Ann. Inst. Statist. Math.
Vol. 50, No. 3, 471–491 (1998)

ON TYLER'S M-FUNCTIONAL OF SCATTER IN HIGH DIMENSION*

LUTZ DÜMBGEN**

Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, D-69120 Heidelberg, Germany

(Received June 4, 1996; revised October 16, 1997)

Abstract. Let $\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n \in \boldsymbol{R}^q$ be independent, identically distributed random vectors with nonsingular covariance matrix Σ, and let $S = S(\boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$ be an estimator for Σ. A quantity of particular interest is the condition number of $\Sigma^{-1}S$. If the $\boldsymbol{y}_i$ are Gaussian and S is the sample covariance matrix, the condition number of $\Sigma^{-1}S$, i.e. the ratio of its extreme eigenvalues, equals $1 + O_p((q/n)^{1/2})$ as $q \to \infty$ and $q/n \to 0$. The present paper shows that the same result can be achieved with two estimators based on Tyler's (1987, *Ann. Statist.*, **15**, 234–251) M-functional of scatter, assuming only elliptical symmetry of $\mathcal{L}(\boldsymbol{y}_i)$ or less. The main tool is a linear expansion for this M-functional which holds uniformly in the dimension q. As a by-product we obtain continuous Fréchet-differentiability with respect to weak convergence.

Key words and phrases: Differentiability, dimensional asymptotics, elliptical symmetry, M-functional, scatter matrix, symmetrization.

1. Introduction

It has been noted by numerous authors that asymptotic results, where the dimension of the underlying model is fixed while the number of observations tends to infinity, are often inappropriate for real applications; e.g. Portnoy (1988) or Girko (1995). In particular, the literature on M-estimation in linear regression models with increasing dimension is vast and still growing; see for instance Huber (1981), Portnoy (1984, 1985), Bai and Wu (1994*a*, 1994*b*), Mammen (1996) and the references cited therein. In the present paper we investigate the related problem of M-estimation of a high-dimensional covariance matrix.

Let $\widehat{P}_n$ be the empirical distribution of independent random vectors $\boldsymbol{y}_{n1}, \boldsymbol{y}_{n2}, \dots, \boldsymbol{y}_{nn}$ in $\boldsymbol{R}^q$ with unknown distribution P_n, and let $S_n = S_n(\widehat{P}_n)$ be an estimator for the covariance matrix Σ_n of P_n, both assumed to be positive definite. Of

* Research supported in part by European Union Human Capital and Mobility Program ERB CHRX-CT 940693.

** Now at Mathematical Institute, Medical University at Luebeck, Wallstrasse 40, D-23560 Luebeck, Germany.

particular interest is the condition number of $\Sigma_n^{-1} S_n$,

$$\gamma(\Sigma_n^{-1} S_n) := \frac{\lambda_1(\Sigma_n^{-1} S_n)}{\lambda_q(\Sigma_n^{-1} S_n)},$$

where $\lambda_1(A) \geq \lambda_2(A) \geq \lambda_3(A) \geq \cdots$ denote the ordered real eigenvalues of $A \in \boldsymbol{R}^{q \times q}$. There are explicit bounds for various scale-invariant functions of S_n and Σ_n such as correlations, partial and canonical correlations, regression coefficients or eigenspaces, all in terms of $\gamma(\Sigma_n^{-1} S_n)$ (cf. Dümbgen (1998)). An example is the following sharp inequality for correlations, where $(\cdot)'$ denotes transposition:

$$\left| \operatorname{arctanh}\left(\frac{x'\Sigma_n y}{\sqrt{x'\Sigma_n x \, y'\Sigma_n y}} \right) - \operatorname{arctanh}\left(\frac{x' S_n y}{\sqrt{x' S_n x \, y' S_n y}} \right) \right| \leq \frac{\log \gamma(\Sigma_n^{-1} S_n)}{2}$$

for arbitrary $x, y \in \boldsymbol{R}^q \setminus \{0\}$. Therefore it is of interest to study the probabilistic behavior of $\gamma(\Sigma_n^{-1} S_n)$. If P_n is multivariate normal and S_n is the sample covariance matrix, a modification of Silverstein's (1985) arguments reveals that

$$\gamma(\Sigma_n^{-1} S_n) = 1 + 4(q/n)^{1/2} + o_{\mathrm{p}}((q/n)^{1/2}); \tag{1.1}$$

see also the proof of Theorem 5.4. In connection with P_n, $\widehat{P}_n$ we assume tacitly that the dimension $q = q_n$ may depend on the sample size n such that $q/n \to 0$. Asymptotic statements refer to $n \to \infty$, unless stated otherwise. Expansions such as (1.1) hold under more general assumptions on the distribution P_n, provided that it has sufficiently light tails (cf. Girko (1995)). On the other hand, the distribution of the extremal eigenvalues of the sample covariance matrix is very sensitive to deviations from normality. The weaker assertion

$$\gamma(\Sigma_n^{-1} S_n) = 1 + O_{\mathrm{p}}((q/n)^{1/2}) \tag{1.2}$$

may be false, even under elliptical symmetry of P_n. It is thus desirable to have an estimator, whose distribution is less model-dependent, such that expansion (1.1) or at least (1.2) holds.

A possible alternative to the sample covariance matrix are M-estimators of scatter as proposed by Maronna (1976) and Tyler (1987). The present paper focuses on two estimators related to Tyler's (1987) M-functional. The latter is defined in Section 2 as a matrix-valued function $Q \mapsto \Sigma(Q)$ on the space of probability measures on $\boldsymbol{R}^q \setminus \{0\}$. Section 3 provides a basic linear expansion for $\Sigma(\cdot)$ with a rather explicit bound for the remainder term. As a by-product we obtain continuous Fréchet-differentiability of $\Sigma(\cdot)$ with respect to the weak topology.

Section 4 describes estimators based on $\Sigma(\cdot)$. One obvious choice is the M-estimator $\Sigma(\widehat{P}_n)$, which is distribution-free if P_n is spherically symmetric around zero. In addition we propose the estimator $\Sigma(\widehat{P}_n^{\mathrm{s}})$, where $\widehat{P}_n^{\mathrm{s}}$ is a symmetrization of $\widehat{P}_n$. This is an intuitively appealing method to get rid of unknown location parameters. The linear expansion of Section 3 implies asymptotic normality of both estimators and consistency of certain bootstrap methods. Some of these results and conclusions are not entirely new but nevertheless stated explicitly for the reader's convenience. In connection with the bootstrap one can use similar arguments as Bickel and Freedman (1981).

In order to prove Fréchet-differentiability for fixed dimension q, one could also apply general methods of Clarke (1983). An advantage of our explicit expansion is that it enables us to investigate the asymptotic behavior of $\Sigma(\widehat{P}_n)$ and $\Sigma(\widehat{P}_n^s)$ as $q = q_n \to \infty$. This is done in Section 5. Under certain regularity assumptions, assertion (1.2) is valid for both estimators $\Sigma(\widehat{P}_n)$ and $\Sigma(\widehat{P}_n^s)$. In particular, if P_n is elliptically symmetric, $\Sigma(\widehat{P}_n)$ is shown to have the same asymptotic behavior as the sample covariance matrix in the Gaussian model, including expansion (1.1).

Proofs are deferred to Section 7. As for Sections 2 and 4 the reader is referred to the technical report Dümbgen (1997), which contains the present material as well as more detailed proofs. This report also reviews Tyler's (1987) approach to the problem of unknown location, re-centering $\widehat{P}_n$ around an estimator $\widehat{\mu}_n = \widehat{\mu}_n(\widehat{P}_n)$ for P_n's "center", in view of dimensional asymptotics.

2. Definition and basic properties of the M-functional $\Sigma(\cdot)$

Let us first introduce some notation. Throughout the set of symmetric matrices in $\boldsymbol{R}^{q\times q}$ is denoted by $\boldsymbol{M}$, while $\boldsymbol{M}^+$ denotes the set of positive definite $M \in \boldsymbol{M}$. For $M \in \boldsymbol{M}^+$ the unique matrix $N \in \boldsymbol{M}^+$ with $NN = M$ is denoted by $M^{1/2}$, and $M^{-1/2} := (M^{-1})^{1/2} = (M^{1/2})^{-1}$. Further we consider the affine subspaces

$$\boldsymbol{M}(t) := \{M \in \boldsymbol{M} : \mathrm{trace}(M) = t\} = (t/q)I + \boldsymbol{M}(0)$$

of $\boldsymbol{M}$, where I stands for the identity matrix in $\boldsymbol{R}^{q\times q}$. Let f be a real or vector-valued function on $\boldsymbol{R}^q$, and let Δ be a signed measure on $\boldsymbol{R}^q$. Then $f(\Delta)$ stands for $\int f(x)\,\Delta(dx)$. This convention will be particularly convenient for functions of several arguments. Further, for $A \in \boldsymbol{R}^{q\times q}$ we denote by $A\Delta$ the transformed signed measure $\Delta \circ A^{-1}$.

Throughout this section let $\boldsymbol{z}$ be a random vector with distribution Q on $\boldsymbol{R}^q \setminus \{0\}$. We regard Q as rotationally symmetric around zero in a weak sense if $C(Q) = \int G(x)\,Q(dx)$ is equal to I, where

$$G(x) := q|x|^{-2}xx' \in \boldsymbol{M}(q) \qquad \text{for} \quad x \in \boldsymbol{R}^q \setminus \{0\};$$

here $|x|$ denotes the standard Euclidean norm $(x'x)^{1/2}$ of x. Note that $G(Q)$ equals q times the matrix of second moments of $|\boldsymbol{z}|^{-1}\boldsymbol{z}$. If Q is spherically symmetric around zero, one easily verifies that in fact $G(Q) = I$. More generally, this equality holds if the vectors $\boldsymbol{z} = (z_i)_{1\le i\le q}$ and $(\epsilon_i z_{\pi(i)})_{1\le i\le q}$ have the same distribution for arbitrary $\epsilon \in \{-1,1\}^q$ and permutations π of $\{1,2,\ldots,q\}$. In general one tries to find $M \in \boldsymbol{M}^+$ such that

$$G(M^{-1/2}Q) = q\int \frac{M^{-1/2}xx'M^{-1/2}}{x'M^{-1}x}\,Q(dx) = I.$$

Note that $G(M^{-1/2}Q) = G((sM)^{-1/2}Q)$ for all $s > 0$, so that $G(\cdot)$ is only useful in connection with scale-invariant functions on $\boldsymbol{M}^+$ such as correlations.

