expectile estimates (solid lines, superimposed with red curves) and credible bands (dashed lines). Grey: level curves of the densities.

Table 3: Mean squared errors computed on 100 repetitions of the Bayesian mode estimator (based on the MED / Gaussian distribution) for different sample sizes  $n$ .

$n$	$\mathbf{X}^{(1)}$	$\mathbf{X}^{(2)}$	$\mathbf{X}^{(3)}$	$\mathbf{X}^{(4)}$
100	0.0588 / 0.0204	0.1054 / 0.0549	0.0553 / 0.0358	0.1139 / 0.0321
500	0.0161 / 0.0037	0.0391 / 0.0408	0.0215 / 0.0253	0.0498 / 0.0188
1,000	0.0086 / 0.0020	0.0347 / 0.0362	0.0109 / 0.0218	0.0352 / 0.0175

Table 4: Empirical coverage probabilities (i.e. the posterior probability that true mode belongs to the credible 95% ellipse) measured on 100 repetitions of the Bayesian mode estimator (based on the MED / Gaussian distribution) for different sample sizes  $n$ .

$n$	$\mathbf{X}^{(1)}$	$\mathbf{X}^{(2)}$	$\mathbf{X}^{(3)}$	$\mathbf{X}^{(4)}$
100	0.99 / 0.95	0.97 / 0.35	1.00 / 0.89	0.94 / 0.61
500	1.00 / 0.96	0.87 / 0.00	0.98 / 0.40	0.79 / 0.01
1,000	1.00 / 0.97	0.74 / 0.00	0.99 / 0.08	0.61 / 0.00

credible regions were provided by sets containing 95% of the 2,000 HMC iterations. These were constructed by using the `dataEllipse` R function from package `car`, which provides the smallest ellipse enveloping the desired proportion of points (95%). The whole procedure was repeated 100 times, leading to Table 3 which displays the mean squared error of the estimates over the 100 repetitions (MED / Gaussian distribution).

Figure 5 represents the results for one of the 100 repetitions obtained with  $n = 1,000$  data points. Each panel corresponds to one of the four distributions, the red cross indicates the true mode, and the blue (resp. green) crosses the estimate from the MED (resp. Gaussian distribution). The ellipses correspond to 95% credible intervals. The results are in accordance with those of Table 3, illustrating that the approach based on the MED provides better calibrated credible intervals than those based on the Gaussian distribution.

### 7.3 Discriminant analysis

We propose a discriminant analysis on real data, by considering the Banknote authentication data set, available at this url<sup>3</sup>. The data of size  $n = 1,372$  contain four covariates ( $X_1 =$  variance,  $X_2 =$  skewness,  $X_3 =$  kurtosis and  $X_4 =$  entropy of wavelet transformed banknote images), and a binary variable indicating whether the banknote is genuine (0) or forged (1). As illustrated in Figure 6, the genuine (blue) vs forged (red) balance in the data set is 56% vs 44%.

We randomly split the data set into a training sample of size  $n = 1,000$  and a testing sample of size 372, and repeat 100 times the experiment. For each experiment, we fit the genuine and forged classes of banknotes with a Gaussian, Skew normal ( $\mathcal{SN}$ ) distribution (Azzalini and Cap-

<sup>3</sup><https://archive.ics.uci.edu/ml/datasets/banknote+authentication>

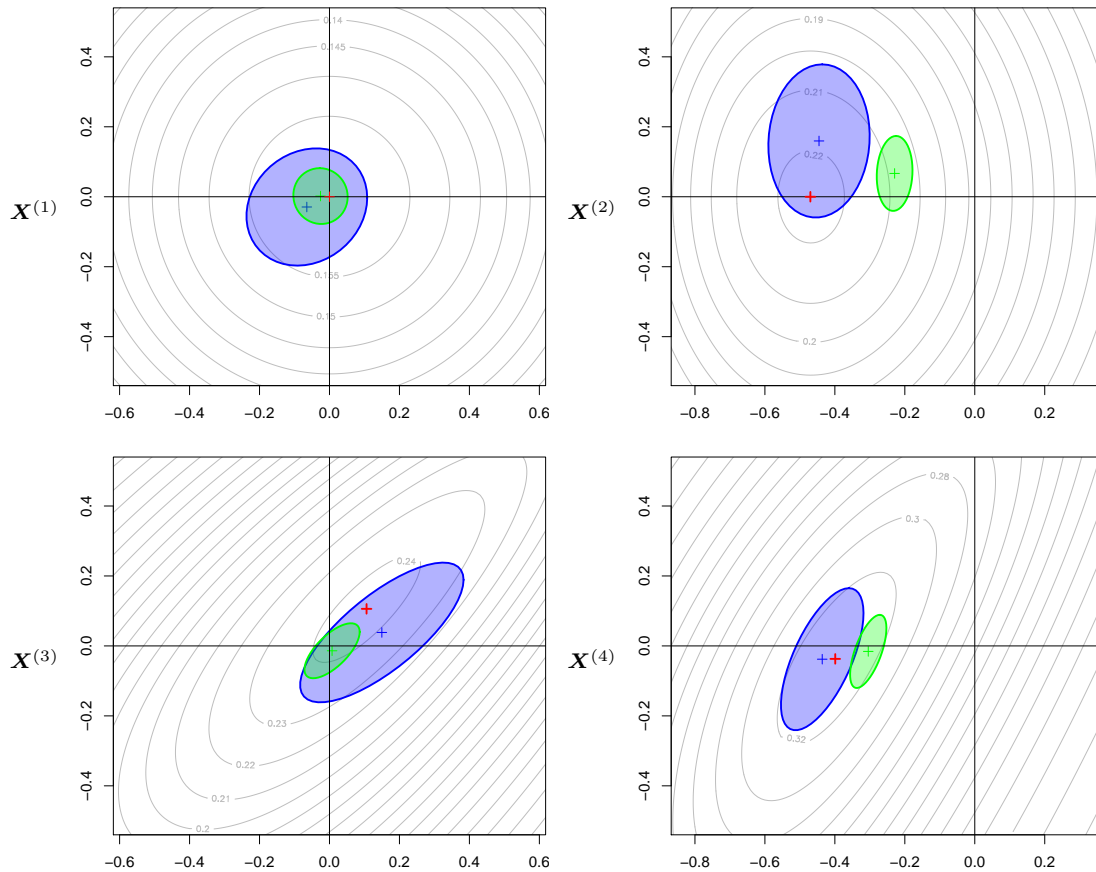


Figure 5: Red, blue and green crosses: theoretical mode, and mode estimates based on the MED and Gaussian distributions, respectively. Blue and green areas: credible ellipses of the mode based on the MED and Gaussian distributions, respectively. Grey: level curves of the densities.

