# On the Convergence of Maronna's M-Estimators of Scatter

Yacine Chitour, Romain Couillet, Frédéric Pascal

# On the convergence of Maronna's $M$-estimators of scatter

Yacine Chitour, Romain Couillet *Member, IEEE,* and Frédéric Pascal *Senior Member, IEEE*

***Abstract*—In this paper, we propose an alternative proof for the uniqueness of Maronna's $M$-estimator of scatter [1] for $N$ vector observations $\mathbf{y}_1, \dots, \mathbf{y}_N \in \mathbf{R}^m$ under a mild constraint of linear independence of any subset of $m$ of these vectors. This entails in particular almost sure uniqueness for random vectors $\mathbf{y}_i$ with a density as long as $N > m$. This approach allows to establish further relations that demonstrate that a properly normalized Tyler's $M$-estimator of scatter [2] can be considered as a limit of Maronna's $M$-estimator. More precisely, the contribution is to show that each $M$-estimator, verifying some mild conditions, converges towards a particular Tyler's $M$-estimator. These results find important implications in recent works on the large dimensional (random matrix) regime of robust $M$-estimation.**

## I. Introduction

Subsequent to Huber's introduction of robust statistics in [3], Maronna proposed in [1] a class of robust estimates for scatter matrices defined as the solution of an implicit equation. In [1], the existence and uniqueness of such a solution are proved, under conditions involving both the ratio $c_N := m/N$ of the population dimension $m$ and the sample size $N$, and the parametrization of the estimate. This constraint was largely relaxed in [4], [5]. With the recent renewed interest in robust $M$-estimation under the random matrix regime $N, m \to \infty$ with $c_N \to c_\infty \in (0,1)$ [6]–[9], alternative proofs of existence and uniqueness have appeared motivated by this assumption of large $m$. While Maronna's original results are valid for any (well-behaved) set of samples satisfying the condition on $c_N$, the results in e.g. [6] are expressed in probabilistic terms and are only valid for all large $m, N$.

Based on the ideas from [10]–[12], the present article proposes an alternative proof to [4] to show existence and uniqueness for all well-behaved set of samples with a known location parameter and for any $c_N \in (0,1)$. More importantly, by a proper parametrization of the weight function appearing in Maronna's estimator, we prove that some sequences of Maronna's $M$-estimators converge to a unique Tyler's distribution-free $M$-estimator of scatter [2]. This result is a novel property of the Tyler's $M$-estimators, rigorously proved in this work. This completes the recent result (Theorem 1 of [13]) stating that the Tyler's $M$-estimator is the Maximum Likelihood estimator (MLE) of the scatter for various complex elliptically symmetric (CES) distributions as well as for the angular central Gaussian (ACG) distributions [14].

The paper is organized as follows: Section II presents our main results as well as Monte-Carlo simulations that corroborate our theoretical claims, the proofs of which are provided in Section III. Section IV draws some conclusions and perspectives of this work.

Chitour is with Laboratoire des Signaux et Systèmes at Supélec, 91192 Gif s/Yvette, France and Université Paris Sud, Orsay, France yacine.chitour@lss.supelec.fr. Couillet is with the Telecommunications Department at Supélec romain.couillet@supelec.fr. Pascal is with the SONDRA laboratory at Supélec frederic.pascal@supelec.fr. Chitour and Pascal's works were partially supported by the iCODE institute, research project of the Idex Paris-Saclay, while Couillet's work is funded by ERC–MORE EC–120133.

## II. Notations and statement of the results

Let $\mathbf{R}_+$ (resp. $\mathbf{R}_+^*$) be the (resp. strictly) positive real line. We use $M_m(\mathbf{R})$ and $\mathrm{Sym}_m$ to denote the vector space of $m \times m$ matrices with real entries and the linear subspace of $M_m(\mathbf{R})$ made of the symmetric matrices, respectively. We also use $\mathrm{Sym}_m^+$ and $\mathrm{PSD}_m$ to denote the non trivial cones in $M_m(\mathbf{R})$ of the non negative symmetric matrices and of the symmetric positive definite matrices, respectively. Also, $(\cdot)^T$ stands for the transpose, $\mathrm{Tr}(\cdot)$ and $\det(\cdot)$ for the trace and the determinant. On $M_m(\mathbf{R})$, we use the inner product defined by the Frobenius norm $\|\mathbf{A}\| = \sqrt{\mathrm{Tr}(\mathbf{A}\mathbf{A}^T)}$. We also use $\leq$ to denote the partial order on $\mathrm{Sym}_m$ and $\mathbf{I}_m$ the $m \times m$ identity matrix. Functions of two non negative real variables $(t, x)$ will be considered. If $f$ is such a function, we use $f_t$, $f_x$, $f_{tx}$, ... to denote (when defined) the partial derivatives of $f$ with respect to $t$ and/or $x$.

