A TYLER-TYPE ESTIMATOR OF LOCATION AND SCATTER LEVERAGING RIEMANNIAN OPTIMIZATION

A. Collas[1], F. Bouchard[2], A. Breloy[3], C. Ren[1], G. Ginolhac[4], J.-P. Ovarlez[1,5]

April 19, 2021

[1]SONDRA, CentraleSupélec, Université Paris-Saclay
[2]CNRS, L2S, CentraleSupélec, Université Paris-Saclay
[3]LEME, Université Paris Nanterre
[4]LISTIC, Université Savoie Mont Blanc
[5]DEMR, ONERA, Université Paris-Saclay

Table of contents

Introduction

Introduction

Many signal processing applications require first and second order statistical moments of the sample set $\{\boldsymbol{x_i}\}_{i=1}^n$. To be robust to heavy-tailed distributions or outliers, [Mar76] proposed the *M*-estimators:

$$\begin{cases} \boldsymbol{\mu} = \left(\sum_{i=1}^{n} u_1(t_i) \right)^{-1} \sum_{i=1}^{n} u_1(t_i) \boldsymbol{x_i} \triangleq \mathcal{H}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ \boldsymbol{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} u_2(t_i)(\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^H \triangleq \mathcal{H}_{\boldsymbol{\Sigma}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{cases} \quad (1)$$

where $t_i \triangleq (\boldsymbol{x_i} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{x_i} - \boldsymbol{\mu})$, u_1 and u_2 are functions that respect Maronna's conditions [Mar76].

Under certain conditions [Mar76],

$$\begin{cases} \boldsymbol{\mu}_{k+1} = \mathcal{H}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\ \boldsymbol{\Sigma}_{k+1} = \mathcal{H}_{\boldsymbol{\Sigma}}(\boldsymbol{\mu}_{k+1}, \boldsymbol{\Sigma}_k) \end{cases} \quad (2)$$

converge towards a unique solution satisfying (1).

Introduction

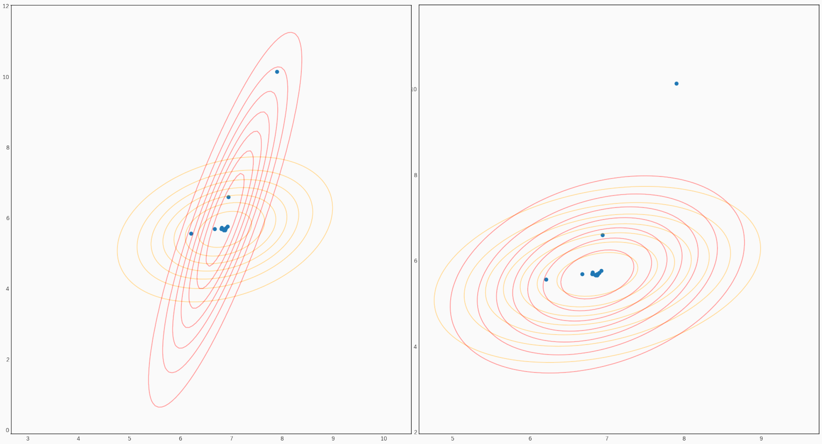

Figure 1: Example of a set of points generated with a heavy-tailed distribution with real probability density function (p.d.f.) in orange. Estimated p.d.f. are in red: Gaussian estimators on the left, our estimators on the right.

Data model

Data model

Let n data points $\boldsymbol{x_i} \in \mathbb{C}^p$ distributed according to the model:

$$\boldsymbol{x_i} \underset{d}{=} \boldsymbol{\mu} + \sqrt{\tau_i}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{u_i} \tag{3}$$

where $\boldsymbol{\mu} \in \mathbb{C}^p$, $\boldsymbol{\tau} \in (\mathbb{R}_*^+)^n$, $\boldsymbol{\Sigma} \in \mathcal{SH}_p^{++}$ and $\boldsymbol{u_i} \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_p)$.
Hence, $\tau_i > 0$, $\boldsymbol{\Sigma} \succ 0$ and $\det(\boldsymbol{\Sigma}) = 1$.