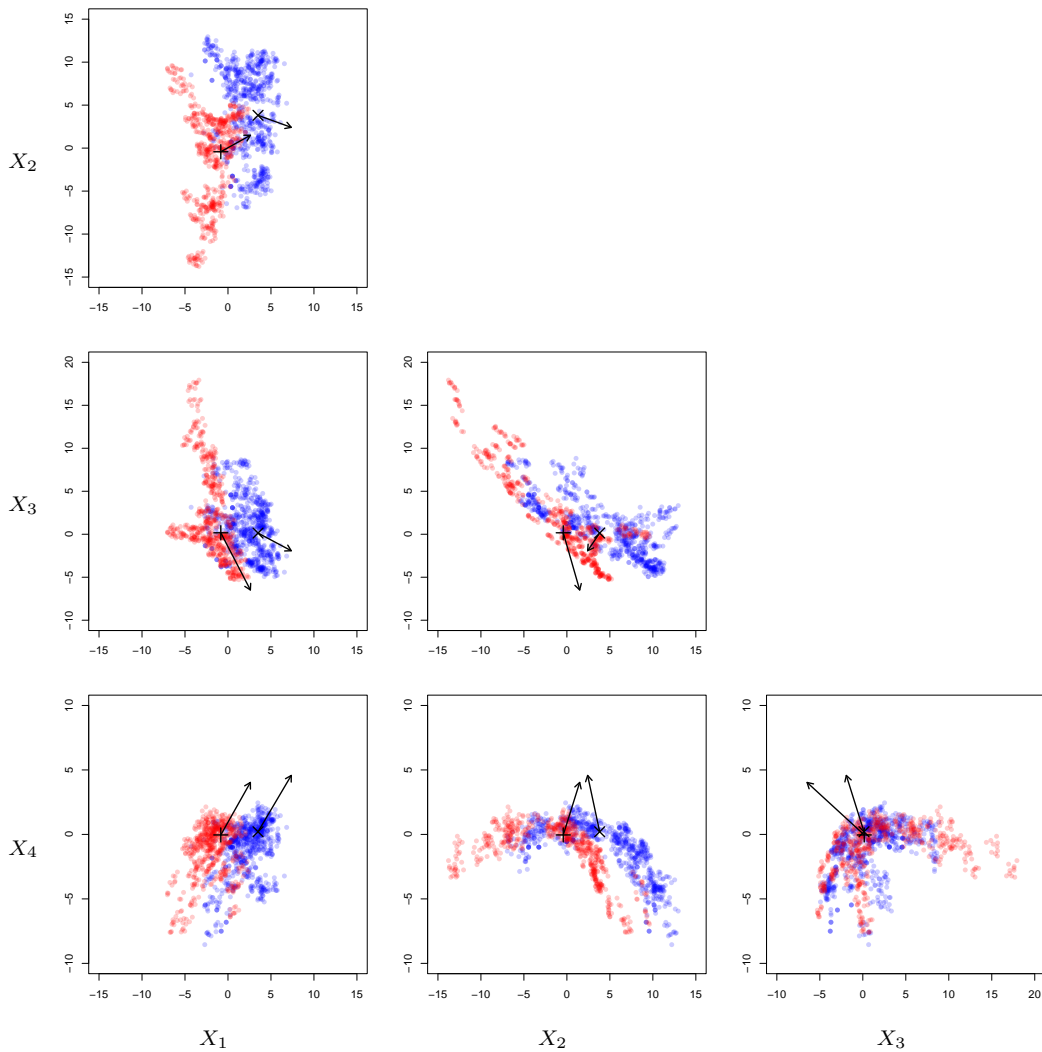


Figure 6: Banknote authentication data set. Covariates  $(X_1, \dots, X_4)$  in blue (genuine) and in red (forged) as bivariate scatterplots. Expectile  $\mu$  estimated with the MED is reported as “ $\times$ ” (genuine) and “ $+$ ” (forged); direction  $\nu$  is represented as an arrow originating from the estimated expectile.

Table 5: Median classification errors computed on the test sample, their standard deviations (SD), their median absolute deviations (MAD), and average AIC and BIC for genuine / forged classes

	% of error	SD	MAD	AIC	BIC
Gaussian	0.538	<b>0.490</b>	0.538	9633 / 7633	9694 / 7691
Skew normal	0.538	0.504	0.538	9501 / <b>7347</b>	9579 / <b>7421</b>
MED	<b>0.269</b>	0.494	<b>0.269</b>	<b>9483</b> / 7509	<b>9565</b> / 7587

itano, 1999) using the R function `msn.mle` from package `sn`, and MED distributions. MLE of the MED parameters  $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho, \boldsymbol{\nu})$  are computed for both classes and reported in Table 7 in Appendix B.3 (let us denote by  $\boldsymbol{\theta}_g$  and  $\boldsymbol{\theta}_f$  the obtained parameters for genuine and forged classes, respectively). Estimated expectiles  $\boldsymbol{\mu}$  and asymmetry directions  $\boldsymbol{\nu}$  are also reported in Figure 6.

We focus on testing the membership of a banknote to the genuine class by using a likelihood-ratio test (with significance level  $\alpha = 0.01$ ). More precisely, for each banknote  $\mathbf{x}_i$  in the test sample ( $i = 1, \dots, 372$ ), we compute

$$\lambda_i = \frac{f_4(\mathbf{x}_i, \boldsymbol{\theta}_g)}{f_4(\mathbf{x}_i, \boldsymbol{\theta}_f)},$$

and the banknote is considered genuine if  $\lambda_i > c_\alpha$ , and forged otherwise. Note that  $c_\alpha$  is the  $\alpha$ -quantile of  $\lambda$ , computed with the genuine banknotes of the training sample. The same approach is adopted for all three distributions. The results (summarized in Table 5) show that the MED has the lowest median error (only 0.269%), while the Skew normal distribution fails to improve the performance of the Gaussian distribution. The mean error, not reported, is also slightly lower for the MED, underlying its flexibility in this application (the MED has lower AIC and BIC values for the genuine class). Note that all the errors of the MED (and the Gaussian distribution) are type I errors, i.e. genuine banknotes wrongly considered as forged (some rare cases of type II errors are obtained with the Skew normal distribution). In comparison with the other distributions, the MED has the advantage of never considering a forged banknote as genuine (like the Skew normal distribution sometimes does), and rarely considers a genuine banknote as forged (less frequently than the Gaussian distribution).

## Acknowledgements

We would like to thank Mario Beraha for improving Stan’s implementation of our Bayesian model. We are grateful to a reviewer who helped us in correcting a typo in the code and in clarifying the manuscript. The work is supported by Inria project [LANDER](#).

## SUPPLEMENTARY MATERIAL

**Title:** The supplementary material contains the R package introduced in the paper.

**R-package MEDdist:** R-package MEDdist containing the functions introduced in Section 5.3.



## References

- Abdous, B. and Remillard, B. (1995). Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics*, 47:371–384.
- Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.
- Andrews, D. W. (1992). Generic uniform convergence. *Econometric Theory*, 8:241–257.
- Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B*, 61(3):579–602.
- Beck, N., Di Bernardino, E., and Mailhot, M. (2021). Semi-parametric estimation of multivariate extreme expectiles. *Journal of Multivariate Analysis*, 184:104758.
- Boumal, N., Absil, P.-A., and Cartis, C. (2018). Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33.
- Brascamp, H. J. and Lieb, E. H. (2002). On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. In *Inequalities*, pages 441–464. Springer.
- Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32.
- Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872.
- Chen, S. and Huang, J. (2014). Rates of convergence of extreme for asymmetric normal distribution. *Statistics & Probability Letters*, 84:158–168.
- Daouia, A., Girard, S., and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B*, 80:263–292.
- Daouia, A. and Paindaveine, D. (2019). From halfspace  $m$ -depth to multiple-output expectile regression. *arXiv preprint arXiv:1905.12718*.
- Fu, M. and Zhou, W. (2016). Non-rigid point set registration via mixture of asymmetric Gaussians with integrated local structures. In *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 999–1004.
- Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *The Annals of Statistics*, 49(6):3358–3382.

- Gradshteyn, I. S. and Ryzhik, I. M. (2014). *Table of integrals, series, and products*. Academic press.
- He, W., Yu, R., Zheng, Y., and Jiang, T. (2018). Image denoising using asymmetric Gaussian mixture models. In *2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI)*, pages 1–4.
- Herrmann, K., Hofert, M., and Mailhot, M. (2018). Multivariate geometric expectiles. *Scandinavian Actuarial Journal*, 2018(7):629–659.
- Holzmann, H. and Klar, B. (2016). Expectile asymptotics. *Electronic Journal of Statistics*, 10:2355–2371.
- Huang, W., Absil, P.-A., and Gallivan, K. A. (2015). A Riemannian symmetric rank-one trust-region method. *Mathematical Programming*, 150(2):179–216.
- Huang, W., Absil, P.-A., Gallivan, K. A., and Hand, P. (2018). ROPTLIB: an object-oriented C++ library for optimization on Riemannian manifolds. *ACM Transactions on Mathematical Software (TOMS)*, 44(4):1–21.
- Huang, W. and Gallivan, K. A. (2022). A Limited-Memory Riemannian Symmetric Rank-One Trust-Region Method with a Restart Strategy. *Journal of Scientific Computing*, 93(1):1–31.
- Kato, T., Omachi, S., and Aso, H. (2002). Asymmetric Gaussian and its application to pattern recognition. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 405–413. Springer.
- Krätschmer, V. and Zähle, H. (2017). Statistical inference for expectile-based risk measures. *Scandinavian Journal of Statistics*, 44:425–454.
- Kuan, C.-M., Yeh, J. H., and Hsu, Y.-C. (2009). Assessing value at risk with CARE, the Conditional Autoregressive Expectile models. *Journal of Econometrics*, 150:261–270.
- Kume, A. and Wood, A. T. (2005). Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika*, 92(2):465–476.
- Majumdar, A. and Paul, D. (2016). Zero expectile processes and Bayesian spatial regression. *Journal of Computational and Graphical Statistics*, 25(3):727–747.
- Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. John Wiley and Sons, New York.
- Martin, S., Raim, A., Huang, W., and Adraghi, K. (2020). ManifoldOptim: An R Interface to the ROPTLIB Library for Riemannian Manifold Optimization. *Journal of Statistical Software*, 93(1):1–32.
- Maume-Deschamps, V., Rullière, D., and Said, K. (2017). Multivariate extensions of expectiles risk measures. *Dependence Modeling*, 5(1):20–44.

- Maume-Deschamps, V., Rullière, D., and Said, K. (2018). Extremes for multivariate expectiles. *Statistics & Risk Modeling*, 35(3-4):111–140.
- Nelsen, R. (2007). *An introduction to copulas*. Springer Science & Business Media.
- Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847.
- Quarteroni, A., Sacco, R., and Saleri, F. (2010). *Numerical mathematics*, volume 37. Springer Science & Business Media.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ritter, G. (2014). *Robust cluster analysis and variable selection*. Chapman and Hall/CRC.
- Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2021). *Lectures on stochastic programming: modeling and theory*. SIAM.
- Tian, Y., Khosoussi, K., and How, J. P. (2019). Block-Coordinate Descent on the Riemannian Staircase for Certifiably Correct Distributed Rotation and Pose Synchronization. *arXiv preprint arXiv:1911.03721*.
- Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494.
- Vladimirova, M., Girard, S., Nguyen, H. D., and Arbel, J. (2020). Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions. *Stat*, 9:e318.
- Ziegel, J. F. (2016). Coherence and elicibility. *Mathematical Finance*, 26:901–918.