**Definition II.1** *A family $(\mathbf{y}_i)_{1\leq i\leq N}$ of vectors in $\mathbf{R}^m$ is admissible if*

- $(C1)$ *for $1 \leq i \leq N$, $\|\mathbf{y}_i\| = 1$;*
- $(C2)$ *the vectors in any subset of size $m$ of $\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}$ are linearly independent*

This definition straightforwardly implies that if $(\mathbf{y}_i)_{1\leq i\leq N}$ is an admissible family of vectors in $\mathbf{R}^m$ and if $m$ vectors (say) $\mathbf{y}_1, \cdots, \mathbf{y}_m$ which are then linearly independent by $(C2)$ are fixed, for $m+1 \leq l \leq N$, we can write $\mathbf{y}_l = \sum_{j=1}^{n} \gamma_{lj}\mathbf{y}_j$. Then, $\gamma_{lj} \neq 0$ for every $1 \leq j \leq m$ and $m+1 \leq l \leq N$.

Let us now consider maps $u : (\mathbf{R}_+^*)^2 \to \mathbf{R}_+$ of class $C^1$ satisfying:

- $(U1)$ $u(t,\cdot)$ is strictly decreasing;
- $(U2)$ for every $t > 0$, $v(t,x) := x \mapsto xu(t,x)$ is increasing on $\mathbf{R}_+$ and $l_t := \sup_{x\geq 0} v(t,x) > m$;

We furthermore define, for every $x > 0$, $u(0,x) = \frac{m}{x}$. Note that, by continuity of $u$, $\forall x > 0$, $\lim_{t\to 0^+} v(t,x) = m$. Also, according to $(U1)$ and $(U2)$, for each $t, x > 0$,

$$v(t,x) = m + tv_1(x) + tw(t,x), \tag{1}$$

with $v_1(\cdot) := v_t(0,\cdot)$ and $\forall x > 0, \lim_{t\to 0} w(t,x) = 0$. By a simple computation, one has that $v_1$ is a nondecreasing function on $\mathbf{R}_+^*$.

For further use, we introduce the following additional notation. Let $x_t > 0$ be the unique positive number such that, $\forall t > 0,\ v(t,x_t) = x_t u(t,x_t) = m$.

We further consider the following assumption

$$(U3) \quad \begin{cases} v_x := dv/dx > 0 \\ v_1 \text{ is increasing} \\ 0 < \liminf_{t\to 0} x_t \leq \limsup_{t\to 0} x_t < \infty. \end{cases}$$

If the latter occurs and $u$ is of class $C^2$, then $w(t,x) = tw_1(x) + o(t)$, with $w_1(\cdot) := w_t(0,\cdot)$ continuous on $(\mathbf{R}_+^*)^2$, the convergence in $(U2)$ is uniform in $x$ on any compact of $\mathbf{R}_+^*$ and $x_t$ converges to the unique solution $x_0$ of $v_1(x) = 0$.

We use $\bar{u}(t,x)$ to denote the particular function

$$\bar{u}(t,x) = \frac{m(1+t)}{x+t} \tag{2}$$

which is analytic on every compact of $(\mathbf{R}_+)^2 \setminus \{(0,0)\}$. Moreover, $\bar{l}_t = m(1+t)$, $\bar{v_1}(x) = m(1-\frac{1}{x})$ and $\bar{w}(t,x) = \frac{-mt}{t+x}$.