Thus, $\boldsymbol{x_i}$ follows a Compound Gaussian distribution, *i.e.*

$$\boldsymbol{x_i} \sim \mathbb{C}\mathcal{N}(\boldsymbol{\mu}, \tau_i\boldsymbol{\Sigma}). \tag{4}$$

Definition

The set of parameters is $\mathcal{M}_{p,n} = \mathbb{C}^p \times (\mathbb{R}_*^+)^n \times \mathcal{SH}_p^{++}$.

Remark

The textures τ_i are assumed to be unknown and deterministic.

Data model - Log-likelihood

Hence, $\forall \theta = (\boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\Sigma}) \in \mathcal{M}_{p,n}$ the negative log-likelihood is

$$L(\theta) = \sum_{i=1}^{n} \left[\log\det\left(\tau_i \boldsymbol{\Sigma}\right) + \frac{(\boldsymbol{x_i} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{x_i} - \boldsymbol{\mu})}{\tau_i} \right] . \tag{5}$$

And the Maximum Likelihood Estimate satisfies

$$\begin{cases} \boldsymbol{\mu} = \left(\sum_{i=1}^{n} \frac{1}{\tau_i} \right)^{-1} \sum_{i=1}^{n} \frac{\boldsymbol{x_i}}{\tau_i} \\ \boldsymbol{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^H}{\tau_i} \\ \tau_i = \frac{1}{p} (\boldsymbol{x_i} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{x_i} - \boldsymbol{\mu}). \end{cases} \tag{6}$$

Remark

(6) *coincides with the fixed point* (1) *for* $u_1(t) = u_2(t) = p/t$ *but does not satisfy Maronna's conditions. The associated fixed-point iterations* (2) *generally diverge in practice !*

Riemannian optimization

Riemannian optimization

A tool of interest for contrained parameters estimation is the Riemannian geometry. Briefly, a Riemannian manifold is a couple $(\mathcal{M}, \langle\cdot,\cdot\rangle_\theta^{\mathcal{M}})$ where

- $\mathcal{M}$ is a *smooth manifold* (*i.e.* a locally Euclidean set).
- $\langle\cdot,\cdot\rangle_\theta^{\mathcal{M}}$ is an inner product, on $T_\theta\mathcal{M}$, called the *Riemannian metric*.

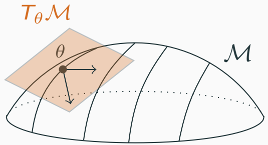

Figure 2: A manifold $\mathcal{M}$ with its tangent space $T_\theta\mathcal{M}$.

The vector space $T_\theta\mathcal{M}$ is called the tangent space and is the linearization of $\mathcal{M}$ at θ.

Remark

With the Riemmanian geometry of $\mathcal{M}$ defined, we can optimize a function $f : \mathcal{M} \to \mathbb{R}$.

For a full review on this topic: Optimization algorithms on matrix manifolds [AMS08; Smi05].

Riemannian optimization

The goal is to minimize the negative log-likelihood:

$$\hat{\theta} = \underset{\theta \in \mathcal{M}_{p,n}}{\arg\min}\, L(\theta). \tag{7}$$

where $\mathcal{M}_{p,n} = \mathbb{C}^p \times (\mathbb{R}_*^+)^n \times \mathcal{SH}_p^{++}$.

Remark

$\mathcal{M}_{p,n}$ is a product manifold of sets which have well known Riemannian manifolds.

The tangent space of $\mathcal{M}_{p,n}$ at θ denoted $T_\theta \mathcal{M}_{p,n}$ is the product of the tangent spaces of $\mathbb{C}^p$, $(\mathbb{R}_*^+)^n$ and $\mathcal{SH}_p^{++}$ i.e,

$$T_\theta \mathcal{M}_{p,n} = \left\{ \xi \in \mathbb{C}^p \times \mathbb{R}^n \times \mathcal{H}_p : \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}}) = 0 \right\}, \tag{8}$$

where $\mathcal{H}_p$ is the Hermitian set.

