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]

[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

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, [1] 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 [1].

Under certain conditions [1],

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

converge towards a unique solution satisfying (1).

Data model

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

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

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 $\det(\boldsymbol{\Sigma}) = 1$. Also, the textures τ_i are assumed to be unknown and deterministic.

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

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

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

Likelihood and MLE

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

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

By derivation we get that the Maximum Likelihood Estimate (MLE) satisfies

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

Thus, (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 geometry

A tool of interest for constrained 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*.

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

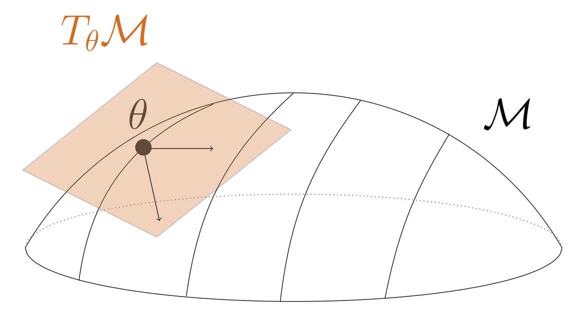

Figure 1. A manifold $\mathcal{M}$ with its tangent space $T_\theta\mathcal{M}$.

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: see [2, 3].

Minimization of the negative log-likelihood L on $\mathcal{M}_{p,n}$

The goal is to minimize the negative log-likelihood (5):

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

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

$\mathcal{M}_{p,n}$ 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,

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

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

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

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

with

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

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

Optimization algorithm

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 [2, §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 [2, §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 [2]

- $\operatorname{grad} L(\theta_k)$ 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 experiments

We estimate location $\boldsymbol{\mu} \in \mathbb{C}^p$ and scatter matrix $\boldsymbol{\Sigma} \in \mathcal{SH}_p^{++}$ from simulated data.

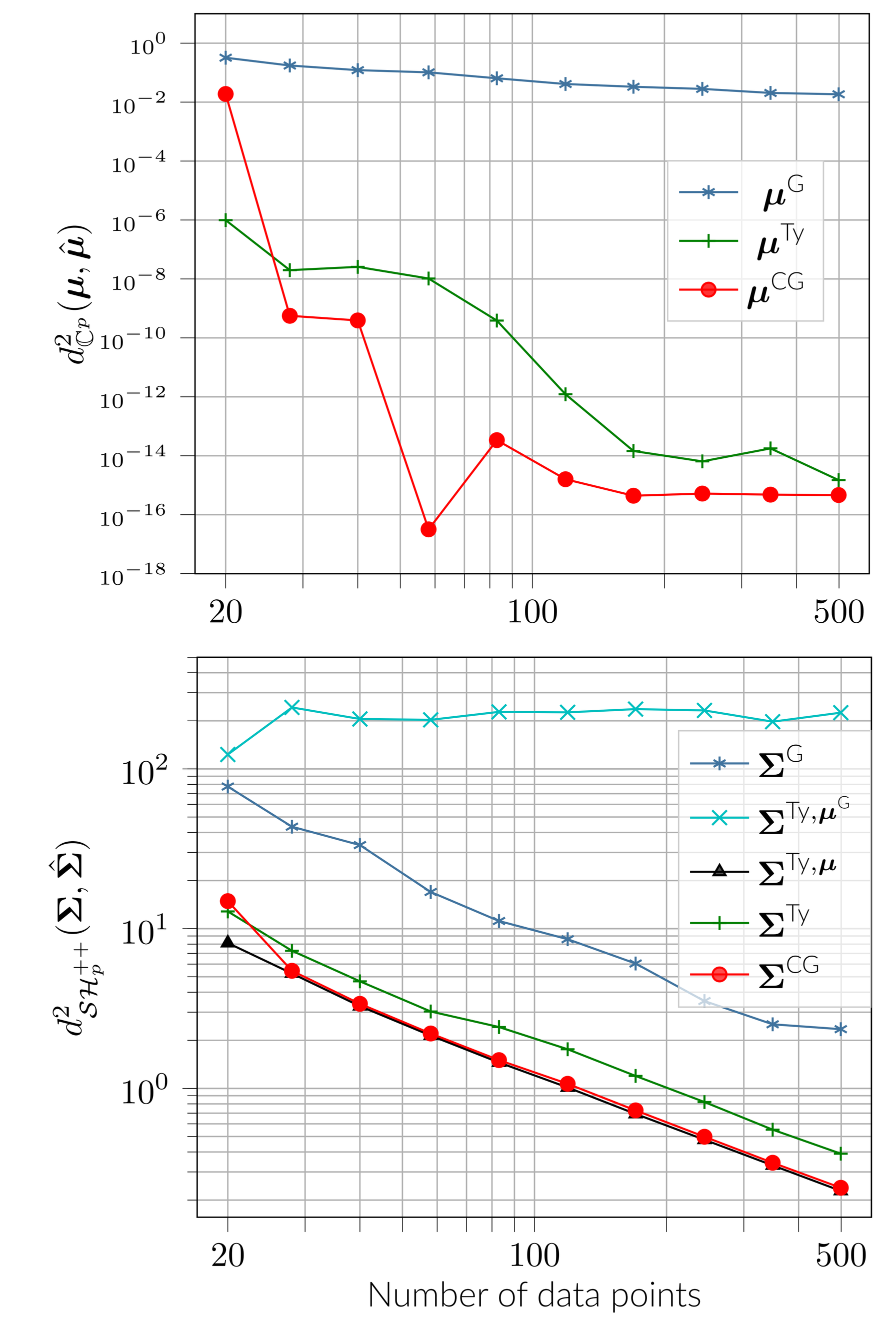

Figure 2. Mean squared errors over 200 simulated sets $\{\boldsymbol{x_i}\}_{i=1}^n$ ($p = 10$) with respect to the number n of samples for the considered estimators $\hat{\boldsymbol{\mu}} \in \{\boldsymbol{\mu}^{G},\boldsymbol{\mu}^{Ty},\boldsymbol{\mu}^{CG}\}$ and $\hat{\boldsymbol{\Sigma}} \in \{\boldsymbol{\Sigma}^{G},\boldsymbol{\Sigma}^{Ty,\boldsymbol{\mu}^G},\boldsymbol{\Sigma}^{Ty,\boldsymbol{\mu}},\boldsymbol{\Sigma}^{Ty},\boldsymbol{\Sigma}^{CG}\}$.

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

$\boldsymbol{\mu}^{CG}$ and $\boldsymbol{\Sigma}^{CG}$, Riemannian Conjugate Gradient estimators, perform better than other estimators.

References

[1] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 4, pp. 51–67, 01 1976.
[2] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2008.
[3] S. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *Signal Processing, IEEE Transactions on*, vol. 53, pp. 1610 – 1630, 06 2005.
[4] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, pp. 234–251, 03 1987.
[5] J. Townsend, N. Koep, and S. Weichwald, "Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation," *J. Mach. Learn. Res.*, vol. 17, p. 4755–4759, Jan. 2016.

antoine.collas@centralesupelec.fr