The objective of the work is to study the solutions of the equation given, for all $t > 0$, by

$$(\mathrm{Eq})_t \qquad \mathbf{M} = \frac{1}{N}\sum_{i=1}^{N} u(t, \mathbf{y}_i^T \mathbf{M}^{-1}\mathbf{y}_i)\mathbf{y}_i\mathbf{y}_i^T.$$

and to characterize them in the limit where $t \to 0$. Taking into account our definitions, if a solution to $(\mathrm{Eq})_t$ exists, it must belong to $\mathrm{PSD}_m$.

Remark that the condition M of [4] also imposes a "strictly" increasing $v$ which excludes e.g. the Huber $M$-estimator.

To state our results, we need to consider the set of solutions of the equation $(\mathrm{Eq})_0$ (that defines the Tyler's $M$-estimator) given by

$$(\mathrm{Eq})_0 \qquad \mathbf{M} = \frac{m}{N}\sum_{i=1}^{N} \frac{1}{\mathbf{y}_i^T\mathbf{M}^{-1}\mathbf{y}_i}\mathbf{y}_i\mathbf{y}_i^T.$$

Recall from [10] that the set of solutions of $(\mathrm{Eq})_0$ is the half-line $\mathbf{R}_+^*\,\mathbf{P}$ in $\mathrm{PSD}_m$, where $\mathbf{P}$ is the unique solution of $(\mathrm{Eq})_0$ with $\mathrm{Tr}(\mathbf{P}) = m$.

Our main result is the following theorem.

**Theorem II.2** *Let $(\mathbf{y}_i)_{1\leq i\leq N}$ be an admissible family of vectors in $\mathbf{R}^m$ and $u : (\mathbf{R}_+)^2 \setminus \{(0,0)\} \to \mathbf{R}_+$ be a $C^1$ function verifying $(U1)$–$(U2)$. Then,*

- $(A)$ *$\forall t > 0$, $(\mathrm{Eq})_t$ admits a unique solution, $\mathbf{M}(t)$.*
- $(B)$ *If, furthermore, $u$ is $C^2$ and satisfies $(U3)$, then the mapping $t \mapsto \mathbf{M}(t)$ is continuous and $\lim_{t\to 0}\mathbf{M}(t) = \mathbf{M}_0$ the solution of $(\mathrm{Eq})_0$ given by $\mathbf{M}_0 = \xi_u\mathbf{P}$ with $\xi_u > 0$ unique solution to*

$$\sum_{i=1}^{N} v_1\left(\frac{\mathbf{y}_i^T\mathbf{P}^{-1}\mathbf{y}_i}{\xi}\right) = 0. \tag{3}$$

*In particular, for $u = \bar{u}$, $\mathbf{M}_0 = \mathbf{P}$, i.e., $\xi_{\bar{u}} = 1$.*

*Proof of Theorem II.2.* The proof is postponed in the next section. ∎

**Remark II.3**

1) The interest of Theorem II.2, in addition to providing an alternative proof for the existence and uniqueness, lies in the convergence of all $M$-estimators to a Tyler's $M$-estimator. This limit can be different (by a scale factor) from one $M$-estimator to another. While this result was expected, this paper rigorously proves it.

2) Moreover, the theorem provides a way of understanding why the Tyler's estimator is the outmost robust[1] $M$-estimator. Indeed, considering a ML approach, the weight function $u(t,x)$ is derived from the observations probability density function (PDF) and in such a case, $t \to 0$ means that the underlying distribution becomes more and more heavy-tailed. For instance, considering $t$ as the exponent parameter of a Generalized Gaussian distribution or of a W-distribution, the smaller the value of $t > 0$ is, the heavier-tailed is the distribution. This is also the case for the degree of freedom of a Student-t distribution or the shape parameters of a K-distribution or of a Compound-Gaussian with inverse Gaussian texture (see [14] for more details). In all these cases, the MLEs satisfy the assumptions of Theorem II.2 (at least for small values of $t$) and should be more robust when the distribution is heavier-tailed. To summarize, this result theoretically motivates the use of the Tyler's estimator, since it will perform similarly as MLEs in heavy-tailed distribution contexts.