DEFINITION. If the equality $G(M^{-1/2}Q) = I$ has a unique solution M in $\boldsymbol{M}^+(q) := \boldsymbol{M}^+ \cap \boldsymbol{M}(q)$, this matrix M is denoted by $\Sigma(Q)$. Otherwise we define arbitrarily $\Sigma(Q) := 0$.

An important property of $\Sigma(\cdot)$ is linear equivariance. For nonsingular $A \in \boldsymbol{R}^{q\times q}$ one can show that

$$\Sigma(AQ) = rA\Sigma(Q)A' \quad \text{with} \quad r := q/\operatorname{trace}(A\Sigma(Q)A'). \tag{2.1}$$

(Recall that AQ stands for the distribution of $A\boldsymbol{z}$.) Necessary and sufficient conditions for $\Sigma(Q) \in \boldsymbol{M}^+$ are as follows.

THEOREM 2.1. *Let $\mathcal{V}$ be the set of proper linear subspaces $\boldsymbol{V}$ of $\boldsymbol{R}^q$, i.e. $1 \le \dim(\boldsymbol{V}) < q$.*

[a] *If $G(M^{-1/2}Q) = I$ for some $M \in \boldsymbol{M}^+$, then*

$$Q(\boldsymbol{V}) \le \dim(\boldsymbol{V})/q \quad \text{for all} \quad \boldsymbol{V} \in \mathcal{V}.$$

[b] *If*

$$Q(\boldsymbol{V}) < \dim(\boldsymbol{V})/q \quad \text{for all} \quad \boldsymbol{V} \in \mathcal{V}, \tag{2.2}$$

then there exists a unique $M \in \boldsymbol{M}^+(q)$ such that $G(M^{-1/2}Q) = I$.

[c] *Suppose that $G(Q) = I$ but $Q(\boldsymbol{V}) = \dim(\boldsymbol{V})/q$ for some $\boldsymbol{V} \in \mathcal{V}$. Then $Q(\boldsymbol{V} \cup \boldsymbol{V}^\perp) = 1$ and*

$$G((a\Pi + b(I - \Pi))^{-1/2}Q) = I \quad \text{for all} \quad a, b > 0,$$

where $\Pi \in \boldsymbol{M}$ describes the orthogonal projection from $\boldsymbol{R}^q$ onto $\boldsymbol{V}$, and $\boldsymbol{V}^\perp$ stands for the orthogonal complement of $\boldsymbol{V}$.

Parts [a] and [b] are due to Tyler (1987) and Kent and Tyler (1988). Their proofs are formulated for empirical distributions Q, but extension to arbitary distributions is mainly straightforward, requiring only notational changes. The only exception is the existence statement in part [b]. For the necessary modifications as well as the proof of part [c] we refer to Dümbgen (1997). Part [c], combined with an equivariance argument, supplements part [b] in that condition (2.2) is even necessary for $\Sigma(Q) \in \boldsymbol{M}^+$. It will be needed in the proof of Theorem 3.2 below.

3. Differentiability of $\Sigma(\cdot)$

For $M \in \boldsymbol{M}$ we define its norm $\|M\| := \max\{|\lambda_1(M)|, |\lambda_q(M)|\}$, commonly referred to as its spectral radius or operator norm. Since the dimension q may vary, this particular choice of a norm is important. It is particularly useful in connection with eigenvalues, because $|\lambda_i(A) - \lambda_i(B)| \le \|A - B\|$ for A, $B \in \boldsymbol{M}$ and $1 \le i \le q$. Generally, we always use the norm

$$\|L\| := \max_{y \in \mathcal{S}(\boldsymbol{B})} \|Ly\|$$

of a linear operator L from a normed vector space $(\boldsymbol{B}, \|\cdot\|)$ into another normed space, where $\boldsymbol{S}(\boldsymbol{B})$ denotes the unit sphere $\{y \in \boldsymbol{B} : \|y\| = 1\}$.

Throughout this section let P, Q be probability distributions on $\boldsymbol{R}^q \setminus \{0\}$. Now we investigate $\Sigma(Q)$ if Q is close to P in a certain sense and $G(P) = I$. By equivariance of $G(\cdot)$ and $\Sigma(\cdot)$ it suffices to consider the latter case.

For any $x \in \boldsymbol{R}^q \setminus \{0\}$ the function $G(M^{-1/2}x)$ is differentiable with respect to $M \in \boldsymbol{M}^+$ with

$$D(x, B) := \frac{\partial}{\partial t}\Big|_{t=0} G((I + tB)^{-1/2}x) = F(x, B) - 2^{-1}(BG(x) + G(x)B),$$
$$F(x, B) := |x|^{-2}x'Bx\, G(x) = q|x|^{-4}x'Bx\, xx'.$$

Note that $D(x, I) = 0$ and $\operatorname{trace}(D(x, B)) = 0$ for all $B \in \boldsymbol{M}$. The next lemma shows that condition (2.2) is closely related to the operator $D(Q, \cdot)$.

LEMMA 3.1. *The operator $D(Q, \cdot)$ is nonsingular on $\boldsymbol{M}(0)$ if, and only if, $Q(\boldsymbol{V} \cup \boldsymbol{V}^{\perp}) < 1$ for arbitrary $\boldsymbol{V} \in \mathcal{V}$. In that case,*

$$\operatorname{trace}(D(Q, B)B) < 0 \quad \textit{for all} \quad B \in \boldsymbol{M}(0) \setminus \{0\}.$$

The inverse operator of $D(Q, \cdot) : \boldsymbol{M}(0) \to \boldsymbol{M}(0)$, if existing, is denoted by $D^{-1}(Q, \cdot)$. Here is our basic linear expansion for $\Sigma(\cdot)$.

THEOREM 3.2. *For any $b < \infty$ there exist constants $\kappa(b) < \infty$ and $\epsilon(b) > 0$ (not depending on q, P or Q) such that*

$$\|\Sigma(Q) - I + D^{-1}(P, G(Q - P))\| \leq \kappa(b)\, \|F(Q - P, \cdot)\| \|G(Q - P)\|$$

whenever

$$\Sigma(P) = I, \quad \|D^{-1}(P, \cdot)\| \leq b \quad \textit{and} \quad \|F(Q - P, \cdot)\| \leq \epsilon(b).$$

The latter two norms $\|\cdot\|$ refer to the linear operators $D^{-1}(P, \cdot)$ on $\boldsymbol{M}(0)$ and $F(Q - P, \cdot)$ on $\boldsymbol{M}$, respectively. Note also that $\|G(Q - P)\| = \|F(Q - P, I)\| \leq \|F(Q - P, \cdot)\|$.

Theorem 3.2, Lemma 3.1 and (2.1) together imply that $\Sigma(\cdot)$ is Fréchet-differentiable with respect to the weak topology. The reason is that $x \mapsto F(x, \cdot)$ is a bounded, continuous mapping from $\boldsymbol{R}^q \setminus \{0\}$ into the finite-dimensional space of linear operators $L : \boldsymbol{M} \to \boldsymbol{M}$, so that $\|F(Q - P, \cdot)\| \to 0$ as $Q \to P$ weakly.

COROLLARY 3.3. *Suppose that $\Sigma(P) = I$. Then, as $Q \to P$ weakly,*

$$G(Q) \to I \quad \textit{and} \quad \Sigma(Q) - I = -D^{-1}(P, G(Q - P)) + o(\|G(Q - P)\|).$$

One can even show that $\Sigma(\cdot)$ is continuously Fréchet-differentiable. Instead of pursuing this issue, we shall prove a related statement about limiting distributions of $\Sigma(\widehat{P}_n)$ and $\Sigma(\widehat{P}_n^s)$ in the next section.

4. Related estimators and their properties in fixed dimension

At this point it is convenient to drop the assumption $Q\{0\} = 0$. We only assume that Q is nondegenerate and define $\Sigma(Q) := \Sigma(Q(\cdot \mid \boldsymbol{R}^q \setminus \{0\}))$.

Suppose first that the distribution P_n has a known "center" $\mu_n \in \boldsymbol{R}^q$. Without loss of generality one may assume that $\mu_n = 0$. Then a straightforward estimator for $\Sigma(P_n)$ is given by $\Sigma(\widehat{P}_n)$. An important example are elliptically symmetric distributions $P_n = \mathcal{L}(R_n \Sigma_n^{1/2} \boldsymbol{u})$, where $R_n > 0$ and $\boldsymbol{u}$ are stochastically independent, $\boldsymbol{u}$ is uniformly distributed on the unit sphere of $\boldsymbol{R}^q$, and $\Sigma_n \in \boldsymbol{M}^+(q)$. Clearly $\Sigma(P_n) = \Sigma_n$, and the empirical distribution $\widehat{P}_n$ satisfies condition (2.2) almost surely if $n > q$. Moreover, the distribution of $\gamma(\Sigma_n^{-1}\Sigma(\widehat{P}_n))$ depends neither on Σ_n nor on $\mathcal{L}(R_n)$ (cf. Tyler (1987)).

The center μ_n, no matter how it is defined, is rarely known in advance. In order to avoid definition and estimation of an unknown location parameter we propose the functional $Q \mapsto \Sigma(Q^{\mathrm{s}})$ with the symmetrized distribution

$$Q^{\mathrm{s}} := \mathcal{L}(\boldsymbol{z}_1 - \boldsymbol{z}_2 \mid \boldsymbol{z}_1 \neq \boldsymbol{z}_2) \quad \text{where} \quad (\boldsymbol{z}_1, \boldsymbol{z}_2) \sim Q \otimes Q.$$

Generally, $\Delta_1 \otimes \Delta_2$ denotes the product measure on $\boldsymbol{R}^q \times \boldsymbol{R}^q$ of (signed) measures Δ_1, Δ_2 on $\boldsymbol{R}^q$. One motivation for the functional $Q \mapsto \Sigma(Q^{\mathrm{s}})$ is the representation $2^{-1}\,\mathbb{E}((\boldsymbol{z}_1 - \boldsymbol{z}_2)(\boldsymbol{z}_1 - \boldsymbol{z}_2)')$ of the covariance matrix of Q. Moreover, if $\boldsymbol{z} \sim Q$ has independent, identically distributed components, then $G(Q^{\mathrm{s}}) = I$, whereas $G(Q)$ may be different from I. Thus symmetrization partly corrects a possible deficiency of M-estimators.

