Numerical Algorithms (2021) 87:77–118
https://doi.org/10.1007/s11075-020-00959-w

ORIGINAL PAPER

Alternatives to the EM algorithm for ML estimation of location, scatter matrix, and degree of freedom of the Student *t* distribution

Marzieh Hasannasab[1] · **Johannes Hertrich**[1] · **Friederike Laus**[2] · **Gabriele Steidl**[1]

Received: 16 October 2019 / Accepted: 2 June 2020 / Published online: 23 September 2020

Abstract

In this paper, we consider maximum likelihood estimations of the degree of freedom parameter ν, the location parameter μ and the scatter matrix Σ of the multivariate Student t distribution. In particular, we are interested in estimating the degree of freedom parameter ν that determines the tails of the corresponding probability density function and was rarely considered in detail in the literature so far. We prove that under certain assumptions a minimizer of the negative log-likelihood function exists, where we have to take special care of the case $\nu \to \infty$, for which the Student t distribution approaches the Gaussian distribution. As alternatives to the classical EM algorithm we propose three other algorithms which cannot be interpreted as EM algorithm. For fixed ν, the first algorithm is an accelerated EM algorithm known from the literature. However, since we do not fix ν, we cannot apply standard convergence results for the EM algorithm. The other two algorithms differ from this algorithm in the iteration step for ν. We show how the objective function behaves for the different updates of ν and prove for all three algorithms that it decreases in each iteration step. We compare the algorithms as well as some accelerated versions by numerical simulation and apply one of them for estimating the degree of freedom parameter in images corrupted by Student t noise.

The original online version of this article was revised due to a retrospective Open Access order.

✉ Marzieh Hasannasab
hasannas@math.tu-berlin.de

Johannes Hertrich
j.hertrich@math.tu-berlin.de

Gabriele Steidl
steidl@math.tu-berlin.de

[1] Institute of Mathematics, Technische Universität Berlin, Straße des 17. Juni 136, 10623, Berlin, Germany

[2] Technische Universität Kaiserslautern, Paul-Ehrlich-Str. 31, 67663, Kaiserslautern, Germany

Keywords Accelerated EM Algorithm · Student *t* distribution · Image denoising · Robust denoising · ML estimators · Cauchy noise · Nonlocal myriad filters

1 Introduction

The motivation for this work arises from certain tasks in image processing, where the robustness of methods plays an important role. In this context, the Student *t* distribution and the closely related Student *t* mixture models became popular in various image processing tasks. In [31] it has been shown that Student *t* mixture models are superior to Gaussian mixture models for modeling image patches and the authors proposed an application in image compression. Image denoising based on Student *t* models was addressed in [17] and image deblurring in [6, 34]. Further applications include robust image segmentation [4, 25, 29] as well as robust registration [8, 35].

In one dimension and for $\nu = 1$, the Student *t* distribution coincides with the one-dimensional Cauchy distribution. This distribution has been proposed to model a very impulsive noise behavior and one of the first papers which suggested a variational approach in connection with wavelet shrinkage for denoising of images corrupted by Cauchy noise was [3]. A variational method consisting of a data term that resembles the noise statistics and a total variation regularization term has been introduced in [23, 28]. Based on an ML approach the authors of [16] introduced a so-called generalized myriad filter that estimates both the location and the scale parameter of the Cauchy distribution. They used the filter in a nonlocal denoising approach, where for each pixel of the image they chose as samples of the distribution those pixels having a similar neighborhood and replaced the initial pixel by its filtered version. We also want to mention that a unified framework for images corrupted by white noise that can handle (range constrained) Cauchy noise as well was suggested in [14].

In contrast to the above pixelwise replacement, the state-of-the-art algorithm of Lebrun et al. [18] for denoising images corrupted by white Gaussian noise restores the image patchwise based on a maximum a posteriori approach. In the Gaussian setting, their approach is equivalent to minimum mean square error estimation, and more general, the resulting estimator can be seen as a particular instance of a best linear unbiased estimator (BLUE). For denoising images corrupted by additive Cauchy noise, a similar approach was addressed in [17] based on ML estimation for the family of Student *t* distributions, of which the Cauchy distribution forms a special case. The authors call this approach generalized multivariate myriad filter.

However, all these approaches assume that the degree of freedom parameter ν of the Student *t* distribution is known, which might not be the case in practice. In this paper we consider the estimation of the degree of freedom parameter based on an ML approach. In contrast to maximum likelihood estimators of the location and/or scatter parameter(s) μ and Σ, to the best of our knowledge the question of existence of a joint maximum likelihood estimator has not been analyzed before and in this paper we provide first results in this direction. Usually the likelihood function of the Student *t* distributions and mixture models are minimized using the EM algorithm derived e.g. in [13, 21, 22, 26]. For fixed ν, there exists an accelerated EM algorithm [12, 24, 32] which appears to be more efficient than the classical one for smaller parameters ν. We examine the convergence of the accelerated version if also the degree of freedom parameter ν has to be estimated. Also for unknown degrees of freedom, there exist an accelerated version of the EM algorithm, the so-called ECME algorithm [20] which differs from our algorithm. Further, we propose two modifications of the ν iteration step which lead to efficient algorithms for a wide range of parameters ν. Finally, we address further accelerations of our algorithms by the squared iterative methods (SQUAREM) [33] and the damped Anderson acceleration with restarts and ϵ-monotonicity (DAAREM) [9].

The paper is organized as follows: In Section 2 we introduce the Student t distribution, the negative log-likelihood function L and their derivatives. The question of the existence of a minimizer of L is addressed in Section 3. Section 4 deals with the solution of the equation arising when setting the gradient of L with respect to ν to zero. The results of this section will be important for the convergence consideration of our algorithms in the Section 5. We propose three alternatives of the classical EM algorithm and prove that the objective function L decreases for the iterates produced by these algorithms. Finally, we provide two kinds of numerical results in Section 5. First, we compare the different algorithms by numerical examples which indicate that the new ν iterations are very efficient for estimating ν of different magnitudes. Second, we come back to the original motivation of this paper and estimate the degree of freedom parameter ν from images corrupted by one-dimensional Student t noise. The code is provided online[1].

2 Likelihood of the multivariate student *t* distribution

The density function of the d-dimensional Student t distribution $T_\nu(\mu, \Sigma)$ with $\nu > 0$ degrees of freedom, *location* paramter $\mu \in \mathbb{R}^d$ and symmetric, positive definite *scatter matrix* $\Sigma \in$ SPD(d) is given by

$$p(x|\nu,\mu,\Sigma) = \frac{\Gamma\left(\frac{d+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\,\nu^{\frac{d}{2}}\,\pi^{\frac{d}{2}}\,|\Sigma|^{\frac{1}{2}}}\,\frac{1}{\left(1+\frac{1}{\nu}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right)^{\frac{d+\nu}{2}}},$$

with the *Gamma function* $\Gamma(s) := \int_0^\infty t^{s-1}\mathrm{e}^{-t}\,\mathrm{d}t$. The expectation of the Student t distribution is $\mathbb{E}(X) = \mu$ for $\nu > 1$ and the covariance matrix is given by $Cov(X) = \frac{\nu}{\nu-2}\Sigma$ for $\nu > 2$; otherwise, the quantities are undefined. The smaller the value of ν, the heavier the tails of the $T_\nu(\mu, \Sigma)$ distribution. For $\nu \to \infty$, the Student t distribution $T_\nu(\mu, \Sigma)$ converges to the normal distribution $\mathcal{N}(\mu, \Sigma)$ and for $\nu = 0$ it is related to the projected normal distribution on the sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$. Figure 1 illustrates this behavior for the one-dimensional standard Student t distribution.

As the normal distribution, the d-dimensional Student t distribution belongs to the class of *elliptically symmetric distributions*. These distributions are stable under linear transforms in the following sense: Let $X \sim T_\nu(\mu, \Sigma)$ and $A \in \mathbb{R}^{d\times d}$ be an

[1] https://github.com/johertrich/Alternatives-EM-Studentt

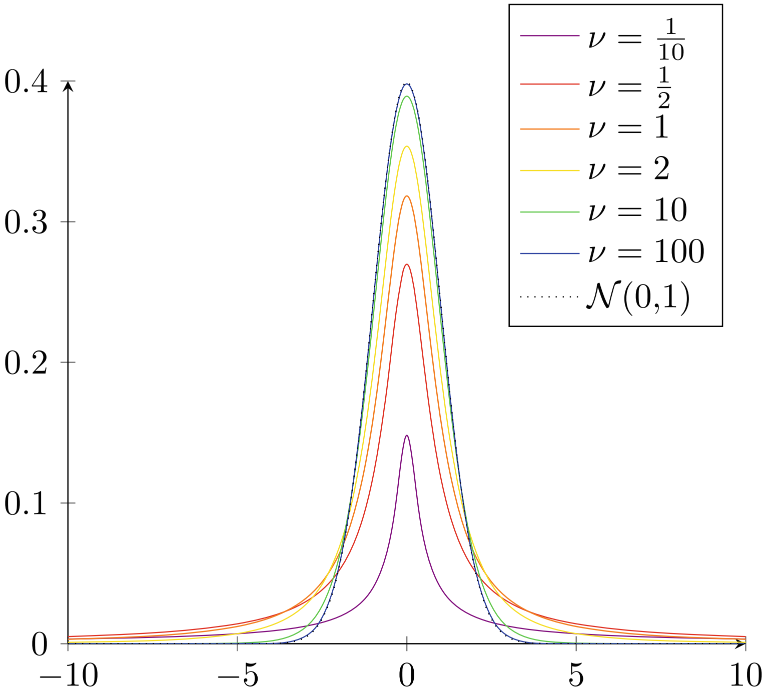


Fig. 1 Standard Student t distribution $T_\nu(0, 1)$ for different values of ν in comparison with the standard normal distribution $\mathcal{N}(0, 1)$

invertible matrix and let $b \in \mathbb{R}^d$. Then $AX+b \sim T_\nu\left(A\mu + b, A\Sigma A^\mathrm{T}\right)$. Furthermore, the Student t distribution $T_\nu(\mu, \Sigma)$ admits the following *stochastic representation*, which can be used to generate samples from $T_\nu(\mu, \Sigma)$ based on samples from the multivariate standard normal distribution $\mathcal{N}(0, I)$ and the Gamma distribution $\Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$: Let $Z \sim \mathcal{N}(0, I)$ and $Y \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ be independent, then

$$X = \mu + \frac{\Sigma^{\frac{1}{2}} Z}{\sqrt{Y}} \sim T_\nu(\mu, \Sigma). \tag{1}$$

For i.i.d. samples $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, the likelihood function of the Student t distribution $T_\nu(\mu, \Sigma)$ is given by

$$\mathcal{L}(\nu,\mu,\Sigma|x_1,\ldots,x_n)=\frac{\Gamma\left(\frac{d+\nu}{2}\right)^n}{\Gamma\left(\frac{\nu}{2}\right)^n(\pi\nu)^{\frac{nd}{2}}\,|\Sigma|^{\frac{n}{2}}}\prod_{i=1}^n\frac{1}{\left(1+\frac{1}{\nu}(x_i-\mu)^\mathrm{T}\Sigma^{-1}(x_i-\mu)\right)^{\frac{d+\nu}{2}}},$$

and the log-likelihood function by

$$\begin{aligned}\ell(\nu, \mu, \Sigma|x_1, \ldots, x_n) \;=\; & n\log\left(\Gamma\left(\tfrac{d+\nu}{2}\right)\right) - n\log\left(\Gamma\left(\tfrac{\nu}{2}\right)\right) - \tfrac{nd}{2}\log(\pi\nu)\\ & -\frac{n}{2}\log|\Sigma| - \tfrac{d+\nu}{2}\sum_{i=1}^n \log\left(1+\frac{1}{\nu}(x_i-\mu)^\mathrm{T}\Sigma^{-1}(x_i-\mu)\right).\end{aligned}$$

In the following, we are interested in the negative log-likelihood function, which up to the factor $\frac{2}{n}$ and weights $w_i = \frac{1}{n}$ reads as

$$\begin{aligned} L(\nu,\mu,\Sigma) &= -2\log\left(\Gamma\left(\tfrac{d+\nu}{2}\right)\right) + 2\log\left(\Gamma\left(\tfrac{\nu}{2}\right)\right) - \nu\log(\nu) \\ &\quad + (d+\nu)\sum_{i=1}^n w_i \log\Big(\nu + (x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)\Big) + \log|\Sigma|. \end{aligned}$$

In this paper, we allow for arbitrary weights from the open probability simplex $\mathring{\Delta}_n := \{w = (w_1,\ldots,w_n) \in \mathbb{R}^n_{>0} : \sum_{i=1}^n w_i = 1\}$. In this way, we might express different levels of confidence in single samples or handle the occurrence of multiple samples. Using $\frac{\partial \log(|X|)}{\partial X} = X^{-1}$ and $\frac{\partial a^{\mathrm{T}}X^{-1}b}{\partial X} = -\big(X^{-\mathrm{T}}\big)ab^{\mathrm{T}}\big(X^{-\mathrm{T}}\big)$ (see [27]), the derivatives of L with respect to μ, Σ and ν are given by

$$\begin{aligned} \frac{\partial L}{\partial \mu}(\nu,\mu,\Sigma) &= -2(d+\nu)\sum_{i=1}^n w_i \frac{\Sigma^{-1}(x_i-\mu)}{\nu + (x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}, \\ \frac{\partial L}{\partial \Sigma}(\nu,\mu,\Sigma) &= -(d+\nu)\sum_{i=1}^n w_i \frac{\Sigma^{-1}(x_i-\mu)(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}}{\nu + (x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)} + \Sigma^{-1}, \\ \frac{\partial L}{\partial \nu}(\nu,\mu,\Sigma) &= \phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu+d}{2}\right) + \sum_{i=1}^n w_i \left(\frac{\nu+d}{\nu + (x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}\right. \\ &\quad \left. - \log\left(\frac{\nu+d}{\nu + (x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}\right) - 1\right), \end{aligned}$$

with

$$\phi(x) := \psi(x) - \log(x), \qquad x > 0$$

and the *digamma function*

$$\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\log(\Gamma(x)) = \frac{\Gamma'(x)}{\Gamma(x)}.$$

Setting the derivatives to zero results in the equations

$$0 = \sum_{i=1}^n w_i \frac{x_i-\mu}{\nu + (x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}, \tag{2}$$

$$I = (d+\nu)\sum_{i=1}^n w_i \frac{\Sigma^{-\frac{1}{2}}(x_i-\mu)(x_i-\mu)^{\mathrm{T}}\Sigma^{-\frac{1}{2}}}{\nu + (x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}, \tag{3}$$

$$\begin{aligned} 0 = F\left(\frac{\nu}{2}\right) &:= \phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu+d}{2}\right) \\ &\quad + \sum_{i=1}^n w_i \left(\tfrac{\nu+d}{\nu+(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)} - \log\left(\tfrac{\nu+d}{\nu+(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}\right) - 1\right). \end{aligned} \tag{4}$$

Computing the trace of both sides of (3) and using the linearity and permutation invariance of the trace operator we obtain

$$\begin{aligned} d &= \operatorname{tr}(I) = (d+\nu)\sum_{i=1}^{n} w_i \frac{\operatorname{tr}\left(\Sigma^{-\frac{1}{2}}(x_i-\mu)(x_i-\mu)^{\mathrm{T}}\Sigma^{-\frac{1}{2}}\right)}{\nu+(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)} \\ &= (d+\nu)\sum_{i=1}^{n} w_i \frac{(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}{\nu+(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}, \end{aligned}$$

which yields

$$1 = (d+\nu)\sum_{i=1}^{n} w_i \frac{1}{\nu+(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)}.$$

We are interested in critical points of the negative log-likelihood function L, i.e., in solutions (μ, Σ, ν) of (2)–(4), and in particular in minimizers of L.

3 Existence of critical points

In this section, we examine whether the negative log-likelihood function L has a minimizer, where we restrict our attention to the case $\mu = 0$. For an approach how to extend the results to arbitrary μ for fixed ν we refer to [17]. To the best of our knowledge, this is the first work that provides results in this direction. The question of existence is, however, crucial in the context of ML estimation, since it lays the foundation for any convergence result for the EM algorithm or its variants. In fact, the authors of [13] observed the divergence of the EM algorithm in some of their numerical experiments, which is in accordance with our observations.

For *fixed* $\nu > 0$, it is known that there exists a unique solution of (3) and for $\nu = 0$ that there exist solutions of (3) which differ only by a multiplicative positive constant (see, e.g., [17]). In contrast, if we do not fix ν, we have roughly to distinguish between the two cases that the samples tend to come from a Gaussian distribution, i.e., $\nu \to \infty$, or not. The results are presented in Theorem 1.

We make the following general assumption:

Assumption 1 *Any subset of less or equal d samples x_i, $i \in \{1, \ldots, n\}$ is linearly independent and* $\max\{w_i : i = 1, \ldots, n\} < \frac{1}{d}$.

For $\mu = 0$, the negative log-likelihood function becomes

$$\begin{aligned} L(\nu, \Sigma) &:= -2\log\left(\Gamma\left(\frac{d+\nu}{2}\right)\right) + 2\log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \nu\log(\nu) \\ &\quad + (d+\nu)\sum_{i=1}^{n} w_i \log\left(\nu + x_i^{\mathrm{T}}\Sigma^{-1}x_i\right) + \log(|\Sigma|) \\ &= -2\log\left(\Gamma\left(\frac{d+\nu}{2}\right)\right) + 2\log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \nu\log(\nu) \\ &\quad + (d+\nu)\log(\nu) + (d+\nu)\sum_{i=1}^{n} w_i \log\left(1 + \frac{1}{\nu}x_i^{\mathrm{T}}\Sigma^{-1}x_i\right) + \log(|\Sigma|). \end{aligned}$$

Further, for a fixed $\nu > 0$, set

$$L_\nu(\Sigma) := (d+\nu)\sum_{i=1}^n w_i \log\left(\nu + x_i^{\mathrm{T}}\Sigma^{-1}x_i\right) + \log(|\Sigma|).$$

To prove the next existence theorem we will need two lemmas, whose proofs are given in the Appendix.

