Journal of Multivariate Analysis 143 (2016) 249–274

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals[☆]

Romain Couillet[a,*], Abla Kammoun[b], Frédéric Pascal[a]

[a] CentraleSupélec, Université Paris Saclay, Gif-sur-Yvette, France

[b] King Abdullah's University of Science and Technology, Saudi Arabia

ARTICLE INFO

Article history:
Received 1 October 2014
Available online 14 September 2015

AMS subject classifications:
62H12
60F05
62G35

Keywords:
Random matrix theory
Robust estimation
Central limit theorem
GLRT

ABSTRACT

A central limit theorem for bilinear forms of the type $a^*\hat{C}_N(\rho)^{-1}b$, where $a, b \in \mathbb{C}^N$ are unit norm deterministic vectors and $\hat{C}_N(\rho)$ a robust-shrinkage estimator of scatter parametrized by ρ and built upon n independent elliptical vector observations, is presented. The fluctuations of $a^*\hat{C}_N(\rho)^{-1}b$ are found to be of order $N^{-\frac{1}{2}}$ and to be the same as those of $a^*\hat{S}_N(\rho)^{-1}b$ for $\hat{S}_N(\rho)$ a matrix of a theoretical tractable form. This result is exploited in a classical signal detection problem to provide an improved detector which is both robust to elliptical data observations (e.g., impulsive noise) and optimized across the shrinkage parameter ρ.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

As an aftermath of the growing interest for large dimensional data analysis in machine learning, in a recent series of articles [14,15,13,32,17], several estimators from the field of robust statistics (dating back to the seventies) started to be explored under the assumption of commensurably large sample (n) and population (N) dimensions. Robust estimators were originally designed to turn classical estimators into outlier- and impulsive noise-resilient estimators, which are of considerable importance in the recent big data paradigm. Among these estimation methods, robust regression was studied in [17] which reveals that, in the large N, n regime, the difference in norm between estimated and true regression vectors (of size N) tends almost surely to a positive constant which depends on the nature of the data and of the robust regressor. In parallel, and of more interest to the present work, Couillet et al. [14,15], Couillet and McKay [13], Zhang et al. [32] studied the limiting behavior of several classes of robust estimators $\hat{C}_N$ of scatter (or covariance) matrices C_N based on independent zero-mean elliptical observations $x_1, \ldots, x_n \in \mathbb{C}^N$. Precisely, Couillet et al. [14] show that, letting $N/n < 1$ and $\hat{C}_N$ be the (almost sure) unique solution to

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i\right)x_ix_i^*$$

☆ Couillet's work is supported by the ERC MORE EC–120133. Pascal's work is supported by the DGA grant no. 2013.60.0011.00.470.75.01.

* Corresponding author.
E-mail addresses: romain.couillet@supelec.fr (R. Couillet), abla.kammoun@kaust.edu.sa (A. Kammoun), frederic.pascal@supelec.fr (F. Pascal).

http://dx.doi.org/10.1016/j.jmva.2015.08.021
0047-259X/© 2015 Elsevier Inc. All rights reserved.

under some appropriate conditions over the nonnegative function u (corresponding to Maronna's M-estimator [24]), $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ in spectral norm as $N, n \to \infty$ with $N/n \to c \in (0, 1)$, where $\hat{S}_N$ follows a standard random matrix model (such as studied in [29,12]). In [32], the important scenario where $u(x) = 1/x$ (referred to as Tyler's M-estimator) is treated. It is in particular shown for this model that for identity scatter matrices the spectrum of $\hat{C}_N$ converges weakly to the Marčenko–Pastur law [23] in the large N, n regime. Finally, for $N/n \to c \in (0, \infty)$, Couillet and McKay [13] studied yet another robust estimation model defined, for each $\rho \in (\max\{0, 1 - n/N\}, 1]$, by $\hat{C}_N = \hat{C}_N(\rho)$, unique solution to

$$\hat{C}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{x_i x_i^*}{\frac{1}{N} x_i^* \hat{C}_N^{-1}(\rho) x_i} + \rho I_N. \tag{1}$$

This estimator, proposed in [27], corresponds to a hybrid robust-shrinkage estimator reminding Tyler's M-estimator of scale [30] and Ledoit–Wolf's shrinkage estimator [22]. This estimator is particularly suited to scenarios where N/n is not small, for which other estimators are badly conditioned if not undefined. For this model, it is shown in [13] that $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$ where $\hat{S}_N(\rho)$ also follows a classical random matrix model.

The aforementioned approximations $\hat{S}_N$ of the estimators $\hat{C}_N$, the structure of which is well understood (as opposed to $\hat{C}_N$ which is only defined implicitly), allow for both a good apprehension of the limiting behavior of $\hat{C}_N$ and more importantly for a better usage of $\hat{C}_N$ as an appropriate substitute for sample covariance matrices in various estimation problems in the large N, n regime. The convergence in norm $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ is indeed sufficient in many cases to produce new consistent estimation methods based on $\hat{C}_N$ by simply replacing $\hat{C}_N$ by $\hat{S}_N$ in the problem defining equations. For example, the results of Couillet et al. [15] led to the introduction of novel consistent estimators based on functionals of $\hat{C}_N$ (of the Maronna type) for power and direction-of-arrival estimation in array processing in the presence of impulsive noise or rare outliers [11]. Similarly, in [13], empirical methods were designed to estimate the parameter ρ which minimizes the expected Frobenius norm $\operatorname{tr}[(\hat{C}_N(\rho) - C_N)^2]$, of interest for various outlier-prone applications dealing with non-small ratios N/n.[1]

Nonetheless, when replacing $\hat{C}_N$ for $\hat{S}_N$ in deriving consistent estimates, if the convergence $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ helps in producing novel consistent estimates, this convergence (which comes with no particular speed) is in general not sufficient to assess the performance of the estimator for large but finite N, n. Indeed, when second order results such as central limit theorems need to be established, say at rate $N^{-\frac{1}{2}}$, to proceed similarly to the replacement of $\hat{C}_N$ by $\hat{S}_N$ in the analysis, one would ideally demand that $\|\hat{C}_N - \hat{S}_N\| = o(N^{-\frac{1}{2}})$; but such a result, we believe, unfortunately does not hold. This constitutes a severe limitation in the exploitation of robust estimators as their performance as well as optimal fine-tuning often rely on second order performance. Concretely, for practical purposes in the array processing application of Couillet [11], one may naturally ask which choice of the u function is optimal to minimize the variance of (consistent) power and angle estimates. This question remains unanswered to this point for lack of better theoretical results.

The main purpose of the article is twofold. From a technical aspect, taking the robust shrinkage estimator $\hat{C}_N(\rho)$ defined by (1) as an example, we first show that, although the convergence $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$ (from [13, Theorem 1]) may not be extensible to a rate $O(N^{1-\varepsilon})$, one has the bilinear form convergence $N^{1-\varepsilon} a^*(\hat{C}_N^k(\rho) - \hat{S}_N^k(\rho))b \xrightarrow{\text{a.s.}} 0$ for each $\varepsilon > 0$, each $a, b \in \mathbb{C}^N$ of unit norm, and each $k \in \mathbb{Z}$. This result implies that, if $\sqrt{N} a^* \hat{S}_N^k(\rho) b$ satisfies a central limit theorem, then so does $\sqrt{N} a^* \hat{C}_N^k(\rho) b$ with the same limiting variance. This result is of fundamental importance to any statistical application based on such quadratic forms. Our second contribution is to exploit this result for the specific problem of signal detection in impulsive noise environments via the generalized likelihood-ratio test, particularly suited for radar signals detection under elliptical noise [10,27]. In this context, we determine the shrinkage parameter ρ which minimizes the probability of false detections and provide an empirical consistent estimate for this parameter, thus improving significantly over traditional sample covariance matrix-based estimators.

The remainder of the article introduces our main results in Section 2 which are proved in Section 3. Technical elements of proof are provided in the Appendix.

Notations: In the remainder of the article, we shall denote $\lambda_1(X), \ldots, \lambda_n(X)$ the real eigenvalues of $n \times n$ Hermitian matrices X. The norm notation $\|\cdot\|$ being considered is the spectral norm for matrices and Euclidean norm for vectors. The symbol $\imath$ is the complex $\sqrt{-1}$.

2. Main results

Let $N, n \in \mathbb{N}$, $c_N \triangleq N/n$, and $\rho \in (\max\{0, 1 - c_N^{-1}\}, 1]$. Let also $x_1, \ldots, x_n \in \mathbb{C}^N$ be n independent random vectors defined by the following assumptions.

[1] Other metrics may also be considered as in e.g. [31] with ρ chosen to minimize the return variance in a portfolio optimization problem.

Assumption 1 (*Data Vectors*). For $i \in \{1, \ldots, n\}$, $x_i = \sqrt{\tau_i} A_N w_i = \sqrt{\tau_i} z_i$, where

- $w_i \in \mathbb{C}^N$ is Gaussian with zero mean and covariance I_N, independent across i;
- $A_N A_N^* \triangleq C_N \in \mathbb{C}^{N\times N}$ is such that $\nu_N \triangleq \frac{1}{N}\sum_{i=1}^N \boldsymbol{\delta}_{\lambda_i(C_N)} \to \nu$ weakly, $\limsup_N \|C_N\| < \infty$, and $\frac{1}{N}\operatorname{tr} C_N = 1$;
- $\tau_i > 0$ are random or deterministic scalars.

Under Assumption 1, letting $\tau_i = \tilde{\tau}_i/\|w_i\|$ for some $\tilde{\tau}_i$ independent of w_i, x_i belongs to the class of elliptically distributed random vectors. Note that the normalization $\frac{1}{N}\operatorname{tr} C_N = 1$ is not a restricting constraint since the scalars τ_i may absorb any other normalization.

It has been well-established by the robust estimation theory that, even if the τ_i are independent, independent of the w_i, and that $\lim_n \frac{1}{n}\sum_{i=1}^n \tau_i = 1$ a.s., the sample covariance matrix $\frac{1}{n}\sum_{i=1}^n x_i x_i^*$ is in general a poor estimate for C_N. Robust estimators of scatter were designed for this purpose [24,30]. In addition, if N/n is non trivial, a linear shrinkage of these robust estimators against the identity matrix often helps in regularizing the estimator as established in e.g., [27,9]. The robust estimator of scatter considered in this work, which we denote $\hat{C}_N(\rho)$, is defined (originally in [27]) as the unique solution to

$$\hat{C}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i \hat{C}_N^{-1}(\rho) x_i} + \rho I_N.$$

2.1. Theoretical results

The asymptotic behavior of this estimator was studied recently in [13] in the regime where $N, n \to \infty$ in such a way that $c_N \to c \in (0,\infty)$. We first recall the important results of this article, which shall lay down the main concepts and notations of the present work. First define

$$\hat{S}_N(\rho) = \frac{1}{\gamma_N(\rho)}\frac{1-\rho}{1-(1-\rho)c_N}\frac{1}{n}\sum_{i=1}^n z_i z_i^* + \rho I_N$$

where $\gamma_N(\rho)$ is the unique solution to

$$1 = \int \frac{t}{\gamma_N(\rho)\rho + (1-\rho)t}\nu_N(dt).$$

For any $\kappa > 0$ small, define $\mathcal{R}_\kappa \triangleq [\kappa + \max\{0, 1-c^{-1}\}, 1]$. Then, from [13, Theorem 1], as $N, n \to \infty$ with $c_N \to c \in (0,\infty)$,

$$\sup_{\rho\in\mathcal{R}_\kappa} \left\| \hat{C}_N(\rho) - \hat{S}_N(\rho) \right\| \xrightarrow{\text{a.s.}} 0.$$

A careful analysis of the proof of Couillet and McKay [13, Theorem 1] (which is performed in Section 3) reveals that the above convergence can be refined as

$$\sup_{\rho\in\mathcal{R}_\kappa} N^{\frac{1}{2}-\varepsilon} \left\| \hat{C}_N(\rho) - \hat{S}_N(\rho) \right\| \xrightarrow{\text{a.s.}} 0 \tag{2}$$

for each $\varepsilon > 0$. This suggests that (well-behaved) functionals of $\hat{C}_N(\rho)$ fluctuating at a slower speed than $N^{-\frac{1}{2}+\varepsilon}$ for some $\varepsilon > 0$ follow the same statistics as the same functionals with $\hat{S}_N(\rho)$ in place of $\hat{C}_N(\rho)$. However, this result is quite weak as most limiting theorems (starting with the classical central limit theorems for independent scalar variables) deal with fluctuations of order $N^{-\frac{1}{2}}$ and sometimes in random matrix theory of order N^{-1}. In our opinion, the convergence speed (2) cannot be improved to a rate $N^{-\frac{1}{2}}$. Nonetheless, thanks to an averaging effect documented in Section 3, the fluctuation of special forms of functionals of $\hat{C}_N(\rho)$ can be proved to be much slower. Although among these functionals we could have considered linear functionals of the eigenvalue distribution of $\hat{C}_N(\rho)$, our present concern (driven by more obvious applications) is rather on bilinear forms of the type $a^*\hat{C}_N^k(\rho)b$ for some $a, b \in \mathbb{C}^N$ with $\|a\| = \|b\| = 1$, $k \in \mathbb{Z}$.

Our first main result is the following.

Theorem 1 (*Fluctuation of Bilinear Forms*). *Let $a, b \in \mathbb{C}^N$ with $\|a\| = \|b\| = 1$. Then, as $N, n \to \infty$ with $c_N \to c \in (0,\infty)$, for any $\varepsilon > 0$ and every $k \in \mathbb{Z}$,*

$$\sup_{\rho\in\mathcal{R}_\kappa} N^{1-\varepsilon} \left| a^*\hat{C}_N^k(\rho)b - a^*\hat{S}_N^k(\rho)b \right| \xrightarrow{\text{a.s.}} 0.$$

Some comments and remarks are in order. First, we recall that central limit theorems involving bilinear forms of the type $a^*\hat{S}_N^k(\rho)b$ are classical objects in random matrix theory (see e.g. [21,25] for $k = -1$), particularly common in signal processing and wireless communications. These central limit theorems in general show fluctuations at speed $N^{-\frac{1}{2}}$. This indicates, taking $\varepsilon < \frac{1}{2}$ in Theorem 1 and using the fact that almost sure convergence implies weak convergence, that $a^*\hat{C}_N^k(\rho)b$ exhibits the same fluctuations as $a^*\hat{S}_N^k(\rho)b$, the latter being classical and tractable while the former is quite intricate at the onset, due to the implicit nature of $\hat{C}_N(\rho)$.

Of practical interest to many applications in signal processing is the case where $k = -1$. In the next section, we present a classical generalized maximum likelihood signal detection in impulsive noise, for which we shall characterize the shrinkage parameter ρ that meets minimum false alarm rates.

2.2. Application to signal detection

In this section, we consider the hypothesis testing scenario by which an N-sensor array receives a vector $y \in \mathbb{C}^N$ according to the following hypotheses

$$y = \begin{cases} x, & \mathcal{H}_0 \\ \alpha p + x, & \mathcal{H}_1 \end{cases}$$

in which $\alpha > 0$ is some unknown scaling factor constant while $p \in \mathbb{C}^N$ is deterministic and known at the sensor array (which often corresponds to a steering vector arising from a specific known angle), and x is an impulsive noise distributed as x_1 according to Assumption 1. For convenience, we shall take $\|p\| = 1$.

Under $\mathcal{H}_0$ (the null hypothesis), a noisy observation from an impulsive source is observed while under $\mathcal{H}_1$ both information and noise are collected at the array. The objective is to decide on $\mathcal{H}_1$ versus $\mathcal{H}_0$ upon the observation y and prior pure-noise observations $x_1, \ldots, x_n$ distributed according to Assumption 1. When $\tau_1, \ldots, \tau_n$ and C_N are unknown, the corresponding generalized likelihood ratio test, derived in [10], reads

$$T_N(\rho) \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} \Gamma$$

for some detection threshold Γ where

$$T_N(\rho) \triangleq \frac{|y^*\hat{C}_N^{-1}(\rho)p|}{\sqrt{y^*\hat{C}_N^{-1}(\rho)y}\sqrt{p^*\hat{C}_N^{-1}(\rho)p}}.$$

More precisely, Conte et al. [10] derived the detector $T_N(0)$ only valid when $n \geq N$. The relaxed detector $T_N(\rho)$ allows for a better conditioning of the estimator, in particular for $n \simeq N$. In [27], $T_N(\rho)$ is used explicitly in a space–time adaptive processing setting but only simulation results were provided. Alternative metrics for similar array processing problems involve the signal-to-noise ratio loss minimization rather than likelihood ratio tests; in [1,4], the authors exploit the estimators $\hat{C}_N(\rho)$ but restrict themselves to the less tractable finite dimensional analysis.

Our objective is to characterize the false alarm performance of the detector. That is, provided $\mathcal{H}_0$ is the actual scenario (i.e. $y = x$), we shall evaluate $P(T_N(\rho) > \Gamma)$. Since it shall appear that, under $\mathcal{H}_0$, $T_N(\rho) \xrightarrow{\text{a.s.}} 0$ for every fixed $\Gamma > 0$ and every ρ, by dominated convergence $P(T_N(\rho) > \Gamma) \to 0$ which does not say much about the actual test performance for large but finite N, n. To avoid such empty statements, we shall then consider the non-trivial case where $\Gamma = N^{-\frac{1}{2}}\gamma$ for some fixed $\gamma > 0$. In this case our objective is to characterize the false alarm probability

$$P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}}\right).$$

Before providing this result, we need some further reminders from [13]. First define

$$\underline{\hat{S}}_N(\rho) \triangleq (1-\rho)\frac{1}{n}\sum_{i=1}^{n} z_i z_i^* + \rho I_N.$$

Then, from [13, Lemma 1], for each $\rho \in (\max\{0, 1-c^{-1}\}, 1]$,

$$\frac{\hat{S}_N(\rho)}{\rho + \frac{1}{\gamma_N(\rho)}\frac{1-\rho}{1-(1-\rho)c}} = \underline{\hat{S}}_N(\underline{\rho})$$

where

$$\underline{\rho} \triangleq \frac{\rho}{\rho + \frac{1}{\gamma_N(\rho)}\frac{1-\rho}{1-(1-\rho)c}}.$$

Moreover, the mapping $\rho \mapsto \underline{\rho}$ is continuously increasing from $(\max\{0, 1-c^{-1}\}, 1]$ onto $(0, 1]$.