[1] Here the robustness has to be understood as the classical property considered in the robust estimation theory literature, see e.g. [15]

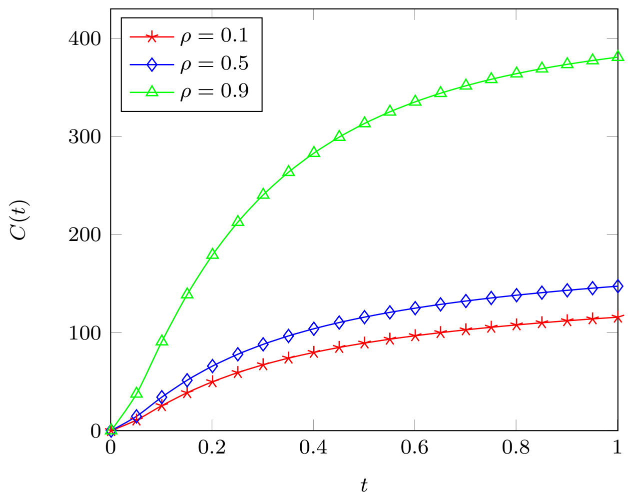


Fig. 1. Convergence of $M(t)$ towards $M_0$ when $t \to 0$ for $N = m + 1 = 51$. The criterion used is the MSE: $C(t) = E\left[\|M(t) - M_0\|_F^2\right]$.

To illustrate Theorem II.2, Figure 1 presents the mean square error $C(t) \triangleq E[\|\mathbf{M}(t) - \mathbf{M}_0\|_F^2]$ between Tyler's $M$-estimator and the Student-t MLE versus the parameter $t$, called the degree of freedom of the multivariate Student-t distribution [14], defined through the weight function $u(t,x) = \frac{m+t}{t+x}$. We take here $N = m+1 = 51$. The data are zero-mean Gaussian distributed with Toeplitz covariance matrix, the $(i,j)$ entry of which is equal to $\rho^{|i-j|}$, for some $\rho \in (0,1)$. As proved in Theorem II.2, Item (A) is illustrated in the case where $N = m+1$ while Item (B) is illustrated for the Student-t MLE for different population covariance matrices.

## III. Proof of Theorem II.2

The strategy of the proof is as follows: for every $t > 0$, we first build a positive functional $H(t,\cdot)$ over $\text{PSD}_m$ whose critical points (if any) are exactly the solutions of $(\text{Eq})_t$. To establish the existence of such critical points, we show that $H(t,\cdot)$ is uniformly bounded and tends to zero at the boundary of $\text{PSD}_m$. To obtain uniqueness, we show that solutions of $(\text{Eq})_t$ are all local strict maxima of $H(t,\cdot)$ and conclude by applying the mountain pass theorem (cf. [16]). This gives Item $(A)$. Item $(B)$ is then obtained using the implicit function theorem and some limiting arguments.

For $t > 0$, we define the function

$$h : \mathbf{R}_+^* \times \mathbf{R}_+ \to \mathbf{R}_+^*, \quad (t,x) \mapsto e^{-\frac{1}{m}\int_{x_t}^{x} u(t,y)dy}. \tag{4}$$

Then $-\frac{h_x}{h} = \frac{u}{m}$ and $h(t,x_t) = 1$. Set $h(0,x) = \frac{1}{x}$ for $x > 0$ and $g : \mathbf{R}_+^* \times \mathbf{R}_+ \to \mathbf{R}_+^*$ with $g(t,x) = xh(t,x)$.

In the case where $u = \bar{u}$, $\forall (t,x) \in (\mathbf{R}_+)^2 \setminus \{(0,0)\}$,

$$x_t \equiv 1, \ \bar{h}(t,x) = \left(\frac{1+t}{x+t}\right)^{1+t}, \ \bar{g}(t,x) = x\left(\frac{1+t}{x+t}\right)^{1+t}.$$

Then, define the functional $H(t,\cdot)$ as

$$H : \mathbf{R}_+^* \times \text{PSD}_n \to \mathbf{R}_+^*, \quad (t,\mathbf{M}) \mapsto \frac{\prod_{i=1}^N h(t,\mathbf{y}_i^T\mathbf{M}^{-1}\mathbf{y}_i)^m}{(\det \mathbf{M})^N} \tag{5}$$

as well as the functional considered in [10]

$$B : \text{PSD}_n \to \mathbf{R}_+^*, \quad \mathbf{M} \mapsto \frac{\prod_{i=1}^N h(0,\mathbf{y}_i^T\mathbf{M}^{-1}\mathbf{y}_i)^m}{(\det \mathbf{M})^N}. \tag{6}$$