Theorem 1 *Let $x_i \in \mathbb{R}^d$, $i = 1,\dots,n$ and $w \in \mathring{\Delta}_n$ fulfill Assumption 1. Then exactly one of the following statements holds:*

- (i) *There exists a minimizing sequence $(\nu_r, \Sigma_r)_r$ of L, such that $\{\nu_r : r \in \mathbb{N}\}$ has a finite cluster point. Then we have $argmin_{(\nu,\Sigma)\in\mathbb{R}_{>0}\times\mathrm{SPD}(d)} L(\nu,\Sigma) \neq \emptyset$ and every $(\hat\nu,\hat\Sigma) \in argmin_{(\nu,\Sigma)\in\mathbb{R}_{>0}\times\mathrm{SPD}(d)} L(\nu,\Sigma)$ is a critical point of L.*
- (ii) *For every minimizing sequence $(\nu_r, \Sigma_r)_r$ of $L(\nu,\Sigma)$ we have $\lim_{r\to\infty}\nu_r = \infty$. Then $(\Sigma_r)_r$ converges to the maximum likelihood estimator $\hat\Sigma = \sum_{i=1}^n w_i x_i x_i^{\mathrm{T}}$ of the normal distribution $\mathcal{N}(0,\Sigma)$.*

Proof **Case 1:** Assume that there exists a minimizing sequence $(\nu_r, \Sigma_r)_r$ of L, such that $(\nu_r)_r$ has a bounded subsequence. In particular, using Lemma 4, we have that $(\nu_r)_r$ has a cluster point $\nu^* > 0$ and a subsequence $(\nu_{r_k})_k$ converging to ν^*. Clearly, the sequence $(\nu_{r_k}, \Sigma_{r_k})_k$ is again a minimizing sequence so that we skip the second index in the following. By Lemma 5, the set $\overline{\{\Sigma_r : r \in \mathbb{N}\}}$ is a compact subset of $\mathrm{SPD}(d)$. Therefore there exists a subsequence $(\Sigma_{r_k})_k$ which converges to some $\Sigma^* \in \mathrm{SPD}(d)$. Now we have by continuity of $L(\nu,\Sigma)$ that

$$L(\nu^*,\Sigma^*) = \lim_{k\to\infty} L(\nu_{r_k},\Sigma_{r_k}) = \min_{(\nu,\Sigma)\in\mathbb{R}_{>0}\times\mathrm{SPD}(d)} L(\nu,\Sigma).$$

Case 2: Assume that for every minimizing sequence $(\nu_r,\Sigma_r)_r$ it holds that $\nu_r \to \infty$ as $r \to \infty$. We rewrite the likelihood function as

$$\begin{aligned} L(\nu,\Sigma) &= 2\log\left(\frac{\Gamma\left(\frac{\nu}{2}\right)\frac{\nu}{2}^{\frac{d}{2}}}{\Gamma\left(\frac{d+\nu}{2}\right)}\right) + d\log(2) \\ &\quad + (d+\nu)\sum_{i=1}^n w_i \log\left(1+\frac{1}{\nu}x_i^{\mathrm{T}}\Sigma^{-1}x_i\right) + \log(|\Sigma|). \end{aligned}$$

Since

$$\lim_{\nu\to\infty}\frac{\Gamma\left(\frac{\nu}{2}\right)\frac{\nu}{2}^{\frac{d}{2}}}{\Gamma\left(\frac{d+\nu}{2}\right)} = 1,$$

we obtain

$$\lim_{r\to\infty} L(\nu_r,\Sigma_r) = d\log(2) + \lim_{\nu_r\to\infty}(d+\nu_r)\sum_{i=1}^n w_i\log\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right) + \log(|\Sigma_r|). \tag{5}$$

Next we show by contradiction that $\overline{\{\Sigma_r : r \in \mathbb{N}\}}$ is in $\mathrm{SPD}(d)$ and bounded: Denote the eigenvalues of Σ_r by $\lambda_{r1} \geq \cdots \geq \lambda_{rd}$. Assume that either $\{\lambda_{r1} : r \in \mathbb{N}\}$ is unbounded or that $\{\lambda_{rd} : r \in \mathbb{N}\}$ has zero as a cluster point. Then, we know by [17, Theorem 4.3] that there exists a subsequence of $(\Sigma_r)_r$, which we again denote by $(\Sigma_r)_r$, such that for any fixed $\nu > 0$ it holds

$$\lim_{r\to\infty} L_\nu(\Sigma_r) = \infty.$$

Since $k \mapsto \left(1+\frac{k}{x}\right)^k$ is monotone increasing, for $\nu_r \geq d+1$ we have

$$\begin{aligned}(d+\nu_r)\sum_{i=1}^n w_i \log\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right) &= \sum_{i=1}^n w_i \log\left(\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)^{\nu_r+d}\right)\\ &\geq \sum_{i=1}^n w_i \log\left(\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)^{\nu_r}\right)\\ &\geq \sum_{i=1}^n w_i \log\left(\left(1+\frac{1}{d+1}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)^{d+1}\right)\\ &= (d+1)\sum_{i=1}^n w_i \log\left(1+\frac{1}{d+1}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)\\ &\geq (d+1)\sum_{i=1}^n w_i \log\left(1+x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)\\ &\quad -\log(d+1)^{d+1}.\end{aligned}$$

By (5) this yields

$$\begin{aligned}\lim_{r\to\infty} L(\nu_r,\Sigma_r) &\geq d\log(2) - \log(d+1)^{d+1}\\ &\quad + \lim_{r\to\infty}(d+1)\sum_{i=1}^n w_i \log\left(1+x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right) + \log(|\Sigma_r|)\\ &= d\log(2) - \log(d+1)^{d+1} + \lim_{r\to\infty} L_1(\Sigma_r) = \infty.\end{aligned}$$

This contradicts the assumption that $(\nu_r, \Sigma_r)_r$ is a minimizing sequence of L. Hence, $\{\Sigma_r : r \in \mathbb{N}\}$ is a bounded subset of $\mathrm{SPD}(d)$.

Finally, we show that any subsequence of $(\Sigma_r)_r$ has a subsequence which converges to $\hat{\Sigma} = \sum_{i=1}^n w_i x_i x_i^{\mathrm{T}}$. Then the whole sequence $(\Sigma_r)_r$ converges to $\hat{\Sigma}$.
Let $\left(\Sigma_{r_k}\right)_k$ be a subsequence of $(\Sigma_r)_r$. Since it is bounded, it has a convergent subsequence $\left(\Sigma_{r_{k_l}}\right)_l$ which converges to some $\tilde{\Sigma} \in \overline{\{\Sigma_r : r \in \mathbb{N}\}} \subset \mathrm{SPD}(d)$. For simplicity, we denote $\left(\Sigma_{r_{k_l}}\right)_l$ again by $(\Sigma_r)_r$. Since $(\Sigma_r)_r$ is converges, we know that also $\left(x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)_r$ converges and is bounded. By $\lim_{r\to\infty}\nu_r = \infty$ we know that the functions $x \mapsto \left(1+\frac{x}{\nu_r}\right)^{\nu_r}$ converge locally uniformly to $x \mapsto \exp(x)$ as $r \to \infty$.

Thus we obtain

$$\begin{aligned}
&\lim_{r\to\infty}(d+\nu_r)\sum_{i=1}^n w_i \log\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)\\
&\quad=\lim_{r\to\infty}\sum_{i=1}^n w_i \log\left(\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)^{d+\nu_r}\right)\\
&\quad=\lim_{r\to\infty}\sum_{i=1}^n w_i \log\left(\lim_{r\to\infty}\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)^{\nu_r}\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)^{d}\right)\\
&\quad=\lim_{r\to\infty}\sum_{i=1}^n w_i \log\left(\lim_{r\to\infty}\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)^{\nu_r}\right)\\
&\quad=\sum_{i=1}^n w_i \log\left(\exp\left(x_i^{\mathrm{T}}\tilde{\Sigma}^{-1}x_i\right)\right)=\sum_{i=1}^n w_i x_i^{\mathrm{T}}\tilde{\Sigma}^{-1}x_i.
\end{aligned}$$

Hence, we have

$$\inf_{(\nu,\Sigma)\in\mathbb{R}_{>0}\times\mathrm{SPD}(d)} L(\nu,\Sigma)=\lim_{r\to\infty}L(\nu_r,\Sigma_r)=d\log(2)+\sum_{i=1}^n w_i x_i^{\mathrm{T}}\tilde{\Sigma}^{-1}x_i+\log\left(\left|\tilde{\Sigma}\right|\right).$$

By taking the derivative with respect to Σ we see that the right-hand side is minimal if and only if $\Sigma=\hat{\Sigma}=\sum_{i=1}^n w_i x_i x_i^{\mathrm{T}}$. On the other hand, by similar computations as above we get

$$\begin{aligned}
\inf_{(\nu,\Sigma)\in\mathbb{R}_{>0}\times\mathrm{SPD}(d)} L(\nu,\Sigma)\ &\le\ \lim_{r\to\infty}L\left(\nu_r,\hat{\Sigma}\right)\\
&=\ d\log(2)+\log\left(\left|\hat{\Sigma}\right|\right)\\
&\qquad+\lim_{\nu_r\to\infty}(d+\nu_r)\sum_{i=1}^n w_i\log\left(1+\frac{1}{\nu_r}x_i^{\mathrm{T}}\hat{\Sigma}^{-1}x_i\right)\\
&=\ d\log(2)+\log\left(\left|\hat{\Sigma}\right|\right)+\sum_{i=1}^n w_i x_i^{\mathrm{T}}\hat{\Sigma}^{-1}x_i+\log\left(\left|\hat{\Sigma}\right|\right),
\end{aligned}$$

so that $\tilde{\Sigma}=\hat{\Sigma}$. This finishes the proof. □

4 Zeros of *F*

In this section, we are interested in the existence of solutions of (4), i.e., in zeros of F for arbitrary fixed μ and Σ. Setting $x:=\frac{\nu}{2}>0$, $t:=\frac{d}{2}$ and

$$s_i:=\frac{1}{2}(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu),\quad i=1,\ldots,n.$$

we rewrite the function F in (4) as

$$\begin{aligned} F(x) &= \phi(x) - \phi(x+t) + \sum_{i=1}^{n} w_i \left(\frac{x+t}{x+s_i} - \log\left(\frac{x+t}{x+s_i}\right) - 1 \right) \\ &= \sum_{i=1}^{n} w_i F_{s_i}(x) = \sum_{i=1}^{n} w_i \left(A(x) + B_{s_i}(x) \right), \end{aligned} \tag{6}$$

where

$$F_s(x) := A(x) + B_s(x) \tag{7}$$

and

$$A(x) := \phi(x) - \phi(x+t), \qquad B_s(x) := \frac{x+t}{x+s} - \log\left(\frac{x+t}{x+s}\right) - 1.$$

The digamma function ψ and $\phi = \psi - \log(\cdot)$ are well examined in the literature (see [1]). The function $\phi(x)$ is the expectation value of a random variable which is $\Gamma(x, x)$ distributed. It holds $-\frac{1}{x} < \phi(x) < -\frac{1}{2x}$ and it is well-known that $-\phi$ is *completely monotone*. This implies that the negative of A is also completely monotone, i.e., for all $x > 0$ and $m \in \mathbb{N}_0$ we have

$$(-1)^{m+1}\phi^{(m)}(x) > 0, \qquad (-1)^{m+1}A^{(m)}(x) > 0,$$

in particular $A < 0$, $A' > 0$ and $A'' < 0$. Further, it is easy to check that

$$\lim_{x\to 0} \phi(x) = -\infty, \qquad \lim_{x\to\infty} \phi(x) = 0^-, \tag{8}$$

$$\lim_{x\to 0} A(x) = -\infty, \qquad \lim_{x\to\infty} A(x) = 0^-. \tag{9}$$

On the other hand, we have that $B(x) \equiv 0$ if $s = t$ in which case $F_s = A < 0$ and has therefore no zero. If $s \neq t$, then B_s is *completely monotone*, i.e., for all $x > 0$ and $m \in \mathbb{N}_0$,

$$(-1)^m B_s^{(m)}(x) > 0,$$

in particular $B_s > 0$, $B_s' < 0$ and $B_s'' > 0$, and

$$B_s(0) = \frac{t}{s} - \log\left(\frac{t}{s}\right) - 1 > 0, \qquad \lim_{x\to\infty} B_s(x) = 0^+.$$

Hence, we have

$$\lim_{x\to 0} F_s(x) = -\infty, \qquad \lim_{x\to\infty} F_s(x) = 0. \tag{10}$$

If $X \sim \mathcal{N}(\mu, \Sigma)$ is a d-dimensional random vector, then $Y := (X-\mu)^{\mathrm{T}}\Sigma^{-1}(X-\mu) \sim \chi^2_d$ with $\mathbb{E}(Y) = d$ and $Var(Y) = 2d$. Thus, we would expect that for samples x_i from such a random variable X the corresponding values $(x_i-\mu)^{\mathrm{T}}\Sigma^{-1}(x_i-\mu)$ lie with high probability in the interval $[d-\sqrt{2d}, d+\sqrt{2d}]$, respective $s_i \in [t-\sqrt{t}, t+\sqrt{t}]$. These considerations are reflected in the following theorem and corollary.

Theorem 2 *For $F_s : \mathbb{R}_{>0} \to \mathbb{R}$ given by (7) the following relations hold true:*

i) If $s \in [t-\sqrt{t}, t+\sqrt{t}] \cap \mathbb{R}_{>0}$, then $F_s(x) < 0$ for all $x > 0$ so that F_s has no zero.

ii) *If $s > 0$ and $s \notin [t - \sqrt{t}, t + \sqrt{t}]$, then there exists x_+ such that $F_s(x) > 0$ for all $x \geq x_+$. In particular, F_s has a zero.*

Proof We have

$$\begin{aligned} F_s'(x) &= \phi'(x) - \phi'(x+t) - \frac{(s-t)^2}{(x+s)^2(x+t)} \\ &= \psi'(x) - \psi'(x+t) - \frac{t}{x(x+t)} - \frac{(s-t)^2}{(x+s)^2(x+t)}. \end{aligned}$$

We want to sandwich F_s' between two rational functions P_s and $P_s + Q$ which zeros can easily be described.

Since the trigamma function ψ' has the series representation

$$\psi'(x) = \sum_{k=0}^{\infty} \frac{1}{(x+k)^2},$$

see [1], we obtain

$$F_s'(x) = \sum_{k=0}^{\infty} \frac{1}{(x+k)^2} - \frac{1}{(x+k+t)^2} - \frac{t}{x(x+t)} - \frac{(s-t)^2}{(x+s)^2(x+t)}. \tag{11}$$

For $x > 0$, we have

$$I(x) = \int_0^{\infty} \underbrace{\frac{1}{(x+u)^2} - \frac{1}{(x+u+t)^2}}_{g(u)} \, du = \frac{1}{x} - \frac{1}{x+t} = \frac{t}{(x+t)x}.$$

Let $R(x)$ and $T(x)$ denote the rectangular and trapezoidal rule, respectively, for computing the integral with step size 1. Then, we verify

$$R(x) = \sum_{k=0}^{\infty} g(k) = \sum_{k=0}^{\infty} \frac{1}{(x+k)^2} - \frac{1}{(x+k+t)^2}$$

so that

$$\begin{aligned} F_s'(x) &= (R(x) - T(x)) + (T(x) - I(x)) - \frac{(s-t)^2}{(x+s)^2(x+t)} \\ &= \frac{1}{2}\left(\frac{1}{x^2} - \frac{1}{(x+t)^2}\right) + (T(x) - I(x)) - \frac{(s-t)^2}{(x+s)^2(x+t)}. \end{aligned}$$

By considering the first and second derivatives of g we see the integrand in $I(x)$ is strictly decreasing and strictly convex. Thus, $P_s(x) < F_s'(x)$, where

$$\begin{aligned} P_s(x) &:= \frac{1}{2}\left(\frac{1}{x^2} - \frac{1}{(x+t)^2}\right) - \frac{(s-t)^2}{(x+s)^2(x+t)} \\ &= \frac{(2tx + t^2)(x+s)^2 - (s-t)^2 x^2 (x+t)}{2x^2(x+s)^2(x+t)^2} \\ &= \frac{p_s(x)}{2x^2(x+s)^2(x+t)^2}. \end{aligned}$$

with $p_s(x) := a_3x^3 + a_2x^2 + a_1x + a_0$ and

$$\begin{aligned} a_0 &= t^2s^2 > 0, & a_1 &= 2st(s+t) > 0, \\ a_2 &= t\left(4s + t - (s-t)^2\right), & a_3 &= 2\left(t - (s-t)^2\right). \end{aligned}$$

For $t \geq 1$, we have

$$a_3 \geq 0 \quad \Longleftrightarrow \quad s \in [t - \sqrt{t}, t + \sqrt{t}] \tag{12}$$

and

$$a_2 \geq 0 \quad \Longleftrightarrow \quad s \in [t + 2 - \sqrt{4+5t}, t + 2 + \sqrt{4+5t}] \supset [t - \sqrt{t}, t + \sqrt{t}].$$

For $t = \frac{1}{2}$, it holds $[t + 2 - \sqrt{4+5t}, t + 2 + \sqrt{4+5t}] \supset [0, t + \sqrt{t}]$.

Thus, for $s \in [t - \sqrt{t}, t + \sqrt{t}]$, by the sign rule of Descartes, $p_s(x)$ has no positive zero which implies

$$0 \leq P_s(x) < F_s'(x) \quad \text{for} \quad s \in [t - \sqrt{t}, t + \sqrt{t}] \cap \mathbb{R}_{>0}.$$

Hence, the continuous function F_s is monotone increasing and by (10) we obtain $F_s(x) < 0$ for all $x > 0$ if $s \in [t - \sqrt{t}, t + \sqrt{t}] \cap \mathbb{R}_{>0}$.