## A Appendix: Proofs

### A.1 Proof of main results

#### Proof of results from Section 2

Remarking that the optimization problems

$$\min_{\mathbf{x} \in \mathbb{R}^d} E(\Lambda_{\rho, \nu, \Sigma}(\mathbf{X} - \mathbf{x})) \quad \text{and} \quad \min_{\mathbf{x} \in \mathbb{R}^d} E(\Lambda_{\rho, \nu, \Sigma}(\mathbf{X} - \mathbf{x}) - \Lambda_{\rho, \nu, \Sigma}(\mathbf{X}))$$

are equivalent, the following Proposition provides a sufficient condition for the well-posedness of (1).

**Proposition A.1.** *Let  $\rho \in [0, 1)$ ,  $\Sigma \in \text{SPD}(d)$  and  $\nu \in \mathbb{R}^d$  such that  $\|\nu\|_{\Sigma^{-1}} = 1$ . If  $E(\|\mathbf{X}\|_{\Sigma^{-1}}) < \infty$ , then  $E(|\Lambda_{\rho, \nu, \Sigma}(\mathbf{X} - \mathbf{x}) - \Lambda_{\rho, \nu, \Sigma}(\mathbf{X})|) < \infty$ .*

*Proof of Proposition A.1.* Let us consider the expansion

$$\begin{aligned}
\Lambda_{\rho,\nu,\Sigma}(\mathbf{X} - \mathbf{x}) - \Lambda_{\rho,\nu,\Sigma}(\mathbf{X}) &= \|\mathbf{X} - \mathbf{x}\|_{\Sigma^{-1}}^2 - \|\mathbf{X}\|_{\Sigma^{-1}}^2 \\
&+ \rho\|\mathbf{X} - \mathbf{x}\|_{\Sigma^{-1}} (\langle \mathbf{X} - \mathbf{x}, \boldsymbol{\nu} \rangle_{\Sigma^{-1}} - \langle \mathbf{X}, \boldsymbol{\nu} \rangle_{\Sigma^{-1}}) \\
&+ \rho\langle \mathbf{X}, \boldsymbol{\nu} \rangle_{\Sigma^{-1}} (\|\mathbf{X} - \mathbf{x}\|_{\Sigma^{-1}} - \|\mathbf{X}\|_{\Sigma^{-1}}) \\
&:= T_1 + T_2 + T_3,
\end{aligned}$$

and consider the three terms separately:

$$\begin{aligned}
|T_1| &= |\langle \mathbf{x}, 2\mathbf{X} - \mathbf{x} \rangle_{\Sigma^{-1}}| \leq \|\mathbf{x}\|_{\Sigma^{-1}} (2\|\mathbf{X}\|_{\Sigma^{-1}} + \|\mathbf{x}\|_{\Sigma^{-1}}), \\
|T_2| &= \rho\|\mathbf{X} - \mathbf{x}\|_{\Sigma^{-1}} |\langle \mathbf{x}, \boldsymbol{\nu} \rangle_{\Sigma^{-1}}| \leq (\|\mathbf{X}\|_{\Sigma^{-1}} + \|\mathbf{x}\|_{\Sigma^{-1}}) \|\mathbf{x}\|_{\Sigma^{-1}}, \\
|T_3| &\leq \|\mathbf{X}\|_{\Sigma^{-1}} \|\mathbf{x}\|_{\Sigma^{-1}},
\end{aligned}$$

by repeated uses of Cauchy–Schwarz and triangle inequalities. The conclusion follows.  $\square$

**Proposition A.2** (Strict convexity of  $\Lambda_{\rho,\nu,\Sigma}$ ). *For any  $\rho \in [0, 1)$ , for any matrix  $\Sigma \in \text{SPD}(d)$  and  $\boldsymbol{\nu} \in \mathbb{R}^d$  such that  $\|\boldsymbol{\nu}\|_{\Sigma^{-1}} = 1$ ,  $\Lambda_{\rho,\nu,\Sigma}$  is strictly convex on  $\mathbb{R}^d$ .*

*Proof of Proposition A.2.* Note that convexity is a property that is preserved by translation and rotation, so that it suffices to prove the result for the standardized MED. Denoting by  $x_1 = \langle \mathbf{x}, \mathbf{u}_1 \rangle$  the first component of  $\mathbf{x}$ , and by  $\mathbf{x}_{-1}$  the other  $d - 1$  entries, we have

$$\Lambda_{\rho}(\mathbf{x}) = \|\mathbf{x}\|^2 + \rho x_1 \|\mathbf{x}\| = x_1^2 + \|\mathbf{x}_{-1}\|^2 + \rho x_1 \sqrt{x_1^2 + \|\mathbf{x}_{-1}\|^2} = \tilde{\Lambda}_{\rho}(x_1, \|\mathbf{x}_{-1}\|), \quad (13)$$

where function  $\tilde{\Lambda}_{\rho} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ ,  $(x, y) \mapsto x^2 + y^2 + \rho x \sqrt{x^2 + y^2}$  coincides with  $\Lambda_{\rho}$  in the bivariate case ( $d = 2$ ). Note that it is strictly convex as the Hessian determinant and trace are both positive, equal respectively to  $3(\rho x / \|\mathbf{x}\| + 1)^2 + 1 - \rho^2$  and  $3\rho x / \|\mathbf{x}\| + 4$ , with  $\mathbf{x} = (x, y)$ . Also, the function of the second component  $y \mapsto \tilde{\Lambda}_{\rho}(x, y)$  has partial derivative  $y \mapsto y(\rho x / \|\mathbf{x}\| + 1)$ , meaning that it is non-decreasing on  $\mathbb{R}_+$  for any  $x \in \mathbb{R}$ . As a result, we have that for any  $\lambda \in (0, 1)$ , and for any  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ ,

$$\begin{aligned}
\Lambda_{\rho}(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) &\stackrel{(a)}{=} \tilde{\Lambda}_{\rho}(\lambda x_1 + (1 - \lambda) y_1, \|\lambda \mathbf{x}_{-1} + (1 - \lambda) \mathbf{y}_{-1}\|) \\
&\stackrel{(b)}{\leq} \tilde{\Lambda}_{\rho}(\lambda x_1 + (1 - \lambda) y_1, \lambda \|\mathbf{x}_{-1}\| + (1 - \lambda) \|\mathbf{y}_{-1}\|) \\
&\stackrel{(c)}{<} \lambda \tilde{\Lambda}_{\rho}(x_1, \|\mathbf{x}_{-1}\|) + (1 - \lambda) \tilde{\Lambda}_{\rho}(y_1, \|\mathbf{y}_{-1}\|) \\
&\stackrel{(a)}{=} \lambda \Lambda_{\rho}(\mathbf{x}) + (1 - \lambda) \Lambda_{\rho}(\mathbf{y}),
\end{aligned}$$

where both equalities (a) are by (13), (b) is by combining the above non-decreasing property of  $y \in \mathbb{R}_+ \mapsto \tilde{\Lambda}_{\rho}(x_1, y)$  with the convexity of the Euclidean norm  $\|\cdot\|$ , and (c) is by the strict convexity of  $\tilde{\Lambda}$ . This concludes the proof.  $\square$

*Proof of Proposition 2.* Part (i) is obvious. For part (ii), the standard MED random variable  $\mathbf{Z} = (Z_1, \mathbf{Z}_{-1})$  has density

$$f_d(\mathbf{z}; \rho) = \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{z_1^2}{2} - \frac{\|\mathbf{z}_{-1}\|^2}{2} - \frac{\rho z_1}{2} \sqrt{z_1^2 + \|\mathbf{z}_{-1}\|^2}\right)$$

The density of the transformed random variable  $\Delta_\rho \mathbf{Z} = (\sqrt{1-\rho}Z_1, \mathbf{Z}_{-1}/\sqrt{2})$  is given by  $2^{(d-1)/2}(1-\rho)^{-1/2}f_d(z_1/\sqrt{1-\rho}, \sqrt{2}\mathbf{z}_{-1}; \rho)$ ; that is:

$$\begin{aligned} & \frac{2^{(d-1)/2}c_d(\rho)}{(2\pi)^{d/2}(1-\rho)^{1/2}} \exp\left(-\frac{z_1^2}{2(1-\rho)} - \|\mathbf{z}_{-1}\|^2 - \frac{\rho z_1}{2\sqrt{1-\rho}} \sqrt{\frac{z_1^2}{1-\rho} + 2\|\mathbf{z}_{-1}\|^2}\right) \\ &= \frac{2^{(d-1)/2}c_d(\rho)}{(2\pi)^{d/2}(1-\rho)^{1/2}} \exp\left(-\|\mathbf{z}_{-1}\|^2 - \frac{z_1^2}{2(1-\rho)} \left[1 + \rho \operatorname{sign}(z_1) \sqrt{1 + \frac{2(1-\rho)\|\mathbf{z}_{-1}\|^2}{z_1^2}}\right]\right) \\ &= \frac{2^{(d-1)/2}c_d(\rho)}{(2\pi)^{d/2}(1-\rho)^{1/2}} \exp\left(-\|\mathbf{z}_{-1}\|^2 - \frac{z_1^2}{2(1-\rho)} \left[1 + \rho \operatorname{sign}(z_1) \left(1 + \frac{(1-\rho)\|\mathbf{z}_{-1}\|^2}{z_1^2} + o(1-\rho)\right)\right]\right) \\ &\underset{\rho \rightarrow 1}{\approx} \frac{2}{(2\pi)^{d/2}} \exp\left(-\frac{\|\mathbf{z}_{-1}\|^2}{2} - \frac{1 + \rho \operatorname{sign}(z_1)}{(1-\rho)} \frac{z_1^2}{2}\right) \\ &\xrightarrow{\rho \rightarrow 1} \frac{2}{(2\pi)^{d/2}} \exp\left(-\frac{\|\mathbf{z}_{-1}\|^2}{2} - \frac{z_1^2}{2}\right) \mathbb{I}\{z_1 < 0\}. \end{aligned}$$