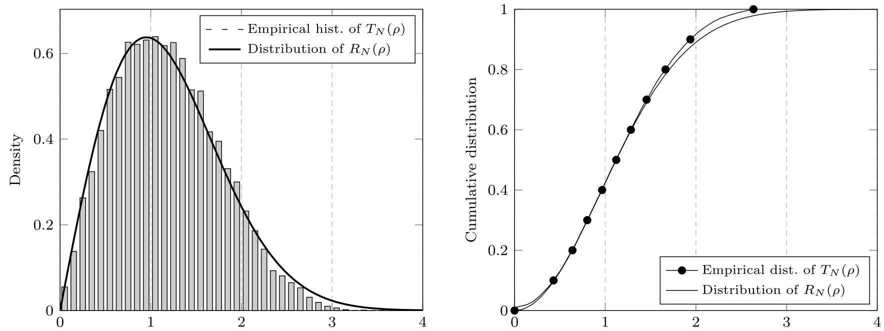

Fig. 1. Histogram distribution function of the $\sqrt{N}T_N(\rho)$ versus $R_N(\rho)$, $N = 20$, $p = N^{-\frac{1}{2}}[1, \dots, 1]^\mathsf{T}$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$, $\rho = 0.2$.

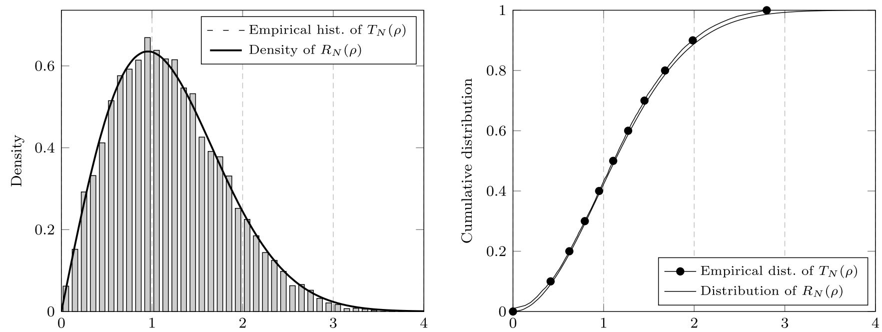

Fig. 2. Histogram distribution function of the $\sqrt{N}T_N(\rho)$ versus $R_N(\rho)$, $N = 100$, $p = N^{-\frac{1}{2}}[1, \dots, 1]^\mathsf{T}$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$, $\rho = 0.2$.

From classical random matrix considerations (see e.g. [28]), letting $Z = [z_1, \dots, z_n] \in \mathbb{C}^{N\times n}$, the empirical spectral distribution[2] of $(1-\rho)\frac{1}{n}Z^*Z$ almost surely admits a weak limit μ. The Stieltjes transform $m(z) \triangleq \int (t-z)^{-1}\mu(dt)$ of μ at $z \in \mathbb{C} \setminus \mathrm{Supp}(\mu)$ is the unique complex solution with positive (resp. negative) imaginary part if $\Im[z] > 0$ (resp. $\Im[z] < 0$) and unique real positive solution if $\Im[z] = 0$ and $\Re[z] < 0$ to

$$m(z) = \left(-z + c\int \frac{(1-\rho)t}{1+(1-\rho)tm(z)}\nu(dt)\right)^{-1}.$$

We denote $m'(z)$ the derivative of $m(z)$ with respect to z (recall that the Stieltjes transform of a positively supported measure is analytic, hence continuously differentiable, away from the support of the measure).

With these definitions in place and with the help of Theorem 1, we are now ready to introduce the main result of this section.

Theorem 2 (*Asymptotic Detector Performance*)**.** *Under hypothesis $\mathcal{H}_0$, as $N, n \to \infty$ with $c_N \to c \in (0, \infty)$,*

$$\sup_{\rho\in\mathcal{R}_\kappa} \left| P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}}\right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\rho)}\right)\right| \to 0$$

where

$$\sigma_N^2(\rho) \triangleq \frac{1}{2}\frac{p^*C_NQ_N^2(\underline{\rho})p}{p^*Q_N(\underline{\rho})p\cdot\frac{1}{N}\operatorname{tr}C_NQ_N(\underline{\rho})\cdot\left(1-c(1-\underline{\rho})^2m(-\underline{\rho})^2\frac{1}{N}\operatorname{tr}C_N^2Q_N^2(\underline{\rho})\right)}$$

with $\rho \mapsto \underline{\rho}$ the aforementioned mapping and $Q_N(\underline{\rho}) \triangleq (I_N + (1-\underline{\rho})m(-\underline{\rho})C_N)^{-1}$.

Otherwise stated, $\sqrt{N}T_N(\rho)$ is uniformly well approximated by a Rayleigh distributed random variable $R_N(\rho)$ with parameter $\sigma_N(\rho)$. Simulation results are provided in Figs. 1 and 2 which corroborate the results of Theorem 2 for $N = 20$ and $N = 100$, respectively (for a single value of ρ though). Comparatively, it is observed, as one would expect, that larger values for N induce improved approximations in the tails of the approximating distribution.

The result of Theorem 2 provides an analytical characterization of the performance of the GLRT for each ρ which suggests in particular the existence of values for ρ which minimize the false alarm probability for given γ. Note in passing that,

2 That is the normalized counting measure of the eigenvalues.

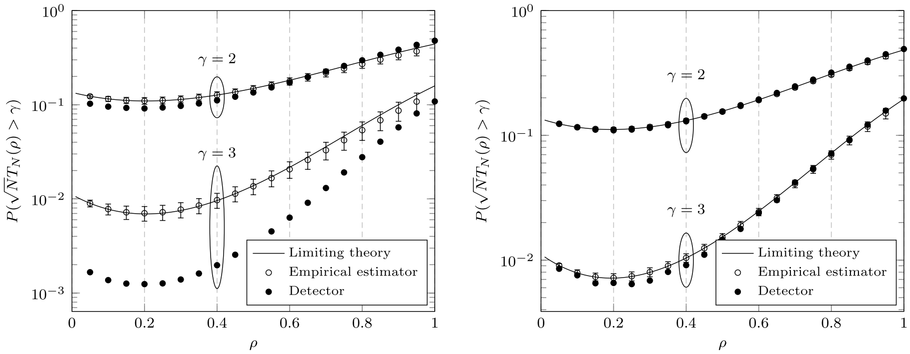

Fig. 3. False alarm rate $P(\sqrt{N}T_N(\rho) > \gamma)$, $N = 20$ (left), $N = 100$ (right), $p = N^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$.

independently of γ, minimizing the false alarm rate is asymptotically equivalent to minimizing $\sigma_N^2(\rho)$ over ρ. However, the expression of $\sigma_N^2(\rho)$ depends on the covariance matrix C_N which is unknown to the array and therefore does not allow for an immediate online choice of an appropriate ρ. To tackle this problem, the following proposition provides a consistent estimate for $\sigma_N^2(\rho)$ based on $\hat{C}_N(\rho)$ and p.

Proposition 1 (*Empirical Performance Estimation*). *For $\rho \in (\max\{0, 1 - c_N^{-1}\}, 1)$, define*

$$\hat{\sigma}_N^2(\rho) \triangleq \frac{1}{2} \frac{1 - \rho \frac{p^* \hat{C}_N^{-2}(\rho) p}{p^* \hat{C}_N^{-1}(\rho) p}}{(1 - c_N + c_N \rho)(1 - \rho)}.$$

Also let $\hat{\sigma}_N^2(1) \triangleq \lim_{\rho \uparrow 1} \hat{\sigma}_N^2(\rho) < \infty$ a.s. Then we have

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| \sigma_N^2(\rho) - \hat{\sigma}_N^2(\rho) \right| \xrightarrow{\text{a.s.}} 0.$$

Since both the estimation of $\sigma_N^2(\rho)$ in Proposition 1 and the convergence in Theorem 2 are uniform over $\rho \in \mathcal{R}_\kappa$, we have the following result.

Corollary 1 (*Empirical Performance Optimum*). *Let $\hat{\sigma}_N^2(\rho)$ be defined as in Proposition 1 and define $\hat{\rho}_N^*$ as any value satisfying*

$$\hat{\rho}_N^* \in \operatorname*{argmin}_{\rho \in \mathcal{R}_\kappa} \left\{ \hat{\sigma}_N^2(\rho) \right\}$$

(this set being in general a singleton). Then, for every $\gamma > 0$,

$$P\left(\sqrt{N}T_N(\hat{\rho}_N^*) > \gamma\right) - \inf_{\rho \in \mathcal{R}_\kappa} \left\{ P\left(\sqrt{N}T_N(\rho) > \gamma\right) \right\} \to 0.$$

This last result states that, for N, n sufficiently large, it is increasingly close-to-optimal to use the detector $T_N(\hat{\rho}_N^*)$ in order to reach minimal false alarm probability. A practical graphical confirmation of this fact is provided in Fig. 3 where, in the same scenario as in Figs. 1–2, the false alarm rates for various values of γ are depicted. In this figure, the black dots correspond to the actual values taken by $P(\sqrt{N}T_N(\rho) > \gamma)$ empirically obtained out of 10^6 Monte Carlo simulations. The plain curves are the approximating values $\exp(-\gamma^2/(2\sigma_N(\rho)^2))$. Finally, the white dots with error bars correspond to the mean and standard deviations of $\exp(-\gamma^2/(2\hat{\sigma}_N(\rho)^2))$ for each ρ, respectively. It is first interesting to note that the estimates $\hat{\sigma}_N(\rho)$ are quite accurate, especially so for N large, with standard deviations sufficiently small to provide good estimates, already for small N, of the false alarm minimizing ρ. However, similar to Figs. 1–2, we observe a particularly weak approximation in the (small) $N = 20$ setting for large values of γ, corresponding to tail events, while for $N = 100$, these values are better recovered. This behavior is obviously explained by the fact that $\gamma = 3$ is not small compared to $\sqrt{N}$ when $N = 20$.

Nonetheless, from an error rate viewpoint, it is observed that errors of order 10^{-2} are rather well approximated for $N = 100$. In Fig. 4, we consider this observation in depth by displaying $P(T_N(\hat{\rho}_N^*) > \Gamma)$ and its approximation $\min_\rho \exp(-N\Gamma^2/(2\sigma_N^2(\rho)))$ for $N = 20$ and $N = 100$, for various values of Γ. This figure shows that even errors of order 10^{-4} are well approximated for large N, while only errors of order 10^{-2} can be evaluated for small N. In this figure, we also consider the case where $[C_N]_{ij} = 0.2^{|i-j|}$, hence "closer" to the identity matrix, and observe better convergence properties.

In Fig. 5, we finally confront our proposed method against existing schemes, in particular based on different plug-in estimators of the sample covariance matrix. Our main comparison is made against the sample covariance matrix (SCM) estimator which estimates C_N as $\frac{1}{n}\sum_{i=1}^n x_i x_i^*$ (we thus maintain $N/n < 1$ to ensure the existence of the inverse). The

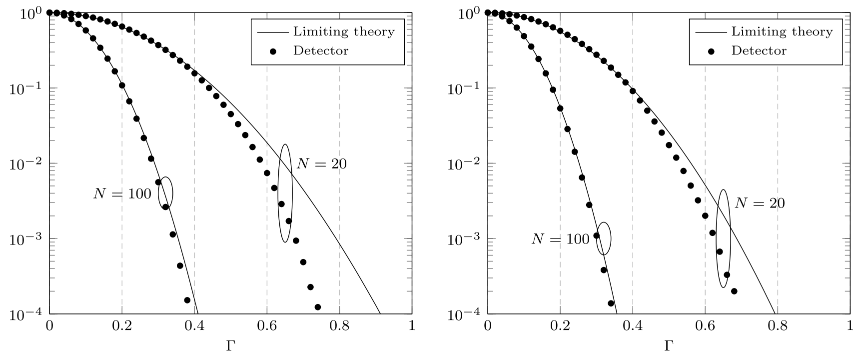

Fig. 4. False alarm rate $P(T_N(\hat{\rho}_N^*) > \Gamma)$ for $N = 20$ and $N = 100$, $p = N^{-\frac{1}{2}}[1, \dots, 1]^\mathsf{T}$, $c_N = 1/2$, and $[C_N]_{ij} = 0.7^{|i-j|}$ (left) or $[C_N]_{ij} = 0.2^{|i-j|}$ (right).

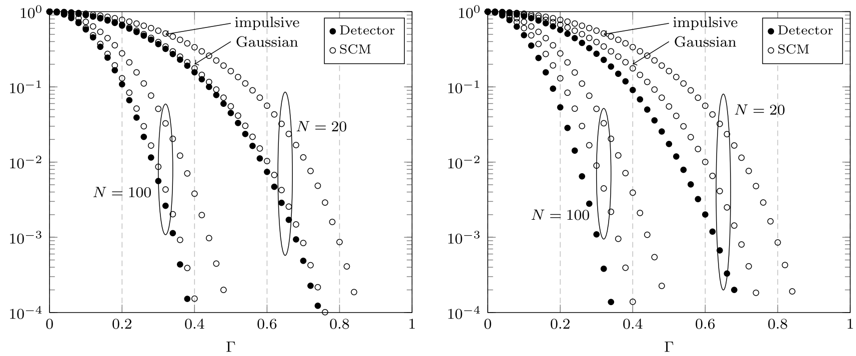

Fig. 5. False alarm performance comparison versus traditional methods, for $N = 20$ and $N = 100$, $p = N^{-\frac{1}{2}}[1, \dots, 1]^\mathsf{T}$, $c_N = 1/2$, and $[C_N]_{ij} = 0.7^{|i-j|}$ (left) or $[C_N]_{ij} = 0.2^{|i-j|}$ (left). Noise taken Gaussian (inner "○" plots) or with Gamma-0.5 impulsions (outer "○" plots). The proposed detector is insensitive to noise impulsions.

performance of the latter is assessed both on Gaussian and elliptical (impulsive) data, for $[C_N]_{ij} = 0.7^{|i-j|}$ (left figure) or $[C_N]_{ij} = 0.2^{|i-j|}$ (right figure). It is seen, as one would expect, that the non-robust SCM suffers from noise impulsiveness. In comparison, the proposed detector is by definition insensitive to this type of noise and thus remains invariant. Note also that the Tyler estimator, which is our proposed estimator, only for $\rho = 0$, performs (in both impulsive and non-impulsive scenarios) almost identically to the Gaussian-noise SCM, which is also theoretically expected; it is thus not displayed explicitly to improve readability. Finally note that, as C_N is closer to the identity matrix, our proposed scheme presents quite large performance gains compared to the non-regularized methods. This is again in line with intuition.

3. Proof

In this section, we shall successively prove Theorems 1, 2, Proposition 1, and Corollary 1. Of utmost interest is the proof of Theorem 1 which shall be the concern of most of the section and of Appendix for the proof of a key lemma.

Before delving into the core of the proofs, let us introduce a few notations that shall be used throughout the section. First recall from [13] that we can write, for each $\rho \in (\max\{0, 1 - c_N^{-1}\}, 1]$,

$$\hat{C}_N(\rho) = \frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \frac{z_i z_i^*}{\frac{1}{N} z_i^* \hat{C}_{(i)}^{-1}(\rho) z_i} + \rho I_N$$

where $\hat{C}_{(i)}(\rho) = \hat{C}_N(\rho) - (1-\rho)\frac{1}{n}\frac{z_i z_i^*}{\frac{1}{N} z_i^* \hat{C}_N^{-1}(\rho) z_i}$.

Now, we define

$$\alpha(\rho) = \frac{1-\rho}{1-(1-\rho)c_N}$$

$$d_i(\rho) = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1}(\rho) z_i = \frac{1}{N} z_i^* \left(\alpha(\rho) \frac{1}{n} \sum_{j \neq i} \frac{z_j z_j^*}{d_j(\rho)} + \rho I_N \right)^{-1} z_i$$

$$\tilde{d}_i(\rho) = \frac{1}{N} z_i^* \hat{S}_{(i)}^{-1}(\rho) z_i = \frac{1}{N} z_i^* \left(\alpha(\rho) \frac{1}{n} \sum_{j \neq i}^n \frac{z_j z_j^*}{\gamma_N(\rho)} + \rho I_N \right)^{-1} z_i.$$

Clearly by uniqueness of $\hat{C}_N$ and by the relation to $\hat{C}_{(i)}$ above, $d_1(\rho), \ldots, d_n(\rho)$ are uniquely defined by their n implicit equations. We shall also discard the parameter ρ for readability whenever not needed.