Let $s > 0$ and $s \notin [t - \sqrt{t}, t + \sqrt{t}]$. By

$$T(x) - I(x) = \sum_{k=0}^{\infty} \left(\frac{1}{2}(g(k+1) + g(k)) - \int_0^1 g(k+u)\, du \right)$$

and Euler's summation formula, we obtain

$$T(x) - I(x) = \sum_{k=0}^{\infty} \frac{1}{12}\left(g'(k+1) - g'(k)\right) - \frac{1}{720} g^{(4)}(\xi_k), \quad \xi_k \in (k, k+1)$$

with $g'(u) = -\frac{2}{(x+u)^3} + \frac{2}{(x+u+t)^3}$ and $g^{(4)}(u) = \frac{5!}{(x+u)^6} - \frac{5!}{(x+u+t)^6}$, so that

$$\begin{aligned} T(x) - I(x) &= -\frac{1}{12} g'(0) + \sum_{k=0}^{\infty} \frac{1}{6} \frac{1}{(x + \xi_k + t)^6} - \frac{1}{6} \frac{1}{(x + \xi_k)^6} \\ &< -\frac{1}{12} g'(0) = \frac{1}{6} \frac{3tx^2 + 3t^2x + t^3}{x^3(x+t)^3}. \end{aligned} \tag{13}$$

Therefore, we conclude

$$F_s'(x) < P_s(x) + \underbrace{\frac{1}{6} \frac{3tx^2 + 3t^2x + t^3}{x^3(x+t)^3}}_{Q(x)} = \frac{p_s(x)x(x+t) + (tx^2 + t^2x + \frac{1}{3}t^3)(x+s)^2}{2x^3(x+s)^2(x+t)^3}.$$

The main coefficient of x^5 of the polynomial in the numerator is $2\left(t - (s-t)^2\right)$ which fulfills (12). Therefore, if $s \notin [t - \sqrt{t}, t + \sqrt{t}]$, then there exists x_+ large enough such that the numerator becomes smaller than zero for all $x \geq x_+$. Consequently, $F_s'(x) \leq P_s(x) + Q(x) < 0$ for all $x \geq x_+$. Thus, F_s is decreasing on $[x_+, \infty)$. By (10), we conclude that F_s has a zero. □

The following corollary states that F_s has exactly one zero if $s > t + \sqrt{t}$. Unfortunately we do not have such a results for $s < t - \sqrt{t}$.

Corollary 1 *Let $F_s : \mathbb{R}_{>0} \to \mathbb{R}$ be given by (7). If $s > t + \sqrt{t}$, $t \geq 1$, then F_s has exactly one zero.*

Proof By Theorem 2ii) and since $\lim_{x\to 0} F_s(x) = -\infty$ and $\lim_{x\to\infty} = 0^+$, it remains to prove that F_s' has at most one zero. Let $x_0 > 0$ be the smallest number such that $F_s'(x_0) = 0$. We prove that $F_s'(x) < 0$ for all $x > x_0$. To this end, we show that $h_s(x) := F_s'(x)(x+s)^2(x+t)$ is strictly decreasing. By (11) we have

$$h_s(x) = (x+s)^2(x+t)\left(\sum_{k=0}^{\infty} \frac{1}{(x+k)^2} - \frac{1}{(x+k+t)^2} - \frac{t}{x(x+t)}\right) - (s-t)^2,$$

and for $s > t$ further

$$\begin{aligned} h_s'(x) &= \Big(2(x+s)(x+t) + (x+s)^2\Big)\left(\sum_{k=0}^{\infty} \frac{1}{(x+k)^2} - \frac{1}{(x+k+t)^2} - \frac{t}{x(x+t)}\right) \\ &\quad + (x+s)^2(x+t)\left(\sum_{k=0}^{\infty} \frac{-2}{(x+k)^3} + \frac{2}{(x+k+t)^3} + \frac{t(2x+t)}{x^2(x+t)^2}\right) \\ &\leq 3(x+s)^2\left(\sum_{k=0}^{\infty} \frac{1}{(x+k)^2} - \frac{1}{(x+k+t)^2} - \frac{t}{x(x+t)}\right) \\ &\quad + (x+s)^2(x+t)\left(\sum_{k=0}^{\infty} \frac{-2}{(x+k)^3} + \frac{2}{(x+k+t)^3} + \frac{t(2x+t)}{x^2(x+t)^2}\right) \\ &= (x+s)^2(R(x) - I(x)), \end{aligned}$$

where $I(x)$ is the integral and $R(x)$ the corresponding rectangular rule with step size 1 of the function $g := g_1 + g_2$ defined as

$$\begin{aligned} g_1(u) &:= 3\left(\frac{1}{(x+u)^2} - \frac{1}{(x+t+u)^2}\right), \\ g_2(u) &:= (x+t)\left(\frac{-2}{(x+u)^3} + \frac{2}{(x+t+u)^3}\right). \end{aligned}$$

We show that $R(x) - I(x) < 0$ for all $x > 0$. Let $T(x)$, $T_i(x)$ be the trapezoidal rules with step size 1 corresponding to $I(x)$ and $I_i(x) = \int_0^\infty g_i(u)du$, $i = 1, 2$. Then it follows

$$R(x) - I(x) = R(x) - T(x) + T(x) - I(x) = R(x) - T(x) + T_1(x) - I_1(x) + T_2(x) - I_2(x).$$

Since g_2 is a decreasing, concave function, we conclude $T_2(x) - I_2(x) < 0$. Using Euler's summation formula in (13) for g_1, we get

$$T_1(x) - I_1(x) = -\frac{1}{12} g_1'(0) - \frac{1}{720}\sum_{k=0}^{\infty} g_1^{(4)}(\xi_k), \quad \xi_k \in (k, k+1).$$

Since $g_1^{(4)}$ is a positive function, we can write

$$\begin{aligned} R(x) - I(x) &< R(x) - T(x) + T_1(x) - I_1(x) \leq \frac{1}{2} g(0) - \frac{1}{12} g_1'(0) \\ &= \frac{3}{2} \left(\frac{1}{x^2} - \frac{1}{(x+t)^2} \right) + \frac{1}{2}(x+t) \left(\frac{-2}{x^3} + \frac{2}{(x+t)^3} \right) \\ &\quad - \frac{1}{2} \left(\frac{-1}{x^3} + \frac{1}{(x+t)^3} \right) \\ &= \frac{t}{2} \frac{(-3t+3)x^2 + \left(-5t^2 + 3t\right) x - 2t^3 + t^2}{x^3 (x+t)^3}. \end{aligned}$$

All coefficients of x are smaller or equal than zero for $t \geq 1$ which implies that h_s is strictly decreasing. □

Theorem 2 implies the following corollary.

Corollary 2 *For $F : \mathbb{R}_{>0} \to \mathbb{R}$ given by (6) and $\delta_i := (x_i - \mu)^{\mathrm{T}} \Sigma^{-1} (x_i - \mu)$, $i = 1, \ldots, n$, the following relations hold true:*

i) If $\delta_i \in [d - \sqrt{2d}, d + \sqrt{2d}] \cap \mathbb{R}_{>0}$ for all $i \in \{1, \ldots, n\}$, then $F(x) < 0$ for all $x > 0$ so that F has no zero.

ii) If $\delta_i > 0$ and $\delta_i \notin [d - \sqrt{2d}, d + \sqrt{2d}]$ for all $i \in \{1, \ldots, n\}$, there exists x_+ such that $F(x) > 0$ for all $x \geq x_+$. In particular, F has a zero.

Proof Consider $F = \sum_{i=1}^n F_{s_i}$. If $\delta_i \in [d - \sqrt{2d}, d + \sqrt{2d}] \cap \mathbb{R}_{>0}$ for all $i \in \{1, \ldots, n\}$, then we have by Theorem 2 that $F_{s_i}(x) < 0$ for all $x > 0$. Clearly, the same holds true for the whole function F such that it cannot have a zero.

If $\delta_i \notin [d - \sqrt{2d}, d + \sqrt{2d}]$ for all $i \in \{1, \ldots, n\}$, then we know by Theorem 2 that there exist $x_{i+} > 0$ such that $F_{s_i}(x) > 0$ for $x \geq x_{i+}$. Thus, $F(x) > 0$ for $x \geq x_+ := \max_i(x_{i+})$. Since $\lim_{x \to 0} F(x) = -\infty$ this implies that F has a zero. □

5 Algorithms

In this section, we propose an alternative of the classical EM algorithm for computing the parameters of the Student t distribution along with convergence results. In particular, we are interested in estimating the degree of freedom parameter ν, where the function F is of particular interest.

Algorithm 1 with weights $w_i = \frac{1}{n}$, $i = 1, \ldots, n$, is the classical EM algorithm. Note that the function in the third M-Step

$$\Phi_r\left(\frac{\nu}{2}\right) := \phi\left(\tfrac{\nu}{2}\right) \underbrace{- \phi\left(\frac{\nu_r + d}{2}\right) + \sum_{i=1}^n w_i \left(\gamma_{i,r} - \log(\gamma_{i,r}) - 1\right)}_{c_r}$$

has a unique zero since by (8) the function $\phi < 0$ is monotone increasing with $\lim_{x \to \infty} \phi(x) = 0^-$ and $c_r > 0$. Concerning the convergence of the EM algorithm it is known that the values of the objective function $L(\nu_r, \mu_r, \Sigma_r)$ are monotone decreasing in r and that a subsequence of the iterates converges to a critical point of $L(\nu, \mu, \Sigma)$ if such a point exists, see [5].

Algorithm 1 EM Algorithm (EM).

```
Input: $x_1, \dots, x_n \in \mathbb{R}^d$, $n \geq d+1$, $w \in \mathring{\Delta}_n$
Initialization: $\nu_0 = \varepsilon > 0$, $\mu_0 = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\Sigma_0 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)(x_i - \mu_0)^{\mathrm{T}}$
for $r = 0, \dots$
    E-Step: Compute the weights
        $\delta_{i,r} = (x_i - \mu_r)^{\mathrm{T}} \Sigma_r^{-1} (x_i - \mu_r)$
        $\gamma_{i,r} = \frac{\nu_r + d}{\nu_r + \delta_{i,r}}$
    M-Step: Update the parameters
        $\mu_{r+1} = \frac{\sum_{i=1}^{n} w_i \gamma_{i,r} x_i}{\sum_{i=1}^{n} w_i \gamma_{i,r}}$
        $\Sigma_{r+1} = \sum_{i=1}^{n} w_i \gamma_{i,r} (x_i - \mu_{r+1})(x_i - \mu_{r+1})^{\mathrm{T}}$
        $\nu_{r+1} = \text{zero of } \phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu_r + d}{2}\right) + \sum_{i=1}^{n} w_i \left(\gamma_{i,r} - \log(\gamma_{i,r}) - 1\right)$
```

Algorithm 2 distinguishes from the EM algorithm in the iteration of Σ, where the factor $\frac{1}{\sum_{i=1}^{n} w_i \gamma_{i,r}}$ is incorporated now. The computation of this factor requires no additional computational effort, but speeds up the performance in particular for smaller ν. Such kind of acceleration was suggested in [12, 24]. *For fixed* $\nu \geq 1$, it was shown in [32] that this algorithm is indeed an EM algorithm arising from another choice of the hidden variable than used in the standard approach, see also [15]. Thus, it follows for fixed $\nu \geq 1$ that the sequence $L(\nu, \mu_r, \Sigma_r)$ is monotone decreasing. However, we also iterate over ν. In contrast to the EM Algorithm 1 our ν iteration step depends on μ_{r+1} and Σ_{r+1} instead of μ_r and Σ_r. This is important for our convergence results. Note that for both cases, the accelerated algorithm can no longer be interpreted as an EM algorithm, so that the convergence results of the classical EM approach are no longer available.

Let us mention that a Jacobi variant of Algorithm 2 for *fixed* ν, i.e.,

$$\Sigma_{r+1} = \sum_{i=1}^{n} \frac{w_i \gamma_{i,r} (x_i - \mu_r)(x_i - \mu_r)^{\mathrm{T}}}{\sum_{i=1}^{n} w_i \gamma_{i,r}},$$

with μ_r instead of μ_{r+1} including a convergence proof was suggested in [17]. The main reason for this index choice was that we were able to prove monotone convergence of a simplified version of the algorithm for estimating the location and scale of Cauchy noise ($d = 1$, $\nu = 1$) which could be not achieved with the variant incorporating μ_{r+1} (see [16]). This simplified version is known as myriad filter in image processing. In this paper, we keep the original variant from the EM algorithm (14) since we are mainly interested in the computation of ν.

Instead of the above algorithms we suggest to take the critical point (4) more directly into account in the next two algorithms.

Algorithm 2 Accelerated EM-like Algorithm (aEM).

Same as Algorithm 1 except for

$$\Sigma_{r+1} = \sum_{i=1}^{n} \frac{w_i \gamma_{i,r} (x_i - \mu_{r+1})(x_i - \mu_{r+1})^{\mathrm{T}}}{\sum_{i=1}^{n} w_i \gamma_{i,r}} \tag{14}$$

$$\nu_{r+1} = \text{zero of } n$$

$$\phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu_r + d}{2}\right) + \sum_{i=1}^{n} w_i \left(\frac{\nu_r + d}{\nu_r + \delta_{i,r+1}} - \log\left(\frac{\nu_r + d}{\nu_r + \delta_{i,r+1}}\right) - 1 \right)$$

Algorithm 3 Multivariate Myriad Filter (MMF).

Same as Algorithm 2 except for

$$\nu_{r+1} = \text{zero of}$$

$$\phi \boxed{\left(\frac{\nu}{2}\right)} - \phi\left(\frac{\nu + d}{2}\right) + \sum_{i=1}^{n} w_i \left(\frac{\nu_r + d}{\nu_r + \delta_{i,r+1}} - \log\left(\frac{\nu_r + d}{\nu_r + \delta_{i,r+1}}\right) - 1 \right)$$

Finally, **Algorithm 4** computes the update of ν by directly finding a zero of the whole function F in (4) given μ_r and Σ_r. The existence of such a zero was discussed in the previous section. The zero computation is done by an inner loop which iterates the update step of ν from Algorithm 3. We will see that the iteration converge indeed to a zero of F.

Algorithm 4 General Multivariate Myriad Filter (GMMF).

Same as Algorithm 2 except for

$$\nu_{r+1} = \text{zero of}$$

$$\phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu + d}{2}\right) + \sum_{i=1}^{n} w_i \left(\frac{\nu + d}{\nu + \delta_{i,r+1}} - \log\left(\frac{\nu + d}{\nu + \delta_{i,r+1}}\right) - 1 \right)$$

for $l = 0, \dots$ **do**

$$\nu_{r,0} = \nu_r$$

$$\nu_{r,l+1} = \text{zero of}$$

$$\phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu + d}{2}\right) + \sum_{i=1}^{n} w_i \left(\frac{\nu_{r,l} + d}{\nu_{r,l} + \delta_{i,r+1}} - \log\left(\frac{\nu_{r,l} + d}{\nu_{r,l} + \delta_{i,r+1}}\right) - 1 \right)$$

In the rest of this section, we prove that the sequence $(L(\nu_r, \mu, r, \Sigma_r))_r$ generated by Algorithms 2 and 3 decreases in each iteration step and that there exists a subsequence of the iterates which converges to a critical point.

We will need the following auxiliary lemma.

Lemma 1 *Let $F_a, F_b\colon \mathbb{R}_{>0} \to \mathbb{R}$ be continuous functions, where F_a is strictly increasing and F_b is strictly decreasing. Define $F := F_a + F_b$. For any initial value $x_0 > 0$ assume that the sequence generated by*

$$x_{l+1} = \textit{zero of } F_a(x) + F_b(x_l)$$

is uniquely determined, i.e., the functions on the right-hand side have a unique zero. Then it holds

i) *If $F(x_0) < 0$, then $(x_l)_l$ is strictly increasing and $F(x) < 0$ for all $x \in [x_l, x_{l+1}]$, $l \in \mathbb{N}_0$.*
ii) *If $F(x_0) > 0$, then $(x_l)_l$ is strictly decreasing and $F(x) > 0$ for all $x \in [x_{l+1}, x_l]$, $l \in \mathbb{N}_0$.*

Furthermore, assume that there exists $x_- > 0$ with $F(x) < 0$ for all $x < x_-$ and $x_+ > 0$ with $F(x) > 0$ for all $x > x_+$. Then, the sequence $(x_l)_l$ converges to a zero x^* of F.

Proof We consider the case i) that $F(x_0) < 0$. Case ii) follows in a similar way.

We show by induction that $F(x_l) < 0$ and that $x_{l+1} > x_l$ for all $l \in \mathbb{N}$. Then it holds for all $l \in \mathbb{N}$ and $x \in (x_l, x_{l+1})$ that $F_a(x) + F_b(x) < F_a(x) + F_b(x_l) < F_a(x_{l+1}) + F_b(x_l) = 0$. Thus $F(x) < 0$ for all $x \in [x_l, x_{l+1}]$, $l \in \mathbb{N}_0$.

Induction step. Let $F_a(x_l) + F_b(x_l) < 0$. Since $F_a(x_{l+1}) + F_b(x_l) = 0 > F_a(x_l) + F_b(x_l)$ and F_a is strictly increasing, we have $x_{l+1} > x_l$. Using that F_b is strictly decreasing, we get $F_b(x_{l+1}) < F_b(x_l)$ and consequently

$$F(x_{l+1}) = F_a(x_{l+1}) + F_b(x_{l+1}) < F_a(x_{l+1}) + F_b(x_l) = 0.$$

Assume now that $F(x) > 0$ for all $x > x_+$. Since the sequence $(x_l)_l$ is strictly increasing and $F(x_l) < 0$ it must be bounded from above by x_+. Therefore it converges to some $x^* \in \mathbb{R}_{>0}$. Now, it holds by the continuity of F_a and F_b that

$$0 = \lim_{l\to\infty} F_a(x_{l+1}) + F_b(x_l) = F_a(x^*) + F_b(x^*) = F(x^*).$$

Hence x^* is a zero of F. □

For the setting in Algorithm 4, Lemma 1 implies the following corollary.

Corollary 3 *Let $F_a(\nu) := \phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu+d}{2}\right)$ and*

$$F_b(\nu) := \sum_{i=1}^n w_i \left(\frac{\nu + d}{\nu + \delta_{i,r+1}} - \log\left(\frac{\nu + d}{\nu + \delta_{i,r+1}}\right) - 1\right), \quad r \in \mathbb{N}_0.$$

Assume that there exists $\nu_+ > 0$ such that $F := F_a + F_b > 0$ for all $\nu \geq \nu_+$. Then the sequence $(\nu_{r,l})_l$ generated by the rth inner loop of Algorithm 4 converges to a zero of F.

Note that by Corollary 2 the above condition on F is fulfilled in each iteration step, e.g., if $\delta_{i,r} \notin [d - \sqrt{2d}, d + \sqrt{2d}]$ for $i = 1, \dots, n$ and $r \in \mathbb{N}_0$.

Proof From the previous section we know that F_a is strictly increasing and F_b is strictly decreasing. Both functions are continuous. If $F(\nu_r) < 0$, then we know from Lemma 1 that $(\nu_{r,l})_l$ is increasing and converges to a zero ν_r^* of F.