**Lemma III.1** *For $t > 0$ and $\mathbf{M} \in \text{PSD}_m$, one has $-\mathbf{M}H_x(t,\mathbf{M})\mathbf{M}/NH(t,\mathbf{M}) = \mathbf{M} - \frac{1}{N}\sum_{i=1}^N u(t,\mathbf{y}_i^T\mathbf{M}^{-1}\mathbf{y}_i)$, with $H_x(t,\mathbf{M})$ the gradient of $H(t,\cdot)$. In particular, $\mathbf{M}$ is a solution of $(\text{Eq})_t$ if and only if $\mathbf{M}$ is a critical point of $H(t,\cdot)$.*

**Lemma III.2** $\forall t > 0, \mathbf{M} \in \text{PSD}_m$, $H(t,\mathbf{M}) \le B(\mathbf{M})$. *As a consequence, $\lim_{\mathbf{M}\to\partial\text{PSD}_m} H(t,\mathbf{M}) = 0$, so that $H(t,\cdot)$ admits critical points.*

*Proof of Lemma III.2.* An immediate calculus yields that $x \mapsto g(t,x)$ reaches its maximum 1 at $x = x_t$. As a consequence, for $t > 0, \mathbf{M} \in \text{PSD}_m$, $H(t,\mathbf{M}) \le B(\mathbf{M})$. Moreover, $\lim_{x\to 0,\infty} g(t,x) = \lim_{x\to 0,\infty} xh(t,x) = 0$. For the limit at $x = 0$, this is obvious. For $x \to \infty$, note that $\ln(g(t,x)) = \frac{1}{m}\int_{x_t}^{x} \frac{m - yu(t,y)}{y}dy$ and, since $m - l_t < 0$, it is equivalent to $(m - l_t)\ln(x)$ as $x \to \infty$. Consider now a sequence $(\mathbf{M}_k)_{k\ge 0}$ in $\text{PSD}_m$ converging to $\partial\text{PSD}_m$. For $k \ge 0$, set $\mathbf{M}_k = \rho_k\mathbf{N}_k$ with $\rho_k = \|\mathbf{M}_k\|$ and $\mathbf{N}_k = \frac{\mathbf{M}_k}{\rho_k}$. Note that $\partial\text{PSD}_m$ is made of matrices either non invertible or with norm going to infinity. Therefore, up to subsequences, either $(i)$ $(\mathbf{N}_k)_{k\ge 0}$ converges itself to $\partial\text{PSD}_m$ or $(b)$ the sequence $(\rho_k)_{k\ge 0}$ converges to zero or infinity and there exists $\exists \alpha > 0, \forall k \ge 0, \mathbf{N}_k \ge \alpha\mathbf{I}_m$. If Case $(i)$ occurs, then $\forall k \ge 0, H(t,\mathbf{M}_k) \le B(\mathbf{N}_k)$, which tends to zero as $k \to \infty$ (cf. [10]). In Case $(ii)$,

$$H(t,\mathbf{M}_k) = \frac{\prod_{i=1}^N h\left(t,x_{i,k}\right)^m}{\rho_k^N \det(\mathbf{N}_k)^N} = B(\mathbf{N}_k)\prod_{i=1}^N g(t,x_{i,k})^m$$

where $x_{i,k} = \mathbf{y}_i^T\mathbf{N}_k^{-1}\mathbf{y}_i/\rho_k$. As $k \to \infty$, $x_{i,k}$ tends either to zero or infinity and we conclude. For $t > 0$, $H(t,\cdot)$ is uniformly bounded over $\text{PSD}_m$ since $B(\cdot)$ is. So $H(t,\cdot)$ has a global maximum which must belong to $\text{PSD}_m$ since $H(t,\mathbf{M}) \to 0$ as $\mathbf{M}$ tends to the boundary of $\text{PSD}_m$. So $H(t,\cdot)$ admits critical points. ∎