Riemannian optimization

Definition

Let $\xi, \eta \in T_\theta \mathcal{M}_{p,n}$, the Riemannian metric at θ is defined as,

$$\langle \xi, \eta \rangle_\theta^{\mathcal{M}_{p,n}} = \langle \boldsymbol{\xi}_{\boldsymbol{\mu}}, \boldsymbol{\eta}_{\boldsymbol{\mu}} \rangle_{\boldsymbol{\mu}}^{\mathbb{C}^p} + \langle \boldsymbol{\xi}_{\boldsymbol{\tau}}, \boldsymbol{\eta}_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n} + \langle \boldsymbol{\xi}_{\boldsymbol{\Sigma}}, \boldsymbol{\eta}_{\boldsymbol{\Sigma}} \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}}, \quad (9)$$

with

- $\langle \boldsymbol{\xi}_{\boldsymbol{\mu}}, \boldsymbol{\eta}_{\boldsymbol{\mu}} \rangle_{\boldsymbol{\mu}}^{\mathbb{C}^p} = \mathfrak{Re}\{\boldsymbol{\xi}_{\boldsymbol{\mu}}^H \boldsymbol{\eta}_{\boldsymbol{\mu}}\}$,
- $\langle \boldsymbol{\xi}_{\boldsymbol{\tau}}, \boldsymbol{\eta}_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n} = (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}})^T (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\eta}_{\boldsymbol{\tau}})$, where $\odot$ and $.^{\odot t}$ denote the elementwise product and power operators respectively,
- $\langle \boldsymbol{\xi}_{\boldsymbol{\Sigma}}, \boldsymbol{\eta}_{\boldsymbol{\Sigma}} \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}} = \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\eta}_{\boldsymbol{\Sigma}}\right)$.

Remark

$\left(\mathcal{M}_{p,n}, \langle \cdot, \cdot \rangle_\cdot^{\mathcal{M}_{p,n}}\right)$ is a Riemannian manifold and all its geometrical elements (exponential mapping, parallel transport, and distance) are derived from Riemannian geometries of $\mathbb{C}^p$, $(\mathbb{R}_*^+)^n$, and $\mathcal{SH}_p^{++}$.

Riemannian optimization

Input : Initial iterate $\theta_1 \in \mathcal{M}_{p,n}$.
Output: Sequence of iterates $\{\theta_k\}$
$k := 1$;
$\xi_1 := -\operatorname{grad} L(\theta_1)$;
while *no convergence* **do**
    Compute a step size α_k (e.g see [AMS08, §4.2]) and set
    $\theta_{k+1} := R_{\theta_k}^{\mathcal{M}_{p,n}}(\alpha_k \xi_k)$;
    Compute β_{k+1} (e.g see [AMS08, §8.3]) and set
    $\xi_{k+1} := -\operatorname{grad} L(\theta_{k+1}) + \beta_{k+1} \mathcal{T}_{\theta_k,\theta_{k+1}}^{\mathcal{M}_{p,n}}(\xi_k)$;
    $k := k + 1$;
end

Algorithm 1: Riemannian conjugate gradient [AMS08]

- $\operatorname{grad} L(\theta_k)$ is the Riemannian gradient, computed in Proposition 1.
- $R_{\theta_k}^{\mathcal{M}_{p,n}}$ is a retraction provided in Section 3.1.
- $\mathcal{T}_{\theta_k,\theta_{k+1}}^{\mathcal{M}_{p,n}}$ is a vector transport provided in Section 3.1.

Numerical experiment

Numerical experiment

We compare the mean squared errors of different estimators on simulated data according to model (3).

1. Gaussian estimators: sample mean $\boldsymbol{\mu}^{\mathrm{G}}$ and SCM denoted $\boldsymbol{\Sigma}^{\mathrm{G}}$.
2. Two-step estimation: the sets $\{\boldsymbol{x_i}\}_{i=1}^{n}$ are centered with $\boldsymbol{\mu}^{\mathrm{G}}$ and then we estimate $\boldsymbol{\Sigma}$ using Tyler's M-estimator [Tyl87]. The estimator is denoted $\boldsymbol{\Sigma}^{\mathrm{Ty},\boldsymbol{\mu}^{\mathrm{G}}}$.
3. Tyler's joint estimators of location and scatter matrix [Tyl87] denoted $\boldsymbol{\mu}^{\mathrm{Ty}}$ and $\boldsymbol{\Sigma}^{\mathrm{Ty}}$. These estimators corresponds to (1) with $u_1(t) = \sqrt{p/t}$ and $u_2(t) = p/t$. It converges in practice unlike fixed-point equations of the MLE.
4. Tyler's M-estimator with location known [Tyl87]. The sets $\{\boldsymbol{x_i}\}_{i=1}^{n}$ are centered with $\boldsymbol{\mu}$ and then we estimate $\boldsymbol{\Sigma}$. The estimator is denoted $\boldsymbol{\Sigma}^{\mathrm{Ty},\boldsymbol{\mu}}$.
5. Our estimators $\boldsymbol{\mu}^{\mathrm{CG}}$ and $\boldsymbol{\Sigma}^{\mathrm{CG}}$: a Riemannian conjugate gradient to minimize (5) on $\mathcal{M}_{p,n}$ performed with the library *Pymanopt* [TKW16].