If $F(\nu_r) > 0$, then we know from Lemma 1 that $(\nu_{r,l})_l$ is decreasing. The condition that there exists $x_- \in \mathbb{R}_{>0}$ with $F(x) < 0$ for all $x < x_-$ is fulfilled since $\lim_{x\to 0} F(x) = -\infty$. Hence, by Lemma 1, the sequence converges to a zero ν_r^* of F. □

To prove that the objective function decreases in each step of the Algorithms 2–4 we need the following lemma.

Lemma 2 *Let $F_a, F_b \colon \mathbb{R}_{>0} \to \mathbb{R}$ be continuous functions, where F_a is strictly increasing and F_b is strictly decreasing. Define $F := F_a + F_b$ and let $G \colon \mathbb{R}_{>0} \to \mathbb{R}$ be an antiderivative of F, i.e., $F = \frac{\mathrm{d}}{\mathrm{d}x} G$. For an arbitrary $x_0 > 0$, let $(x_l)_l$ be the sequence generated by*

$$x_{l+1} = \text{zero of } F_a(x) + F_b(x_l).$$

Then the following holds true:

i) *The sequence $(G(x_l))_l$ is monotone decreasing with $G(x_l) = G(x_{l+1})$ if and only if x_0 is a critical point of G. If $(x_l)_l$ converges, then the limit x^* fulfills*

$$G(x_0) \geq G(x_1) \geq G(x^*),$$

with equality if and only if x_0 is a critical point of G.

ii) *Let $F = \tilde{F}_a + \tilde{F}_b$ be another splitting of F with continuous functions $\tilde{F}_a, \tilde{F}_b$, where the first one is strictly increasing and the second one strictly decreasing. Assume that $\tilde{F}_a'(x) > F_a'(x)$ for all $x > 0$. Then holds for $y_1 :=$ zero of $\tilde{F}_a(x) + \tilde{F}_b(x_0)$ that $G(x_0) \geq G(y_1) \geq G(x_1)$ with equality if and only if x_0 is a critical point of G.*

Proof i) If $F(x_0) = 0$, then x_0 is a critical point of G.

Let $F(x_0) < 0$. By Lemma 1 we know that $(x_l)_l$ is strictly increasing and that $F(x) < 0$ for $x \in [x_r, x_{r+1}]$, $r \in \mathbb{N}_0$. By the Fundamental Theorem of calculus it holds

$$G(x_{l+1}) = G(x_l) + \int_{x_l}^{x_{l+1}} F(\nu) d\nu.$$

Thus, $G(x_{l+1}) < G(x_l)$.

Let $F(x_0) > 0$. By Lemma 1 we know that $(x_l)_l$ is strictly decreasing and that $F(x) > 0$ for $x \in [x_{r+1}, x_r]$, $r \in \mathbb{N}_0$. Then

$$G(x_l) = G(x_{l+1}) + \int_{x_{l+1}}^{x_l} F(\nu) d\nu.$$

implies $G(x_{l+1}) < G(x_l)$. Now, the rest of assertion i) follows immediately.

ii) It remains to show that $G(x_1) \leq G(y_1)$. Let $F(x_0) < 0$. Then we have $y_1 \geq x_0$ and $x_1 \geq x_0$. By the Fundamental Theorem of calculus we obtain

$$F(x_0) + \int_{x_0}^{x_1} F_a'(x)dx = F_a(x_0) + \int_{x_0}^{x_1} F_a'(x)dx + F_b(x_0) = F_a(x_1) + F_b(x_0) = 0,$$

and

$$F(x_0) + \int_{x_0}^{y_1} \tilde{F}_a'(x)dx = \tilde{F}_a(x_0) + \int_{x_0}^{y_1} \tilde{F}_a'(x)dx + \tilde{F}_b(x_0) = \tilde{F}_a(y_1) + \tilde{F}_b(x_0) = 0.$$

This yields

$$\int_{x_0}^{x_1} F_a'(x)dx = \int_{x_0}^{y_1} \tilde{F}_a'(x)dx,$$

and since $\tilde{F}_a'(x) > F_a'(x)$ further $y_1 \leq x_1$ with equality if and only if $x_0 = x_1$, i.e., if x_0 is a critical point of G. Since $F(x) < 0$ on (x_0, x_1) it holds

$$G(x_1) = G(y_1) + \int_{y_1}^{x_1} F(x)dx \leq G(y_1),$$

with equality if and only if $x_0 = x_1$. The case $F(x_0) > 0$ can be handled similarly. □

Lemma 2 implies the following relation between the values of the objective function L for Algorithms 2–4.

Corollary 4 *For the same fixed $\nu_r > 0, \mu_r \in \mathbb{R}^d, \Sigma_r \in \mathrm{SPD}(d)$ define μ_{r+1}, Σ_{r+1}, ν_{r+1}^{aEM}, ν_{r+1}^{MMF} and $\nu_{r+1}^{\mathrm{GMMF}}$ by Algorithm 2, 3 and 4, respectively. For the GMMF algorithm assume that the inner loop converges. Then it holds*

$$\begin{aligned} L(\nu_r, \mu_{r+1}, \Sigma_{r+1}) &\geq L(\nu_{r+1}^{\mathrm{aEM}}, \mu_{r+1}, \Sigma_{r+1}) \geq L(\nu_{r+1}^{\mathrm{MMF}}, \mu_{r+1}, \Sigma_{r+1}) \\ &\geq L(\nu_{r+1}^{\mathrm{GMMF}}, \mu_{r+1}, \Sigma_{r+1}). \end{aligned}$$

Equality holds true if and only if $\frac{\mathrm{d}}{\mathrm{d}\nu} L(\nu_r, \mu_{r+1}, \Sigma_{r+1}) = 0$ and in this case $\nu_r = \nu_{r+1}^{\mathrm{aEM}} = \nu_{r+1}^{\mathrm{MMF}} = \nu_{r+1}^{\mathrm{GMMF}}$.

Proof For $G(\nu) := L(\nu, \mu_{r+1}, \Sigma_{r+1})$, we have $\frac{\mathrm{d}}{\mathrm{d}\nu} L(\nu, \mu_{r+1}, \Sigma_{r+1}) = F(\nu)$, where

$$F(\nu) := \phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu+d}{2}\right) + \sum_{i=1}^n w_i \left(\frac{\nu+d}{\nu+\delta_{i,r+1}} - \log\left(\frac{\nu+d}{\nu+\delta_{i,r+1}}\right) - 1\right).$$

We use the splitting

$$F = F_a + F_b = \tilde{F}_a + \tilde{F}_b$$

with

$$F_a(\nu) := \phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu+d}{2}\right), \quad \tilde{F}_a := \phi\left(\frac{\nu}{2}\right),$$

$$F_b(\nu) := \sum_{i=1}^n w_i \left(\frac{\nu+d}{\nu+\delta_{i,r+1}} - \log\left(\frac{\nu+d}{\nu+\delta_{i,r+1}}\right) - 1\right),$$

and

$$\tilde{F}_b(\nu) := -\phi\left(\frac{\nu+d}{2}\right) + F_b(\nu).$$

By the considerations in the previous section we know that F_a, $\tilde{F}_a$ are strictly increasing and F_b, $\tilde{F}_b$ are strictly decreasing. Moreover, since $\phi' > 0$ we have $\tilde{F}_a' > F_a'$. Hence it follows from Lemma 2(ii) that

$$L(\nu_r, \mu_{r+1}, \Sigma_{r+1}) \geq L\left(\nu_r^{\mathrm{aEM}}, \mu_{r+1}, \Sigma_{r+1}\right) \geq L\left(\nu_r^{\mathrm{MMF}}, \mu_{r+1}, \Sigma_{r+1}\right).$$

Finally, we conclude by Lemma 2(i) that

$$L\left(\nu_r^{\mathrm{MMF}}, \mu_{r+1}, \Sigma_{r+1}\right) \geq L\left(\nu_r^{\mathrm{GMMF}}, \mu_{r+1}, \Sigma_{r+1}\right).$$

□

Concerning the convergence of the three algorithms we have the following result.

Theorem 3 *Let $(\nu_r, \mu_r, \Sigma_r)_r$ be sequence generated by Algorithm 2, 3 or 4, respectively starting with arbitrary initial values $\nu_0 > 0, \mu_0 \in \mathbb{R}^d, \Sigma_0 \in \mathrm{SPD}(d)$. For the GMMF algorithm we assume that in each step the inner loop converges. Then it holds for all $r \in \mathbb{N}_0$ that*

$$L(\nu_r, \mu_r, \Sigma_r) \geq L(\nu_{r+1}, \mu_{r+1}, \Sigma_{r+1}),$$

with equality if and only if $(\nu_r, \mu_r, \Sigma_r) = (\nu_{r+1}, \mu_{r+1}, \Sigma_{r+1})$.

Proof By the general convergence results of the accelerated EM algorithm for fixed ν, see also [17], it holds

$$L(\nu_r, \mu_{r+1}, \Sigma_{r+1}) \leq L(\nu_r, \mu_r, \Sigma_r),$$

with equality if and only if $(\mu_r, \Sigma_r) = (\mu_{r+1}, \Sigma_{r+1})$. By Corollary 4 it holds

$$L(\nu_{r+1}, \mu_{r+1}, \Sigma_{r+1}) \leq L(\nu_r, \mu_{r+1}, \Sigma_{r+1}),$$

with equality if and only if $\nu_r = \nu_{r+1}$. The combination of both results proves the claim. □

Lemma 3 *Let $T = (T_1, T_2, T_3) : \mathbb{R}_{>0} \times \mathbb{R}^d \times SPD(d) \to \mathbb{R}_{>0} \times \mathbb{R}^d \times SPD(d)$ be the operator of one iteration step of Algorithm 2 (or 3). Then T is continuous.*

Proof We show the statement for Algorithm 3. For Algorithm 2 it can be shown analogously. Clearly the mapping $(T_2, T_3)(\nu, \mu, \Sigma)$ is continuous. Since

$$T_1(\nu, \mu, \Sigma) = \text{zero of } \Psi(x, \nu, T_2(\nu, \mu, \Sigma), T_3(\nu, \mu, \Sigma)),$$

where

$$\Psi\ (x,\ \nu, \mu, \Sigma) = \phi\left(\frac{x}{2}\right) - \phi\left(\frac{x+d}{2}\right) + \sum_{i=1}^n w_i \left(\frac{\nu+d}{\nu+(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)} - \log\left(\frac{\nu+d}{\nu+(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)}\right) - 1\right).$$

It is sufficient to show that the zero of Ψ depends continuously on ν, T_2 and T_3. Now the continuously differentiable function Ψ is strictly increasing in x, so that $\frac{\partial}{\partial x}\Psi(x,\nu,T_2,T_3) > 0$. By $\Psi(T_1,\nu,T_2,T_3) = 0$, the Implicit Function Theorem yields the following statement: There exists an open neighborhood $U \times V$ of (T_1,ν,T_2,T_3) with $U \subset \mathbb{R}_{>0}$ and $V \subset \mathbb{R}_{>0} \times \mathbb{R}^d \times SPD(d)$ and a continuously differentiable function $G\colon V \to U$ such that for all $(x,\nu,\mu,\Sigma) \in U \times V$ it holds

$$\Psi(x,\nu,\mu,\Sigma) = 0 \quad \text{if and only if} \quad G(\nu,\mu,\Sigma) = x.$$

Thus the zero of Ψ depends continuously on ν, T_2 and T_3. □

This implies the following theorem.

Theorem 4 *Let $(\nu_r,\mu_r,\Sigma_r)_r$ be the sequence generated by Algorithm 2 or 3 with arbitrary initial values $\nu_0 > 0$, $\mu_0 \in \mathbb{R}^d$, $\Sigma_0 \in \mathrm{SPD}(d)$. Then every cluster point of $(\nu_r,\mu_r,\Sigma_r)_r$ is a critical point of L.*

Proof The mapping T defined in Lemma 3 is continuous. Further we know from its definition that (ν,μ,Σ) is a critical point of L if and only if it is a fixed point of T. Let $(\hat\nu,\hat\mu,\hat\Sigma)$ be a cluster point of $(\nu_r,\mu_r,\Sigma_r)_r$. Then there exists a subsequence $(\nu_{r_s},\mu_{r_s},\Sigma_{r_s})_s$ which converges to $(\hat\nu,\hat\mu,\hat\Sigma)$. Further we know by Theorem 3 that $L_r = L(\nu_r,\mu_r,\Sigma_r)$ is decreasing. Since $(L_r)_r$ is bounded from below, it converges. Now it holds

$$\begin{aligned} L\left(\hat\nu,\hat\mu,\hat\Sigma\right) &= \lim_{s\to\infty} L\left(\nu_{r_s},\mu_{r_s},\Sigma_{r_s}\right) \\ &= \lim_{s\to\infty} L_{r_s} = \lim_{s\to\infty} L_{r_s+1} \\ &= \lim_{s\to\infty} L\left(\nu_{r_s+1},\mu_{r_s+1},\Sigma_{r_s+1}\right) \\ &= \lim_{s\to\infty} L\left(T\left(\nu_{r_s},\mu_{r_s},\Sigma_{r_s}\right)\right) = L\left(T\left(\hat\nu,\hat\mu,\hat\Sigma\right)\right). \end{aligned}$$

By Theorem 3 and the definition of T we have that $L(\nu,\mu,\Sigma) = L(T(\nu,\mu,\Sigma))$ if and only if $(\nu,\mu,\Sigma) = T(\nu,\mu,\Sigma)$. By the definition of the algorithm this is the case if and only if (ν,μ,Σ) is a critical point of L. Thus $(\hat\nu,\hat\mu,\hat\Sigma)$ is a critical point of L. □

6 Numerical results

In this section we give two numerical examples of the developed theory. First, we compare the four different algorithms in Section 6.1. Then, in Section 6.2, we address further accelerations of our algorithms by SQUAREM [33] and DAAREM [9] and show also a comparison with the ECME algorithm [20]. Finally, in Section 6.3, we provide an application in image analysis by determining the degree of freedom parameter in images corrupted by Student t noise. We run all experiments on a HP Probook with Intel i7-8550U Quad Core processor. The code is provided online[2].

6.1 Comparison of algorithms

In this section, we compare the numerical performance of the classical EM algorithm 1 and the proposed Algorithms 2, 3, and 4. To this aim, we did the following Monte Carlo simulation: Based on the stochastic representation of the Student t distribution, see (1), we draw $n = 1000$ i.i.d. realizations of the $T_\nu(\mu, \Sigma)$ distribution with location parameter $\mu = 0$ and different scatter matrices Σ and degrees of freedom parameters ν. Then, we used Algorithms 2, 3, and 4 to compute the ML estimator $(\hat{\nu}, \hat{\mu}, \hat{\Sigma})$.

We initialize all algorithms with the sample mean for μ and the sample covariance matrix for Σ. Furthermore, we set $\nu = 3$ and in all algorithms the zero of the respective function is computed by Newton's method. As a stopping criterion we use the following relative distance:

$$\frac{\sqrt{\|\mu_{r+1} - \mu_r\|^2 + \|\Sigma_{r+1} - \Sigma_r\|_F^2}}{\sqrt{\|\mu_r\|^2 + \|\Sigma_r\|_F^2}} + \frac{\sqrt{(\log(\nu_{r+1}) - \log(\nu_r))^2}}{|\log(\nu_r)|} < 10^{-5}.$$

We take the logarithm of ν in the stopping criterion, because $T_\nu(\mu, \Sigma)$ converges to the normal distribution as $\nu \to \infty$ and therefore the difference between $T_\nu(\mu, \Sigma)$ and $T_{\nu+1}(\mu, \Sigma)$ becomes small for large ν.

To quantify the performance of the algorithms, we count the number of iterations until the stopping criterion is reached. Since the inner loop of the GMMF is potentially time consuming we additionally measure the execution time until the stopping criterion is reached. This experiment is repeated $N = 10.000$ times for different values of $\nu \in \{1, 2, 5, 10\}$. Afterward we calculate the average number of iterations and the average execution times. The results are given in Tables 1 and 2. We observe that the performance of the algorithms depends on Σ. Further we see, that the performance of the aEM algorithm is always better than those of the classical EM algorithm. Further all algorithms need a longer time to estimate large ν. This seems to be natural since the likelihood function becomes very flat for large ν. Further, the GMMF needs the lowest number of iterations. But for small ν the execution time of the GMMF is larger than those of the MMF and the aEM algorithm. This can be explained by the fact, that the ν step has a smaller relevance for small ν but is still time consuming in the GMMF. The MMF needs slightly more iterations than the GMMF but if ν is not extremely large the execution time is smaller than for the GMMF and for the aEM algorithm. In summary, the MMF algorithm is proposed as algorithm of choice.