3.1. Bilinear form equivalence

In this section, we prove Theorem 1. As shall become clear, the proof unfolds similarly for each $k \in \mathbb{Z} \setminus \{0\}$ and we can therefore restrict ourselves to a single value for k. As Theorem 2 relies on $k = -1$, for consistency, we take $k = -1$ from now on. Thus, our objective is to prove that, for $a, b \in \mathbb{C}^N$ with $\|a\| = \|b\| = 1$, and for any $\varepsilon > 0$,

$$\sup_{\rho \in \mathcal{R}_\kappa} N^{1-\varepsilon} \left| a^* \hat{C}_N^{-1}(\rho) b - a^* \hat{S}_N^{-1}(\rho) b \right| \xrightarrow{\text{a.s.}} 0.$$

For this, forgetting for some time the index ρ, first write

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{\alpha}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b \tag{3}$$

$$= \frac{\alpha}{n} \sum_{i=1}^n a^* \hat{C}_N^{-1} z_i \frac{d_i - \gamma_N}{\gamma_N d_i} z_i^* \hat{S}_N^{-1} b. \tag{4}$$

In [13], where it is shown that $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ (that is the spectral norm of the inner parenthesis in (3) vanishes), the core of the proof was to show that $\max_{1 \le i \le n} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$ which, along with the convergence of γ_N away from zero and the almost sure boundedness of $\| \frac{1}{n} \sum_{i=1}^n z_i z_i^* \|$ for all large N (from e.g. [2]), gives the result. A thorough inspection of the proof in [13] reveals that $\max_{1 \le i \le n} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$ may be improved into $\max_{1 \le i \le n} N^{\frac{1}{2} - \varepsilon} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$ for any $\varepsilon > 0$ but that this speed cannot be further improved beyond $N^{\frac{1}{2}}$. The latter statement is rather intuitive since γ_N is essentially a sharp deterministic approximation for $\frac{1}{N} \operatorname{tr} \hat{C}_N^{-1}$ while d_i is a quadratic form on $\hat{C}_{(i)}^{-1}$; classical random matrix results involving fluctuations of such quadratic forms, see e.g. [21], indeed show that these fluctuations are of order $N^{-\frac{1}{2}}$. As a consequence, $\max_{1 \le i \le n} N^{1-\varepsilon} |d_i - \gamma_N|$ and thus $N^{1-\varepsilon} \|\hat{C}_N - \hat{S}_N\|$ are not expected to vanish for small ε.

This being said, when it comes to bilinear forms, for which we shall naturally have $N^{\frac{1}{2}-\varepsilon} |a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b| \xrightarrow{\text{a.s.}} 0$, seeing the difference in absolute values as the n-term average (4), one may expect that the fluctuations of $d_i - \gamma_N$ are sufficiently loosely dependent across i to further increase the speed of convergence from $N^{\frac{1}{2}-\varepsilon}$ to $N^{1-\varepsilon}$ (which is the best one could expect from a law of large numbers aspect if the $d_i - \gamma_N$ were truly independent). It turns out that this intuition is correct.

Nonetheless, to proceed with the proof, it shall be quite involved to work directly with (4) which involves the rather intractable terms d_i (as the random solutions to an implicit equation). As in [13], our approach will consist in first approximating d_i by a much more tractable quantity. Letting γ_N be this approximation is however not good enough this time since $\gamma_N - d_i$ is a non-obvious quantity of amplitude $O(N^{-\frac{1}{2}})$ which, due to intractability, we shall not be able to average across i into a $O(N^{-1})$ quantity. Thus, we need a refined approximation of d_i which we shall take to be $\tilde{d}_i$ defined above. Intuitively, since $\tilde{d}_i$ is also a quadratic form closely related to d_i, we expect $d_i - \tilde{d}_i$ to be of order $O(N^{-1})$, which we shall indeed observe. With this approximation in place, d_i can be replaced by $\tilde{d}_i$ in (4), which now becomes a more tractable random variable (as it involves no implicit equation) that fluctuates around γ_N at the expected $O(N^{-1})$ speed.

Let us then introduce the variable $\tilde{d}_i$ in (3) to obtain

$$\begin{aligned} a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b &= a^* \hat{C}_N^{-1} \left(\frac{\alpha}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b + a^* \hat{C}_N^{-1} \left(\frac{\alpha}{n} \sum_{i=1}^n \left[\frac{1}{\tilde{d}_i} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b \\ &\triangleq \xi_1 + \xi_2. \end{aligned}$$

We will now show that $\xi_1 = \xi_1(\rho)$ and $\xi_2 = \xi_2(\rho)$ vanish at the appropriate speed and uniformly so on $\mathcal{R}_\kappa$.

Let us first progress in the derivation of $\xi_1(\rho)$ from which we wish to discard the explicit dependence on $\hat{C}_N$. We have

$$\begin{aligned} \xi_1 &= a^* \hat{C}_N^{-1} \left(\frac{\alpha}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b \\ &= a^* \hat{S}_N^{-1} \left(\frac{\alpha}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b + a^* (\hat{C}_N^{-1} - \hat{S}_N^{-1}) \left(\frac{\alpha}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b \\ &= a^* \hat{S}_N^{-1} \left(\frac{\alpha}{n} \sum_{i=1}^n \frac{\tilde{d}_i - \gamma_N}{\gamma_N^2} z_i z_i^* \right) \hat{S}_N^{-1} b - a^* \hat{S}_N^{-1} \left(\frac{\alpha}{n} \sum_{i=1}^n \frac{(\tilde{d}_i - \gamma_N)^2}{\gamma_N^2 \tilde{d}_i} z_i z_i^* \right) \hat{S}_N^{-1} b \end{aligned}$$

$$
\begin{aligned}
&+ a^*(\hat{C}_N^{-1} - \hat{S}_N^{-1}) \left(\frac{\alpha}{n} \sum_{i=1}^{n} \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b \\
&\triangleq \xi_{11} + \xi_{12} + \xi_{13}.
\end{aligned}
$$

The terms ξ_{12} and ξ_{13} exhibit products of two terms that are expected to be of order $O(N^{-\frac{1}{2}})$ and which are thus easily handled. As for ξ_{11}, it no longer depends on $\hat{C}_N$ and is therefore a standard random variable which, although involved, is technically tractable via standard random matrix methods. In order to show that $N^{1-\varepsilon} \max\{|\xi_{12}|, |\xi_{13}|\} \xrightarrow{\text{a.s.}} 0$ uniformly in ρ, we use the following lemma.

Lemma 1. *For any $\varepsilon > 0$,*

$$
\max_{1 \le i \le n} \sup_{\rho \in \mathcal{R}_\kappa} N^{\frac{1}{2}-\varepsilon} |\tilde{d}_i(\rho) - \gamma_N(\rho)| \xrightarrow{\text{a.s.}} 0
$$

$$
\max_{1 \le i \le n} \sup_{\rho \in \mathcal{R}_\kappa} N^{\frac{1}{2}-\varepsilon} |d_i(\rho) - \gamma_N(\rho)| \xrightarrow{\text{a.s.}} 0.
$$

Note that, while the first result is a standard, easily established, random matrix result, the second result is the aforementioned refinement of the core result in the proof of Couillet and McKay [13, Theorem 1].

Proof of Lemma 1. We start by proving the first identity. From [13, p. 17] (taking $w = -\gamma_N \rho \alpha^{-1}$), we have, for each $p \ge 2$ and for each $1 \le k \le n$,

$$
\mathrm{E}\left[\left| \tilde{d}_k(\rho) - \gamma_N(\rho) \right|^p \right] = O\left(N^{-\frac{p}{2}} \right)
$$

where the bound does not depend on $\rho > \max\{0, 1 - 1/c\} + \kappa$. Let now $\max\{0, 1 - 1/c\} + \kappa = \rho_0 < \cdots < \rho_{\lceil \sqrt{n} \rceil} = 1$ be a regular sampling of $\mathcal{R}_\kappa$ in $\lceil \sqrt{n} \rceil$ intervals. We then have, from Markov inequality and the union bound on $n(\lceil \sqrt{n} \rceil + 1)$ events, for $C > 0$ given,

$$
P\left(\max_{1 \le k \le n, 0 \le i \le \lceil \sqrt{n} \rceil} \left| N^{\frac{1}{2}-\varepsilon} (\tilde{d}_k(\rho_i) - \gamma_N(\rho_i)) \right| > C \right) \le K N^{-p\varepsilon + \frac{3}{2}}
$$

for some $K > 0$ only dependent on p and C. From the Borel Cantelli lemma, we then have $\max_{k,i} |N^{\frac{1}{2}-\varepsilon}(\tilde{d}_k(\rho_i) - \gamma_N(\rho_i))| \xrightarrow{\text{a.s.}} 0$ as long as $-p\varepsilon + 3/2 < -1$, which is obtained for $p > 5/(2\varepsilon)$. Using $|\gamma_N(\rho) - \gamma_N(\rho')| \le K|\rho - \rho'|$ for some constant K and each $\rho, \rho' \in \mathcal{R}_\kappa$ (see [13, top of Section 5.1]) and similarly $\max_{1 \le k \le n} |\tilde{d}_k(\rho) - \tilde{d}_k(\rho')| \le K|\rho - \rho'|$ for all large n a.s. (obtained by explicitly writing the difference and using the fact that $\|z_k\|^2/N$ is asymptotically bounded almost surely), we get

$$
\begin{aligned}
\max_{1 \le k \le n} \sup_{\rho \in \mathcal{R}_\kappa} N^{\frac{1}{2}-\varepsilon} |\tilde{d}_k(\rho) - \gamma_N(\rho)| &\le \max_{k,i} N^{\frac{1}{2}-\varepsilon} |\tilde{d}_k(\rho_i) - \gamma_N(\rho_i)| + K N^{-\varepsilon} \\
&\xrightarrow{\text{a.s.}} 0.
\end{aligned}
$$

The second result relies on revisiting the proof of Couillet and McKay [13, Theorem 1] incorporating the convergence speed on $\tilde{d}_k - \gamma_N$. For convenience and compatibility with similar derivations that appear later in the proof, we slightly modify the original proof of Couillet and McKay [13, Theorem 1]. We first define $f_i(\rho) = d_i(\rho)/\gamma_N(\rho)$ and relabel the $d_i(\rho)$ in such a way that $f_1(\rho) \le \cdots \le f_n(\rho)$ (the ordering may then depend on ρ). Then, we have by definition of $d_n(\rho) = \gamma_N(\rho) f_n(\rho)$

$$
\begin{aligned}
\gamma_N(\rho) f_n(\rho) &= \frac{1}{N} z_n^* \left(\alpha(\rho) \frac{1}{n} \sum_{i<n} \frac{z_i z_i^*}{\gamma_N(\rho) f_i(\rho)} + \rho I_N \right)^{-1} z_n \\
&\le \frac{1}{N} z_n^* \left(\alpha(\rho) \frac{1}{f_n(\rho)} \frac{1}{n} \sum_{i<n} \frac{z_i z_i^*}{\gamma_N(\rho)} + \rho I_N \right)^{-1} z_n
\end{aligned}
$$

where we used $f_n(\rho) \ge f_i(\rho)$ for each i. The above is now equivalent to

$$
\gamma_N(\rho) \le \frac{1}{N} z_n^* \left(\alpha(\rho) \frac{1}{n} \sum_{i<n} \frac{z_i z_i^*}{\gamma_N(\rho)} + f_n(\rho) \rho I_N \right)^{-1} z_n.
$$

We now make the assumption that there exist $\eta > 0$ and a sequence $\{\rho^{(n)}\} \in \mathcal{R}_\kappa$ such that $f_n(\rho^{(n)}) > 1 + N^{\eta - \frac{1}{2}}$ infinitely often, which is equivalent to saying $d_n(\rho^{(n)}) > \gamma_N(\rho^{(n)})(1 + N^{\eta - \frac{1}{2}})$ infinitely often (i.o.). Then, from these assumptions and the above first convergence result

$$\begin{aligned}\gamma_N(\rho^{(n)}) &\le \frac{1}{N} z_n^* \left(\alpha(\rho^{(n)}) \frac{1}{n} \sum_{i<n} \frac{z_i z_i^*}{\gamma_N(\rho^{(n)})} + \rho^{(n)} \left(1 + N^{\eta-\frac{1}{2}}\right) I_N \right)^{-1} z_n \\ &= \tilde{d}_n(\rho^{(n)}) - N^{\eta-\frac{1}{2}} \frac{1}{N} z_n^* \left(\frac{1}{n} \sum_{i<n} \frac{\alpha(\rho^{(n)}) z_i z_i^*}{\rho^{(n)} \gamma_N(\rho^{(n)})} + \left(1 + N^{\eta-\frac{1}{2}}\right) I_N \right)^{-1} \\ &\quad \times \left(\frac{1}{n} \sum_{i<n} \frac{\alpha(\rho^{(n)}) z_i z_i^*}{\gamma_N(\rho^{(n)})} + \rho^{(n)} I_N \right)^{-1} z_n. \end{aligned} \tag{5}$$

Now, by the first result of the lemma, letting $0 < \varepsilon < \eta$, we have

$$\left| \tilde{d}_n(\rho^{(n)}) - \gamma_N(\rho^{(n)}) \right| \le \max_{\rho \in \mathcal{R}_\kappa} \left| \tilde{d}_n(\rho) - \gamma_N(\rho) \right| \le N^{\varepsilon - \frac{1}{2}}$$

for all large n a.s., so that, for these large n, $\tilde{d}_n(\rho^{(n)}) \le \gamma_N(\rho^{(n)}) + N^{\varepsilon-\frac{1}{2}}$. Applying this inequality to the first right-end side term of (5) and using the almost sure boundedness of the rightmost right-end side term entails

$$0 \le N^{\varepsilon-\frac{1}{2}} - K N^{\eta - \frac{1}{2}}$$

for some $K > 0$ for all large n a.s. But, $N^{\varepsilon/2-1/2} - KN^{\eta/2-1/2} < 0$ for all large N, which contradicts the inequality. Thus, our initial assumption is wrong and therefore, for each $\eta > 0$, we have for all large n a.s., $d_n(\rho) < \gamma_N(\rho) + N^{\eta-\frac{1}{2}}$ uniformly on $\rho \in \mathcal{R}_\kappa$. The same calculus can be performed for $d_1(\rho)$ by assuming that $f_1(\rho'^{(n)}) < 1 - N^{\eta-\frac{1}{2}}$ i.o. over some sequence $\rho'^{(n)}$; by reverting all inequalities in the derivation above, we similarly conclude by contradiction that $d_1(\rho) > \gamma_N(\rho) - N^{\eta-\frac{1}{2}}$ for all large n, uniformly so in $\mathcal{R}_\kappa$. Together, both results finally lead, for each $\varepsilon > 0$, to

$$\max_{1 \le k \le n} \sup_{\rho \in \mathcal{R}_\kappa} \left| N^{\frac{1}{2}-\varepsilon} \left(d_k(\rho) - \gamma_N(\rho) \right) \right| \xrightarrow{\text{a.s.}} 0$$

obtained by fixing ε, taking η such that $0 < \eta < \varepsilon$, and using $\max_k \sup_\rho |d_k(\rho) - \gamma_N(\rho)| < N^{\eta-\frac{1}{2}}$ for all large n a.s.

Thanks to Lemma 1, expressing $\hat{C}_N^{-1}(\rho) - \hat{S}_N^{-1}(\rho)$ as a function of $d_i(\rho) - \gamma_N(\rho)$ and using the (almost sure) boundedness of the various terms involved, we finally get $N^{1-\varepsilon}\xi_{12} \xrightarrow{\text{a.s.}} 0$ and $N^{1-\varepsilon}\xi_{13} \xrightarrow{\text{a.s.}} 0$ uniformly on ρ.

It then remains to handle the more delicate term ξ_{11}, which can be further expressed as

$$\begin{aligned}\xi_{11} &= \frac{\alpha}{\gamma_N^2} a^* \hat{S}_N^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\tilde{d}_i - \gamma_N) z_i z_i^* \right) \hat{S}_N^{-1} b \\ &= \frac{\alpha}{\gamma_N^2} \frac{1}{n} \sum_{i=1}^n a^* \hat{S}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left(\tilde{d}_i - \gamma_N \right).\end{aligned}$$

For that, we will resort to the following lemma, whose proof is postponed to Appendix.

Lemma 2. *Let c and d be random or deterministic vectors, independent of $z_1, \ldots, z_n$, such that $\max\left(\mathrm{E}[\|c\|^k], \mathrm{E}[\|d\|^k]\right) \le K$ for some $K > 0$ and all integers k. Then, for each integer p,*

$$\mathrm{E}\left[\left| \frac{1}{n} \sum_{i=1}^n c^* \hat{S}_N^{-1} z_i z_i^* \hat{S}_N^{-1} d \left(\frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i - \gamma_N(\rho) \right) \right|^{2p} \right] = O\left(N^{-2p}\right).$$

By the Markov inequality and the union bound, similar to the proof of Lemma 1, we get from Lemma 2 (with $a = c$ and $d = b$) that, for each $\eta > 0$ and for each integer $p \ge 1$,

$$P\left(\sup_{\rho \in \{\rho_0 < \cdots < \rho_{\lceil \sqrt{n} \rceil}\}} N^{1-\varepsilon} |\xi_{11}| > \eta \right) \le K N^{-p\varepsilon + \frac{1}{2}}$$

with K only function of η and $\rho_0 < \cdots < \rho_{\lceil \sqrt{n} \rceil}$ a regular sampling of $\mathcal{R}_\kappa$. Taking $p > 3/(2\varepsilon)$, we finally get from the Borel Cantelli lemma that

$$N^{1-\varepsilon} \xi_{11} \xrightarrow{\text{a.s.}} 0$$

uniformly on $\{\rho_0, \dots, \rho_{\lceil\sqrt{n}\rceil}\}$ and finally, using Lipschitz arguments as in the proof of Lemma 1, uniformly on $\mathcal{R}_\kappa$. Putting all results together, we finally have

$$\sup_{\rho\in\mathcal{R}_\kappa} N^{1-\varepsilon}|\xi_1(\rho)| \xrightarrow{\text{a.s.}} 0$$

which concludes the first part of the proof.