**Lemma III.3** *Let $t > 0$. Then all critical points of $H(t,\cdot)$ are local strict maxima.*

*Proof of Lemma III.3.* We show that, if $\mathbf{M}$ is a critical point then the Hessian of $H(t,\cdot)$ at $\mathbf{M}$ is a negative definite quadratic form implying that $\mathbf{M}$ is a local strict maximum of $H(t,\cdot)$. Let $\mathbf{M} \in \mathrm{PSD}_m$ be a critical point of $H(t,\cdot)$. Then, one gets that for every $\mathbf{Q} \in \mathrm{Sym}_m$,

$$\langle \mathbf{Q}, \mathrm{Hess}_{\mathbf{M}}(\mathbf{Q})\rangle = -NH(t,\mathbf{M})\Big[\langle \mathbf{Q}, \mathbf{M}^{-1}\mathbf{Q}\mathbf{M}^{-1}\rangle + \frac{1}{N}\sum_{i=1}^{N} u_x(t,\mathbf{y}_i^T\mathbf{M}^{-1}\mathbf{y}_i)(\mathbf{y}_i^T\mathbf{M}^{-1}\mathbf{Q}\mathbf{M}^{-1}\mathbf{y}_i)^2\Big].$$

Let $\mathbf{R} := \mathbf{M}^{-1/2}\mathbf{Q}\mathbf{M}^{-1/2}$ and $\mathbf{d}_i := \mathbf{M}^{-1/2}\mathbf{y}_i$, one has

$$-\frac{\langle \mathbf{Q}, \mathrm{Hess}_{\mathbf{M}}(\mathbf{Q})\rangle}{NH(t,\mathbf{M})} = \|\mathbf{R}\|^2 + \frac{1}{N}\sum_{i=1}^{N} u_x(t,\|\mathbf{d}_i\|^2)(\mathbf{d}_i^T\mathbf{R}\mathbf{d}_i)^2. \quad (7)$$

Recall that $\mathbf{M}$ is a critical point of $H(t,\cdot)$ and thus a solution of $(\mathrm{Eq})_t$, i.e.,

$$\mathbf{I}_m = \frac{1}{N}\sum_{i=1}^{N} u(t,\|\mathbf{d}_i\|^2)\mathbf{d}_i\mathbf{d}_i^T. \quad (8)$$

Multiplying (8) by $\mathbf{R}$ on both left and right, taking the trace and plugging the result into (7) gives

$$(7) = \frac{1}{N}\sum_{i=1}^{N} u(t,\|\mathbf{d}_i\|^2)\|\mathbf{R}\mathbf{d}_i\|^2 + u_x(t,\|\mathbf{d}_i\|^2)(\mathbf{d}_i^T\mathbf{R}\mathbf{d}_i)^2.$$

Let $I_{\mathbf{Q}} = \{i \in \{1,\cdots,N\}, \mathbf{R}\mathbf{d}_i \neq 0\}$. Then

$$(7) = \frac{1}{N}\sum_{i\in I_{\mathbf{Q}}} \|\mathbf{R}\mathbf{d}_i\|^2\left[u(t,\|\mathbf{d}_i\|^2) + \|\mathbf{d}_i\|^2 u_x(t,\|\mathbf{d}_i\|^2) r_i\right]$$

where $r_i := (\mathbf{d}_i^T\mathbf{R}\mathbf{d}_i/[\|\mathbf{d}_i\|\|\mathbf{R}\mathbf{d}_i\|])^2$. Using $0 \leq r_i \leq 1$ (by Cauchy-Schwarz's inequality) and $u_x \leq 0$ (since $u$ is of class $C^1$ and verifies $(U1)$), we have $r_i u_x(\cdot,\cdot) \geq u_x(\cdot,\cdot)$. Then, recalling that $v(t,x) = xu(t,x)$,

$$(7) \geq \frac{1}{N}\sum_{i\in I_{\mathbf{Q}}} \|\mathbf{R}\mathbf{d}_i\|^2 v_x(t,\|\mathbf{d}_i\|^2) \geq 0.$$

Moreover, if $\mathbf{Q} \neq 0$, $I_{\mathbf{Q}} \neq \emptyset$ and there exists $\bar{i}$ such that $v_x(t,\|\mathbf{d}_{\bar{i}}\|^2) > 0$. Therefore $\langle \mathbf{Q}, \mathrm{Hess}_{\mathbf{M}}(\mathbf{Q}) < 0$, i.e., $\mathrm{Hess}_{\mathbf{M}}$ is negative definite, concluding the proof. ■