In Fig. 2 we exemplarily show the functional values $L(\nu_r, \mu_r, \Sigma_r)$ of the four algorithms and samples generated for different values of ν and $\Sigma = I$. Note that the x-axis of the plots is in log-scale. We see that the convergence speed (in terms

[2] https://github.com/johertrich/Alternatives-EM-Studentt

Table 1 Average number of iterations (lowest in bold) and the corresponding standard deviations of the different algorithms

Σ	ν	EM	aEM	MMF	GMMF
$\begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$	1	62.32 ± 2.50	23.44 ± 0.79	22.16 ± 0.75	$\mathbf{20.61 \pm 0.70}$
	2	46.17 ± 1.82	26.42 ± 1.08	21.48 ± 0.94	$\mathbf{17.79 \pm 0.80}$
	5	50.42 ± 11.22	49.97 ± 7.48	25.28 ± 2.61	$\mathbf{12.14 \pm 1.73}$
	10	122.62 ± 31.74	117.40 ± 31.65	38.16 ± 4.51	$\mathbf{14.32 \pm 0.96}$
	100	531.07 ± 91.41	528.14 ± 92.19	53.66 ± 6.98	$\mathbf{10.76 \pm 2.07}$
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1	62.34 ± 2.52	23.43 ± 0.78	22.16 ± 0.75	$\mathbf{20.59 \pm 0.70}$
	2	46.20 ± 1.81	26.43 ± 1.07	21.49 ± 0.94	$\mathbf{17.79 \pm 0.80}$
	5	50.68 ± 10.86	50.06 ± 7.42	25.31 ± 2.58	$\mathbf{12.06 \pm 1.75}$
	10	122.72 ± 31.65	117.51 ± 31.56	38.18 ± 4.50	$\mathbf{14.28 \pm 0.97}$
	100	531.75 ± 90.98	528.84 ± 91.75	53.62 ± 6.94	$\mathbf{10.64 \pm 2.02}$
$\begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$	1	62.35 ± 2.55	23.44 ± 0.78	22.15 ± 0.76	$\mathbf{20.59 \pm 0.71}$
	2	46.27 ± 1.82	26.45 ± 1.08	21.51 ± 0.95	$\mathbf{17.81 \pm 0.80}$
	5	50.71 ± 11.21	50.15 ± 7.61	25.34 ± 2.63	$\mathbf{12.08 \pm 1.78}$
	10	122.44 ± 30.66	117.19 ± 30.56	38.17 ± 4.46	$\mathbf{14.27 \pm 0.96}$
	100	533.21 ± 89.80	530.27 ± 90.57	53.64 ± 6.93	$\mathbf{10.62 \pm 2.01}$
$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$	1	62.32 ± 2.55	23.43 ± 0.78	22.15 ± 0.76	$\mathbf{20.60 \pm 0.70}$
	2	46.22 ± 1.82	26.43 ± 1.09	21.50 ± 0.94	$\mathbf{17.80 \pm 0.80}$
	5	50.76 ± 11.12	50.21 ± 7.52	25.35 ± 2.59	$\mathbf{12.09 \pm 1.75}$
	10	122.37 ± 31.01	117.17 ± 30.92	38.13 ± 4.49	$\mathbf{14.30 \pm 0.96}$
	100	530.89 ± 91.36	527.96 ± 92.15	53.68 ± 7.07	$\mathbf{10.75 \pm 2.08}$

of number of iterations) of the EM algorithm is much slower than those of the MMF/GMMF. For small ν the convergence speed of the aEM algorithm is close to the GMMF/MMF, but for large ν it is close to the EM algorithm.

In Fig. 3 we show the histograms of the ν-output of 1000 runs for different values of ν and $\Sigma = I$. Since the ν-outputs of all algorithms are very close together we only plot the output of the GMMF. We see that the accuracy of the estimation of ν decreases for increasing ν. This can be explained by the fact, that the likelihood function becomes very flat for large ν such that the estimation of ν becomes much harder.

6.2 Comparison with other accelerations of the EM algortihm

In this section, we compare our algorithms with the Expectation/Conditional Maximization Either (ECME) algorithm [19, 20] and apply the SQUAREM acceleration [33] as well as the damped Anderson Acceleration (DAAREM) [9] to our algorithms.

Table 2 The execution times (lowest in bold) and the corresponding standard deviations of the different algorithms

Σ	ν	EM	aEM	MMF	GMMF
$\begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$	1	0.008469 ± 0.00111	0.003511 ± 0.00044	**0.003498** ± **0.00044**	0.006954 ± 0.00114
	2	0.006428 ± 0.00069	0.003995 ± 0.00042	**0.003409** ± **0.00036**	0.005388 ± 0.00061
	5	0.007237 ± 0.00208	0.007768 ± 0.00181	0.004133 ± 0.00085	**0.003752** ± **0.00100**
	10	0.017421 ± 0.00532	0.017991 ± 0.00567	0.006187 ± 0.00122	**0.005796** ± **0.00110**
	100	0.070024 ± 0.01306	0.075191 ± 0.01418	0.008146 ± 0.00131	**0.005601** ± **0.00097**
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1	0.008645 ± 0.00090	0.003581 ± 0.00034	**0.003572** ± **0.00036**	0.007126 ± 0.00098
	2	0.006431 ± 0.00074	0.003989 ± 0.00044	**0.003417** ± **0.00039**	0.005427 ± 0.00071
	5	0.006883 ± 0.00162	0.007352 ± 0.00128	0.003939 ± 0.00058	**0.003550** ± **0.00079**
	10	0.016434 ± 0.00439	0.016964 ± 0.00470	0.005869 ± 0.00089	**0.005493** ± **0.00077**
	100	0.072309 ± 0.01507	0.077724 ± 0.01624	0.008363 ± 0.00155	**0.005773** ± **0.00117**
$\begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$	1	0.008839 ± 0.00108	0.003664 ± 0.00043	**0.003639** ± **0.00042**	0.007217 ± 0.00104
	2	0.006516 ± 0.00075	0.004054 ± 0.00048	**0.003449** ± **0.00039**	0.005428 ± 0.00065
	5	0.007293 ± 0.00207	0.007799 ± 0.00180	0.004149 ± 0.00082	**0.003740** ± **0.00098**
	10	0.020598 ± 0.00659	0.021193 ± 0.00683	0.007228 ± 0.00167	**0.006834** ± **0.00155**
	100	0.078682 ± 0.01969	0.084275 ± 0.02087	0.009039 ± 0.00213	**0.006246** ± **0.00160**
$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$	1	0.008837 ± 0.00107	0.003648 ± 0.00039	**0.003641** ± **0.00041**	0.007207 ± 0.00104
	2	0.006481 ± 0.00070	0.004016 ± 0.00041	**0.003433** ± **0.00036**	0.005413 ± 0.00061
	5	0.006968 ± 0.00167	0.007440 ± 0.00129	0.003965 ± 0.00055	**0.003561** ± **0.00077**
	10	0.016608 ± 0.00442	0.017107 ± 0.00468	0.005920 ± 0.00092	**0.005499** ± **0.00076**
	100	0.072354 ± 0.01509	0.077586 ± 0.01619	0.008385 ± 0.00153	**0.005715** ± **0.00114**

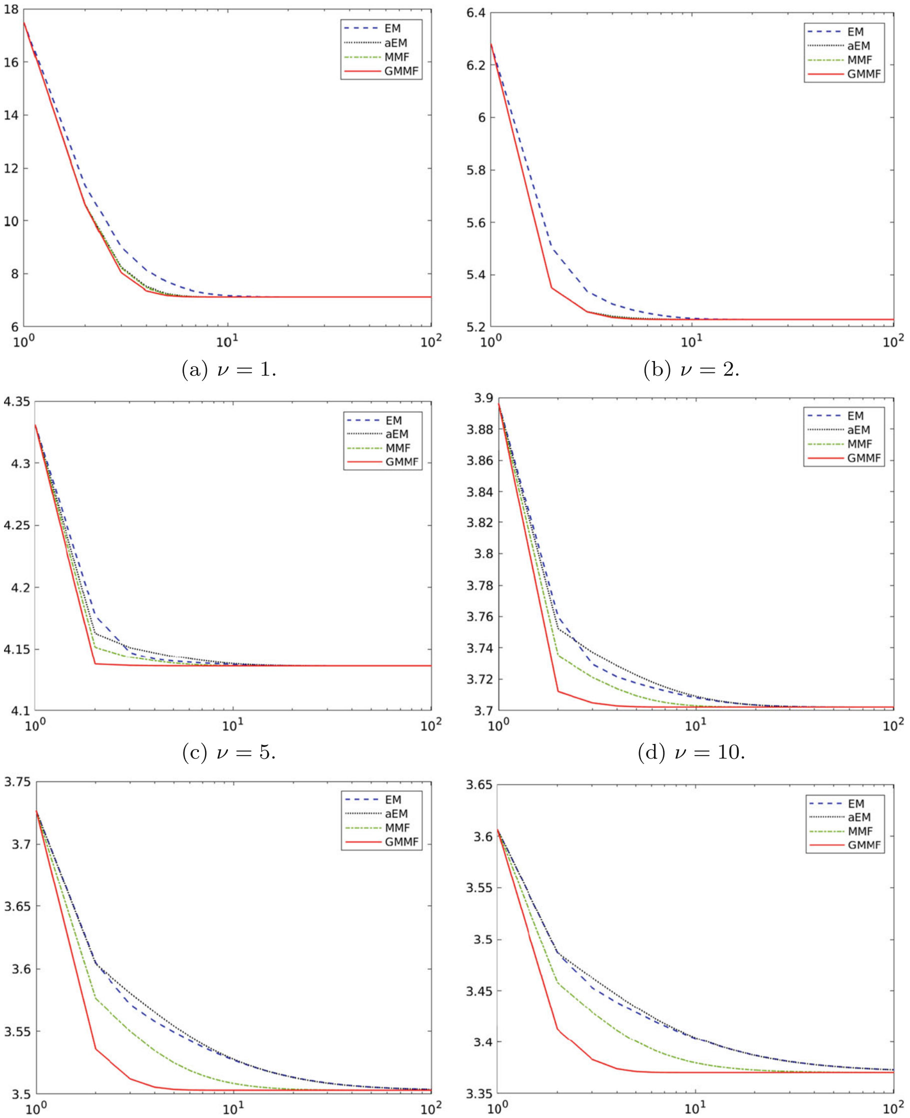
ECME algorithm: The ECME algorithm was first proposed in [19]. Some numerical examples of the behavior of the ECME algorithm for estimating the parameters (ν, μ, Σ) of a Student t distribution $T_\nu(\mu, \Sigma)$ are given in [20]. The idea of ECME is first to replace the M-Step of the EM algorithm by the following update of the parameters (ν_r, μ_r, Σ_r): first, we fix $\nu = \nu_r$ and compute the update $(\mu_{r+1}, \Sigma_{r+1})$ of the parameters (μ_r, Σ_r) by performing one step of the EM algorithm for fixed degree of freedom (CM1-Step). Second, we fix $(\mu, \Sigma) = (\mu_r, \Sigma_r)$ and compute the update ν_{r+1} of ν_r by maximizing the likelihood function with respect to ν (CM2-Step). The resulting algorithm is given in Algorithm 5. It is similar to the GMMF (Algorithm 4), but uses the Σ-update of the EM algorithm (Algorithm 5) instead of the Σ-update of the aEM algorithm (Algorithm 2). The authors of [19] showed a similar convergence result as for the EM algorithm. Alternatively, we could prove Theorem 3 for the ECME algorithm analogously as for the GMMF algorithm.

Algorithm 5 ECME Algorithm (ECME).

```
Input: $x_1, \dots, x_n \in \mathbb{R}^d$, $n \ge d+1$, $w \in \mathring{\Delta}_n$
Initialization: $\nu_0 = \varepsilon > 0$, $\mu_0 = \frac{1}{n}\sum_{i=1}^n x_i$, $\Sigma_0 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu_0)(x_i - \mu_0)^{\mathrm{T}}$
for $r = 0, \dots$
    E-Step: Compute the weights
        $\delta_{i,r} = (x_i - \mu_r)^{\mathrm{T}} \Sigma_r^{-1} (x_i - \mu_r)$
        $\gamma_{i,r} = \frac{\nu_r + d}{\nu_r + \delta_{i,r}}$
    CM1-Step: Update the parameters
        $\mu_{r+1} = \frac{\sum_{i=1}^n w_i \gamma_{i,r} x_i}{\sum_{i=1}^n w_i \gamma_{i,r}}$
        $\Sigma_{r+1} = \sum_{i=1}^n w_i \gamma_{i,r} (x_i - \mu_{r+1})(x_i - \mu_{r+1})^{\mathrm{T}}$
    CM2-Step: Update the parameter
        $\nu_{r+1} =$ zero of
        $\phi\left(\frac{\nu}{2}\right) - \phi\left(\frac{\nu+d}{2}\right) + \sum_{i=1}^n w_i \left(\frac{\nu+d}{\nu+\delta_{i,r+1}} - \log\left(\frac{\nu+d}{\nu+\delta_{i,r+1}}\right) - 1\right)$
```

Next, we consider two acceleration schemes of arbitrary fixed point algorithms $\vartheta_{r+1} = G(\vartheta_r)$. In our case $\vartheta \in \mathbb{R}^p$ is given by (ν, μ, Σ) and G is given by one step of Algorithm 1, 2, 3, 4, or 5.

SQUAREM Acceleration: The first acceleration scheme, called squared iterative methods (SQUAREM) was proposed in [33]. The idea of SQUAREM is to update the parameters $\vartheta_r = (\nu_r, \mu_r, \Sigma_r)$ in the following way: we compute $\vartheta_{r,1} = G(\vartheta_r)$ and $\vartheta_{r,2} = G(\vartheta_{r,1})$. Then, we calculate $s = \vartheta_{r,1} - \vartheta_r$ and $v = (\vartheta_{r,2} - \vartheta_{r,1}) - s$. Now we set $\vartheta' = \vartheta_r - 2\alpha r + \alpha^2 v$ and define the update $\vartheta_{r+1} = G(\vartheta')$, where α is chosen as follows. First, we set $\alpha = \min(-\frac{\|r\|_2}{\|v\|_2}, -1)$. Then we compute ϑ' as described before. If $L(\vartheta') < L(\vartheta_r)$, we keep our choice of α. Otherwise we update α by $\alpha = \frac{\alpha-1}{2}$. Note that this scheme terminates as long a ϑ_r is not a critical point of L by the following argument: it holds that $\vartheta_r + 2r + v = \vartheta_{r,2}$, which implies that it holds that $\lim_{\alpha \to -1} L(\vartheta_r - 2\alpha + \alpha^2 v) = L(\vartheta_{r,2}) \le L(\vartheta_r)$ with equality if and only if ϑ_r is a critical point of L, since all our algorithms have the property that $L(\vartheta) \ge L(G(\vartheta))$ with equality if and only if ϑ is a critical point of L. By construction this scheme ensures that the negative log-likelihood values of the iterates are decreasing.



(a) $\nu = 1$.

(b) $\nu = 2$.

(c) $\nu = 5$.

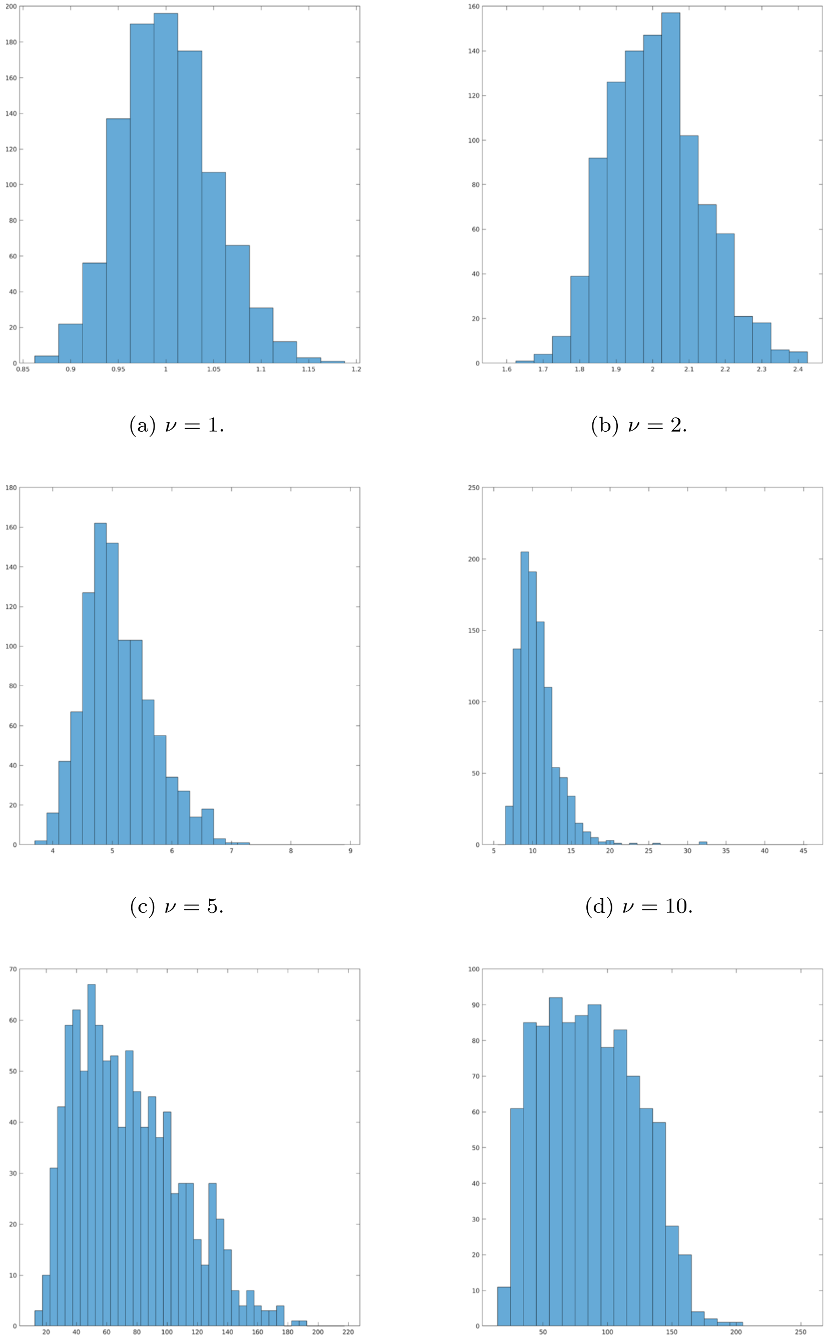
(d) $\nu = 10$.

(e) $\nu = 100$.

(f) $\nu = 200$.

Fig. 2 Plots of $L(\nu_r, \mu_r, \Sigma_r)$ on the *y*-axis and r on the *x*-axis for all algorithms

Damped Anderson Acceleration with Restarts and ϵ-Monotonicity (DAAREM): The DAAREM acceleration was proposed in [9]. It is based on the Anderson acceleration, which was introduced in [2]. As for the SQUAREM acceleration want to solve the fixed point equation $\vartheta = G(\vartheta)$ with $\vartheta = (\nu, \mu, \Sigma)$ using the iteration $\vartheta_{r+1} = G(\vartheta_r)$. We also use the equivalent formulation to solve $f(\vartheta) = 0$, where $f(\vartheta) = G(\vartheta) - \vartheta$. For a fixed parameter $m \in \mathbb{N}_{>0}$, we define $m_r = \min(m, r)$. Then, one update of ϑ_r using the Anderson Acceleration is given by



(a) $\nu = 1.$

(b) $\nu = 2.$

(c) $\nu = 5.$

(d) $\nu = 10.$

(e) $\nu = 100.$

(f) $\nu = 200.$

Fig. 3 Histograms of the output ν from the algorithms

$$\begin{aligned}\vartheta_{r+1} &= G(\vartheta_r) - \sum_{j=1}^{m_r}(G(\vartheta_{r-m_r+j}) - G(\vartheta_{r-m_r+j-1}))\gamma_j^{(r)} \\ &= \vartheta_r + f(\vartheta_r) - \sum_{j=1}^{m_r}((\vartheta_{r-m_r+j} - \vartheta_{r-m_r+j-1}) - (f(\vartheta_{r-m_r+j}) \\ &\qquad -f(\vartheta_{r-m_r+j-1})))\gamma_j^{(r)}, \qquad (14)\end{aligned}$$

with $\gamma^{(r)} = \left(\mathcal{F}_r^{\mathrm{T}}\mathcal{F}_r\right)^{-1}\mathcal{F}_r^{\mathrm{T}} f(\vartheta_r)$, where the columns of $\mathcal{F}_r \in \mathbb{R}^{p\times m_r}$ are given by $f(\vartheta_{r-m_r+j+1}) - f(\vartheta_{r-m_r+j})$ for $j = 0, \dots, m_r - 1$. An equivalent formulation of update step (14) is given by

$$\vartheta_{r+1} = \vartheta_r + f(\vartheta_r) - (\mathcal{X}_r + \mathcal{F}_r)\gamma^{(r)},$$

where the columns of $\mathcal{X}_r \in \mathbb{R}^{p\times m_r}$ are given by $\vartheta_{r-m_r+j+1} - \vartheta_{r-m_r+j}$ for $j = 0, \dots, m_r - 1$. The Anderson acceleration can be viewed as a special case of a multisecant quasi-Newton procedure to solve $f(\vartheta) = 0$. For more details we refer to [7, 9].