Numerical experiment

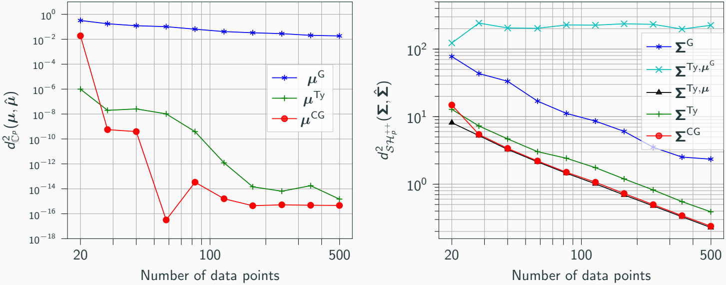

Figure 3: Mean squared errors over 200 simulated sets $\{\boldsymbol{x_i}\}_{i=1}^{n}$ ($p = 10$) with respect to the number n of samples for the considered estimators $\hat{\boldsymbol{\mu}} \in \{\boldsymbol{\mu}^{\mathrm{G}}, \boldsymbol{\mu}^{\mathrm{Ty}}, \boldsymbol{\mu}^{\mathrm{CG}}\}$ and $\hat{\boldsymbol{\Sigma}} \in \{\boldsymbol{\Sigma}^{\mathrm{G}}, \boldsymbol{\Sigma}^{\mathrm{Ty},\boldsymbol{\mu}^{\mathrm{G}}}, \boldsymbol{\Sigma}^{\mathrm{Ty},\boldsymbol{\mu}}, \boldsymbol{\Sigma}^{\mathrm{Ty}}, \boldsymbol{\Sigma}^{\mathrm{CG}}\}$.

Remark

$\boldsymbol{\mu}^{\mathrm{CG}}$ and $\boldsymbol{\Sigma}^{\mathrm{CG}}$, Riemannian Conjugate Gradient estimators, perform better than other estimators. For $n \geq 3p$, $\boldsymbol{\Sigma}^{\mathrm{CG}}$ perform as good as Tyler's estimator with $\boldsymbol{\mu}$ known, $\boldsymbol{\Sigma}^{\mathrm{Ty},\boldsymbol{\mu}}$, [Tyl87] !

Conclusion

This paper has proposed an efficient Riemannian optimization-based procedure to jointly estimate the location and scatter matrix of a Compound Gaussian distribution. A Riemannian geometry of the parameter manifold $\mathcal{M}_{p,n}$ has been described in order to derive a Riemannian conjugate gradient optimizer. This algorithm reaches performance close to the MLE of the "known location" case, which illustrates the interest of the proposed approach.

References

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2008. ISBN: 0691132984, 9780691132983.

R. A. Maronna. "Robust M-Estimators of Multivariate Location and Scatter". In: *Ann. Statist.* 4.1 (Jan. 1976), pp. 51–67. DOI: 10.1214/aos/1176343347. URL: https://doi.org/10.1214/aos/1176343347.

S. Smith. "Covariance, Subspace, and Intrinsic Cramér-Rao Bounds". In: *Signal Processing, IEEE Transactions on* 53 (June 2005), pp. 1610–1630. DOI: 10.1109/TSP.2005.845428.

J. Townsend, N. Koep, and S. Weichwald. "Pymanopt: A Python Toolbox for Optimization on Manifolds Using Automatic Differentiation". In: *J. Mach. Learn. Res.* 17.1 (Jan. 2016), pp. 4755–4759. ISSN: 1532-4435.

D. E. Tyler. "A Distribution-Free M-Estimator of Multivariate Scatter". In: *Ann. Statist.* 15.1 (Mar. 1987), pp. 234–251. DOI: 10.1214/aos/1176350263. URL: https://doi.org/10.1214/aos/1176350263.