One easily verifies that $Q \mapsto \Sigma(Q^{\mathrm{s}})$ is affinely invariant in that

$$A\Sigma(Q^{\mathrm{s}})A' = r\,\Sigma((\mu + AQ)^{\mathrm{s}}) \quad \text{with} \quad r := \mathrm{trace}(A\Sigma(Q^{\mathrm{s}})A')/q \tag{4.1}$$

for any nonsingular $A \in \boldsymbol{R}^{q\times q}$ and $\mu \in \boldsymbol{R}^q$. Here $\mu + AQ$ stands for $\mathcal{L}(\mu + A\boldsymbol{z})$. If Q is elliptically symmetric around μ with scatter matrix $\Sigma_o \in \boldsymbol{M}^+(q)$, then Q^{s} is elliptically symmetric around zero with the same scatter matrix Σ_o.

An application of Theorem 3.2 utilizing the explicit error bound is the following Central Limit Theorem for the distribution of $\Sigma(\widehat{P}_n)$ and $\Sigma(\widehat{P}_n^{\mathrm{s}})$.

Corollary 4.1. *Suppose that P_n converges weakly to some distribution P on $\boldsymbol{R}^q$.*

[a] *Let $P\{0\} = 0$ and $\Sigma(P) = I$. Let $L_n(\cdot \mid P_n)$ denote the distribution of*

$$n^{1/2}(\Sigma(\Sigma(P_n)^{-1/2}\widehat{P}_n) - I)$$

(provided that $\Sigma(P_n) \in \boldsymbol{M}^+$). Then $\Sigma(P_n) \to I$ and

$$L_n(\cdot \mid P_n) \to_{\mathrm{w}} \mathcal{L}(W),$$

where W is a random matrix with centered Gaussian distribution on $\boldsymbol{M}(0)$ and the same covariance function as $D^{-1}(P, G(\boldsymbol{y}) - I)$, $\boldsymbol{y} \sim P$.

[b] *Let $P\{\mu\} = 0$ for all $\mu \in \boldsymbol{R}^q$ and $\Sigma(P^{\mathrm{s}}) = I$. Let $L_n^{\mathrm{s}}(\cdot \mid P_n)$ denote the distribution of*

$$n^{1/2}(\Sigma(\Sigma(P_n^{\mathrm{s}})^{-1/2}\widehat{P}_n^{\mathrm{s}}) - I)$$

(provided that $\Sigma(P_n^{\mathrm{s}}) \in \boldsymbol{M}^+$). Then $\Sigma(P_n^{\mathrm{s}}) \to I$ and

$$L_n^{\mathrm{s}}(\cdot \mid P_n) \to_{\mathrm{w}} \mathcal{L}(W^{\mathrm{s}}),$$

where W^{s} is a random matrix with centered Gaussian distribution on $\boldsymbol{M}(0)$ and the same covariance function as $2D^{-1}(P^{\mathrm{s}}, \widetilde{G}(y,P) - I)$, $y \sim P$. Here $\widetilde{G}(x,y) := G(x-y)$.

Remark 4.2. The covariance function of a random matrix $\widetilde{W} \in \boldsymbol{M}(0)$ is defined as the function $(A,B) \mapsto \mathrm{Cov}(\mathrm{trace}(\widetilde{W}A), \mathrm{trace}(\widetilde{W}B))$ on $\boldsymbol{M}(0) \times \boldsymbol{M}(0)$.

Remark 4.3. In case of P being spherically symmetric around zero one can deduce from equations (7.9) and (7.10) in Lemma 7.4 that

$$\mathbb{E}(\mathrm{trace}(WA)\,\mathrm{trace}(WB)) = 2(1+2/q)\,\mathrm{trace}(AB) \qquad \text{for} \quad A, B \in \boldsymbol{M}(0).$$

Remark 4.4. If $P_n \to P$ weakly, then the emprirical distribution $\widehat{P}_n$ converges weakly to P in probability. Precisely, $d_{\mathrm{w}}(\widehat{P}_n, P)$ converges to zero in probability, where $d_{\mathrm{w}}(\cdot,\cdot)$ metrizes weak convergence of probability measures on $\boldsymbol{R}^q$. Consequently, the bootstrap distributions $L_n(\cdot \mid \widehat{P}_n)$ and $L_n^{\mathrm{s}}(\cdot \mid \widehat{P}_n)$ are consistent estimators of $L_n(\cdot \mid P_n)$ and $L_n^{\mathrm{s}}(\cdot \mid P_n)$, respectively.

Remark 4.5. Utilizing the equivariance properties of $\Sigma(\cdot)$, (2.1) and (4.1), one can deduce from Corollary 4.1 that

$$n^{1/2}(\gamma(\Sigma(P_n)^{-1}\Sigma(\widehat{P}_n)) - 1) \to_{\mathcal{L}} (\lambda_1 - \lambda_q)(W_o) \qquad \text{in part [a]},$$
$$n^{1/2}(\gamma(\Sigma(P_n^{\mathrm{s}})^{-1}\Sigma(\widehat{P}_n^{\mathrm{s}})) - 1) \to_{\mathcal{L}} (\lambda_1 - \lambda_q)(W_o^{\mathrm{s}}) \qquad \text{in part [b]}.$$

5. Asymptotic behavior of $\Sigma(\widehat{P}_n)$ and $\Sigma(\widehat{P}_n^{\mathrm{s}})$ in high dimension

Now we consider the case where

$$q = q_n \to \infty \qquad \text{but} \qquad q/n \to 0.$$

For the sake of simplicity it is assumed that P_n has no atoms.

THEOREM 5.1. *Suppose that $\Sigma(P_n) = I$ for all n. Let*

$$\kappa_n^2 := \max_{u \in \boldsymbol{S}(\boldsymbol{R}^q)} \int (u'G(y)u)^2 \, P_n(dy) = O(1),$$
$$\sigma_n^2 := \max_{B \in \boldsymbol{S}(\boldsymbol{M}(0))} \int \left(\frac{y'By}{y'y}\right)^2 P_n(dy) = o(1).$$

Further let $q = O(n^{1/2})$. Then

$$\mathbb{E}\,\|G(\widehat{P}_n) - I\| = o(1) \qquad \textit{and} \qquad \mathbb{E}\,\|\Sigma(\widehat{P}_n) - G(\widehat{P}_n)\| = o(\mathbb{E}\,\|G(\widehat{P}_n) - I\|).$$

If in addition $q = O(n^{1/3})$, then

$$\mathbb{E}\,\|G(\widehat{P}_n) - I\| = O((q/n)^{1/2}).$$

Remark 5.2. Suppose that $\boldsymbol{y}_{n1} = (y_{n,i})_{1\le i\le q} \sim P_n$ has independent, identically distributed components with continuous, symmetric distribution such that $\mathbb{E}(y_{n,1}^2) = 1$ and $\mathbb{E}(y_{n,1}^4) = O(1)$. Then $\kappa_n^2 = O(1)$ and $\sigma_n^2 = O(q^{-1})$. For it follows from the one-sided version of Bennett's (1962) inequality that $\mathbb{P}\{|\boldsymbol{y}_{n1}|^2/q \le 1/2\} \le \exp(-a_n q)$ for some number a_n depending on the fourth moment of $y_{n,1}$ and q such that $\liminf_{n\to\infty} a_n > 0$. Therefore, since $(u'G(y)u)^2 \le q^2$ and $(y'By)^2/(y'y)^2 \le 1$, one may replace these integrands of κ_n^2 and σ_n^2 with $4(u'y)^4$ and $4q^{-2}(y'By)^2$, respectively. Then the assertion follows from tedious but elementary moment calculations.

Remark 5.3. The conclusions of Theorem 5.1 and Remark 5.2 remain valid if $(P_n, \widehat{P}_n)$ is replaced with $(P_n^{\mathrm{s}}, \widehat{P}_n^{\mathrm{s}})$, where the symmetry condition in Remark 5.2 becomes superfluous. For the proof of Theorem 5.1 consists essentially of bounding $\mathbb{E}(\|F(\widehat{P}_n - P_n, \cdot)\|^2)$ and $\mathbb{E}(\|G(\widehat{P}_n - P_n)\|^2)$. But $F(\widehat{P}_n^{\mathrm{s}}, B)$ can be written as a matrix-valued U-statistic

$$\binom{n}{2}^{-1} \sum_{1\le i<j\le n} F(\boldsymbol{y}_{ni} - \boldsymbol{y}_{nj}, B).$$

Let $\check{P}_n^{\mathrm{s}}$ be the empirical distribution of $\boldsymbol{y}_{n1}^{\mathrm{s}}, \boldsymbol{y}_{n2}^{\mathrm{s}}, \ldots, \boldsymbol{y}_{nm}^{\mathrm{s}}$, where $m = m_n := \lfloor n/2 \rfloor$ and $\boldsymbol{y}_{ni}^{\mathrm{s}} := \boldsymbol{y}_{n,2i-1} - \boldsymbol{y}_{n,2i}$. Then a simple convexity argument due to Hoeffding (1963) yields

$$\begin{aligned} &\mathbb{E}(\|G(\widehat{P}_n^{\mathrm{s}} - P_n^{\mathrm{s}})\|^2) \le \mathbb{E}(\|G(\check{P}_n^{\mathrm{s}} - P_n^{\mathrm{s}})\|^2), \\ &\mathbb{E}(\|F(\widehat{P}_n^{\mathrm{s}} - P_n^{\mathrm{s}}, \cdot)\|^2) \le \mathbb{E}(\|F(\check{P}_n^{\mathrm{s}} - P_n^{\mathrm{s}}, \cdot)\|^2); \end{aligned} \tag{5.1}$$

see also equation (7.18) in Section 7. Now the signed measure $\check{P}_n^{\mathrm{s}} - P_n^{\mathrm{s}}$ can be handled analogously as $\widehat{P}_n - P_n$.

Under spherical symmetry of P_n, restrictions on q beyond $q = o(n)$ are superfluous, and one can obtain rather precise expansions.

THEOREM 5.4. *Suppose that P_n is spherically symmetric around zero for all n.*

[a] *Then*

$$\begin{aligned} &\mathbb{E}\,\|G(\widehat{P}_n) - I\| = O((q/n)^{1/2}), \\ &\mathbb{E}\,\|\Sigma(\widehat{P}_n) - I - (1 + 2/q)(G(\widehat{P}_n) - I)\| = O(\log(n/q)q/n). \end{aligned}$$

Moreover, one can couple $\Sigma(\widehat{P}_n)$ with a standard Wishart matrix $M_n \in \boldsymbol{M}$ with n degrees of freedom such that

$$\mathbb{E}\,\|\Sigma(\widehat{P}_n) - n^{-1}M_n\| = o((q/n)^{1/2}).$$

In particular, $\gamma(\Sigma(\widehat{P}_n)) = 1 + 4(q/n)^{1/2} + o_{\mathrm{p}}((q/n)^{1/2})$.