We now continue with $\xi_2(\rho)$. In order to prove $N^{1-\varepsilon}\xi_2(\rho) \xrightarrow{\text{a.s.}} 0$ uniformly on $\rho \in \mathcal{R}_\kappa$, it is sufficient (thanks to the boundedness of the various terms involved) to prove that

$$\max_{1\le i\le n}\sup_{\rho\in\mathcal{R}_\kappa}\left|N^{1-\varepsilon}\left(\tilde{d}_i(\rho) - d_i(\rho)\right)\right| \xrightarrow{\text{a.s.}} 0.$$

To obtain this result, we first need the following fundamental proposition.

Proposition 2. *For any $\varepsilon > 0$,*

$$\max_{1\le k\le n}\sup_{\rho\in\mathcal{R}_\kappa}\left|N^{1-\varepsilon}\left(\tilde{d}_k(\rho) - \frac{1}{N}z_k^*\left(\alpha(\rho)\frac{1}{n}\sum_{i\neq k}\frac{z_iz_i^*}{\tilde{d}_i(\rho)} + \rho I_N\right)^{-1} z_k\right)\right| \xrightarrow{\text{a.s.}} 0.$$

Proof. By expanding the definition of $\tilde{d}_k$, first observe that

$$\tilde{d}_k - \frac{1}{N}z_k^*\left(\alpha\frac{1}{n}\sum_{i\neq k}\frac{z_iz_i^*}{\tilde{d}_i} + \rho I_N\right)^{-1} z_k = \alpha\frac{1}{n}\sum_{i\neq k}\frac{1}{N}z_k^*\hat{S}_{(k)}^{-1}z_iz_i^*\frac{\gamma_N - \tilde{d}_i}{\gamma_N\tilde{d}_i}\left(\alpha\frac{1}{n}\sum_{j\neq k}\frac{z_jz_j^*}{\tilde{d}_j} + \rho I_N\right)^{-1} z_k.$$

Similar to the derivation of ξ_1, we now proceed to approximating $\tilde{d}_i$ in the central denominator and each $\tilde{d}_j$ in the rightmost inverse matrix by the non-random γ_N. We obtain (from Lemma 1)

$$\tilde{d}_k - \frac{1}{N}z_k^*\left(\alpha\frac{1}{n}\sum_{i\neq k}\frac{z_iz_i^*}{\tilde{d}_i} + \rho I_N\right)^{-1} z_k = \frac{\alpha}{\gamma_N^2}\frac{1}{n}\sum_{i\neq k}\frac{1}{N}z_k^*\hat{S}_{(k)}^{-1}z_iz_i^*(\gamma_N - \tilde{d}_i)\hat{S}_{(k)}^{-1}z_k + o(N^{\varepsilon-1})$$

almost surely, for $\varepsilon > 0$ and uniformly so on ρ.

The objective is then to show that the first right-hand side term is $o(N^{\varepsilon-1})$ almost surely and that this holds uniformly on k and ρ. This is achieved by applying Lemma 2 with $c = d = z_k$. Indeed, Lemma 2 ensures that, for each integer p,[3]

$$\mathrm{E}\left[\left|\frac{1}{n}\sum_{i\neq k}\frac{1}{N}z_k^*S_{(k)}^{-1}(\rho)z_iz_i^*S_{(k)}^{-1}(\rho)z_k\left(\frac{1}{N}z_i^*S_{(i,k)}^{-1}(\rho)z_i - \gamma_N(\rho)\right)\right|^p\right] = O(N^{-p}).$$

From this lemma, applying Markov's inequality, we have for each k,

$$P\left(N^{1-\varepsilon}\left|\frac{1}{n}\sum_{i\neq k}\frac{1}{N}z_k^*\hat{S}_{(k)}^{-1}z_iz_i^*\hat{S}_{(k)}^{-1}z_k\left(\frac{1}{N}z_i^*\hat{S}_{(i,k)}^{-1}z_i - \gamma_N\right)\right| > \eta\right) \le KN^{-p\varepsilon}$$

for some $K > 0$ only dependent on $\eta > 0$. Applying the union bound on the $n(n + 1)$ events for $k = 1, \dots, n$ and for $\rho \in \{\rho_0, \dots, \rho_n\}$, regular n-discretization of $\mathcal{R}_\kappa$, we then have

$$P\left(\max_{k,j} N^{1-\varepsilon}\left|\frac{1}{n}\sum_{i\neq k}\frac{1}{N}z_k^*\hat{S}_{(k)}^{-1}z_iz_i^*\hat{S}_{(k)}^{-1}z_k\left(\frac{1}{N}z_i^*\hat{S}_{(i,k)}^{-1}z_i - \gamma_N(\rho_j)\right)\right| > \eta\right) \le KN^{-p\varepsilon+2}.$$

Taking $p > 3/\varepsilon$, by the Borel Cantelli lemma the above convergence holds almost surely, we finally get

$$\max_{k,j}\left|N^{1-\varepsilon}\left(\tilde{d}_k(\rho_j) - \frac{1}{N}z_k^*\left(\alpha(\rho_j)\frac{1}{n}\sum_{i\neq k}\frac{z_iz_i^*}{\tilde{d}_i(\rho_j)} + \rho_jI_N\right)^{-1} z_k\right)\right| \xrightarrow{\text{a.s.}} 0.$$

Using the ρ-Lipschitz property (which holds almost surely so for all large n a.s.) on both terms in the above difference concludes the proof of the proposition.

[3] Note that Lemma 2 can strictly be applied here for $n - 1$ instead of n; but since $1/n - 1/(n - 1) = O(n^{-2})$, this does not affect the result.

The crux of the proof for the convergence of ξ_2 starts now. In a similar manner as in the proof of Lemma 1, we define $\tilde{f}_i(\rho) = d_i(\rho)/\tilde{d}_i(\rho)$ and reorder the indexes in such a way that $\tilde{f}_1(\rho) \leq \cdots \leq \tilde{f}_n(\rho)$ (this ordering depending on ρ). Then, by definition of $d_n(\rho) = \tilde{f}_i(\rho)\tilde{d}_i(\rho)$,

$$\begin{aligned}\tilde{d}_n(\rho)\tilde{f}_n(\rho) &= \frac{1}{N} z_n^* \left(\alpha(\rho)\frac{1}{n}\sum_{i<n}\frac{z_i z_i^*}{\tilde{d}_i(\rho)\tilde{f}_i(\rho)} + \rho I_n\right)^{-1} z_n \\ &\leq \frac{1}{N} z_n^* \left(\alpha(\rho)\frac{1}{\tilde{f}_n(\rho)}\frac{1}{n}\sum_{i<n}\frac{z_i z_i^*}{\tilde{d}_i(\rho)} + \rho I_n\right)^{-1} z_n\end{aligned}$$

where we used $\tilde{f}_n(\rho) \geq \tilde{f}_i(\rho)$ for each i. This inequality is equivalent to

$$\tilde{d}_n(\rho) \leq \frac{1}{N} z_n^* \left(\alpha(\rho)\frac{1}{n}\sum_{i<n}\frac{z_i z_i^*}{\tilde{d}_i(\rho)} + \tilde{f}_n(\rho)\rho I_n\right)^{-1} z_n.$$

Assume now that, over some sequence $\{\rho^{(n)}\} \in \mathcal{R}_\kappa$, $\tilde{f}_n(\rho^{(n)}) > 1 + N^{\eta-1}$ infinitely often for some $\eta > 0$ (or equivalently, $d_n(\rho^{(n)}) > \tilde{d}_n(\rho^{(n)}) + N^{\eta-1}$ i.o.). Then we would have

$$\begin{aligned}\tilde{d}_n(\rho^{(n)}) &\leq \frac{1}{N} z_n^* \left(\alpha(\rho^{(n)})\frac{1}{n}\sum_{i<n}\frac{z_i z_i^*}{\tilde{d}_i(\rho^{(n)})} + \rho^{(n)}(1+N^{\eta-1})I_N\right)^{-1} z_n \\ &= \tilde{d}_n(\rho^{(n)}) - N^{\eta-1}\frac{1}{N} z_n^* \left(\frac{1}{n}\sum_{i<n}\frac{\alpha(\rho^{(n)}) z_i z_i^*}{\rho^{(n)}\tilde{d}_i(\rho^{(n)})} + (1+N^{\eta-1})I_N\right)^{-1} \left(\frac{1}{n}\sum_{i<n}\frac{\alpha(\rho^{(n)}) z_i z_i^*}{\tilde{d}_i(\rho^{(n)})} + \rho I_N\right)^{-1} z_n.\end{aligned}$$

But, by Proposition 2, letting $0 < \varepsilon < \eta$, we have, for all large n a.s.,

$$\frac{1}{N} z_n^* \left(\alpha(\rho^{(n)})\frac{1}{n}\sum_{i<n}\frac{z_i z_i^*}{\tilde{d}_i(\rho^{(n)})} + \rho^{(n)} I_n\right)^{-1} z_n \leq \tilde{d}_n(\rho^{(n)}) + N^{\varepsilon-1}$$

which, along with the uniform boundedness of the $\tilde{d}_i$ away from zero, leads to

$$\tilde{d}_n(\rho^{(n)}) \leq \tilde{d}_n(\rho^{(n)}) + N^{\varepsilon-1} - KN^{\eta-1}$$

for some $K > 0$. But, as $N^{\varepsilon-1} - KN^{\eta-1} < 0$ for all large N, we obtain a contradiction. Hence, for each $\eta > 0$, we have for all large n a.s., $d_n(\rho) < \tilde{d}_n(\rho) + N^{\eta-1}$ uniformly on $\rho \in \mathcal{R}_\kappa$. Proceeding similarly with $d_1(\rho)$, and exploiting $\limsup_n \sup_\rho \max_i |\tilde{d}_i(\rho)| = O(1)$ a.s., we finally have, for each $0 < \varepsilon < \frac{1}{2}$, that

$$\max_{1\leq k\leq n}\sup_{\rho\in\mathcal{R}_\kappa}\left|N^{1-\varepsilon}\left(d_k(\rho) - \tilde{d}_k(\rho)\right)\right| \xrightarrow{\text{a.s.}} 0$$

(for this, take an η such that $0 < \eta < \varepsilon$ and use $\max_k \sup_\rho |d_k(\rho) - \tilde{d}_k(\rho)| < N^{\eta-1}$ for all large n a.s.).

Getting back to ξ_2, we now have

$$N^{1-\varepsilon}|\xi_2(\rho)| = N^{1-\varepsilon}\left|a^*\hat{C}_N^{-1}(\rho)\left(\frac{\alpha(\rho)}{n}\sum_{i=1}^n\frac{d_i(\rho) - \tilde{d}_i(\rho)}{d_i(\rho)\tilde{d}_i(\rho)} z_i z_i^*\right)\hat{S}_N^{-1}(\rho) b\right|.$$

But, from the above result,

$$\begin{aligned}N^{1-\varepsilon}\left\|\frac{\alpha(\rho)}{n}\sum_{i=1}^n\frac{d_i(\rho) - \tilde{d}_i(\rho)}{d_i(\rho)\tilde{d}_i(\rho)} z_i z_i^*\right\| &\leq N^{1-\varepsilon}\max_{1\leq k\leq n}\left|\frac{d_k(\rho) - \tilde{d}_k(\rho)}{d_k(\rho)\tilde{d}_k(\rho)}\right|\left\|\frac{\alpha(\rho)}{n}\sum_{i=1}^n z_i z_i^*\right\| \\ &\xrightarrow{\text{a.s.}} 0\end{aligned}$$

uniformly so on $\rho \in \mathcal{R}_\kappa$ which, along with the boundedness of $\|\hat{C}_N^{-1}\|$, $\|\hat{S}_N^{-1}\|$, $\|a\|$, and $\|b\|$, finally gives $N^{1-\varepsilon}\xi_2 \xrightarrow{\text{a.s.}} 0$ uniformly on $\rho \in \mathcal{R}_\kappa$ as desired.

We have then proved that for each $\varepsilon > 0$,

$$\sup_{\rho\in\mathcal{R}_\kappa}\left|N^{1-\varepsilon}\left(a^*\hat{C}_N^{-1}(\rho)b - a^*\hat{S}_N^{-1}(\rho)b\right)\right| \xrightarrow{\text{a.s.}} 0$$

which proves Theorem 1 for $k = -1$. The generalization to arbitrary k is rather immediate. Writing recursively $\hat{C}_N^k - \hat{S}_N^k = \hat{C}_N^{k-1}(\hat{C}_N - \hat{S}_N) + (\hat{C}_N^{k-1} - \hat{S}_N^{k-1})\hat{S}_N$, for positive k or $\hat{C}_N^k - \hat{S}_N^k = \hat{C}_N^k(\hat{S}_N - \hat{C}_N)\hat{S}_N^{-1} + (\hat{C}_N^{k-1} - \hat{S}_N^{k-1})\hat{S}_N^{-1}$, (3) becomes a finite sum of terms that can be treated almost exactly as in the proof. This concludes the proof of Theorem 1.

3.2. Fluctuations of the GLRT detector

This section is devoted to the proof of Theorem 2, which shall fundamentally rely on Theorem 1. The proof will be established in two steps. First, we shall prove the convergence for each $\rho \in \mathcal{R}_\kappa$, which we then generalize to the uniform statement of the theorem.

Let us then fix $\rho \in \mathcal{R}_\kappa$ for the moment. In anticipation of the eventual replacement of $\hat{C}_N(\rho)$ by $\underline{\hat{S}}_N(\underline{\rho})$, we start by studying the fluctuations of the bilinear forms involved in $T_N(\rho)$ but with $\hat{C}_N(\rho)$ replaced by $\underline{\hat{S}}_N(\underline{\rho})$ (note that $T_N(\rho)$ remains constant when scaling $\hat{C}_N(\rho)$ by any constant, so that replacing $\hat{C}_N(\rho)$ by $\underline{\hat{S}}_N(\underline{\rho})$ instead of by $\underline{\hat{S}}_N(\underline{\rho}) \cdot \frac{1}{N} \operatorname{tr} \hat{S}_N(\rho)$ as one would expect comes with no effect).

Our first goal is to show that the vector $\sqrt{N}(\Re[y^* \underline{\hat{S}}_N^{-1}(\underline{\rho})p], \Im[y^* \underline{\hat{S}}_N^{-1}(\underline{\rho})p])$ is asymptotically well approximated by a zero mean Gaussian vector with given covariance matrix. To this end, let us denote $A = [y\ p] \in \mathbb{C}^{N\times 2}$ and $Q_N = Q_N(\underline{\rho}) = (I_N + (1-\underline{\rho})m(-\underline{\rho})C_N)^{-1}$. Then, from [8, Lemma 5.3] (adapted to our current notations and normalizations), for any Hermitian $B \in \mathbb{C}^{2\times 2}$ and for any $u \in \mathbb{R}$,

$$\mathrm{E}\left[\exp\left(\imath\sqrt{N}u \operatorname{tr} BA^*\left[\underline{\hat{S}}_N(\underline{\rho})^{-1} - \frac{1}{\underline{\rho}}Q_N(\underline{\rho})\right]A\right)\ \middle|\ y\right] = \exp\left(-\frac{1}{2}u^2 \Delta_N^2(B;y;p)\right) + O\left(N^{-\frac{1}{2}}\right) \tag{6}$$

where we denote by $\mathrm{E}[\cdot|y]$ the conditional expectation with respect to the random vector y and where

$$\Delta_N^2(B;y;p) \triangleq \frac{cm(-\underline{\rho})^2(1-\underline{\rho})^2 \operatorname{tr}\left(ABA^*C_N Q_N^2(\underline{\rho})\right)^2}{\underline{\rho}^2\left(1 - cm(-\underline{\rho})^2(1-\underline{\rho})^2 \frac{1}{N}\operatorname{tr} C_N^2 Q_N^2(\underline{\rho})\right)}.$$

Also, we have from classical central limit results on Gaussian random variables

$$\mathrm{E}\left[\exp\left(\imath\sqrt{N}u \operatorname{tr} B\left[A^* Q_N(\underline{\rho})A - \Gamma_N\right]\right)\right] = \exp\left(-\frac{1}{2}u^2 \Delta_N'^2(B;p)\right) + O\left(N^{-\frac{1}{2}}\right)$$

where

$$\Gamma_N \triangleq \begin{bmatrix} \frac{1}{N}\operatorname{tr} C_N Q_N(\underline{\rho}) & 0 \\ 0 & p^* Q_N(\underline{\rho}) p \end{bmatrix}$$

$$\Delta_N'^2(B;p) \triangleq \frac{B_{11}^2}{\underline{\rho}^2}\frac{1}{N}\operatorname{tr} C_N^2 Q_N^2(\underline{\rho}) + \frac{2|B_{12}|^2}{\underline{\rho}^2} p^* C_N Q_N^2(\underline{\rho}) p.$$

Besides, the $O(N^{-\frac{1}{2}})$ terms on the right-hand side of (6) remain $O(N^{-\frac{1}{2}})$ under expectation over y (for this, see the proof of Chapon et al. [8, Lemma 5.3]).