**Lemma III.4** *Let $t > 0$. Then $(\mathrm{Eq})_t$ admits a unique solution, $\mathbf{M}(t)$, the unique strict maximum of $H(t,\cdot)$.*

*Proof of Lemma III.4.* We reason by contradiction assuming $H(t,\cdot)$ admits at least two local strict maxima. Applying the mountain-pass theorem [16] to the functional $1/H(t,\cdot)$ which tends to infinity in the vicinity of $\partial\mathrm{PSD}_m$, we obtain the existence of a saddle point of $\mathbf{F}$ in $\mathrm{PSD}_m$ which is contradictory to Lemma III.3. ■

We next prove that $\mathbf{M}(t)$ is uniformly bounded in $\mathrm{PSD}_m$ as $t \to 0$, i.e.

**Lemma III.5** *There exists $0 < a \leq b$ and $t_0 > 0$ such that, for every $t \in (0,t_0)$, $aI_m \leq \mathbf{M}(t) \leq b\mathbf{I}_m$.*

*Proof of Lemma III.5.* Let $\mathbf{P}$ be the unique matrix of $\mathrm{PSD}_m$ satisfying $B(\mathbf{P}) = \max_{\mathbf{M}\in\mathrm{PSD}_m} B(\mathbf{M})$ and $\mathrm{Tr}(\mathbf{M}) = m$. Then, for every $t > 0$, $H(t,\mathbf{P}) \leq H(t,\mathbf{M}(t))$ and $B(\mathbf{M}(t)) \leq B(\mathbf{P})$. Multiplying both inequalities, after simplifications, we get $\prod_{i=1}^N g(t,\mathbf{y}_i^T\mathbf{P}^{-1}\mathbf{y}_i) \leq \prod_{i=1}^N g(t,\mathbf{y}_i^T\mathbf{M}(t)^{-1}\mathbf{y}_i) \leq 1$, with $\prod_{i=1}^N g(t,\mathbf{y}_i^T\mathbf{P}^{-1}\mathbf{y}_i) \to 1$ as $t \to 0$. So there exists $t_0 > 0$ such that, for every $t \in (0,t_0)$ and $1 \leq i \leq N$, $1/2 \leq g(t,\mathbf{y}_i^T\mathbf{M}(t)^{-1}\mathbf{y}_i)$, and, since $(U3)$ holds true, there exists $0 < a \leq b$ s.t. for every $t \in (0,t_0)$ and $1 \leq i \leq N$, $a \leq \mathbf{y}_i^T\mathbf{M}(t)^{-1}\mathbf{y}_i \leq b$. This implies that, for every $t \in (0,t_0)$ and $1 \leq i \leq N$, $u(t,b) \leq u(t,\mathbf{y}_i^T\mathbf{M}(t)^{-1}\mathbf{y}_i)) \leq u(t,a)$, hence $u(t,b)\mathbf{C} \leq \mathbf{M}(t) \leq u(t,a)\mathbf{C}$ with $\mathbf{C} := \frac{m}{N}\sum_{i=1}^N \mathbf{y}_i\mathbf{y}_i^T$. One concludes easily. ■

**Lemma III.6** *Under the conditions of Theorem II.2, $\lim_{t\to 0}\mathbf{M}(t) = \mathbf{M}_0$ solution of $(\mathrm{Eq})_0$ given by $\mathbf{M}_0 = \xi_u\mathbf{P}$, where $\xi_u > 0$ is the unique solution of* (3)*.*