[b] *As for* $\widehat{P}_n^{\mathrm{s}}$,

$$\mathbb{E}\,\|\Sigma(\widehat{P}_n^{\mathrm{s}}) - I\| = O((q/n)^{1/2}),$$
$$\mathbb{E}\,\|\Sigma(\widehat{P}_n^{\mathrm{s}}) - I - n^{-1}\textstyle\sum_{i=1}^n H_n(|\boldsymbol{y}_{ni}|)(G(\boldsymbol{y}_{ni}) - I)\| = o((q/n)^{1/2}),$$

where H_n *is an increasing function from* $[0,\infty[$ *into* $[0,2[$. *If in addition* $|\boldsymbol{y}_{n1}|^2/q$ *converges in probability to a constant* $\kappa_o > 0$, *then*

$$\mathbb{E}\,\|\Sigma(\widehat{P}_n^{\mathrm{s}}) - \Sigma(\widehat{P}_n)\| = o((q/n)^{1/2}).$$

6. Some final remarks

In principle different M-estimators such as in Maronna (1976) could be treated similarly. But this would require stronger regularity assumptions (not to mention more complicated notation) without giving substantially better results.

An interesting special case is the maximum likelihood estimator for the multivariate Cauchy distribution on $\boldsymbol{R}^{q-1}$. Suppose that $\boldsymbol{y}_{ni} = (\widetilde{\boldsymbol{y}}_{ni}', 1)'$ with random vectors $\widetilde{\boldsymbol{y}}_{ni} \in \boldsymbol{R}^{q-1}$ having distribution $\widetilde{P}_n$. If we write

$$\Sigma(\widehat{P}_n) = (\Sigma(\widehat{P}_n))_{qq}\begin{pmatrix} \check{\Sigma}_n - \check{\mu}_n\check{\mu}_n' & \check{\mu}_n \\ \check{\mu}_n' & 1 \end{pmatrix}$$

with $\check{\mu}_n = \check{\mu}_n(\widehat{P}_n) \in \boldsymbol{R}^{q-1}$ and $\check{\Sigma}_n = \check{\Sigma}_n(\widehat{P}_n) \in \boldsymbol{R}^{(q-1)\times(q-1)}$, then $(\check{\mu}_n, \check{\Sigma}_n)$ is the maximum likelihood estimator for $(\widetilde{\mu}_n, \widetilde{\Sigma}_n)$ under the model assumption that

$$(6.1)\quad \widetilde{P}_n(d\widetilde{y}) = \text{const.}(q-1)\,\det(\widetilde{\Sigma}_n)^{-1/2}(1 + (\widetilde{y} - \widetilde{\mu}_n)'\widetilde{\Sigma}_n^{-1}(\widetilde{y} - \widetilde{\mu}_n))^{-q/2}\,d\widetilde{y};$$

see Kent and Tyler (1991). The results of the present paper can be used directly to derive asymptotic properties of $(\check{\mu}_n, \check{\Sigma}_n)$, where (6.1) is replaced with general regularity conditions on $\widetilde{P}_n$.

7. Proofs

7.1 *Proofs for Section* 3

PROOF OF LEMMA 3.1. For any $B \in \boldsymbol{M}(0)$,

$$\text{trace}(D(Q,B)B) = q\int (|x|^{-4}(x'Bx)^2 - |x|^{-2}x'B^2x)\,Q(dx).$$

By the Cauchy-Schwarz inequality, $(x'Bx)^2 \le |x|^2(x'B^2x)$ with equality if, and only if, x is an eigenvector of B. Hence, if $\lambda_{(1)} > \cdots > \lambda_{(m)}$ are the distinct eigenvalues of B, and if $\boldsymbol{V}_i := \{x \in \boldsymbol{R}^q : Bx = \lambda_{(i)}x\}$, then $\text{trace}(D(Q,B)B) \le 0$ with equality if, and only if,

$$Q(\boldsymbol{V}_1 \cup \cdots \cup \boldsymbol{V}_m) = 1.$$

Now the assertion follows from the fact that $m > 1$ whenever $B \neq 0$. □

In order to prove Theorem 3.2 one needs explicit bounds for the norm of the remainder term $G((I+B)^{-1/2}Q) - G(Q) - D(Q,B)$.

LEMMA 7.1. *There is a universal constant $\kappa_o \in \boldsymbol{R}^+$ (not depending on Q or q) such that*

$$\|G((I+B)^{-1/2}Q) - G(Q) - D(Q,B)\| \le \kappa_o \, \|G(Q)\| \|B\|^2$$

for arbitrary $B \in \boldsymbol{M}$ with $\|B\| \le 1/2$.

PROOF OF LEMMA 7.1. For $A \in \boldsymbol{M}$ with $\lambda_1(A) < 1$ define

$$K(x,A) := G((I-A)x) - G(x) - 2D(x,A).$$

Then for $y := |x|^{-1}x \in \boldsymbol{S}(\boldsymbol{R}^q)$,

$$\begin{aligned}
K(x,A) &= \left(\frac{(I-A)G(y)(I-A)}{y'(I-A)^2y} - G(y)\right) - 2D(y,A) \\
&= \frac{(I-A)G(y)(I-A) - (1-2y'Ay+y'A^2y)G(y)}{y'(I-A)^2y} - 2D(y,A) \\
&= \frac{AG(y)A - y'A^2y\,G(y) + 2D(y,A)}{y'(I-A)^2y} - 2D(y,A) \\
&= \frac{AG(y)A - y'A^2y\,G(y) + 2(2y'Ay - y'A^2y)D(y,A)}{y'(I-A)^2y}.
\end{aligned}$$

The denominator $y'(I-A)^2y$ is not smaller than $\lambda_q((I-A)^2)$. As for the numerator, given any unit vector u, pick $v \in \boldsymbol{S}(\boldsymbol{R}^q)$ such that $Au = |Au|v$. Then

$$\begin{aligned}
|u'(AG(y)A - y'A^2y\,G(y))u| &\le \|A\|^2(v'G(y)v + u'G(y)u), \\
|u'(2y'Ay - y'A^2y)D(y,A)u| &\le (2\|A\| + \|A\|^2)|u'D(y,A)u| \\
&\le (2\|A\|^2 + \|A\|^3)(u'G(y)u + |u'G(y)v|).
\end{aligned}$$

Further there are orthonormal vectors $\widetilde{u}$, $\widetilde{v}$ such that

$$\begin{aligned}
u &= ((1+u'v)/2)^{1/2}\widetilde{u} + ((1-u'v)/2)^{1/2}\widetilde{v}, \\
v &= ((1+u'v)/2)^{1/2}\widetilde{u} - ((1-u'v)/2)^{1/2}\widetilde{v},
\end{aligned}$$

so that

$$\begin{aligned}
|u'G(y)v| &= 2^{-1}|(1+u'v)\widetilde{u}'G(y)\widetilde{u} - (1-u'v)\widetilde{v}'G(y)\widetilde{v}| \\
&\le 2^{-1}(1+u'v)\widetilde{u}'G(y)\widetilde{u} + 2^{-1}(1-u'v)\widetilde{v}'G(y)\widetilde{v}.
\end{aligned}$$

Hence

$$\begin{aligned}
\|K(Q,A)\| &\le \max_{u \in \boldsymbol{S}(\boldsymbol{R}^q)} \int |u'K(x,A)u| \, Q(dx) \\
&\le \lambda_1((I-A)^{-2})(10\|A\|^2 + 4\|A\|^3) \max_{u \in \boldsymbol{S}(\boldsymbol{R}^q)} u'G(Q)u \\
&= \lambda_1((I-A)^{-2})(10\|A\|^2 + 4\|A\|^3)\,\|G(Q)\|. \qquad (7.1)
\end{aligned}$$

Moreover, since $\|F(Q,\cdot)\| = \|G(Q)\|$,

$$\|D(Q,\cdot)\| \leq 2\|G(Q)\|. \tag{7.2}$$

Now let $B \in \boldsymbol{M}$ with $\|B\| \leq 1/2$ and define $A := I - (I+B)^{-1/2}$, i.e. $I + B = (I-A)^{-2}$. Then it follows from the spectral representation of B and A, together with a Taylor expansion of the function $t \mapsto 1-(1+t)^{-1/2}$, that $\lambda_1((I-A)^{-2}) \leq 1 + \|B\|$, $\|2A - B\| \leq (3/4)\|B\|^2 + \kappa'\|B\|^3$ and $\|A\| \leq \|B\|/2 + \kappa''\|B\|^2$ for universal constants $\kappa', \kappa'' \in \boldsymbol{R}^+$. Hence (7.1) and (7.2) imply that

$$\begin{aligned}
&\|G((I+B)^{-1/2}Q) - G(Q) - D(Q,B)\| \\
&\quad \leq \|K(Q,A)\| + \|D(Q, 2A-B)\| \\
&\quad \leq (1+\|B\|)(10\|A\|^2 + 4\|A\|^3)\|G(Q)\| + 2\|G(Q)\|\|2A-B\| \\
&\quad \leq \|G(Q)\|(4\|B\|^2 + \kappa'''\|B\|^3)
\end{aligned}$$

for suitable $\kappa''' = \kappa'''(\kappa', \kappa'')$. □

Proof of Theorem 3.2. For notational convenience let $L := D^{-1}(P,\cdot)$. Suppose that $\|L\| \leq b < \infty$ and $\|F(Q-P,\cdot)\| \leq \epsilon \in]0,1]$. Now $f(B) := L(G((I+B)^{-1/2}Q) - I)$ defines a continuous mapping from $\Omega := \{B \in \boldsymbol{M}(0) : \|B\| \leq \rho\}$ into $\boldsymbol{M}(0)$, where $\rho \in]0,1/2]$ is some constant. One can write

$$\begin{aligned}
f(B) &= LG(Q-P) + L(G((I+B)^{-1/2}Q) - G(Q)) \\
&= LG(Q-P) + B + LD(Q-P,B) \\
&\quad + L(G((I+B)^{-1/2}Q) - G(Q) - D(Q,B)) \\
&= LG(Q-P) + B + R(B),
\end{aligned}$$