Altogether, we then have

$$\mathrm{E}\left[\exp\left(\imath\sqrt{N}u \operatorname{tr} B\left[A^* \underline{\hat{S}}_N^{-1}(\underline{\rho})A - \Gamma_N\right]\right)\right] = \mathrm{E}\left[\exp\left(-\frac{1}{2}u^2 \Delta_N^2(B;y;p)\right)\right]\exp\left(-\frac{1}{2}u^2 \Delta_N'^2(B;p)\right) + O\left(N^{-\frac{1}{2}}\right).$$

Note now that

$$A^* C_N Q_N^2(\underline{\rho})A - \Upsilon_N \xrightarrow{\text{a.s.}} 0$$

where

$$\Upsilon_N \triangleq \begin{bmatrix} \frac{1}{N}\operatorname{tr} C_N^2 Q_N^2(\underline{\rho}) & 0 \\ 0 & p^* C_N Q_N^2(\underline{\rho}) p \end{bmatrix}$$

so that, by dominated convergence, we obtain

$$\mathrm{E}\left[\exp\left(\imath\sqrt{N}u \operatorname{tr} B\left[A^* \underline{\hat{S}}_N^{-1}(\underline{\rho})A - \Gamma_N\right]\right)\right] = \exp\left(-\frac{1}{2}u^2\left[\Delta_N^2(B;p) + \Delta_N'^2(B;p)\right]\right) + o(1)$$

where we defined

$$\Delta_N^2(B;p) \triangleq \frac{cm(-\underline{\rho})^2(1-\underline{\rho})^2 \operatorname{tr}(B\Upsilon_N)^2}{\underline{\rho}^2\left(1 - cm(-\underline{\rho})^2(1-\underline{\rho})^2 \frac{1}{N}\operatorname{tr} C_N^2 Q_N^2(\underline{\rho})\right)}.$$

By a generalized Lévy's continuity theorem argument (see e.g. [18, Proposition 6]) and the Cramer–Wold device, we conclude that

$$\sqrt{N}\left(y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})y, \Re[y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p], \Im[y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p]\right) - Z_N = o_P(1)$$

where Z_N is a Gaussian random vector with mean and covariance matrix prescribed by the above approximation of $\sqrt{N}\operatorname{tr} BA^*\hat{\underline{S}}_N^{-1}A$ for each Hermitian B. In particular, taking $B_1 \in \left\{\begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0\end{bmatrix}, \begin{bmatrix} 0 & \frac{\imath}{2} \\ -\frac{\imath}{2} & 0\end{bmatrix}\right\}$ to retrieve the asymptotic variances of $\sqrt{N}\Re[y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p]$ and $\sqrt{N}\Im[y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p]$, respectively, gives

$$\Delta_N^2(B_1;p) = \frac{1}{2\underline{\rho}^2}p^*C_NQ_N^2(\underline{\rho})p\frac{cm(-\underline{\rho})^2(1-\underline{\rho})^2\frac{1}{N}\operatorname{tr} C_N^2Q_N^2(\underline{\rho})}{1-cm(-\underline{\rho})^2(1-\underline{\rho})^2\frac{1}{N}\operatorname{tr} C_N^2Q_N^2(\underline{\rho})}$$
$$\Delta_N'^2(B_1;p) = \frac{1}{2\underline{\rho}^2}p^*C_NQ_N^2(\underline{\rho})p$$

and thus $\sqrt{N}(\Re[y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p], \Im[y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p])$ is asymptotically equivalent to a Gaussian vector with zero mean and covariance matrix

$$(\Delta_N^2(B_1;p)+\Delta_N'^2(B_1;p))I_2 = \frac{1}{2\underline{\rho}^2}\frac{p^*C_NQ_N^2(\underline{\rho})p}{1-cm(-\underline{\rho})^2(1-\underline{\rho})^2\frac{1}{N}\operatorname{tr} C_N^2Q_N^2(\underline{\rho})}I_2.$$

We are now in position to apply Theorem 1. Reminding that $\hat{S}_N^{-1}(\rho)(\rho+\frac{1}{\gamma_N(\rho)}\frac{1-\rho}{1-(1-\rho)c}) = \hat{\underline{S}}_N^{-1}(\underline{\rho})$, we have by Theorem 1 for $k=-1$

$$\sqrt{N}A^*\left[\hat{C}_N^{-1}(\rho) - \frac{\hat{\underline{S}}_N(\underline{\rho})^{-1}}{\rho+\frac{1}{\gamma_N(\rho)}\frac{1-\rho}{1-(1-\rho)c}}\right]A \xrightarrow{\text{a.s.}} 0.$$

Since almost sure convergence implies weak convergence, $\sqrt{N}A^*\hat{C}_N^{-1}(\rho)A$ has the same asymptotic fluctuations as $\sqrt{N}A^*\hat{\underline{S}}_N^{-1}(\underline{\rho})A/(\frac{1}{N}\operatorname{tr}\hat{S}_N(\rho))$. Also, as $T_N(\rho)$ remains identical when scaling $\hat{C}_N^{-1}(\rho)$ by $\frac{1}{N}\operatorname{tr}\hat{S}_N(\rho)$, only the fluctuations of $\sqrt{N}A^*\hat{\underline{S}}_N^{-1}(\underline{\rho})A$ are of interest, which were previously derived. We then finally conclude by the delta method (or more directly by Slutsky's lemma) that

$$\sqrt{\frac{N}{y^*\hat{C}_N^{-1}(\rho)yp^*\hat{C}_N^{-1}(\rho)p}}\begin{bmatrix}\Re\left[y^*\hat{C}_N^{-1}(\rho)p\right] \\ \Im\left[y^*\hat{C}_N^{-1}(\rho)p\right]\end{bmatrix} - \sigma_N(\rho)Z' = o_P(1)$$

for some $Z' \sim \mathcal{N}(0, I_2)$ and

$$\sigma_N^2(\rho) \triangleq \frac{1}{2}\frac{p^*C_NQ_N^2(\underline{\rho})p}{p^*Q_N(\underline{\rho})p\cdot\frac{1}{N}\operatorname{tr} C_NQ_N(\underline{\rho})\cdot\left(1-cm(-\underline{\rho})^2(1-\underline{\rho})^2\frac{1}{N}\operatorname{tr} C_N^2Q_N^2(\underline{\rho})\right)}.$$

It unfolds that, for $\gamma > 0$,

$$P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}}\right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\rho)}\right) \to 0 \tag{7}$$

as desired.

The second step of the proof is to generalize (7) to uniform convergence across $\rho \in \mathcal{R}_\kappa$. To this end, somewhat similar to above, we shall transfer the distribution $P(\sqrt{N}T_N(\rho) > \gamma)$ to $P(\sqrt{N}\underline{T}_N(\rho) > \gamma)$ by exploiting the uniform convergence of Theorem 1, where we defined

$$\underline{T}_N(\rho) \triangleq \frac{\left|y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p\right|}{\sqrt{y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})y}\sqrt{p^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p}}$$

and exploit a ρ-Lipschitz property of $\sqrt{N}\underline{T}_N(\rho)$ to reduce the uniform convergence over $\mathcal{R}_\kappa$ to a uniform convergence over finitely many values of ρ.

The ρ-Lipschitz property we shall need is as follows: for each $\varepsilon > 0$

$$\lim_{\delta\to 0}\lim_{N\to\infty} P\left(\sup_{\substack{\rho,\rho'\in\mathcal{R}_\kappa\\ |\rho-\rho'|<\delta}} \sqrt{N}\left|T_N(\rho)-T_N(\rho')\right| > \varepsilon\right) = 0. \tag{8}$$

Let us prove this result. By Theorem 1, since almost sure convergence implies convergence in distribution, we have

$$P\left(\sup_{\rho\in\mathcal{R}_\kappa}\sqrt{N}\left|T_N(\rho)-\underline{T}_N(\rho)\right| > \varepsilon\right) \to 0.$$

Applying this result to (8) induces that it is sufficient to prove (8) for $\underline{T}_N(\rho)$ in place of $T_N(\rho)$. Let $\eta > 0$ be small and $\mathcal{A}_N^\eta \triangleq \{\exists \underline{\rho} \in \mathcal{R}_\kappa, y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})yp^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p < \eta\}$. Developing the difference $\underline{T}_N(\rho) - \underline{T}_N(\rho')$ and isolating the denominator according to its belonging to $\mathcal{A}_N^\eta$ or not, we may write

$$P\left(\sup_{\substack{\rho,\rho'\in\mathcal{R}_\kappa\\ |\rho-\rho'|<\delta}} \sqrt{N}\left|\underline{T}_N(\rho)-\underline{T}_N(\rho')\right| > \varepsilon\right) \le P\left(\mathcal{A}_N^\eta\right) + P\left(\sup_{\substack{\rho,\rho'\in\mathcal{R}_\kappa\\ |\rho-\rho'|<\delta}} \sqrt{N}V_N(\rho,\rho') > \varepsilon\eta\right)$$

where

$$V_N(\rho,\rho') \triangleq \left|y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p\right|\sqrt{y^*\hat{\underline{S}}_N^{-1}(\underline{\rho}')y}\sqrt{p^*\hat{\underline{S}}_N^{-1}(\underline{\rho}')p} - \left|y^*\hat{\underline{S}}_N^{-1}(\underline{\rho}')p\right|\sqrt{y^*\hat{\underline{S}}_N^{-1}(\underline{\rho})y}\sqrt{p^*\hat{\underline{S}}_N^{-1}(\underline{\rho})p}.$$

From classical random matrix results, $P(\mathcal{A}_N^\eta) \to 0$ for a sufficiently small choice of η. To prove that $\lim_\delta \limsup_n P(\sup_{|\rho-\rho'|<\delta}\sqrt{N}V_N(\rho,\rho') > \varepsilon\eta) = 0$, it is then sufficient to show that

$$\lim_{\delta\to 0}\limsup_{n} P\left(\sup_{\substack{\rho,\rho'\in\mathcal{R}_\kappa\\ |\rho-\rho'|<\delta}} \sqrt{N}|y^*\hat{\underline{S}}_N(\underline{\rho})^{-1}p - y^*\hat{\underline{S}}_N(\underline{\rho}')^{-1}p| > \varepsilon'\right) = 0 \tag{9}$$

for any $\varepsilon' > 0$ and similarly for $y^*\hat{\underline{S}}_N(\underline{\rho})^{-1}y - y^*\hat{\underline{S}}_N(\underline{\rho}')^{-1}y$ and $p^*\hat{\underline{S}}_N(\underline{\rho})^{-1}p - p^*\hat{\underline{S}}_N(\underline{\rho}')^{-1}p$. Let us prove (9), the other two results following essentially the same line of arguments. For this, by Kallenberg [19, Corollary 16.9] (see also [5, Theorem 12.3]), it is sufficient to prove, say

$$\sup_{\substack{\rho,\rho'\in\mathcal{R}_\kappa\\ \rho\neq\rho'}}\sup_n \frac{\mathrm{E}\left[\sqrt{N}\left|y^*\hat{\underline{S}}_N(\underline{\rho})^{-1}p - y^*\hat{\underline{S}}_N(\underline{\rho}')^{-1}p\right|^2\right]}{|\rho-\rho'|^2} < \infty.$$

But then, remarking that

$$\sqrt{N}y^*\hat{\underline{S}}_N(\underline{\rho})^{-1}p - y^*\hat{\underline{S}}_N(\underline{\rho}')^{-1}p = (\underline{\rho}'-\underline{\rho})\sqrt{N}y^*\hat{\underline{S}}_N(\underline{\rho})^{-1}\left(I_N - \frac{1}{n}\sum_{i=1}^n z_iz_i^*\right)\hat{\underline{S}}_N(\underline{\rho}')^{-1}p$$

this reduces to showing that

$$\sup_{\rho,\rho'\in\mathcal{R}_\kappa}\sup_n \mathrm{E}\left[N\left|y^*\hat{\underline{S}}_N(\underline{\rho})^{-1}\left(I_N - \frac{1}{n}\sum_{i=1}^n z_iz_i^*\right)\hat{\underline{S}}_N(\underline{\rho}')^{-1}p\right|^2\right] < \infty.$$

Conditioning first on $z_1,\ldots,z_n$, this further reduces to showing

$$\sup_{\rho,\rho'\in\mathcal{R}_\kappa}\sup_n \mathrm{E}\left[\left\|\hat{\underline{S}}_N(\underline{\rho})^{-1}\left(I_N - \frac{1}{n}\sum_{i=1}^n z_iz_i^*\right)\hat{\underline{S}}_N(\underline{\rho}')^{-1}p\right\|^2\right] < \infty.$$

But this is yet another standard random matrix result, obtained e.g., by noticing that

$$\left\|\hat{\underline{S}}_N(\underline{\rho})^{-1}\left(I_N - \frac{1}{n}\sum_{i=1}^n z_iz_i^*\right)\hat{\underline{S}}_N(\underline{\rho}')^{-1}p\right\|^2 \le \frac{1}{\kappa^4}\left\|I_N - \frac{1}{n}\sum_{i=1}^n z_iz_i^*\right\|^2$$

which remains of uniformly finite expectation (left norm is vector Euclidean norm, right norm is matrix spectral norm). This completes the proof of (8).

Getting back to our original problem, let us now take $\varepsilon > 0$ arbitrary, $\rho_1 < \cdots < \rho_K$ be a regular sampling of $\mathcal{R}_\kappa$, and $\delta = 1/K$. Then by (7), K being fixed, for all $n > n_0(\varepsilon)$,

$$\max_{1\le k\le K}\left|P\left(T_N(\rho_i) > \frac{\gamma}{\sqrt{N}}\right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\rho_i)}\right)\right| < \varepsilon. \tag{10}$$

Also, from (8), for small enough δ,

$$\max_{1\le k\le K} P\left(\sup_{\substack{\rho\in\mathcal{R}_\kappa\\ |\rho-\rho_k|<\delta}} \sqrt{N}|T_N(\rho) - T_N(\rho_k)| > \gamma\zeta\right) \le P\left(\sup_{\substack{\rho,\rho'\in\mathcal{R}_\kappa\\ |\rho-\rho'|<\delta}} \sqrt{N}|T_N(\rho) - T_N(\rho')| > \gamma\zeta\right) < \varepsilon$$

for all large $n > n_0'(\varepsilon, \zeta) > n_0(\varepsilon)$ where $\zeta > 0$ is also taken arbitrarily small. Thus we have, for each $\rho \in \mathcal{R}_\kappa$ and for $n > n_0'(\varepsilon, \zeta)$

$$\begin{aligned} P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}}\right) &\le P\left(T_N(\rho_i) > \frac{\gamma(1-\zeta)}{\sqrt{N}}\right) + P\left(\sqrt{N}|T_N(\rho) - T_N(\rho_i)| > \gamma\zeta\right) \\ &\le P\left(T_N(\rho_i) > \frac{\gamma(1-\zeta)}{\sqrt{N}}\right) + \varepsilon \end{aligned}$$

for $i \le K$ the unique index such that $|\rho - \rho_i| < \delta$ and where the inequality holds uniformly on $\rho \in \mathcal{R}_\kappa$. Similarly, reversing the roles of ρ and ρ_i,

$$P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}}\right) \ge P\left(T_N(\rho_i) > \frac{\gamma(1+\zeta)}{\sqrt{N}}\right) - \varepsilon.$$

As a consequence, by (10), for $n > n_0'(\varepsilon, \zeta)$, uniformly on $\rho \in \mathcal{R}_\kappa$,

$$\begin{aligned} P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}}\right) &\le \exp\left(-\frac{\gamma^2(1-\zeta)^2}{2\sigma_N^2(\rho_i)}\right) + 2\varepsilon \\ P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}}\right) &\ge \exp\left(-\frac{\gamma^2(1+\zeta)^2}{2\sigma_N^2(\rho_i)}\right) - 2\varepsilon \end{aligned}$$

which, by continuity of the exponential and of $\rho \mapsto \sigma_N(\rho)$,[4] letting ζ and δ small enough (up to growing $n_0'(\varepsilon, \zeta)$), leads to

$$\sup_{\rho\in\mathcal{R}_\kappa}\left|P\left(\sqrt{N}T_N(\rho) > \gamma\right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\rho)}\right)\right| \le 3\varepsilon$$

for all $n > n_0'(\varepsilon, \zeta)$, which completes the proof.

3.3. Around empirical estimates

This section is dedicated to the proof of Proposition 1 and Corollary 1.

We start by showing that $\hat{\sigma}_N^2(1)$ is well defined. It is easy to observe that the ratio defining $\hat{\sigma}_N^2(\rho)$ converges to an undetermined form (zero over zero) as $\rho \uparrow 1$. Applying l'Hospital's rule to the ratio, using the equivalence between $\hat{C}_N(\rho)$ and $\hat{\underline{S}}_N(\underline{\rho})$, i.e.,

$$\sup_{\rho\in(\max\{0,1-c^{-1}\},1]}\left\|\frac{\hat{C}_N(\rho)}{\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)} - \hat{\underline{S}}_N(\underline{\rho})\right\| \xrightarrow{\text{a.s.}} 0$$

then the differentiation $\frac{d}{d\underline{\rho}}\hat{\underline{S}}_N^{-1}(\underline{\rho}) = -\hat{\underline{S}}_N^{-2}(\underline{\rho})(I_N - \frac{1}{n}\sum_i z_i z_i^*)$ and finally the limit $\hat{\underline{S}}_N^{-1}(\underline{\rho}) \to I_N$ as $\rho \uparrow 1$, we end up with

$$\hat{\sigma}_N^2(\underline{\rho}) \to \frac{1}{2}\frac{p^*\left(\frac{1}{n}\sum_{i=1}^n z_i z_i^*\right)p}{\frac{1}{N}\operatorname{tr}\left(\frac{1}{n}\sum_{i=1}^n z_i z_i^*\right)}.$$

[4] Note that it is unnecessary to ensure $\liminf_N \sigma_N(\rho) > 0$ as the exponential would tend to zero anyhow in this scenario.