The DAAREM acceleration modifies the Anderson acceleration in three points. The first modification is to restart the algorithm after m steps. That is, to set $m_r = \min(m, c_r)$ instead of $m_r = \min(m, r)$, where $c_r \in \{1, \dots, m\}$ is defined by $c_r = r \bmod m$. The second modification is to add damping term in the computation coefficients $\gamma^{(r)}$. This means, that $\gamma^{(r)}$ is given by $\gamma^{(r)} = (\mathcal{F}_r^{\mathrm{T}}\mathcal{F}_r + \lambda_r I)^{-1}\mathcal{F}_r^{\mathrm{T}} f(\vartheta_r)$ instead of $\gamma^{(r)} = (\mathcal{F}_r^{\mathrm{T}}\mathcal{F})^{-1}\mathcal{F}_r^{\mathrm{T}} f(\vartheta_r)$. The parameter λ_r is chosen such that

$$\|(\mathcal{F}_r^{\mathrm{T}}\mathcal{F}_r + \lambda_r I)^{-1}\mathcal{F}_r^{\mathrm{T}} f(\vartheta_r)\|_2^2 = \delta_r\|(\mathcal{F}_r^{\mathrm{T}}\mathcal{F}_r)^{-1}\mathcal{F}_r^{\mathrm{T}} f(\vartheta_r)\|_2^2 \qquad (15)$$

for some damping parameters δ_r. We initialize the δ_r by $\delta_1 = \frac{1}{1+\alpha^\kappa}$ and decrease the exponent of α in each step by 1 up to a minimum of $\kappa - D$ for some parameter $D \in \mathbb{N}_{>0}$. The third modification is to enforce that for the negative log-likelihood function L does not increase more than ϵ in one iteration step. To do this, we compute the update ϑ_{r+1} using the Anderson acceleration. If $L(\vartheta_{r+1}) > L(\vartheta_r) + \epsilon$, we use our original fixed point algorithm in this step, i.e., we set $\vartheta_{r+1} = G(\vartheta_r)$.

We summarize the DAAREM acceleration in Algorithm 6. In our numerical experiments we use for the parameters the values suggested by [9], that is $\epsilon = 0.01$, $\epsilon_c = 0$, $\alpha = 1.2$, $\kappa = 25$, $D = 2\kappa$ and $m = \min(\lceil \frac{p}{2} \rceil, 10)$, where p is the number of parameters in ϑ.

Algorithm 6 DAAREM acceleration.

```
Input: Parameters $\epsilon \geq 0, \epsilon_c \geq 0, \alpha > 1, \kappa \geq 0, D \geq 0, m \geq 1$
Initialization: Initialize $\vartheta_0 = (\nu_0, \mu_0, \Sigma_0)$ as in the corresponding fixed point algorithm.
Set $c_1 = 1, s_1 = 0, \vartheta_1 = \vartheta_0 + f(\vartheta_0), L^* = L(x_1)$.
for r=1,2,... do
    Set $m_r = \min(m, c_r)$, $\delta_r = \frac{1}{1+\alpha^{\kappa - s_r}}$ and compute $f_r = f(\vartheta_r)$.
    Define the columns of $\mathcal{F}_r, \mathcal{X}_r \in \mathbb{R}^{p \times m_k}$ by $f_{r-m_r+j+1} - f_{r-m_r+j}$ and $\vartheta_{r-m_r+j+1} - \vartheta_{r-m_r+j}$ respectively, $j = 0, \ldots, m_r - 1$.
    Define $\lambda_r$ by (16) and set $\gamma^{(r)} = (\mathcal{F}_r^{\mathrm{T}} \mathcal{F}_r + \lambda_r I)^{-1} \mathcal{F}_r^{\mathrm{T}} f_r$.
    Set $t_{r+1} = \vartheta_r + f_r - (\mathcal{X}_r + \mathcal{F}_r)\gamma^{(r)}$
    if $L(t_{r+1}) \leq L(\vartheta_r) + \epsilon$ then
        Set $\vartheta_{r+1} = t_{r+1}$ and $s_{\mathrm{new}} = s_r + 1$.
    else
        Set $\vartheta_{r+1} = \vartheta_r + f_r$ and $s_{\mathrm{new}} = s_r$.
    if $k \bmod m = 0$ then
      if $L(\vartheta_{r+1}) > L^* + \epsilon_c$ then
          Set $s_{\mathrm{new}} = \max\{s_{\mathrm{new}} - m, -D\}$
        Set $c_{k+1} = 1$ and $L^* = L(\vartheta_{k+1})$.
    else
        Set $c_{r+1} = c_r + 1$.
```

Simulation Study: To compare the performance of all of these algorithms we perform again a Monte Carlo simulation. As in the previous section we draw $n = 100$ i.i.d. realizations of $T_\nu(\mu, \Sigma)$ with $\mu = 0$, $\Sigma = 0.1\,\mathrm{Id}$ and $\nu \in \{1, 2, 5, 10, 100\}$. Then, we use each of the Algorithms 1, 2, 3, 4 and 5 to compute the ML estimator $(\hat{\nu}, \hat{\mu}, \hat{\Sigma})$. We use each of these algorithms with no acceleration, with SQUAREM acceleration and with DAAREM acceleration.

We use the same initialization and stopping criteria as in the previous section and repeat this experiment $N = 1.000$ times. To quantify the performance of the algorithms, we count the number of iterations and measure the execution time. The results are given in Tables 3 and 4. Since the DAAREM and SQUAREM accelerations were proposed originally for an absolute stopping criteria, we redo the experiments with the stopping criteria

$$\sqrt{\|\mu_{r+1} - \mu_r\|^2 + \|\Sigma_{r+1} - \Sigma_r\|_F^2 + (\log(\nu_{r+1}) - \log(\nu_r))^2} < 10^{-8}.$$

The results are given in Tables 5 and 6.

We observe that for nearly any choice of the parameters the performance of the GMMF is better than the performance of the ECME. For small ν, the performance of the SQUAREM-aEM is also very good. On the other hand, for large ν the SQUAREM-GMMF behaves very well. Further, for any choice of ν the performance of the SQUAREM-MMF is close to the best algorithm.

Table 3 Average number of iterations (lowest in bold) and the corresponding standard deviations of the different algorithms using a relative stopping criterion

Algorithm	$\nu = 1$	$\nu = 2$	$\nu = 5$	$\nu = 10$	$\nu = 100$
EM	62.24 ± 2.47	46.20 ± 1.84	50.14 ± 11.01	122.45 ± 30.81	530.72 ± 89.11
aEM	23.39 ± 0.75	26.46 ± 1.08	49.60 ± 7.55	117.21 ± 30.74	527.77 ± 89.92
MMF	22.13 ± 0.73	21.51 ± 0.96	25.12 ± 2.63	38.17 ± 4.47	53.98 ± 7.06
GMMF	20.56 ± 0.67	17.79 ± 0.79	12.06 ± 1.73	14.35 ± 0.97	10.86 ± 2.10
ECME	60.81 ± 2.41	40.73 ± 1.97	29.07 ± 1.81	22.12 ± 3.81	12.81 ± 2.96
DAAREM-EM	22.09 ± 4.05	22.26 ± 4.59	20.39 ± 5.42	24.72 ± 6.34	28.09 ± 6.93
DAAREM-aEM	15.52 ± 1.57	14.90 ± 2.39	15.35 ± 3.22	17.84 ± 4.41	20.07 ± 3.68
DAAREM-MMF	15.16 ± 1.45	14.02 ± 2.09	13.12 ± 2.09	14.99 ± 3.62	66.86 ± 630.74
DAAREM-GMMF	14.11 ± 1.04	12.81 ± 1.46	9.61 ± 1.27	9.84 ± 1.46	10.15 ± 2.10
DAAREM-ECME	22.69 ± 4.71	19.15 ± 3.50	17.06 ± 3.33	16.89 ± 3.75	12.35 ± 3.90
SQUAREM-EM	26.36 ± 2.25	21.77 ± 4.56	21.43 ± 3.13	46.01 ± 10.72	111.24 ± 40.47
SQUAREM-aEM	15.32 ± 0.98	14.86 ± 0.86	22.87 ± 2.26	43.57 ± 8.29	38.56 ± 35.35
SQUAREM-MMF	15.47 ± 1.09	14.05 ± 1.40	14.18 ± 1.56	18.40 ± 1.21	22.41 ± 9.39
SQUAREM-GMMF	**13.30 ± 1.49**	**11.99 ± 0.16**	**9.02** ± **0.49**	**8.90** ± **0.80**	**8.28** ± **1.29**
SQUAREM-ECME	24.25 ± 2.79	19.20 ± 1.96	18.48 ± 3.12	17.98 ± 3.33	13.41 ± 3.41

6.3 Unsupervised estimation of noise parameters

Next, we provide an application in image analysis. To this aim, we consider images corrupted by one-dimensional Student t noise with $\mu = 0$ and unknown $\Sigma \equiv \sigma^2$ and ν. We provide a method that allows to estimate ν and σ in an unsupervised way. The basic idea is to consider constant areas of an image, where the signal to noise ratio is weak and differences between pixel values are solely caused by the noise.

Constant area detection: In order to detect constant regions in an image, we adopt an idea presented in [30]. It is based on Kendall's τ-coefficient, which is a measure of rank correlation, and the associated z-score, see [10, 11]. In the following, we briefly summarize the main ideas behind this approach. For finding constant regions we proceed as follows: First, the image grid $\mathcal{G}$ is partitioned into K small, non-overlapping regions $\mathcal{G} = \bigcup_{k=1}^{K} R_k$, and for each region we consider the hypothesis testing problem

$$\mathcal{H}_0\colon R_k \text{ is constant} \qquad \text{vs.} \qquad \mathcal{H}_1 : R_k \text{ is not constant.}$$

To decide whether to reject $\mathcal{H}_0$ or not, we observe the following: Consider a fixed region R_k and let $I, J \subseteq R_k$ be two disjoint subsets of R_k with the same cardinality. Denote with u_I and u_J the vectors containing the values of u at the positions indexed by I and J. Then, under $\mathcal{H}_0$, the vectors u_I and u_J are uncorrelated (in fact even independent) for all choices of $I, J \subseteq R_k$ with $I \cap J = \emptyset$ and $|I| = |J|$. As a consequence, the rejection of $\mathcal{H}_0$ can be reformulated as the question whether we can find I, J such that u_I and u_J are significantly correlated, since in this case there

Table 4 The execution times (lowest in bold) and the corresponding standard deviations of the different algorithms using a relative stopping criterion

Algorithm	$\nu = 1$	$\nu = 2$	$\nu = 5$	$\nu = 10$	$\nu = 100$
EM	0.00890 ± 0.00163	0.00644 ± 0.00074	0.00682 ± 0.00158	0.01659 ± 0.00432	0.07076 ± 0.01350
aEM	0.00365 ± 0.00056	0.00401 ± 0.00049	0.00732 ± 0.00128	0.01706 ± 0.00465	0.07513 ± 0.01416
MMF	0.00369 ± 0.00075	0.00342 ± 0.00039	0.00390 ± 0.00052	0.00589 ± 0.00085	0.00834 ± 0.00151
GMMF	0.00763 ± 0.00193	0.00540 ± 0.00061	0.00355 ± 0.00074	0.00551 ± 0.00063	0.00599 ± 0.00112
ECME	0.01998 ± 0.00343	0.01214 ± 0.00137	0.00927 ± 0.00114	0.00801 ± 0.00105	0.00684 ± 0.00157
DAAREM-EM	0.00728 ± 0.00163	0.00726 ± 0.00158	0.00652 ± 0.00180	0.00796 ± 0.00218	0.00905 ± 0.00233
DAAREM-aEM	0.00554 ± 0.00095	0.00519 ± 0.00097	0.00530 ± 0.00124	0.00613 ± 0.00160	0.00687 ± 0.00141
DAAREM-MMF	0.00553 ± 0.00090	0.00500 ± 0.00084	0.00463 ± 0.00082	0.00529 ± 0.00137	0.02410 ± 0.22518
DAAREM-GMMF	0.00837 ± 0.00185	0.00679 ± 0.00091	0.00491 ± 0.00081	0.00601 ± 0.00086	0.00772 ± 0.00201
DAAREM-ECME	0.01527 ± 0.00351	0.01061 ± 0.00175	0.00968 ± 0.00171	0.00993 ± 0.00189	0.00825 ± 0.00207
SQUAREM-EM	0.00456 ± 0.00081	0.00372 ± 0.00077	0.00375 ± 0.00068	0.00831 ± 0.00220	0.02299 ± 0.00837
SQUAREM-aEM	**0.00291 ± 0.00050**	0.00269 ± 0.00029	0.00441 ± 0.00065	0.00913 ± 0.00203	0.00795 ± 0.00621
SQUAREM-MMF	0.00308 ± 0.00059	**0.00268 ± 0.00035**	**0.00270 ± 0.00041**	**0.00373 ± 0.00041**	0.00474 ± 0.00184
SQUAREM-GMMF	0.00569 ± 0.00129	0.00400 ± 0.00040	0.00304 ± 0.00042	0.00375 ± 0.00046	**0.00420 ± 0.00080**
SQUAREM-ECME	0.01153 ± 0.00222	0.00722 ± 0.00086	0.00717 ± 0.00112	0.00761 ± 0.00090	0.00727 ± 0.00182

Table 5 Average number of iterations (lowest in bold) and the corresponding standard deviations of the different algorithms using an absolute stopping criterion

Algorithm	$\nu = 1$	$\nu = 2$	$\nu = 5$	$\nu = 10$	$\nu = 100$
EM	87.56 ± 3.22	60.61 ± 2.93	58.16 ± 10.53	126.97 ± 30.12	535.17 ± 88.15
aEM	29.00 ± 1.27	30.98 ± 1.44	53.40 ± 7.24	119.34 ± 30.30	531.86 ± 89.26
MMF	28.07 ± 1.23	26.58 ± 1.13	29.24 ± 2.39	41.21 ± 4.49	55.30 ± 7.15
GMMF	26.54 ± 1.19	22.97 ± 1.15	15.84 ± 1.77	17.31 ± 1.04	12.32 ± 2.19
ECME	86.02 ± 3.32	54.99 ± 2.92	36.30 ± 2.72	25.92 ± 4.33	13.94 ± 3.37
DAAREM-EM	30.48 ± 7.45	29.11 ± 8.37	24.50 ± 6.35	27.65 ± 6.44	29.13 ± 6.63
DAAREM-aEM	19.96 ± 2.05	19.19 ± 3.15	18.74 ± 4.30	20.45 ± 4.92	21.34 ± 3.90
DAAREM-MMF	19.44 ± 1.80	18.20 ± 2.61	15.94 ± 2.44	17.65 ± 4.30	62.17 ± 546.96
DAAREM-GMMF	18.49 ± 1.50	16.29 ± 1.86	12.26 ± 1.54	12.15 ± 1.67	11.50 ± 2.18
DAAREM-ECME	30.87 ± 8.03	24.74 ± 5.14	20.88 ± 4.38	19.68 ± 4.66	13.95 ± 4.42
SQUAREM-EM	34.97 ± 3.73	28.91 ± 4.59	25.45 ± 3.00	49.97 ± 10.56	111.47 ± 38.98
SQUAREM-aEM	20.73 ± 1.08	18.02 ± 0.78	26.56 ± 2.05	46.94 ± 8.43	41.79 ± 31.88
SQUAREM-MMF	21.04 ± 0.51	17.84 ± 0.96	18.07 ± 1.24	21.59 ± 1.34	25.51 ± 10.29
SQUAREM-GMMF	$\mathbf{17.09 \pm 1.39}$	$\mathbf{15.02 \pm 0.21}$	$\mathbf{12.12 \pm 0.73}$	$\mathbf{12.01 \pm 0.97}$	$\mathbf{11.28 \pm 1.34}$
SQUAREM-ECME	33.28 ± 4.73	25.84 ± 2.73	22.82 ± 2.77	21.12 ± 3.36	16.13 ± 3.40

has to be some structure in the image region R_k and it cannot be constant. Now, in order to quantify the correlation, we adopt an idea presented in [30] and make use of Kendall's τ-coefficient, which is a measure of rank correlation, and the associated z-score, see [10, 11]. The key idea is to focus on the rank (i.e., on the relative order) of the values rather than on the values themselves. In this vein, a block is considered homogeneous if the ranking of the pixel values is uniformly distributed, regardless of the spatial arrangement of the pixels. In the following, we assume that we have extracted two disjoint subsequences $x = u_I$ and $y = u_J$ from a region R_k with I and J as above. Let (x_i, y_i) and (x_j, y_j) be two pairs of observations. Then, the pairs are said to be

$$\begin{cases} \text{concordant} & \text{if } x_i < x_j \text{ and } y_i < y_j \\ & \text{or } x_i > x_j \text{ and } y_i > y_j, \\ \text{discordant} & \text{if } x_i < x_j \text{ and } y_i > y_j \\ & \text{or } x_i > x_j \text{ and } y_i < y_j, \\ \text{tied} & \text{if } x_i = x_j \text{ or } y_i = y_j. \end{cases}$$