where

$$\begin{aligned}
\|R(B)\| &\leq b\|D(Q-P,\cdot)\|\|B\| + b\kappa_o\|G(Q)\|\|B\|^2 \\
&\leq 2b\|F(Q-P,\cdot)\|\|B\| + 2b\kappa_o\|B\|^2 \\
&\leq 2b(\epsilon + \kappa_o\rho)\|B\|,
\end{aligned} \tag{7.3}$$

according to Lemma 7.1. Since $\|LG(Q-P)\| \leq b\epsilon$, this implies that

$$\|R(B)\| \leq \|B\|/2 \quad \text{and} \quad \|B - f(B)\| \leq \rho \quad \text{for all} \quad B \in \Omega,$$

provided that $b\epsilon$, $b\rho$ and ϵ/ρ are sufficiently small. Then Brouwer's Fixed Point theorem shows that $f(B_o) = 0$ for some $B_o \in \Omega$. If $f(B_1) = 0$ for some point $B_1 \in \Omega$, which is equivalent to $G((I+B_1)^{-1/2}Q) = I$, then $\|B_1\| \leq \|LG(Q-P)\| + \|R(B_1)\| \leq b\|G(Q-P)\| + \|B_1\|/2$, whence

$$\|B_1\| \leq 2b\|G(Q-P)\| \leq 2b\epsilon. \tag{7.4}$$

Combined with inequality (7.3) this yields

$$\begin{aligned}\|B_o + LG(Q-P)\| &= \|R(B_o)\| \\ &\le 4b^2\|F(Q-P,\cdot)\|\|G(Q-P)\| + 8b^3\kappa_o\|G(Q-P)\|^2 \\ &\le 4b^2(1+2b\kappa_o)\,\|F(Q-P,\cdot)\|\|G(Q-P)\|.\end{aligned}$$

It remains to be shown that $\Sigma(Q) = I + B_o$, i.e. that Q satisfies condition (2.2). Suppose the contrary. Then, by Theorem 2.1 [c] and (2.1), there exists a proper projection matrix $\Pi \in \boldsymbol{M}$ such that $G(M^{-1/2}Q) = I$ with $M = (I + B_o)^{1/2}(a\Pi + b(I-\Pi))(I+B_o)^{1/2}$ for arbitrary $a, b > 0$. But then one easily verifies that for suitable $a, b > 0$ the matrix $B_1 := M - I$ belongs to $\partial\Omega$, i.e. $\|B_1\| = \rho$. For sufficiently small ϵ/ρ this is in contradiction to (7.4). □

7.2 *Proofs for Section* 5

The proofs of Theorem 5.1 and Theorem 5.4 utilize the following two results.

LEMMA 7.2. *For any normed vector space* $(\boldsymbol{B}, \|\cdot\|)$ *let* $\boldsymbol{F}(\boldsymbol{B})$ *be a maximal subset of the sphere* $\boldsymbol{S}(\boldsymbol{B})$ *such that* $\|x-y\| > 1/3$ *for different* $x, y \in \boldsymbol{F}(\boldsymbol{B})$. *Then*

$$\#(\boldsymbol{F}(\boldsymbol{B})) \le \exp(2\,\dim(\boldsymbol{B})) \quad \text{and} \quad \|L\| \le (3/2) \max_{x \in \boldsymbol{F}(\boldsymbol{B})} \|Lx\|$$

for any linear function L from $\boldsymbol{B}$ *into another normed space. In particular,*

$$\|M\| \le 3 \max_{v \in \boldsymbol{F}(\boldsymbol{R}^q)} |v'Mv| \quad \text{for all} \quad M \in \boldsymbol{M}.$$

LEMMA 7.3. *For any finite collection of functions* $g_1, g_2, \ldots, g_m \in \mathcal{L}^1(P_n)$ *and arbitrary numbers* $t > 0$,

$$\left(\mathbb{E} \max_{1\le j\le m} g_j(\widehat{\Delta}_n)^2\right)^{1/2} \le 2\left(\frac{\log(2m)}{n} + \max_{1\le j\le m} \log \mathbb{E}\cosh(tg_j(\boldsymbol{y}_{n1}))\right) \Big/ t,$$

where $\widehat{\Delta}_n := \widehat{P}_n - P_n$.

In Lemma 7.2 the bound $\exp(2\,\dim(\boldsymbol{B}))$ for the cardinality of $\boldsymbol{F}(\boldsymbol{B})$ is standard and follows from considering balls of radius 1/6 with center in $\boldsymbol{F}(\boldsymbol{B})$ (cf. Pollard (1990), Section 4). The bounds for $\|L\|$ and $\|M\|$ are elementary.

PROOF OF LEMMA 7.3. This inequality is a modification of Pisier's (1983) Lemma 1.6, which is tailored for our purposes. It follows from Jensen's inequality and convexity of $\exp(\cdot)$ that

$$\begin{aligned}(\mathbb{E}\exp(\pm ntg_j(\widehat{\Delta}_n)/2))^{1/n} &= \mathbb{E}\exp(\pm t(g_j(\boldsymbol{y}_{n1}) - g_j(P_n))/2) \\ &= \mathbb{E}\exp(\mathbb{E}(\pm t(g_j(\boldsymbol{y}_{n1}) - g_j(\boldsymbol{y}_{n2}))/2 \mid \boldsymbol{y}_{n1})) \\ &\le \mathbb{E}\exp(\pm t(g_j(\boldsymbol{y}_{n1}) - g_j(\boldsymbol{y}_{n2}))/2) \\ &\le (\mathbb{E}\exp(\pm tg_j(\boldsymbol{y}_{n1})) + \mathbb{E}\exp(\mp tg_j(\boldsymbol{y}_{n2})))/2 \\ &= \mathbb{E}\cosh(tg_j(\boldsymbol{y}_{n1})).\end{aligned}$$

Thus $\mathbb{E}\,\psi(g_j(\widehat{\Delta}_n)^2) \le (\mathbb{E}\cosh(tg_j(y_{n1})))^n$, where $\psi(x) := \cosh(ntx^{1/2}/2)$ is convex and increasing in $x \ge 0$. Since $\psi^{-1}(y) \le (2\log(2y)/(nt))^2$ for $y \ge 1$, a second application of Jensen's inequality yields

$$\begin{aligned}
\mathbb{E}\max_j g_j(\widehat{\Delta}_n)^2 &\le \psi^{-1}\left(\mathbb{E}\max_j \psi(g_j(\widehat{\Delta}_n)^2)\right) \\
&\le \psi^{-1}\left(\sum_j \mathbb{E}\,\psi(g_j(\widehat{\Delta}_n)^2)\right) \\
&\le \psi^{-1}\left(m\max_j \mathbb{E}\,\psi(g_j(\widehat{\Delta}_n)^2)\right) \\
&\le \left((2/nt)\log(2m) + (2/t)\max_j \log\mathbb{E}\cosh(tg_j(y_{n1}))\right)^2 . \qquad \square
\end{aligned}$$

PROOF OF THEOREM 5.1. Note first that $\|D(P_n,B)+B\| = \|F(P_n,B)\| \le \sigma_n\kappa_n\|B\|$ for all $B \in \boldsymbol{M}(0)$, by the Cauchy-Schwarz inequality. Thus $\sup_{B\in \boldsymbol{S}(\boldsymbol{M}(0))}\|D^{-1}(P_n,B)+B\|$ converges to zero. Therefore, according to Theorem 3.2, it suffices to show that

$$\begin{aligned}
&\mathbb{E}\ \|F(\widehat{\Delta}_n,\cdot)\|^2 = o(1),\\
&\mathbb{E}\ \|G(\hat{\Delta}_n)\|^2 = O(q/n) \qquad \text{if} \quad q = O(n^{1/3}),
\end{aligned}$$

where $\widehat{\Delta}_n = \widehat{P}_n - P_n$. Lemma 7.2 yields

$$\begin{aligned}
&\|G(\widehat{\Delta}_n)\| \le 3 \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} |v'G(\widehat{\Delta}_n)v|,\\
&\|F(\widehat{\Delta}_n,\cdot)\| \le \|G(\widehat{\Delta}_n)\| + \|F(\widehat{\Delta}_n,\cdot)|_{\boldsymbol{M}(0)}\|,\\
&\|F(\widehat{\Delta}_n,\cdot)\|_{\boldsymbol{M}(0)}\| \le (9/2) \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q),B\subset \boldsymbol{F}(\boldsymbol{M}(0))} |v'F(\widehat{\Delta}_n,B)v|.
\end{aligned}$$

In order to bound the latter maximum we use a truncation argument. For any constant $K \ge 0$,

$$\begin{aligned}
|v'F(\widehat{\Delta}_n,B)v| &\le \left|\int |x|^{-2}x'Bx\,1\{v'G(x)v < K\}v'G(x)v\,\widehat{\Delta}_n(dx)\right| \\
&\quad + \int 1\{v'G(x)v \ge K\}v'G(x)v\,(\widehat{P}_n + P_n)(dx) \\
&= K|f(\widehat{\Delta}_n \,|\, v,B,K)| + g(\widehat{\Delta}_n \,|\, v,K) + 2g(P_n \,|\, v,K),
\end{aligned}$$

where

$$\begin{aligned}
&g(x \,|\, v,K) := 1\{v'G(x)v \ge K\}v'G(x)v,\\
&f(x \,|\, v,B,K) := |x|^{-2}x'Bx\,1\{v'G(x)v < K\}v'G(x)v/K
\end{aligned}$$

(and $f(x \mid v, B, 0) := 0$). Note that $v'G(\hat{\Delta}_n)v = g(\hat{\Delta}_n \mid v, 0)$ and $g(P_n \mid v, K) \leq \kappa_n^2/K$ for all $K > 0$. Thus it suffices to show that for arbitrary fixed $K \geq 0$,

$$\mathbb{E} \max_{v \in F(R^q)} g(\widehat{\Delta}_n \mid v, K)^2 = \begin{cases} o(1), & \\ O(q/n) & \text{if} \quad q = O(n^{1/3}), \end{cases} \tag{7.5}$$

$$\mathbb{E} \max_{v \in F(R^q), B \in F(M(0))} f(\widehat{\Delta}_n \mid v, B, K)^2 = o(1). \tag{7.6}$$