This corresponds to a half-standard Gaussian restricted to negative values of the first component. Lastly, for part (iii) with  $d = 1$ , consider without loss of generality, the  $\text{MED}(\mu, \sigma^2, 1, \rho)$ . Its density is made of two half-Gaussians on  $(-\infty, \mu]$  and on  $[\mu, \infty)$  with respective variances  $\sigma_-^2 = \frac{\sigma^2}{1-\rho}$  and  $\sigma_+^2 = \frac{\sigma^2}{1+\rho}$ . This corresponds to the one-dimensional skew-Gaussian distribution of [Kato et al. \(2002\)](#), shifted by  $\mu$ , with variance  $\sigma_+^2$  (for the right tail) and ratio of variances  $r^2 = \frac{1+\rho}{1-\rho}$ .  $\square$

*Proof of Proposition 3.* (i) The mode(s) of the  $\text{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$  is (are) equivalently defined by

$$\arg \min_{\mathbf{x} \in \mathbb{R}^d} \Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{x} - \boldsymbol{\mu}) = \boldsymbol{\mu} + \arg \min_{\mathbf{y} \in \mathbb{R}^d} \Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{y}).$$

From Proposition A.2,  $\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}$  is strictly convex on  $\mathbb{R}^d$  and therefore the mode is unique. Moreover,  $\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{0}) = 0$  and  $\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{y}) \geq 0$ , for any  $\mathbf{y} \in \mathbb{R}^d$ , in view of Cauchy–Schwarz inequality. It follows that

$$\arg \min_{\mathbf{y} \in \mathbb{R}^d} \Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{y}) = \mathbf{0}$$

and the first result is proved.

(ii) By definition, the unique multivariate geometric expectile is given by

$$\begin{aligned} \mathbf{e}_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{X}) &= \arg \min_{\mathbf{x} \in \mathbb{R}^d} E(\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{X} - \mathbf{x})) \\ &= \arg \min_{\mathbf{x} \in \mathbb{R}^d} \int_{\mathbb{R}^d} \Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{t} - \mathbf{x}) \exp(-\frac{1}{2}\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{t} - \boldsymbol{\mu})) d\mathbf{t} =: \arg \min_{\mathbf{x} \in \mathbb{R}^d} J_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}, \boldsymbol{\mu}}(\mathbf{x}). \end{aligned}$$

To prove that  $\mathbf{e}_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{X}) = \boldsymbol{\mu}$ , it is then sufficient to check that the gradient of  $J_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}, \boldsymbol{\mu}}$  vanishes at  $\mathbf{x} = \boldsymbol{\mu}$ :

$$\begin{aligned} \nabla J_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}, \boldsymbol{\mu}}|_{\mathbf{x}=\boldsymbol{\mu}} &= - \int_{\mathbb{R}^d} \nabla \Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{t} - \boldsymbol{\mu}) \exp(-\frac{1}{2}\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{t} - \boldsymbol{\mu})) d\mathbf{t} \\ &= - \int_{\mathbb{R}^d} \nabla \Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{y}) \exp(-\frac{1}{2}\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{y})) d\mathbf{y} = 2 \int_{\mathbb{R}^d} \nabla \exp(-\frac{1}{2}\Lambda_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\mathbf{y})) d\mathbf{y} \\ &= 2 \int_{\mathbb{R}^d} \nabla \exp(-\frac{1}{2}\Lambda_{\rho, \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\nu}, \mathbf{I}_d}(\boldsymbol{\Sigma}^{-1/2} \mathbf{y})) d\mathbf{y} \propto \int_{\mathbb{R}^d} \nabla \exp(-\frac{1}{2}\Lambda_{\rho, \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\nu}, \mathbf{I}_d}(\mathbf{z})) d\mathbf{z}. \end{aligned}$$

For any  $j = 1, \dots, d$ , integrating along the  $j$ th component provides the difference of the limits of  $\exp(-\frac{1}{2}\Lambda_{\rho, \Sigma^{-1/2}\boldsymbol{\nu}, \mathbf{I}_d}(\mathbf{z}))$  as  $z_j$ , the  $j$ th coordinate of  $\mathbf{z}$ , goes to  $+\infty$  and  $-\infty$ , which are both zero since  $\Lambda_{\rho, \Sigma^{-1/2}\boldsymbol{\nu}, \mathbf{I}_d}$  is coercive, see [Herrmann et al. \(2018, Theorem 4.5\)](#). As a consequence,  $\nabla J_{\rho, \boldsymbol{\nu}, \Sigma, \boldsymbol{\mu}}|_{\mathbf{x}=\boldsymbol{\mu}} = \mathbf{0}$  and the result follows.

(iii) The proof is similar to that of Lemma A.1 in Appendix A.2. Recalling that  $\mathbf{Q}$  is an orthogonal matrix such that  $\mathbf{Q}\Sigma^{-1/2}\boldsymbol{\nu} = \mathbf{u}_1$ , one can write:

$$E[\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})] = \frac{c_d(\rho)}{(2\pi)^{d/2}} 2^{\frac{d-1}{2}} \Gamma\left(\frac{d+1}{2}\right) \mathbf{Q}^\top \int_{\mathcal{S}^{d-1}} \frac{\boldsymbol{\theta} d\mathcal{S}^{d-1}(\boldsymbol{\theta})}{(1 + \rho\langle \boldsymbol{\theta}, \mathbf{u}_1 \rangle)^{(d+1)/2}}.$$

Inspired by the proof of Lemma A.1, changes of variables lead to:

$$E[\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})] = \frac{c_d(\rho)}{(2\pi)^{d/2}} 2^{\frac{d-1}{2}} \Gamma\left(\frac{d+1}{2}\right) \mathcal{A}^{d-2} \mathbf{Q}^\top \mathbf{u}_1 \int_0^\pi \frac{\cos(\psi) (\sin(\psi))^{d-2}}{(1 + \rho \cos(\psi))^{\frac{d+1}{2}}} d\psi.$$

Since  $\mathbf{Q}^\top \mathbf{u}_1 = \Sigma^{-1/2}\boldsymbol{\nu}$ , and

$$\frac{1}{\Gamma\left(\frac{d-1}{2}\right)} \int_0^\pi \frac{\cos(\psi) (\sin(\psi))^{d-2}}{(1 + \rho \cos(\psi))^{\frac{d+1}{2}}} d\psi = -\frac{d+1}{4} \sqrt{\pi} \rho \frac{{}_2F_1\left(\frac{d+3}{4}, \frac{d+5}{4}, \frac{d+2}{2}; \rho^2\right)}{\Gamma\left(\frac{d+2}{2}\right)},$$

we get the expected result.  $\square$

*Proof of Proposition 4.* Let  $\mathbf{X} \sim \text{MED}(\rho_1)$ . By definition,  $\mathbf{e}_{\rho_2}(\mathbf{X})$  is the solution of a slightly modified Equation (1) involving two asymmetry parameters  $\rho_1$  and  $\rho_2$ :

$$\mathbf{e}_{\rho_2}(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} E(\Lambda_{\rho_2}(\mathbf{X} - \mathbf{x})), \quad (14)$$

where expectation is taken with respect to the  $\text{MED}(\rho_1)$ . In other words, the expectation is taken with respect to  $\mathbf{X} \sim \text{MED}(\rho_1)$ , but the order of the expectile is  $\rho_2$ . Since the objective function in the above optimization task is strictly convex by Proposition A.2, it has a unique solution. By cylindrical symmetry of the distribution of  $\mathbf{X}$  along the first axis (the direction of  $\mathbf{u}_1$  through the origin), the expectile  $\mathbf{e}_{\rho_2}(\mathbf{X})$  is also supported on this axis, so there exists some  $m \in \mathbb{R}$  such that  $\mathbf{e}_{\rho_2}(\mathbf{X}) = m\mathbf{u}_1$ . Let  $\psi$  be the function with arguments  $(m, \rho_2) \in \mathbb{R} \times [0, 1)$  defined from the objective function above as:

$$\psi(m, \rho_2) := E[\Lambda_{\rho_2}(\mathbf{X} - m\mathbf{u}_1)] = E[\|\mathbf{X} - m\mathbf{u}_1\|(\|\mathbf{X} - m\mathbf{u}_1\| + \rho_2(X_1 - m))].$$