Next, let $x, y \in \mathbb{R}^n$ be two sequences without tied pairs and let n_c and n_d be the number of concordant and discordant pairs, respectively. Then, *Kendall's τ coefficient* [10] is defined as $\tau \colon \mathbb{R}^n \times \mathbb{R}^n \to [-1, 1]$,

$$\tau(x, y) = \frac{n_c - n_d}{\frac{n(n-1)}{2}}.$$

From this definition we see that if the agreement between the two rankings is perfect, i.e., the two rankings are the same, then the coefficient attains its maximal value 1. On the other extreme, if the disagreement between the two rankings is perfect, that

Table 6 The execution times (lowest in bold) and the corresponding standard deviations of the different algorithms using an absolute stopping criterion

Algorithm	$\nu = 1$	$\nu = 2$	$\nu = 5$	$\nu = 10$	$\nu = 100$
EM	0.01578 ± 0.00421	0.00995 ± 0.00203	0.01008 ± 0.00296	0.02179 ± 0.00743	0.08227 ± 0.01929
aEM	0.00581 ± 0.00159	0.00565 ± 0.00117	0.01012 ± 0.00257	0.02222 ± 0.00769	0.08872 ± 0.02065
MMF	0.00596 ± 0.00164	0.00503 ± 0.00102	0.00582 ± 0.00136	0.00817 ± 0.00226	0.00978 ± 0.00205
GMMF	0.01119 ± 0.00344	0.00751 ± 0.00167	0.00541 ± 0.00150	0.00770 ± 0.00216	0.00682 ± 0.00161
ECME	0.03058 ± 0.00874	0.01698 ± 0.00354	0.01314 ± 0.00327	0.01080 ± 0.00309	0.00802 ± 0.00253
DAAREM-EM	0.01286 ± 0.00455	0.01147 ± 0.00395	0.00995 ± 0.00338	0.01116 ± 0.00356	0.01078 ± 0.00305
DAAREM-aEM	0.00923 ± 0.00267	0.00817 ± 0.00203	0.00835 ± 0.00274	0.00897 ± 0.00298	0.00846 ± 0.00197
DAAREM-MMF	0.00912 ± 0.00254	0.00783 ± 0.00181	0.00718 ± 0.00186	0.00790 ± 0.00284	0.02663 ± 0.24134
DAAREM-GMMF	0.01298 ± 0.00407	0.00976 ± 0.00219	0.00752 ± 0.00204	0.00866 ± 0.00247	0.00937 ± 0.00272
DAAREM-ECME	0.02314 ± 0.00759	0.01508 ± 0.00394	0.01402 ± 0.00424	0.01365 ± 0.00410	0.00995 ± 0.00322
SQUAREM-EM	0.00779 ± 0.00239	0.00603 ± 0.00152	0.00579 ± 0.00171	0.01164 ± 0.00416	0.02634 ± 0.00987
SQUAREM-aEM	**0.00505 ± 0.00149**	**0.00406 ± 0.00090**	0.00662 ± 0.00171	0.01262 ± 0.00417	0.00977 ± 0.00623
SQUAREM-MMF	0.00537 ± 0.00158	0.00418 ± 0.00094	**0.00443 ± 0.00121**	0.00552 ± 0.00146	0.00609 ± 0.00253
SQUAREM-GMMF	0.00834 ± 0.00271	0.00561 ± 0.00120	0.00467 ± 0.00128	**0.00551 ± 0.00152**	**0.00554 ± 0.00139**
SQUAREM-ECME	0.01701 ± 0.00529	0.01023 ± 0.00249	0.01031 ± 0.00288	0.01043 ± 0.00302	0.00863 ± 0.00234

is, one ranking is the reverse of the other, then the coefficient has value -1. If the sequences x and y are uncorrelated, we expect the coefficient to be approximately zero. Denoting with X and Y the underlying random variables that generated the sequences x and y, we have the following result, whose proof can be found in [10].

Theorem 5 *Let X and Y be two arbitrary sequences under $\mathcal{H}_0$ without tied pairs. Then, the random variable $\tau(X, Y)$ has an expected value of 0 and a variance of $\frac{2(2n+5)}{9n(n-1)}$. Moreover, for $n \to \infty$, the associated z-score $z\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,*

$$z(x, y) = \frac{3\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}\tau(x, y) = \frac{3\sqrt{2}(n_c - n_d)}{\sqrt{n(n-1)(2n+5)}}$$

is asymptotically standard normal distributed,

$$z(X, Y) \overset{n\to\infty}{\sim} \mathcal{N}(0, 1).$$

With slight adaption, Kendall's τ coefficient can be generalized to sequences with tied pairs (see [11]). As a consequence of Theorem 5, for a given significance level $\alpha \in (0, 1)$, we can use the quantiles of the standard normal distribution to decide whether to reject $\mathcal{H}_0$ or not. In practice, we cannot test any kind of region and any kind of disjoint sequences. As in [30], we restrict our attention to quadratic regions and pairwise comparisons of neighboring pixels. We use four kinds of neighboring relations (horizontal, vertical and two diagonal neighbors) thus perform in total four tests. We reject the hypothesis $\mathcal{H}_0$ that the region is constant as soon as one of the four tests rejects it. Note that by doing so, the final significance level is smaller than the initially chosen one. We start with blocks of size 64 × 64 whose side-length is incrementally decreased until enough constant areas are found.

Parameter estimation. In each constant region we consider the pixel values in the region as i.i.d. samples of a univariate Student t distribution $T_\nu(\mu, \sigma^2)$, where we estimate the parameters using Algorithm 3.

After estimating the parameters in each found constant region, the estimated location parameters μ are discarded, while the estimated scale and degrees of freedom parameters σ respective ν are averaged to obtain the final estimate of the global noise parameters. At this point, as both ν and σ influence the resulting distribution in a multiplicative way, instead of an arithmetic mean, one might use a geometric which is slightly less affected by outliers.

In Fig. 4 we illustrate this procedure for two different noise scenarios. The left column in each figure depicts the detected constant areas. The middle and right column show histograms of the estimated values for ν respective σ. For the constant area detection we use the code of [30][3]. The true parameters used to generate the noisy images where $\nu = 1$ and $\sigma = 10$ for the top row and $\nu = 5$ and $\sigma = 10$ for the bottom row, while the obtained estimates are (geometric mean in brackets) $\hat{\nu} = 1.0437$ (1.0291) and $\hat{\sigma} = 10.3845$ (10.3111) for the top row and $\hat{\nu} = 5.4140$ (5.0423) and $\hat{\sigma} = 10.5500$ (10.1897) for the bottom row.

[3] https://github.com/csutour/RNLF

A further example is given in Fig. 5. Here, the obtained estimates are (geometric mean in brackets) $\hat{\nu} = 1.0075$ (0.99799) and $\hat{\sigma} = 10.2969$ (10.1508) for the top row and $\hat{\nu} = 5.4184$ (5.1255) and $\hat{\sigma} = 10.2295$ (10.1669) for the bottom row.

Appendix. Auxiliary lemmas

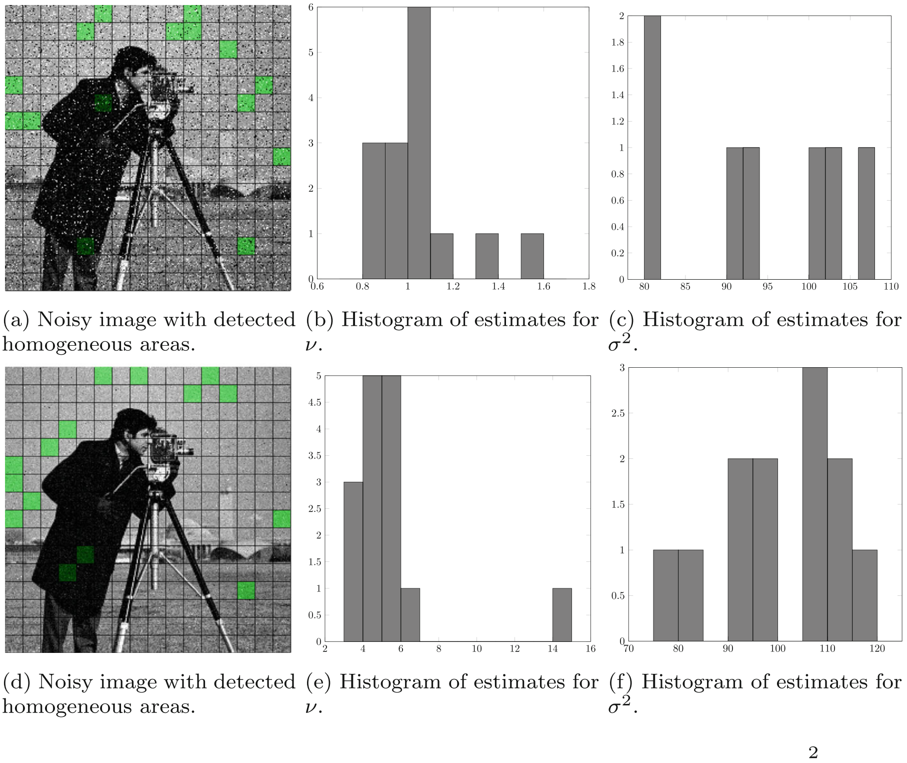
Lemma 4 *Let $x_i \in \mathbb{R}^d$, $i = 1, \dots, n$ and $w \in \mathring{\Delta}_n$ fulfill Assumption 1. Let $(\nu_r, \Sigma_r)_r$ be a sequence in $\mathbb{R}_{>0} \times \mathrm{SPD}(d)$ with $\nu_r \to 0$ as $r \to \infty$ (or if $\{\nu_r\}_r$ has a subsequence which converges to zero). Then $(\nu_r, \Sigma_r)_r$ cannot be a minimizing sequence of $L(\nu, \Sigma)$.*

Proof We write

$$L(\nu, \Sigma) = g(\nu) + L_\nu(\Sigma),$$

where

$$g(\nu) = 2\log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - 2\log\left(\Gamma\left(\frac{d+\nu}{2}\right)\right) - \nu\log(\nu).$$



(a) Noisy image with detected homogeneous areas.

(b) Histogram of estimates for ν.

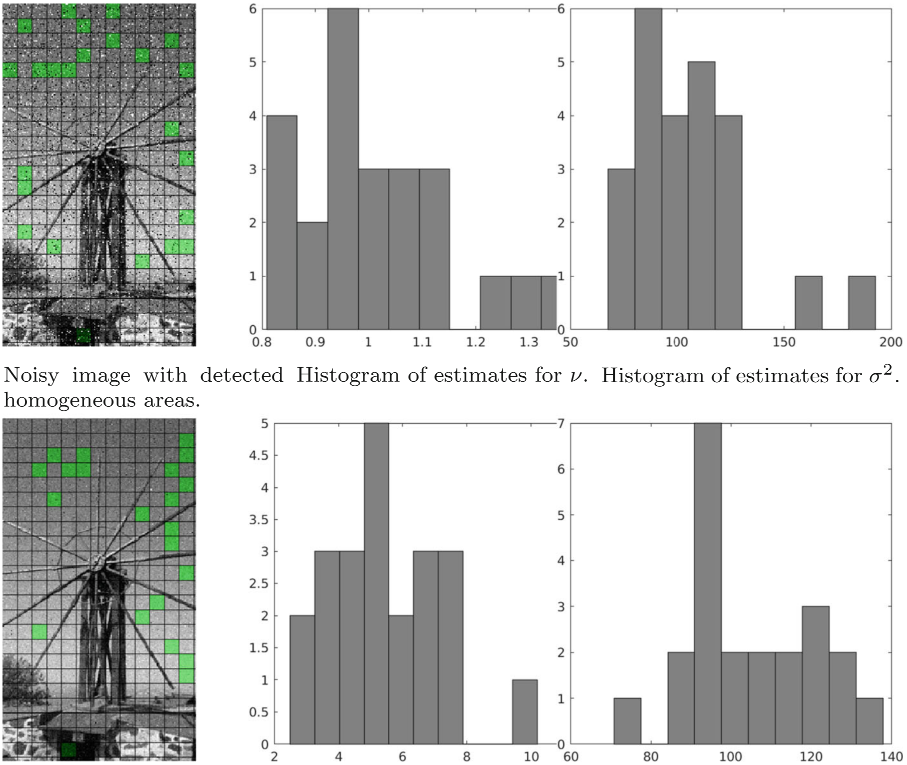
(c) Histogram of estimates for σ^2.

(d) Noisy image with detected homogeneous areas.

(e) Histogram of estimates for ν.

(f) Histogram of estimates for σ^2.

Fig. 4 Unsupervised estimation of the noise parameters ν and σ^2



Noisy image with detected homogeneous areas. Histogram of estimates for ν. Histogram of estimates for σ^2.

Noisy image with detected homogeneous areas. Histogram of estimates for ν. Histogram of estimates for σ^2.

Fig. 5 Unsupervised estimation of the noise parameters ν and σ^2

Then it holds $\lim_{\nu \to 0} g(\nu) = \infty$. Hence it is sufficient to show that $(\nu_r, \Sigma_r)_r$ has a subsequence $(\nu_{r_k}, \Sigma_{r_k})$ such that $\left(L_{\nu_{r_k}}(\Sigma_{r_k})\right)_r$ is bounded from below. Denote by $\lambda_{r1} \geq \ldots \geq \lambda_{rd}$ the eigenvalues of Σ_r.

Case 1: Let $\{\lambda_{r,i} : r \in \mathbb{N}, i = 1, \ldots, d\} \subseteq [a, b]$ for some $0 < a \leq b < \infty$. Then it holds $\liminf_{r\to\infty} \log |\Sigma_r| \geq \log(a^d) = d \log(a)$ and

$$\begin{aligned}\liminf_{r\to\infty}(d+\nu_r)\sum_{i=1}^{n} w_i \log(\nu_r + x_i^{\mathrm{T}} \Sigma_r^{-1} x_i) &\geq \lim_{r\to\infty}(d+\nu_r)\sum_{i=1}^{n} w_i \log\left(\frac{1}{b} x_i^{\mathrm{T}} x_i\right)\\ &= d\sum_{i=1}^{n} w_i \log\left(\frac{1}{b} x_i^{\mathrm{T}} x_i\right).\end{aligned}$$

Note that Assumption 1 ensures $x_i \neq 0$ and $x_i^{\mathrm{T}} x_i > 0$ for $i = 1, \ldots, n$. Then we get

$$\begin{aligned}\liminf_{r\to\infty} L_{\nu_r}(\Sigma_r) &= \liminf_{r\to\infty}(d+\nu_r)\sum_{i=1}^n w_i \log(\nu_r + x_i^{\mathrm{T}}\Sigma_r^{-1}x_i) + \log|\Sigma_r| \\ &\geq d\sum_{i=1}^n w_i \log\left(\frac{1}{b}x_i^{\mathrm{T}}x_i\right) + d\log(a).\end{aligned}$$

Hence $(L_{\nu_r}(\Sigma_r))_r$ is bounded from below and (ν_r, Σ_r) cannot be a minimizing sequence.

Case 2: Let $\{\lambda_{r,i} : r \in \mathbb{N}, i = 1, \ldots, d\} \not\subseteq [a, b]$ for all $0 < a \leq b < \infty$. Define $\rho_r = \|\Sigma_r\|_F$ and $P_r = \frac{\Sigma_r}{\rho_r}$. Then, by concavity of the logarithm, it holds

$$\begin{aligned}L_{\nu_r}(\Sigma_r) &= (d+\nu_r)\sum_{i=1}^n w_i \log(\nu_r + x_i^{\mathrm{T}}\Sigma_r^{-1}x_i) + \log(|\Sigma_r|) \\ &\geq d\sum_{i=1}^n w_i \log(x_i^{\mathrm{T}}\Sigma_r^{-1}x_i) + \nu_r\log(\nu_r) + \log(|\Sigma_r|) \\ &\geq d\sum_{i=1}^n w_i \log\left(\frac{1}{\rho_r}x_i^{\mathrm{T}}P_r^{-1}x_i\right) + \log(\rho_r^d\,|P_r|) + \text{const} \\ &= \underbrace{d\sum_{i=1}^n w_i \log(x_i^{\mathrm{T}}P_r^{-1}x_i) + \log(|P_r|)}_{=:L_0(P_r)} + \text{const}. \qquad (16)\end{aligned}$$

Denote by $p_{r,1} \geq \ldots \geq p_{r,d} > 0$ the eigenvalues of P_r. Since $\{P_r : r \in \mathbb{N}\}$ is bounded there exists some $C > 0$ with $C \geq p_{r,1}$ for all $r \in \mathbb{N}$. Thus one of the following cases is fulfilled:

i) There exists a constant $c > 0$ such that $p_{r,d} > c$ for all $r \in \mathbb{N}$.
ii) There exists a subsequence $(P_{r_k})_k$ of $(P_r)_r$ which converges to some $P \in \partial SPD(d)$.

Case 2i) Let $c > 0$ with $p_{r,d} \geq c$ for all $r \in \mathbb{N}$. Then $\liminf_{r\to\infty} \log(|P_r|) \geq \log(c^d) = d\log(c)$ and

$$\liminf_{r\to\infty} d\sum_{i=1}^n w_i \log(x_i^{\mathrm{T}}P_r^{-1}x_i) \geq d\sum_{i=1}^n w_i \log\left(\frac{1}{C}x_i^{\mathrm{T}}x_i\right).$$

By (16) this yields

$$\begin{aligned}\liminf_{r\to\infty} L_{\nu_r}(\Sigma_r) &\geq \liminf_{r\to\infty} d\sum_{i=1}^n w_i \log(x_i^{\mathrm{T}}P_r^{-1}x_i) + \log(|P_r|) + \text{const} \\ &\geq d\sum_{i=1}^n w_i \log\left(\frac{1}{C}x_i^{\mathrm{T}}x_i\right) + d\log(c) + \text{const}.\end{aligned}$$

Hence, $(L_{\nu_r}(\Sigma_r))_r$ is bounded from below and (ν_r, Σ_r) cannot be a minimizing sequence.