Since $|g(\cdot)| \leq q$, $\mathbb{E}(g(y_{n1} \mid v, K)^2) \leq \kappa_n^2$, $|f(\cdot)| \leq 1$ and $\mathbb{E}(f(y_{n1} \mid v, B, K)^2) \leq \sigma_n^2$, one obtains

$$\begin{aligned} \mathbb{E}\cosh(tg(y_{n1} \mid v, K)) &\leq 1 + \sum_{k=1}^{\infty} t^{2k} \kappa_n^2 q^{2k-2}/(2k)! \\ &= 1 + (\kappa_n/q)^2(\cosh(qt) - 1) \\ &\leq \exp((\kappa_n/q)^2(\cosh(qt) - 1)), \\ \mathbb{E}\cosh(tf(y_{n1} \mid v, B, K)) &\leq \exp(\sigma_n^2(\cosh(t) - 1)) \end{aligned}$$

for all $t > 0$. Combining this with Lemmas 7.2 and 7.3 we get

$$\begin{aligned} &\left(\mathbb{E} \max_{v \in F(R^q)} g(\widehat{\Delta}_n \mid v, K)^2\right)^{1/2} \\ &\quad \leq 2 \min_{t>0} ((2q+1)/n + (\kappa_n/q)^2(\cosh(qt) - 1))/t \\ &\quad \leq 2\kappa_n^2/q \min_{r>0} ((3/\kappa_n^2)q^3/n + \cosh(r) - 1)/r \\ &\quad = 2\kappa_n^2/q \; h((3/\kappa_n^2)q^3/n), \end{aligned} \tag{7.7}$$

$$\begin{aligned} &\left(\mathbb{E} \max_{v \in F(R^q), B \in F(M(0))} f(\widehat{\Delta}_n \mid v, B, K)^2\right)^{1/2} \\ &\quad \leq 2 \min_{t>0} ((2q + q(q+1) + 1)/n + \sigma_n^2(\cosh(t) - 1))/t \\ &\quad \leq 2\sigma_n^2 \; h(5q^2/(n\sigma_n^2)). \end{aligned} \tag{7.8}$$

Here $h(a) := \min_{r>0}(a + \cosh(r) - 1)/r$ is increasing in $a > 0$ with

$$h(a) = \begin{cases} (2a)^{1/2}\,(1 + o(1)) & \text{as} \quad a \to 0, \\ a/\log a\,(1 + o(1)) & \text{as} \quad a \to \infty. \end{cases}$$

Consequently, (7.7) and (7.8) imply (7.5) and (7.6). □

The next lemma summarizes some (in)equalities for spherically symmetric distributions.

LEMMA 7.4. *Let $y \sim \mathcal{N}_q(0, I)$. Then $\mathbb{E}\, G(y) = I$ and*

$$\mathbb{E}(\operatorname{trace}(G(y)A)\operatorname{trace}(G(y)B)) = \frac{\operatorname{trace}(A)\operatorname{trace}(B) + 2\operatorname{trace}(AB)}{1 + 2/q} \tag{7.9}$$

for all $A, B \in \boldsymbol{M}$. In particular, for spherically symmetric distributions P on $\boldsymbol{R}^q \setminus \{0\}$,

$$(7.10) \qquad D(P,B) = -(1+2/q)^{-1}B \quad \textit{for all} \quad B \in \boldsymbol{M}(0).$$

Moreover,

$$(7.11) \qquad \mathbb{E}((v'G(\boldsymbol{y})v)^k) \le \mathbb{E}((v'\boldsymbol{y})^{2k}) \le 2^k k! \qquad \textit{for arbitrary} \quad v \in \boldsymbol{S}(\boldsymbol{R}^q) \quad \textit{and integers} \quad k \ge 1,$$

$$(7.12) \qquad \begin{aligned} \mathbb{E}\exp(\operatorname{trace}(G(\boldsymbol{y})B)) &\le \mathbb{E}\exp(\boldsymbol{y}'B\boldsymbol{y}) \\ &\le \exp\left(\operatorname{trace}(B) + \frac{\operatorname{trace}(B^2)}{(1-2\|B\|)^+}\right) \end{aligned} \qquad \textit{for arbitrary} \quad B \in \boldsymbol{M}.$$

Proof of Lemma 7.4. The key point is that $G(\boldsymbol{y}) = G(\boldsymbol{u})$ and $\boldsymbol{y}\boldsymbol{y}' = q^{-1}|\boldsymbol{y}|^2 G(\boldsymbol{u})$, where $\boldsymbol{u} := |\boldsymbol{y}|^{-1}\boldsymbol{y}$ and $|\boldsymbol{y}|^2$ are stochastically independent with $\mathcal{L}(|\boldsymbol{y}|^2) = \chi^2_q$. Consequently, for $A, B \in \boldsymbol{M}$,

$$\begin{aligned} &\mathbb{E}(\operatorname{trace}(G(\boldsymbol{y})A)\operatorname{trace}(G(\boldsymbol{y})B)) \\ &\quad = \mathbb{E}(\operatorname{trace}(\boldsymbol{y}\boldsymbol{y}'A)\operatorname{trace}(\boldsymbol{y}\boldsymbol{y}'B))/\,\mathbb{E}(q^{-2}|\boldsymbol{y}|^4) \\ &\quad = (1+2/q)^{-1}\,\mathbb{E}(\operatorname{trace}(\boldsymbol{y}\boldsymbol{y}'A)\operatorname{trace}(\boldsymbol{y}\boldsymbol{y}'B)) \\ &\quad = (1+2/q)^{-1}\,\mathbb{E}(\boldsymbol{y}'A\boldsymbol{y}\,\boldsymbol{y}'B\boldsymbol{y}) \\ &\quad = (1+2/q)^{-1}(\operatorname{trace}(A)\operatorname{trace}(B) + 2\operatorname{trace}(AB)), \end{aligned}$$

where the last equality follows from elementary moment calculations. In particular, if $A, B \in \mathcal{M}(0)$, then

$$\operatorname{trace}(F(P,B)A) = q^{-1}\,\mathbb{E}(\operatorname{trace}(G(\boldsymbol{y})A)\operatorname{trace}(G(\boldsymbol{y})B)) = 2(q+2)^{-1}\operatorname{trace}(AB).$$

Hence $F(P,B) = 2(2+q)^{-1}B$ and $D(P,B) = -(1+2/q)^{-1}B$.

Generally, for any convex function $\psi : \boldsymbol{M} \to \boldsymbol{R}$, Jensen's inequality yields

$$\mathbb{E}\,\psi(G(\boldsymbol{y})) - \mathbb{E}\,\psi(\mathbb{E}(\boldsymbol{y}\boldsymbol{y}' \mid \boldsymbol{u})) \le \mathbb{E}\,\mathbb{E}(\psi(\boldsymbol{y}\boldsymbol{y}') \mid \boldsymbol{u}) - \mathbb{E}\,\psi(\boldsymbol{y}\boldsymbol{y}').$$

In particular, for $v \in \boldsymbol{S}(\boldsymbol{R}^q)$ and integers $k \ge 1$,

$$\mathbb{E}(|v'G(\boldsymbol{y})v|^k) \le \mathbb{E}((v'\boldsymbol{y})^{2k}) = \mathbb{E}(y_1^{2k}) = \prod_{j=0}^{k-1}(1+2j) \le 2^k k!,$$

while for any $B \in \boldsymbol{M}$ and $\boldsymbol{y} = (y_i)_{1 \le i \le q}$,

$$\begin{aligned} \mathbb{E}\exp(\operatorname{trace}(G(\boldsymbol{y})B)) &\le \mathbb{E}\exp(\boldsymbol{y}'B\boldsymbol{y}) \\ &= \mathbb{E}\exp\left(\sum_{i=1}^{q}\lambda_i(B)y_i^2\right) \end{aligned}$$

$$= \mathbb{E}\exp\left(-2^{-1}\sum_{i=1}^{q}\log((1-2\lambda_i(B))^+)\right)$$
$$= \mathbb{E}\exp\left(\sum_{i=1}^{q}\sum_{k=1}^{\infty}2^{k-1}\lambda_i(B)^k/k\right)$$
$$\leq \mathbb{E}\exp\left(\sum_{i=1}^{q}\left(\lambda_i(B)+\lambda_i(B)^2\sum_{k=0}^{\infty}(2\|B\|)^k\right)\right)$$
$$= \exp(\mathrm{trace}(B)+\mathrm{trace}(B^2)/(1-2\|B\|)^+). \qquad \Box$$

PROOF OF THEOREM 5.4 [a]. Since $M \mapsto G(M^{-1/2}\widehat{P}_n)$ depends only on the directions $|\boldsymbol{y}_{ni}|^{-1}\boldsymbol{y}_{ni}$, which are uniformly distributed on $\boldsymbol{S}(\boldsymbol{R}^q)$, one may assume without loss of generality that P_n is a standard normal distribution on $\boldsymbol{R}^q$. With the same notation as in the proof of Theorem 5.1, the first two assertions in part [a] follow from Theorem 3.2 and (7.10), provided that the following two claims are true:

$$\mathbb{E}\max_{v\in\boldsymbol{F}(\boldsymbol{R}^q)} g(\widehat{\Delta}_n\,|\,v,K)^2 = O(q/n) \quad \text{uniformly in} \quad K\geq 0, \tag{7.13}$$

$$\max_{v\in\boldsymbol{F}(\boldsymbol{R}^q)} g(P_n\,|\,v,K_n)^2 + K_n^2\,\mathbb{E}\max_{v\in\boldsymbol{F}(\boldsymbol{R}^q),\,B\in\boldsymbol{F}(\boldsymbol{M}(0))} f(\widehat{\Delta}_n\,|\,v,B,K_n)^2 = O(\log(n/q)^2\,q/n) \tag{7.14}$$

for suitable numbers K_n in $\boldsymbol{R}^+$.