Differentiating with respect to  $m$  yields

$$\frac{\partial}{\partial m} \psi(m, \rho_2) = E\left[2(m - X_1) - \rho_2 \left(\frac{(m - X_1)^2}{\|\mathbf{X} - m\mathbf{u}_1\|} + \|\mathbf{X} - m\mathbf{u}_1\|\right)\right].$$

The value in  $m = 0$  is equal to

$$\frac{\partial}{\partial m} \psi(m = 0, \rho_2) = E\left[-2X_1 - \rho_2 \left(\frac{X_1^2}{\|\mathbf{X}\|} + \|\mathbf{X}\|\right)\right], \quad (15)$$

which is an affine function with negative slope. Additionally, this function vanishes when  $\rho_2$  equals  $\rho_1$ , by definition of the expectile of order  $\rho_1$  being the minimizer of (1). Hence, the equivalence of the statement: if  $\rho_2 > \rho_1$ , then the derivative in (15) is negative, therefore going along the first axis in the positive direction decreases the objective function, which is necessarily the direction of  $\mathbf{e}_{\rho_2}(\mathbf{X})$  by convexity; so  $m > 0$ . The other equivalences are derived similarly.  $\square$

### Proof of results from Section 3

*Proof of Proposition 5.* One can write the density of  $\mathbf{Z} \sim \text{MED}(\rho)$  as:

$$f_d(\mathbf{z}; \rho) = \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\mathbf{z}\|^2}{2} \left[1 + \rho \left\langle \frac{\mathbf{z}}{\|\mathbf{z}\|}, \mathbf{u}_1 \right\rangle\right]\right),$$

and thus, noting that the dot product  $\langle \cdot, \cdot \rangle$  lies in  $[-1, 1]$  provides the simple bounds

$$\frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\mathbf{z}\|^2}{2} (1 + \rho)\right) \leq f_d(\mathbf{z}; \rho) \leq \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\mathbf{z}\|^2}{2} (1 - \rho)\right). \quad (16)$$

Let  $\mathbf{X} = \Sigma^{1/2} \mathbf{Q}^\top \mathbf{Z} + \boldsymbol{\mu} \sim \text{MED}(\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho)$  from Corollary 1. By change of variable, the density of  $\mathbf{X}$  is given by

$$f_d(\mathbf{x}; \boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}, \rho) = \frac{1}{|\Sigma|^{1/2}} f_d(\mathbf{Q}\Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu}); \rho).$$

By replacement in (16), the result is proved.  $\square$

*Proof of Corollary 2.* Let  $\mathbf{Z} = \mathbf{Q}\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$  be the standardized version of  $\mathbf{X}$ ; see Corollary 1. For any non-zero  $\boldsymbol{\beta} \in \mathbb{R}^d$  and  $p > 0$ , one has

$$E(|\langle \boldsymbol{\beta}, \mathbf{Z} \rangle|^p) = \int_{\mathbb{R}^d} |\langle \boldsymbol{\beta}, \mathbf{z} \rangle|^p f_d(\mathbf{z}; \rho) d\mathbf{z},$$

and thus, in view of Proposition 5,

$$m_d(\rho) E(|Y_+|^p) \leq E(|\langle \boldsymbol{\beta}, \mathbf{Z} \rangle|^p) \leq M_d(\rho) E(|Y_-|^p),$$

where  $Y_-$  and  $Y_+$  are univariate centered Gaussian random variables with variance  $1/(1 - \rho)$  and  $1/(1 + \rho)$ , respectively. It follows that

$$m_d(\rho)(1 + \rho)^{-p/2} E(|Y_0|^p) \leq E(|\langle \boldsymbol{\beta}, \mathbf{Z} \rangle|^p) \leq M_d(\rho)(1 - \rho)^{-p/2} E(|Y_0|^p), \quad (17)$$

where  $Y_0$  is a standard Gaussian random variable whose moments are given by  $E(|Y_0|^p) = 2^{p/2} \pi^{-1/2} \Gamma((p + 1)/2)$ . By  $E(|\langle \boldsymbol{\beta}, \mathbf{Z} \rangle|^p) = E(|\langle \Sigma^{1/2} \boldsymbol{\beta}, \mathbf{X} - \boldsymbol{\mu} \rangle_{\Sigma^{-1}}|^p)$ , and noting that the bounds of (17) do not depend on  $\boldsymbol{\beta} \neq 0$ , the conclusion follows.  $\square$

*Proof of Proposition 6.* A similar calculation as that of Proposition 3 leads to

$$E [\|\mathbf{X} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2] = d \frac{(1-\rho)^{d/2} \left(1 + \sqrt{\frac{1+\rho}{1-\rho}}\right)^d (\rho^2 - 1)}{2^{d-1} (\rho^2 - 1 - \sqrt{1-\rho^2})} \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} \pi^{-1/2} \int_{-1}^1 \frac{(1-x^2)^{\frac{d-3}{2}}}{(1+\rho x)^{1+d/2}} dx.$$

The result is deduced via the identity

$$\frac{1}{\Gamma(\frac{d-1}{2})} \int_{-1}^1 \frac{(1-x^2)^{(d-3)/2}}{(1+\rho x)^{1+d/2}} dx = \frac{1}{\Gamma(1+\frac{d}{2})} \frac{2^{d/2-2} \sqrt{\pi} (d - d\rho^2 + 2\rho^2 + d\sqrt{1-\rho^2})}{(1+\sqrt{1-\rho^2})^{d/2} (1-\rho^2)^{3/2}}.$$

□

*Proof of Proposition 7.* (i) is a direct application of the asymptotic equivalent for the  ${}_2F_1$  function from Lemma A.2 in Appendix A.2, while (ii) comes from the fact that the second term in the squared norm expectation in Proposition 6 dominates the first. □

#### Proof of results from Section 4

*Proof of Proposition 8.* Let  $H_{d-1}$  be the hyperplane orthogonal to  $\mathbf{u}_1$ . (i) Consider the one-dimensional margin  $Z_1$  of  $\mathbf{Z}$ :

$$f_{Z_1}(z_1) = \int_{H_{d-1}} f_d(z_1, \mathbf{z}_{-1}) d\mathbf{z}_{-1} = \frac{c_d(\rho)}{(2\pi)^{d/2}} \int_{H_{d-1}} \exp\left(-\frac{z_1^2}{2} - \frac{\|\mathbf{z}_{-1}\|^2}{2} - \frac{\rho z_1}{2} \sqrt{z_1^2 + \|\mathbf{z}_{-1}\|^2}\right) d\mathbf{z}_{-1}.$$

The first result is obtained by integrating the bounds provided by Proposition 5 on  $H_{d-1}$ . Secondly, let us assume  $z_1 > 0$ , the case  $z_1 < 0$  being obtained by replacing  $\rho$  by  $-\rho$ . Consider the change of variable  $\mathbf{z}_{-1} = z_1 \mathbf{x}$ :

$$f_{Z_1}(z_1) = \frac{c_d(\rho)}{(2\pi)^{d/2}} z_1^{d-1} \int_{H_{d-1}} \exp\left(-\frac{z_1^2}{2} (1 + \|\mathbf{x}\|^2 + \rho \operatorname{sign}(z_1) \sqrt{1 + \|\mathbf{x}\|^2})\right) d\mathbf{x}.$$

Switching to polar coordinates yields:

$$f_{Z_1}(z_1) = \mathcal{A}^{d-2} \frac{c_d(\rho)}{(2\pi)^{d/2}} z_1^{d-1} \int_0^\infty \exp\left(-\frac{z_1^2}{2} (1 + r^2 + \rho \sqrt{1 + r^2})\right) r^{d-2} dr,$$

where  $\mathcal{A}^{d-2} = 2\pi^{(d-1)/2} / \Gamma((d-1)/2)$  is the surface of the unit sphere  $\mathcal{S}^{d-2}$ . Introducing

$$I_d(z_1) = \int_0^\infty \exp\left(-\frac{z_1^2}{2} \varphi(r)\right) r^{d-2} dr$$

with  $\varphi(r) = 1 + r^2 + \rho \sqrt{1 + r^2}$ , and applying Lemma A.3 with  $\varphi(0) = 1 + \rho$  and  $\varphi''(0) = 2 + \rho$ , it follows that,

$$f_{Z_1}(z_1) \underset{z_1 \rightarrow +\infty}{\approx} \frac{c_d(\rho)}{(2\pi)^{1/2}} \left(\frac{1}{1 + \rho/2}\right)^{(d-1)/2} \exp\left(-\frac{z_1^2}{2} (1 + \rho)\right),$$



while, in view of the above remarks,

$$f_{Z_1}(z_1) \underset{z_1 \rightarrow -\infty}{\approx} \frac{c_d(\rho)}{(2\pi)^{1/2}} \left( \frac{1}{1 - \rho/2} \right)^{(d-1)/2} \exp\left(-\frac{z_1^2}{2}(1 - \rho)\right).$$