*Proof of Lemma III.6.* Since $\mathbf{M}(\cdot)$ is uniformly bounded in $\mathrm{PSD}_m$ as $t \to 0$, its accumulation points still belong to $\mathrm{PSD}_m$ and are necessarily of the form $\mu\mathbf{P}$ where $\mu > 0$ and $\mathbf{P}$ is the solution of $(\mathrm{Eq})_0$ with trace $m$. Taking the trace in (8), one gets $m = \frac{1}{N}\sum_{i=1}^N v(t,\|\mathbf{d}_i(t)\|^2)$, where $\mathbf{d}_i(t) = \mathbf{M}(t)^{-1/2}\mathbf{y}_i$ for $1 \leq i \leq N$. Using (1) and $(U3)$, one deduces that, for every $t > 0$, $\sum_{i=1}^N v_1(\|\mathbf{d}_i(t)\|^2) + t\sum_{i=1}^N w_1(\|\mathbf{d}_i(t)\|^2) + o(t) = 0$. Consider an accumulation point $\mu\mathbf{P}$ of $\mathbf{M}(\cdot)$ as $t \to 0$. Then, up to a subsequence, $\lim_{t\to 0}\mathbf{M}(t) = \mu\mathbf{P}$ and, for $1 \leq i \leq N$, $\lim_{t\to 0}\mathbf{d}_i(t) = \mathbf{P}^{-1/2}\mathbf{y}_i/\sqrt{\mu}$. According to $(U3)$, the second sum in the previous equation tends to zero as $t \to 0$ and we are left with $\sum_{i=1}^N v_1(\mathbf{y}_i^T\mathbf{P}^{-1}\mathbf{y}_i/\mu) = 0$. Since the left-hand side of the latter defines a decreasing function of $\mu$, it has a unique solution denoted $\xi_u > 0$, which concludes the proof since $\mathbf{M}(\cdot)$ admits a unique accumulation point as $t \to 0$. ■

## IV. Conclusions

In this paper, an alternative proof for existence and uniqueness for the Maronna's $M$-estimators is provided. More importantly, using this particular approach leads to draw some connections between Maronna's and Tyler's estimators by expressing (properly scaled) Tyler's estimator in terms of a limit of a class of Maronna's estimators. This result may also find interest in studies of Tyler's $M$-estimator in the large random matrix regime.

## REFERENCES


[1] R. A. Maronna, "Robust $M$-estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.

[2] D.E. Tyler, "A distribution-free $m$-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.

[3] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[4] J. T. Kent and D. E. Tyler, "Redescending $M$-estimates of multivariate location and scatter," *Annals of Statistics*, vol. 19, no. 4, pp. 2102–2119, December 1991.

[5] Teng Zhang, Ami Wiesel, and Maria Sabrina Greco, "Multivariate generalized gaussian distribution: Convexity and graphical models," *Signal Processing, IEEE Transactions on*, vol. 61, no. 16, pp. 4141–4148, 2013.

[6] R. Couillet, F. Pascal, and J.W. Silverstein, "Robust M-Estimation for Array Processing: A Random Matrix Approach," *Information Theory, IEEE Transactions on (to appear)*, 2013. [Online]. Available: arXiv:1204.5320.

[7] R Couillet, F Pascal, and J W Silverstein, "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples," *Journal of Multivariate Analysis (to appear)*, 2013. [Online]. Available: arXiv:1311.7034.

[8] Teng Zhang, Xiuyuan Cheng, and Amit Singer, "Marchenko-Pastur Law for Tyler's and Maronna's M-estimators," *arXiv preprint arXiv:1401.3424*, 2014.

[9] Ilya Soloveychik and Ami Wiesel, "Non-asymptotic Error Analysis of Tyler's Scatter Estimator," *arXiv preprint arXiv:1401.6926*, 2014.

[10] F. Pascal, Y. Chitour, J.P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: existence and algorithm analysis," *Signal Processing, IEEE Transactions on*, vol. 56, no. 1, pp. 34–48, Jan. 2008.

[11] Y. Chitour and F. Pascal, "Exact maximum likelihood estimates for SIRV covariance matrix: existence and algorithm analysis," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4563–4573, Oct. 2008.

[12] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to STAP detection problem," *Signal Processing, IEEE Transactions on (to appear)*, 2013. [Online]. Available: arXiv:1311.6567.

[13] E. Ollila and D. E. Tyler, "Distribution-free detection under complex elliptically symmetric clutter distribution," in *IEEE Sensor Array and Multichannel Signal Processing Workshop - SAM 2012*, Hoboken, NJ, USA, June 2012.

[14] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *Signal Processing, IEEE Transactions on*, vol. 60, no. 11, pp. 5597 –5625, nov. 2012.

[15] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, *Robust statistics: the approach based on influence functions*, John Wiley & Sons New York, 1986.

[16] Michael Struwe, *Variational methods: applications to nonlinear partial differential equations and Hamiltonian systems*, vol. 34, Springer, 4 edition, 2008.