Note first that

$$\mathbb{E}\cosh(tg(\boldsymbol{y}_{n1}\,|\,v,K)) \leq \mathbb{E}\cosh(tv'G(\boldsymbol{y}_{n1})v) \leq \sum_{k=0}^{\infty}(2t)^{2k} = (1-4t^2)^{-1},$$

according to (7.11). Thus Lemmas 7.2 and 7.3 yield

$$\left(\mathbb{E}\max_{v\in\boldsymbol{F}(\boldsymbol{R}^q)} g(\widehat{\Delta}_n\,|\,v,K)^2\right)^{1/2} \leq (2/t_n)(2q/n-\log(1-4t_n^2)) = O((q/n)^{1/2})$$

if $t_n := \min\{(q/n)^{1/2}, 1/2\}$. Moreover, it follows from (7.12) that

$$g(P_n\,|\,v,K_n) \leq K_n\exp(-K_n/3)\,\mathbb{E}\exp(v'G(\boldsymbol{y}_{n1})v/3)$$
$$\leq K_n\exp(-K_n/3)\exp(2/3) \quad \text{for all} \quad v\in\boldsymbol{S}(\boldsymbol{R}^q)$$

whenever $K_n\geq 3$, because $x\exp(-x/3)$ is decreasing in $x\geq 3$. Setting K_n equal to $(3/2)\log(n/q)$ shows that a sufficient condition for (7.14) is given by

$$\mathbb{E}\max_{v\in\boldsymbol{F}(\boldsymbol{R}^q),\,B\in\boldsymbol{F}(\boldsymbol{M}(0))} f(\widehat{\Delta}_n\,|\,v,B,K_n)^2 = O(q/n). \tag{7.15}$$

But for any $B\in\boldsymbol{S}(\boldsymbol{M}(0))$ and $0<t\leq q/2$,

$$\mathbb{E}\cosh(tf(\boldsymbol{y}_{n1}\,|\,v,B,K_n)) \leq \mathbb{E}\cosh((t/q)\,\mathrm{trace}(G(\boldsymbol{y}_{n1})B))$$
$$\leq \exp((t/q)^2\,\mathrm{trace}(B^2)/(1-2t/q))$$
$$\leq \exp((t^2/q)/(1-2t/q)),$$

according to (7.12). Now (7.15) follows from Lemma 7.2 and 7.3 if $t = t_n$ is taken to be $q \min\{(q/n)^{1/2}, 1/2\}$.

As for the coupling with a Wishart matrix, the preceding results show that $\Sigma(\widehat{P}_n)$ may be replaced with $G(\widehat{P}_n)$. The matrix $M_n := n \int xx' \widehat{P}_n(dx)$ has the desired Wishart distribution. Further,

$$\begin{aligned}
&\mathbb{E}(G(y_{n1}) - y_{n1}y_{n1}') = 0,\\
&\mathbb{E}(|v'(G(y_{n1}) - y_{n1}y_{n1}')v|^2)\\
&\qquad = \mathbb{E}((v'G(y_{n1})v)^2)\operatorname{Var}(|y_{n1}|^2/q) = 6/(q+2),\\
&\mathbb{E}(|v'(G(y_{n1}) - y_{n1}y_{n1}')v|^k)\\
&\qquad \le (\mathbb{E}((2v'G(y_{n1})v)^k) + \mathbb{E}((2v'y_{n1}y_{n1}'v)^k))/2 \le 4^k k!,
\end{aligned}$$

see (7.9) and (7.11). Consequently,

$$\mathbb{E}\cosh(tv'(G(y_{n1}) - y_{n1}y_{n1}')v) \le 1 + 3t^2/q + (4t)^4/(1-16t^2) \quad \text{for} \quad 0 < t \le 1/4.$$

If we take $t = t_n = \min\{\epsilon^{-1}(q/n)^{1/2}, 1/4\}$ for arbitrarily small $\epsilon > 0$, it follows from Lemmas 7.2 and 7.3 that $\mathbb{E}(\|G(\widehat{P}_n) - n^{-1}M_n\|^2) = o(q/n)$.

As for the assertion about the eigenvalues of $\Sigma(\widehat{P}_n)$, one can modify Silverstein's (1985) arguments in order to show that

$$\|n^{-1}M_n - T_qD_nT_q'\| = O_p((\log(q)/n)^{1/2}),$$

where T_q is Haar-distributed on the group of orthonormal matrices in $\boldsymbol{R}^{q\times q}$, while D_n denotes the non-random tridiagonal matrix

$$D_n := n^{-1}\begin{pmatrix}
n & (n(q-1))^{1/2} & 0 & \cdots & 0\\
 & n+q-2 & ((n-1)(q-2))^{1/2} & \ddots & \vdots\\
 & & n+q-4 & \ddots & 0\\
 & & & \ddots & (n-q+2)^{1/2}\\
\text{(symm.)} & & & & n-q+2
\end{pmatrix}.$$

Silverstein (1985) derived from Geršgorin's theorem that

$$\lambda_1(D_n) \le (1 + (q/n)^{1/2})^2 \quad \text{and} \quad \lambda_q(D_n) \ge (1 - (q/n)^{1/2})^2.$$

On the other hand consider unit vectors

$$\begin{aligned}
u_{n,+} &:= k^{-1/2}(0,1,1,\ldots,1,0,\ldots,0)',\\
u_{n,-} &:= k^{-1/2}(0,-1,1,-1,\ldots,(-1)^k,0,\ldots,0)'
\end{aligned}$$

in $\boldsymbol{R}^q$ with $k = k_n = q^{1/2} + O(1)$ nonzero coefficients. Then $\lambda_1(D_n) \geq u'_{n,+} D_n u_{n,+}$ and $\lambda_q(D_n) \leq u'_{n,-} D_n u_{n,-}$ with

$$\begin{aligned} u'_{n,\pm} D_n u_{n,\pm} &= (kn)^{-1} \left(\sum_{i=2}^{k+1} (n+q-2i) \ \pm\ 2 \sum_{i=2}^{k} ((n-i)(q-i-1))^{1/2} \right) \\ &= (1 \pm (q/n)^{1/2})^2 + O(n^{-1/2}). \end{aligned}$$

□

Proof of Theorem 5.4 [b]. Because of (5.1) and the proof of part [a], we know that $\|\Sigma(\widehat{P}_n^s) - G(\widehat{P}_n^s)\|$ has expectation $o((q/n)^{1/2})$. Thus it suffices to analyze $G(\widehat{P}_n^s)$ in more detail. This is just a matrix-valued U-statistic with Hoeffding-decomposition

$$G(\widehat{P}_n^s) = \widetilde{G}(\widehat{U}_n) = I + 2\widetilde{G}(\widehat{\Delta}_n \otimes P_n) + \widetilde{G}(\widehat{R}_n),$$

where $\widetilde{G}(x,y) = G(x-y)$ and

$$\widehat{U}_n := \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \delta_{y_{ni}} \otimes \delta_{y_{nj}} \qquad \text{and} \qquad \widehat{R}_n := \widehat{U}_n - 2\widehat{P}_n \otimes P_n + P_n \otimes P_n.$$

Now we show that

$$\mathbb{E}\, \|\widetilde{G}(\widehat{R}_n)\| \leq 3\, \mathbb{E} \max_{v \in \boldsymbol{F}(\boldsymbol{R}^q)} |v'\widetilde{G}(\widehat{R}_n)v| = o((q/n)^{1/2}).$$

For that purpose we use once more a truncation argument. Let $g(x \mid v, K)$ be defined as in the proof of Theorem 5.1 and $h(x \mid v, K) := 1\{v'G(x)v < K\} v'G(x)v = v'G(x)v - g(x \mid v, K)$. Further let $\widetilde{g}(x, y \mid v, K) := g(x-y \mid v, K)$ and $\widetilde{h}(x, y \mid v, K) := h(x-y \mid v, K)$. Then it suffices to show that for suitable positive numbers K_n,

$$\mathbb{E} \max_{v \in \boldsymbol{F}(\boldsymbol{R}^q)} |\widetilde{g}(\widehat{R}_n \mid v, K_n)| = o_{\mathrm{p}}((q/n)^{1/2}), \tag{7.16}$$

$$\mathbb{E} \max_{v \in \boldsymbol{F}(\boldsymbol{R}^q)} |\widetilde{h}(\widehat{R}_n \mid v, K_n)| = o_{\mathrm{p}}((q/n)^{1/2}). \tag{7.17}$$

In order to prove (7.16), let π be uniformly distributed on the set of permutations of $\{1, 2, \ldots, n\}$ and independent from $(\boldsymbol{y}_{ni})_{1 \leq i \leq n}$. Then

$$\begin{aligned} \widehat{U}_n &= \mathbb{E} \left(m^{-1} \sum_{i=1}^{n} \delta_{y_{n,\pi(2i-1)}} \otimes \delta_{y_{n,\pi(2i)}} \middle| \boldsymbol{y}_{n1}, \boldsymbol{y}_{n2}, \ldots, \boldsymbol{y}_{nn} \right), \\ \widehat{P}_n \otimes P_n &= \mathbb{E} \left(m^{-1} \sum_{i=1}^{n} \delta_{y_{n,\pi(2i-1)}} \otimes P_n \middle| \boldsymbol{y}_{n1}, \boldsymbol{y}_{n2}, \ldots, \boldsymbol{y}_{nn} \right), \\ m^{-1} \sum_{i=1}^{n} & \delta_{y_{n,\pi(2i-1)}} \otimes P_n \\ &= \mathbb{E} \left(m^{-1} \sum_{i=1}^{n} \delta_{y_{n,2i-1}} \otimes \delta_{y_{n,2i}} \middle| \boldsymbol{y}_{n2}, \boldsymbol{y}_{n4}, \boldsymbol{y}_{n6}, \ldots \right), \end{aligned} \tag{7.18}$$

where $m = m_n = \lfloor n/2 \rfloor$. Consequently, applying Jensen's inequality three times while using the fact that $\mathcal{L}((\boldsymbol{y}_{ni})_i) = \mathcal{L}((\boldsymbol{y}_{n,\pi(i)})_i)$ gives us

$$\begin{aligned}
&\mathbb{E} \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} |\widetilde{g}(\widehat{R}_n \,|\, v, K_n)| \\
&\quad \le \mathbb{E} \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} |\widetilde{g}(\widehat{U}_n - P_n \otimes P_n \,|\, v, K_n)| \\
&\qquad + 2\,\mathbb{E} \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} |\widetilde{g}(\widehat{P}_n \otimes P_n - P_n \otimes P_n \,|\, v, K_n)| \\
&\quad \le 3\,\mathbb{E} \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} \left| \widetilde{g}\left(m^{-1} \sum_{i=1}^{n} \delta_{\boldsymbol{y}_{n,2i-1}} \otimes \delta_{\boldsymbol{y}_{n,2i}} - P_n \otimes P_n \,|\, v, K_n \right) \right| \\
&\quad = 3\,\mathbb{E} \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} |g(\check{P}_n^{\mathrm{s}} - P_n^{\mathrm{s}} \,|\, v, K_n)| \\
&\quad \le 6\left((2q+1)/m + \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} \log \mathbb{E} \cosh(tg(\boldsymbol{y}_{n1} - \boldsymbol{y}_{n2} \,|\, v, K_n)) \right) \Big/ t
\end{aligned}$$