Since $p^*\frac{1}{n}\sum_i z_iz_i^*p - p^*C_Np \xrightarrow{\text{a.s.}} 0$, $\frac{1}{N}\operatorname{tr}\frac{1}{n}\sum_i z_iz_i^* \xrightarrow{\text{a.s.}} 1$ as $n \to \infty$, we have, by a uniform variation of l'Hospital's rule applied to $\hat{\sigma}_N^2(\rho)$ and $\sigma_N^2(\rho)$ [20, Lemma 13],

$$\lim_{\rho\uparrow 1}\limsup_N \left|\hat{\sigma}_N^2(\rho) - \hat{\sigma}_N^2(\rho)\right| \xrightarrow{\text{a.s.}} 0.$$

Hence, letting $\varepsilon, \kappa > 0$ small,

$$\sup_{\rho\in(1-\kappa,1]}\left|\hat{\sigma}_N^2(\rho) - \sigma_N^2(\rho)\right| \le \varepsilon$$

for all large n almost surely. From now on, it then suffices to prove Proposition 1 on the complementary set $\mathcal{R}'_\kappa \triangleq [\kappa + \min\{0, 1-c^{-1}\}, 1-\kappa]$. For this, we first recall the following results borrowed from [13]. For $z \in \mathbb{C}\setminus\mathbb{R}^+$, defining

$$\hat{\underline{\underline{S}}}_N(z) \triangleq (1-\underline{\rho})\frac{1}{n}\sum_{i=1}^n z_iz_i^* - zI_N$$

(so in particular $\hat{\underline{\underline{S}}}_N(-\underline{\rho}) = \hat{\underline{S}}_N(\underline{\rho})$, for all $\underline{\rho} \in \mathcal{R}_\kappa$), we have, with $\mathcal{C}$ a compact set of $\mathbb{C}\setminus\mathbb{R}^+$ and any integer k,

$$\sup_{\bar{z}\in\mathcal{C}}\left|\frac{d^k}{dz^k}\left\{\frac{1}{N}\operatorname{tr}\hat{\underline{\underline{S}}}_N^{-1}(z) - \frac{1}{N}\operatorname{tr}\left(-z\left[I_N + (1-\underline{\rho})m_N(z)C_N\right]\right)^{-1}\right\}_{z=\bar{z}}\right| \xrightarrow{\text{a.s.}} 0$$
$$\sup_{\bar{z}\in\mathcal{C}}\left|\frac{d^k}{dz^k}\left\{p^*\hat{\underline{\underline{S}}}_N^{-1}(z)p - p^*\left(-z\left[I_N + (1-\underline{\rho})m_N(z)C_N\right]\right)^{-1}p\right\}_{z=\bar{z}}\right| \xrightarrow{\text{a.s.}} 0$$

where $m_N(z)$ is defined as the unique solution with positive (resp. negative) imaginary part if $\Im[z] > 0$ (resp. $\Im[z] < 0$) or unique positive solution if $z < 0$ of

$$m_N(z) = \left(-z + c\int\frac{(1-\underline{\rho})t}{1+(1-\underline{\rho})tm_N(z)}\nu_N(dt)\right)^{-1}$$

(this follows directly from [28]).

This expression of $m_N(z)$ can be rewritten under the more convenient form

$$\begin{aligned} m_N(z) &= -\frac{1-c}{z} + c\int\frac{\nu_N(dt)}{-z - z(1-\underline{\rho})tm_N(z)} \\ &= -\frac{1-c}{z} + c\frac{1}{N}\operatorname{tr}\left(-z\left[I_N + (1-\underline{\rho})m_N(z)C_N\right]\right)^{-1} \end{aligned}$$

so that, with $\sup_\rho\left\|\frac{\hat{C}_N(\rho)}{\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)} - \hat{\underline{\underline{S}}}_N(-\underline{\rho})\right\| \xrightarrow{\text{a.s.}} 0$,

$$\sup_{\rho\in\mathcal{R}'_\kappa}\left|m_N(-\underline{\rho}) - \left(\frac{1-c_N}{\underline{\rho}} + c_N\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho)\cdot\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)\right)\right| \xrightarrow{\text{a.s.}} 0$$
$$\sup_{\rho\in\mathcal{R}'_\kappa}\left|\int\frac{t\nu_N(dt)}{1+(1-\underline{\rho})m_N(-\underline{\rho})t} - \frac{1-\underline{\rho}\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho)\cdot\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)}{(1-\underline{\rho})m_N(-\underline{\rho})}\right| \xrightarrow{\text{a.s.}} 0.$$

Differentiating along z the first defining identity of $m_N(z)$, we also recall that

$$m'_N(z) = \frac{m_N^2(z)}{1 - c\int\frac{m_N(z)^2(1-\underline{\rho})^2t^2\nu_N(dt)}{(1-(1-\underline{\rho})tm_N(-\underline{\rho}))^2}}.$$

Now, remark that

$$p^*\hat{\underline{\underline{S}}}_N(-\underline{\rho})^{-2}p = \frac{d}{dz}\left[p^*\hat{\underline{\underline{S}}}_N(z)^{-1}p\right]_{z=-\underline{\rho}}$$

which (by analyticity) is uniformly well approximated by

$$\begin{aligned} \frac{d}{dz}\left[p^*\left(-z\left[I_N + (1-\underline{\rho})m_N(z)C_N\right]\right)^{-1}p\right]_{z=-\underline{\rho}} &= \frac{1}{\underline{\rho}^2}p^*Q_N(\underline{\rho})p - \frac{1}{\underline{\rho}}(1-\underline{\rho})m'_N(-\underline{\rho})p^*C_NQ_N^2(\underline{\rho})p \\ &= \frac{1}{\underline{\rho}^2}p^*Q_N(\underline{\rho})p - \frac{1}{\underline{\rho}}(1-\underline{\rho})\frac{m_N^2(-\underline{\rho})p^*C_NQ_N^2(\underline{\rho})p}{1-cm_N(-\underline{\rho})^2(1-\underline{\rho})^2\frac{1}{N}\operatorname{tr}Q_N^2(\underline{\rho})}. \end{aligned}$$

(recall that $Q_N(\underline{\rho}) = \left(I_N + (1-\underline{\rho})m_N(-\underline{\rho})C_N\right)^{-1}$). We then conclude

$$\sup_{\rho\in\mathcal{R}'_\kappa}\left|\frac{p^*C_NQ_N^2(\underline{\rho})p}{1-cm_N(-\underline{\rho})^2(1-\underline{\rho})^2\frac{1}{N}\operatorname{tr}Q_N^2(\underline{\rho})} - \frac{p^*\hat{C}_N^{-1}(\rho)p\cdot\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho) - \underline{\rho}p^*\hat{C}_N^{-2}(\rho)p\cdot\left(\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)\right)^2}{(1-\underline{\rho})m_N(-\underline{\rho})^2}\right| \xrightarrow{\text{a.s.}} 0.$$

Putting all results together, we obtain

$$\sup_{\rho\in\mathcal{R}'_\kappa}\left|\frac{1}{2}\frac{1-\underline{\rho}\frac{p^*\hat{C}_N^{-2}(\rho)}{p^*\hat{C}_N^{-1}(\rho)p}\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho)\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)}{\left(1-c_N+c_N\underline{\rho}\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho)\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)\right)\left(1-\underline{\rho}\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho)\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)\right)} - \sigma_N^2(\rho)\right| \xrightarrow{\text{a.s.}} 0.$$

Noting that $\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho) = 1$ by definition of $\hat{C}_N$ for all ρ and that $\sup_\rho|\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho) - \rho\underline{\rho}^{-1}| \xrightarrow{\text{a.s.}} 0$, we retrieve the expected result.

It now remains to prove Corollary 1. This is easily performed thanks to Theorem 2 and Proposition 1. From these, we indeed have the three relations

$$P\left(\sqrt{N}T_N(\hat{\rho}_N^*) > \gamma\right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\hat{\rho}_N^*)}\right) \xrightarrow{\text{a.s.}} 0$$
$$P\left(\sqrt{N}T_N(\rho_N^*) > \gamma\right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\rho_N^*)}\right) \to 0$$
$$\exp\left(-\frac{\gamma^2}{2\sigma_N^2(\hat{\rho}_N^*)}\right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^{*2}}\right) \xrightarrow{\text{a.s.}} 0$$

where we denoted ρ_N^* any element in the argmin over ρ of $P(\sqrt{N}T_N(\rho) > \gamma)$ and σ_N^{*2} the minimum of $\sigma_N(\underline{\rho})$ (i.e. the minimizer for $\exp(-\frac{\gamma^2}{2\sigma_N^2(\underline{\rho})})$). Note that the first two relations rely fundamentally on the uniform convergence $\sup_{\rho\in\mathcal{R}_\kappa}|P\left(\sqrt{N}T_N(\rho) > \gamma\right) - \exp\left(-\gamma^2/(2\sigma_N^2(\rho))\right)| \xrightarrow{\text{a.s.}} 0$. By definition of ρ_N^* and σ_N^{*2}, we also have

$$\exp\left(-\frac{\gamma^2}{2\sigma_N^{*2}}\right) \le \min\left\{\exp\left(-\frac{\gamma^2}{2\sigma_N^2(\hat{\rho}_N^*)}\right), \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\rho_N^*)}\right)\right\}$$
$$P\left(\sqrt{N}T_N(\rho_N^*) > \gamma\right) \le P\left(\sqrt{N}T(\hat{\rho}_N^*) > \gamma\right).$$

Putting things together then gives

$$P\left(\sqrt{N}T(\hat{\rho}_N^*) > \gamma\right) - P\left(\sqrt{N}T_N(\rho_N^*) > \gamma\right) \xrightarrow{\text{a.s.}} 0$$

which is the expected result.

Appendix. Proof of Lemma 2

This section is devoted to the proof of the key Lemma 2. The proof relies on an appropriate decomposition of the quantity under study as a sum of martingale differences. Before delving into the core of the proofs, we introduce some notations along with some of the key-lemmas that will be extensively used in this section.

In this section, E_j will denote the conditional expectation with respect to the σ-field $\mathcal{F}_j$ generated by the vectors $(z_\ell, 1 \le \ell \le j)$. By convention, $\mathrm{E}_0 = \mathrm{E}$.

Useful lemmas. We shall review two key lemmas that will be extensively used, namely the generalized Hölder inequality (obtained by induction of the standard Hölder inequality) as well as an instance of Jensen's inequality.

Lemma 3 (*Jensen Inequality, Boyd and Vandenberghe [6]*). *Let $\mathcal{I}$ be a discrete set of elements of $\{1,\ldots,n\}$ with finite cardinality denoted by $|\mathcal{I}|$. Let $(\theta_i)_{i\in\mathcal{I}}$ be a sequence of complex scalars indexed by the set $\mathcal{I}$. Then, for any $p \ge 1$,*

$$\left|\sum_{i\in\mathcal{I}}\theta_i\right|^p \le |\mathcal{I}|^{p-1}\sum_{i=1}^n|\theta_i|^p.$$

Lemma 4 (*Generalized Hölder Inequality, See e.g., [16, Th. 5.1.2]*). *Let $X_1, \ldots, X_k$ be k complex random variables with finite moments of order k. Then,*

$$\left| \mathrm{E}\left[\prod_{i=1}^{k} X_i \right] \right| \leq \prod_{i=1}^{k} \left(\mathrm{E}\left[|X_i|^k \right] \right)^{\frac{1}{k}}.$$

It remains to introduce the Burkhölder inequalities on which the proof relies.

Lemma 5 (*Burkhölder Inequality [7]*). *Let $(X_k)_{k=1}^n$ be a sequence of complex martingale differences sequence. For every $p \geq 1$, there exists K_p dependent only on p such that:*

$$\mathrm{E}\left[\left| \sum_{k=1}^{n} X_k \right|^{2p} \right] \leq K_p n^p \max_k \mathrm{E}\left[|X_k|^{2p} \right].$$

Letting $X_k = (\mathrm{E}_k - \mathrm{E}_{k-1})\, z_k^* A_k z_k$ where A_k is independent of z_k and noting that $\mathrm{E}\left[|X_k|^{2p}\right] \leq \mathrm{E}\left[\|A_k\|_{\mathrm{Fro}}^{2p}\right]$, with $\|A\|_{\mathrm{Fro}} \triangleq \sqrt{\operatorname{tr} AA^*}$, we get in particular the following corollary (see e.g., [2]).

Lemma 6 (*Burkhölder Inequality for Quadratic Forms*). *Let $z_1, \ldots, z_n \in \mathbb{C}^{N\times 1}$ be n independent random vectors with mean 0 and covariance C_N. Let $\left(A_j\right)_{j=1}^n$ be a sequence of $N \times N$ random matrices where for all j, A_j is independent of z_j. Define X_j as*

$$X_j = \left(\mathrm{E}_j - \mathrm{E}_{j-1}\right) z_j^* A_j z_j = z_j^* \mathrm{E}_j A_j z_j - \operatorname{tr} \mathrm{E}_{j-1} C_N A_j.$$

Then,

$$\mathrm{E}\left[\left| \sum_{j=1}^{n} X_j \right|^{2p} \right] \leq K_p \, \|C_N\|_{\mathrm{Fro}}^{2p} \, n^p \max_j \mathrm{E}\left[\left\|A_j C_N\right\|_{\mathrm{Fro}}^{2p} \right].$$

Preliminaries. We start the proof by some preliminary results.

Lemma 7. *Let $z_1, \ldots, z_n$ be as in Assumption 1. Let $c \in \mathbb{C}^{N\times 1}$ be independent of $z_1, \ldots, z_n$ and such that $\mathrm{E}\|c\|^k$ is bounded uniformly in N for all order k. Then, for any integer p, there exists K_p such that*

$$\mathrm{E}\left[\left| z_i^* \hat{S}_N^{-1} c \right|^p \right] \leq \mathrm{E}\left[\left| z_i^* \hat{S}_{(i)}^{-1} c \right|^p \right] \leq K_p.$$

Proof. The first inequality can be obtained from the following decomposition:

$$\hat{S}_N^{-1} z_i = \frac{\hat{S}_{(i)}^{-1} z_i}{1 + \frac{\alpha(\rho)}{\gamma_N(\rho)} \frac{1}{n} z_i^* \hat{S}_{(i)}^{-1} z_i}$$

while the second follows by noticing that $\mathrm{E}\left|z_i^* c\right|^p \leq \mathrm{E}\left(c^* C_N c\right)^{\frac{p}{2}}$.

Using the same kind of calculations, we can also control the order of magnitude of some interesting quantities.

Lemma 8. *The following statements hold true:*

1. *Denote by $\Delta_{i,j}$ the quantity:*

$$\Delta_{i,j} = \frac{1}{n} z_j^* \hat{S}_{(i,j)}^{-1} z_j - \frac{1}{n} \operatorname{tr} C_N \hat{S}_{(i,j)}^{-1}.$$

   Then, for any $p \geq 2$.

$$\mathrm{E}\left|\Delta_{i,j}\right|^p = O\left(n^{-\frac{p}{2}}\right).$$

2. *Let i and j be two distinct integers from $\{1, \ldots, n\}$. Then,*

$$\mathrm{E}\left| z_i^* \hat{S}_{(i,j)}^{-1} z_j \right|^p = O\left(n^{\frac{p}{2}}\right).$$

3. *Let $z_i \in \mathbb{C}^{N\times 1}$ be as in Assumption 1 and A be a $N\times N$ random matrix independent of z_i and having a bounded spectral norm. Then,*
$$\mathrm{E}\left|z_i^* A z_i\right|^p = O(n^p).$$
4. *Let $j\in\{1,\ldots,n\}$ and i and k two distinct integers different from j. Then:*
$$\mathrm{E}\left|z_i^*\hat{S}_{(i,j)}^{-1}\hat{S}_{(j,k)}^{-1}z_k\right|^p = O\left(n^{\frac{p}{2}}\right).$$

Proof. Items 1. and 3. are standard results that are a by-product of Bai and Silverstein [3, Lemma B.26], while Item 2. can be easily obtained from Lemma 7. As for item 4., it follows by first decomposing $\hat{S}_{(i,j)}^{-1}$ and $\hat{S}_{(j,k)}^{-1}$ as:
$$\hat{S}_{(i,j)}^{-1} = \hat{S}_{(i,j,k)}^{-1} - \frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{\hat{S}_{(i,j,k)}^{-1}z_k z_k^*\hat{S}_{(i,j,k)}^{-1}}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_k^*\hat{S}_{(i,j,k)}^{-1}z_k}$$
$$\hat{S}_{(j,k)}^{-1} = \hat{S}_{(i,j,k)}^{-1} - \frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{\hat{S}_{(i,j,k)}^{-1}z_i z_i^*\hat{S}_{(i,j,k)}^{-1}}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_i^*\hat{S}_{(i,j,k)}^{-1}z_i}.$$
The above relations serve to better control the dependencies of $\hat{S}_{(i,j)}^{-1}$ and $\hat{S}_{(j,k)}^{-1}$ on z_k and z_i. Plugging the above decompositions on $z_i^*\hat{S}_{(i,j)}^{-1}\hat{S}_{(j,k)}^{-1}z_k$, we obtain
$$\begin{aligned}z_i^*\hat{S}_{(i,j)}^{-1}\hat{S}_{(j,k)}^{-1}z_k &= z_i^*\hat{S}_{(i,j,k)}^{-2}z_k - \frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{z_i^*\hat{S}_{(i,j,k)}^{-1}z_k z_k^*\hat{S}_{(i,j,k)}^{-2}z_k}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_k^*\hat{S}_{(i,j,k)}^{-1}z_k} - \frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{z_i^*\hat{S}_{(i,j,k)}^{-2}z_i z_i^*\hat{S}_{(i,j,k)}^{-1}z_k}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_i^*\hat{S}_{(i,j,k)}^{-1}z_i}\\ &\quad+\frac{1}{n^2}\left(\frac{\alpha(\rho)}{\gamma_N(\rho)}\right)^2\frac{z_i^*\hat{S}_{(i,j,k)}^{-1}z_k z_k^*\hat{S}_{(i,j,k)}^{-2}z_i z_i^*\hat{S}_{(i,j,k)}^{-1}z_k}{\left(1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_k^*\hat{S}_{(i,j,k)}^{-1}z_k\right)\left(1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_i^*\hat{S}_{(i,j,k)}^{-1}z_i\right)}.\end{aligned}$$
The control of these four terms follows from a direct application of item 2. and 3. along with possibly the use of the generalized Hölder inequality in Lemma 4.