Case 2ii) We use similar arguments as in the proof of [17, Theorem 4.3]. Let $(P_{r_k})_k$ be a subsequence of $(P_r)_r$ which converges to some $P \in \partial SPD(d)$. For simplicity we denote $(P_{r_k})_k$ again by $(P_r)_r$. Let $p_1 \geq \ldots \geq p_d \geq 0$ be the eigenvalues of P. Since $\|P\|_F = \lim\limits_{r\to\infty} \|P_r\|_F = 1$ it holds $p_1 > 0$. Let $q \in 1, \ldots, d-1$ such that

$$p_1 \geq \ldots \geq p_q > p_{q+1} = \ldots = p_d = 0.$$

By $e_{r,1}, \ldots, e_{,rd}$, we denote the orthonormal eigenvectors corresponding to $p_{r,1}, \ldots, p_{r,d}$. Since $(\mathbb{S}^d)^d$ is compact we can assume (by going over to a subsequence) that $(e_{r,1}, \ldots, e_{r,d})_r$ converges to orthonormal vectors $(e_1, \ldots, e_d)$. Define $S_0 := \{0\}$ and for $k = 1, \ldots, d$ set $S_k := \operatorname{span}\{e_1, \ldots, e_k\}$. Now, for $k = 1, \ldots, d$ define

$$W_k := S_k \backslash S_{k-1} = \{y \in \mathbb{R}^d : \langle y, e_k \rangle \neq 0, \langle y, e_l \rangle = 0 \text{ for } l = k+1, \ldots, d\}.$$

Further, let

$$\tilde{I}_k := \{i \in \{1, \ldots, n\} : x_i \in S_k\} \quad \text{and} \quad I_k := \{i \in \{1, \ldots, n\} : x_i \in W_k\}.$$

Because of $S_k = W_k \dot{\cup} S_{k-1}$ we have $\tilde{I}_k = I_k \dot{\cup} \tilde{I}_{k-1}$ for $k = 1, \ldots, d$. Due to Assumption 1 we have $|I_k| \leq \left|\tilde{I}_k\right| \leq \dim(S_k) = k$ for $k = 1, \ldots, d-1$. Defining for $j = 1, \ldots, d$,

$$L_j(P_r) := d \sum_{i \in I_j} w_i \log(x_i^{\mathrm{T}} P_r^{-1} x_i) + \log(p_{rj}),$$

it holds $L_0(P_r) = \sum_{j=1}^d L_j$. For $j \leq q$ we get

$$\liminf_{r\to\infty} L_j(P_r) \geq \liminf_{r\to\infty} d \sum_{i \in I_j} w_i \log\left(\frac{1}{C} x_i^{\mathrm{T}} x_i\right) + \log(p_{r,j}) = d \sum_{i \in I_j} w_i \log\left(\frac{1}{C} x_i^{\mathrm{T}} x_i\right) + \log(p_j).$$

Since for $k \in \{1, \ldots, d\}$ and $i \in I_k$,

$$x_i^{\mathrm{T}} P_r^{-1} x_i = \sum_{j=1}^d \frac{1}{p_{r,j}} \left\langle x_i, e_{r,j} \right\rangle^2 \geq \frac{1}{p_{r,k}} \left\langle x_i, e_{rk} \right\rangle^2,$$

and $\lim_{r\to\infty} \langle x_i, e_{rk} \rangle = \langle x_i, e_k \rangle \neq 0$, we obtain

$$\liminf_{r\to\infty} p_{r,k} x_i^{\mathrm{T}} P_r x_i \geq \liminf_{r\to\infty} \left\langle y, e_{r,k} \right\rangle \geq \langle y, e_k \rangle^2 > 0.$$

Hence, it holds for $j \geq q+1$ that

$$\begin{aligned} L_j(P_r) &= d \sum_{i \in I_j} w_i \left[\log(x_i^{\mathrm{T}} P_r^{-1} x_i) + \log(p_{r,j})\right] + \left(1 - d \sum_{i \in I_j} w_i\right) \log(p_{r,j}) \\ &= d \sum_{i \in I_j} w_i \log(p_{r,j} x_i^{\mathrm{T}} P_r^{-1} x_i) + \left(1 - d \sum_{i \in I_j} w_i\right) \log(p_{r,j}). \end{aligned}$$

Thus, we conclude

$$\begin{aligned}
\liminf_{r\to\infty} L_0(P_r) &= \liminf_{r\to\infty} \sum_{j=1}^{d} L_j(P_r) \geq \sum_{j=1}^{q} \liminf_{r\to\infty} L_j(P_r) + \liminf_{r\to\infty} \sum_{j=q+1}^{d} L_j(P_r) \\
&\geq \sum_{j=1}^{q} d \sum_{i\in I_j} w_i \log\left(\frac{1}{C} x_i^{\mathrm{T}} x_i\right) + \log(p_j) + \liminf_{r\to\infty} \sum_{j=q+1}^{d} d \sum_{i\in I_j} w_i \log(p_{rj} x_i^{\mathrm{T}} P_r^{-1} x_i) \\
&\quad + \liminf_{r\to\infty} \sum_{j=q+1}^{d} \left(1 - d\sum_{i\in I_j} w_i\right) \log(p_{rj}) \\
&\geq \sum_{j=1}^{q} d \sum_{i\in I_j} w_i \log(\frac{1}{C} x_i^{\mathrm{T}} x_i) + \log(p_j) + \sum_{j=q+1}^{d} d \sum_{i\in I_j} w_i \log(\langle x_i, e_j \rangle) \\
&\quad + \liminf_{r\to\infty} \sum_{j=q+1}^{d} \left(1 - d\sum_{i\in I_j} w_i\right) \log(p_{r,j}) \\
&= \text{const} + \liminf_{r\to\infty} \sum_{j=q+1}^{d} \left(1 - d\sum_{i\in I_j} w_i\right) \log(p_{r,j}).
\end{aligned}$$

It remains to show that there exist $\tilde{c} > 0$ such that

$$\liminf_{r\to\infty} \sum_{j=q+1}^{d} \left(1 - d\sum_{i\in I_j} w_i\right) \log(p_{r,j}) \geq \tilde{c}. \tag{17}$$

We prove for $k \geq q+1$ by induction that for sufficiently large $r \in \mathbb{N}$ it holds

$$\sum_{j=k}^{d} \left(1 - d\sum_{i\in I_j} w_i\right) \log(p_{rj}) \geq \left(d \sum_{i\in \tilde{I}_{k-1}} w_i - (k-1)\right) \log(p_{r,k}). \tag{18}$$

Induction basis $k = d$**:** Since $\tilde{I}_k = I_k \cup \tilde{I}_{k-1}$ we have

$$\sum_{i\in \tilde{I}_k} w_i - \sum_{i\in \tilde{I}_{k-1}} w_i = \sum_{i\in I_k} w_i,$$

and further

$$1-d\sum_{i\in I_d} w_i = 1-d\left(\sum_{i\in \tilde{I}_d} w_i - \sum_{i\in \tilde{I}_{d-1}} w_i\right) = 1-d\left(1-\sum_{i\in \tilde{I}_{d-1}} w_i\right) = d\sum_{i\in \tilde{I}_{d-1}} w_i - (d-1).$$

If we multiply both sides with $\log(p_{rd})$ this yields (18) for $k = d$.
Induction step: Assume that (18) holds for some $k+1$ with $d \geq k+1 > q+1$, i.e.,

$$\sum_{j=k+1}^{d} \left(1 - d\sum_{i\in I_j} w_i\right) \log(p_{r,j}) \geq d\left(\sum_{i\in \tilde{I}_k} w_i - \frac{k}{d}\right) \log(p_{r,k+1}).$$

Then we obtain

$$\begin{aligned}
&\sum_{j=k}^{d}\left(1-d\sum_{i\in I_j}w_i\right)\log(p_{r,j})\\
&=\sum_{j=k+1}^{d}\left(1-d\sum_{i\in I_j}w_i\right)\log(p_{r,j})+\left(1-d\sum_{i\in I_k}w_i\right)\log(p_{r,k})\\
&\geq d\left(\sum_{i\in\tilde{I}_k}w_i-\frac{k}{d}\right)\log(p_{r,k+1})+\left(1-d\sum_{i\in I_k}w_i\right)\log(p_{r,k}).
\end{aligned}$$

and since $\sum_{i\in\tilde{I}_k}w_i<\left|\tilde{I}_k\right|\frac{1}{d}\leq\frac{k}{d}$ by Assumption 1 and $p_{r,k+1}\leq p_{r,k}<1$ finally

$$\begin{aligned}
&\geq d\left(\sum_{i\in\tilde{I}_k}w_i-\frac{k}{d}\right)\log(p_{r,k})+\left(1-d\sum_{i\in I_k}w_i\right)\log(p_{r,k})\\
&=\left(d\sum_{i\in\tilde{I}_{k-1}}w_i-(k-1)\right)\log(p_{r,k}).
\end{aligned}$$

This shows (18) for $k\geq q+1$. Using $k=q+1$ in (17) we get

$$\liminf_{r\to\infty}\sum_{j=q+1}^{d}\left(1-d\sum_{i\in I_j}w_i\right)\log(p_{rj})\geq\liminf_{r\to\infty}\underbrace{\left(d\sum_{i\in\tilde{I}_q}w_i-q\right)}_{<0}\underbrace{\log(p_{r,q+1})}_{\text{bounded from above}}>-\infty.$$

This finishes the proof. □

Lemma 5 *Let $(\nu_r,\Sigma_r)_r$ be a sequence in $\mathbb{R}_{>0}\times\mathrm{SPD}(d)$ such that there exists $\nu_-\in\mathbb{R}_{>0}$ with $\nu_-\leq\nu_r$ for all $r\in\mathbb{N}$. Denote by $\lambda_{r,1}\geq\cdots\geq\lambda_{r,d}$ the eigenvalues of Σ_r. If $\{\lambda_{1,r}:r\in\mathbb{N}\}$ is unbounded or $\{\lambda_{d,r}:r\in\mathbb{N}\}$ has zero as a cluster point, then there exists a subsequence $(\nu_{r_k},\Sigma_{r_k})_k$ of $(\nu_r,\Sigma_r)_r$, such that $\lim\limits_{k\to\infty}L(\nu_{r_k},\Sigma_{r_k})=\infty$.*

Proof Without loss of generality we assume (by considering a subsequence) that either $\lambda_{r1}\to\infty$ as $r\to\infty$ and $\lambda_{rd}\geq c>0$ for all $r\in\mathbb{N}$ or that $\lambda_{rd}\to 0$ as $r\to\infty$. By [17, Theorem 4.3] for fixed $\nu=\nu_-$, we have $L_{\nu_-}(\Sigma_r)\to\infty$ as $r\to\infty$.

The function $h\colon\mathbb{R}_{>0}\to\mathbb{R}$ defined by $\nu\mapsto(d+\nu)\log(\nu+k)$ is monotone increasing for all $k\in\mathbb{R}_{\geq 0}$. This can be seen as follows: The derivative of h fulfills

$$h'(\nu)=\frac{d+\nu}{k+\nu}+\log(\nu+k)\geq\frac{1+\nu}{k+\nu}+\log(\nu+k),$$

and since

$$\frac{\partial}{\partial k}\left(\frac{1+\nu}{k+\nu}+\log(\nu+k)\right)=\frac{k-1}{(k+\nu)^2},$$

the later function is minimal for $k = 1$, so that

$$h'(\nu) \geq \frac{1+\nu}{k+\nu} + \log(\nu+k) \geq \frac{1+\nu}{1+\nu} + \log(\nu+1) = 1 + \log(1+\nu) > 0.$$

Using this relation, we obtain

$$(d+\nu_r)\sum_{i=1}^{n} w_i \log\left(\nu_r + x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right) \geq (d+\nu_-)\sum_{i=1}^{n} w_i \log\left(\nu_- + x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right)$$

and further

$$\begin{aligned} L(\nu_r, \Sigma_r) &= (d+\nu_r)\sum_{i=1}^{n} w_i \log\left(\nu_r + x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right) + \log(|\Sigma_r|) \\ &\geq (d+\nu_-)\sum_{i=1}^{n} w_i \log\left(\nu_- + x_i^{\mathrm{T}}\Sigma_r^{-1}x_i\right) + \log(|\Sigma_r|) \\ &= L_{\nu_-}(\Sigma_r) \to \infty \qquad \text{as} \quad r \to \infty. \end{aligned}$$ □

Acknowledgments The authors want to thank the anonymous referees for bringing certain accelerations of the EM algorithm to our attention.

Funding Open Access funding enabled and organized by Projekt DEAL. This work received funding from the German Research Foundation (DFG) within the project STE 571/16-1.

References

1. Abramowitz, M., Stegun, I.A.: Handbook of mathematical functions: with formulas, graphs, and mathematical tables, volume 55 Courier Corporation (1965)
2. Anderson, D.G.: Iterative procedures for nonlinear integral equations. J. Assoc. Comput. Mach. **12**, 547–560 (1965)
3. Antoniadis, A., Leporini, D., Pesquet, J.-C.: Wavelet thresholding for some classes of non-Gaussian noise. Statis. Neerlandica **56**(4), 434–453 (2002)
4. Banerjee, A., Maji, P.: Spatially constrained Student's t-distribution based mixture model for robust image segmentation. J. Mathe. Imag. Vision **60**(3), 355–381 (2018)
5. Byrne, C.L.: The EM algorithm: theory, applications and related methods. Lecture notes university of massachusetts (2017)
6. Ding, M., Huang, T., Wang, S., Mei, J., Zhao, X.: Total variation with overlapping group sparsity for deblurring images under Cauchy noise. Appl. Math. Comput. **341**, 128–147 (2019)
7. Fang, H.-R., Saad, Y.: Two classes of multisecant methods for nonlinear acceleration. Numer. Linear Algebra Appli. **16**(3), 197–221 (2009)
8. Gerogiannis, D., Nikou, C., Likas, A.: The mixtures of Student's t-distributions as a robust framework for rigid registration. Image Vis. Comput. **27**(9), 1285–1294 (2009)

9. Henderson, N.C., Varadhan, R.: Damped Anderson acceleration with restarts and monotonicity control for accelerating EM and EM-like algorithms. J. Comput. Graph. Stat. **28**(4), 834–846 (2019)
10. Kendall, M.G.: A new measure of rank correlation. Biometrika **30**(1/2), 81–93 (1938)
11. Kendall, M.G.: The treatment of ties in ranking problems. Biometrika 239–251 (1945)
12. Kent, J.T., Tyler, D.E., Vard, Y.: A curious likelihood identity for the multivariate *t*-distribution. Communications in Statistics-Simulation and Computation **23**(2), 441–453 (1994)
13. Lange, K.L., Little, R.J., Taylor, J.M.: Robust statistical modeling using the t distribution. J. Am. Stat. Assoc. **84**(408), 881–896 (1989)
14. Lanza, A., Morigi, S., Sciacchitano, F., Sgallari, F.: Whiteness constraints in a unified variational framework for image restoration. J. Mathe. Imag. Vision **60**(9), 1503–1526 (2018)
15. Laus, F.: Statistical Analysis and Optimal Transport for Euclidean and Manifold-Valued Data. PhD Thesis, TU Kaiserslautern (2020)
16. Laus, F., Pierre, F., Steidl, G.: Nonlocal myriad filters for Cauchy noise removal. J. Math. Imag. Vision **60**(8), 1324–1354 (2018)
17. Laus, F., Steidl, G.: Multivariate myriad filters based on parameter estimation of student-t distributions. SIAM J Imaging Sci **12**(4), 1864–1904 (2019)
18. Lebrun, M., Buades, A., Morel, J.-M.: A nonlocal Bayesian image denoising algorithm. SIAM J. Imag. Sci. **6**(3), 1665–1688 (2013)
19. Liu, C., Rubin, D.B.: The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. Biometrika **81**(4), 633–648 (1994)
20. Liu, C., Rubin, D.B.: ML estimation of the *t* distribution using EM and its extensions, ECM and ECME. Stat. Sin. **5**(1), 19–39 (1995)
21. McLachlan, G., Krishnan, T.: The EM algorithm and extensions. John wiley and sons inc (1997)
22. McLachlan, G., Peel, D.: Robust cluster analysis via mixtures of multivariate *t*-distributions. volume 1451 of Lecture Notes in Computer Science. Springer, New York (1998)
23. Mei, J.-J., Dong, Y., Huang, T.-Z., Yin, W.: Cauchy noise removal by nonconvex ADMM with convergence guarantees. J. Sci. Comput. **74**(2), 743–766 (2018)
24. Meng, X.-L., Van Dyk, D.: The EM algorithm - an old folk-song sung to a fast new tune. J. Royal Statis. Soc. :, Series B (Statis. Methodol.) **59**(3), 511–567 (1997)
25. Nguyen, T.M., Wu, Q.J.: Robust Student's-*t* mixture model with spatial constraints and its application in medical image segmentation. IEEE Trans. Med. Imaging **31**(1), 103–116 (2012)
26. Peel, D., McLachlan, G.J.: Robust mixture modelling using the *t* distribution. Stat. Comput. **10**(4), 339–348 (2000)
27. Petersen, K.B., Pedersen, M.S.: The Matrix Cookbook. Technical University of Denmark, Lecture Notes (2008)
28. Sciacchitano, F., Dong, Y., Zeng, T.: Variational approach for restoring blurred images with Cauchy noise. SIAM J. Imag. Sci. **8**(3), 1894–1922 (2015)
29. Sfikas, G., Nikou, C., Galatsanos, N.: Robust image segmentation with mixtures of Student's *t*-distributions. In: 2007 IEEE International Conference on Image Processing, volume 1, pages I – 273–I –276 (2007)
30. Sutour, C., Deledalle, C.-A., Aujol, J.-F.: Estimation of the noise level function based on a nonparametric detection of homogeneous image regions. SIAM J. Imag. Sci. **8**(4), 2622–2661 (2015)
31. Van Den Oord, A., Schrauwen, B.: The Student-*t* mixture as a natural image patch prior with application to image compression. J. Mach. Learn. Res. **15**(1), 2061–2086 (2014)
32. Van Dyk, D.A.: Construction, Implementation, and Theory of Algorithms Based on Data Augmentation and Model Reduction. The University of Chicago, PhD Thesis (1995)
33. Varadhan, R., Roland, C.: Simple and globally convergent methods for accelerating the convergence of any EM algorithm. Scandinavian. J. Statis. Theory Appli **35**(2), 335–353 (2008)
34. Yang, Z., Yang, Z., Gui, G.: A convex constraint variational method for restoring blurred images in the presence of alpha-stable noises. Sensors **18**(4), 1175 (2018)
35. Zhou, Z., Zheng, J., Dai, Y., Zhou, Z., Chen, S.: Robust non-rigid point set registration using Student's-*t* mixture model. PloS one **9**(3), e91381 (2014)