$$= 6\left((2q+1)/m + \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} \log \mathbb{E} \cosh(tg(\boldsymbol{y}_{n1} \,|\, v, K_n)) \right) \Big/ t \tag{7.19}$$

for arbitrary $t > 0$, where $\check{P}_n^{\mathrm{s}}$ was defined in Remark 5.3. The last inequality follows from Lemmas 7.2 and 7.3, applied to $(m, \check{P}_n^{\mathrm{s}}, P_n^{\mathrm{s}})$ in place of $(n, \widehat{P}_n, P_n)$. The last equality is due to $G(\boldsymbol{y}_{n1} - \boldsymbol{y}_{n2})$ being distributed as $G(\boldsymbol{y}_{n1})$. Now we deduce from (7.11) that

$$\begin{aligned}
\mathbb{E}(g(\boldsymbol{y}_{n1} \,|\, v, K_n)^k) &\le \mathbb{E}((v'G(\boldsymbol{y}_{n1})v)^k) \le 2^k k!, \\
\mathbb{E}(g(\boldsymbol{y}_{n1} \,|\, v, K_n)^2) &\le \mathbb{E}((v'G(\boldsymbol{y}_{n1})v)^4)/K_n^2 \le C/K_n^2,
\end{aligned}$$

whence $\log \mathbb{E} \cosh(tg(\boldsymbol{y}_{n1} \,|\, v, K_n)) \le Ct^2/K_n^2 + 16t^4/(1-4t^2)_+$. Consequently, if $K_n \to \infty$ but $K_n^4 \le n/q$, then (7.16) follows by setting $t = K_n(q/n)^{1/2}$ in (7.19).

As for (7.17), an exponential inequality for degenerate U-statistics yields

$$\mathbb{E} \cosh(cn\widetilde{h}(\widehat{R}_n \,|\, v, K_n)/K_n) \le e$$

for some universal constant $c > 0$. This follows from Nolan and Pollard ((1987), Section 2) or Arcones and Giné ((1994), Proposition 2.3(d)). Hence, with $\psi(x) := \cosh(cnx^{1/2}/K_n)$ for $x \ge 0$, one can conclude from Lemmas 7.2 and Pisier's (1983) Lemma 1.6 that

$$\begin{aligned}
&(\mathbb{E} \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} \widetilde{h}(\widehat{R}_n \,|\, v, K_n)^2)^{1/2} \\
&\quad \le \left(\psi^{-1}\left(\exp(2q) \max_{v\in \boldsymbol{F}(\boldsymbol{R}^q)} \mathbb{E}\, \psi(\widetilde{h}(\widehat{R}_n \,|\, v, K_n)^2) \right) \right)^{1/2} \\
&\quad \le 2K_n(q+1)/(nc) \\
&\quad = O((q/n)^{3/4}),
\end{aligned}$$

because $\psi^{-1}(y) \le (K_n \log(2y)/(nc))^2$.

Now we consider the random matrix $\widetilde{G}(\widehat{\Delta}_n \otimes P_n)$ in more detail. For fixed $x \in \boldsymbol{R}^q$ let $v_1, v_2, \ldots, v_q$ be an orthonormal basis of $\boldsymbol{R}^q$ such that $x = |x|v_1$. Then with $\boldsymbol{y}_{n1} = (y_{n,i})_{1<i<n}$ and $h_n(r) := \mathbb{E}((r - y_{n,1})^2/((r - y_{n,1})^2 + |\boldsymbol{y}_{n1}|^2 - y_{n,1}^2))$,

$$
\begin{aligned}
v_1'\widetilde{G}(x, P_n)v_1 &= q\,\mathbb{E}((|x| - v_1'\boldsymbol{y}_{n1})^2/((|x| - v_1'\boldsymbol{y}_{n1})^2 + |\boldsymbol{y}_{n1}|^2 - (v_1'\boldsymbol{y}_{n1})^2)) \\
&= qh_n(|x|), \\
v_i'\widetilde{G}(x, P_n)v_i &= q\,\mathbb{E}((v_i'\boldsymbol{y}_{n1})^2/((|x| - v_1'\boldsymbol{y}_{n1})^2 + |\boldsymbol{y}_{n1}|^2 - (v_1'\boldsymbol{y}_{n1})^2)) \\
&= q(q-1)^{-1}\,\mathbb{E}((|\boldsymbol{y}_{n1}|^2 - y_{n,1}^2)/((|x| - y_{n,1})^2 + |\boldsymbol{y}_{n1}|^2 - y_{n,1}^2)) \\
&= q(q-1)^{-1}(1 - h_n(|x|)) \quad \text{for} \quad 2 \le i \le q, \\
v_i'\widetilde{G}(x, P_n)v_j &= 0 \quad \text{for} \quad 1 \le i < j \le q.
\end{aligned}
$$

Hence $\widetilde{G}(x, P_n) - I$ can be written as

$$
qh_n(|x|)v_1v_1' + q(q-1)^{-1}(1 - h_n(|x|))(I - v_1v_1') - I = 2^{-1}H_n(|x|)(G(x) - I),
$$

where

$$
H_n(r) := 2(qh_n(r) - 1)/(q - 1) \in [0, 2[.
$$

This leads to the representation $n^{-1}\sum_{i=1}^n H_n(|\boldsymbol{y}_{ni}|)(G(\boldsymbol{y}_{ni}) - I)$ of $2\widetilde{G}(\widehat{\Delta}_n \otimes P_n)$.

Finally, suppose that $|\boldsymbol{y}_{n1}|^2/q \to_{\mathrm{p}} \kappa_o > 0$. Then one easily verifies that $h_n(|\boldsymbol{y}_{n1}|) \to_{\mathrm{p}} 1/2$ and thus $H_n(|\boldsymbol{y}_{n1}|) \to_{\mathrm{p}} 1$. Since $0 \le H_n < 2$, this implies that

$$
\begin{aligned}
&\mathbb{E}(|(H_n(|\boldsymbol{y}_{n1}|) - 1)v'(G(\boldsymbol{y}_{n1}) - I)v|^k) \\
&\quad = \mathbb{E}(|H_n(|\boldsymbol{y}_{n1}|) - 1|^k)\,\mathbb{E}(|v'(G(\boldsymbol{y}_{n1}) - I)v|^k) \\
&\quad \le \epsilon_n 4^k k!
\end{aligned}
$$

for $k > 2$, where $\epsilon_n \to 0$. Thus $\log \mathbb{E}\cosh(t(H_n(|\boldsymbol{y}_{n1}|) - 1)v'(G(\boldsymbol{y}_{n1}) - I)v)$ is not greater than $16\epsilon_n t^2/(1 - 16t^2)^+$, and a final application of Lemmas 7.2 and 7.3 gives us

$$
\mathbb{E}\left\|G(\widehat{P}_n) - I - n^{-1}\sum_{i=1}^n H_n(|\boldsymbol{y}_{ni}|)(G(\boldsymbol{y}_{ni}) - I)\right\| = o((q/n)^{1/2}).
$$

But part [a] provides the expansion $\mathbb{E}\,\|\Sigma(\widehat{P}_n) - G(\widehat{P}_n)\| = o_{\mathrm{p}}((q/n)^{1/2})$. Consequently, $\mathbb{E}\,\|\Sigma(\widehat{P}_n^{\mathrm{s}}) - \Sigma(\widehat{P}_n)\| = o_{\mathrm{p}}((q/n)^{1/2})$. □

Acknowledgements

I am grateful to a referee for numerous suggestions concerning the exposition.

References

Arcones, M. A. and Gine, E. (1993). Limit theorems for *U*-processes, *Ann. Probab.*, **21**, 1494–1542.

Bai, Z. D. and Wu, Y. (1994*a*). Limiting behavior of *M*-estimators of regression coefficients in high dimensional linear models. I. Scale-dependent case, *J. Multivariate Anal.*, **51**, 211–239.

Bai, Z. D. and Wu, Y. (1994*b*). Limiting behavior of *M*-estimators of regression coefficients in high dimensional linear models. II. Scale-invariant case, *J. Multivariate Anal.*, **51**, 240–251.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.*, **57**, 33–45.

Bickel, P. and Freedman, D. (1981). Some asymptotic theory for the bootstrap, *Ann. Statist.*, **9**, 1196–1217.

Clarke, B. R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations, *Ann. Statist.*, **11**, 1196–1205.

Dümbgen, L. (1997). The asymptotic behavior of Tyler's *M*-estimator of scatter in high dimension, Beitrag zur Statistik # 23, Heidelberg University.

Dümbgen, L. (1998). Perturbation inequalities and confidence sets for functions of a scatter matrix, *J. Multivariate Anal.*, **65**, 19–35.

Girko, V. L. (1995). *Statistical Analysis of Observations of Increasing Dimension*, Kluwer, Dordrecht.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, **58**, 13–30.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Kent, J. T. and Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution, *Journal of Applied Statistics,* **15**, 247–254.

Kent, J. T. and Tyler, D. E. (1991). Redescending *M*-estimates of multivariate location and scatter, *Ann. Statist.*, **19**, 2102–2119.

Mammen, E. (1996). Empirical process of residuals for high-dimensional linear models, *Ann. Statist.*, **24**, 307–335.

Maronna, R. A. (1976). Robust *M*-estimators of multivariate location and scatter, *Ann. Statist.*, **4**, 51–67.

Nolan, D. and Pollard, D. (1987). *U*-processes: rates of convergence, *Ann. Statist.*, **15**, 780–799.

Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis, Banach Spaces, Harmonic Analysis, and Probability Theory (eds. R. C. Blei and S. J. Sidney), *Lecture Notes in Math.*, **995**, 123–154, Springer, Berlin.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*, NSF-CBMS Regionial Conference Series in Probability and Statistics, **2**, IMS, Hayward, California.

Portnoy, S. (1984). Asymptotic behavior *M*-estimators of p regression parameters when p^2/n is large, I. Consistency, *Ann. Statist.*, **12**, 1298–1309.

Portnoy, S. (1985). Asymptotic behavior *M*-estimators of p regression parameters when p^2/n is large, II. Asymptotic normality, *Ann. Statist.*, **13**, 1403–1417.

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity, *Ann. Statist.*, **16**, 356–366.

Silverstein, J. W. (1985). The smallest eigenvalue of a large dimensional Wishart matrix, *Ann. Probab.*, **13**, 1364–1368.

Tyler, D. E. (1987). A distribution-free *M*-estimator of multivariate scatter, *Ann. Statist.*, **15**, 234–251.