(ii) For any  $\mathbf{z} \in \mathbb{R}^d$ , write  $\mathbf{z} = (z_1, \mathbf{z}_{-1})$  so that

$$\begin{aligned} f_{\mathbf{Z}_{-1}}(\mathbf{z}_{-1}) &= \int_{\mathbb{R}} f_d(z_1 \mathbf{u}_1 + \mathbf{z}_{-1}) dz_1 \\ &= \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\mathbf{z}_{-1}\|^2}{2}\right) \int_{\mathbb{R}} \exp\left(-\frac{z_1^2}{2} - \frac{\rho z_1}{2} \sqrt{z_1^2 + \|\mathbf{z}_{-1}\|^2}\right) dz_1 \\ &=: \tilde{f}(\|\mathbf{z}_{-1}\|), \end{aligned}$$

which proves that the marginal distribution of  $\mathbf{Z}_{-1}$  is an elliptically contoured distribution. Let  $z = \|\mathbf{z}_{-1}\|$ . A simple linear change of variables decouples  $z$  and  $z_1$  in the integrand above

$$\int_{\mathbb{R}} \exp\left(-\frac{z_1^2}{2} - \frac{\rho z_1}{2} \sqrt{z_1^2 + \|\mathbf{z}_{-1}\|^2}\right) dz_1 = z \int_{\mathbb{R}} \exp\left(-\frac{z^2}{2} \left(z_1^2 + \rho z_1 \sqrt{z_1^2 + 1}\right)\right) dz_1.$$

Let function  $h$  be defined on  $\mathbb{R}$  by  $h(z_1) = z_1^2 + \rho z_1 \sqrt{z_1^2 + 1}$ . We know from Proposition A.2 that  $h$  is strictly convex, thus it admits a single minimum on  $\mathbb{R}$ . Simple algebra yields that the arg min and minimum are  $t_\rho = -\left(\frac{1}{2} \left(\frac{1}{\sqrt{1-\rho^2}} - 1\right)\right)^{1/2}$  and  $h(t_\rho) = \frac{1}{2}(\sqrt{1-\rho^2} - 1)$ . As a result, Laplace's method provides the following large  $z$  equivalent to the integral

$$\int_{\mathbb{R}} \exp\left(-\frac{z^2}{2} \left(z_1^2 + \rho z_1 \sqrt{z_1^2 + 1}\right)\right) dz_1 \underset{z_1 \rightarrow +\infty}{\approx} \sqrt{\frac{2\pi}{z^2 h''(t_\rho)/2}} \exp\left(-\frac{z^2}{2} h(t_\rho)\right),$$

where

$$h''(t_\rho) = 2 - \frac{1}{\sqrt{2}} \rho \sqrt{\frac{1}{\sqrt{1-\rho^2}} - 1} \left(\frac{1}{\sqrt{1-\rho^2}} + 2\right) \left(1/2 \left(\frac{1}{\sqrt{1-\rho^2}} - 1\right) + 1\right)^{-3/2}.$$

Together with the above identities, we get the large  $\|\mathbf{z}_{-1}\|$  distribution

$$\begin{aligned} f_{\mathbf{Z}_{-1}}(\mathbf{z}_{-1}) &= \frac{c_d(\rho)}{(2\pi)^{d/2}} \|\mathbf{z}_{-1}\| \exp\left(-\frac{\|\mathbf{z}_{-1}\|^2}{2}\right) \exp\left(-\frac{\|\mathbf{z}_{-1}\|^2}{2} \frac{1}{2}(\sqrt{1-\rho^2} - 1)\right) \\ &\underset{\|\mathbf{z}_{-1}\| \rightarrow +\infty}{\approx} C_d(\rho) \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\mathbf{z}_{-1}\|^2}{2} \frac{1}{2}(\sqrt{1-\rho^2} + 1)\right), \end{aligned}$$

where

$$C_d(\rho) = \sqrt{\frac{4\pi}{h''(t_\rho)}} = \frac{\sqrt{\pi}}{2^{d-2}} \frac{(1 + \sqrt{1-\rho^2})^{d-3/2}}{\sqrt{1-\rho^2}}.$$

Additionally,  $C_d(0) = \sqrt{2\pi}$ , since  $h''(t_\rho) = 2$  when  $\rho = 0$ . □

*Proof of Proposition 9.* (i) For any  $z \in \mathbb{R}$ , the density  $f_{\langle \mathbf{w}, \mathbf{z} \rangle | P_{\mathbf{w}^\perp}(\mathbf{z})=0}(z)$  is equal to:

$$\frac{f_d(z\mathbf{w}; \rho)}{f_{P_{\mathbf{w}^\perp}(\mathbf{z})}(\mathbf{0})} \propto f_d(z\mathbf{w}; \rho) \propto \exp\left(-\frac{1}{2}z^2 - \frac{\rho}{2}\langle \mathbf{w}, \mathbf{u}_1 \rangle z|z|\right) = \exp\left(-\frac{1}{2}(1 + \rho\langle \mathbf{w}, \mathbf{u}_1 \rangle \text{sign}(z))z^2\right),$$

which is the one-dimensional MED of the statement.

(ii) Conditioning on  $Z_1 = 0$  cancels the non-quadratic term in the exponential term of the density. As a consequence, this conditional density is proportional to  $\exp(-\frac{1}{2}\|z_{-1}\|^2)$ , which identifies the asserted multivariate Gaussian distribution.  $\square$

## Proofs of results from Section 5

*Proof of Proposition 10.* The proof is a direct application of Theorem 5.3 of [Shapiro et al. \(2021\)](#). The only non-trivial criterion to be checked is that

$$\sup_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} |n^{-1}\ell_n(\tilde{\boldsymbol{\theta}}) - E \log f_d(\mathbf{X}; \tilde{\boldsymbol{\theta}})| \xrightarrow{a.s.} 0,$$

as  $n \rightarrow \infty$ , for compact set  $\bar{\Theta}$ . We can apply Theorem 6 of [Andrews \(1992\)](#) to verify the condition. To do so, we verify that  $n^{-1}\ell_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{a.s.} E \log f_d(\mathbf{X}; \tilde{\boldsymbol{\theta}})$ , pointwise. This is possible via a trivial application of the strong law of large numbers combined with Proposition 5 and Corollary 2. To complete the proof, we must demonstrate the existence of a function  $\alpha(\mathbf{x})$ , such that  $\alpha(\mathbf{x}) \geq \sup_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} |\log f_d(\mathbf{x}; \tilde{\boldsymbol{\theta}})|$  and  $E(\alpha(\mathbf{X})) < \infty$ . Such a function can be obtained as follows. Use Proposition 5 to obtain the inequalities:

$$m_d(\rho) \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1+\rho}\right) \leq f_d(\mathbf{x}; \tilde{\boldsymbol{\theta}}) \leq M_d(\rho) \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1-\rho}\right),$$

and thus

$$\left| \log f_d(\mathbf{x}; \tilde{\boldsymbol{\theta}}) \right| \leq |\log m_d(\rho)| + |\log M_d(\rho)| + \left| \log \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1+\rho}\right) \right| + \left| \log \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1-\rho}\right) \right|,$$

from which we set

$$\alpha(\mathbf{x}) = \sup_{\rho \in [0, \gamma]} \{|\log m_d(\rho)| + |\log M_d(\rho)|\} + \sup_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} \left| \log \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1+\rho}\right) \right| + \sup_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} \left| \log \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1-\rho}\right) \right|,$$

where  $\gamma$  above is defined in  $\bar{\Theta}$  in (12). The function  $\alpha$  is well-defined (and non-infinite) by construction of the compact set  $\bar{\Theta}$ . We apply Proposition 5, again, to obtain

$$\begin{aligned} E \sup_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} \left| \log \phi_d\left(\mathbf{X}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1-\rho}\right) \right| &= \int \sup_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} \left| \log \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1-\rho}\right) \right| f(\mathbf{x}; \tilde{\boldsymbol{\theta}}) \, d\mathbf{x} \\ &\leq M_d(\rho) \int \sup_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} \left| \log \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1-\rho}\right) \right| \phi_d\left(\mathbf{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1-\rho}\right) \, d\mathbf{x}, \end{aligned}$$

where the last integral is finite since the normal distribution has finite first and second moments. The analogous result holds for  $E \sup_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} \left| \log \phi_d\left(\mathbf{X}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1+\rho}\right) \right|$ , and thus we have verified that  $\alpha(\mathbf{x})$  satisfies the necessary properties.  $\square$

## A.2 Auxiliary results

**Lemma A.1.** For any  $\rho \in [0, 1)$ , for any  $\Sigma \in \text{SPD}(d)$  and  $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^d$  such that  $\|\boldsymbol{\nu}\|_{\Sigma^{-1}} = 1$ ,

$$\frac{c_d(\rho)}{(2\pi)^{d/2} |\Sigma|^{1/2}} \int_{\mathbb{R}^d} \exp(-\Lambda_{\rho, \boldsymbol{\nu}, \Sigma}(\mathbf{x} - \boldsymbol{\mu})) d\mathbf{x} = 1,$$

where  $c_d(\rho)$  is given in (2).