Core of the proof. With these preliminary results at hand, we are now in position to get into the core of the proof. Let β_N be given by
$$\beta_N = \frac{1}{n}\sum_{i=1}^n c^*\hat{S}_N^{-1}z_i z_i^*\hat{S}_N^{-1}d\left(\frac{1}{N}z_i^*\hat{S}_{(i)}^{-1}z_i - \gamma_N(\rho)\right).$$
Decompose β_N as
$$\begin{aligned}\beta_N &= \frac{1}{n}\sum_{i=1}^n c^*\hat{S}_N^{-1}z_i z_i^*\hat{S}_N^{-1}d\left(\frac{1}{N}z_i^*\hat{S}_{(i)}^{-1}z_i - \frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i)}^{-1}\right) + \frac{1}{n}\sum_{i=1}^n c^*\hat{S}_N^{-1}z_i z_i^*\hat{S}_N^{-1}d\left(\frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i)}^{-1} - \gamma_N(\rho)\right)\\ &\triangleq \beta_{N,1}+\beta_{N,2}.\end{aligned}$$
The control of $\beta_{N,2}$ follows from a direct application of Lemmas 3 and 4, that is
$$\begin{aligned}\mathrm{E}\left[\left|\beta_{N,2}\right|^{2p}\right] &\le \frac{n^{2p-1}}{n^{2p}}\sum_{i=1}^n \mathrm{E}\left|c^*\hat{S}_N^{-1}z_i\right|^{2p}\left|z_i^*\hat{S}_N^{-1}d\right|^{2p}\left|\frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i)}^{-1} - \gamma_N(\rho)\right|^{2p}\\ &\le \frac{n^{2p-1}}{n^{2p}}\sum_{i=1}^n\left(\mathrm{E}\left|c^*\hat{S}_N^{-1}z_i\right|^{6p}\right)^{\frac{1}{3}}\left(\mathrm{E}\left|z_i^*\hat{S}_N^{-1}d\right|^{6p}\right)^{\frac{1}{3}}\left(\mathrm{E}\left|\frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i)}^{-1} - \gamma_N(\rho)\right|^{6p}\right)^{\frac{1}{3}}.\end{aligned}$$
By standard results from random matrix theory (e.g. [26, Prop. 7.1]), we know that
$$\mathrm{E}\left|\frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i)}^{-1} - \gamma_N(\rho)\right|^{6p} = O(n^{-6p}).$$
Hence, by Lemma 7, we finally get:
$$\mathrm{E}\left|\beta_{N,2}\right|^{2p} = O(n^{-2p}).$$
While the control of $\beta_{N,2}$ requires only the manipulation of conventional moment bounds due to the rapid convergence of $\frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i)}^{-1} - \gamma_N(\rho)$, the analysis of $\beta_{N,1}$ is more intricate since
$$\mathrm{E}\left|\frac{1}{N}z_i^*\hat{S}_{(i)}^{-1}z_i - \frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i)}^{-1}\right|^p = O\left(n^{-\frac{p}{2}}\right)$$

a convergence rate which seems insufficient at the onset. The averaging occurring in $\beta_{N,2}$ shall play the role of improving this rate. To control $\beta_{N,1}$, one needs to resort to advanced tools based on Burkhölder inequalities. First, decompose $\beta_{N,1}$ as

$$\beta_{N,1} = \overset{o}{\beta}_{N,1} + \mathrm{E}\left[\beta_{N,1}\right].$$

As in Lemma 8, define $\Delta_i \triangleq \frac{1}{n} z_i^* \hat{S}_{(i)}^{-1} z_i - \frac{1}{n}\operatorname{tr} C_N \hat{S}_{(i)}^{-1}$. Using the relation

$$\hat{S}_N^{-1} z_i = \frac{\hat{S}_{(i)}^{-1} z_i}{1 + \frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)} z_i^* \hat{S}_{(i)}^{-1} z_i}$$

we get

$$\begin{aligned}
\mathrm{E}\left[\beta_{N,1}\right] &= \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^n \frac{c^*\hat{S}_{(i)}^{-1} z_i z_i^* \hat{S}_{(i)}^{-1} d}{\left(1 + \frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)} z_i^* \hat{S}_{(i)}^{-1} z_i\right)^2}\Delta_i\right] \\
&= \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^n \frac{c^*\hat{S}_{(i)}^{-1} z_i z_i^* \hat{S}_{(i)}^{-1} d}{\left(1 + \frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)} \operatorname{tr} C_N \hat{S}_{(i)}^{-1}\right)^2}\Delta_i\right] \\
&\quad - \frac{\alpha(\rho)}{\gamma_N(\rho)}\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^n \frac{c^*\hat{S}_{(i)}^{-1} z_i z_i^* \hat{S}_{(i)}^{-1} d \Delta_i^2 \left(2 + \left(\frac{\alpha(\rho)}{\gamma_N(\rho)}\right)\left(\frac{1}{n} z_i^* \hat{S}_{(i)}^{-1} z_i + \frac{1}{n}\operatorname{tr} C_N \hat{S}_{(i)}^{-1}\right)\right)}{\left(1 + \frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr} C_N \hat{S}_{(i)}^{-1}\right)^2\left(1 + \frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{1}{n} z_i^* \hat{S}_{(i)}^{-1} z_i\right)^2}\right] \\
&\triangleq \beta_{N,1,1} + \beta_{N,1,2}.
\end{aligned}$$

Since $\mathrm{E}\left[w^* A w\left(w^* B w - \operatorname{tr} B\right)\right] = \mathrm{E}\operatorname{tr} AB$ when w is standard complex Gaussian vector and A, B random matrices independent of w, we have

$$\mathrm{E}\left[\beta_{N,1,1}\right] = \frac{1}{Nn}\mathrm{E}\left[\operatorname{tr}\frac{C_N \hat{S}_{(i)}^{-1} C_N \hat{S}_{(i)}^{-1} d c^* \hat{S}_{(i)}^{-1}}{\left(1 + \frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{1}{n}\operatorname{tr} C_N \hat{S}_{(i)}^{-1}\right)^2}\right] = O(n^{-1}).$$

As for $\beta_{N,1,2}$, we have for some $K > 0$, again by Lemma 8

$$\begin{aligned}
\left|\beta_{N,1,2}\right| &\le \frac{K}{n}\sum_{i=1}^n \left(\mathrm{E}\left|c^*\hat{S}_{(i)}^{-1} z_i\right|^4\right)^{\frac{1}{4}}\left(\mathrm{E}\left|z_i^*\hat{S}_{(i)}^{-1} d\right|^4\right)^{\frac{1}{4}}\left(\mathrm{E}\left|\Delta_i\right|^8\right)^{\frac{1}{4}}\left(\mathrm{E}\left|2 + \frac{\alpha(\rho)}{\gamma_N(\rho)}\left(\frac{1}{n} z_i^*\hat{S}_{(i)}^{-1} z_i + \frac{1}{n}\operatorname{tr} C_N \hat{S}_{(i)}^{-1}\right)\right|^4\right)^{\frac{1}{4}} \\
&= O\left(\frac{1}{n}\right).
\end{aligned}$$

We therefore have

$$\left|\mathrm{E}\left[\beta_{N,1}\right]\right|^{2p} = O(n^{-2p}).$$

Let us turn to the control of $\overset{o}{\beta}_{N,1}$. For that, we decompose $\overset{o}{\beta}_{N,1}$ as a sum of martingale differences as

$$\overset{o}{\beta}_{N,1} = \sum_{j=1}^n \left(\mathrm{E}_j - \mathrm{E}_{j-1}\right)\beta_{N,1}.$$

The control of $\mathrm{E}\left[\left|\overset{o}{\beta}_{N,1}\right|^p\right]$ requires the convergence rate of two kinds of martingale differences:

- Sum of martingale differences with a quadratic form representation of the form

$$\sum_{j=1}^n \left(\mathrm{E}_j - \mathrm{E}_{j-1}\right) z_j^* A_j z_j.$$

  For these terms, from Lemma 6, it will be sufficient to show that $\max_j \mathrm{E}\|A_j\|_{\mathrm{Fro}}^{2p} = O(n^{-3p})$ in order to obtain the required convergence rate.

- Sum of martingale differences with more than one occurrence of z_j and z_j^*. In this case, this sum is given by:

$$\sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1}) \sum_{i=1, i\neq j}^n \varepsilon_i$$

where ε_j are small random quantities depending on $z_1, \dots, z_n$. According to Lemma 5, we have

$$\mathrm{E}\left|\sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1}) \sum_{i=1, i\neq j}^n \varepsilon_i\right|^{2p} = O(n^{-2p})$$

provided that

$$\mathrm{E}\left|\sum_{i=1, i\neq j} \varepsilon_i\right|^{2p} = O(n^{-3p}).$$

The control of the above sum will rely on successively using Lemma 3 to get

$$\mathrm{E}\left|\sum_{i=1, i\neq j} \varepsilon_i\right|^{2p} \leq n^{2p-1} \sum_{i=1}^n \mathrm{E}\,|\varepsilon_i|^{2p}$$

and controlling $\max_i \mathrm{E}\,|\varepsilon_i|^{2p}$.

With this explanation at hand, we will now get into the core of the proofs. We first have

$$\begin{aligned}
\overset{o}{\beta}_{N,1} &= \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1}) \frac{1}{N} \sum_{i=1}^n c^* \hat{S}_N^{-1} z_i z_i^* d \Delta_i \\
&= \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1}) c^* \hat{S}_N^{-1} z_j z_j^* d \Delta_j + \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1}) \frac{1}{N} \sum_{i=1, i\neq j}^n c^* \hat{S}_N^{-1} z_i z_i^* d \Delta_i \\
&\triangleq \sum_{j=1}^n W_{j,1} + \sum_{j=1}^n W_{j,2}.
\end{aligned}$$

In order to prove that $\mathrm{E}\left|\sum_{j=1}^n W_{j,1}\right| = O(n^{-2p})$, it is sufficient to show

$$\mathrm{E}\left|W_{j,1}\right| = O(n^{-3p})$$

a statement which holds true since, by Lemma 4

$$\begin{aligned}
\mathrm{E}\left|W_{j,1}\right|^{2p} &\leq \frac{K}{n^{2p}} \mathrm{E}\left|c^* \hat{S}_N^{-1} z_j\right|^{2p} \left|z_j^* \hat{S}_N^{-1} d\right|^{2p} \Delta_j^{2p} \\
&\leq \frac{K}{n^{2p}} \left(\mathrm{E}\left|c^* \hat{S}_N^{-1} z_j\right|^{6p}\right)^{\frac{1}{3}} \left(\mathrm{E}\left|z_j^* \hat{S}_N^{-1} d\right|^{6p}\right)^{\frac{1}{3}} \left(\mathrm{E}\Delta_j^{6p}\right)^{\frac{1}{3}} \\
&= O(n^{-3p}).
\end{aligned}$$

We now consider the more involved term $\sum_{j=1}^n W_{j,2}$. Using the relation

$$\hat{S}_N^{-1} = \hat{S}_{(j)}^{-1} - \frac{\alpha(\rho)}{\gamma_N(\rho)} \frac{1}{n} \frac{\hat{S}_{(j)}^{-1} z_j z_j^* \hat{S}_{(j)}^{-1}}{1 + \frac{\alpha(\rho)}{\gamma_N(\rho)} \frac{1}{n} z_j^* \hat{S}_{(j)}^{-1} z_j}$$

to let the independent $\hat{S}_{(j)}^{-1}$ and z_j variables appear, we write

$$\begin{aligned}
\sum_{j=1}^n W_{j,2} &= \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1}) \frac{1}{n} \sum_{i=1, i\neq j}^n c^* \hat{S}_{(j)}^{-1} z_i z_i^* \hat{S}_{(j)}^{-1} d \left(\frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i - \frac{1}{N} \operatorname{tr} C_N \hat{S}_{(i)}^{-1}\right) \\
&\quad - \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1}) \frac{\alpha(\rho)}{\gamma_N(\rho)} \frac{1}{n^2} \sum_{i=1, i\neq j}^n \frac{c^* \hat{S}_{(j)}^{-1} z_j z_j^* \hat{S}_{(j)}^{-1} z_i z_i^* \hat{S}_{(j)}^{-1} d}{1 + \frac{1}{n} \frac{\alpha(\rho)}{\gamma_N(\rho)} z_j^* \hat{S}_{(j)}^{-1} z_j} \left(\frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i - \frac{1}{N} \operatorname{tr} C_N \hat{S}_{(i)}^{-1}\right) \\
&\quad - \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1}) \frac{\alpha(\rho)}{\gamma_N(\rho)} \frac{1}{n^2} \sum_{i=1, i\neq j}^n \frac{c^* \hat{S}_{(j)}^{-1} z_i z_i^* \hat{S}_{(j)}^{-1} z_j z_j^* \hat{S}_{(j)}^{-1} d}{1 + \frac{1}{n} \frac{\alpha(\rho)}{\gamma_N(\rho)} z_j^* \hat{S}_{(j)}^{-1} z_j} \left(\frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i - \frac{1}{N} \operatorname{tr} C_N \hat{S}_{(i)}^{-1}\right)
\end{aligned}$$

$$+\sum_{j=1}^{n}(\mathrm{E}_j-\mathrm{E}_{j-1})\left(\frac{\alpha(\rho)}{\gamma_N(\rho)}\right)^2\frac{1}{n^3}\sum_{i=1,i\neq j}^{n}\frac{c^*\hat{S}_{(j)}^{-1}z_jz_j^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}z_jz_j^*\hat{S}_{(j)}^{-1}d}{\left(1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_j^*\hat{S}_{(j)}^{-1}z_j\right)^2}\left(\frac{1}{N}z_i^*\hat{S}_{(i)}^{-1}z_i-\frac{1}{N}\operatorname{tr}C_N\hat{S}_{(i)}^{-1}\right)$$

$$\triangleq \chi_1+\chi_2+\chi_3+\chi_4.$$

Next, we will sequentially control $\chi_i, i=1,\ldots,4$.

Control of χ_1. Using the relation

$$\hat{S}_{(i)}^{-1}=\hat{S}_{(i,j)}^{-1}-\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{1}{n}\frac{\hat{S}_{(i,j)}^{-1}z_jz_j^*\hat{S}_{(i,j)}^{-1}}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_j^*\hat{S}_{(i,j)}^{-1}z_j}$$

the quantity χ_1 can be decomposed as

$$\chi_1=\sum_{j=1}^{n}-(\mathrm{E}_j-\mathrm{E}_{j-1})\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{1}{n^2N}\sum_{i=1,i\neq j}^{n}\frac{c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}d\left|z_i^*\hat{S}_{(i,j)}^{-1}z_j\right|^2}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_j^*\hat{S}_{(i,j)}^{-1}z_j}$$
$$+\sum_{j=1}^{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}(\mathrm{E}_j-\mathrm{E}_{j-1})\frac{1}{n^2N}\sum_{i=1,i\neq j}^{n}\frac{c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}dz_j^*\hat{S}_{(i,j)}^{-1}C_N\hat{S}_{(i,j)}^{-1}z_j}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_j^*\hat{S}_{(i,j)}^{-1}z_j}$$
$$\triangleq \chi_{1,1}+\chi_{1,2}$$

where we used the fact that for r_j random quantity independent of z_j, $(\mathrm{E}_j-\mathrm{E}_{j-1})(r_j)=0$. We will begin by controlling $\chi_{1,1}$. To handle the quadratic forms in the denominator, we further develop $\chi_{1,1}$ as

$$\chi_{1,1}=-\sum_{j=1}^{n}(\mathrm{E}_j-\mathrm{E}_{j-1})\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{1}{n^2N}\sum_{i=1,i\neq j}^{n}\frac{c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}d\left|z_i^*\hat{S}_{(i,j)}^{-1}z_j\right|^2}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}C_N\hat{S}_{(i,j)}^{-1}}$$
$$+\sum_{j=1}^{n}(\mathrm{E}_j-\mathrm{E}_{j-1})\left(\frac{\alpha(\rho)}{\gamma_N(\rho)}\right)^2\frac{1}{n^2N}\sum_{i=1,i\neq j}^{n}\frac{c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}\left|z_i^*\hat{S}_{(i,j)}^{-1}z_j\right|^2\Delta_{i,j}}{\left(1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}C_N\hat{S}_{(i,j)}^{-1}\right)\left(1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}z_j^*\hat{S}_{(i,j)}^{-1}z_j\right)}$$
$$=\sum_{j=1}^{n}X_{j,1}+\sum_{j=1}^{n}X_{j,2}.$$

To control $\sum_{j=1}^{n}X_{j,1}$, we resort to Lemma 6. Indeed, $X_{j,1}$ can be written as

$$X_{j,1}=-\frac{\alpha(\rho)}{\gamma_N(\rho)}(\mathrm{E}_j-\mathrm{E}_{j-1})z_j^*A_jz_j$$

where A_j is given by

$$A_j=\frac{1}{n^2N}\sum_{i=1,i\neq j}^{n}\frac{c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}d}{1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}C_N\hat{S}_{(i,j)}^{-1}}\hat{S}_{(i,j)}^{-1}z_iz_i^*\hat{S}_{(i,j)}^{-1}.$$