*Proof.* Let us calculate the integral

$$J_d(\rho) = |\Sigma|^{-1/2} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\|\mathbf{z}_{-1} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2\right) \exp\left(-\frac{\rho}{2}\langle \mathbf{z}_{-1} - \boldsymbol{\mu}, \boldsymbol{\nu} \rangle_{\Sigma^{-1}} \|\mathbf{z}_{-1} - \boldsymbol{\mu}\|_{\Sigma^{-1}}\right) d\mathbf{z}_{-1}.$$

The change of variable  $\mathbf{z} = \Sigma^{-1/2}(\mathbf{z}_{-1} - \boldsymbol{\mu})$  yields

$$J_d(\rho) = \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\|\mathbf{z}\|^2\right) \exp\left(-\frac{\rho}{2}\langle \mathbf{z}, \Sigma^{-1/2}\boldsymbol{\nu} \rangle \|\mathbf{z}\|\right) d\mathbf{z}.$$

By assumption,  $\Sigma^{-1/2}\boldsymbol{\nu}$  lies on the unit sphere, so there exists an orthogonal matrix, denoted  $\mathbf{Q}$ , such that  $\mathbf{Q}\Sigma^{-1/2}\boldsymbol{\nu} = \mathbf{u}_1$ , the first unit axis vector. Then, the change of variable  $\mathbf{x} = \mathbf{Q}\mathbf{z}$  leads to

$$J_d(\rho) = \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \exp\left(-\frac{\rho}{2}\langle \mathbf{x}, \mathbf{u}_1 \rangle \|\mathbf{x}\|\right) d\mathbf{x}.$$

We use polar coordinates:  $\mathbf{x} = r\boldsymbol{\theta}$ ,  $r \in \mathbb{R}_+$  and  $\boldsymbol{\theta} \in \mathcal{S}^{d-1}$  leading to  $d\mathbf{x} = r^{d-1}dr d\mathcal{S}^{d-1}(\boldsymbol{\theta})$  and

$$\begin{aligned} J_d(\rho) &= \int_{\mathcal{S}^{d-1}} \int_0^\infty \exp\left(-\frac{r^2}{2}(1 + \rho\langle \boldsymbol{\theta}, \mathbf{u}_1 \rangle)\right) r^{d-1} dr d\mathcal{S}^{d-1}(\boldsymbol{\theta}) \\ &= \int_0^\infty \exp\left(-\frac{u^2}{2}\right) u^{d-1} du \int_{\mathcal{S}^{d-1}} \frac{d\mathcal{S}^{d-1}(\boldsymbol{\theta})}{(1 + \rho\langle \boldsymbol{\theta}, \mathbf{u}_1 \rangle)^{d/2}} =: J_d^{(1)} J_d^{(2)}(\rho). \end{aligned}$$

Clearly, the first integral  $J_d^{(1)}$  equals  $2^{d/2-1}\Gamma(d/2)$  by definition of the gamma function. To deal with the second integral  $J_d^{(2)}(\rho)$ , write  $\boldsymbol{\theta}$  as  $(\theta_1, \boldsymbol{\theta}')$ . Introduce now polar coordinates for both  $\theta_1$  and  $\boldsymbol{\theta}'$ :  $\theta_1 = r_1\eta$  and  $\boldsymbol{\theta}' = r_2\boldsymbol{\zeta}$ , where  $\eta \in \mathcal{S}^0 = \{-1, 1\}$  and  $\boldsymbol{\zeta}' \in \mathcal{S}^{d-2}$ . Since  $r_1^2 + r_2^2 = 1$  and both  $r_1$  and  $r_2$  are non-negative, introduce  $\psi \in [0, \pi/2]$  such that  $r_1 = \cos \psi$  and  $r_2 = \sin \psi$ . Then Equation (A5) of [Kume and Wood \(2005\)](#) yields

$$d\mathcal{S}^{d-1}(\boldsymbol{\theta}) = (\sin \psi)^{d-2} d\psi d\mathcal{S}^0(\eta) d\mathcal{S}^{d-2}(\boldsymbol{\zeta}).$$

Noting that  $x_1 = \langle \mathbf{x}, \mathbf{u}_1 \rangle = r\langle \boldsymbol{\theta}, \mathbf{u}_1 \rangle = r\theta_1$  on one side, and  $\theta_1 = x_1/r_1 \cos \psi = \eta \cos \psi$  on the other side, thus  $\langle \boldsymbol{\theta}, \mathbf{u}_1 \rangle = \eta \cos \psi$ , and

$$\begin{aligned} J_d^{(2)}(\rho) &= \int_{\mathcal{S}^{d-1}} \frac{d\mathcal{S}^{d-1}(\boldsymbol{\theta})}{(1 + \rho\theta_1)^{d/2}} = \int_{\mathcal{S}^{d-2}} \int_{\mathcal{S}^0} \int_0^{\pi/2} \frac{(\sin \psi)^{d-2}}{(1 + \rho\eta \cos \psi)^{d/2}} d\psi d\mathcal{S}^0(\eta) d\mathcal{S}^{d-2}(\boldsymbol{\zeta}) \\ &= \mathcal{A}^{d-2} \int_{\mathcal{S}^0} \int_0^{\pi/2} \frac{(\sin \psi)^{d-2}}{(1 + \rho\eta \cos \psi)^{d/2}} d\psi d\mathcal{S}^0(\eta) \\ &= \mathcal{A}^{d-2} \int_0^{\pi/2} \frac{(\sin \psi)^{d-2}}{(1 + \rho \cos \psi)^{d/2}} + \frac{(\sin \psi)^{d-2}}{(1 - \rho \cos \psi)^{d/2}} d\psi = \mathcal{A}^{d-2} \int_0^\pi \frac{(\sin \psi)^{d-2}}{(1 + \rho \cos \psi)^{d/2}} d\psi, \end{aligned}$$

where we have successively integrated out, with respect to  $\zeta$ , leading to the surface of  $\mathcal{S}^{d-2}$  denoted by  $\mathcal{A}^{d-2} = 2\pi^{(d-1)/2}/\Gamma((d-1)/2)$ , then to  $\eta$ . As a conclusion,

$$J_d(\rho) = 2^{d/2}\pi^{(d-1)/2} \frac{\Gamma(d/2)}{\Gamma((d-1)/2)} \int_0^\pi \frac{(\sin \psi)^{d-2}}{(1 + \rho \cos \psi)^{d/2}} d\psi.$$

Further, tedious algebra yields:

$$\frac{1}{\Gamma((d-1)/2)} \int_0^\pi \frac{(\sin \psi)^{d-2}}{(1 + \rho \cos \psi)^{d/2}} d\psi = \frac{\sqrt{\pi}}{\Gamma(d/2)\sqrt{1-\rho^2}} \left( \frac{2}{1 + \sqrt{1-\rho^2}} \right)^{\frac{d}{2}-1},$$

and finally  $J_d(\rho) = (2\pi)^{d/2}/c_d(\rho)$ , which concludes the proof.  $\square$

**Lemma A.2.** *If  $a, b, c$  are positive reals such that  $a + b > c$ , then the hypergeometric function satisfies the following asymptotic equivalent when  $z \rightarrow 1$ :*

$${}_2F_1(a, b; c; z) \underset{z \rightarrow 1}{\approx} \frac{\Gamma(a+b-c)\Gamma(c)}{\Gamma(a)\Gamma(b)} \frac{1}{(1-z)^{a+b-c}}.$$

*Proof.* The hypergeometric function admits the following linear combination (Gradshteyn and Ryzhik, 2014, Chapter 7.5):

$$\begin{aligned} {}_2F_1(a, b; c, z) &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} {}_2F_1(a, b; a+b+1-c; 1-z) \\ &+ \frac{\Gamma(a+b-c)\Gamma(c)}{\Gamma(a)\Gamma(b)} (1-z)^{c-a-b} {}_2F_1(c-a, c-b; 1+c-a-b; 1-z). \end{aligned}$$

The result is proved by observing that in the  $z \rightarrow 1$  limit, both hypergeometric functions in the r.h.s. linear combination tend to one.  $\square$

**Lemma A.3.** *Let  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$  be a twice continuously differentiable function such that  $\varphi'(0) = 0$  and  $\varphi''(x) \geq m > 0$  for any  $x \in \mathbb{R}_+$ . Introduce, for any  $t \in \mathbb{R}$  and  $d \geq 2$ ,  $I_d(t) = \int_0^\infty \exp\left(-\frac{t^2}{2}\varphi(x)\right) x^{d-2} dx$ . Then,*

$$I_d(t) \underset{|t| \rightarrow \infty}{\approx} 2^{d-2} \Gamma((d-1)/2) \left( \frac{1}{\varphi''(0)} \right)^{(d-1)/2} \frac{1}{|t|^{d-1}} \exp\left(-\frac{t^2}{2}\varphi(0)\right).$$