According to Lemma 6, it is sufficient to prove that $\mathrm{E}\left\|A_j\right\|_{\mathrm{Fro}}^{2p}=O(n^{-3p})$. Expanding $\mathrm{E}\left\|A_j\right\|_{\mathrm{Fro}}^{2p}$, we indeed get

$$\mathrm{E}\left\|A_j\right\|_{\mathrm{Fro}}^{2p}\leq\frac{K}{n^{6p}}\mathrm{E}\left|\sum_{i\neq j}\sum_{k\neq j}\frac{\left|z_k^*\hat{S}_{(j,k)}^{-1}\hat{S}_{(i,j)}^{-1}z_i\right|^2c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}dd^*\hat{S}_{(j)}^{-1}z_kz_k^*\hat{S}_{(j)}^{-1}c}{\left(1+\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{1}{n}\operatorname{tr}C_N\hat{S}_{(i,j)}^{-1}\right)\left(1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}C_N\hat{S}_{(j,k)}^{-1}\right)}\right|^p$$
$$\leq\frac{K}{n^{6p}}\mathrm{E}\left|\sum_{i\neq j}\left|z_i^*\hat{S}_{(i,j)}^{-2}z_i\right|^2\left|c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}d\right|^2\right|^p$$
$$+\frac{K}{n^{6p}}\mathrm{E}\left|\sum_{i\neq j}\sum_{\substack{k\neq j\\k\neq i}}\frac{\left|z_k^*\hat{S}_{(j,k)}^{-1}\hat{S}_{(i,j)}^{-1}z_i\right|^2c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}dd^*\hat{S}_{(j)}^{-1}z_kz_k^*\hat{S}_{(j)}^{-1}c}{\left(1+\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{1}{n}\operatorname{tr}C_N\hat{S}_{(i,j)}^{-1}\right)\left(1+\frac{1}{n}\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}C_N\hat{S}_{(j,k)}^{-1}\right)}\right|^p$$
$$\leq\frac{Kn^{p-1}}{n^{6p}}\mathrm{E}\left|z_i^*\hat{S}_{(i,j)}^{-2}z_i\right|^{2p}\left|c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}d\right|^{2p}$$

$$
\begin{aligned}
&+\frac{Kn^{2(p-1)}}{n^{6p}}\sum_{i\neq j}\sum_{\substack{k\neq j\\k\neq i}}\mathrm{E}\left|z_k^*\hat{S}_{(j,k)}^{-1}\hat{S}_{(i,j)}^{-1}z_i\right|^{2p}\left|c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}d\right|^{p}\left|d^*\hat{S}_{(j,k)}^{-1}z_kz_k^*\hat{S}_{(j,k)}^{-1}c\right|^{p}\\
&\leq \frac{Kn^{p-1}}{n^{6p}}\sum_{i\neq j}\left(\mathrm{E}\left|z_i^*\hat{S}_{(i,j)}^{-2}z_i\right|^{6p}\right)^{\frac13}\left(\left|c^*\hat{S}_{(j)}^{-1}z_i\right|^{6p}\right)^{\frac13}\left(\left|z_i^*\hat{S}_{(j)}^{-1}d\right|^{6p}\right)^{\frac13}\\
&+\frac{Kn^{2(p-1)}}{n^{6p}}\sum_{i\neq j}\sum_{\substack{k\neq j\\k\neq i}}\left(\mathrm{E}\left|z_k^*\hat{S}_{(j,k)}^{-1}\hat{S}_{(i,j)}^{-1}z_i\right|^{10p}\right)^{\frac15}\left(\mathrm{E}\left|c^*\hat{S}_{(i,j)}^{-1}z_i\right|^{5p}\right)^{\frac15}\\
&\times\left(\mathrm{E}\left|z_i^*\hat{S}_{(j)}^{-1}d\right|^{5p}\right)^{\frac15}\left(\mathrm{E}\left|d^*\hat{S}_{(j,k)}^{-1}z_k\right|^{5p}\right)^{\frac15}\left(\mathrm{E}\left|z_k^*\hat{S}_{(j,k)}^{-1}c\right|^{5p}\right)^{\frac15}\\
&=O(n^{-3p}).
\end{aligned}
$$

As for $X_{j,1}$, we can show that $\mathrm{E}\left|X_{j,1}\right|^{2p}=O(n^{-3p})$. Indeed, we have

$$
\begin{aligned}
\mathrm{E}\left|X_{j,2}\right|^{2p}&\leq\frac{Kn^{2p-1}}{n^{6p}}\sum_{i\neq j}\left(\mathrm{E}\left|c^*\hat{S}_{(j)}^{-1}z_i\right|^{8p}\right)^{\frac14}\left(\mathrm{E}\left|z_i^*\hat{S}_{(j)}^{-1}d\right|^{8p}\right)^{\frac14}\left(\mathrm{E}\left|z_i^*\hat{S}_{(j)}^{-1}z_j\right|^{16p}\right)^{\frac14}\left(\mathrm{E}\left|\Delta_{i,j}\right|^{8p}\right)^{\frac14}\\
&=O(n^{-3p}).
\end{aligned}
$$

The Burkhölder inequality shows that this rate of convergence of the moment of $X_{j,1}$ and $X_{j,2}$ is sufficient to finally ensure that $\mathrm{E}\left|\chi_{1,1}\right|^{2p}=O(n^{-2p})$.

We study next $\chi_{1,2}$. First, decompose $\chi_{1,2}$ as

$$
\begin{aligned}
\chi_{1,2}&=\sum_{j=1}^n(\mathrm{E}_j-\mathrm{E}_{j-1})\frac{1}{n^2N}\sum_{i\neq j}\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}dz_j^*\hat{S}_{(i,j)}^{-1}C_N\hat{S}_{(i,j)}^{-1}z_j}{1+\frac1n\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}C_N\hat{S}_{(i,j)}^{-1}}\\
&-\sum_{j=1}^n(\mathrm{E}_j-\mathrm{E}_{j-1})\frac{1}{n^2N}\sum_{i\neq j}\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac{c^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}d\Delta_{i,j}z_j^*\hat{S}_{(i,j)}^{-1}C_N\hat{S}_{(i,j)}^{-1}z_j}{\left(1+\frac{\alpha(\rho)}{\gamma_N(\rho)}\frac1nz_j^*\hat{S}_{(i,j)}^{-1}z_j\right)\left(1+\frac1n\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}C_N\hat{S}_{(i,j)}^{-1}\right)}\\
&\triangleq\sum_{j=1}^nY_{j,1}+\sum_{j=1}^nY_{j,2}.
\end{aligned}
$$

The quantities $\sum_{j=1}^nY_{j,1}$ and $\sum_{j=1}^nY_{j,2}$ are differences of martingales whose controls follow the same procedure as above. While $\sum_{j=1}^nY_{j,1}$ can be controlled using Lemma 6, the convergence of $\sum_{j=1}^nY_{j,2}$ is faster due to the term $\Delta_{i,j}$. Details are thus omitted.

Control of χ_2. The control of χ_2 cannot be exactly dealt with using the same procedure. As for χ_1, one works out χ_2 by substituting $\frac1nz_j^*\hat{S}_{(j)}^{-1}z_j$ by its approximate $\frac1n\operatorname{tr}C_N\hat{S}_{(j)}^{-1}$ and using the decomposition of $\hat{S}_{(i)}^{-1}$ as a function of $\hat{S}_{(i,j)}^{-1}$ to get

$$
\chi_2=-\frac{\alpha(\rho)}{\gamma_N(\rho)}\sum_{j=1}^n(\mathrm{E}_j-\mathrm{E}_{j-1})\frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}\frac{c^*\hat{S}_{(j)}^{-1}z_jz_j^*\hat{S}_{(j)}^{-1}z_iz_i^*\hat{S}_{(j)}^{-1}d\left(\frac1Nz_i^*\hat{S}_{(i,j)}^{-1}z_i-\frac1N\operatorname{tr}C_N\hat{S}_{(i,j)}^{-1}\right)}{1+\frac1n\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}C_N\hat{S}_{(j)}^{-1}}+\varepsilon
$$

where we easily obtain that $\mathrm{E}[|\varepsilon|^{2p}]=O(n^{-2p})$. We omit the details of this step, since the calculations are the same as those used for the control of χ_1. The control of the Frobenius norm of the underlying matrices using the same techniques as above does not yield the required convergence rate. We will thus pursue a different approach. Precisely, we write χ_2 as

$$
\chi_2=-\frac{\alpha(\rho)}{\gamma_N(\rho)}\sum_{j=1}^n(\mathrm{E}_j-\mathrm{E}_{j-1})T_j+\varepsilon
$$

with

$$
T_j=\frac{1}{n^2}\frac{c^*\hat{S}_{(j)}^{-1}z_jz_j^*\hat{S}_{(j)}^{-1}Z_jD_jZ_j^*\hat{S}_{(j)}^{-1}d}{1+\frac1n\frac{\alpha(\rho)}{\gamma_N(\rho)}\operatorname{tr}\hat{S}_{(j)}^{-1}}
$$

where $Z_j=\left[z_1,\ldots,z_{j-1},z_{j+1},\ldots,z_n\right]$ and D_j is a diagonal matrix with diagonal elements: $\left[D_j\right]_{i,i}=\frac nN\Delta_{j,i}$. Hence, by Lemma 5

$$
\begin{aligned}
\mathrm{E}\left|T_j\right|^{2p}&\leq\frac{1}{n^{4p}}\mathrm{E}\left|c^*\hat{S}_{(j)}^{-1}z_j\right|^{2p}\left|z_j^*\hat{S}_{(j)}^{-1}Z_jD_jZ_j^*\hat{S}_{(j)}^{-1}d\right|^{2p}\\
&\leq\frac{1}{n^{4p}}\left(\mathrm{E}\left|c^*\hat{S}_{(j)}^{-1}z_j\right|^{4p}\right)^{\frac12}\left(\mathrm{E}\left|z_j^*\hat{S}_{(j)}^{-1}Z_jD_jZ_j^*d\right|^{4p}\right)^{\frac12}.
\end{aligned}
$$

Since D_j is independent of z_j, applying the inequality $\mathrm{E}\left|z_j^* u\right|^p \leq \mathrm{E}\left(u^* C_N u\right)^{\frac{p}{2}}$, we finally get

$$\begin{aligned}
\mathrm{E}\left|T_j\right|^{2p} &\leq \frac{K}{n^{4p}}\left(\mathrm{E}\left|d^*\hat{S}_{(j)}^{-1}Z_j D_j Z_j^*\hat{S}_{(j)}^{-1}C_N\hat{S}_{(j)}^{-1}Z_j D_j Z_j^*\hat{S}_{(j)}^{-1}d\right|^{2p}\right)^{\frac{1}{2}} \\
&= \frac{K}{n^{3p}}\left(\mathrm{E}\left|d^*\hat{S}_{(j)}^{-1}Z_j D_j\frac{Z_j^*\hat{S}_{(j)}^{-1}C_N\hat{S}_{(j)}^{-1}Z_j}{n}D_j Z_j^*\hat{S}_{(j)}^{-1}d\right|^{2p}\right)^{\frac{1}{2}} \\
&\overset{(a)}{\leq} \frac{K}{n^{3p}}\left(\mathrm{E}\left\|D_j Z_j^*\hat{S}_{(j)}^{-1}d\right\|^{4p}\right)^{\frac{1}{2}}
\end{aligned}$$

where (a) follows since $\left\|\frac{Z_j^*\hat{S}_{(j)}^{-1}C_N\hat{S}_{(j)}^{-1}Z_j}{n}\right\|$ is bounded. In order to prove that $\mathrm{E}\left|T_j\right|^{2p} = O(n^{-3p})$, it suffices to check that $\mathrm{E}\left\|D_j Z_j^*\hat{S}_{(j)}^{-1}d\right\|^{4p}$ is uniformly bounded in N. Expanding this quantity, we indeed get

$$\begin{aligned}
\mathrm{E}\left\|D_j Z_j^*\hat{S}_{(j)}^{-1}d\right\|^{4p} &= \mathrm{E}\left|\sum_{\substack{i=1\\ i\neq j}}^{n}\left(\frac{1}{N}z_i^*\hat{S}_{(i,j)}^{-1}z_i - \frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i,j)}^{-1}\right)^2\left|z_i^*\hat{S}_{(j)}^{-1}d\right|^2\right|^{2p} \\
&\leq n^{2p-1}\sum_{i=1}^{n}\mathrm{E}\left(\frac{1}{N}z_i^*\hat{S}_{(i,j)}^{-1}z_i - \frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i,j)}^{-1}\right)^{4p}\left|z_i^*\hat{S}_{(j)}^{-1}d\right|^{4p} \\
&\leq n^{2p-1}\sum_{i=1}^{n}\left(\mathrm{E}\left(\frac{1}{N}z_i^*\hat{S}_{(i,j)}^{-1}z_i - \frac{1}{N}\operatorname{tr} C_N\hat{S}_{(i,j)}^{-1}\right)^{8p}\right)^{\frac{1}{2}}\left(\mathrm{E}\left|z_i^*\hat{S}_{(j)}^{-1}d\right|^{8p}\right)^{\frac{1}{2}} \\
&= O(1).
\end{aligned}$$

The control of χ_3 is similar to that of χ_2, while that of χ_4 follows immediately by using sequentially Lemma 3 along with the generalized Hölder inequality in Lemma 4. This completes the proof.

References

[1] Y. Abramovich, O. Besson, Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 1: The over-sampled case, IEEE Trans. Signal Process. 61 (23) (2013) 5807–5818.
[2] Z.D. Bai, J.W. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices, Ann. Probab. 26 (1) (1998) 316–345.
[3] Z.D. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, second ed., in: Springer Series in Statistics, Springer, 2009, New York, NY, USA.
[4] O. Besson, Y. Abramovich, Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 2: The under-sampled case, IEEE Trans. Signal Process. 61 (23) (2013) 5819–5829.
[5] P. Billingsley, Convergence of Probability Measures, John Wiley and Sons, Inc., Hoboken, NJ, 1968.
[6] S.P. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[7] D.L. Burkhölder, Distribution function inequalities for martingales, Ann. Probab. 1 (1) (1973) 19–41.
[8] F. Chapon, R. Couillet, W. Hachem, X. Mestre, The outliers among the singular values of large rectangular random matrices with additive fixed rank deformation, Markov Process. Related Fields 20 (2014) 183–228.
[9] Y. Chen, A. Wiesel, A.O. Hero, Robust shrinkage estimation of high-dimensional covariance matrices, IEEE Trans. Signal Process. 59 (9) (2011) 4097–4107.
[10] E. Conte, M. Lops, G. Ricci, Asymptotically optimum radar detection in compound-gaussian clutter, IEEE Trans. Aerosp. Electron. Syst. 31 (2) (1995) 617–625.
[11] R. Couillet, Robust spiked random matrices and a robust G-MUSIC estimator, J. Multivariate Anal. 140 (2015) 139–151.
[12] R. Couillet, W. Hachem, Analysis of the limit spectral measure of large random matrices of the separable covariance type, Random Matrix Theory Appl. 3 (4) (2014) 1–23.
[13] R. Couillet, M. McKay, Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators, J. Multivariate Anal. 131 (2014) 99–120.
[14] R. Couillet, F. Pascal, J.W. Silverstein, Robust estimates of covariance matrices in the large dimensional regime, IEEE Trans. Inform. Theory 60 (11) (2014) 7269–7278.
[15] R. Couillet, F. Pascal, J.W. Silverstein, The random matrix regime of Maronna's M-estimator with elliptically distributed samples, J. Multivariate Anal. 139 (2015) 56–78.
[16] R.M. Dudley, Real Analysis and Probability, Vol. 74, Cambridge University Press, 2002.
[17] N. El Karoui, Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results, 2013. ArXiv Preprint arXiv:1311.2445.
[18] W. Hachem, O. Khorunzhy, P. Loubaton, J. Najim, L.A. Pastur, A new approach for capacity analysis of large dimensional multi-antenna channels, IEEE Trans. Inform. Theory 54 (9) (2008) 3987–4004.
[19] O. Kallenberg, Foundations of Modern Probability, Springer, 2002.
[20] A. Kammoun, R. Couillet, F. Pascal, M.-S. Alouini, Optimal design of the adaptive normalized matched filter detector, 2015. ArXiv Preprint arXiv:1501.06027.
[21] A. Kammoun, M. Kharouf, W. Hachem, J. Najim, A central limit theorem for the sinr at the lmmse estimator output for large-dimensional signals, IEEE Trans. Inform. Theory 55 (11) (2009) 5048–5063.
[22] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal. 88 (2) (2004) 365–411.

[23] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues for some sets of random matrices, Math. USSR-Sb. 1 (4) (1967) 457–483.
[24] R.A. Maronna, Robust M-estimators of multivariate location and scatter, Ann. Statist. 4 (1976) 51–67.
[25] X. Mestre, On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices, IEEE Trans. Signal Process. 56 (11) (2008) 5353–5368.
[26] J. Najim, J.F. Yao, Gaussian fluctuations for linear spectral statistics of large random covariance matrices, 2013. ArXiv Preprint arXiv:1309.3728.
[27] F. Pascal, Y. Chitour, Y. Quek, Generalized robust shrinkage estimator and its application to STAP detection problem, IEEE Trans. Signal Process. 62 (21) (2014) 5640–5651.
[28] J.W. Silverstein, Z.D. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal. 54 (2) (1995) 175–192.
[29] J.W. Silverstein, S. Choi, Analysis of the limiting spectral distribution of large dimensional random matrices, J. Multivariate Anal. 54 (2) (1995) 295–309.
[30] D.E. Tyler, A distribution-free M-estimator of multivariate scatter, Ann. Statist. 15 (1) (1987) 234–251.
[31] L. Yang, R. Couillet, M. McKay, Minimum variance portfolio optimization with robust shrinkage covariance estimation, in: Proc. IEEE Asilomar Conference on Signals, Systems, and Computers. Pacific Grove, CA, USA, 2014.
[32] T. Zhang, X. Cheng, A. Singer, Marchenko–Pastur law for Tyler's and Maronna's M-estimators, 2014. http://arxiv.org/abs/1401.3424.