*Proof.* Clearly,  $I_d(\cdot)$  is an even function, we thus focus on the situation where  $t > 0$ . Consider for any  $x \geq 0$  the Taylor expansion:

$$\varphi(x) = \varphi(0) + x\varphi'(0) + \frac{x^2}{2}R(x) = \varphi(0) + \frac{x^2}{2}R(x),$$

where  $R(x) = \varphi''(\theta x)$  for some  $\theta = \theta(x) \in (0, 1)$ . By replacement, we get

$$\begin{aligned} I_d(t) &= \exp\left(-\frac{t^2}{2}\varphi(0)\right) \int_0^\infty x^{d-2} \exp\left(-\frac{t^2 x^2}{4}R(x)\right) dx \\ &= \left(\frac{2}{t}\right)^{d-1} \exp\left(-\frac{t^2}{2}\varphi(0)\right) \int_0^\infty v^{d-2} \exp(-v^2 R(2v/t)) dv, \end{aligned}$$

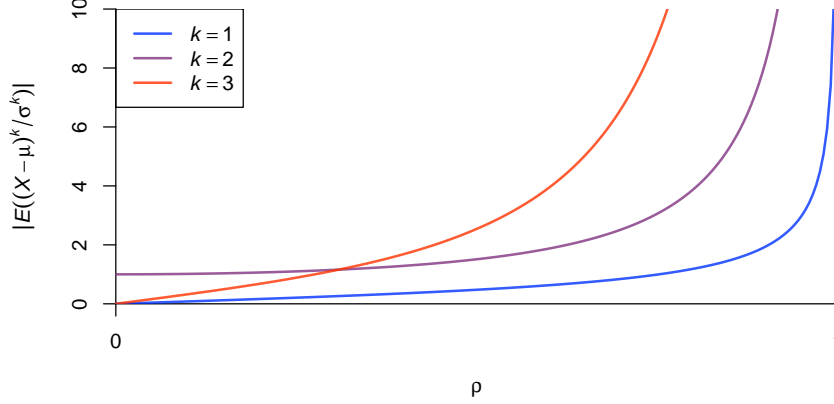


Figure 7: Absolute value of first three moments centered with respect to parameter  $\mu$  for the 1-dimension MED.

using the change of variable  $v = tx/2$ . Since  $R(x) \rightarrow \varphi''(0)$  as  $x \rightarrow 0$  and  $\varphi''$  is lower bounded on  $\mathbb{R}_+$ , Lebesgue dominated convergence theorem thus entails

$$\begin{aligned} I_d(t) &\underset{t \rightarrow +\infty}{\approx} \left(\frac{2}{t}\right)^{d-1} \exp\left(-\frac{t^2}{2}\varphi(0)\right) \int_0^\infty v^{d-2} \exp(-v^2\varphi''(0)) \, dv \\ &= \left(\frac{2}{\varphi''(0)}\right)^{(d-1)/2} \frac{1}{t^{d-1}} \exp\left(-\frac{t^2}{2}\varphi(0)\right) \int_0^\infty u^{d-2} \exp(-u^2/2) \, du. \end{aligned}$$

Finally, classical results regarding the standard Gaussian distribution yield

$$\int_0^\infty u^{d-2} \exp(-u^2/2) \, du = 2^{(d-3)/2} \Gamma((d-1)/2)$$

and the result is proved. □

## B Appendix: Implementation details and additional application results

### B.1 First three moments in dimension $d = 1$

Figure 7 represents the first three moments centered with respect to the  $\mu$  parameter for the 1-dimension MED. Expressions are provided in Section 2.

### B.2 Bayesian implementation in Stan

In this section, we describe an implementation of the Bayesian model of Section 5.2 in Stan probabilistic programming language (Carpenter et al., 2017). Stan starts by describing the data:

```
1 data {
2   int<lower=1> d;           // data dimension
```

```

3 int<lower=0> n;           // number of observations
4 vector<d> Y[n];         // observations
5 }

```

Then parameters are declared. We also adopt the parameterization of Section 5.1, which replaces  $\nu$  in  $\theta$ , by  $\tilde{\nu} = \Sigma^{-1/2}\nu$ . Instead of the covariance matrix, it is more efficient to work with the precision matrix  $\Omega = \Sigma^{-1} \in \text{SPD}(d)$ .

```

1 parameters {
2   vector<d> mu;           // mean
3   cov_matrix<d> Sigma_inv; // precision matrix
4   real<lower=0, upper=1> rho; // asymmetry parameter
5   unit_vector<d> nu_tilde; // direction (tilde version)
6   vector<lower=0><d> diag_sigma_sq; // mu hyperprior cov mat
7 }
8
9 transformed parameters {
10  matrix<d, d> sigma_inv_sqrt = square_root(Sigma_inv, d);
11 }

```

Specific functions used in order to compute the square root of a symmetric and positive definite matrix, compute the log normalizing constant, and define the log p.d.f. of the MED density are then declared:

```

1 functions {
2   // assumes A is SPD
3   matrix square_root(matrix A, int R) {
4     matrix[R, R] eigvecs = eigenvectors_sym(A);
5     vector[R] eigvals = eigenvalues_sym(A);
6     return eigvecs * diag_matrix(sqrt(eigvals)) / eigvecs;
7   }
8
9   real log_norm_const(int d, real rho) {
10    real num = (d-1)*log(2) + log(1+inv_sqrt(1-pow(rho, 2)));
11    real den = d * log(sqrt(1 - rho) + sqrt(1 + rho));
12    return num - den;
13  }
14
15  real mymed_lpdf(vector x, vector m, matrix sigma_inv,
16                 matrix sigma_inv_sqrt, vector nu, real r, int d) {
17    real out = multi_normal_prec_lpdf(x | m, sigma_inv);
18    out -= (log_norm_const(d, r));
19    out -= r / 2 * (x-m)' * sigma_inv_sqrt * nu * sqrt(quad_form(sigma_inv, x
20    -m));
21    return out;
22  }
23 }

```

The Bayesian model is finally complemented with the prior distributions and the likelihood. In the case where no specific expert information is available, the prior of Section 5.2 can be adapted as follows:

```

1 model {
2   mu ~ normal(rep_vector(0, d), diag_sigma_sq);
3   for (j in 1:d) {

```

```

4   diag_sigma_sq[j] ~ inv_gamma(1, 1);
5   }
6   rho ~ uniform(0, 1);
7   Sigma_inv ~ wishart(d, diag_matrix(rep_vector(1, d)));
8   for (i in 1:n) {
9     Y[i] ~ mymed(mu, Sigma_inv, sigma_inv_sqrt, nu_tilde, rho, d);
10  }
11 }

```

### B.3 Bayesian applications: miscellanea

The data distributions used in the Bayesian applications of Section 7.1 and Section 7.2 are described in Table 6. Let us recall that Gumbel copula is defined for all  $(u, v) \in [0, 1]^2$  and  $\theta \geq 1$  by  $C_\theta(u, v) = \exp(-[(-\log u)^\theta + (-\log v)^\theta]^{1/\theta})$ . In particular,  $C_1(u, v) = uv$  is the independence copula, see Nelsen (2007) for further details. Besides,  $\mathcal{SN}$  denotes the skew normal distribution (Azzalini and Capitanio, 1999). Finally, the MLE results of the discriminant analysis of Section 7.3 are provided in Table 7.

Table 6: Underlying data distributions for  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(4)}$  in Section 7.1 and 7.2.

Vector	Copula	$\mathbf{X}_1^{(i)} \sim$	$\mathbf{X}_2^{(i)} \sim$
$\mathbf{X}^{(1)} = (X_1^{(1)}, X_2^{(1)})$	Independence	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 1)$
$\mathbf{X}^{(2)} = (X_1^{(2)}, X_2^{(2)})$	Independence	$\mathcal{SN}(-1, 1, 2)$	$t_4$
$\mathbf{X}^{(3)} = (X_1^{(3)}, X_2^{(3)})$	Gumbel, $\theta = 2$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 1)$
$\mathbf{X}^{(4)} = (X_1^{(4)}, X_2^{(4)})$	Gumbel, $\theta = 2$	$\mathcal{SN}(-1, 1, 2)$	$t_4$

Table 7: Maximum likelihood estimators computed with the genuine and forged records averaged over 100 train samples.

	$\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$	$\boldsymbol{\nu}$	$\rho$
Genuine	$\begin{pmatrix} 3.50 \\ 3.94 \\ 0.09 \\ 0.17 \end{pmatrix}$	$\begin{pmatrix} 3.01 & -1.92 & -1.76 & 0.56 \\ -1.92 & 22.21 & -10.46 & -5.76 \\ -1.76 & -10.46 & 8.92 & 2.97 \\ 0.56 & -5.76 & 2.97 & 2.93 \end{pmatrix}$	$\begin{pmatrix} 0.96 \\ -0.07 \\ -0.65 \\ 1.00 \end{pmatrix}$	0.74
Forged	$\begin{pmatrix} -0.85 \\ -0.36 \\ 0.12 \\ -0.07 \end{pmatrix}$	$\begin{pmatrix} 2.73 & 0.94 & -3.53 & 0.07 \\ 0.94 & 23.94 & -20.65 & -5.39 \\ -3.53 & -20.65 & 21.57 & 4.26 \\ 0.07 & -5.39 & 4.26 & 3.05 \end{pmatrix}$	$\begin{pmatrix} 0.83 \\ 0.53 \\ -1.72 \\ 1.01 \end{pmatrix}$	0.73